

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

TESIS DE GRADO

**“DATAWAREHOUSING Y SU APLICACIÓN COMO APOYO A LA
TOMA DE DECISIONES EN LA ESPOL”**

Previa a la obtención del Título de:

**INGENIERO EN COMPUTACIÓN ESPECIALIZACIÓN
SISTEMAS TECNOLÓGICOS**

Presentada por:

Juan Carlos Bustamante Collaguazo

José Francisco Rodríguez Rojas

GUAYAQUIL – ECUADOR

2004 -2005

AGRADECIMIENTO

A Dios por su guía amorosa, a la ESPOLE por formarnos, al Ing. Fabricio Echeverría por su invaluable dirección, a nuestros compañeros del CSI por su desinteresada colaboración y a nuestros amigos por su indispensable apoyo. Gracias por hacer de esos momentos un verdadero vivir.

DEDICATORIA

Esta tesis es una parte de mi vida y comienzo de otras, por haber sido mi fuente de inspiración, la dedico con todo mi amor y cariño a Leticia, Carlos, Lalo, Xavico y Cafú, y a todos aquellos que me han devuelto una sonrisa, a todos aquellos que me ofrecieron su apoyo en tiempos difíciles, a todos aquellos que han puesto de su parte para que el trajín diario sea más llevadero y muy en especial a DIOS que me ha dado tanto.....

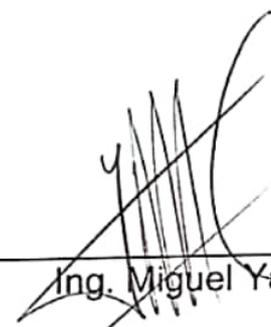
Juan C. Bustamante

DEDICATORIA

A mis padres José y Mary por su confianza, apoyo y comprensión, a mis hermanos Nidia, Priscilla y Marlon por su tierna compañía, a Adriana por su apoyo incondicional y a mis abuelos Francisco⁺ y Napoleón⁺ que desde el cielo siempre iluminan mi camino.

José F. Rodríguez

TRIBUNAL DE GRADUACIÓN



Ing. Miguel Yapur
SUB-DECANO DE LA FIEC



Ing. Fabricio Echeverría
DIRECTOR DE TESIS



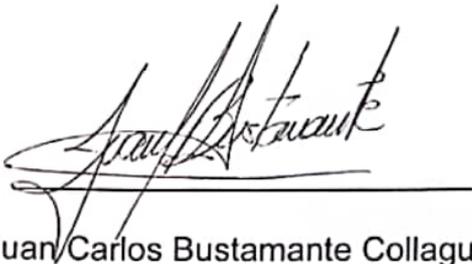
Ing. Otilia Alejandro
MIEMBRO DEL TRIBUNAL



Ing. Galo Valverde
MIEMBRO DEL TRIBUNAL

DECLARACIÓN EXPRESA

“La responsabilidad del contenido de esta Tesis de Grado, nos corresponde exclusivamente; y el patrimonio intelectual de la misma a la Escuela Superior Politécnica del Litoral”



Juan Carlos Bustamante Collaguazo



José Francisco Rodríguez Rojas

“Construir castillos en el aire no es perder el tiempo. Ahí es en donde se construyen.

Luego, ponedle los cimientos.”

Henry David Thoreau

RESUMEN

El proyecto a exponerse, diseña e implementa un Datawarehouse y su aplicación como apoyo a la toma de decisiones en la ESPOL. El sistema incluye componentes de ubicación de fuentes de datos, extracción, limpieza, montado, sumarización y consultas de cubos de información. Los diversos módulos están diseñados para poder ser ejecutados de manera independiente; los componentes de administración y control como una aplicación bajo Windows cliente-servidor y el componente de reportes y consultas por intermedio de un navegador web.

Vale mencionar que esta implementación corresponde a una plataforma de información que está en capacidad de ser accesada a través de herramientas de terceros o propietarias.

En el primer capítulo, se revisan todos los conceptos y definiciones que introducen a las ventajas del uso de un Datawarehouse. Se centra principalmente en aspectos organizacionales y culturales, causales de fracaso, criterios de éxito, imperativos culturales y demás aspectos de su organización.

En el segundo capítulo, se realiza un análisis detallado de los justificativos, metas y objetivos, alcance, asunciones y limitantes, beneficios y costos que deben ser considerados en una implementación de datawarehouse para una institución como la ESPOL.

En el tercer capítulo, se presenta la metodología de implementación del datawarehouse que abarca aspectos tales como identificación de etapas, determinación de roles y responsabilidades y tiempos para culminación de tareas.

El cuarto capítulo contiene las características del diseño e implementación del datawarehouse identificando aspectos tales como definición de la arquitectura, necesidades de información, análisis del área objetivo, estudio de las fuentes de datos, diseño de las transformaciones, base de datos física, acceso de usuarios finales y la selección de las herramientas para el desarrollo del proyecto.

En el quinto capítulo, se presenta el desarrollo del repositorio, las implementaciones de interfaces de administración y consulta, dimensiones y modelos del datawarehouse, procesos de carga, poblado y control de los datos, así como las seguridades implementadas y el hardware recomendado.

En el sexto capítulo, se presentan las conclusiones del trabajo y las recomendaciones apropiadas para el desarrollo del mismo.

Por último, en el séptimo capítulo se desarrollan los anexos y apéndices que abarcan documentos informativos tales como diagramas de trabajo, directorio de contacto del equipo, evaluación de riesgo, impacto del proyecto y un glosario de términos.

ÍNDICE GENERAL

RESUMEN.....	VII
INDICE GENERAL	X
INDICE DE FIGURAS	XVII
INDICE DE TABLAS	XX
1. ANTECEDENTES	1
1.1. ¿Qué es Datawarehousing?.....	2
1.1.1. Definiciones de Datawarehouse	5
1.1.2. Extracción de Información Estratégica	7
1.1.3. Requerimientos de Procesamiento Analítico.....	9
1.1.4. Superalmacén de Datos e Información	14
1.1.5. Tecnología de Datawarehouse	20
1.2. Renacimiento para la Tecnología de Información	23
1.2.1. La Nueva Concepción y sus Visionarios.....	27
1.2.2. Estableciendo una Infraestructura	32
1.2.3. Repositorios de Datos	38
1.2.4. Una Estrategia para la Construcción de Datawarehouse.....	42
1.3. Aspectos Organizacionales y Culturales	51

1.3.1. Datawarehousing del Mundo Real: El Cambio en la Administración.....	51
1.3.2. Aspectos de la Administración.....	53
1.3.3. Situación Actual.....	55
1.3.4. Imperativos Culturales	58
1.3.5. Criterios de Éxito	59
1.3.6. Causales de Fracaso.....	62
1.3.7. Roles y Responsabilidades en la Organización.....	73
1.3.8. Determinación de Áreas y Datos Iniciales	81
1.3.9. Cambios en la Estructura de Poder en la Organización.....	83
1.3.10. Aspectos a Resolver	88
2. JUSTIFICACIÓN DEL DATAWAREHOUSE	89
2.1. Metas y Objetivos	90
2.1.1. Metas y Objetivos del Negocio.....	90
2.1.2. Metas y Objetivos del Proyecto	92
2.2. Alcance	94
2.2.1. Definición del Alcance	96
2.2.2. Productos del Proyecto – Lista de Entregables	97
2.2.3. Hitos del Proyecto	99
2.2.4. Áreas del Negocio involucradas	101
2.3. Asunciones	102
2.3.1. Asunciones del Proyecto	102

2.4.	Limitaciones	104
2.4.1.	Limitantes del Proyecto	104
2.4.2.	Proyectos Relacionados	106
2.4.3.	Dependencias Críticas	109
2.5.	Propuesta de Valor Comercial del Datawarehousing	110
2.5.1.	Retorno de Inversión	110
2.5.1.1.	Beneficios Tangibles	113
2.5.1.2.	Beneficios Intangibles	115
2.5.2.	Presupuesto del Datawarehouse	120
2.5.2.1.	Evaluación de Costos de Tecnologías	121
2.5.2.2.	Estimación de Costos de Recursos	123
3.	METODOLOGIA DE IMPLEMENTACION DEL DATAWAREHOUSE	126
3.1.	Metodología de Desarrollo	127
3.1.1.	Identificación de etapas del proyecto	127
3.1.2.	Determinar objetivos temporales	137
3.1.3.	Determinar tareas y responsabilidades	138
3.1.4.	Asignar tiempos para culminación de tareas	154
4.	ANALISIS Y DISEÑO DEL DATAWAREHOUSE	155
4.1.	Iniciación del Proyecto	156
4.1.1.	Alcance de Datos	156
4.1.2.	Alcance Temporal	158
4.1.3.	Razones Organizacionales	160

4.1.4.	Enfoque Completo del DW	163
4.1.5.	Participantes y sus Roles	164
4.1.6.	Asunciones y Limitantes	166
4.1.7.	Estrategia de Administración del Proyecto	169
4.2.	Definición de la Arquitectura.....	172
4.2.1.	Arquitectura de las Aplicaciones.....	174
4.2.2.	Dimensionamiento inicial	175
4.2.3.	Entorno Técnico Definido.....	176
4.2.4.	Entorno Técnico Preparado.....	177
4.3.	Necesidades de los Tomadores de Decisiones.....	178
4.3.1.	Requerimientos de Información	178
4.3.2.	Requerimientos Organizacionales.....	184
4.3.3.	Requerimientos de Acceso.....	187
4.3.4.	Prototipos.....	189
4.4.	Análisis del Área Objetivo	191
4.4.1.	Alcance del Área Objetivo	191
4.4.2.	Nivel de Detalle Requerido	192
4.4.3.	Tablas de Resumen Preliminares	194
4.4.4.	Modelo de Datos del DW	195
4.5.	Análisis de los Sistemas de Legado (Fuentes de Información).....	202
4.5.1.	Fuentes Candidatas.....	202
4.5.2.	Problemas Operacionales y de Integridad.....	205

4.5.3.	Evaluación	208
4.5.4.	Nuevos Atributos	210
4.5.5.	Requerimientos de la Metadata	211
4.5.6.	Disponibilidad Histórica.....	215
4.5.7.	Requerimientos de Control	217
4.6.	Diseño de las Transformaciones	219
4.6.1.	Especificaciones de las Transformaciones	221
4.6.2.	Procesos de Transformación	223
4.6.3.	Diseño de Controles	226
4.6.4.	Procesos de Sumarización	227
4.6.5.	Procesos de Conversión Histórica	228
4.6.6.	Comprobación de Datos	230
4.7.	Diseño de la Base de Datos Física	232
4.7.1.	Tablas de Dominios	234
4.7.2.	Tablas de Sumarización.....	235
4.7.3.	Tablas de Hechos y Esquema Estrella	236
4.7.4.	Índices	243
4.7.5.	Políticas de Respaldo y Recuperación.....	245
4.8.	Herramientas de los Usuarios Finales	247
4.8.1.	Diseño de Accesos	247
4.8.2.	Definición de Accesos	249
4.8.3.	Desarrollo de Accesos.....	251

4.9.	Elección de las Herramientas de Desarrollo	253
4.9.1.	Posibles Herramientas para el desarrollo del proyecto	255
4.9.1.1.	Hardware	255
4.9.1.2.	Software de Almacenamiento (SGBD)	256
4.9.1.3.	Software de extracción y manipulación de datos.....	258
4.9.2.	Criterios para la elección de las herramientas	261
4.9.3.	Justificación para la elección de las Herramientas seleccionadas	
	264	
5.	IMPLEMENTACION DEL DATAWAREHOUSE.....	267
5.1.	Desarrollo del Repositorio.....	268
5.1.1.	Desarrollo de la Transformación.....	268
5.1.2.	Carga de Datos Preliminar	269
5.1.3.	Procesos de Control y Auditoría	271
5.1.4.	Carga de la Metadata	272
5.1.5.	Pruebas del Sistema.....	274
5.2.	Implementación del Proyecto	276
5.2.1.	Metodología Implementada	276
5.2.2.	Implementación de los modelos del negocio.....	280
5.2.3.	Implementación de las interfaces	282
5.2.4.	Implementación de las dimensiones en la Base de Datos	296
5.3.	Poblado del Datawarehouse	306
5.3.1.	Corrida del Poblado del Repositorio	306

5.3.2.	Implementación del Software en Usuarios del Sistema.....	308
5.3.3.	Entrenamiento de Acceso y de Datos.....	309
5.3.4.	Prueba de Aceptación de los Usuarios del Sistema.....	311
5.3.5.	Control y Verificación Auditada	313
5.4.	Seguridad Implementada	314
5.5.	Hardware recomendado.....	316
6.	CONCLUSIONES Y RECOMENDACIONES	318
6.1.	Conclusiones.....	319
6.2.	Recomendaciones	322
7.	ANEXOS Y APÉNDICES.....	324
7.1.	Apéndice A - Plan de Trabajo del Proyecto	325
7.2.	Apéndice B - Gráfico del Proyecto Network-PERT	326
7.3.	Apéndice C - Directorio de Contacto del Equipo del Proyecto	327
7.4.	Apéndice F - Reporte de Impacto del Proyecto	328
7.5.	Apéndice G – Glosario de Términos	334
	ANEXOS.....	344
	Anexo 1-A.....	345
	Anexo 1-B.....	348
	Anexo 1-C	350
	Anexo 1-D	353
	REFERENCIAS BIBLIOGRÁFICAS.....	354

ÍNDICE DE FIGURAS

Figura 1: Integración en un Datawarehouse, primera característica ...	15
Figura 2 : Orientación a un tema, segunda característica de un Datawarehouse	17
Figura 3 : Característica histórica en un Datawarehouse	18
Figura 4 : Comparativo de operaciones en un Datawarehouse	19
Figura 5 : Esquema Bidimensional Alcance vs. Redundancia de Datos	109
Figura 6: Arquitectura Típica de un Datawarehouse	174
Figura 7: Pirámide de Tipos de Información	178
Figura 8: Modelo de Datos de Registros Académicos	197
Figura 9: Modelo de Datos de Graduados Académicos	199
Figura 10: Modelo de Datos de Personal	201
Figura 11: Muestra del modelo de datos de la Metadata	214
Figura 12: Ejemplo de Transformación.....	221
Figura 13: Vista del Log de ejecución de Procesos del datawarehouse	231
Figura 14: Vista de alto nivel de un modelo Entidad-Relación	236
Figura 15: Modelo detallado multidimensional para un punto de ventas.	239
Figura 16: Tipos de almacenamiento	246

Figura 17: Metodología Genérica para el Desarrollo de un Datawarehouse	277
Figura 18: Vista del Control Center de DB2. Herramienta de administración de la base de datos del EDW	284
Figura 19: Administración de Usuarios a través del EDW-Security...	285
Figura 20: Administración de Perfiles del EDW-Security.....	286
Figura 21: Interfaz de Control de Operaciones del EDW-Security.....	286
Figura 22: Vista del componente EDW-OLAP	287
Figura 23: Interfaz de conexión de datos con el EDW-OLAP	288
Figura 24: Vista del Journal Center de DB2. Ejecutor de tareas de limpieza, exportación y cargado de datos para el EDW	289
Figura 25: Vista del Scheduler. Programador de ejecución del EDW	290
Figura 26: Interfaz de consulta del EDW	291
Figura 27: Muestra de Pivoting del EDW-Consult.....	292
Figura 28: Muestra de Slicing Dicing del EDW-Consult	293
Figura 29: Muestra de Roll-up del EDW-Consult	294
Figura 30: Muestra de Drill down del EDW-Consult	295
Figura 31: Modelo de consultas de Registros	297
Figura 32: Reporte de Registros por Campus, Carrera, Año y Sectores	298
Figura 33: Modelo de consultas de Graduados.....	299

Figura 34: Reporte de Graduados y Egresados Completo	300
Figura 35: Modelo de consulta de Estudiantes	301
Figura 36: Reporte de Estudiantes por Ingreso y Datos Personales.	302
Figura 37: Modelo de consultas de Profesores	302

ÍNDICE DE TABLAS

Tabla 1: Comparativo entre tecnologías de Almacenamiento	22
Tabla 2: Funciones de Gerencia	75
Tabla 3: Funciones de Elaboración de Requerimientos	77
Tabla 4: Funciones de Desarrollo de Base de Datos	79
Tabla 5: Funciones de Desarrollo de la Aplicación.....	80
Tabla 6: Estimado de Costos para la Implementación del Datawarehouse	122
Tabla 7: Inversión de Recursos para un Datawarehouse.....	125
Tabla 8: Entorno Técnico Definido en la ESPOL	176
Tabla 9: Entorno Técnico Preparado para el Datawarehouse	177
Tabla 10: Fuentes Candidatas del Datawarehouse de la ESPOL	203
Tabla 11: Otras Fuentes de Datos de la ESPOL	204
Tabla 12: Procesos de transformación de datos	225
Tabla 13: Detalle de dimensiones del datawarehouse.....	241
Tabla 14: Jerarquías de las dimensiones	242
Tabla 15: Preguntas planteadas para elegir la herramienta adecuada	263
Tabla 16: Modelo de Negocio del Datawarehouse de la ESPOL	281
Tabla 17: Actividades realizadas en la migración de datos.....	305

CAPÍTULO I

ANTECEDENTES

1.1. ¿Qué es Datawarehousing?

El concepto de Datawarehousing ha evolucionado desde sus orígenes, a principios de los años ochenta. Antes de mostrar una definición aceptable, se listará algunas de las definiciones más usuales:

- ✓ “Un datawarehousing es un repositorio de datos históricos referidos a un tema en particular”, Inmon ¹(1992).
- ✓ “Un datawarehousing es una colección de datamarts² más pequeños”, Dyché³,(2001).
- ✓ “Cualquier plataforma de hardware separada – ya sea un servidor o una PC – que permita a un hombre de negocios tomar una decisión puede ser considerado un datawarehousing”, Kimball ⁴(2002).
- ✓ “Un datawarehouse es un almacén de información integrada, disponible para las consultas y análisis. Se extraen datos e información de las fuentes heterogéneas como ellos se

¹ Bill Inmon es uno de los creadores y gurús de la tecnología de Datawarehousing al establecer los criterios y bases para la implementación y gestión de éstos

² Datamarts son repositorios de datos más pequeños orientados a un tema o departamento específico

³ Jill Dyché es socia del Baseline Consulting Group Inc., una compañía especializada en la creación de soluciones de inteligencia de negocios adaptadas a diversas industrias. Autora del libro e-data: Transformando los en información con Datawarehousing.

⁴ Ralph Kimball fue co-inventor de la Xerox Star Workstation, el primer producto que utilizó ratón, iconos y ventanas. Fue vicepresidente de aplicaciones de la Metaphor Computer Systems y es fundador y CEO de Red Brick Systems. Actualmente trabajo como consultor independiente en el diseño de grandes Datawarehouses. Usted puede saber más acerca de Kimball a través de su página web <http://www.ralphkimball.com>

generan.... Esto hace que la corrida de consultas sean más fáciles y más eficaces sobre los datos que originalmente vinieron de fuentes diferentes”, según la Universidad de Stanford⁵.

Según el entorno del que se trate, todas estas definiciones pueden ser correctas.

Olvidando las definiciones complejas sobre datos históricos, granularidad⁶ temporal, y servidores interconectados, se detallan los cuatro principales principios de un datawarehousing, según Inmon(1996). Los mismos son válidos independientemente de la plataforma, de la cantidad de datos, y del software utilizado, y deberían bastar para que el usuario común haga su propia idea de lo que es un datawarehouse.

1.- “Un datawarehouse es normalmente – pero no necesariamente – una computadora o plataforma de hardware aislada”. Esta plataforma puede ser grande (un servidor) o pequeña (una estación de trabajo). En algunos casos puede llegar a ser una colección de plataformas distribuidas. En otros,

⁵ Más información en <http://www.stanford.edu>

⁶ Corresponde al nivel de atomicidad elegido para una tabla de hechos o entidad

podría residir en una serie de nodos de una plataforma informática mayor.

2.- “Los datos de un datawarehouse se usan para la toma de decisiones”. Esta es la principal naturaleza de un proyecto de datawarehouse.

3.- “Los datawarehouse duplican datos que ya están en otra parte de la empresa”. Aunque esta redundancia parezca un desperdicio, en realidad está bien.

4.- “Un datawarehouse no es algo más que una computadora ubicada en algún lugar en las entrañas del centro de cómputo de la empresa. Es la combinación de hardware, software especializado y datos”. Normalmente, cuando alguien dice “nuestro datawarehouse”, en realidad está hablando acerca de un equipo de hardware, una colección de productos y herramientas de software, y montones de datos.

En resumen, un datawarehouse es un repositorio de información extraído de otros sistemas corporativos – sean éstos sistemas transaccionales, bases de datos departamentales, o la Intranet de la compañía – a la que los hombres de negocios de la empresa pueden acceder.

1.1.1. Definiciones de Datawarehouse

El datawarehouse, es actualmente, el centro de atención de las grandes instituciones, porque provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.

Inmon (1992), considerado como el “padre del datawarehousing”, indica que “un datawarehouse es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales”.

Según Kimball (1996), “un datawarehouse es una colección de datos en la cual se encuentra integrada la información de una Institución y que se usa como soporte para el proceso de toma de decisiones gerenciales”.

SOFTGAL (2000)⁷ afirma que Datawarehouse es un proceso, no es un producto. “Nosotros siempre decimos que estos

⁷ SOFTGAL es una empresa de consultoría y desarrollo de Tecnologías de la Información. Véase www.softgal.com

sistemas de información no se pueden comprar, hay que construirlos, pero en su diseño y construcción es necesario emplear una metodología adecuada”.

La genialidad del datawarehouse es crear un entorno con una arquitectura bien definida moviendo la toma de decisiones (la salida del proceso) a este nuevo entorno, sin tener que reescribir la entrada (transacciones).

Su valor inmediato es satisfacer la demanda de los gestores en cuanto a tener una visión integrada de la empresa y su entorno.

1.1.2. Extracción de Información Estratégica

Existen funciones dentro de la empresa que tienen que ver con el planeamiento, previsión y administración de la organización. Estas funciones son también críticas para la supervivencia de la organización ya que son vitales para la toma de decisiones en un entorno comercial de rápidos cambios.

Las funciones como "planificación de marketing", "planeamiento de ingeniería" y "análisis financiero", requieren de sistemas de información especializados que los soporten. Pero estas funciones son diferentes a las de los sistemas operacionales y/o transaccionales, por lo que los tipos de sistemas y la información requerida son también diferentes.

Mientras las necesidades de los datos operacionales se enfocan normalmente hacia una sola área, los datos para el soporte de decisiones, con frecuencia, toma un número de áreas diferentes y necesita cantidades grandes de datos operacionales relacionadas y procesadas.

Son estos sistemas en los que se basa la tecnología datawarehousing, orientados a soportar la toma de decisiones, facilitan la labor de la dirección, proporcionándole un soporte básico, en forma de mejor información, para la toma de decisiones. Según Dyché (2001), se caracterizan porque son sistemas sin carga periódica de trabajo, es decir, su utilización no es predecible, al contrario de los casos anteriores, cuya utilización es periódica.

Dyché (2001) destaca entre estos sistemas: los Sistemas de Información Gerencial (MIS), Sistemas de Información Ejecutivos (EIS), Sistemas de Información Georeferencial (GIS), Sistemas de Simulación de Negocios (BIS y que en la práctica son sistemas expertos o de Inteligencia Artificial-AI).

Todo esto da lugar a comprender la importancia y variedad de uso que puede llegar a tener un datawarehouse como plataforma de otros sistemas informáticos orientados a negocios.

1.1.3. *Requerimientos de Procesamiento Analítico*

En la actualidad, las tecnologías de la información han automatizado los procesos de carácter típicamente operativo o administrativo, haciendo uso de lo que se conoce como sistemas de información operacionales. Se entiende por aplicaciones operacionales, aquellas que resuelven las necesidades de funcionamiento de la empresa. En este tipo de sistemas, los conceptos más importantes son la actualización y el tiempo de respuesta, según Inmon (1992).

Los procesos que son automatizados una vez implementados, tienen que tener indicadores o índices que permitan medir su gestión con respecto a los líderes de la organización, por lo que estos permitirán conocer cómo se encuentra el (los) proceso(s) en un momento determinado.

Una vez satisfechas las necesidades operacionales más urgentes, surge un nuevo grupo de necesidades sobre los sistemas de la empresa, a las cuales se calificarán como

necesidades de información. Por necesidades de información, se entiende a aquellas que tienen por objeto obtener la información necesaria, que sirva de base para la toma de decisiones tanto a escala estratégica como táctica. Estas necesidades de información se basan en gran medida en el análisis de un número enorme de datos, en el que es tan importante obtener un valor muy detallado de negocio como el valor totalizado para el mismo.

Es fundamental también la visión histórica de todas las variables analizadas, y el análisis de los datos del entorno. Estos requerimientos no son, a priori, difíciles de resolver dado que la información está efectivamente en los sistemas operacionales. Cualquier actividad que realiza la empresa está reflejada de forma minuciosa en sus bases de datos. La realidad, sin embargo, es distinta, puesto que al atender las necesidades de tipo informacional, los responsables de sistemas se tropiezan con múltiples problemas.

En primer lugar, al realizar consultas masivas de información (con el fin de conseguir el ratio⁸, valor agrupado o grupo de valores solicitados), se puede ver perjudicado el nivel de servicio del resto de sistemas, dado que las consultas de las que se habla, suelen ser bastante costosas en recursos. Además, las necesidades se ven insatisfechas por la limitada flexibilidad a la hora de navegar por la información y a su inconsistencia debido a la falta de una visión global (cada visión particular del dato está almacenada en el sistema operacional que lo gestiona).

En esta situación, Inmon (1992) afirma “el siguiente paso evolutivo ha venido siendo la generación de un entorno gemelo del operativo, que se ha denominado comúnmente Centro de Información, en el cual la información se refresca con menor periodicidad que en los entornos operacionales y los requerimientos en el nivel de servicio al usuario son más flexibles”.

⁸ Un ratio expresa numéricamente la relación que existe entre una magnitud y otra de la misma clase.

Con esta estrategia se resuelve el problema de la planificación de recursos ya que las aplicaciones que precisan un nivel de servicio alto usan el entorno operacional y las que precisan consultas masivas de información trabajan en el Centro de Información. Otro beneficio de este nuevo entorno, es la no interferencia con las aplicaciones operacionales.

Pero no terminan aquí los problemas de accesibilidad, inflexibilidad, inconsistencia y nivel de servicio. La información mantiene la misma estructura que en las aplicaciones operacionales o transaccionales por lo que las operaciones que se realizan son de tipo consultas y deben de acceder a diferentes lugares para obtener el conjunto de datos que formarán la información. El tiempo de respuesta a las solicitudes de información tiende a ser excesivamente elevado debido a que estas consultas deben agrupar grandes cantidades de datos y realizar procesos de sumarización⁹, filtrado, ordenamiento y agrupamiento. Adicionalmente, al proceder la información de distintos sistemas, con visiones distintas y distintos objetivos, en muchas ocasiones no es

⁹ Sumarización es el proceso mediante el cual se obtienen subtotaes a partir de procesamiento de registros individuales.

posible obtener la información deseada de una forma fácil y además carece de la necesaria fiabilidad.

De cara al usuario estos problemas se traducen en que no dispone a tiempo de la información solicitada y que debe dedicarse con más intensidad a la obtención de la información que al análisis de la misma, que es donde aporta mayor valor agregado.

1.1.4. Superalmacén de Datos e Información

Tras las dificultades de los sistemas tradicionales en satisfacer las necesidades informacionales, surge el concepto de Datawarehouse, como “solución a las necesidades informacionales globales de la empresa”. Este término acuñado por Inmon (1992), se traduce literalmente como Almacén de Datos en español. No obstante si el Datawarehouse fuese exclusivamente un almacén de datos, los problemas seguirían siendo los mismos que en los Centros de Información.

La ventaja principal de este tipo de sistemas se basa en su concepto fundamental, la estructura de la información, según Inmon (1992). Este concepto significa el almacenamiento de información homogénea y fiable, en una estructura basada en la consulta y el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales. Según definió Inmon (1992), el Datawarehouse se caracteriza por ser:

Integrado:

“Los datos almacenados en el Datawarehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas”. La información suele estructurarse en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

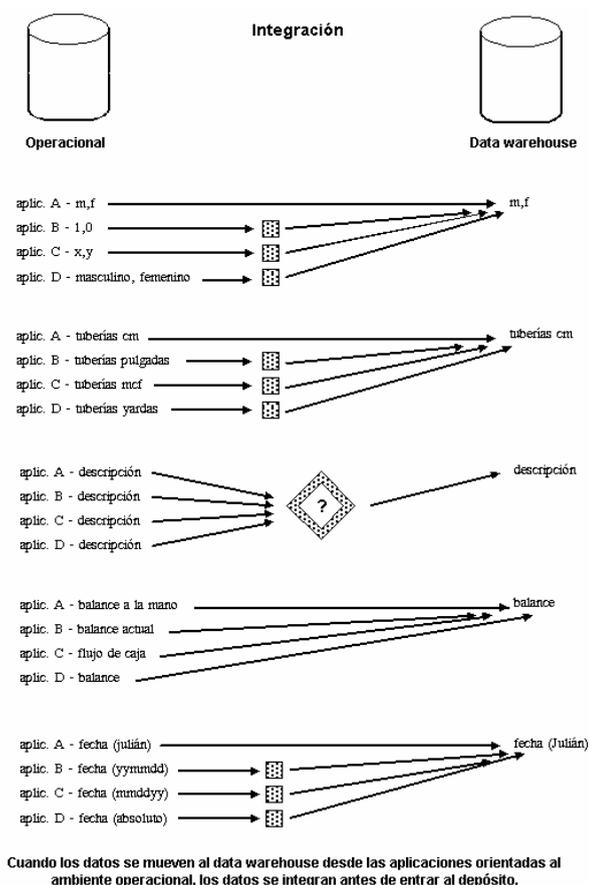


Figura 1: Integración en un Datawarehouse, primera característica

Fuente: <http://www.inei.gob.pe/cpi -mapa/bancopub/libfree/lib619/132.HTM>

Temático:

“Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional y transaccional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales”. Por ejemplo, todos los datos sobre estudiantes pueden ser consolidados en una única tabla del Datawarehouse. De esta forma, las peticiones de información sobre estudiantes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

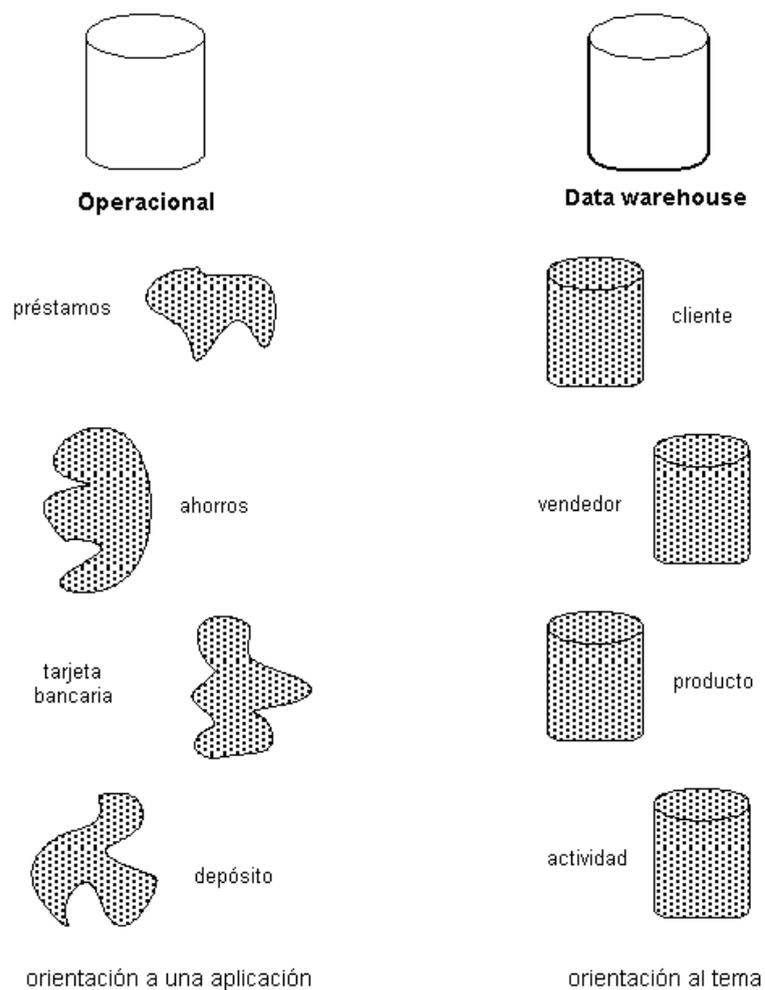


Figura 2 : Orientación a un tema, segunda característica de un Datawarehouse

Fuente: <http://www.inei.gob.pe/cpi -mapa/bancopub/libfree/lib619/132.HTM>

Histórico:

“El tiempo es parte implícita de la información contenida en un Datawarehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Datawarehouse sirve, entre otras cosas, para realizar análisis de tendencias”. Por lo tanto, un datawarehouse se carga con los distintos valores que toma una variable(o indicador) en el tiempo para permitir comparaciones.

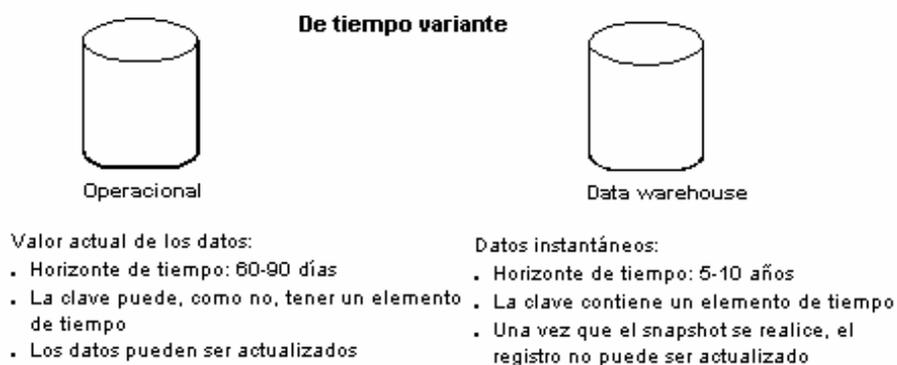


Figura 3 : Característica histórica en un Datawarehouse

Fuente: <http://www.inei.gob.pe/cpi -mapa/bancopub/libfree/lib619/132.HTM>

No volátil:

“El almacén de información de un Datawarehouse existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del Datawarehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía”. Todos los datos son cargados en el repositorio y accedido desde ahí, pero una vez que la confirmación esta hecha los datos en el depósito no cambian. Normalmente su actualización, inserción o reemplazo se realiza registro por registro.

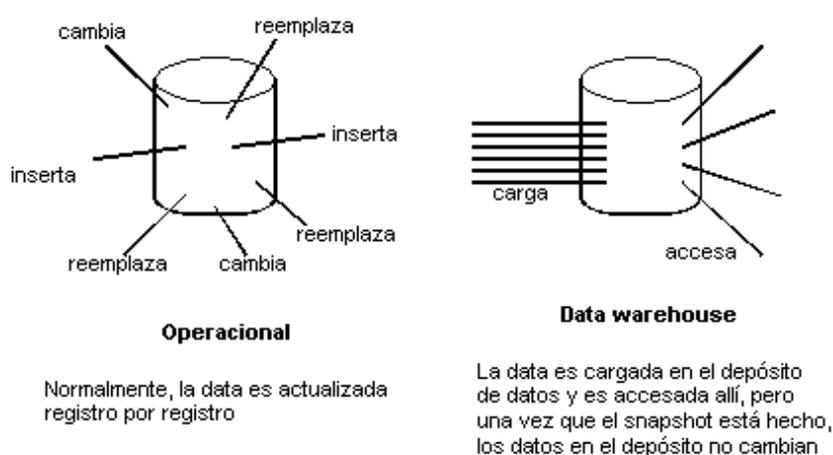


Figura 4 : Comparativo de operaciones en un Datawarehouse

Fuente: <http://www.inei.gob.pe/cpi -mapa/bancopub/libfree/lib619/132.HTM>

1.1.5. Tecnología de Datawarehouse

En el entorno informático, especialmente los usuarios finales, se considera a un datawarehouse como una herramienta de software más que como un componente primordial de la infraestructura tecnológica de la empresa. Vale recalcar que el datawarehouse abarca hardware, software, redes y datos. Según Dyché (2001) distingue dos tipos básicos de datawarehouse: “el datawarehouse empresarial que usan múltiples organizaciones a lo largo de la compañía para diferentes propósitos; y el datamart que por lo general es un datawarehouse asignado a una función única y usado por una sola organización”.

La arquitectura de un datawarehouse empresarial “se caracteriza por el tamaño de la plataforma de hardware, la cantidad de sistemas de origen que lo alimentan y la cantidad de usuarios”, según Dyché (2001).

A la inversa, la arquitectura de un datamart puede variar según su uso específico. En el caso de un datamart dependiente, el datamart está conectado a un datawarehouse empresarial y

cumple una función comercial específica usando un subconjunto de datos tomados del datawarehouse.

El datamart independiente “es un sistema aislado”, según Dyché (2001). Sus datos pueden provenir de sistemas operacionales, como los de un datawarehouse empresarial, pero también pueden ser generados o, en algunos casos, hasta ingresados manualmente.

“Datawarehousing es el centro de la arquitectura para los sistemas de información en la década de los '90”, según INEI (1997). Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico, informático sobre una amplia perspectiva de tiempo.

Se puede caracterizar un datawarehouse haciendo un contraste de cómo los datos de un negocio almacenados en un datawarehouse, difieren de los datos operacionales usados por las aplicaciones de producción.

Base de Datos Operacional	Datawarehouse
Datos Operacionales	Datos del negocio para Información
Orientado a la aplicación	Orientado al sujeto
Actual	Actual + histórico
Detallada	Detallada + más resumida
Cambia continuamente	Estable

Tabla 1: Comparativo entre tecnologías de Almacenamiento

Fuente: <http://www.inei.gob.pe/cpi -mapa/bancopub/libfree/lib619/132.HTM>

1.2. Renacimiento para la Tecnología de Información

Desde que se inició la era de la computación, las organizaciones han usado los datos desde los sistemas operacionales para atender las necesidades de información. Algunas proporcionan acceso directo a la información contenida dentro de las aplicaciones operacionales. Otras, han extraído los datos desde sus bases de datos operacionales para combinarlos de varias formas no estructuradas¹⁰, en su intento por atender a los usuarios en sus necesidades de información.

Ambos métodos han evolucionado a través del tiempo y ahora las organizaciones manejan una data no limpia e inconsistente, sobre las cuales, en la mayoría de las veces, se toman decisiones importantes.

La gestión administrativa reconoce que una manera de elevar su eficiencia está en hacer el mejor uso de los recursos de información que ya existen dentro de la organización. Sin embargo, a pesar de que esto se viene intentando desde hace muchos años, no se tiene todavía un uso efectivo de los mismos.

¹⁰ Formas no estructuradas: Se refiere a formas que no guardan una jerarquía o un orden específico. Un ejemplo clásico de una forma no estructurada es el pensamiento humano.

La razón principal es la manera en que han evolucionado las computadoras, basadas en las tecnologías de información y sistemas. Inmon (1992) afirma que “la mayoría de las organizaciones hacen lo posible por conseguir buena información, pero el logro de ese objetivo depende fundamentalmente de su arquitectura actual, tanto de hardware como de software”.

El datawarehouse, es actualmente, el centro de atención de las grandes instituciones, porque provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.

Aunque diversas organizaciones y personas individuales logran comprender el enfoque de un datawarehouse, la experiencia ha demostrado que existen muchas dificultades potenciales de carácter cultural, organizacional y operativo, que serán explicadas en la sección 1.3.6.

Reunir los elementos de datos apropiados desde diversas fuentes de aplicación en un ambiente integral centralizado, simplifica el problema de acceso a la información y en consecuencia, acelera el

proceso de análisis, consultas y el menor tiempo de uso de la información.

Las aplicaciones para soporte de decisiones basadas en un datawarehousing, pueden hacer más práctica y fácil la explotación de datos para una mayor eficacia del negocio, que no se logra cuando se usan sólo los datos que provienen de las aplicaciones operacionales o transaccionales (que ayudan en la operación de la empresa en sus operaciones cotidianas), en los que la información se obtiene realizando procesos independientes y muchas veces complejos.

Un datawarehouse se crea al extraer datos desde una o más bases de datos de aplicaciones operacionales o transaccionales. La data extraída es transformada para eliminar inconsistencias y resumir si es necesario y luego, cargadas en el datawarehouse. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información. Este nuevo enfoque ayuda a las personas individuales, en todos los niveles de la empresa, a efectuar su toma de decisiones con más responsabilidad.

La innovación de la Tecnología de Información dentro de un ambiente datawarehousing, puede permitir a cualquier organización hacer un uso más óptimo de los datos, como un ingrediente clave para un proceso de toma de decisiones más efectivo. Las organizaciones tienen que aprovechar sus recursos de información para crear la información de la operación del negocio, pero deben considerarse las estrategias tecnológicas necesarias para la implementación de una arquitectura completa de datawarehouse.

1.2.1. La Nueva Concepción y sus Visionarios

El datawarehousing se expandió dramáticamente a fines de los años ochenta, cuando las empresas comenzaron a darse cuenta del valor de sus datos, según Dyché (2001). En esos tiempos la atención se centraba en cómo poder integrar un datawarehouse en la tecnología existente de una compañía, más que en su valor comercial.

En el pasado, según Dyché (2001) las motivaciones para el datawarehousing se podían hallar en un número de factores separados:

- ✓ *Aliviar la carga de los mainframes¹¹. Frecuentemente, el mainframe de una compañía ya estaba suficientemente ocupado, y hacer actualizaciones – agregar más hardware y poder de procesamiento – podía ser tan caro como comprar un datawarehouse nuevo. Estos mainframes solían ser sistemas transaccionales asignados a funciones críticas como facturación o depósitos bancarios. Las compañías*

¹¹ Mainframes son grandes equipos de computación provistos de capacidad de procesamiento y almacenamiento adecuado y suficiente para la gestión de usuarios y terminales.

eran renuentes a exigir más de sus mainframes, y la redundancia de datos y el procesamiento separado que proveían los datawarehouses eran tentadores.

- ✓ *Datos sucios. Los diferentes datos de toda la empresa – desde cuentas por pagar hasta cálculo de comisiones o información sobre los clientes – no solamente solía ser difíciles de hallar, sino imposibles de entender, y a menudo directamente erróneos. El datawarehouse brindaba la posibilidad de una única plataforma para cargar estos datos “heterogéneos”. También ofrece un pretexto para depurar los datos – controlar su calidad – y reformatearlos de modo que fueran comprensibles y útiles antes de ser cargados en el repositorio.*

 - ✓ *Seguridad. Proteger sus sistemas operacionales críticos. La mayor parte de las compañías limitaba su acceso a algunos pocos expertos. En vez de poner los datos de esos sistemas a disposición de la totalidad de la empresa, los expertos se los daban a cuentagotas a aquellos que los requerían cuando se*
-

volvían necesarios. Los datawarehouses ofrecieron un entorno para que la información frecuentemente buscada fuera accesible de forma más general.

Antes del datawarehouse, compañías minoristas como Wall-Mart¹² y bancos como el Citibank¹³ no podían comprar computadoras suficientemente grandes para llevar el control de sus inventarios. Las compañías de teléfono no tenían donde guardar los números de todos sus clientes en un solo lugar. Los hospitales no tenían dónde poner los registros de todos sus pacientes. Los datawarehouses mostraron tanto brillantes ventajas funcionales, así como hacer técnicamente posible que una empresa pueda almacenar grandes cantidades de información dispar en un mismo lugar. El descenso de los precios de las computadoras lo hicieron progresivamente más factible.

A medida que las compañías pioneras comenzaron a hacer uso del soporte de decisión y la computación se hacía más barata, el valor comercial del datawarehouse quedó más claro, y pronto

¹² WallMart empresa Norteamericana con sede en varios países del mundo dedicada al comercio.

¹³ Citibank: Banco internacional con operaciones comerciales a nivel mundial. Maneja un importante volumen de datos de sus transacciones.

no hizo falta que una compañía estuviera incluida en el Fortune¹⁴ 100 para que se sumara al clamor por nuevas formas de ver los datos.

Las empresas empezaron a darse cuenta que sus ejecutivos no eran los únicos que podían tomar decisiones, y que la demanda de información tenía bases sólidas.

Dyché (2001) afirma que la lógica del proceso se puede justificar con premisas como:

- *En nuestra empresa todos pueden tomar decisiones.*
- *Las decisiones se basan tanto en hechos duros como blandos.*
- *Nuestra empresa tiene un enorme inventario de datos;*

Aunque estas viejas justificaciones para el datawarehousing son muy reales, existen nuevas ventajas que toman en cuenta más la dirección de los negocios que la alineación de los datos con la tecnología. Entre algunos de los ejemplos que cita Dyché

¹⁴ Fortune: Revista Internacional que enfoca y resalta el crecimiento de las principales empresas a nivel mundial. Referente para inversiones y posicionamiento de mercado.

(2001) se mencionan la verdad, implementación más simple y relación íntima con el cliente.

Con respecto a la verdad se puede decir que con un datawarehouse, no existirán múltiples respuestas para la misma pregunta. El uso de un datawarehouse consolida datos heterogéneos que permiten a las compañías responder a diferentes preguntas con la misma información.

Un datawarehouse simplifica la tarea de toma de decisiones, al ser sistemas informativos; al momento de implementar un datawarehouse se puede pasar por alto restricciones de seguridad, de diseño y de tecnología que tienen los sistemas operacionales.

Finalmente lograr una relación íntima con el cliente es considerado otro justificativo para iniciar un proyecto de datawarehouse. Conocer las necesidades del cliente en el momento oportuno brinda mejores oportunidades de negocio.

1.2.2. Estableciendo una Infraestructura

Según Dyché (2001), “la infraestructura de un datawarehouse es un compendio de varios elementos, como Sistemas Operacionales, Extracción, Limpieza, Carga de datos, Metadata, Acceso de usuario final, Plataforma del datawarehouse y Datos externos, que integrados definen un datawarehouse”. En concordancia con lo expuesto en los apartados previos la definición abarca todos los elementos que garantizan el correcto funcionamiento de un datawarehouse.

Los elementos que menciona esta autora son los siguientes:

a) Sistemas Operacionales

“Los datos administrados por los sistemas de aplicación operacionales son la fuente principal de datos para el datawarehouse”.

Para un proyecto de datawarehouse se pueden considerar bases de datos operacionales organizadas como archivos indexados (Foxpro), bases de datos de redes/jerárquicas (XML) o sistemas de base de datos relacionales (DB2, Oracle, Informix, etc.).

b) Extracción, Limpieza, Transformación y Carga de los Datos

“Se requieren herramientas de gestión de datos para extraer datos desde bases de datos y/o archivos operacionales, luego es necesario manipular o transformar los datos antes de cargar los resultados en el datawarehouse”.

Tomar los datos desde varias bases de datos operacionales y transformarlos en datos requeridos para el depósito, se refiere a la transformación o a la integración de datos. Las bases de datos operacionales, diseñadas para el soporte de varias aplicaciones de producción, frecuentemente difieren en el formato.

“Los mismos elementos de datos, si son usados por aplicaciones diferentes o administrados por diferentes software DBMS, pueden definirse al usar nombres de elementos inconsistentes, que tienen formatos inconsistentes y/o ser codificados de manera diferente”.

Todas estas inconsistencias deben resolverse antes que

los elementos de datos sean almacenados en el datawarehouse.

c) Metadata

“Otro paso necesario es crear la metadata. La metadata (es decir, datos acerca de datos) describe los contenidos del datawarehouse. La metadata consiste de definiciones de los elementos de datos en el depósito, sistema(s) del (os) elemento(s) fuente. Como la data, se integra y transforma antes de ser almacenada en información similar”.

La metadata es muy útil al momento de describir los datos contenidos dentro de un datawarehouse, debido a que provee información de origen, transformación y destino de los datos operacionales extraídos.

d) Acceso de usuario final

“Los usuarios acceden al datawarehouse por medio de herramientas de productividad basadas en GUI (Graphical User Interface - Interfase gráfica de usuario)”.

Estos pueden incluir software de consultas, generadores de reportes, procesamiento analítico en línea, herramientas data/visual mining, etc., dependiendo de los tipos de usuarios y los requerimientos particulares. Sin embargo, una sola herramienta no satisface todos los requerimientos, por lo que es necesaria la integración de una serie de herramientas.

e) Plataforma del datawarehouse

“La plataforma para el datawarehouse es casi siempre un servidor de base de datos relacional. Cuando se manipulan volúmenes muy grandes de datos puede requerirse una configuración en bloque de servidores UNIX con multiprocesador simétrico (SMP) o un servidor con procesador paralelo masivo (MPP) especializado”.

Los extractos de la data integrada/transformada se cargan en el datawarehouse. La elección de la plataforma es crítica. El depósito crecerá y hay que comprender los requerimientos después de 3 o 5 años.

Muchas de las organizaciones quieran o no, escogen una plataforma por diversas razones: “el Sistema X es nuestro sistema elegido o el Sistema Y está ya disponible sobre un sistema UNIX que nosotros ya tenemos”. Uno de los errores más grandes que las organizaciones cometen al seleccionar la plataforma, es que presumen que el sistema (hardware y/o DBMS) escalará con los datos.

El sistema de depósito ejecuta las consultas que se pasa a los datos por el software de acceso a los datos del usuario. Aunque un usuario visualiza las consultas desde el punto de vista de un GUI, las consultas típicamente se formulan como pedidos SQL, porque SQL es un lenguaje universal y el estándar para el acceso a datos.

f) Datos Externos

“Dependiendo de las necesidades de información de la organización, pueden existir datos externos a los sistemas transaccionales que alimentan el datawarehouse y que son accesibles por medio de archivos planos o en hojas electrónicas y/o vía Internet. La integración de estos datos

al datawarehouse expande el alcance y disponibilidad a los usuarios”.

Esto se debe en gran medida a que en las organizaciones todavía existen islas de información que representan sectores no integrados que manejan datos de interés común. Su inclusión en el consolidado de información que reposará en el datawarehouse es de vital importancia para garantizar la homogeneidad y fiabilidad del sistema.

1.2.3. Repositorios de Datos

Los datawarehouses (conjuntamente con los sistemas de soporte de decisión [Decision Support Systems - DSS] y las aplicaciones cliente/servidor), fueron los primeros éxitos para el DBMS relacional (Relational Data Base Management Systems - RDBMS), según INEI (1997).

Mientras la gran parte de los sistemas operacionales fueron resultados de aplicaciones basadas en antiguas estructuras de datos, los depósitos y sistemas de soporte de decisiones aprovecharon el RDBMS por su flexibilidad y capacidad para efectuar consultas con un único objetivo concreto.

La misma fuente indica que “los RDBMS son muy flexibles cuando se usan con una estructura de datos normalizada”. En una base de datos normalizada, las estructuras de datos son no redundantes y representan las entidades básicas y las relaciones descritas por los datos (por ejemplo productos, comercio y transacción de ventas). Pero un procesamiento analítico en línea (OLAP) típico de consultas que involucra

varias estructuras, requiere varias operaciones de unión para colocar los datos juntos.

El performance¹⁵ de los RDBMS tradicionales es mejor para consultas basadas en claves, según INEI (1997). Por ejemplo "Encuentre cuenta de cliente #2014" que para consultas basadas en el contenido "Encuentre a todos los clientes con un ingreso sobre \$ 10,000 que hayan comprado un automóvil en los últimos seis meses".

Para el soporte de depósitos a gran escala y para mejorar el interés hacia las aplicaciones OLAP, los proveedores han añadido nuevas características al RDBMS tradicional. Estas, también llamadas características super relacionales, incluyen el soporte para hardware de base de datos especializada, tales como la máquina de base de datos Teradata¹⁶.

¹⁵ Performance es una característica que denota desempeño o rendimiento. Término en inglés que es utilizado comúnmente en el lenguaje computacional.

¹⁶ Teradata El Sistema de Gestión de Base de Datos Relacional (RDBMS) Teradata de NCR, es una base de datos relacional paralela de soporte de decisiones muy poderosa para datawarehousing. Con el servidor basado en el soporte para UNIX SVR4 y Microsoft Windows NT, Teradata ofrece soporte a las empresas para la escalabilidad desde gigabytes a terabytes, pentabytes y más.

Los modelos super relacionales también soportan extensiones para almacenar formatos y operaciones relacionales (ofrecidas por proveedores como RedBrick) y diagramas de indexación especializados, tales como aquellos usados por Sybase IQ. Estas técnicas pueden mejorar el rendimiento para las recuperaciones basadas en el contenido, al pre-juntar tablas usando índices o mediante el uso de listas de índice totalmente invertidos.

Muchas de las herramientas de acceso a los datawarehouses explotan la naturaleza multidimensional del datawarehouse. Por ejemplo, los analistas de marketing necesitan buscar en los volúmenes de ventas por producto, por mercado, por período de tiempo, por promociones y niveles anunciados y por combinaciones de estos diferentes aspectos.

La estructura de los datos en una base de datos relacional tradicional, facilita consultas y análisis a lo largo de dimensiones¹⁷ diferentes que han llegado a ser comunes. Estos

¹⁷ Dimensiones: determina la granularidad para la determinación de los hechos. Ej. : producto, fecha, almacén

esquemas podrían usar tablas múltiples e indicadores para simular una estructura multidimensional. Algunos productos DBMS, tales como Essbase y Gentium, implementan técnicas de almacenamiento y operadores que soportan estructuras de datos multidimensionales.

Mientras las bases de datos multidimensionales (Multidimensional Databases - MDDBs) ayudan directamente a manipular los objetos de datos multidimensionales (por ejemplo, la rotación fácil de los datos para verlos entre dimensiones diferentes, o las operaciones de drill down¹⁸ que sucesivamente exponen los niveles de datos más detallados), se debe identificar estas dimensiones cuando se construya la estructura de la base de datos. Así, agregar una nueva dimensión o cambiar las vistas deseadas, puede ser engorroso y costoso. Algunos MDDBs requieren un recargue completo de la base de datos cuando ocurre una reestructuración.

¹⁸ Drill Down: Operación inversa: muestra información detallada de cada agrupamiento. Por Ej. Analizar las ventas de tipo de producto a las ventas por producto.

1.2.4. Una Estrategia para la Construcción de Datawarehouse

En esta sección se detalla una metodología probada por Dyché (2001) en varias empresas de diversa orientación de negocios, si bien es cierto no existe una fórmula de garantía real para el éxito de la construcción de un datawarehouse, muestra muchos puntos que contribuyen a ese objetivo.

ESTRATEGIAS PARA LA PLANIFICACIÓN DE UN DATAWAREHOUSE

A continuación, se indican algunos puntos claves que deben considerarse según el INEI (1997) para la planificación de un Datawarehouse:

- 1. Establecer una asociación de usuarios, gestión y grupos.*
 - 2. Seleccionar una aplicación piloto con una alta probabilidad de éxito.*
 - 3. Construir prototipos rápida y frecuentemente.*
 - 4. Implementación incremental.*
 - 5. Reportar activamente y publicar los casos exitosos.*
-

Desarrollando los pasos expuestos se puede mencionar que es esencial involucrar tanto a los usuarios como a la gestión para asegurar que el datawarehouse contenga información que satisfaga los requerimientos de la empresa. Los usuarios y la gestión justifican los costos del datawarehouse sobre cómo será el ambiente y está basado primero en lo esperado y segundo, en el valor comercial real.

Una aplicación piloto de alcance limitado, con un reembolso medible para los usuarios y la gestión, establecerá el datawarehouse como una tecnología clave para la empresa. Estos mismos criterios (alcance limitado, reembolso medible y beneficios claros para la empresa) se aplican a cada fase de la implementación de un datawarehouse.

La única manera para asegurar que el datawarehouse reúna las necesidades de los usuarios, es hacer el prototipo a lo largo del proceso de implementación y aún más allá, así como agregar los nuevos datos y/o los modelos en forma permanente. El trabajo continuo con los usuarios y la gestión es, nuevamente, la clave.

La implementación incremental reduce riesgos y asegura que el tamaño del proyecto permanezca manejable en cada fase.

La retroalimentación de los usuarios ofrece una excelente oportunidad para publicar los hechos exitosos dentro de una organización. La retroalimentación del usuario también ayuda a comprender cómo evoluciona la implementación del datawarehouse a través del tiempo para reunir requerimientos de usuario nuevamente identificados.

ESTRATEGIAS PARA EL DISEÑO DE UN DATAWAREHOUSE

El diseño de los datawarehouses es muy diferente al diseño de los sistemas operacionales tradicionales. De acuerdo al INEI (1997), se pueden considerar los siguientes puntos:

1. Los usuarios de los datawarehouses usualmente no conocen mucho sobre sus requerimientos y necesidades como los usuarios operacionales.

2. El diseño de un datawarehouse, con frecuencia

involucra lo que se piensa en términos más amplios y con conceptos del negocio más difíciles de definir que en el diseño de un sistema operacional.

3. Finalmente, la estrategia de diseño ideal para un datawarehousing es generalmente de afuera hacia adentro (outside-in) a diferencia de arriba hacia abajo (top-down).

Vale indicar que a pesar de que el diseño de un datawarehouse es diferente al usado en los diseños de sistemas tradicionales, no es menos importante. El hecho de que los usuarios finales tengan dificultad en definir lo que ellos necesitan, no lo hace menos necesario.

En la práctica, los diseñadores de datawarehouses tienen que usar muchos trucos para ayudar a los usuarios a visualizar sus requerimientos. Por ello, son esenciales los prototipos de trabajo, que no son otra cosa que herramientas que permiten la evaluación de un modelo o sistema

ESTRATEGIAS PARA EL DESARROLLO DE UN DATAWAREHOUSE

Antes de desarrollar un datawarehouse, es crítica la elaboración de una estrategia equilibrada que sea apropiada a las necesidades y los usuarios.

Las preguntas que deben tenerse en cuenta para el desarrollo de la estrategia son:

¿Quién es el auditorio?

¿Cuál es el alcance?

¿Qué tipo de datawarehouse debería construirse?

Las respuestas a estas interrogantes permitirán tener una concepción más clara sobre la orientación que deberá tomar el datawarehouse.

Según Bischoff¹⁹ (1997), existe un número de estrategias mediante las cuales las organizaciones pueden conseguir sus datawarehouses:

¹⁹ Joyce Bischoff es Presidenta de Bischoff Consulting, Inc. Empresa dedicada a la consultoría de metodologías en EEUU, Europa y Canadá de diseño, implementación y desempeño de bases de datos.

1. *Establecer un ambiente "datawarehouse virtual".*
2. *Construir una copia de los datos operacionales desde un sistema operacional único y posibilitar al datawarehouse de una serie de herramientas de acceso a la información.*
3. *Finalmente, la estrategia datawarehousing óptima es seleccionar el número de usuarios basados en el valor de la empresa y hacer un análisis de sus puntos, preguntas y necesidades de acceso a datos.*

Se puede mencionar que para el establecimiento de un ambiente de datawarehouse virtual es necesario considerar:

- ✓ Instalación de un conjunto de facilidades para acceso a datos, directorio de datos y gestión de proceso.
- ✓ Entrenamiento de usuarios finales.
- ✓ Control de cómo se usan realmente las instalaciones del datawarehouse.

Basados en el uso actual, crear un datawarehouse físico debe poder soportar los pedidos de alta frecuencia.

La estrategia de seleccionar los usuarios del sistema y las necesidades del mismo tiene la ventaja de ser simple y rápida.

Una vez se tenga un consenso general sobre las necesidades, entonces se consiguen los datos provenientes de los sistemas operacionales existentes a través de la empresa y/o desde fuentes externas de datos y se cargan al datawarehouse.

Si se requieren herramientas de acceso a la información, se puede también permitir a los usuarios finales tener acceso a los datos requeridos usando sus herramientas favoritas propias, o facilitar la creación de sistemas de acceso a la información multidimensional de alta performance, usando el núcleo del datawarehouse como base.

En conclusión, no se tiene un enfoque único para construir un datawarehouse que se adapte a las necesidades de las empresas, debido a que las necesidades de cada una de ellas son diferentes, al igual que su contexto.

ESTRATEGIAS PARA LA GESTION DE UN DATAWAREHOUSE

La gestión de un datawarehouse comprende todo aquello que se relacione con la administración y mantenimiento del mismo; es por esto que requiere una gestión muy cuidadosa. Bischoff (1997) recomienda considerar lo siguiente para este apartado:

1. Un datawarehouse es una inversión buena sólo si los usuarios finales realmente pueden conseguir información vital más rápida y más barata de lo que obtienen con la tecnología actual.

2. La administración debe reconocer que el mantenimiento de la estructura del datawarehouse es tan crítico como el mantenimiento de cualquier otra aplicación de misión-crítica.

De hecho, la experiencia ha demostrado que los datawarehouses llegarán a ser rápidamente uno de los sistemas más usados en cualquier organización.

3. La gestión debe comprender también que si se embarcan sobre un programa datawarehousing, se crearán nuevas demandas sobre los sistemas operacionales.

Como consecuencia, la gestión de una datawarehouse tiene que pensarse seriamente, definir la forma en que requieren los depósitos para un eficaz desempeño y la forma de llegar a los usuarios finales, son de vital importancia.

1.3. Aspectos Organizacionales y Culturales

1.3.1. *Datawarehousing del Mundo Real: El Cambio en la Administración*

Al implementar un datawarehouse por primera vez, las compañías deben ser capaces de desbloquear el valor de sus datos históricos. No sólo deben recuperar datos, deben usarlos para futuros proyectos que exploren tendencias de ventas, necesidades y mercados.

Fosdick²⁰ (1997) menciona que para muchas compañías, “datawarehousing representa una nueva dimensión en la utilidad de los sistemas de información”.

Los gerentes de estas compañías toman decisiones de administración y mercadeo basados en el particionamiento de datos hecho desde muchos ángulos y perspectivas.

Dyché (2001) menciona casos de clientes que han ido más allá de limitarse a consultar los datos y que han logrado poner en

²⁰ Howard Fosdick es un analista y consultor independiente de industrias y negocios que ha escrito cientos de artículos y 6 libros de administración.

manos de sus hombres de negocios un soporte de decisión suficiente para la acción.

Un informe realizado por el Palo Alto Management Group²¹ (1999) confirma que los informes básicos todavía son la norma en los datawarehouses instalados en gran medida. El informe descubrió que el análisis financiero y la realización de reportes en general eran las dos aplicaciones más populares utilizadas.

El informe también menciona una tendencia a que la realización de reportes de inteligencia de negocios cruce gradualmente las fronteras de los “trabajadores del conocimiento” y caiga en manos de empleados de toda la empresa.

En resumen, la concepción está dada porque los usuarios son los que deciden los datos que necesitan. La confianza que deben tener estos datos es fundamental por cuanto sirven de base para la toma de decisiones del negocio.

²¹ Palo Alto Management Group, Business Intelligence and Data Warehousing: Crossing the Millenium, 1999,12. Para un resumen del informe, véase <http://www.pamg.com>

1.3.2. Aspectos de la Administración

La clave para justificar la implementación de un datawarehouse es aclarar los aspectos de la administración que se beneficiarán y determinar las ganancias que según Dyché (2001) son:

- Un mejoramiento en la toma de decisiones.
- Mejorar los Servicios de Información.
- Optimizar el manejo de recursos.
- Perfeccionar los procesos organizacionales.

Por otra parte, hay otras funciones dentro de la empresa que tienen que ver con el planeamiento, previsión y administración de la organización. Estas funciones son también críticas para la supervivencia de la organización, especialmente en nuestro mundo de rápidos cambios.

Sin lugar a duda, en la actualidad se vive un ambiente muy acelerado, lo cual, repercute en la dificultad para responder a las oportunidad y amenazas.

Es por ello que hoy en día los administradores de cada una de las empresas se enfrentan a una gran cantidad de retos para lograr mantener su empresa en el mercado.

Los sistemas de soporte a la decisión, ayudan a las empresas con su objetivo de crear una ventaja competitiva, debido a que son sistemas basados en computadoras que brindan a los usuarios la información necesaria para la toma de decisiones,

A través de estos sistemas, el usuario puede responder de manera más concreta en la toma de decisiones, ya que le brinda una panorámica sobre el problema y una serie de alternativas propias para la resolución del mismo. Asimismo, los beneficios que se pueden obtener de los mismos son: calidad de decisión, mejora de comunicación, reducción de costos, incrementos en productividad, etc.

1.3.3. *Situación Actual*

A continuación se revisará el ambiente y el entorno en el cual se desarrollan las actividades de negocios en la ESPOL con el ánimo de tener una perspectiva introductoria (información tomada de archivos históricos publicados por el Centro de Servicios Informáticos de la ESPOL):

- En los primeros días de los sistemas implantados en la ESPOL, el diseño de los módulos de información estaba marcado por las necesidades puntuales del día a día de diferentes áreas (Académica, Contabilidad, Presupuesto, Tesorería, Personal, etc.).
 - Pronto se dio comienzo el mantenimiento de esas aplicaciones. Las aplicaciones necesitaban cambios por muchas razones:
 - Nuevos requisitos.
 - Cambios en los procesos del negocio.
 - Nuevas oportunidades.
 - Al mismo tiempo que comenzaba el mantenimiento surgía la necesidad de extraer más información de estas aplicaciones.
-

- El primer intento para satisfacer esta necesidad fue la escritura de módulos que listen o presenten reportes.
- La primera limitación de estos reportes es que accedían a una información aislada. Hubo que definir interfaces entre las aplicaciones para que pudiesen compartir datos entre ellas (scripts²², actualizaciones entre tablas de Roles y Financieras, imports/exports²³ de registros).
- La segunda limitación es que los reportes había que modificarlos de forma constante.

Todo lo anterior se traduce en aspectos puntuales como:

- Reportes de áreas relacionadas no presentan información consistente.
- Los datos no son totalmente confiables.
- Los usuarios no cuentan con herramientas que les permita personalizar sus reportes. Dependen exclusivamente del personal del departamento de sistemas para esto.
- Los requerimientos de los usuarios en cuanto a reportes no pueden ser atendidos de manera inmediata, puesto que en

²² Scripts son pequeños programas o porciones de programación que ejecutan operaciones de procesamiento sobre un servidor.

²³ Imports/Exports son procedimientos de exportación de datos en una localidad específica para luego importarlos (insertarlos y/o actualizarlos) en el destino seleccionado.

el departamento de sistemas existen prioridades en base a la programación y desarrollo de proyectos.

- Los usuarios no conocen qué información existe y cómo hacer uso de ella. Esto provoca que constantemente llamen al departamento de sistemas para resolver sus dudas al respecto.
 - Falta de procedimientos administrativos claramente establecidos, que impiden que los sistemas desarrollados cumplan con la funcionalidad requerida.
 - La Dirección General conformada por los directivos principales de áreas como Rectorado, Vicerrectorados, CRECE y demás unidades académicas de apoyo, no tienen acceso a información resumida que le permita realizar una gestión administrativa de manera ágil y efectiva.
-

1.3.4. Imperativos Culturales

Las organizaciones son tan eficientes como lo son sus procesos. En una institución pública como la ESPOL en la que los procesos administrativos deben estar sujetos a controles legales establecidos resulta más complejo crear una verdadera cultura en el acceso de la información que sirve de soporte para la Toma de Decisiones. Una alternativa para potenciar la manera en que actualmente se toman decisiones es emprender un proyecto de Datawarehouse.

Entre algunos de los imperativos culturales identificados en esta investigación se pueden citar los siguientes:

1. Limitada disponibilidad de tiempo del grupo de desarrollo para el proyecto con las habilidades requeridas.
 2. Nueva tecnología en la curva de aprendizaje, y aspectos en la implementación del sistema.
 3. Permanente cooperación y soporte en el funcionamiento y gestión del datawarehouse.
-

1.3.5. *Criterios de Éxito*

Los criterios de éxito forman parte de los aspectos culturales y organizacionales a ser tomados en cuenta. Adelman²⁴ (1997) menciona 10 criterios de éxito en un proyecto de datawarehouse:

1. El datawarehouse debe ser capaz de satisfacer los requerimientos de los usuarios.
2. El datawarehouse debe tener una contribución significativa en el éxito de los negocios. (por ejemplo mejores decisiones deben trasladarse en más ventas y alta productividad)
3. La organización debe ser capaz de establecer cierta jerarquía de proyectos. En un ambiente donde un datawarehouse sea considerado efectivo, siempre existirán proyectos relacionados que demanden su atención y dependencia.
4. El datawarehouse es percibido por la alta administración y dirección como una contribución significativa en la

²⁴ Sid Adelman ha trabajado activamente en consultorías sobre sistemas de soporte en la toma de decisiones. Publicador de revistas especializadas de datawarehouse como “Data Base Management Magazines”, “Computerworld”, entre otras.

manera de hacer negocios. Si la administración no reconoce los beneficios, el proyecto de datawarehouse podría fracasar.

5. Los usuarios deben acoger el datawarehouse y usarlo activamente.
 6. Los beneficios no deben exceder los costos.
 7. Un presupuesto adecuado debe ser destinado para el datawarehouse ya que no son baratos. Los fondos destinados para este proyecto serán utilizados en compra de software, hardware y un adecuado entrenamiento.
 8. Las habilidades apropiadas para la planificación, diseño, implementación y gestión deben estar definidas y deben dedicarse al proyecto del datawarehouse.
 9. La implementación del datawarehouse no debe ser causa de otros problemas que opaquen los beneficios del mismo. Puede darse el caso que un datawarehouse afecte el desempeño de sistemas transaccionales críticos.
 10. Una programación razonable debe ser establecida. Un proyecto exitoso es aquel que ha cumplido un cronograma establecido. A menudo, tiempos
-

inadecuados son permitidos para un proyecto de datawarehouse, especialmente considerando que algunas actividades son realizadas por primera vez.

El cumplimiento de estos criterios es de utilidad al momento de instaurar con éxito un datawarehouse dentro de una organización. Para el desarrollo del datawarehouse de la ESPOC se tomará muy en cuenta el cumplimiento de todos los criterios con excepción del criterio 7 ya que por formar parte de un proyecto de investigación en su fase inicial no se contempla la adquisición de software o hardware adicional.

1.3.6. Causales de Fracaso

La gran mayoría de los fracasos de datawarehouses y datamarts se caracterizan porque un número de malas decisiones obligaron a las empresas a empezar de nuevo antes que por un esfuerzo de desarrollo totalmente mal dirigido. A partir de estos intentos fallidos en los proyectos de datawarehouses se pueden sacar valiosas lecciones.

En diversos foros y documentos en línea²⁵ acerca del tema se insiste repetidamente en causales de fracaso que contienen los siguientes puntos:

1. Diseñar el datawarehouse de manera diferente a la manera en que se diseñan los sistemas operacionales y transaccionales.
2. Asegurar que no se limitará a cargar gran cantidad de datos sin comprender su significado y su valor comercial.

²⁵ Se pueden mencionar los siguientes foros de discusión sobre tecnología datawarehousing, algunos independientes y otros particulares <http://www.datawarehousing.com>, <http://forums.oracle.com> <http://www.warriorforum.com>, etc.

3. Contar con un patrocinador o “padrino” del área de negocios comprometido con el proyecto.
4. No pasarse meses o años desarrollando un modelo de datos para toda la empresa.
5. No fijar metas poco realistas.
6. Si quiere fracasar en datawarehousing, no tome en cuenta a los usuarios.

A continuación se listan los riesgos latentes del datawarehousing según Dyché (2001), lo cuales no se centran exclusivamente a la implementación (muy criticada y culpada) sino también se considera el uso.

Riesgo N°1: “El Síndrome del datawarehouse como Panacea²⁶”

Uno de los problemas más grandes de la ciencia es hacer que una nueva tecnología sea tan buena o mejor que la tecnología que presume reemplazar. Por esto, muchos proyectos de datawarehouses fueron emprendidos con buena fe al ser considerados como panaceas. Pero en realidad, a veces un datawarehouse no es la mejor solución al problema.

²⁶ Medicamento que se creía eficaz para todas las enfermedades.

Por esto una de las primeras preguntas que se deben hacer para evaluar la oportunidad de un datawarehouse es “¿Realmente su problema comercial se resuelve con datos?”. A veces el problema está en los procesos administrativos o en la cultura de la empresa. Si la respuesta es “no”, o incluso “quizá”, se recomienda que antes de comprometer dinero en el proyecto, la empresa haga una evaluación adicional.

Riesgo N°2: “Hablar con los Usuarios Finales ¡Pero con los Equivocados!”

Pudiera parecer que esta posibilidad es poco probable, pero en la práctica sucede todo el tiempo.

En las reuniones de recolección de requerimientos los usuarios que suelen aparecer son los más locuaces, o se elige al subconjunto más técnico de los usuarios finales esperando que se imponga la regla del 80-20 (es decir, el 20% de los usuarios representa el 80% de los requerimientos).

A menudo, las personas que se saben expresar mejor para explicar el problema comercial delegarán el uso del datawarehouse a otros con prioridades enteramente diferentes.

Es importante la comparación de las respuestas a la encuesta con entrevistas personales o con los hallazgos obtenidos en talleres de recolección de requerimientos. Hay que decidir correctamente quiénes deberían formar el núcleo del grupo de entrevistados. Luego, hacer un análisis para tratar de descubrir si fuera de ese grupo hay otros individuos aislados a los que también hay que tener en cuenta y entrevistarlos.

Riegos N°3: “Dedicar demasiado tiempo a investigar y perder de vista a los destinatarios”.

Esto se interpreta como el gran error que representaría en una implementación de datawarehouse intentar comprender hasta los aspectos más detallados de esta tecnología con el consumo de tiempo y recursos excesivo, lo cual restaría dedicación al objetivo primordial y de valor para la empresa como son los usuarios.

Riesgo N°4: “Estancar un proyecto en la creación de Metadatos.”

Puede ocurrir que el equipo de TI ocupe demasiado tiempo y recursos nada más que para ingresar definiciones y reglas comerciales dentro de una base de datos, mientras se deja el desarrollo suspendido.

En realidad, lo recomendable es implementar la metadata de manera incremental, sabiendo que tanto ellos como los datos que describen irán evolucionando lentamente con el tiempo.

Es importante la implementación iterativa. Aunque la metadata es importante, su implementación jamás debe tener prioridad sobre el desarrollo de un sistema fácil de usar.

Riesgo N° 5: “Desviarse en análisis de cosas interesantes.”

Esto apunta a la tendencia que tienen algunas compañías a hacer uso de tecnologías estratégicas para propósitos triviales y estratégicos. Aunque no hay nada

de malo en lo estratégico, muchos usuarios finales se limitan a consultar el datawarehouse por el datawarehouse en sí mismo, experimentando si pueden obtener respuesta a una pregunta dada, aunque ésta tenga poco valor comercial.

El riesgo en este caso es que se pierde el sentido de utilidad al emplear los recursos de esta plataforma en actividades poco productivas para la actividad del negocio, y que posteriormente cuando se requiera una actualización de hardware o una nueva herramienta de software, la gerencia requerirá una explicación sobre la forma en que esta saturación del datawarehouse ayudó a incrementar los ingresos.

Riesgo N°6: “Adoptar soporte de decisión sin soportar decisiones”.

El problema es que los usuarios no tienen autoridad dentro de la organización para promover nuevas estrategias comerciales. En ocasiones ni siquiera se les permite tomar decisiones sugeridas por los informes de soporte de decisión obtenidos del datawarehouse.

Un datawarehouse y la información que provee debería no solamente servir para encontrar problemas, sino también para mejorar procesos de negocios y motivar acciones innovadoras.

Si el datawarehouse se limita a corroborar hipótesis ya aceptadas, o si los usuarios finales pueden mirar pero no tocar, entonces la empresa no está haciendo con él todo lo que podría.

Riesgo N°7: "Codicia en las organizaciones de desarrollo"

La repartición de los costos de nuevos sistemas o información de valor de una organización se ha vuelto una práctica común entre organizaciones de informática que tienen que priorizar proyectos de implementación a la vez que se ven forzados a lidiar con presupuestos cada vez más reducidos.

El problema con este enfoque es que los departamentos de desarrollo informático empiezan a ponerse codiciosos y le ponen un recargo a la realización de nuevas

aplicaciones. Algunos grupos llegan a agregarle un 75% a lo que realmente cuesta crear una nueva aplicación.

Pero las unidades comerciales se están volviendo astutas, y cada vez más acuden a consultores/proveedores externos para que les provean de soluciones de inteligencia comercial. Estas soluciones pueden requerir nuevas tecnologías y quizás no se adhieran a los estándares fijados por el área informática de la empresa, pero permiten disponer de la funcionalidad por menos dinero de lo que costaría un proyecto interno similar.

Las empresas deberían estar dispuestas a pagar por sus necesidades de soporte de decisión, pero se les debería cobrar al costo. De esta manera, nadie pierde dinero, y además se puede mantener una estrategia centralizada duradera, lo que permite que el datawarehouse madure en una única dirección integrada.

Riego N°8: "Falta de Relaciones públicas internas"

Los mentalizadores del datawarehouse, tanto por el área tecnológica como por el área comercial, tienden a

olvidarse del concepto de relaciones públicas internas, es decir, el marketing constante de las capacidades y características del datawarehouse dirigido a las principales organizaciones y miembros del personal. El objetivo que se apunta es aumentar continuamente la funcionalidad y prestaciones de datos, incrementando a la vez el número de usuarios finales.

La difusión continua de los beneficios y funcionalidades de un datawarehouse y de sus beneficios engendra el apoyo necesario para mantenerlo vivo y en crecimiento.

Riesgo N°9: “No darse cuenta de que las aplicaciones de soporte de decisión son finitas.”

La conjetura habitual aquí es que una vez que se programa una aplicación de soporte de decisión y se la distribuye a los usuarios finales, ésta durará para siempre. Más aún, solamente el hecho de tenerlas en su lugar tiene costos de mantenimiento asociados. El hecho es que, por una variedad de razones, muchas aplicaciones de variada índole dejan de ser útiles (y en consecuencia, todos dejan de usarlas). No es que la

aplicación o herramienta sea mala, es que ya cumplió su ciclo de vida.

Mediciones tales como los registros de uso (que indican las estadísticas de acceso y uso del sistema), notas de reclamos técnicos, y pedidos de mejoras deberían figurar en la decisión de continuar manteniendo una aplicación o dejarla morir.

Riesgo N°10: “Prestarle demasiada atención al desarrollo e ignorar el empleo.”

Es probable que se haya visto uno de esos complicados diagramas del proceso de implementación de un datawarehouse, con todos sus cuadros y sus flechas que apuntan de un lado para el otro, y se haya preguntado ¿cómo hacen los datos para terminar en mi hoja de cálculo?

Ésta es una pregunta válida, ya que existen muchas tecnologías para poner los datos en el datawarehouse. Los profesionales de datawarehousing experimentados saben que la parte difícil del ciclo es la extracción,

transformación y carga en la plataforma de datawarehouse elegida de los datos provenientes de los diversos sistemas operacionales. Así que allí se quedan.

Pero una vez cargados los datos tienen una infinidad de actividades que realizar, de las que crear reportes y consultas es solamente una de sus actividades. Una vez creada la aplicación, hay diversas decisiones que tomar en cuanto a pruebas, mantenimiento, soporte, entrenamiento, documentación y seguridades.

1.3.7. Roles y Responsabilidades en la Organización

Definiciones de roles y responsabilidades son importantes sólo si todos los roles requeridos son entendidos y asignados. Las asignaciones pueden fácilmente ser olvidadas o pasadas por alto.

En casos particulares, existen organizaciones cuyo desarrollo del datawarehouse quizás no requiera cada una de las funciones descritas a continuación, se pueden usar las tablas siguientes para evaluar si asignará cada uno de los papeles descritos a alguien, y cómo lo hará. Cada tabla del sistema de toma de decisiones abarca los papeles y responsabilidades fundamentales que están implicados en cada una de las funciones que aparecen en la figura abajo mostrada.

Las descripciones pueden servir como una lista de verificación para el equipo de desarrollo que ya tiene, o como una guía de desarrollo para cuando se implemente otro proyecto en particular.

Las descripciones de tareas en las tablas son específicas del desarrollo del datawarehouse y por eso no incluyen funciones más generales tales como “arquitecto en jefe” o “gerente de control de calidad”. Estas funciones pueden verse implicadas en el desarrollo de un datawarehouse, pero no son de su exclusividad.

Los títulos marcados con un asterisco (*) señalan puestos que pueden ser ocupados por consultores externos. Todos estos puestos de trabajo deberían salir de la compañía (y permanecer en ella), es decir, los deberían ocupar miembros del personal permanente de tiempo completo.

Las funciones de Gerencia tienen que ver con labores de planificación de recursos, administración de fondos, establecimiento de criterios de avance, coordinación de reuniones y entrevistas, supervisión y control de las actividades involucradas en el datawarehouse.

La siguiente tabla detalla cada uno de los roles y responsabilidades.

TÍTULO	DESCRIPCIÓN	RESPONSABILIDADES TÍPICAS
Patrocinador ejecutivo	El patrocinador ejecutivo es normalmente un miembro del personal de la gerencia de la empresa o de la línea comercial (como opuesto al área de informática). Lo ideal es que el patrocinador ejecutivo sea el director de uno o más departamentos en los que trabajan los potenciales usuarios finales del datawarehouse. Es el punto focal de la toma de decisiones de alto nivel y usualmente aporta los fondos para el desarrollo, en su totalidad o parte.	Es representante del datawarehouse y de su estado de desarrollo ante la gerencia superior. Establece la visión del datawarehouse. Asiste en la autorización de objetos del proyecto de datawarehouse. Participa en actividades de planificación de proyectos de alto nivel. Participa en la contratación de colaboradores fundamentales.
Dirección técnica del proyecto	La dirección técnica del proyecto trabaja con los grupos de desarrollo de la base de datos y de la aplicación para definir las actividades del proyecto y establecer criterios para medir el grado de avance. Usualmente está al frente de la planificación de la arquitectura general de la solución. También está involucrado en determinar la capacitación requerida para cada función y en entrevistar y reclutar al personal de desarrollo.	Alinear el plan de proyecto con la metodología de desarrollo elegida. Verificar los componentes técnicos del plan de proyecto. Entrevistar y reclutar al personal. Participar en reuniones informativas técnicas y en revisiones de diseño.
Director administrativo del proyecto*	El director administrativo del proyecto coordina y dirige el proyecto de desarrollo en general, supervisa las actividades diarias del proyecto, y mantiene el plan de proyecto. Trabaja con el patrocinador ejecutivo y el director técnico de proyecto para definir los marcos temporales y controlar avances y desviaciones.	Crear/mantener el plan de proyecto. Celebrar reuniones periódicas acerca del estado del proyecto. Comunicar el progreso del proyecto al patrocinador ejecutivo. Organizar los medios necesarios. Llevar un control permanente de los problemas que surjan y de su resolución.

Tabla 2: Funciones de Gerencia

Fuente: "E-data. Transformando datos en información", Dyché, Prentice Hall

Las funciones de elaboración de requerimientos están enfocadas a la dirección y participación en reuniones de recolección de necesidades de usuarios, comprobación de resultados y provisión de retroalimentación, participación en pruebas y prototipos.

TÍTULO	DESCRIPCIÓN	RESPONSABILIDADES TÍPICAS
Enlace con los usuarios/comunicaciones del proyecto	Esta función hace de colchón entre los usuarios finales y el personal de desarrollo, así como de recurso central al que los usuarios finales pueden comunicar sus problemas y preocupaciones. Este puesto lo puede ocupar una persona que cumpla múltiples funciones, desde análisis comercial a soporte, o un equipo de gente creado especialmente para este propósito.	Dirigir reuniones periódicas de informe de situación del proyecto con los usuarios finales y los principales interesados. Participar en reuniones y entrevistas para la recolección de los requerimientos de los usuarios. Comunicar a los usuarios finales la agenda de desarrollo y las fechas de lanzamiento.
Experto en la materia	Siempre debe ser un recurso interno, y estar tan familiarizado con la visión del soporte de decisión como con los requerimientos comerciales, el experto en la materia varía según la aplicación en construcción. Por ejemplo, podría ser un analista financiero, o un especialista en distribución, así como un experto en un sistema interno u otra tecnología.	Participar de la prueba de aceptación de los datos una vez cargados. Prover de retroalimentación (feedback) una vez terminada la construcción. Servir como recurso para preguntas acerca de los requerimientos y cuestiones de "aparición" (look and feel ²⁷). Prover asesoramiento permanente acerca de la aplicación terminada.
Facilitador*	Esta persona es el conductor de toda reunión de descubrimiento comercial, definición de alcance y recolección de requerimientos	Agendar y dirigir reuniones de recolección de requerimientos. Establecer metas claras

²⁷ Término acuñado en computación a sistemas e interfaces que son visualizadas y funcionan de la misma manera en que fueron diseñadas.

	que involucre múltiples participantes. Aunque en la mayoría de los casos el analista comercial también sirve como facilitador, en proyectos más complejos el facilitador dirigirá la reunión mientras el analista comercial hace preguntas o toma nota de las respuestas.	para las sesiones de recolección de requerimientos. Colaborar con el alcance con los usuarios para la creación de consenso. Documentar el resultado de las reuniones y las cuestiones.
Analista comercial*	Contacto principal para las actividades de recolección de requerimientos. Dependiendo del alcance del datawarehouse, el analista comercial puede en realidad dirigir un equipo de analistas especializados en facilitación, tomado de notas y creación de consenso. El analista comercial también puede ser requerido para crear la documentación de requerimientos.	Agendar y conducir talleres para la recolección de requerimientos de los usuarios. Comunicar al director de proyecto cambios en el alcance o en los requerimientos. Trasladar los requerimientos comerciales al diseñador de la base de datos. Participar en pruebas de prototipos de aplicación con los usuarios finales para asegurar la continuidad.

Tabla 3: Funciones de Elaboración de Requerimientos

Fuente: E-data. Transformando datos en información”, Dyché, Prentice Hall

Las funciones de desarrollo de base de datos se encargan de construir modelos lógicos con los datos obtenidos de los usuarios, asegurar la integridad y calidad de los datos, administrar y controlar las herramientas y los accesos.

TÍTULO	DESCRIPCIÓN	RESPONSABILIDADES TÍPICAS
Modelador/diseñador de datos*	El modelador/diseñador de las bases de datos puede cumplir más de una función en el desarrollo de la base de datos, y a menudo dirigirá un equipo de diseño formado por gente con diferentes habilidades: modelo lógico de datos, experiencia en software de modelado, o experiencia en DBMS ²⁸ . El diseñador debe tener experiencia en el diseño de bases de datos relacionales para soporte de decisión, incluyendo la interacción directa con hombres de negocios y el personal de DBA por igual, y debe comprender la convención de modelado elegida.	Construir modelos lógicos con los datos obtenidos del analista comercial y de los usuarios finales. Trabajar con el DBA para crear el diseño físico de la base de datos, basándose en las necesidades de acceso. Participar en actividades escogidas de recolección de requerimientos para comprender los problemas comerciales. Elegir y usar una herramienta de diseño de bases de datos para documentar el modelo. Trabajar con el DBA para realizar cambios periódicos de diseño a fin de mejoras al rendimiento y ajustes.
Administrador de datos (también llamado data steward, mayordom o de datos)	El administrador de datos se concentra en cuestiones relacionadas a los datos, su aplicabilidad a los requerimientos, su ubicación tanto antes como después de la carga, y su definición. Es responsable de encargarse de cuestiones tales como establecer definiciones y reglas comerciales generales para toda la empresa, definir y mantener los metadatos ²⁹ , y ofrecer requerimientos adicionales de datos a medida que evolucionan las necesidades de la empresa. Dependiendo del tamaño de la organización y del alcance del proyecto, el administrador de	Reunir los requerimientos de datos a medida que se definen los requerimientos comerciales. Asegurar la calidad y la integridad de los datos. Participar junto al diseñador de las bases de datos en las decisiones relativas a la transformación de los datos. Crear y mantener definiciones de los metadatos. Identificar y definir los sistemas de origen. Controlar el uso de los

²⁸ DBMS es una abreviación en inglés de DataBase Management System ó Sistema Manejador de Bases de Datos. Es lo que se conoce como el “motor” de una base de datos.

²⁹ Metadatos son datos que definen datos. Similar a un diccionario de datos en el cual se definen la naturaleza y comportamiento de los diferentes tipos de información en un sistema manejados.

	datos también puede crear y mantener los modelos de datos, en ausencia de un diseñador de bases de datos tiempo completo.	datos y actuar como la persona indicada para todas las preguntas de los usuarios relacionadas con datos.
Administrador de base de datos (DBA ³⁰)	Aunque el DBA puede ser un consultor externo, lo ideal es que sea un recurso permanente, ya que es responsable de la administración del datawarehouse y de su mantenimiento continuo. Aunque es normalmente responsable de las tablas físicas en el datawarehouse (las que se diseñan a partir del modelo lógico), también debería estar familiarizado con las tareas de mantenimiento operacionales asociadas con el datawarehouse y con el producto RDBMS ³¹ . También debería comprender las cuestiones relativas al rendimiento y al ajuste, y el uso de diversas herramientas de soporte a la producción y de administración.	Trabajar con el administrador de datos para validar los datos físicos. Trabajar con el diseñador de la base de datos para realizar el diseño físico, incluyendo ajustes de rendimiento. Administrar las diversas herramientas de diseño y CASE ³² . Generar y mantener los objetos de bases de datos, incluyendo tablas e índices. Implementar prácticas continuas de operación tales como copia de seguridad/restauración, seguridad de datos, etc. Realizar la planificación de la capacidad y control de uso de forma continua.
Cargador de datos*	Dependiendo de la complejidad de los datos y la cantidad de sistemas de origen involucrados, la carga de datos puede ser realizada ya sea por el DBA o por un cargador de datos aparte. El cargador de datos es responsable de implementar las herramientas o el código necesarios para acceder, depurar, y transferir datos al datawarehouse.	Realizar la extracción, transformación y carga permanente de los datos. Evaluar los datos de los sistemas de origen. Trabajar con el administrador de datos para definir reglas de transformación de los datos y estándares de calidad.

Tabla 4: Funciones de Desarrollo de Base de Datos

Fuente: "E-data. Transformando datos en información", Dyché, Prentice Hall

³⁰ Database Administrator o Administrador de bases de datos. Profesional de sistemas encargado del manejo y control de un sistema de bases de datos.

³¹ RDBMS (Relational Database Management System) Manejador de bases de datos Relacional.

³² CASE es una filosofía que se orienta a la mejor comprensión de los modelos de la empresa, sus actividades y el desarrollo de los sistemas de información.

Las funciones de desarrollo de la aplicación son las encargadas de la programación de los requerimientos, revisión de documentos, planes y procedimientos.

TÍTULO	DESCRIPCIÓN	RESPONSABILIDADES TÍPICAS
Programador de aplicaciones*	El programador debería tener experiencia en el desarrollo de aplicaciones frontales de interfaz gráfica y tener conocimientos del entorno de software de aplicación específico. También debería estar familiarizado con, al menos, un lenguaje de programación. También es deseable que tenga experiencia en la definición de procedimientos de prueba, en la creación de prototipos y en proyectos rápidos de desarrollo de aplicaciones.	Desarrollar la aplicación frontal necesaria, usando la o las herramientas elegidas. Administrar la herramienta de aplicación para que acepte a nuevos usuarios. Participar en pruebas de la aceptación de los usuarios o en demos provisionales.
Probador de la aplicación*	El probador también debería tener experiencia en el desarrollo de herramientas frontales de interfaz gráfica, así como en procedimientos de prueba funcional y de integración.	Revisar la documentación de los requerimientos de los usuarios finales para confirmar el cumplimiento de los objetivos de la aplicación. Definir planes/procedimientos de prueba. Realizar pruebas de la aceptación de los usuarios y dirigir demos provisionales.

Tabla 5: Funciones de Desarrollo de la Aplicación

Fuente: “E-data. Transformando datos en información”, Dyché, Prentice Hall

1.3.8. Determinación de Áreas y Datos Iniciales

La más importante contribución para el éxito es tener en mente cuáles departamentos usarán el datawarehouse y qué información será cubierta.

El proyecto en si mismo debe tener un tamaño razonable, lo suficientemente grande como para que la organización aprenda de la experiencia y que los resultados sean apreciables. Esto es importante ya que el proyecto inicial hace una contribución importante al negocio. El datawarehouse no debe ser algo desechable ni mucho menos una simple prueba de conceptos, debe tener algún significado.

Luego de realizar un estudio y después de varias reuniones con los Directores de las áreas más críticas en cuanto a necesidades de información, se determinaron áreas y datos iniciales considerando primordialmente aspectos tales como:

- Alcance del área (usuarios, accesos)
 - Nivel de detalle (granularidad específica para cada entidad)
-

- Reportes preliminares (basados en las necesidades de información más cotidianas)
- Modelo de datos del Datawarehouse
- Elementos específicos

Con estos criterios previos al diseño formal, puede realizarse el establecimiento de una infraestructura que soporte de manera efectiva el Datawarehouse planificado.

1.3.9. Cambios en la Estructura de Poder en la Organización

Muchas organizaciones encaran el cambio de sistemas como una mera cuestión tecnológica, poniendo en éste todas las expectativas que supuestamente genera. Es así como frecuentemente se invierten muchos esfuerzos (no solamente recursos económicos) que generan frustraciones y desencantos, al percibir realmente que los resultados buscados no se consiguen o están lejos de conseguirse.

Cabe entonces preguntarse por qué se presenta esta disociación entre las expectativas y las reales posibilidades surgidas de un proyecto de sistemas.

Y conviene aquí hacer algo de historia. Según Mohapatra (2002), consultor de Synergytech³³, hace 30 años la inversión en tecnología informática era muy importante (se hablaba de ciento de miles de dólares, o aún millones). Por ello todo proyecto de sistemas era exhaustivamente evaluado, para lo cual se elaboraba un Plan de Sistemas.

³³ Synergytech es una empresa que desarrolla soluciones de Inteligencia de Negocios a sus clientes. Más información en www.synergytech.com

Mohapatra (2002) comenta que “el plan de sistemas partía de los requerimientos funcionales de la organización y de la estrategia de negocios de la empresa”. De esta forma ambos conceptos debían alinearse para focalizar la búsqueda en sistemas (de biblioteca o desarrollos) en una tecnología que se adecue a las necesidades detectadas. A partir de la búsqueda, se elaboraba un documento explicativo de las características funcionales y tecnológicas requeridas. A ello se unía una estimación de su costo/beneficio y cronograma de implantación. Todo ello configuraba el denominado Plan de Sistemas.

A partir del auge de las PC's³⁴, los proyectos de sistemas se simplificaron en demasía y, lo que antes constituía una guía para el desarrollo del proyecto, hoy se plantea simplificada como una cuestión meramente tecnológica. A ello cabe agregar la invasión de profesiones de formación tecnológica, en los que los principios organizacionales y de cambio son totalmente ignorados. En esta línea, proyectos culturales como el desarrollo de un Tablero de Comando se reducen a la implementación de un Datawarehouse. Es cultural porque las variables definidas para el tablero tienen su origen

³⁴ PC, Personal Computer. Computadoras Personales

en el proceso de toma de decisiones y éste, a su vez, en el modelo (mental) instalado en la empresa de acuerdo a la etapa de evolución organizacional en la que se encuentre. La tecnología se circunscribe en este caso sólo a la captación de la información referente a las variables de control, para luego ser procesado y presentado en forma de índices o graficados en cuadros de tendencia. De esto surge entonces que un proyecto de sistemas no debe ser considerado como un proyecto de innovación tecnológica, sino fundamentalmente como un proyecto de cambio cultural.

Mohapatra (2002) afirma que “toda organización cuenta con una cultura operativa instalada, entendiendo por tal a todas las prácticas instauradas para el manejo del negocio”. Muchas prácticas pueden ser sistematizadas a partir de computadoras, otras manualmente y muchas son de existencia implícita en las actividades del negocio. La lista puede ser larga, pero pueden individualizarse dos tipos de preguntas, según Mohapatra (2002):

1. Aquellas que hacen al control de las operaciones de la empresa
 2. Aquellas referidas al proceso de toma de decisiones.
-

Las primeras están relacionadas con los circuitos administrativos instalados, como: Compras, Stock, Cuentas por Pagar, Pagos, Ventas, Cuentas a Cobrar, Cobranzas, Contabilidad General, Tesorería, Bancos, Manufactura, etc.

Las segundas son propias de cómo la empresa toma sus decisiones. Estas se integran verticalmente y son en gran parte captadas en el momento de operar los circuitos administrativos.

En efecto, Softgal (2002) muestra una visión diferente, pero por ello no menos importante, considerar a una organización como “una relación de poder estructurada entre los componentes de la misma. Esta estructura de poder genera normalmente una desproporcionada acumulación de funciones y responsabilidades; funciones que juntamente con sus propias responsabilidades serán reformuladas en la implantación de una nueva solución de sistemas”. Es por ello que todo proyecto de sistemas deberá considerar el impacto que tendrá sobre la estructura de poder de la organización, siendo consciente (uno mismo) que el desequilibrio se generará y derivará en conflictos

de luchas hasta lograr nuevamente equilibrarse en un nuevo lugar.

De lo anterior surge entonces que tomar un proyecto de sistemas como una cuestión meramente tecnológica es, por lo menos en principio, una terrible exageración. Si así se lo considera se corre el riesgo de que el proyecto fracase, o por lo menos, que las expectativas generadas no se equiparen a las posibilidades de éxito.

1.3.10. *Aspectos a Resolver*

En el desarrollo de un proyecto de datawarehouse surgen varias interrogantes en torno a las expectativas y riesgos creados por lo que es importante conocerlos para poder monitorearlos, según Dyché (2001), y son:

- *Expectativas de los usuarios*

Muchas veces el éxito depende de la diferencia entre lo que los usuarios esperan y lo que ellos perciben que les es entregado. Es crítico que el equipo de Datawarehouse comunique las expectativas acerca de lo que será entregado muy claramente y ayude al usuario final a entender la naturaleza iterativa de construir un Datawarehouse.

- *Experiencia con Datawarehouses.*

Este riesgo se puede reducir con el uso juicioso de experiencias de proveedores y consultores.

- *Dirección estratégica*

Es relativamente lógico definir un punto de inicio lógico para el Datawarehouse.

CAPÍTULO II

JUSTIFICACIÓN DEL DATAWAREHOUSE

2.1. Metas y Objetivos

2.1.1. Metas y Objetivos del Negocio

Las siguientes son algunas de las metas institucionales de la Escuela Superior Politécnica del Litoral y sirven de justificativo válidos para la implementación del sistema planificado.

“La administración politécnica se ejecutará bajo los modernos principios de gestión, descentralización y desconcentración”,

“La situación financiera institucional se sustentará en la autogeneración de recursos y en los réditos de los fondos patrimoniales”,

“El ‘cliente interno’ es una preocupación permanente en la ESPOL. Si nuestros estamentos realizan sus actividades académicas y administrativas en un ambiente idóneo, los resultados serán positivos. Crear y desarrollar este ambiente es un deber institucional”.

Estos estatutos son parte de la Visión Administrativa Académico - Financiera de la ESPOL hasta la fecha en la que se desarrolló esta tesis. El mejoramiento continuo, eficiente y acorde con los cambios tecnológicos es una meta propuesta.

2.1.2. Metas y Objetivos del Proyecto

El objetivo de este proyecto es proveer a la ESPOL innovación de la Tecnología de Información dentro de un ambiente de Datawarehousing que le permita hacer un uso más óptimo de la información, como un ingrediente clave para un proceso de toma de decisiones más efectivo, con soporte al procesamiento analítico e informático sobre una amplia perspectiva de tiempo provisto por una sólida plataforma a partir de datos históricos.

Se termina este apartado, resumiendo los beneficios que un Datawarehouse puede aportar:

- Poner tanta información de interés general y comercial como sea posible en manos de tantos usuarios diferentes como sea posible.
 - Mejorar el tiempo de espera que se consumen en la generación de los informes habituales.
 - Monitorear el comportamiento de los clientes internos.
 - Producir un aumento justificado de la productividad.
-

- Incrementar y distribuir las responsabilidades.
 - Proporcionar una herramienta que permita tomar las decisiones en cualquier área funcional, basándose en información integrada y global del negocio.
 - Facilitar la aplicación de técnicas estadísticas de análisis y modelamiento para encontrar relaciones en los datos del almacén; obteniendo un valor agregado para el negocio de dicha información.
 - Proporcionar la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
 - Simplificar dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
 - Supone una optimización tecnológica y económica en entornos de Centro de Información, estadística o de generación de informes con retornos de la inversión considerables.
-

2.2. Alcance

Construir un datawarehouse es una gran tarea. No es recomendable emprender el desarrollo del datawarehouse de la empresa como un proyecto cualquiera. Más bien, se recomienda que los requerimientos de una serie de fases se desarrollen e implementen en modelos consecutivos que permitan un proceso de implementación más gradual e iterativa.

Los datos en el datawarehouse no son volátiles y es un repositorio de datos de sólo lectura (en general). Sin embargo, pueden añadirse nuevos elementos sobre una base regular para que el contenido siga la evolución de los datos en la base de datos fuente, tanto en los contenidos como en el tiempo.

Uno de los desafíos de mantener un datawarehouse, es idear métodos para identificar datos nuevos o modificados en las bases de datos operacionales. Algunas maneras para identificar estos datos incluyen insertar fecha/tiempo en los registros de base de datos y entonces crear copias de registros actualizados y copiar información de los registros de transacción y/o base de datos diarios.

Estos elementos de datos nuevos y/o modificados son extraídos, integrados, transformados y agregados al datawarehouse en pasos periódicos programados. Como se añaden las nuevas ocurrencias de datos, los datos antiguos son eliminados. Por ejemplo, para el caso específico de la carga de calificaciones de estudiantes existen situaciones en las que se realizan recalificaciones y/o modificaciones en los valores de las notas, al ser el datawarehouse una fuente de datos confiable, dichos cambios deben trasladarse desde los sistemas transaccionales hacia el repositorio depurado. Para esto se ha implementado cargas totales de los datos cada cierto tiempo (que depende de la volatilidad de los datos) en los que se realiza un vaciado de datos y repoblado de la base de datos.

Adicionalmente todos los registros en los sistemas operacionales cuentan con un campo informacional del tiempo de cambio (ULTIMO_CAMBIO), el cual es consultado por procesos de escrutinio periódicos que cargarán en el datawarehouse los valores recientemente modificados.

2.2.1. Definición del Alcance

El alcance de un datawarehouse puede ser tan amplio como toda la información estratégica de la empresa desde su inicio, o puede ser tan limitado como un datawarehouse personal para un solo gerente durante un año.

En la práctica, en la amplitud del alcance, el mayor valor del datawarehouse es para la empresa y lo más caro y consumidor de tiempo es crear y mantenerlo.

Como consecuencia de ello, la mayoría de las organizaciones comienzan con datawarehouses funcionales, departamentales o divisionales y luego los expanden como usuarios que proveen retroalimentación. Es por esto que este proyecto contempla la cobertura de una de las áreas de mayor impacto en la Institución como lo es la Académica. Dentro de este ámbito se crearán modelos de datos iniciales (cubos de información³⁵) que intenten resolver las necesidades de información más críticas y consultadas actualmente.

³⁵ Un Cubo de información representa el agrupamiento de un tema del negocio con todas las características relacionadas.

2.2.2. Productos del Proyecto – Lista de Entregables

La Tesis abarca la planificación, diseño e implementación de este tipo de tecnologías dentro de una organización como la ESPOL.

El desarrollo de las herramientas de procesamiento, resolución de operaciones agregadas (sumarización y promedios) y consultas de información están contempladas en este proyecto.

El incremento paulatino de las áreas de alcance como parte de la metodología de implementación de un Datawarehouse será considerado ya que inicialmente se escogió el área Académica, pero que se extenderá al ámbito de Personal y Financiero.

- **DW-Security** (Administración de Usuarios, Operaciones, Metadata)
 - **DW-Feeder** (Módulo del DW-Security. Contiene información de procesos de carga, tablas dimensionales, tablas facts, frecuencias, etc.)
 - **DW-OLAP** (Crea y procesa los cubos de datos)
 - **DW-Consult** (Reporteador Web para consultar el E-DW)
-

- **Scheduler** (Configuración de la Base de Datos del DW para procesos de carga)
 - **Templates** (Plantillas de Reportes Predefinidos)
-

2.2.3. Hitos del Proyecto

En este apartado se muestran los hitos más importantes en el desarrollo del proyecto de Datawarehouse de la ESPOL según las etapas involucradas en este proceso. El detalle a continuación:

1. Fase de requerimientos y administración del Proyecto
 - a. Planificación y administración del Datawarehouse
 - b. Plan de desarrollo de la comunicación
 - c. Definición de requerimientos de usuarios
 2. Diseño de datos
 - a. Modelamiento multidimensional
 - b. Análisis de las fuentes de datos
 - c. Creación del modelo de negocio y estándares de nomenclatura
 - d. Creación de Grupo de definición de seguridad de usuarios
 - i. Definir tareas y objetivos técnicos de seguridad
 - ii. Implementar medidas de seguridad de orden táctico.
-

iii. Desarrollar un Plan de Seguridad
Estratégico

3. Construcción del Datawarehouse

- a. Construir el esquema estrella en ERwin³⁶
- b. Diseñar y desarrollar los estados de los datos
- c. Poblar y validar la base de datos

4. Construcción de Herramientas de Consulta (DW-Consult)

- a. Especificaciones de la aplicación
- b. Desarrollo de la aplicación
 - i. Prototipo inicial
 - ii. Modelo avanzado
 - iii. Revisión Final
- c. Capacitación sobre la herramienta

5. Distribución e instalación

- a. Planificación de Instalación
- b. Evaluación general del sistema
- c. Entrenamiento coordinado
- d. Distribución de cuentas de acceso

6. Revisión del Datawarehouse

- a. Revisión por parte del comité promotor
- b. Reportes

³⁶ ERwin: Herramienta CASE de modelamiento de bases de datos relacionales y multidimensionales, distribuido por LogicWork.

2.2.4. Áreas del Negocio involucradas

Inicialmente se incluirán las áreas o departamentos con mayores expectativas y problemas que se evidenciaron en la fase de recopilación de requerimientos y necesidades, como lo son: Administración General y CRECE³⁷ .

Cada una de éstas a su vez está integrada por módulos o sectores que determinan actividades o procesos en común dentro de cada departamento. Así:

- Administración académica
- Unidades académicas
- Vicerrectorados
- Bienestar Estudiantil
- Profesores
- Estudiantes

³⁷ CRECE es el Centro de Registros y Calificaciones de Estudiantil de la ESPOL.

2.3. Asunciones

2.3.1. *Asunciones del Proyecto*

Hay algunas consideraciones adicionales que deben tenerse en cuenta al construir y administrar el datawarehouse. La primera consideración es respecto al índice. La información de los niveles de esquematización más altos puede ser libremente indexada, mientras que las de los niveles más bajos de detalle, por ser tan voluminosa, pueden ser indexadas moderadamente.

Por lo mismo, los datos en los niveles más altos de detalle pueden ser reestructurados fácilmente, mientras que el volumen de datos en los niveles más inferiores es tan grande, que los datos no pueden ser fácilmente reestructurados.

Por consiguiente, el modelo de datos y el diseño clásico fundamentan que el datawarehouse se aplique casi exclusivamente al nivel actual de detalle. En otras palabras, las

actividades de modelamiento de datos no se aplican a los niveles de esquematización, en casi todos los casos.

Otra consideración estructural es la partición de la información en el datawarehouse. El nivel de detalle actual es casi siempre particionado.

La partición puede hacerse de dos maneras: al nivel de DBMS y al nivel de la aplicación. En la partición DBMS, se conoce las particiones y se administra por consiguiente. En el caso de la partición de las aplicaciones, sólo los programadores de las mismas conocen las particiones y la responsabilidad de su administración es asignada a ellos.

Al interior de las particiones DBMS, mucho de los trabajos de infraestructura se hacen automáticamente. Pero existe un elevado grado de rigidez asociada con la gestión automática de las particiones. En el caso de las particiones de las aplicaciones del datawarehouse, la mayor parte del trabajo recae sobre el programador, pero el resultado final es que la gestión de datos es más flexible.

2.4. Limitaciones

2.4.1. *Limitantes del Proyecto*

Mientras que los componentes del datawarehouse trabajan de acuerdo al modelo descrito para casi todos los datos, hay pocas excepciones o limitantes útiles que necesitan ser discutidos.

Una de ellas es la *data resumida pública*, que es la data que ha sido calculada fuera del datawarehouse pero es usada a través de la empresa. La data resumida pública se almacena y administra en el datawarehouse, aunque su cálculo se haya hecho fuera de él.

Un ejemplo clásico de data resumida pública es el archivo trimestral hecho por cada compañía pública. Los contadores trabajan para producir cantidades como rentas trimestrales, gastos trimestrales, ganancias trimestrales y otros. El trabajo hecho por los contadores está fuera del datawarehouse. Sin embargo, esas cantidades referenciales producidas por ellos se usan ampliamente dentro de la corporación para marketing,

ventas, etc. Una vez que se haya hecho el archivo, los datos se almacenan en el datawarehouse.

Otra excepción no considerada en este documento es la *data externa*.

Otro excepcional tipo de datos a veces encontrados en un datawarehouse es el *detalle de los datos permanentes*, que resulta de la necesidad de una empresa para almacenar la data a un nivel detallado permanentemente por razones éticas o legales.

El detalle de datos permanentes comparte muchas de las mismas consideraciones como otro datawarehouse, excepto que:

- El medio donde se almacena la data debe ser tan seguro como sea posible.
 - Los datos deben permitir ser restaurados.
 - Los datos necesitan un tratamiento especial en su indexación, ya que de otra manera los datos pueden no ser accesibles aunque se haya almacenado con mucha seguridad.
-

2.4.2. *Proyectos Relacionados*

Como el depósito evoluciona y los datos que contiene llegan a ser más accesible, los empleados externos al depósito podrían descubrir también el valor de sus datos. Al enlazar el datawarehouse a otros sistemas (tanto internos como externos a la organización), se puede compartir información con otras entidades comerciales con poco o sin desarrollo. Los mensajes E-mail, servidores Web y conexiones Intranet/Internet, pueden entregar listas por niveles a los proveedores o según la condición, a los socios de negocio.

Como los datawarehouses continúan creciendo en sofisticación y uso, los datos acumulados dentro de una empresa llegarán a ser más organizados, más interconectados, más accesibles y, en general, más disponibles a más empleados.

El resultado será la obtención de mejores decisiones en el negocio, más oportunidades y más claridad de trabajo.

Teniendo claro que el primer paso para implantar soluciones de análisis de negocio es garantizar la calidad de la información en

su origen. Sólo con información consolidada, normalizada, no duplicada y consistente se puede medir, controlar y gestionar adecuadamente el negocio.

Por esta razón, el proceso de depuración de los datos pasa por la utilización de una herramienta de Extracción, Transformación y Carga de datos (ETL) y por la construcción de un almacén para ellos (DataWareHouse) o DataMart, que permitirá estructurar los datos de forma sencilla, pensando en el rendimiento, rapidez y facilidad de uso.

Otras soluciones de Inteligencia de Negocios que pueden ser tomados como proyectos relacionados son:

- La implantación de herramientas de soporte a la toma de decisiones para la realización de consultas, informes y análisis
 - Sistemas de Información para Ejecutivos, como los cuadros de mando, donde se recogen los indicadores necesarios para el control y gestión de la empresa
-

- Los Balanced ScoreCard³⁸ que le permitirán implantar su estrategia de negocio
- La construcción de herramientas de Data mining³⁹ estaría soportada por una plataforma con datos limpios y fácilmente consultables.

³⁸ Balanced ScoreCard: Es una metodología que permite implementar la estrategia y la misión de una Organización a partir de un conjunto de Objetivos Estratégicos.

³⁹ Data mining :“Extracción de información oculta y predecible de grandes bases de datos”. Predicción automatizada de tendencias y comportamientos. Descubrimiento automatizado de modelos previamente desconocidos.

2.4.3. Dependencias Críticas

Hay muchas maneras para desarrollar datawarehouses como tantas organizaciones existen. Sin embargo, hay un número de dependencias diferentes que necesitan ser consideradas:

- Alcance de un datawarehouse
- Redundancia de datos
- Tipo de usuario final

La figura 5 muestra un esquema bidimensional para analizar las opciones básicas. La dimensión horizontal indica el alcance del depósito y la vertical muestra la cantidad de datos redundantes que deben almacenarse y mantenerse.

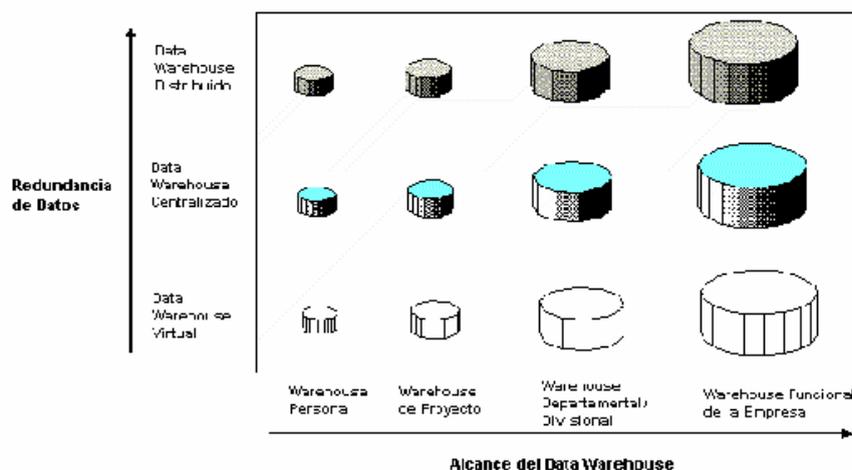


Figura 5 : Esquema Bidimensional Alcance vs. Redundancia de Datos
Fuente: <http://www.inei.gob.pe/cpi-mapa/bancopub/libfree/lib619/132.HTM>

2.5. Propuesta de Valor Comercial del Datawarehousing

2.5.1. Retorno de Inversión

En datawarehousing el cálculo del retorno de inversión es una actividad muy debatida pero poco practicada. Con aplicaciones operacionales tradicionales tales como sistemas de facturación o sistemas de reserva, un equipo de desarrollo tiene metas para cumplir que han sido delineadas desde el principio y son relativamente sencillas y medibles. Es raro que un nuevo sistema transaccional se construya sin hacer antes algún tipo de cálculo financiero.

Pero el costo de un sistema de soporte de decisiones es mucho más difícil de justificar por adelantado. Por un lado, las áreas y procesos comerciales a los que estos sistemas apuntan están en constante transición, y es rutinario que sus objetivos cambien a mitad del río. Por otro lado, aunque el soporte de decisión puede a la larga producir miles de millones de dólares en ahorros directos o aumento de ingresos, la empresa no suele reconocer esto como algo crítico. Más aún, prácticamente existen criterios establecidos para medir el éxito del soporte de decisión. En un inicio, suele ser algún aislado gerente visionario

el que puede ver los beneficios que el datawarehousing podría darle a su compañía y el que se da cuenta de que esos beneficios son solamente la punta de un iceberg.

El hecho de que el datawarehouse y la correspondiente e-data⁴⁰ mejoren la capacidad de toma de decisiones y pongan información acerca de los clientes a disposición de sus usuarios debería parecer suficiente justificación. Sin embargo, algunas compañías siguen pidiendo pruebas concretas, especialmente a la luz de los frecuentemente publicitados fracasos del datawarehousing. Algunos ejecutivos – no visionarios, pero sí pragmáticos con la vista puesta en los resultados – no gastarán ni un centavo en un datawarehousing a menos que vean en él la promesa de un rédito medible. Esto es más cierto aún cuando el datawarehouse se presenta internamente como una tecnología en vez de cómo una solución comercial. De hecho el uso del término “datawarehouse” en lugar de los otros de corte más comercial, “soporte de decisión” o “inteligencia comercial”, hace que sea aún más probable que la gerencia exija tener estimaciones de retorno de inversión, aunque el resultado final siga siendo el mismo.

⁴⁰ E-data es un término de inteligencia de negocios empleado por Jill Dyché (2001) y que se refiere a los datos que han sido extraídos, limpiados, procesados y almacenados en un datawarehouse.

Sin embargo, pocos de estos ejecutivos consideran el costo de oportunidad del datawarehousing, es decir, qué ganancias financieras podría producir la misma inversión si fuera aplicada en otra parte, es decir, en un nuevo sistema de distribución.

Ya se trate de un datamart o de un datawarehouse, todo tipo de evaluación de beneficios debería distinguir entre “retorno de inversión duro” (hard ROI), es decir, ahorros y aumentos de ingresos tangibles, y “retorno de inversión blando” (soft ROI), por ejemplo mayores niveles de satisfacción, aumento del prestigio, y mayor a pesar los costos contra los beneficios financieros generales, el retorno de inversión blando, aunque más difícil, es igual de crítico y suele tener un impacto mayor.

2.5.1.1. Beneficios Tangibles

A veces, calcular el retorno de inversión duro de un datawarehouse es fácil. La proyección en términos amplios del costo del datawarehouse versus su rédito puede ser fácil.

Ya sea que se trate de determinar el retorno de inversión para un nuevo datawarehouse o para una nueva aplicación en un datawarehouse existente, lo primero podría ayudar a mejorar el negocio, o lo que se podría llamar sweet spots del datawarehousing, los puntos justos. Estos puntos juntos pueden ser procesos comerciales, nuevas iniciativas corporativas, o áreas funcionales tales como la gestión de campaña o el soporte de comercio electrónico.

El datawarehousing produce dos beneficios principales:

- Ahorros directos, reflejados en reducción de costos de producción, de distribución, de personal y de suministros.
 - Aumentos en los ingresos, a través de la venta cruzada, ajustes de precios para incrementar los
-

márgenes, adquisición de nuevos clientes, o cancelación de productos o servicios. Aunque esta última estrategia puede implicar disminución de ingresos debido a la pérdida de las ventas del producto cancelado, también puede implicar un aumento de ganancias debido al ahorro de costos asociado con no tener que desarrollar el producto u ofrecerlo en el mercado. Aquí la meta es saber qué productos ya no justifican el gasto que tienen asociado.

La estrategia de reducción de costos suele ser la primera que practican las organizaciones que se inician en datawarehousing. La premisa misma de mejorar las cuentas finales sugiere administrar con más eficiencia procesos comerciales tales como el control de inventario, los costos por proveedores, o la publicidad. El beneficio del datawarehousing es que permite hacer una mirada histórica sobre estos procesos.

2.5.1.2. Beneficios Intangibles

Es difícil ponerle valor monetario a mejoras en la calidad de los datos o más atribuciones para los usuarios. ¿Son cosas que hacen a los hombres de negocios más productivos o las operaciones más eficientes? ¿Tuvieron un impacto positivo o negativo en las ventas? ¿Y cómo se miden realmente estos factores?

Bueno, quizá no haya necesidad de hacerlo, sobre todo si el datawarehouse en cuestión produce un retorno de inversión duro demostrable. Pero los datawarehouses también pueden producir mejoras “blandas”, cuyo impacto probablemente sea tan efectivo como una mejora en los resultados económicos. Se examinan algunas de ellas en más detalle a continuación:

Mejoras en la calidad de los datos

Un datawarehouse mejora la calidad de los datos operacionales al obligar a depurar los datos antes de su carga. El hecho de que los datos que llegan al datawarehouse sean de mayor calidad implica que los

hombres de negocios que los usen pueden confiar más en ellos. Este aumento de la confiabilidad de los datos tiene el beneficio adicional de que los usuarios finales puedan tomar sus decisiones con más seguridad.

Accesibilidad de la información comercial

Por definición, un datawarehouse consolida información separada de una variedad de fuentes diferentes en un solo lugar. En otras palabras, libera a los diversos usuarios de la necesidad de ir a múltiples lugares para hallar la información que necesitan, ya que les permite encontrar los datos comerciales fundamentales en un solo lugar.

Ya sea que la plataforma en sí misma esté centralizada o que esté distribuida en varios servidores, el hecho de que los datos estén a disposición de los usuarios de negocios en una forma consistente puede resultar en un aumento de la producción y un ahorro de costo incalculables.

Optimización de la cadena de suministros

Un datawarehouse puede optimizar la cadena de suministros, parcial o totalmente. Por ejemplo, el personal de reparto puede ingresar información acerca de los puntos de entrega en un dispositivo manual que transmita los datos a las oficinas centrales de la compañía para su eventual seguimiento mediante el datawarehouse. O quizá el análisis de secuencias de clics puede indicarle a un e-tailer (un comercio electrónico) que encontrar un determinado producto lleva demasiado tiempo y que el producto tendría que estar puesto en un lugar del sitio web donde fuera más accesible.

Un subproducto de la mejora de la cadena de suministros es que los vendedores probablemente se volverán más eficientes, lo que redundará en un mejor servicio y hasta en descuentos para los clientes.

Satisfacción de los empleados

Ya se habló acerca del potencial cultural de un datawarehouse. Los usuarios de negocios, probablemente acostumbrados a esperar semanas o meses para recibir información, ahora pueden obtenerla con un simple clic del ratón. Pero más allá de eso, los empleados que antes de que existiera el datawarehouse solamente estaban implicados en distribución de información – armando informes para la gerencia o engorrosas hojas de cálculo – ahora pueden hacerse cargo de tareas adicionales y, en muchos casos, actuar realmente sobre el contenido de esa información por sí mismos en vez de tener que dispersarla a una y otra parte para su revisión.

Satisfacción de los clientes y servicio al cliente mejorado

El empleado de un centro de atención que puede ver información acerca de un cliente mientras lo tiene en línea probablemente no sepa que lo que se le provee es el datawarehouse.

La mayoría de las compañías que adquieren un datawarehouse lo hacen para reducir costos y en definitiva producir ingresos.

La satisfacción de los usuarios se puede medir cuando las compañías pueden predecir situaciones futuras dentro del negocio, en el caso del área académica de la ESPOL, poder predecir cuántos estudiantes se van a registrar en una materia para planificar la cantidad de paralelos necesarios es un indicador de satisfacción a los clientes.

Imagen de liderazgo tecnológico

Este beneficio muchas veces olvidado puede significar realmente la diferencia entre un líder reconocido de la industria y un adoptante rezagado. Muchas compañías antes de adquirir tecnología estratégica miran a sus colegas de la industria que ejemplifican “mejores prácticas”, y en cada industria hay unos pocos líderes reconocidos que estuvieron a la vanguardia del datawarehousing.

2.5.2. Presupuesto del Datawarehouse

Elaborar un esquema de costos para un datawarehouse es más complejo que limitarse a superponer las guesstimates⁴¹ típicas en informática respecto a los costos de hardware y las licencias de las bases de datos. Un presupuesto típico incluye los costos de personal y costos permanentes de infraestructura, así como los costos de hardware y software.

Por supuesto, antes de empezar a evaluar el costo de un datawarehouse se debería responder a algunas preguntas generales:

- ¿Qué tan grande sería este datawarehouse?
- ¿Cuánta gente hará falta para mantenerlo?
- ¿Cuántos usuarios tendrá?
- ¿Qué habilidades tiene el personal interno de la compañía?
- Aproximadamente, ¿cuántos datos habrá a corto y largo plazo?
- ¿Dónde residirá físicamente el datawarehouse?

⁴¹ Mitad estimación, mitad adivinanza.

2.5.2.1. Evaluación de Costos de Tecnologías

La tabla 6 muestra una estimación típica de costos previstos, elemento por elemento.

El mayor error que se comete al elaborar presupuestos para tecnología es ignorar los costos de infraestructura, cosas tan mundanas como las estaciones de trabajo en los escritorios de los usuarios finales o las actualizaciones de PC. En muchos casos, para crear aplicaciones de inteligencia comercial primero habrá que adquirir o mejorar tecnologías. El costo de adquirir o actualizar tecnologías periféricas debería ser incluido dentro de la estimación general de costos, a fin de encontrarse con sorpresas desagradables durante la implementación.

ESTIMACION DE COSTOS PARA TECNOLOGIA	
Tecnología	
Hardware para el datawarehouse	
2 CPU / 1GB de memoria, 70 GB de RAID	\$ 25.000
Mantenimiento y actualizaciones al 17%	\$ 4.250
Licencia del software para el DBMS	
150 copias a \$100/copia (usuarios concurrentes)	\$ 150.000
Mantenimiento y actualizaciones al 15%	\$ 22.500
Herramienta de Aplicación	
150 copias a \$87/copia	\$ 131.250
Soporte al 15%	\$ 19.688
Puerta de Enlace para las consultas	\$ 50.000
Soporte al 15%	\$ 7.500
Actualizaciones de las estaciones de trabajo (memoria) \$ 250 por usuario , para 100 usuarios	\$ 25.000
Herramientas de Desarrollo	
Suite de Administración de datos (modelado, transformación/metadatos)	\$ 100.000
Actualización de software de control de versiones	\$ 8.000
Herramientas varias de administración de sistemas	\$ 50.000

Tabla 6: Estimado de Costos para la Implementación del Datawarehouse

Autor: José Rodríguez Rojas

2.5.2.2. Estimación de Costos de Recursos

Lo más importante a tener en cuenta respecto de la estimación de costos de recursos es que los recursos en personal cambian con mucha más frecuencia que la tecnología, así que una estimación de este tipo debería tener en cuenta todas las actividades de desarrollo que se tenga previsto hacer durante el año. Cada nueva aplicación de datawarehouse requerirá personal adicional, ya se trate de programadores, analistas comerciales o cargadores de datos.

La tabla 7 muestra un sencillo ejemplo de los costos de recursos humanos a lo largo de un año. Observe que el presupuesto incluye tanto a los miembros internos del personal como los consultores externos, y que también tiene en cuenta las actividades de recolección de requerimientos y entrenamientos.

Una vez estimados los costos del datawarehouse, la tarea de determinar el retorno de inversión puede estar en cualquier lugar entre lo fácil y lo imposible. La facilidad

es directamente proporcional a los datos respecto de gastos que ya estén disponibles. Si está buscando una fórmula a grandes trazos pero le falta información sobre los gastos, le aguarda una gran sorpresa: Es imposible determinar los futuros ahorros de costo a menos que se entienda lo que se gasta actualmente. Y tampoco es realista medir los costos en relación con beneficios blandos tales como “mejores pronósticos de venta” o “campañas promocionales más exitosas”.

ESTIMACION DE COSTOS DE RECURSOS	
Recursos	
Internos* (costos reales internos)	\$ 128.000
Director de proyecto a 80/hora	\$ 120.000
1 DBA a 75/hora	\$ 120.000
Administrador de datos a 75/hora	\$ 96.000
Administrador de sistema a 60/hora	\$ 14.400
Administrador de red a 60/hora por 6 semanas (esporádicos)	\$ 12.000
Modelador de datos a 75/hora por 4 semanas	\$ 12.000
Programador senior a 75/hora	\$ 120.000
Programador a 60/hora	\$ 96.000
Cargador de datos a 60/hora	\$ 96.000
Total recursos internos	\$ 802.400
Externos*	
Consultor en management: análisis comercial, 2 a 250/hora por 6 semanas	\$ 120.000
Consultor del proveedor de la herramienta de aplicación: 185/hora por 4 semanas	\$ 29.600
Diseñador de bases de datos: 175/hora por 3 semanas	\$ 21.000
Programador SQL: 150/hora por 4 semanas	\$ 24.000
Programadores a contrato: 3 a 125/hora por 4 meses	\$ 240.000
Total recursos externos	\$ 434.600
Total recursos	\$ 1'237.000
*Empleados internos a 40 semanas/año, a menos que se indique lo contrario	
**Consultores tomando como base 42 semanas al año, a menos que se indique lo contrario	
Varios	
Entrenamiento en la aplicación (50 usuarios)	\$ 40.000
Total de costos del primer año:	\$ 2'723.438

Tabla 7: Inversión de Recursos para un Datawarehouse

Autor: José Rodríguez Rojas

CAPÍTULO III

METODOLOGÍA DE IMPLEMENTACIÓN DEL DATAWAREHOUSE

3.1. Metodología de Desarrollo

3.1.1. *Identificación de etapas del proyecto*

La metodología de implementación del datawarehouse escogida corresponde al modelo de Plan de Proyecto Incremental⁴², el cual tiene tres fases principales: (1) análisis comercial, (2) desarrollo de la base de datos, y (3) desarrollo de la aplicación.

Según Dyché (2001) estas fases aseguran que se le preste atención a cada área, que las fases –no los pasos- sean lineales, y que la salida de una fase sirva como entrada para otra. A continuación se examina cada fase por separado:

1. Análisis comercial: Esta fase asegura que la organización logre una comprensión cabal del o de los problemas de la empresa a cuya resolución está destinado el datawarehouse. El análisis comercial asegura que se reúnan, se comprendan, se documenten y se autoricen los requerimientos antes

⁴² Metodología de Desarrollo de sistemas basado en un crecimiento paulatino de las áreas abarcadas

de dar comienzo formal a la implementación. Más aún, esta fase asegura que todas las piezas estén en su lugar antes de comenzar el desarrollo.

2. Desarrollo de la base de datos: Como la base de datos es el cimiento del datawarehouse, el objetivo de esta fase es imponer el rigor necesario a la selección, el diseño, la extracción, la transformación, y la carga de los datos. También asegura que el diseño de la base de datos se realice en forma separada de la aplicación, lo que resulta en el mejor acercamiento, en vez de solo el más fácil.
 3. Desarrollo de la aplicación: Esta fase garantiza que el desarrollo de las herramientas usadas para acceder a los datos sea tal que éstas puedan coexistir y adaptarse sin fisuras a la arquitectura general del datawarehouse. Además, impone la realización de tareas específicas relacionadas con la programación y distribución de la aplicación del datawarehouse a los usuarios finales. (Hasta los profesionales experimentados suelen olvidarse de la prueba, distribución y entrenamiento en el uso de la aplicación).
-

LOS PASOS DEL ANALISIS COMERCIAL

Los cinco pasos siguientes que forman la fase de análisis comercial de la metodología del proyecto:

1. *Exploración*: Este paso asegura que el datawarehousing sea realmente la respuesta correcta a las preguntas de la compañía. Durante la exploración, una compañía puede decidir realizar una cierta cantidad de actividades diferentes que quizás no estén directamente relacionadas con el datawarehousing, pero que de todos modos ofrecen indicios acerca de si el datawarehousing es la solución a la que le llegó su momento. Estas actividades pueden incluir comparaciones con empresas líderes, o *best-practice benchmarking*, tareas de alineación estratégica de tecnología, rediseño de procesos de la empresa, o la fijación de criterios de evaluación equilibrados.
 2. *Descubrimiento comercial*: El descubrimiento comercial es la examinación a alto nivel de la necesidad, la inquietud, o el problema relativo, documentando los motores comerciales y los requerimientos de alto nivel que pueden justificar un datawarehouse. Este proceso confirma que el problema que
-

la compañía trata de encarar con su datawarehouse realmente puede ser resuelto con e-data (en vez de, por ejemplo, impulsando mejoras en los procesos). Aunque en la mayoría de los casos el descubrimiento comercial se hace cuando se trata de datawarehouses empresariales, también los datamarts pueden disfrutar de los beneficios de esta actividad. El descubrimiento comercial identifica una jerarquía de necesidades y define los límites del proyecto de implementación inicial.

3. *Definición de alcance del proyecto*: Definir el alcance de un proyecto de datawarehouse debería ser un paso separado que se realice antes de la planificación del proyecto. Aunque para muchas organizaciones existe la tentación de tratarla como una mera formalidad, en realidad la definición del alcance ayuda a reducir los tiempos y los recursos específicos necesarios para el proyecto y define los resultados que se pueden esperar en un principio. Y lo más importante, la definición de alcance asegura que los miembros del personal estén listos antes de que empiece el desarrollo real. También ayuda a fundamentar las estimaciones presupuestarias.
-

4. *Planificación del proyecto y de la infraestructura:* Este paso prepara a la organización para el desarrollo de un datawarehouse, al asegurar que la metodología elegida se refleje en un plan de proyecto táctico, y que la infraestructura necesaria esté en su lugar lista para brindar soporte. Esto implica desde asegurar que las estaciones de trabajo para el desarrollo sean compatibles con los productos y las versiones del software.
 5. *Recolección de requerimientos:* Esta etapa abarca la creación de consenso, la realización de entrevistas, y a menudo sesiones de quejas. La recolección de requerimientos deja sentados los objetivos de decisión y funcionales para cada aplicación de soporte de decisión, sea grande o pequeña, y para muchos es la pieza clave de la implementación. La recolección de requerimientos puede definir objetivos, por ejemplo, cargar tres años de historial de ventas en el datawarehouse o mostrar resultados en forma de gráficos de torta. Identifica datos iniciales respecto de las áreas temáticas y debería incluir la documentación de objetivos y criterios de medición de éxito.
-

LOS PASOS DE ESTRUCTURACIÓN Y ADMINISTRACIÓN DE DATOS

Dada la afirmación inicial de que la base de datos está en el centro del datawarehouse, la base de datos debería ser comprendida e implementada antes que la construcción de la aplicación. La fase de desarrollo de la estructura y administración de la base de datos abarca los pasos siguientes:

1. *Arquitectura:* El trazado de la arquitectura implica no solamente evaluar y elegir la base de datos y las tecnologías relacionadas específicas sino también planificar cómo colaborarán esas tecnologías entre sí, y con los sistemas ya existentes de la compañía. Este paso asegura que el datawarehouse sea tanto funcional como compatible con la infraestructura tecnológica existente.
 2. *Diseño de la base de datos:* El diseño de la base de datos comienza con la implementación de la arquitectura, ya que el diseño inicial se concentrará en los requerimientos comerciales y luego evolucionará para incorporar los datos reales que haya disponibles. La tarea de diseñar la base de datos abarca tanto el desarrollo de un modelo de datos como un diseño físico que define la creación efectiva de las
-

tablas e implica ajustes al rendimiento. Las actividades de diseño también establecen el nivel de detalle.

3. *Extracción, transformación y carga:* El propósito de estos pasos es examinar el diseño físico de la base de datos para proceder a la extracción de los elementos de datos específicos, así como a la creación efectiva de las rutinas de transformación y carga. En los mejores casos, la extracción y carga de datos pueden ser realizadas usando utilidades de algún proveedor de bases de datos, lo que hace necesario programar poco o nada. Sin embargo, muchas compañías todavía dedican tiempo a codificar programas a medida para la extracción de los datos de sus diversos sistemas de origen. En cualquier caso, este paso nunca es tan fácil como parece a primera vista.
 4. *Implementación y Prueba:* Probablemente, el DBA (Administrador de bases de datos) cargará la o las bases de datos en el datawarehouse varias veces antes de la carga final. Se someterá la base de datos a prueba para asegurar que refleje con precisión las necesidades informativas de la empresa y puedan brindar soporte a sus procesos. En la parte de validación de datos de la prueba de la base de datos suelen también estar involucrados los usuarios finales.
-

Una vez terminada la prueba, puede tener lugar la carga final para comenzar la creación de la aplicación.

LOS PASOS EN EL DESARROLLO DE LA APLICACIÓN

Una vez probado el diseño de la base de datos, puede comenzar el desarrollo de la aplicación. Esta fase abarca los siguientes pasos:

1. *Arquitectura*: Así como el desarrollo de la base de datos implica la selección e integración de las tecnologías pertinentes, el trazado de la arquitectura de la aplicación implica la identificación de las tecnologías de desarrollo y soporte. Estas tecnologías no se limitan al software de aplicación que los usuarios finales usarán para hacer consultas; también incluyen las herramientas para sincronizar las versiones diferentes del software, ayudar a los programadores a integrar sus programas, y habilitar el acceso remoto, entre otras tareas. La arquitectura de la aplicación también implica probar las tecnologías propuestas con las tecnologías de bases de datos ya elegidas, para asegurar una integración sin fisuras.
 2. *Diseño*: Entre tantas responsabilidades de desarrollo, los programadores suelen olvidarse del diseño de la aplicación.
-

Esta tarea hace uso de los requerimientos recogidos, así como otros datos obtenidos de los usuarios finales, para evaluar de qué forma los usuarios esperan ver y usar la información en sus escritorios. Este paso ayuda a los programadores a definir técnicas de presentación de datos y asegura que la aplicación tenga una apariencia (look-and-feel) aceptable.

3. *Implementación*: La implementación de la aplicación corresponde a la programación real del software de aplicación y puede incluir revisiones del diseño, creación de prototipos y desarrollo de consultas.

4. *Prueba*: Un desarrollo completo de la aplicación requiere varios pasos de prueba:

- *Diseño y desarrollo de pruebas*
- *Prueba funcional*
- *Prueba de integración*
- *Prueba de aceptación de los usuarios*

La realización de pruebas, que involucra a programadores y a usuarios finales, emplea varios métodos de validación y autorización.

5. *Distribución y soporte*: La entrega del código de la aplicación y el acceso a la base de datos a los usuarios recibe el nombre de distribución de la aplicación (deployment⁴³). La distribución no implica poner la aplicación a disposición de un conjunto de usuarios una sola vez, constituye un enfoque iterativo de ajuste y corrección constante de las aplicaciones para hacerlas más funcionales y usables con el tiempo. Brindar soporte a las aplicaciones puede significar desde designar a un experto en la materia para ayudar a los usuarios finales a adoptar y a aprender la aplicación hasta armar una mesa de ayuda para proveer asistencia permanente. Esta etapa también tiene en cuenta el entrenamiento de los usuarios y la documentación de la aplicación.

⁴³ Término inglés que describe la distribución y difusión de un servicio o recurso

3.1.2. Determinar objetivos temporales

Los objetivos temporales (a priori) de esta implementación tienen relación con las áreas involucradas y las expectativas de información que tienen.

Como ya se mencionó, la disponibilidad de reportes y/o consultas que con regularidad son requeridos por los usuarios en un formato estándar, de fácil manipulación, portable⁴⁴ y de acceso general es algo crucial.

El poder contar con una plataforma sobre la cual poder realizar análisis estadístico por medio de herramientas de terceros y/o técnicas de agrupamiento especiales (dentro de hojas electrónicas, por ejemplo) es un objetivo de primera mano que puede ser alcanzado en esta primera fase del desarrollo del Datawarehouse.

⁴⁴ Portable se refiere al hecho de poder ser trasladable de un lugar a otro.

3.1.3. Determinar tareas y responsabilidades

El equipo de implementación debe tener una buena reputación con el departamento de TI⁴⁵ y los usuarios, una reputación ganada con el desarrollo exitoso de otros sistemas.

Los miembros de este sistema deben tener una buena reputación con los usuarios finales. Por otra parte, los usuarios deben dar el tiempo o las personas necesarias para realizar un proyecto exitoso.

Las definiciones de roles y responsabilidades son importantes solamente si todos estos son entendidos y asignados. Las asignaciones pueden ser fácilmente olvidadas o pasadas por alto. Las designaciones en esta sección son mostradas a manera de ejemplo. Las organizaciones pueden cambiar estas asignaciones y los nombres de los roles en función de su conveniencia interna, pero en este proceso se debería estar seguro de que todos los roles sean asignados y que las responsabilidades estén claramente entendidas y las responsabilidades primarias identificadas.

⁴⁵ TI son los departamentos o áreas de sistemas de Información, encargados de realizar el estudio, investigación, aplicación y seguimiento del software, hardware y la interconexión de redes.

Cada uno de estos roles requieren entrenamiento y experiencia. Un número de estos roles pueden aparecer en más de un área funcional, indicando un potencial para las responsabilidades conjuntas.

Adelman (1997) define los siguiente roles y responsabilidades que deben ser considerados en el desarrollo de un datawarehouse.

- Usuarios finales: Los usuarios finales tienen diferentes niveles de interés en sus requerimientos de consultas. Algunos de ellos están ansiosos de escribir sus propias consultas, mientras que otros utilizarán sólo los que se encuentran predefinidos.

Los usuarios finales deben aceptar las siguientes responsabilidades:

- Identificar requerimientos de Datos
 - Proveer definiciones de Datos.
 - Ser responsable en la determinación de seguridad, acceso, capacidad de actualización, nivel requerido de disponibilidad y desempeño.
 - Identificar prioridades para la identificación.
-

- Determinar la forma de los datos (detallados, resumizados, agregados, derivados e históricos).
 - Validar la confiabilidad de reportes y consultas producidas desde el Datawarehouse.
 - Identificar reglas de transformación desde los sistemas fuente.
- Arquitectura Estratégica: La arquitectura estratégica puede ser vista como un proyecto aparte del Datawarehouse y no necesariamente para la implementación inicial del mismo. La arquitectura estratégica debería tener también las siguientes responsabilidades:
- Determinar y validar las metas y objetivos del Datawarehouse y asegurarse de que ellas conforman las metas y objetivos de la empresa.
 - Desarrollar la arquitectura del Datawarehouse incluyendo estándares, productos y suites de herramientas.
-

- Establecer propiedades basadas en la dirección estratégica de la organización.
- Administrador de Proyectos de Datawarehouse: Hay quienes piensan que la tecnología es todo lo que se necesita para el éxito de un datawarehouse. Aquellos con un poco más de experiencia entienden que un Administrador de Proyectos es crucial. El Administrador de Proyectos debe tener experiencia y entender muy bien el negocio. Su participación dentro del proyecto incluye lo siguiente:
- Asumir totalmente las responsabilidades para el Datawarehouse.
 - Estimar los costos y beneficios.
 - Actuar como un revisor de calidad a lo largo del proyecto.
 - Ayudar con la planificación del proyecto, determinación de entregables y establecimiento de tiempos de entrega.
 - Planificar la integración de los datos con otros sistemas.
-

- Planificar la incorporación e integración de fuentes de datos externas.
 - Medir los beneficios, costos y proveer a la administración la información justificativa sobre los costos.
 - Realizar recomendaciones en vías de mejorar el Datawarehouse.
- Usuario de Enlace o Aliado: El usuario aliado debe estar relacionado con el equipo de TI y los usuarios finales. Su trabajo puede variar dependiendo de la experiencia personal. La mayoría de los usuarios de enlace exitosos usualmente tienen una combinación de ambos conocimientos (TI y departamentos). El usuario de enlace debe trabajar en estrecha relación con el equipo del proyecto en las cuatro mayores áreas de responsabilidad definidas por Adelman (1997).
1. Recopilación de los requerimientos de usuarios finales.
 - Identificar áreas de alcance (financiero, recursos humanos, servicio al cliente, datos regionales, etc.).
-

- Determinar el conjunto de clientes para cada área de alcance, quienes serán los que accederán a cada uno de ellas.
- Identificar nuevos requerimientos de datos.
- Determinar la forma de los datos.
- Validar la calidad de los datos.
- Identificar reglas de transformación para los datos.
- Identificar prioridades para la implementación.

2. Monitorear la implementación de la infraestructura del Datawarehouse.

- Proveer requerimientos del cliente para la selección de la arquitectura requerida por el entorno del Datawarehouse.
 - Ayudar con la planificación del proyecto, determinación de entregables y establecimiento de tiempos.
-

- Ayudar en el establecimiento de criterios para la selección de herramientas de consulta del usuario final.
- Participación en la capacitación de usuarios en el uso de las herramientas de consulta.
- Educar nuevos usuarios en la existencia y contenido del Datawarehouse.
- Identificar inconsistencias de datos y establecer procesos para resolverlas.
- Comunicar nuevos requerimientos al equipo del proyecto.

3. Trabajar en la provisión de calidad de datos.

- Proveer definiciones de datos.
 - Identificar problemas de calidad de datos, redundancias, inconsistencias y sus fuentes.
 - Comunicar requerimientos de cambio en la programación al equipo de desarrollo para que provean calidad de datos.
-

4. Desarrollar pruebas de aceptación sobre las herramientas de usuario final, entorno cliente de consultas, consultas predefinidas y desarrollar expectativa en la producción de resultados confiables,

- Soporte a Usuarios finales: El soporte de usuario final debe entender que no todos los usuarios tienen los mismos conocimientos deseados en computación. Los usuarios que ejecutan las aplicaciones tienen diferentes niveles por lo que deben ser tratados de manera diferenciada.

Este es el grupo más visible a nivel de usuarios finales, los cuales al no encontrar satisfactorio su trabajo dejarán de usar el sistema y el Datawarehouse fallará. Con el Datawarehouse, este rol necesita incluir muchas más capacidades y responsabilidades:

- Manejar y administrar las consultas.
 - Evaluar las herramientas de consulta finales.
 - Monitorear el desempeño de las consultas.
-

- Entrenamiento a los usuarios finales (determinar requerimientos, evaluar tiempos y cronogramas de entrega, coordinar y administrar sesiones de clase y crear manuales de usuario y de referencia).
 - Proveer Help Desk⁴⁶
 - Desarrollar pruebas de aceptación.
 - Coordinar la integración de software y sistemas operativos.
 - Entender los datos en el Datawarehouse.
- Administración de Datos: La administración de datos no tiene mucha influencia en la mayoría de organizaciones en los EEUU, pero el Datawarehouse sugiere un número de nuevas responsabilidades para las cuales la administración de datos es el único calificado. El éxito final del Datawarehouse se debe en gran medida a cómo este rol cumplirá con lo siguiente:
 - Consultar con los equipos de modelamiento del Datawarehouse y los sistemas de legado.
 - Modelar los datos.

⁴⁶ Help Desk: Soporte de Escritorio. Se refiere al equipo del área de sistemas que provee ayuda al usuario en la resolución de problemas de hardware, software y comunicaciones más comunes.

- Asegurar la calidad e integridad del modelo del Datawarehouse.
 - Alinear los requerimientos de información con los modelos del Datawarehouse.
 - Administrar los modelos.
 - Proveer soporte interno de las herramientas de modelamiento de datos.
 - Servir de interfaz con el DBA.
 - Administrar y controlar el repositorio de datos, incluyendo quien tiene permitido el acceso y modificaciones a los objetos.
 - Facilitar la unión entre el personal del TI y los usuarios finales en la determinación de requerimientos.
 - Ser responsable en conjunto con el equipo de desarrollo para la selección de cuales fuentes de datos serán usadas cuando hay múltiples orígenes de datos.
- Analista de Datos: El Analista de Datos puede formar parte del equipo de desarrollo o puede ser un subrol de la administración de datos. En cualquiera de los
-

casos, los siguientes roles son importantes y necesitan ser desempeñados:

- Actuar como un especialista de diseño técnico.
 - Construir las reglas de negocio dentro de las herramientas del Datawarehouse.
 - Mirar por oportunidades de reutilización e integración de datos.
 - Trasladar Datos Operacionales a modelos de Datawarehouse.
 - Construir e implementar datos y metadata dentro del Datawarehouse.
- Equipo de Desarrollo (Desarrollador de Aplicaciones):

Los desarrolladores de aplicación están más familiarizados con las fuentes de datos y lo duro que es identificarlos. Mientras la migración y transformación de datos a través de herramientas que ofrecen los vendedores representan en este proceso cuando menos algo de involucramiento en la programación, la experiencia que ellos tienen es relevante en la codificación de aplicaciones en los casos no triviales. Los siguientes son los roles que estos deberían abarcar:
-

- Participar en la determinación de los significados de las columnas tanto de las fuentes de datos como del Datawarehouse.
 - Escribir las más complejas consultas que exploten las capacidades de los usuarios finales.
 - Identificar los datos que deberían ser migrados al Datawarehouse.
 - Desarrollar y mantener los modelos y requerimientos del negocio.
 - Ser responsable de las fórmulas de computación en los procesos de transformación.
 - Ser responsable en conjunto con el Administrador de Datos de la selección de cual origen de datos usar cuando hay múltiples fuentes.
 - Oficial de Seguridad: Muchas organizaciones tienen una oficina o grupo responsable de la seguridad. Las opciones de seguridad (embebida en la mayoría de las aplicaciones usadas dentro del Datawarehouse) son únicamente efectivas si hay el entendimiento y
-

apropiada selección de lo implementado. El oficial de seguridad tiene las siguientes responsabilidades:

- Educar a los usuarios sobre la necesidad de seguridad y ayudarlos a entender su rol en el establecimiento de los requerimientos de seguridad.
 - Determinar quien debe tener autoridad para leer, crear, actualizar y borrar datos específicos.
 - Establecer estrategias, diseñar políticas de seguridad, desarrollar e implementar procedimientos seguros.
 - Determinar los requerimientos de seguridad.
 - Explorar las capacidades de seguridad del DBMS y formar una estrategia de implementación de varias opciones de seguridad con el DBA.
 - Explorar las capacidades de seguridad de los usuarios finales.
 - Explorar las capacidades de seguridad de las herramientas de migración, el repositorio y las herramientas de desarrollo.
-

- Administrar los procedimientos de seguridad (asignación y revocación de autoridad).
 - Administración de la Base de Datos: La mayoría de los roles y responsabilidades del DBA no están muy distantes de sus funciones tradicionales, pero la cantidad de trabajo es significativa especialmente cuando las bases de datos llegan a ser muy grandes y estando distribuidas su desempeño llega a ser un problema. Los roles y las responsabilidades para el DBA son:
 - Servir de interfaz en la administración de datos incluyendo la sincronización de los modelos físicos y lógicos.
 - Generar el diseño físico de la Base de Datos para el Datawarehouse.
 - Participar en la migración de datos desde los sistemas tradicionales.
 - Responsabilizarse del desempeño del Datawarehouse.
 - Monitorear los recursos en el DBMS.
 - Manejar los respaldos y restauraciones de las bases de datos.
-

- Implementar la seguridad en las bases de datos.
 - Responsabilizarse por la distribución de datos del Datawarehouse.
 - Interactuar con el proveedor del DBMS.
 - Entender como las herramientas de usuario final acceden al DBMS.
- Servicio Técnico: Los servicios técnicos son responsables de establecer la infraestructura sobre la cual se fundamenta los componentes del Datawarehouse. Son siete los principales roles en la creación de una infraestructura para Datawarehouse:
 - Establecer la arquitectura técnica en conjunto con otros grupos.
 - Desarrollar planes de capacidad y rendimiento de redes, servidores, plataformas, DBMS, etc.
 - Monitorear el desempeño de redes, servidores, plataformas, DBMS así como de las consultas.
 - Desarrollar planes de contingencia para restauración de desastres y monitoreo de disponibilidad de redes, servidores, plataformas, DBMS, etc.
-

- Recomendar estrategias de conectividad de redes en protocolos y equipos de interconexión.
- Manejar la instalación y mantenimiento del hardware y software del sistema.
- Desarrollar pruebas de aceptación de todos los componentes del Datawarehouse.

La gran mayoría de los roles expuestos pueden dar la impresión de que son necesarios grandes equipos de trabajo para implementar el Datawarehouse. Aunque en la práctica la mayoría de proyectos exitosos han sido ejecutados por pequeños equipos de personas quienes son inteligentes, entrenadas en sus respectivos roles, motivados, dedicados al proyecto de Datawarehouse y bien administrados.

3.1.4. *Asignar tiempos para culminación de tareas*

La programación del desarrollo del proyecto de datawarehouse con los hitos y fechas más importantes se refiere al Plan de Trabajo del Proyecto.

El detalle de este proceso se muestra en el capítulo 7 de Anexos 7.1 apéndice A

CAPÍTULO IV

ANÁLISIS Y DISEÑO DEL DATAWAREHOUSE

4.1. Iniciación del Proyecto

4.1.1. Alcance de Datos

Como se mencionó anteriormente en el apartado de “Alcance del Datawarehouse”⁴⁷ este proyecto contempla la cobertura de una de las áreas de mayor impacto en la ESPOL como lo es la Académica.

Dentro de este ámbito se crearán modelos de datos iniciales (cubos de información) que intenten resolver las necesidades de información más críticas y consultadas actualmente.

El datawarehouse abarcará tres (3) modelos de datos distribuidos por áreas de la siguiente manera:

Área Académica:

- Cubo de Registros
- Cubo de Graduados

Área de Personal

- Cubo de Personal Docente

⁴⁷ Se refiere a la sección 2.2.1 del capítulo de Justificación del Datawarehouse.

Para las áreas contempladas se ha dispuesto considerar información demográfica (situación geográfica, estado civil, sexo, nacionalidad, tipo de sangre, etc.), de manera general pero haciendo énfasis en características particulares de la entidad de datos correspondiente; por ejemplo para el cubo de Graduados se considera la modalidad y el tipo de grado.

4.1.2. Alcance Temporal

Debido a que el datawarehouse evoluciona y los datos que contiene llegan a ser más accesibles, los usuarios de la ESPOL que no usan el Sistema inicialmente, podrían descubrir también el valor de los datos. Al enlazar el datawarehouse a otros sistemas (tanto internos como externos de la ESPOL), se puede compartir información con otras entidades con poco o sin desarrollo, como son el caso de Facultades, Institutos y otras unidades investigación y apoyo.

En el momento que se incremente el número de usuarios que hagan uso del datawarehouse, será necesario ampliar los cubos de información contenidos dentro del mismo, con lo cual se tendrá mayor sofisticación y utilidad; los datos acumulados dentro de la ESPOL llegarán a ser más organizados, más interconectados, más accesibles y, en general, disponible a más empleados.

El resultado esperado será la obtención de mejores decisiones en el negocio, más oportunidades y más claridad de trabajo.

Temporalmente y como ya se mencionó en el apartado 2.2, el datawarehouse abarcará dentro del área académica las entidades de registros y graduados con toda su demografía y características propias de su naturaleza. De igual manera se ha tomado en consideración la información de empleados y profesores basados en los continuos requerimientos de este tipo de información por parte de usuarios internos⁴⁸ como externos⁴⁹.

⁴⁸ Usuarios Internos: Rectorado, Vicerrectorado Administrativo Financiero, Departamento de Personal

⁴⁹ Usuarios externos: Tribunal Electoral del Guayas, CONESUP, CONUEP, Contraloría del ESTADO.

4.1.3. Razones Organizacionales

Por definición decidir es el proceso de elegir entre varias alternativas posibles, a fin de que se tome cierta determinación, según Stoner y Freeman (1994). La vida, ya sea en la esfera profesional, la académica e incluso en la personal, está marcada por situaciones y hechos que requieren de la adopción de decisiones al respecto.

Generalmente estas decisiones se toman de forma intuitiva, o no requieren de mucho análisis para su aceptación. Esto se debe a que la mejor alternativa es clara o a que la decisión a tomar no es lo suficientemente importante y, no vale la pena dedicar tiempo y esfuerzo al análisis sistemático y riguroso de las diferentes alternativas existentes.

Por el contrario, existen otras situaciones en las que estos esfuerzos para analizar las alternativas y racionalizar las decisiones tiene su importancia. Es al momento de tomar estas decisiones que se necesita pensar y reflexionar antes de actuar en cualquier sentido.

La gestión de la empresa consiste precisamente en eso, en tomar decisiones y, al igual que en la vida cotidiana, unas veces se hace de forma intuitiva y otras, a través de procesos sistemáticos de reflexión. Hablar de organización y de toma de decisiones en forma separada es tan solo un artificio pedagógico para explicar ambos conceptos, según Koontz & O'donnell⁵⁰ (1985), organizar es “repartir el poder de decidir”.

En el caso de la ESPOL se necesitan constantemente tomar decisiones con respecto a las siguientes actividades:

- Área Académica: .- Las unidades y el CRECE tienen que tomar diversas decisiones con respecto a esta área; por ejemplo al inicio de cada Período Académico tienen que decidir cuantos paralelos se van a dictar de cada materia, esta planificación la hacen en base a consultar del Período Académico anterior, y no en base a un análisis de los datos históricos. Con la implementación del datawarehouse, se busca que las unidades tengan la información oportuna al momento de realizar la planificación.

⁵⁰ Koontz, H., O'donnell, C. (1985). Administración. México: Mc.Graw-Hill.

Para la implementación del datawarehouse, se han tomado en consideración los siguientes tópicos:

- Registros
 - Graduados y Egresados
 - Demografía de Estudiantes
- Área de Personal: Es importante conocer la información de los empleados con los que cuenta la ESPOL, como están distribuidos a través de los diferentes Centros Administrativos de la ESPOL, y se considerarán los siguientes tópicos para el desarrollo del datawarehouse:
 - Demografía de Profesores
-

4.1.4. *Enfoque Completo del DW*

Luego de haber analizado las necesidades de información y tomando en cuenta la disponibilidad de recursos con los que se cuenta, se ha planificado el desarrollo de un datawarehouse flexible, accesible, escalable y confiable como plataforma a herramientas de análisis, extracción, consulta, administración y distribución. Por estas razones se ha considerado el desarrollo de los módulos del datawarehouse de la ESPOL a continuación descritos:

- DW-Security (Administración de Usuarios, Operaciones, Metadata)
 - DW-Feeder (Módulo del DW-Security. Contiene información de procesos de carga, tablas dimensionales, tablas facts, frecuencias, etc.)
 - DW-OLAP (Crea y procesa los cubos de datos)
 - DW-Consult (Reporteador Web para consultar el E-DW)
 - Scheduler (Configuración de la Base de Datos del DW para procesos de carga)
 - Templates (Plantillas de Reportes Predefinidos)
-

4.1.5. *Participantes y sus Roles*

Los roles de los integrantes en la implementación del EDW en la ESPOL se basan en lo ya expuesto en la sección 1.3.7 y relacionándolos con los integrantes del CSI serán:

- Padrino/Sponsor⁵¹ – Directora del Centro de Servicios Informáticos (CSI)
- Administrador del Proyecto – Administrador de los Datos del CSI
- Modelador de Datos – Analista Programador y DBA
- Analistas del Negocio – Director del Departamento Financiero y Coordinadores de las áreas académicas (CRECE, Contabilidad, Tesorería y Presupuesto)
- Analistas de Sistemas – Coordinadora del Sistema Financiero y Analistas Programadores
- Desarrolladores – Analistas Programadores y Documentadores
- Usuarios – Personal de áreas académicas y financieras

⁵¹ Auspiciante

Todas estas personas intervendrán en las diferentes etapas que comprende el proyecto, teniendo obviamente mayor énfasis y participación en aquellas que les son más familiares.

El traslado de un modelo Lógico al Diseño de la base de Datos física estará a cargo del DBA y del Modelador de Datos toda vez que se haya aprobado un entorno lógico adecuado (análisis dimensional correcto).

4.1.6. *Asunciones y Limitantes*

Para el desarrollo del Datawarehouse de la ESPOL se tuvieron que considerar algunos aspectos que de alguna manera u otra ayudan o entorpecen la labor de desarrollo. La naturaleza de la contribución o limitante de un aspecto esta dada por factores culturales, organizacionales y tecnológicos.

En base a lo recomendado por Kosar⁵² (1997) con relación a los siete errores mortales de una implementación de datawarehouse; las asunciones deben ser identificadas y documentadas al inicio del proyecto. Es por esto que se mencionan las siguientes asunciones:

- Los proyectos usados como ejemplos han sido aprobados.
- El equipo de TI conoce lo que es un Datawarehouse y lo que comprende la tecnología del mismo.
- Existe apoyo organizacional.
- El Datawarehouse es seriamente aceptado como un proyecto de tecnología de información.

⁵² Denis Kosar actualmente trabaja en el Case Manhattan Bank como vicepresidente de Arquitectura de Información Empresarial. El es responsable del establecimiento de la estrategia del Datawarehouse empresarial. También es miembro de la IDUG (DB2 User Group).

- Que la estructura de los modelos de datos esté bien definida y no registrar cambios profundos a lo largo de la implementación del sistema, ya que esto conllevaría a un rediseño de los cubos de datos y un retraso en el desarrollo del proyecto; esto debe cumplirse al menos en la etapa inicial.
- Que los modelos de datos desarrollados y los reportes predefinidos se basen exclusivamente en los requerimientos de información de los usuarios del Datawarehouse.
- Que la cantidad de datos que serán cargados inicialmente en el datawarehouse esté establecida. El área abarcada es la académica y no existirán de momento cobertura para otras áreas.

De igual manera Kosar (1997) menciona el establecimiento y enunciado de limitantes que pueden ocasionar conflicto con el desarrollo de un datawarehouse. Para la implementación de la ESPOL se consideran los siguientes limitantes que influyen radicalmente en su entorno:

- No considerar herramientas comerciales que pudieran haber facilitado el desarrollo del proyecto, debido a su alto costo y
-

a la necesidad de independencia de un determinado producto que se requería.

- El DBMS para este proyecto es DB2, siguiendo los estándares de plataforma de base de datos establecidos por el Centro de Servicios Informáticos de la ESPOL para todos los sistemas transaccionales que se encuentran en producción (esto es, con licencia operativa vigente).
 - Debido a que el concepto de datawarehouse es poco conocido para los usuarios y dado que este sistema será utilizado por un grupo pequeño de usuarios su difusión no es amplia y requieren cierta capacitación.
-

4.1.7. Estrategia de Administración del Proyecto

No existe una fórmula de garantía real para el éxito de la construcción de un datawarehouse, pero hay muchos puntos que contribuyen a ese objetivo, según Kimball (1996).

A continuación, se indican algunos puntos claves que se consideran en el datawarehouse de la ESPOL:

1. "Establecer una asociación de usuarios, gestión y grupos"

Es esencial involucrar tanto a los usuarios como a la gestión para asegurar que el datawarehouse contenga información que satisfaga los requerimientos de la empresa.

La gestión puede ayudar a priorizar la fase de la implementación del datawarehouse, así como también la selección de herramientas del usuario. Los usuarios y la gestión justifican los costos del datawarehouse sobre cómo será "su ambiente" y está basado primero en lo esperado y segundo, en el valor comercial real.

2. “Seleccionar una aplicación piloto con una alta probabilidad de éxito”

Una aplicación piloto de alcance limitado, con un reembolso medible para los usuarios y la gestión, establecerá el datawarehouse como una tecnología clave para la empresa.

Estos mismos criterios (alcance limitado, reembolso medible y beneficios claros para la empresa) se aplican a cada fase de la implementación de un datawarehouse.

3. “Construir prototipos rápida y frecuentemente”

La única manera para asegurar que el datawarehouse reúna las necesidades de los usuarios, es hacer el prototipo a lo largo del proceso de implementación y aún más allá, así como agregar los nuevos datos y/o los modelos en forma permanente. El trabajo continuo con los usuarios y la gestión es, nuevamente, la clave.

4. “Implementación incremental”

La implementación incremental reduce riesgos y asegura que el tamaño del proyecto permanezca manejable en cada fase.

5. “Reportar activamente y publicar los casos exitosos”

La retroalimentación de los usuarios ofrece una excelente oportunidad para publicar los hechos exitosos dentro de una organización. La publicidad interna sobre cómo el datawarehouse ha ayudado a los usuarios a operar más efectivamente puede apoyar la construcción del datawarehouse a lo largo de una empresa.

4.2. Definición de la Arquitectura

Los pasos envueltos en el desarrollo de una arquitectura de alto nivel que establezcan una infraestructura técnica y operativa para el Datawarehouse son requeridos cuando una organización como la ESPOL esta enfocada en sus necesidades de información. Esto ocurre como resultado de ejercicios de reingeniería de negocios o simplemente cuando la carencia de información integrada afecta la parte operacional, así como la decisión de desarrollar nuevas habilidades.

Las pasos para la definición de la arquitectura muestra las subsecuentes actividades dentro de valiosos pasos. El equipo de desarrollo debe usar técnicas que relacionen los requerimientos del negocio con las áreas involucradas o las necesidades de información de alto nivel.

El dimensionamiento inicial se hace para guiar las decisiones del equipo de desarrollo y definir rendimiento. El resultado de esta actividad brinda requerimientos de alto nivel para el procesamiento y el almacenamiento.

La definición y preparación del ambiente técnico se hace cuando el nuevo equipo de desarrollo y el software esta en orden. Eso es cuando el equipo está siendo ordenado e instalado.

Ladley⁵³ (1997) indica que la arquitectura física debe ser definida en términos de estos principales componentes:

- Fuentes o Sistemas de Legado: ¿De dónde los datos serán obtenidos en términos de hardware y DBMS?
- Transporte: ¿Qué implantación de red soportará la transferencia de datos?
- Destino: ¿Qué hardware y DBMS almacenará los datos?
- Metadata: ¿Cómo serán almacenadas y accesadas las definiciones de datos, relaciones, reglas y transformaciones?
- Acceso: ¿Cómo los datos serán vistos por el usuario final?
- Transformaciones: ¿Existe alguna herramienta o aplicación especial requerida para transformar los datos y construir la metadata?

Finalmente, las actividades de la arquitectura intentarán asegurar que los varios componentes lleguen a trabajar juntos.

⁵³ John Ladley es actualmente Director de Programas de Metagroup, especializado en Datawarehouses y aspectos organizacionales. Fue Director de Administración de Información para Alliance Blue Cross & Blue Shield, St. Louis, MO, donde fue responsable de la definición y provisión de un entorno de toma de decisiones para datos corporativos. Más información sobre él, puede ser encontrada john_ladley@metagroup.com.

4.2.1. Arquitectura de las Aplicaciones

Aplicando lo anteriormente expuesto en el apartado 1.2.2 referente a la plataforma de la ESPOL, a continuación se muestra algunos de los tipos de operaciones que se efectúan dentro de un ambiente de datawarehousing:

- Sistemas Operacionales
- Extracción, Transformación y Carga de los Datos
- Metadata
- Acceso de usuario final
- Plataforma del datawarehouse
- Datos Externos

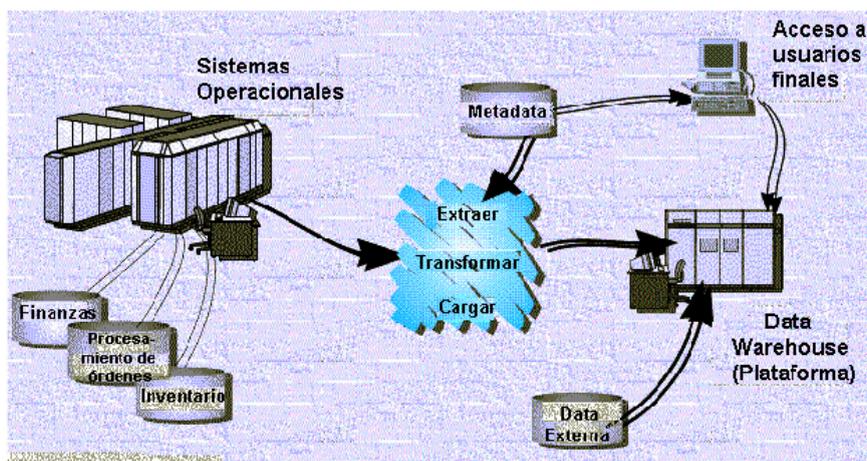


Figura 6: Arquitectura Típica de un Datawarehouse

Fuente: <http://www.inei.gob.pe/cpi-mapa/bancopub/libfree/lib619/152.HTM>

4.2.2. Dimensionamiento inicial

El dimensionamiento inicial de este proyecto se realiza durante el análisis del área objetivo (prototipo).

En dicho análisis se consideran cantidades lógicas, requerimientos, componentes disponibles, alcance y demás características.

Como anteriormente se mencionó la cobertura inicial de este proyecto será del área académica abarcando entidades como registros, graduados, estudiantes y profesores.

4.2.3. Entorno Técnico Definido

A continuación se detalla la plataforma principal de las bases de datos existentes en la ESPOL, que son utilizadas para la alimentación del datawarehouse. Corresponden a los sistemas fuentes donde operan las aplicaciones transaccionales.

SERVIDOR:	Servidor SPARC Sunfire 280R con sistema operativo Solaris 8 <ul style="list-style-type: none"> • 2 procesadores • 4-GB RAM • 436 GB disk storage
NETWORK:	TCP/IP
RDBMS:	DB2 Universal Database Enterprise Edition for Solaris Versión 6.1.00
HERRAMIENTAS DE ADQUISICIÓN DE DATOS:	<ul style="list-style-type: none"> • CAE • SQL scripts • Control Center
HERRAMIENTAS DEL USUARIO FINAL:	<ul style="list-style-type: none"> • Sistema Académico de la ESPOL (SAAC) • Polimático⁵⁴ • Audio ESPOL⁵⁵ • Académico en Línea • Lotofone⁵⁶

Tabla 8: Entorno Técnico Definido en la ESPOL

Autor: Juan Carlos Bustamante

⁵⁴ Sistema parecido al cajero automático que emite certificados académicos en la ESPOL.

⁵⁵ Sistema de Audio Respuesta que a través del teléfono permite a estudiantes y proveedores consultar datos académicos y de pagos.

⁵⁶ Sistema de consultas académicas a través del celular.

4.2.4. Entorno Técnico Preparado

A continuación se detalla la plataforma a emplearse en la implementación del datawarehouse de la ESPOL. Algunos componentes pueden ser modificados en base a las necesidades que surjan en el desarrollo del mismo.

SERVIDOR:	Servidor DELL con sistema operativo Windows 2000 Advanced Server <ul style="list-style-type: none"> • 1 procesador • 1-GB RAM • 80 GB disk storage
NETWORK:	TCP/IP
RDBMS:	DB2 Universal Database Enterprise Edition for Windows 2000 Versión 6.1.00
HERRAMIENTAS DE ADQUISICIÓN DE DATOS:	<ul style="list-style-type: none"> • CAE • SQL scripts • Control Center
HERRAMIENTAS DEL USUARIO FINAL:	<ul style="list-style-type: none"> • Microsoft Office (Excel) • EDW - Consult • GXQuery • Web browsers (Internet Explorer, Navigator, Neoplanet)
MEJORAS PLANEADAS/ UPGRADES	<ul style="list-style-type: none"> • En planificación la ampliación a los Sistemas Financiero y de Personal

Tabla 9: Entorno Técnico Preparado para el Datawarehouse

Autor: Juan Carlos Bustamante

4.3. Necesidades de los Tomadores de Decisiones

4.3.1. *Requerimientos de Información*

Los requerimientos de información en las empresas se han dividido de acuerdo al siguiente esquema, de acuerdo a la clasificación publicada por el INEI (1997):

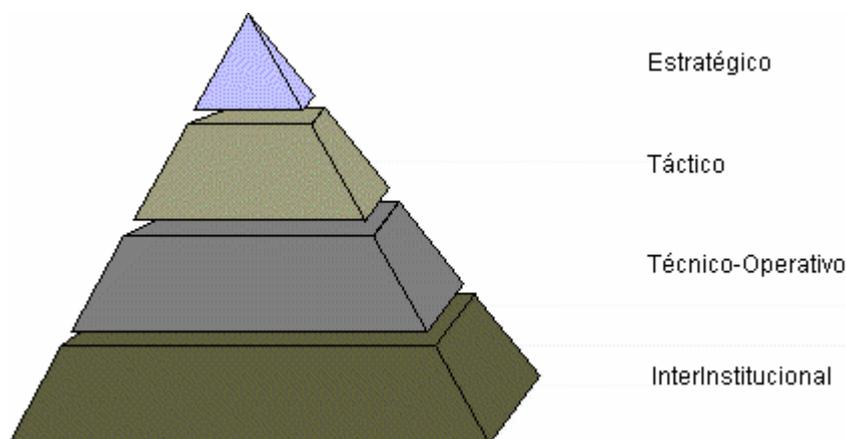


Figura 7: Pirámide de Tipos de Información

Fuente: <http://www.inei.gob.pe/cpi-mapa/bancopub/libfree/lib619/12.htm>

Bajo esta jerarquía se han desarrollado sistemas y aplicaciones orientadas a la cobertura de las necesidades de información particulares de cada empresa, los cuales se mencionan a continuación:

Sistemas Estratégicos, “orientados a soportar la toma de decisiones, facilitan la labor de la dirección, proporcionándole un soporte básico, en forma de mejor información, para la toma de decisiones”. Se caracterizan porque son sistemas sin carga periódica de trabajo, es decir, su utilización no es predecible, al contrario de los casos posteriores, cuya utilización es periódica. Vale mencionar entre estos sistemas: los Sistemas de Información Gerencial (MIS), Sistemas de Información Ejecutivos (EIS), Sistemas de Información Georeferencial (GIS), Sistemas de Simulación de Negocios (BIS y que en la práctica son sistemas expertos o de Inteligencia Artificial-AI).

Sistemas Tácticos, “diseñados para soportar las actividades de coordinación de actividades y manejo de documentación, definidos para facilitar consultas sobre información almacenada en el sistema, proporcionar informes”. En resumen, estos sistemas facilitan la gestión independiente de la información por parte de los niveles intermedios de la organización.

Destacan entre ellos: los Sistemas Ofimáticos⁵⁷ (OA), Sistemas de Transmisión de Mensajería (E-mail y Fax Server), coordinación y control de tareas (Work Flow) y tratamiento de

⁵⁷ Ofimáticos es un término que hace referencia a elementos computacionales de oficina como hojas electrónicas, editores de texto, administrador de diapositivas y demás.

documentos (Imagen, Trámite y Bases de Datos Documentarios).

Sistemas Técnico-Operativos, “que cubren el núcleo de operaciones tradicionales de captura masiva de datos (Data Entry) y servicios básicos de tratamiento de datos, con tareas predefinidas (contabilidad, facturación, almacén, presupuesto, personal y otros sistemas administrativos)”. Son los encargados de las actividades operativas y cotidianas de la empresa. Estos sistemas están evolucionando con la incursión de sensores, autómatas, sistemas multimedia, bases de datos relacionales más avanzadas y datawarehousing.

Sistemas Interinstitucionales, “es consecuencia del desarrollo organizacional orientado a un mercado de carácter global, el cual obliga a pensar e implementar estructuras de comunicación más estrechas entre la organización y el mercado (Empresa Extendida, Organización Inteligente e Integración Organizacional)”.

Este último nivel de sistemas de información recién está surgiendo, todo esto a partir de la generalización de las redes informáticas de alcance nacional y global (INTERNET), que se

convierten en vehículo de comunicación entre la organización y el mercado, no importa dónde esté la organización (INTRANET), el mercado de la institución (EXTRANET) y el mercado (Red Global).

Sin embargo, la tecnología datawarehousing basa sus conceptos y diferencias, según INEI (1997), entre dos tipos fundamentales de sistemas de información en todas las organizaciones: los sistemas técnico-operacionales y los sistemas de soporte de decisiones. Este último es la base de un datawarehouse.

Sistemas técnico-operacionales, como indica su nombre, “son los sistemas que ayudan a manejar la empresa con sus operaciones cotidianas. Estos son los sistemas que operan sobre el "backbone" (columna vertebral) de cualquier empresa o institución, entre las que se tiene sistemas de ingreso de órdenes, inventario, fabricación, planilla y contabilidad, entre otros.”

Debido a su volumen e importancia en la organización, los sistemas operacionales siempre han sido las primeras partes de

la empresa a ser computarizados. A través de los años, estos sistemas operacionales se han extendido, revisados, mejorados y mantenidos al punto que hoy, ellos son completamente integrados en la organización.

Desde luego, la mayoría de las organizaciones grandes de todo el mundo, actualmente no podrían operar sin sus sistemas operacionales y los datos que estos sistemas mantienen.

Sistemas de Soporte de Decisiones, “son sistemas interactivos basados en la computadora, que integran sistemas de información geográfica (SIG) con modelos que describen y predicen procesos productivos (Viglizzo, 1999), diseñados para ayudar a los decisores a utilizar datos y modelos para identificar y resolver problemas y tomar decisiones (Ascough, 1997). “

Estos sistemas tienen cuatro características principales: incorporan datos y modelos, son diseñados para asistir a los gerentes en los procesos de decisión, no reemplazan, sino que apoyan a la toma de decisiones y su objetivo es mejorar la efectividad de las decisiones y no la eficiencia con la que esas

decisiones son tomadas (Ferreira, 1995; Shaffer, 1998; Milne, 1998).

Son estos sistemas sobre los que se basa la tecnología datawarehousing. A través de estos sistemas, el usuario puede responder de manera más concreta en la toma de decisiones, ya que le brinda una panorámica sobre el problema y una serie de alternativas propias para la resolución del mismo. Asimismo, los beneficios que se pueden obtener de los mismos son: calidad de decisión, mejora de comunicación, reducción de costos, incrementos en productividad, etc.

4.3.2. *Requerimientos Organizacionales*

Dentro de una organización como la ESPOL, el grupo de TI tiene como objetivo sincronizar su misión con la de la organización, a fin de cubrir sus requerimientos y necesidades.

Las siguientes son algunas de las metas institucionales de la Escuela Superior Politécnica del Litoral⁵⁸ y sirven de justificativo válidos para la implementación del Sistema planificado:

- “La administración politécnica se ejecutará bajo los más modernos principios de gestión, descentralización y desconcentración”,
- “La situación financiera institucional se sustentará en la autogeneración de recursos y en los réditos de los fondos patrimoniales”,
- “El *cliente interno* es una preocupación permanente en la ESPOL. Si nuestros estamentos realizan sus actividades académicas y administrativas en un ambiente idóneo, los

⁵⁸ Metas institucionales de la ESPOL pueden ser consultadas en el sitio <http://www.espol.edu.ec>

resultados serán positivos. Crear y desarrollar este ambiente es un deber institucional”.

- Brindar una eficiencia administrativa y de transferencia de información, en cumplimiento con la política de calidad ISO 9001:2000 de la ESPOL ⁵⁹.

La misión del Centro de Servicios Informáticos⁶⁰ está basada en la misión institucional de la ESPOL e incluye objetivos tales como:

- Implementar mejores servicios de información enfocados hacia las necesidades de clientes internos y externos.
- Innovar continuamente los servicios que ofrece.
- Proveer información corporativa a todos los usuarios que necesitan tomar decisiones estratégicas, para una eficaz y eficiente Gestión Institucional.
- Instalar soluciones que integren en una Red de Información Corporativa, las soluciones heterogéneas de hardware, software y comunicaciones.
- Proveer aplicaciones analíticas que permitan a los usuarios monitorear procesos, analizar comportamientos y

⁵⁹ Política de Calidad ISO 9001:2000 de la ESPOL: Mayor información en el sitio web: <http://www.iso9001.espol.edu.ec>

⁶⁰ Mayor información de la misión del Centro de Servicios Informáticos consultar en <http://www.csi.espol.edu.ec>

tendencias, y determinar dónde tomar medidas para corregir desvíos no deseados.

- Integrar todas las bases de datos existentes en la ESPOL, para crear una Base de Datos Institucional.

En fiel seguimiento de estos requerimientos es que resulta justificable la implementación del Datawarehouse para la ESPOL. Es una meta de la Institución y por ende de sus centros de Apoyo (en este caso del Centro de Servicios Informáticos), buscar eficiencia en la gestión administrativa apoyada en tecnología de información.

4.3.3. *Requerimientos de Acceso*

De la misma forma que hay una gran cantidad de maneras para organizar un datawarehouse, es importante notar que también hay una gama cada vez más amplia de usuarios finales.

De acuerdo a la clasificación de necesidades de información realizada por el INEI (1997), se puede ubicar tres grandes categorías de usuarios dentro de la ESPOL:

- “Estratégicos” o “Ejecutivos y gerentes”.- En la ESPOL es posible identificar dentro de este rol al Rector, Vicerrectores, Decanos y Directores.
- "Tácticos" o "Coordinadores de Actividades".- En el caso de la ESPOL existen los Programadores del CRECE, consultores e investigadores.
- “Operativos” o “Usuarios técnicos”.- Que comprenden los usuarios de oficina, administrativos, etc.

Cada una de estas categorías diferentes de usuario tiene su propio conjunto de requerimientos para los datos, acceso, flexibilidad y facilidad de uso.

La experiencia adquirida por el equipo de desarrollo de TI de otros proyectos dentro de la ESPOL (Sistema Académico, Sistema Financiero, Sistema de Control de Admisiones) y la tendencia de ofrecer tecnología de punta, da lugar a poder sugerir como principales fuentes de acceso las siguientes:

- Un Sitio Web
 - Excel
 - Acceso a datos vía ODBC
 - Aplicaciones desarrolladas por terceros (Genexus, Brio, Powerplay, Cognos)
-

4.3.4. Prototipos

El prototipo desarrollado brinda a los usuarios finales una aproximación de lo que el Datawarehouse les puede proporcionar en un período de tiempo corto, para que el grupo de TI pueda demostrar los beneficios del Datawarehouse a los usuarios y recolectar en el menor tiempo la retroalimentación crítica de los usuarios.

Siguiendo las recomendaciones de Inmon (1992) en lo referente a implementación de datawarehouses, se concluye que los datos considerados en este proyecto deben ser:

- Llevados, integrados y cargados en estructuras de información apropiadas.
- Distribuidos por medio de herramientas de acceso a usuario finales y aplicaciones para realizar consultas.
- Soportar la creación de herramientas de Toma de Decisiones en caso de ser aplicable.

Con la aplicación de estas consideraciones, la implementación del datawarehouse de la ESPOl puede tener mayor aceptación

por parte de los usuarios, ya que tienen una idea general de la funcionalidad del sistema y principalmente se sentirán identificados con el mismo al ser tomados en cuenta en su desarrollo.

4.4. Análisis del Área Objetivo

4.4.1. Alcance del Área Objetivo

Tomando en consideración que el área objetivo inicial para este proyecto de Datawarehouse fue el Centro de Registros y Calificaciones de Estudiantes (CRECE), que es el ente encargado del control, administración, manejo y difusión de la información académica generada por todos los sistemas transaccionales de la ESPOL: se priorizó su elección como área objetivo dada la necesidad de proveer una plataforma que soporte un manejo de nivel superior de dicha información en lo relacionado a consultas, reportes y estadísticas.

Adicionalmente, la información del área académica reúne condiciones favorables para ser considerada como base para un Datawarehouse, dado que puede extraerse importante información de carácter estratégico y de gestión en campos como registros, graduados, egresados, demografía estudiantil, factores socioeconómicos, tendencias étnicas y migratorias.

4.4.2. Nivel de Detalle Requerido

El nivel de detalle de información esta supeditado a las necesidades particulares del área de negocio elegida, que para el desarrollo de este proyecto es el ámbito académico.

Los usuarios de esta área están interesados en información a nivel de estudiantes y profesores con características tanto académicas (matrícula, factor p, promedio, número de materias aprobadas) como personales (tipo de sangre, sexo, dirección, estado civil).

Para la definición de tiempo el más bajo nivel de granularidad será semestral/bimestral/trimestral según la carrera y el esquema a seguir en un período académico será Año, Término y Ciclo.

En el ámbito de calificaciones se hila al nivel de exámenes parciales, promedios por materia y por carrera.

Geográficamente la granularidad permite navegar a través del espectro país, provincia, ciudad, sector y ubicación específica del estudiante y/o profesor.

Según la clasificación académica en cuanto a carreras se detalla hasta el nivel de especialización en el formato división, carrera, especialización.

Las materias cuentan con características de categoría, tipo de financiamiento, créditos teóricos y prácticos, horas que se dicta y unidad responsable.

Es crucial el soporte de sumarización y procesamiento a partir de cálculos en las actividades cotidianas de los elementos que conforman esta área, según sea la naturaleza de las categorías de registros, graduados, estudiantes y profesores.

4.4.3. Tablas de Resumen Preliminares

Las Tablas de Resumen Preliminares son estructuras de datos físicas o virtuales que agrupan, totalizan o condensan de manera general los datos con la finalidad de proveer eficiencia y rapidez en las consultas a partir de información procesada.

Estas pueden ser escritadas de los resultados obtenidos a partir de los requerimientos de los usuarios en actividades preliminares.

Se identifican estructuras sumariadas preliminares de entidades como:

- Totales de registros y graduados por períodos académicos, carreras, campus.
 - Totales de materias aprobadas, reprobadas, tomadas.
 - Totales de profesores y estudiantes por sexo, estado civil, tipo de sangre.
 - Totales de estudiantes por factores socioeconómicos (factor P, sectores), procedencia y residencia.
 - Otros.
-

4.4.4. *Modelo de Datos del DW*

De acuerdo a Ullman⁶¹ (1999), “un modelo de datos es un sistema formal y abstracto que permite describir los datos de acuerdo con reglas y convenios predefinidos”. Es formal pues los objetos del sistema se manipulan siguiendo reglas perfectamente definidas y utilizando exclusivamente los operadores definidos en el sistema, independientemente de lo que estos objetos y operadores puedan significar.

El tipo de implementación seleccionado para modelar los datos es la Star Scheme o “Esquema Estrella” , “Compuesto de una tabla central -tabla de hechos- y un conjunto de tablas mostradas en una forma radial alrededor de ésta -tablas dimensión” según Kimball (1996).

Todo esto forma parte del modelamiento multidimensional explicado con mayor detalle en los apartados 4.7.1 y 4.7.3.

⁶¹ Ullman, Jeffrey y Widom, Jennifer, *Introducción a los Sistemas de Bases de Datos*. Editorial Prentice Hall, México 1999, ISBN: 970-17-0256-5.

A continuación se citan los principales modelos de datos definidos inicialmente para el desarrollo del datawarehouse de la ESPOL:

Registros Académicos.- Este modelo permite consultar información acerca de los registros de los estudiantes a través del tiempo, se han considerado como filtros de consulta:

- Carrera del estudiante
- Factor Socioeconómico (Factor P)
- Procedencia
- Lugar de residencia
- Colegio de Procedencia
- Campus
- Periodo de ingreso
- Materias registradas
- Estado del registro.
- Sexo.
- Estado civil.
- Tipo de Sangre.
- Estado del Estudiante.

Los valores que se pueden consultar son:

- Cantidad de materias aprobadas.
-

- Cantidad de Materias tomadas.
- Promedios generales y por materia.
- Calificaciones por exámenes.
- Créditos teóricos y prácticos.
- Verificaciones si es estudiante novato.

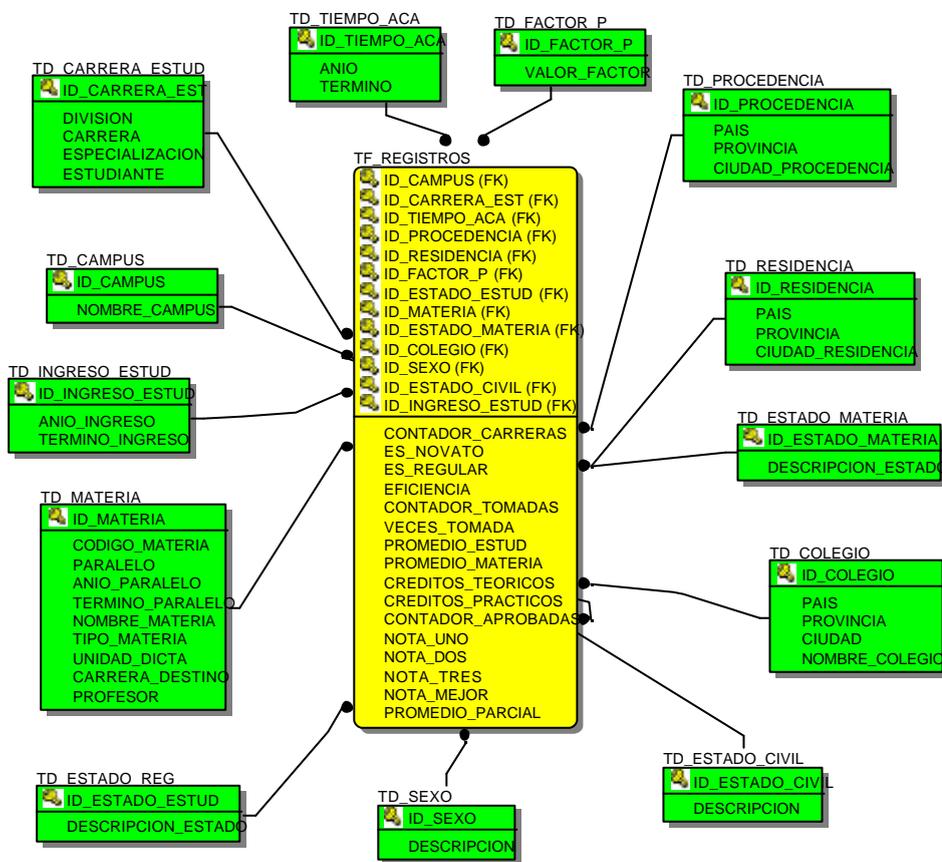


Figura 8: Modelo de Datos de Registros Académicos

Autor: Juan Carlos Bustamante

Graduados Académicos.- El cubo de graduados académicos contiene información cronológica de los estudiantes graduados de la ESPOL, se definieron filtros como:

- Tiempo.
- Campus.
- Factor Socioeconómico
- Lugar de procedencia.
- Residencia.
- Colegio de procedencia.
- Título.
- Periodo Académico que egresó el estudiante.
- Periodo de Ingreso.
- Fecha de graduación.
- Estado Civil.
- Modalidad de grado.
- Sexo.
- Carrera del estudiante.

La información que se puede consultar es:

- Promedio final del estudiante
 - Eficiencia
 - Cantidad de materias tomadas
-

- Cantidad de materias aprobadas
- Cantidad de materias reprobadas.

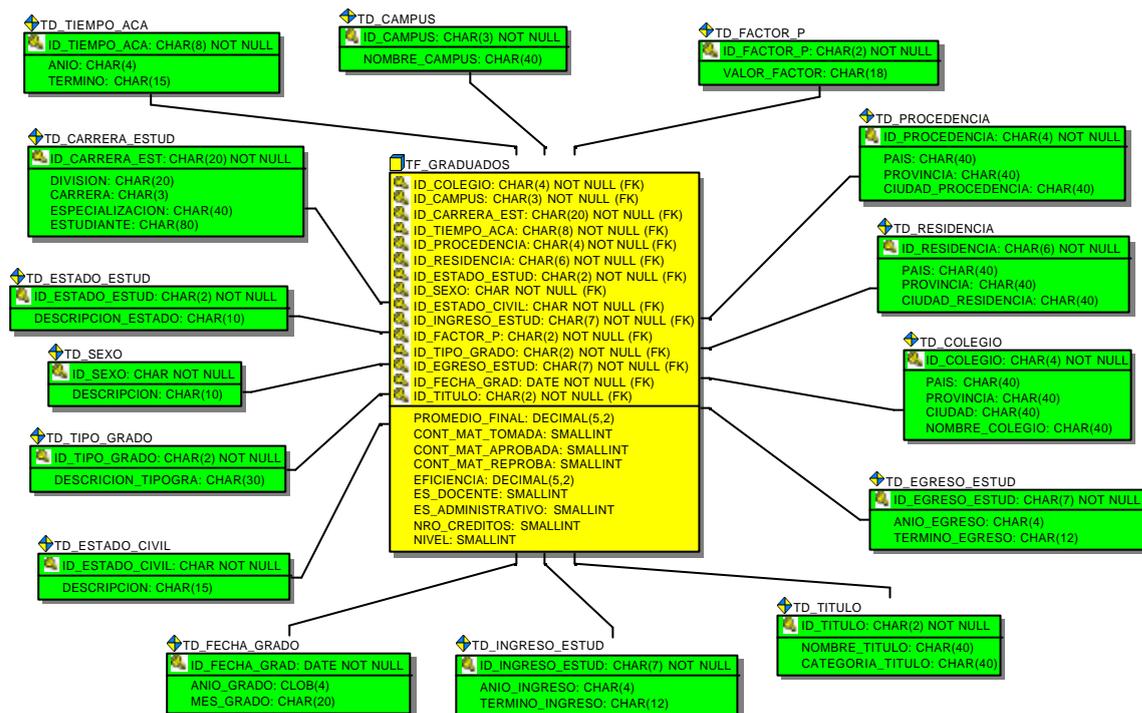


Figura 9: Modelo de Datos de Graduados Académicos

Autor: Juan Carlos Bustamante

Personal.- El cubo de personal contiene información demográfica de los profesores de la ESPOL, se definieron filtros como:

- Apellidos y Nombres
- Procedencia.
- Residencia.
- Sexo
- Título.
- Unidad
- Nacionalidad
- Estado Civil.
- Tipo de Contrato
- Tipo de Sangre

La información que se puede consultar es:

- Total de paralelos
 - Cantidad de materias
-

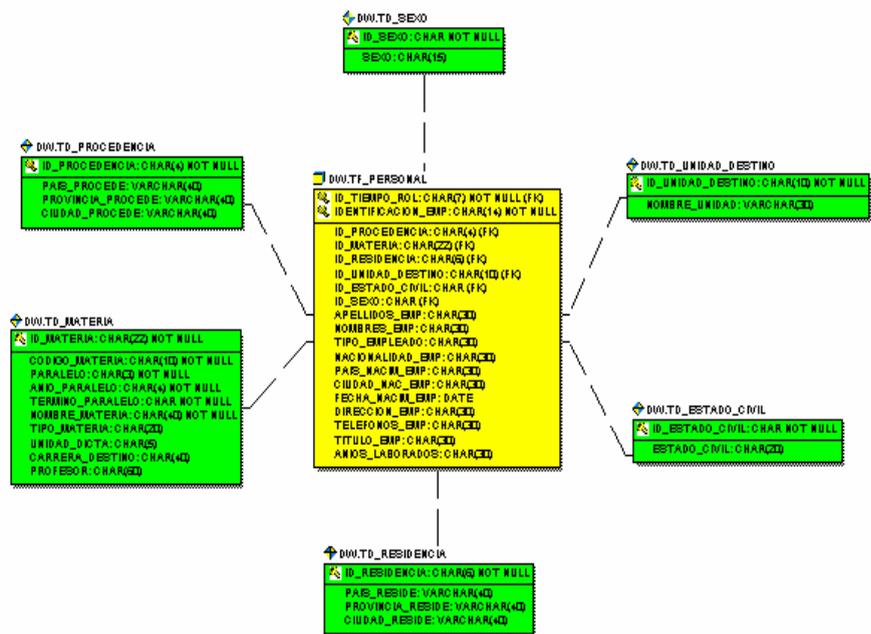


Figura 10: Modelo de Datos de Personal

Autor: Juan Carlos Bustamante

4.5. Análisis de los Sistemas de Legado⁶² (Fuentes de Información)

4.5.1. Fuentes Candidatas

Las fuentes candidatas que aportan con datos al E-DW comprenden las bases de datos, archivos y documentos que contienen información relacionada de alguna manera con las operaciones que serán involucradas en el proceso de diseño de los modelos.

El sistema de base de datos actual instalado en la ESPOL es DB2 Enterprise Edition de IBM, cuya versión en Producción es la 6.1 sobre plataforma operativa Solaris 8.0.

Este motor de base de datos corre sobre un servidor Sun FIRE 280R con 4GB. de memoria RAM, 2 procesadores, arreglo en configuración RAID 0 y 5 de discos para una capacidad instalada aproximada de 432 GB de almacenamiento.

El conjunto de bases de datos abarcado por esta infraestructura contiene los siguientes elementos:

⁶² Sistemas de Legado son comúnmente conocidos como Sistemas Fuentes, Orígenes de datos y Transaccionales, es decir, todo aquel que pueda contribuir con datos al datawarehouse.

Nombre	Descripción
SAF	Base de datos Financiera
SAR	Base de datos de Roles y Personal
SAAC	Base de datos Académica
DBWEB	Base de datos para aplicaciones Web

Tabla 10: Fuentes Candidatas del Datawarehouse de la ESPOL

Autor: José Rodríguez Rojas

La configuración de estas bases fue realizada de manera independiente al momento de crearlas como respuesta a las necesidades de información de sectores específicos, es por esto que, a pesar de residir en un mismo equipo, mantienen información redundante y duplicada que se mantiene actualizada mediante la ejecución de procesos de replicación en horas muertas (horas de la madrugada) o cuando la ocasión amerite su sincronización.

Adicionalmente a estas bases existen otras bases de datos diseminadas en las diferentes unidades y centros de la ESPOL como parte de sistemas desarrollados y/o adquiridos de manera independiente.

Realizando un recorrido por estos centros se han determinado las siguientes fuentes de información existentes:

Nombre	Descripción	DBMS
BIBLIO	Base de datos de Biblioteca – Administración	DB2 EE Versión 6.1 sobre Solaris 8.0
BIBDIG	Base de datos de Biblioteca Digital	DB2 EE Versión 6.1 sobre Solaris 8.0
SIDW	Base de Datos del Datawarehouse de la ESPOL	DB2 EE Versión 6.1 sobre Windows 2000 Server
ESPOL50	Base de Datos de la Fundación ESPOL 50 Años -Contactos	SQL Server 2000 sobre Windows XP
CEPROEM	Base de Datos del Sistema de Control de Graduados, egresados, Empresas y Contactos	SQL Server 2000 sobre Windows 2000
SCA	Base de Datos del Sistema de Control de Admisiones del Prepolitécnico	SQL Server 2000 sobre Windows 2000 Server
FUNDESPOL	Base de Datos de Administrativa de FUNDESPOL	SQL Server 2000 sobre Windows NT
SQLESPOL	Base de Datos del Sistema de AudioESPOL	SQL Server 2000 sobre Windows 2000 Server
Archivos Planos	Reportes, Archivos de sistemas viejos, Lector Biométrico	PC's
Documentos impresos	Rectorado, CRECE, Unidades	PC's
Hojas de Cálculo	Rectorado, CRECE, Unidades	PC's

Tabla 11: Otras Fuentes de Datos de la ESPOL

Autor: Juan Carlos Bustamante

4.5.2. Problemas Operacionales y de Integridad

Según el INEI (1997), el aspecto más importante del ambiente datawarehousing es que la información encontrada al interior está siempre integrada.

Afirman también que a través de los años, los diseñadores de las diferentes aplicaciones han tomado sus propias decisiones sobre cómo se debería construir una aplicación. Los estilos y diseños personalizados se muestran de muchas maneras.

La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros.

Para la implementación del datawarehouse de la ESPOL se tuvieron que resolver diversos problemas operacionales y de integridad en base a los criterios sugeridos por el INEI (1997). Entre los más importantes se pueden citar:

Codificación: En los datos contenidos en los sistemas fuentes de la ESPOL existen varias formas de codificar un campo en particular que muchas veces dada la naturaleza de la aplicación no guarda un formato estándar, lo cual origina codificaciones diversas e inconsistencias. Por ejemplo el campo SEXO en algunas tablas se representa como una "M" y una "F", mientras que en otras se representa como una "H" y una "M".

No importa mucho cómo el SEXO llega al datawarehouse. Probablemente "M" y "F" sean tan buenas como cualquier otra representación. Lo importante es que sea de cualquier fuente de donde venga, el SEXO debe llegar al datawarehouse en un estado integrado uniforme.

Por lo tanto, cuando el SEXO se carga en el datawarehouse desde una aplicación, donde ha sido representado en formato "M" y "F", los datos deben convertirse al formato del datawarehouse.

Medida de atributos: En el análisis del área académica se encontró unidades de medidas de los PROMEDIOS en una variedad de formatos. Existen reportes que muestran los

valores de los promedios en una escala entre 1 y 100, en tanto que otros reportes muestran como PROMEDIO valores en una escala entre 0.00 y 10.00.

Cualquiera que sea la fuente, cuando la información del PROMEDIO llegue al datawarehouse necesitará ser expresado de la misma manera.

Convenciones de Nombramiento.- El mismo elemento es frecuentemente referido por nombres diferentes en las diversas aplicaciones. Tal es el caso de la IDENTIFICACION de un estudiante que en algunas tablas aparece como IDENTIFICACION y en otras como NRO_IDENTIFICACION. El proceso de transformación asegura que se use preferentemente IDENTIFICACION.

Cuando el analista de sistema de soporte de decisiones observe el datawarehouse, su enfoque deberá estar en el uso de los datos que se encuentre en el depósito, antes que preguntarse sobre la confiabilidad o consistencia de los datos.

4.5.3. Evaluación

La tarea de evaluación del sistema puede actualmente tomar una gran cantidad de la programación del proyecto lo cual depende directamente del alcance, recursos y cambios en los requerimientos.

Kimball (1996) afirma que las múltiples fuentes que alimentan un datawarehouse pueden crear una larga lista de criterios de evaluación así como de resultados.

La evaluación recomienda una depuración⁶³ minuciosa de los problemas que sean descubiertos para garantizar fidelidad en los datos.

En la implementación del datawarehouse de la ESPOL, la verificación de los procesos de extracción, limpieza y carga de datos se realiza a través de reportes de logs⁶⁴. Una vista de la interfaz de evaluación es mostrada en el apartado 5.2.3.

⁶³ Se refiere a la purificación y limpieza de datos erróneos que afecten la confiabilidad de los mismos.

⁶⁴ Logs: Son bitácoras históricas y cronológicas de hechos y actividades realizadas dentro de un sistema en la ejecución de un proceso.

Los reportes de logs asociados a la comprobación de totales de información con reportes generados en el Sistema Académico constituyen las principales herramientas de evaluación con las que cuenta esta implementación de datawarehouse. Cualquier anomalía o discordancia de datos será debidamente identificada y corregida.

Por ejemplo en el proceso de carga de REGISTROS ACADÉMICOS la verificación de que todos los datos recuperados del sistema fuente hayan sido cargados en el datawarehouse es realizada con la verificación de totales de imports y exports. Adicionalmente se puede comprobar dichos totales por medio de un reporte de registros del sistema académico original.

4.5.4. Nuevos Atributos

Dyché (2001) recomienda que la incorporación de nuevos atributos en un sistema de transaccional de legado o fuentes de datos venga dada principalmente por la necesidad de alcanzar los objetivos de información obtenidos a partir de los requerimientos de los usuarios.

Estos atributos soportarán las implementaciones o cambios adicionales a fin de facilitar la estructuración de sumalizaciones, resúmenes, detalles y demás esquemas típicos de un datawarehouse.

Basados en las necesidades de información de la ESPOL para el actual proyecto, no se consideraron nuevos atributos ya que los datos existentes ofrecían facilidades para la implementación adecuada de estas estructuras.

4.5.5. Requerimientos de la Metadata

Kimball (1996) manifiesta que un aspecto de la arquitectura de datawarehouse es crear soporte a la metadata. Metadata “es la información sobre los datos que se alimenta, se transforma y existe en el datawarehouse”. Metadata es un concepto genérico, pero cada implementación de la metadata usa técnicas y métodos específicos.

Estos métodos y técnicas son dependientes de los requerimientos de cada organización, de las capacidades existentes y de los requerimientos de interfaces de usuario. Hasta ahora, no hay normas para la metadata, por lo que la metadata debe definirse desde el punto de vista del software datawarehousing, seleccionado para una implementación específica.

Kimball (1996) indica que típicamente, la metadata incluye los siguientes ítems:

- Las estructuras de datos que dan una visión de los datos al administrador de datos.
-

- Las definiciones del sistema de registro desde el cual se construye el datawarehouse.
- Las especificaciones de transformaciones de datos que ocurren tal como la fuente de datos se replica al datawarehouse.
- El modelo de datos del datawarehouse (es decir, los elementos de datos y sus relaciones).
- Un registro de cuando los nuevos elementos de datos se agregan al datawarehouse y cuando los elementos de datos antiguos se eliminan o se resumen.
- Los niveles de sumalización, el método de sumalización y las tablas de registros de su datawarehouse.

Algunas implementaciones de la metadata también incluyen definiciones de la(s) vista(s) presentada(s) a los usuarios del datawarehouse. Típicamente, se definen vistas múltiples para favorecer las preferencias variadas de diversos grupos de usuarios. En otras implementaciones, estas descripciones se almacenan en un Catálogo de Información.

Los esquemas y subesquemas para bases de datos operacionales, forman una fuente óptima de entrada cuando se

crea la metadata. Hacer uso de la documentación existente, especialmente cuando está disponible en forma electrónica, puede acelerar el proceso de definición de la metadata del ambiente datawarehousing.

La metadata del datawarehouse de la ESPOL está almacenada en la misma base de datos del repositorio, bajo un esquema de nomenclatura diferenciado del resto de tablas para facilitar su manejo. Este esquema, estandarizado por los desarrolladores del proyecto, antepone a las tablas el sufijo MD_NOMBRE_TABLA.

La información contenida en la metadata de ESPOL abarca aspectos tales como:

- Fuentes de datos (orígenes)
 - Acceso a información (usuarios, perfiles, operaciones y sus relaciones)
 - Elementos del modelo multidimensional (tablas facts, dimensionales y sus relaciones)
 - Reportes predefinidos
 - Diccionario de datos
 - Otra información de interés general.
-

En la figura a continuación mostrada se puede notar una parte del modelo de datos de la metadata,

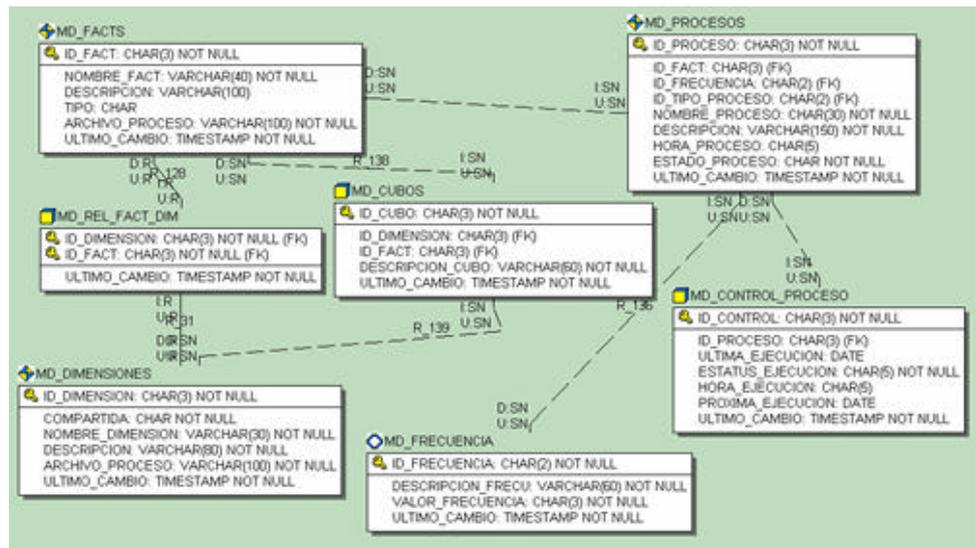


Figura 11: Muestra del modelo de datos de la Metadata

Autor: José Francisco Rodríguez Rojas

En resumen, la metadata sirve como el corazón del ambiente datawarehousing. Crear definiciones de metadata completa y efectiva puede ser un proceso que consume tiempo, pero estos esfuerzos darán como resultado un mantenimiento eficiente del datawarehouse.

4.5.6. Disponibilidad Histórica

Ladley (1997), indica que los datawarehouses tienden a crecer en gran medida a través del tiempo. Para evitar problemas con el desempeño cuando llegan a ser lentos y el consumo de espacio físico es alto, el almacenamiento de datos rara vez accedidos es lo recomendable.

Cuando el datawarehouse necesita datos de años anteriores, generalmente es necesario recuperar valores agregados (sumarizados, procesados en resumen).

De ser necesario contar con datos que no se encuentren en una sola tabla de hecho, es válido recurrir a estructuras de unión de tablas históricas separadas para los casos de consultas de datos antiguos.

Para el Datawarehouse de la ESPOL, dado que la cantidad de información a la fecha almacenada es manejable eficazmente por estar en el orden de Gigas⁶⁵ no ha sido necesario realizar

⁶⁵ Gigabytes es una medida computacional que define la cantidad de almacenamiento. Significa 'mil millones (10^9) de veces'. Con nombres de unidades de medida forma el múltiplo correspondiente, en este caso bytes.

particionamiento, pero en el caso de que la cantidad de información afecte los tiempos de respuesta de las consultas y el trabajo de procesamiento, el diseño puede soportar particionamiento de las tablas grandes en divisiones parciales de acuerdo a la naturaleza de la entidad.

En el caso de proyectos prospectivos⁶⁶, una conversión histórica no es necesaria, y esta característica puede ser omitida en el desarrollo de un proyecto de esta clase.

⁶⁶ Prospectivo se refiere a un conjunto de análisis y estudios realizados con el fin de explorar o de predecir el futuro, en una determinada materia.

4.5.7. Requerimientos de Control

Los requerimientos de control garantizan a los usuarios de un datawarehouse la veracidad y credibilidad en los datos que consultarán.

En un estudio realizado sobre las principales implementaciones de datawarehouse en compañías, Bischoff (1997) afirma que con frecuencia estos requerimientos son pasados por alto en la construcción de un datawarehouse. Esto es, cómo se puede asegurar que todos los datos cargados fueron obtenidos de dónde se supone deberían tomarse. Los usuarios no aceptarán ni confiarán en la información transformada a menos que este proceso sea absolutamente probado.

Considerando las recomendaciones de Bischoff (1997) en cuanto a la importancia del control y por lo cual muchos proyectos de datawarehouse han fracasado por omitir estos requerimientos.

En el datawarehouse de la ESPOL se han considerado como principales requerimientos de control tanto los probados en

sistemas transaccionales como los estimados en el estudio de esta implementación, esto es:

- Verificación de totales y conteo de registros en el proceso de carga.
- Revisión periódica de logs para comprobar la correcta ejecución de procesos automáticos de exportación, transformación, limpieza y cargado de datos.
- Comparación de reportes de los sistemas transaccionales con los de datawarehouse.
- Pruebas aleatorias de datos específicos para seguimiento minucioso a lo largo del proceso de migración.

Los requerimientos y las implementaciones de control son propios de la naturaleza de cada datawarehouse. Si bien es cierto, la mayoría de los controles son iguales a los empleados por los sistemas transaccionales, existen otros que dependen de las características propias de la empresa (en este caso la ESPOL) y del grado de control que se requiera en el establecimiento de dichos controles.

4.6. Diseño de las Transformaciones

El éxito o el fracaso de un negocio dependen en muchas ocasiones de una correcta toma de decisión a su debido tiempo. Por ello los Sistemas de Información juegan un papel primordial en cualquier organización.

Ladley (1997) indica que las empresas tradicionalmente se han centrado en sistemas orientados a las operaciones, que no son útiles cuando se precisa información homogénea e integrada para la toma de decisiones estratégicas que involucran a diferentes áreas y operaciones. En general la información reside en diferentes sistemas y con diversos formatos, lo cual dificulta su análisis integrado.

La solución a este problema según Ladley (1997) y Dyché (2001) es extraer dicha información y consolidarla en un único repositorio de datos, Datawarehouse, a través de interfaces. Éstas normalmente se implementan mediante DTS o productos

existentes en el mercado. Es en las interfaces donde se unifican los formatos de los datos provenientes de distintos sistemas.

En estos contextos la Extracción, Transformación y Carga de los Datos (ETL) se convierten en un punto central. Uno de los desafíos de cualquier implementación de Datawarehouse, es el problema de transformar los datos. La transformación se encarga de las inconsistencias en los formatos de datos y de homogeneizar la información. Se requiere una planificación cuidadosa y detallada para transformar datos inconsistentes en conjuntos de datos conciliables y consistentes.

Manteniendo todos los recaudos la implementación se realizará con éxito y la empresa podrá utilizar esta herramienta como una ayuda vital para el crecimiento de los negocios.

Con referencia a las recomendaciones dadas por Ladley (1997) se desarrollan las transformaciones requeridas para el datawarehouse de la ESPOL en los apartados 4.6.1 y 4.6.2.

4.6.1. Especificaciones de las Transformaciones

Uno de los desafíos de cualquier implementación de datawarehouse, es el problema de transformar los datos. La transformación se encarga de las inconsistencias en los formatos de datos y la codificación, que pueden existir dentro de una base de datos única y que casi siempre existen cuando múltiples bases de datos contribuyen al datawarehouse.

En la Figura 12 se ilustra una forma de inconsistencia, en la cual el género se codifica de manera diferente en tres bases de datos diferentes. Los procesos de transformación de datos se desarrollan para direccionar estas inconsistencias.

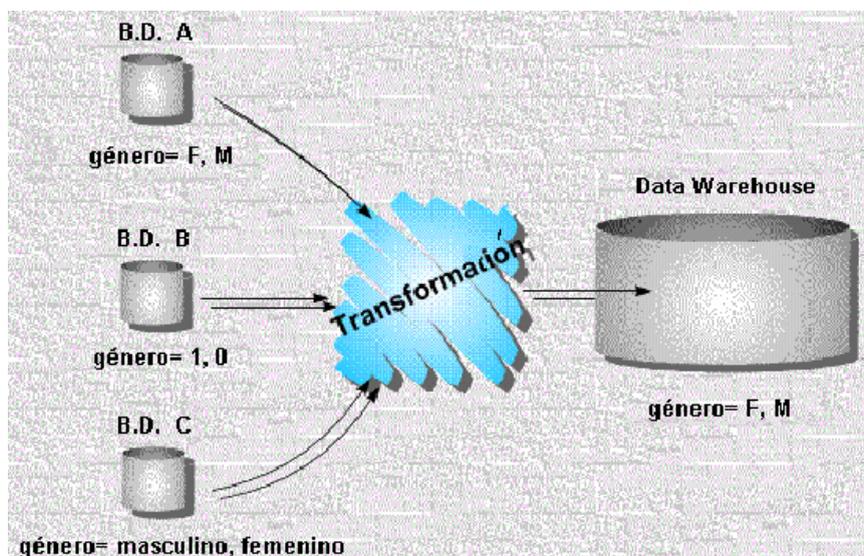


Figura 12: Ejemplo de Transformación

Fuente: <http://www.inei.gob.pe/cpi-mapa/bancopub/libfree/lib619/2234.HTM>

La transformación de datos también se encarga de las inconsistencias en el contenido de datos. Una vez que se toma la decisión sobre que reglas de transformación serán establecidas, deben crearse e incluirse las definiciones en las rutinas de transformación.

Se requiere una planificación cuidadosa y detallada para transformar datos inconsistentes en conjuntos de datos conciliables y consistentes para cargarlos en el datawarehouse.

Para el caso de la ESPOL las especificaciones de las transformaciones vienen en respuesta a los operacionales y de integridad mencionados en el apartado 4.5.2. Estas transformaciones concilian inconsistencias de codificación, convenciones de nombramiento y medidas de atributos.

4.6.2. Procesos de Transformación

Una vez determinadas las especificaciones de las transformaciones, el siguiente paso es implementar dichos procesos a fin de garantizar la consistencia de los datos. Dyché (2001) afirma que la limpieza de los datos es un proceso multifacético y complejo. Los lineamientos a seguir según Dyché (2001) son los siguientes:

1° Analizar sus datos corporativos para descubrir inexactitudes, anomalías y otros problemas.

2° Transformar los datos para asegurar que sean precisos y coherentes.

3° Asegurar la integridad referencial, que es la capacidad del datawarehouse, para identificar correctamente al instante cada objeto del negocio, tales como un producto, un cliente o un empleado.

4° Validar los datos que usa la aplicación del datawarehouse para realizar las consultas de prueba.

5° Producir la metadata, una descripción del tipo de datos, formato y el significado relacionado al negocio de cada campo.

6° Finalmente, viene el paso crucial de la documentación del proceso completo para que se pueda ampliar, modificar y arreglar los datos en el futuro con más facilidad.

En la práctica, se tendría que realizar múltiples pasos como parte de una operación única o cuando use una sola herramienta. En particular, limpiar la data y asegurar la integridad referencial son procesos interdependientes.

Los programas de limpieza de datos no proporcionan mucho razonamiento, por lo que las compañías necesitan tomar sus decisiones en forma manual, basados en información importante y reportes de auditoría de datos.

La metodología empleada en este proceso de transformación para el caso de la ESPOL condensa los seis lineamientos anteriores en tres pasos que limpian y aseguran la integridad de los datos en procesos interdependientes.

Los pasos seguidos en el proceso de transformación de datos del datawarehouse de la ESPOL quedan resumidos en la siguiente tabla:

SECUENCIA	DESCRIPCION	MUESTRA
Paso Uno	En base a los requerimientos diseñar y estudiar la implementación de un nuevo objeto en el E-DW	TD_NEW-DIMENSION TF_NEW-FACT
Paso Dos	Crear proceso de limpieza, filtrado, conversión y estandarización de la Fact o Dimensión que será creada. Se lo hace en SQL y es almacenado en un archivo en el correspondiente directorio de Archivos de Procesamiento ETL . Físicamente corresponde a una vista que consulta los datos requeridos y procesados.	V_TD_NEW-DIMENSION.VIEW V_TF_NEW-FACT.VIEW
Paso Tres	Editar un archivo de extracción de los datos en base a la vista del paso anterior y ubicarlo en el directorio de Procesos de Extracción de Datos	TD_NEW-DIMENSION.EXT TF_NEW-FACT.EXP TD_NEW-DIMENSION.EXR

Tabla 12: Procesos de transformación de datos

Autor: Juan Carlos Bustamante

En resumen los procesos de transformación pueden ser realizados por herramientas comerciales que ayuden en cada uno de estos pasos. Sin embargo, es posible escribir programas propios para hacer el mismo trabajo.

4.6.3. *Diseño de Controles*

Para cada uno de los procesos de transformación de la ESPOLE se han realizado validaciones en cuanto a la existencia o no de datos, dado que existen muchos casos en los que la información no es completa o es imprecisa.

Estos controles han permitido identificar datos sucios o erróneos como fechas con rango no válidos, estados inexistentes y otros casos particulares que se tomaron en cuenta.

Se puede mencionar la implementación de validaciones de fecha en las pantallas de ingreso de datos, la creación de listas predefinidas de estado para ciertos atributos y demás controles que permitieron identificar y corregir fuentes de información inexactas con lo que se evita nuevas incongruencias.

4.6.4. Procesos de Sumarización

Una etapa previa a la carga de datos y continuando con el concepto de transformación, es la sumarización de ciertos atributos, característica primordial dentro de un datawarehouse, que no es sino la aplicación de operaciones aritméticas, lógicas y personalizadas que permiten obtener datos relevantes en función de grupo.

El datawarehouse de la ESPOL está basado en el método de implementación de vistas de sumarización y procedimientos almacenados de procesamiento para obtener datos desde los sistemas fuentes, se generaron campos calculados a partir de ellos, un ejemplo de esto puede ser apreciado en los totales de materias tomadas, aprobadas, reprobadas del cubo de graduados.

El empleo de operadores lógicos ha sido puesto de manifiesto en ciertas instancias como lo son el reconocimiento de un estudiante como novato dentro de un período académico para el cubo de registro, así para muchos otros casos donde la sumarización fue requerida.

4.6.5. *Procesos de Conversión Histórica*

El objetivo de estos procesos es estandarizar la medida de una dimensión dentro de todas las posibles representaciones encontradas, esto se refiere a que en un proceso de transformación se pueden encontrar datos cuyo valor ha perdido significado relevante a través del tiempo debido a cambio de reglamentos, cambio de necesidades o cambios dentro de la organización.

El proceso de conversión histórica dentro del área académica de la ESPOL abarcó la transformación de los períodos académicos que habían sido llevados a lo largo de su historia con dos componentes que eran año y término, hacia un nuevo esquema en el que se incorpora la entidad del período que representa la duración de un ciclo dependiendo de la naturaleza de la carrera bajo la que esté sustentada.

La ejecución de esta conversión significó para años anteriores al actual, la reclasificación de los datos basado en las fechas en los que ocurrieron los eventos para así poder encasillarlos en el nuevo orden establecido. Se encontró información incompleta

que no pudo ser obtenida y en la que se procedió a realizar procedimientos basados en promedios o valores por defecto, intentando con esto completar el formato de datos establecido.



4.6.6. *Comprobación de Datos*

Ladley (1997) recuerda que las comprobaciones de datos en cuanto a confiabilidad, están basadas en obtener totales tanto en el datawarehouse como en los sistemas OLTP⁶⁷ de los cuales se extrajeron; dicha comprobación no es sino la verificación de estos totales en cada una de la entidades que compongan el conjunto de información.

Para el caso de la ESPOLE existe un log⁶⁸ que detalla la cantidad de datos tanto exportados (obtenidos de las fuentes de legado) como importados (cargados en el repositorio del datawarehouse) donde puede ser verificable la exactitud de este proceso e identificable cualquier diferencia entre ellos.

Estas comprobaciones son hechas a todas las tablas que componen el repositorio de datos para cada uno de los procesos en las diferentes frecuencias en que son ejecutados.

⁶⁷ OLTP On Line Transaccion Processing: Término inglés que se refiere a los sistemas de procesamiento en línea.

⁶⁸ Log: Bitácora que registra operaciones y/o acciones efectuadas por un componente de hardware o software.

La revisión de los resultados es realizada a través de la herramienta empleada para la ejecución de procesos, el Journal Center de DB2, que lleva dicho detalle en orden cronológico por tarea, estado y demás atributos.

La figura 13 muestra una imagen de verificación de detalles de procesamiento de la *tabla de hecho de registros*.

The screenshot shows the DB2 Journal Center interface. The main window displays a list of jobs with columns for Date, Time, Owner, Job ID, Job description, Script description, Status, and Elapsed time. A job with ID 129 is highlighted, and a pop-up window titled 'Job results - Job ID 129' is open, showing the following details:

Date	Time	Owner	Job ID	Job description	Script description	Status	Elapsed time
10/25/2004	3:00:48 AM	db2dw	140	Importación de ...	Carga de Dimension TO MATERIA	Completed	00:01:02
10/24/2004	8:00:04 PM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:05:59
10/24/2004	8:00:41 PM	db2dw	129	Importación de ...	Importación de Fact de Registros	Completed	00:01:57
10/24/2004	6:00:08 PM	db2dw	129	Importación de ...	Importación de Fact de Registros	Completed	00:15:06
10/24/2004	5:00:31 PM	db2dw	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:18
10/24/2004	4:00:12 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:07:59
10/24/2004	3:10:48 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:04
10/24/2004	3:00:48 AM	db2dw	129	Importación de ...	Importación de Fact de Registros	Completed	00:01:13
10/24/2004	3:00:48 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	3:00:42 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	3:00:37 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	3:00:35 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	3:00:29 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	3:00:23 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	3:00:17 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:01:11
10/24/2004	3:00:09 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	3:00:04 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:04
10/24/2004	3:00:02 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:04
10/24/2004	2:55:44 AM	dwuser	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	2:51:13 AM	db2dw	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	2:50:27 AM	db2dw	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:03
10/24/2004	2:49:04 AM	db2dw	129	Importación de ...	Importación de Fact de Registros	Completed	00:02:36
10/24/2004	2:42:08 AM	db2dw	129	Importación de ...	Importación de Fact de Registros	Completed	00:00:15

The pop-up window 'Job results - Job ID 129' contains the following text:

```

Job output:
Database server = DB2SUN6.1.0
SQL authorization ID = DWUSER
Local database alias = SACDWF

export to f_registros.dat of del messages: esregistros.log

Number of rows exported: 1232668
  
```

Figura 13: Vista del Log de ejecución de Procesos del datawarehouse

Autor: Juan Carlos Bustamante

4.7. Diseño de la Base de Datos Física

A pesar que el diseño del datawarehouse es diferente al usado en los diseños tradicionales, no es menos importante. El hecho que bs usuarios finales tengan dificultad en definir lo que ellos necesitan, no lo hace menos necesario.

Dyché (2001) refiere que el diseño de los datawarehouses es muy diferente al diseño de los sistemas operacionales tradicionales. De dicho texto se puede resumir lo siguiente:

- Los usuarios de los datawarehouses usualmente no conocen mucho sobre sus requerimientos y necesidades como los usuarios operacionales.
 - El diseño de un datawarehouse, con frecuencia involucra lo que se piensa en términos más amplios y con conceptos del negocio más difíciles de definir que en el diseño de un sistema operacional.
 - Finalmente, la estrategia de diseño ideal para un datawarehousing es generalmente de afuera hacia adentro (outside-in) a diferencia de arriba hacia abajo (top-down).
-

En la práctica, los diseñadores de datawarehouses tienen que usar muchos "trucos" para ayudar a sus usuarios a "visualizar" sus requerimientos. Por ello, son esenciales los prototipos de trabajo.

El diseño de la base de datos física del datawarehouse de la ESPOL sigue el modelo multidimensional que a menudo cae bajo la misma denominación que las bases de datos relacionales, con las que tienen en común el hecho de tener una estructura de tablas formadas por filas y columnas, las base de datos multidimensionales están construidas de un modo diferente que las bases de datos relacionales tradicionales. Desde el apartado 4.7.1 hasta el apartado 4.7.5 se trata en detalle este tema.

4.7.1. Tablas de Dominios

Las tablas de dominios según Ladley (1997) son tablas que almacenan muchos elementos que representan códigos reales, valores válidos, u otro tipo de parámetro del negocio. Durante el modelamiento del datawarehouse nuevos elementos pueden ser descubiertos y deben representarse con nuevos códigos o por el código correcto. Estos son definidos y añadidos al modelo de datos.

Para el caso de la ESPOL inicialmente este proyecto no ofrece un conjunto de códigos o cambios en el dominio, por lo cuál no fue necesaria la implementación de tablas de dominios.

4.7.2. Tablas de Sumarización

Las tablas de sumarización según Ladley (1997) son creadas para potenciar el desempeño de un datawarehouse. Por consiguiente, ellas deberían ser creadas donde es conocido que el acceso a los datos requiere un alto nivel de desempeño. Esto tiende a ser la medida de negocios de las dimensiones más populares, por ejemplo la agrupación y cálculo de totales por mes, por regiones, por carreras, etc.

Para el datawarehouse de la ESPOL en la tablas de hechos de registros y graduados se sumarizan las materias (aprobadas, reprobadas y tomadas), los créditos (teóricos y prácticos), promedios (parciales y generales), como un buen ejemplo de tablas sumarizadas.

4.7.3. Tablas de Hechos y Esquema Estrella

Ladley (1997) y Kimball (1996) concuerdan en que el Esquema Estrella y las tablas de Hechos (el diseño de las tablas granulares) son un aspecto que puede levantar controversia. Los DBA tradicionales tendrán un duro trabajo en la creación de la redundancia que ellos perciban para estas tablas. Lo mejor es utilizar un DBA experimentado en Datawarehouse, o involucrar al DBA tempranamente en el desarrollo del proyecto para que los conceptos no le sean desconocidos. Además el concepto de esquema estrella tiene que ser entendido para hacer un uso efectivo de esta técnica.

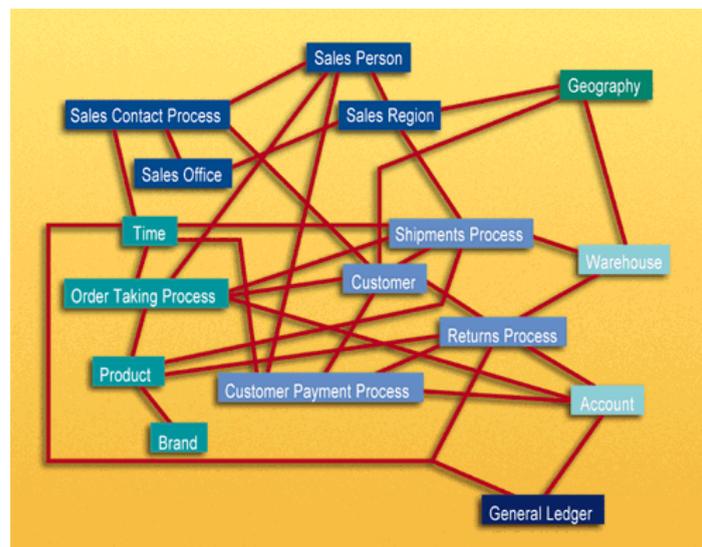


Figura 14: Vista de alto nivel de un modelo Entidad-Relación

Fuente: <http://www.dbmsmag.com/9708d15.html#figure1>

En un modelo multidimensional, los datos están almacenados como tablas de hecho y dimensiones antes que como filas y columnas. Un esquema estrella emplea tablas grandes llamadas “tablas de hecho” o “tablas fact”. que contienen datos cuantitativos y cualitativos acerca del negocio. Este además utiliza tablas pequeñas llamadas “tablas dimensionales” que pueden estar en la tercera forma normal y mantienen datos descriptivos. Una dimensión representa una perspectiva individual de un dato. Esta es una medida individual o una cara de un cubo.

Este tipo de modelo es a menudo llamado cúbico, estrella de facts y dimensiones, o modelo dimensional. Representa la visión conceptual del analista de negocios de los datos.

Una dimensión representa el criterio “POR”. Por ejemplo si se desea ver los registros por período académico, graduados por carrera, o estudiantes por sectores.

Las dimensiones son las tablas más pequeñas alrededor de una tabla de hecho en un esquema estrella. Ellas son tablas de

referencia y pueden ser jerárquicas en naturaleza. Una dimensión jerárquica puede incluir grupo de productos, tipo de producto y producto. La jerarquía de la dimensión de tiempo puede incluir día, mes, año. Para el modelo de registros la jerarquía de la dimensión del tiempo incluye año, término y período.

La tabla de hecho es usualmente grande y esta altamente renormalizada. En efecto datos para múltiples períodos de tiempo pueden aparecer en una simple tabla. Numerosos productos en el mercado están diseñados para proveer al usuario amigables herramientas de consulta que permitan pivotar⁶⁹ la tabla sobre cualquier dimensión y partición de datos dependiendo de los requerimientos del usuario.

⁶⁹ Pivotear o pivotar: Moverse o apoyarse sobre un pivote, que es un extremo o parte de una pieza, donde se apoya o inserta otra, bien con carácter fijo o bien de manera que una de ellas pueda girar u oscilar con facilidad respecto de la otra.

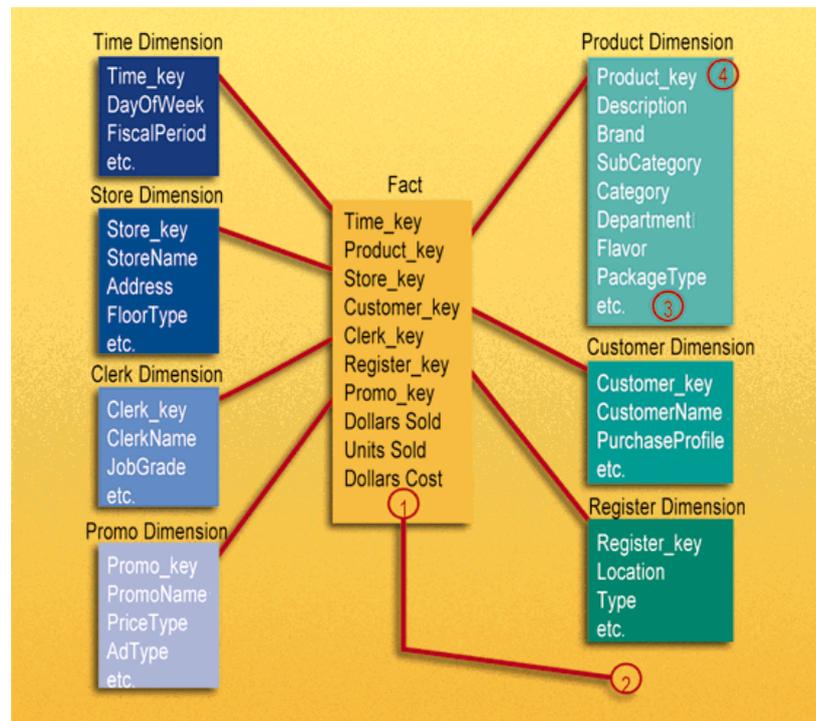


Figura 15: Modelo detallado multidimensional para un punto de ventas.

Fuente: <http://www.dbmsmag.com/9708d15.html#figure2>

Para alcanzar un buen desempeño se debe procurar que el número de dimensiones este en un nivel manejable, probablemente entre cinco y siete dimensiones. Cuando el número de dimensiones es incrementado el número de filas en la tabla de hecho se incrementará.

En algunos casos las tablas dimensionales son combinadas con la principal tabla de hecho.

Ladley (1997) define como pasos envueltos en el diseño de tablas multidimensionales los siguientes:

1. Identificar requerimientos.
2. Identificar facts (tabla de hechos) de los requerimientos.
3. Identificar las dimensiones o los criterios “POR”
4. Identificar los niveles de roll-up⁷⁰ (históricos de agregación)
5. Diseñar las tablas de hecho y dimensionales.
6. Determinar requerimientos de tamaño y retención de datos.
7. Identificar requerimientos de particionamiento.
8. Validar el diseño contra el DBMS y las herramientas de usuario final que serán usadas.
9. Modificar el diseño con los requerimientos de cambio.

La aplicación de estos preceptos en el diseño de los modelos implementados en el datawarehouse de la ESPOL, contribuye al aprovechamiento de las principales características y funcionalidades de la multidimensionalidad de una plataforma de consultas.

⁷⁰ Roll Up: se agrupa por alguna dimensión determinada. Por Ej. : Analizar las ventas de producto a las ventas por tipo de producto

Las facts (tabla de hechos) y dimensiones del datawarehouse de la ESPOL son detalladas en la tabla siguiente:

Entidad	Descripción	Tipo
TIEMPO	Periodos Académicos	Dimensión
RESIDENCIA	País, Provincia y Ciudad de Residencia de Estudiantes	Dimensión
REGISTROS	Información de Registros directamente vinculada con las dimensiones relacionadas.	Fact
CAMPUS	Información de Localidades y Campus de la ESPOL	Dimensión
PROCEDENCIA	País, Provincia y Ciudad de Procedencia de Estudiantes	Dimensión
ESTUDIANTE	Información Personal de Estudiantes	Fact
FACTOR P	Información Socio-Económica	Dimensión
CARRERA	Datos sobre las diferentes carrera que oferta la ESPOL	Dimensión
SEXO	Sexo de Estudiantes, Profesores y Administrativos	Dimensión
ESTADO CIVIL	Estado civil de Estudiantes, Profesores y Administrativos	Dimensión
GRADUADOS	Información recabada de graduados y egresados de la ESPOL desde 1959 con la relación respectiva de las dimensiones o criterios aplicables	Fact
COLEGIO	Información de País, Provincia, Ciudad y Colegio de Procedencia de los Estudiantes	Dimensión
MATERIA	Información y características de las materias con el detalle del paralelo para cada periodo académico.	Dimensión
UNIDAD	Centro de Dictados de Materias	Dimensión
PERSONAL	Información personal de Empleados. Incluye docentes y Administrativos	Dimensión

Tabla 13: Detalle de dimensiones del datawarehouse

Autor: Juan Carlos Bustamante

Cabe mencionar el empleo de dimensiones jerárquicas⁷¹ que incluyen agrupamiento y especialización de un criterio de selección, como es el caso de las dimensiones Procedencia, Residencia, Colegio, Tiempo, entre otras.

Dimensión	Detalle de Jerarquía
Procedencia	País, Provincia y Ciudad de Procedencia
Residencia	País, Provincia y Ciudad de Residencia
Colegio	País, Provincia, Ciudad y Tipo del Colegio
Carrera Estudiante	División, Carrera, Especialización del Estudiante
Tiempo académico ⁷²	Año, Periodo y Término de Registros Año y Término de Egreso, Fecha y Mes de Grado

Tabla 14: Jerarquías de las dimensiones

Autor: Juan Carlos Bustamante

⁷¹ Jerarquías: Las dimensiones se asocian con sus jerarquías y especifican distintos niveles de agrupamiento

⁷² Existen para este proyecto dos dimensiones de tiempo diferentes; la una para el modelo de Registros y la otra para el de Graduados.

4.7.4. Índices

La información de los niveles de esquematización más altos puede ser libremente indexada, mientras que las de los niveles más bajos de detalle, por ser tan voluminosa, pueden ser indexadas moderadamente.

Por lo mismo, los datos en los niveles más altos de detalle pueden ser reestructurados fácilmente, mientras que el volumen de datos en los niveles más inferiores es tan grande, que los datos no pueden ser fácilmente reestructurados.

Por consiguiente, el modelo de datos y el diseño clásico fundamentan que el datawarehouse se aplique casi exclusivamente al nivel actual de detalle. En otras palabras, las actividades de modelamiento de datos no se aplican a los niveles de esquematización, en casi todos los casos

Las tablas que tienen índices clusterizados⁷³ son aquellas que no poseen muchas actualizaciones y/o inserciones, caso particular de las tablas de un datawarehouse.

⁷³ Cluster: Término computacional que denota un grupo de cosas menudas dispuestas con alguna semejanza de racimo y que en conjunto realizan una acción común.

Para el caso de la ESPOL las tablas donde se encuentran concentrada la información resumida por varios parámetros como registros, estudiantes y graduados no tienen índices clusterizados porque se realizarán inserciones masivas y esto degradaría el tiempo de respuesta. Se le han creado índices normales para que el tiempo de acceso sea óptimo y así satisfacer los diferentes requerimientos.

4.7.5. Políticas de Respaldo y Recuperación

Cuanto más se usa un datawarehouse, más valiosos se vuelven los datos y la información en él contenidos. Para asegurar la disponibilidad permanente de los datos, un administrador de sistema o de base de datos debe hacer periódicamente copias de seguridad de los datos, lo que significa copiar total o parcialmente los datos críticos a cinta.

Esta medida de precaución asegura que, si algo sale mal y los datos quedan inutilizables, se podrá restaurar al datawarehouse la información salvada en la última copia de seguridad. Cuanto más frecuentes sean las copias de seguridad, más actuales serán los datos restaurados.

El símbolo mostrado en la figura 16 para medios de almacenamiento de información antigua es la cinta magnética, que puede usarse para almacenar este tipo de información.

De hecho hay una amplia variedad de medios de almacenamiento que deben considerarse para almacenar datos más antiguos. En la figura se muestra algunos de esos medios.

Dependiendo del volumen de información, la frecuencia de acceso, el costo de los medios y el tipo de acceso, es probable que otros medios de almacenamiento sirvan a las necesidades del nivel de detalle más antiguo en el datawarehouse. En la ESPOL el tipo de respaldo establecido por diversos factores es la cinta.



Los medios de almacenamiento para la porción voluminosa del data warehouse puede ser de una amplia variedad de tipos de almacenamiento

Figura 16: Tipos de almacenamiento

Fuente: <http://www.inei.gob.pe/cpi - mapa/bancopub/libfree/lib619/1-7.HTM>

4.8. Herramientas de los Usuarios Finales

4.8.1. *Diseño de Accesos*

Durante el diseño de accesos de los usuarios finales, el tipo de acceso al Datawarehouse es especificado y establecido. Si un prototipo hecho tempranamente está disponible, es refinado. Si algún nivel de desarrollo del acceso de los usuarios finales es requerido se debe determinar como los datos serán accedados desde un punto de vista externo. El análisis dimensional es utilizado por aplicaciones del Datawarehouse que a menudo soportan sistemas de toma de decisiones.

Ocasionalmente, una organización puede querer desarrollar sus propios accesos y esta actividad especifica la metodología que los usuarios usarán para la visión de varias medidas del negocio. El acceso a la metadata es un aspecto único que será diseñado también. Este debe ser considerado mandatorio para el acceso a las reglas de transformación, definición de datos y medidas, cálculos de sumarización, e historia de cambios en los datos son provistos a los usuarios que accedan al datawarehouse. Si esto no está ofrecido debería ser solicitado.

Existe una gran cantidad de poderosas herramientas de consulta y reporte en el mercado (Ver Anexo 1-A). Algunos proveedores ofrecen productos que permiten tener más control sobre qué procesamiento de consulta es hecho en el cliente y qué procesamiento en el servidor.

La ESPOLE considera en el diseño de sus accesos los requerimientos revisados anteriormente en el apartado 4.3.3 y que definían como posibles fuentes de acceso recursos web, Excel, acceso a datos vía ODBC y aplicaciones desarrolladas por terceros.

4.8.2. Definición de Accesos

Durante la definición de accesos componentes técnicos específicos como procedimientos almacenados y funciones son diseñadas y desarrolladas.

Ladley (1997) considera que para el Datawarehouse un componente que puede ser considerado único es la preparación del entorno para las tareas en batch. Reportes empaquetados dentro de entornos cliente-servidor son eventos raros, y existen facilidades (como el KRON en Unix) que son mejores en comparación con los más maduros entornos batch.

Por consiguiente el diseño de facilidades para la producción de reportes programados no debe ser tomado a la ligera.

El control de acceso (access tracking) significa monitorear cuáles son los datos del datawarehouse que miran los usuarios, y permite a los administradores de las bases de datos llevar control de los datos que generan mayor demanda, planificar actualizaciones periódicas de los datos y saber la frecuencia

con la que habría que hacer copia de seguridad de ciertos datos.

El registro de uso (usage logging) significa mantener un registro en línea de las personas que se conectan al datawarehouse, ya sea directamente o a través de una aplicación específica. Esta medida permite a los administradores de las bases de datos saber exactamente quiénes son los usuarios que acceden al datawarehouse y lo usan, y con qué frecuencia lo hacen.

Para el caso de la ESPOL los componentes batch programados son realizados en el servidor de la plataforma del datawarehouse. El acceso a las fuentes de datos es realizado y controlado por medio del Journal Center de DB2, el cual es una aplicación cliente servidor bajo Windows que ejecuta estos procesos previa validación de un usuario con permisos de operación.

El control de acceso implementado para la ESPOL registra las operaciones de entrada y salida de los usuarios del sistema los cuál ayudará a determinar el uso que se le da al sistema y la frecuencia con que se lo hace.

4.8.3. Desarrollo de Accesos

La metodología expuesta por Ladley (1997) en cuanto al desarrollo de accesos de usuario trata de las características y particularidades del desarrollo y las unidades de evaluación de los reportes y paneles de acceso de usuarios.

La única característica es que cuando este paso está completo, hay aún un gran camino que recorrer hacia la aceptación del usuario del Datawarehouse.

El uso de los datos en conjunción con su acceso debe aún ser verificado. La meta en este punto es que la posición del acceso de usuario evolucione con el uso de los datos.

El datawarehouse de la ESPOL provee diversas maneras de acceder, las cuales fueron mencionadas en el apartado 4.8.1 siendo la más simple la consulta de información a través del sitio web <http://www.edw.csi.espol.edu.ec>.

Vale mencionar que sea cual fuere el tipo de acceso, éste es controlado siempre por el usuario otorgado por la ESPOLE para estudiantes, profesores y administrativos.

4.9. Elección de las Herramientas de Desarrollo

La información estratégica sobre clientes importantes o un exitoso lanzamiento de producto, se almacena en gigabytes de datos de marketing o índice de transacciones de venta. Esa información debe ser extraída de alguna forma para la toma de decisiones.

En este caso se necesita software especializado que permita capturar los datos relevantes en forma rápida y pueda verse a través de diferentes dimensiones de los datos. El software no debería limitarse únicamente al acceso a los datos, si no también, al análisis significativo de los datos. En efecto, transformar los datos de la información cruda o no procesada, en información útil para la empresa.

Brooks⁷⁴ (1997) sugiere que el software usado en un datawarehouse se clasifica en Herramientas de Consulta y Reporte, Herramientas de Base de Datos Multidimensionales/Olap (On Line Analytical Processing), Sistemas de Información

⁷⁴ Peter Brooks es consultor de administración en el Advanced Technology Group of Coopers & Librand Consulting. Trabajó en IBM como consultor en Sistemas de Soporte a Toma de Decisiones; es bien conocido dentro del campo de Datawarehouse.

Ejecutivos, Herramientas Data Mining y los Sistemas de Gestión de Bases de Datos propiamente.

En la implementación de la ESPOL se consideraron los criterios antes mencionados así como la experiencia del equipo de desarrollo, plataforma de software establecida, convenios de licenciamiento y disponibilidad de recursos.

4.9.1. Posibles Herramientas para el desarrollo del proyecto

4.9.1.1. Hardware

Un componente fundamental a la hora de poder contar con un Datawarehouse que responda a las necesidades analíticas avanzadas de los usuarios, es el poder contar con una infraestructura hardware que la soporte.

Estos sistemas suelen comenzar con una funcionalidad limitada, que se va expandiendo con el tiempo, es necesario que los sistemas sean *escalables* para dar soporte a las necesidades crecientes de equipamiento. En este sentido, será conveniente el optar por una arquitectura abierta, que permita aprovechar lo mejor de cada fabricante.

El hardware utilizado en el datawarehouse de la ESPOL corresponde a la plataforma técnica preparada descrita anteriormente en el apartado 4.2.4.

4.9.1.2. Software de Almacenamiento (SGBD)

Como se ha comentado, el sistema que gestione el almacenamiento de la información (Sistema de Gestión de Base de Datos o SGBD), es otro elemento clave en un Datawarehouse. Independientemente de que la información almacenada en el Datawarehouse se pueda analizar mediante visualización multidimensional, el SGBD puede estar realizado utilizando tecnología de Bases de Datos Relacionales o Multidimensionales.

Según Dyché (2001), las bases de datos relacionales se han popularizado en los sistemas operacionales, pero se han visto incapaces de enfrentarse a las necesidades de información de los entornos Datawarehouse. Por ello, y puesto que, como se ha comentado, las necesidades de información suelen atender a consultas multidimensionales, parece que unas Bases de Datos multidimensionales, parten con ventaja. En este sentido son de aplicación los comentarios que se realizaron en el

apartado de hardware, por requerimientos de prestaciones, escalabilidad y consolidación tecnológica.

Queda fuera del ámbito de esta sección el detallar cómo los proveedores de bases de datos han optimizado los accesos a los índices, o las nuevas posibilidades que ofrece la compresión de datos (menos espacio para la misma información lo que implica, entre otras ventajas, que más información se puede tener en caché), para lo cual se toma como referencia a la prensa especializada o a las publicaciones de los fabricantes, ver Anexo 1-D.

La ESPOL siguiendo la infraestructura establecida utiliza como software de almacenamiento de datos DB2 Enterprise Edition para Windows 2000 versión 6.1

4.9.1.3. Software de extracción y manipulación de datos

En este apartado se analiza un componente esencial a la hora de implantar un Datawarehouse, la extracción y manipulación. Para esta labor, que entra dentro del ámbito de los profesionales de tecnologías de la información, es crítico el poder contar con herramientas que permitan controlar y automatizar los continuos "mimos" y necesidades de actualización del Datawarehouse.

Estas herramientas deberán proporcionar las siguientes funcionalidades según Dyché (2001):

- Control de la extracción de los datos y su automatización, disminuyendo el tiempo empleado en el descubrimiento de procesos no documentados, minimizando el margen de error y permitiendo mayor flexibilidad.
-

- Acceso a diferentes tecnologías, haciendo un uso efectivo del hardware, software, datos y recursos humanos existentes.
- Acceso a una gran variedad de fuentes de datos diferentes.
- Manejo de excepciones.
- Planificación, logs, interfaces a schedulers de terceros.
- Interfaz independiente de hardware.

A veces, no se suele prestar la suficiente atención a esta fase de la gestión del Datawarehouse, aun cuando supone una gran parte del esfuerzo en la construcción de un Datawarehouse.

Existen multitud de herramientas disponibles en el mercado que automatizan parte del trabajo, para lo cual se recomienda la visita a la página Internet:

<http://pwp.starnetinc.com/larryg/clean.html>

En la que se proporciona una lista de más de 100 herramientas de extracción y manipulación de datos, con links a sus páginas Internet, y una somera descripción de la funcionalidad cubierta por cada herramienta.

El software de extracción y manipulación de datos utilizado en el datawarehouse de la ESPOL fue desarrollado empleando SQL nativo y las estructuras soportadas por el DBMS tales como vistas, funciones y procedimientos almacenados. El funcionamiento de esta herramienta puede ser revisado en el apartado 4.6.2.

4.9.2. Criterios para la elección de las herramientas

El datawarehouse de la ESPOL toma en consideración algunos criterios de selección de acuerdo a la naturaleza de la herramienta. La mayoría de las herramientas según Dyché (2001) usadas para implementar un datawarehouse y acceder a él pueden entrar en la categoría de “herramientas de desarrollo”.

A continuación se notan algunas categorías de estas herramientas:

Herramientas de modelado y diseño de datos: Estas herramientas deben cumplir con lo siguiente, según Dyché (2001):

- Elaborar una visión lógica de las entidades comerciales.
 - Traslado de un diseño de base de datos a definiciones de tablas en la realidad.
 - Definir metadatos y generar ciertos códigos de programación que determinan la estructura de las tablas para el DBMS.
-

- Ayudar a delinear la estructura de las tablas del datawarehouse.

Herramientas de extracción y carga de datos: Estas herramientas no se limitan a tomar datos de un lugar y ponerlos en otro; también deben ser capaces de hacer lo siguiente:

- Reunir datos de sistemas operacionales separados.
- Convertir los datos de un formato a otro.
- Modificar los datos para que sean más comprensibles o más completos.
- Cargar los datos modificados en el datawarehouse de destino.

Herramientas de Gestión y de Administración: Sirven para afinar el entorno de datawarehouse, particularmente por el lado de la base de datos. Las principales tareas de gestión y administración son las siguientes:

- Planificación de capacidad.
 - Copias de seguridad y restauraciones.
 - Administración de seguridad
 - Control de acceso y registro de uso.
 - Ajuste y control de rendimiento.
-

En la siguiente tabla se definen los parámetros a tener en cuenta para la elección de las herramientas adecuadas.

Tipo de Herramienta	Pregunta básica	Modelo de Salida	Usuario típico
Consulta y Reporte	¿Qué sucedió?	Reportes de ventas mensuales; histórico de inventario	Necesita data histórica puede tener aptitud técnica limitada
Procesamiento analítico En línea (OLAP)	¿Qué sucedió y Por qué?	Ventas mensuales vs. Cambios de precio de los competidores	Necesita ir de una visión estática de los datos a "slicing and dicing" ⁷⁵ técnicamente astuto
Sistema de Información Ejecutiva (SIE)	¿Qué necesito conocer ahora?	Libros electrónicos; Centros de comandos	Necesita información resumida o de alto nivel puede no ser técnicamente astuto
Data mining	¿Qué es interesante? ¿Qué podría pasar?	Modelos predictivos	Necesita extraer la relación y tendencias de la data ininteligible Técnicamente astuto.

Tabla 15: Preguntas planteadas para elegir la herramienta adecuada

Fuente: "Datawarehouse: Practical Advice from the Experts"

⁷⁵ Slicing Dicing: se selecciona algún subconjunto del cubo. Analizar el cubo de datos restringiéndolo para algunos proveedores, productos y fechas

4.9.3. Justificación para la elección de las Herramientas seleccionadas

Por ser un proyecto de desarrollo no se considera el uso de herramientas de terceros para el datawarehouse de la ESPOL; salvo aquellas que vengan incorporadas en el DBMS.

La selección de herramientas de modelado de datos es una de las pocas áreas en las que no hace falta preguntarles a los usuarios finales. De hecho, la discusión de su selección depende de los administradores de bases de datos antes que del sector comercial. Por este motivo se escogió la herramienta CASE Erwin 3.5.1.

La herramienta que se encargará de extraer los datos de los sistemas de origen y cargarlos en el datawarehouse esta implementada por medio de componentes propios de la base de datos DB2 como lo son: vistas, funciones y procedimientos almacenados. Adicionalmente la manipulación de datos por

medio de sentencias SQL complementa la labor de estos elementos.

Las herramientas de Gestión y Administración del datawarehouse de la ESPOL comprende varios módulos siendo unos desarrollados y otros propietarios del DBMS.

La administración de usuarios, operaciones y metadata es realizada por el DW-Security; aplicación desarrollada para este proyecto.

Las tareas de administración de los objetos de la base de datos, control de acceso a los componentes, backups y restauraciones y afinamiento son realizadas por el Control Center de DB2.

La confección de los cubos de datos y su posterior procesamiento son realizados por el DW-OLAP, aplicación que almacena en formato XML las estructuras multidimensionales desarrolladas para este proyecto.

El DW-Consult es la herramienta basada en web que permite a un gran número de personas usando tecnología de Internet existentes consultar el datawarehouse. Esta herramienta basa su funcionamiento en templates o plantillas predefinidas que almacenan las consultas más utilizadas por los usuarios según su demanda.

CAPÍTULO V

IMPLEMENTACIÓN DEL DATAWAREHOUSE

5.1. Desarrollo del Repositorio

5.1.1. *Desarrollo de la Transformación*

Haciendo mención al apartado 4.6 que detalla el diseño de las transformaciones del datawarehouse de la ESPOL se llega a la meta de este paso, que es producir los procesos que construyan el datawarehouse y verificar que las reglas de transformación y los cálculos estén trabajando eficientemente.

Las etapas de desarrollo y ensamblaje de los procesos de transformación (ej. Trabajos en batch, scripts, etc.) fueron realizados por los componentes desarrollados para este proyecto.

Este proyecto será en resumen, datos granulares, tablas sumariadas, tablas de hecho y dimensiones para consultas y reportes predefinidos. Los procesos de transformación no son complicados para el área escogida y se encuentran ya desarrollados.

5.1.2. Carga de Datos Preliminar

Para la implementación del datawarehouse de la ESPOL hay que considerar la carga preliminar de datos, después de que las tareas de integración, desarrollo y evaluación están completas, ya que es vital cargar los archivos de destino y permitir a los usuarios acceder y verificar la veracidad de los procesos. No hay que chequear en este paso la usabilidad o aceptación del sistema.

Esta actividad sigue dos propósitos según Dyché (2001):

1. El usuario final permanece involucrado durante la porción del proyecto en la cual los problemas de datos adicionales pueden aparecer.
2. Los problemas de integridad que no son determinados en el momento usualmente es imperfecto.

Los usuarios deben estar envueltos en la solución de los problemas con los datos, ya que esto a menudo ocasiona un gran impacto sobre el proyecto.

La carga de datos preliminar para el datawarehouse de la ESPOL se realizó de manera integra abarcando todas las tablas facts y dimensionales en todo el espectro de tiempo almacenado en los sistemas fuentes.

5.1.3. *Procesos de Control y Auditoría*

De acuerdo a lo expuesto en el apartado 4.6.3 sobre diseño de controles para el datawarehouse de la ESPOL, Los procesos de auditoria y control deben ser completos y puestos en sitios claves según indica Ladley (1997). Dado que algunos son automáticos, fueron desarrollados junto con los otros componentes del Datawarehouse, por ejemplo el control de exports e imports.

Los procesos de control manual fueron documentados, examinados, y probados sobre los datos generados por el equipo de desarrollo.

Fue necesario conocer como estas herramientas realizan este control antes de aceptar su empleo definitivo.

El desarrollo de cualquier proceso automático y/o manual que verifique que todos los registros fueron hechos en un periodo de tiempo específico fue parte de esta actividad.

5.1.4. *Carga de la Metadata*

Para el caso del datawarehouse de la ESPOL, la composición de la metadata según el apartado 4.5.5 que trata sobre los requerimientos de la metadata se considera lo siguiente:

- Las estructuras de información que dan una visión de los datos al administrador a través de un visor relacional que muestre los componentes de las entidades de manera específica. Esta acción la realiza el módulo DW-Security
- Las definiciones del sistema de registro desde el cual se construye el datawarehouse, mostrado a través de un diccionario de datos fácil acceso y consulta en el sitio web disponible para este proyecto (www.edw.csi.espol.edu.ec).
- Las especificaciones de transformaciones de datos que ocurren tal como la fuente de datos se replica al datawarehouse, por medio de los scripts que son ejecutados en procesos batch⁷⁶ a través del módulo Scheduler.

⁷⁶ Procesos Batch: Son procesos que se ejecutan de manera automática en horas programadas, generalmente son establecidos en horas de la noche y abarcan tareas como replicaciones, actualizaciones masivas y operaciones sobre los datos en general.

- El modelo de datos del datawarehouse (es decir, los elementos de datos y sus relaciones).
 - Un control de nuevos elementos de datos agregados al datawarehouse así como los elementos de datos antiguos que se eliminan o se modifican, esta labor se la realiza en el DW-Security.
 - Control de la frecuencia de actualización de datos.
-

5.1.5. Pruebas del Sistema

Las pruebas del sistema ocurren después de que la mayoría de los subprocesos dentro del Datawarehouse han sido construidos, el propósito es verificar que todos los componentes desarrollados recientemente trabajen juntos.

Esta tarea no es hecha sino hasta después de la aprobación de la carga preliminar de datos granulares. Las evaluaciones del sistema necesitan ser realizadas cuando los cambios en la programación y codificación son mínimos.

Idealmente la evaluación del sistema debe ser un simple paso que asegure que todas sus piezas trabajen juntas y mantengan la integridad de los datos.

Las pruebas del sistema fueron realizadas una vez cargados los datos en el datawarehouse de la ESPOL por medio de los usuarios involucrados en este proyecto. Un caso de esto fue lo realizado con el personal del CRECE al realizar consultas a esta plataforma haciendo uso de una herramienta de consulta de datos adquirida por dicho departamento. Esta herramienta es

el Genexus Query cuyos resultados obtenidos fueron validados con reportes del Sistema Académico, comprobando la integridad de los datos.

5.2. Implementación del Proyecto

5.2.1. Metodología Implementada

Por definición una metodología es la definición formal de los procesos requeridos para brindar una solución de Tecnología de Información desde una idea inicial hasta un resultado exitoso, según Ladley (1997). La metodología es un concepto que tiene grandes implicaciones en las organizaciones y empresas de negocios.

Una metodología existe para guiar un proyecto de software. Cada proyecto es diferente. Similar a un gran chef que consulta un conjunto de sus recetas favoritas para crear una buena comida, la metodología es consultada para entender que es requerido para el éxito y adaptación a condiciones que pueden variar desde una norma percibida.

En este punto del proyecto, varias asunciones son hechas y concentran actividades prácticas tales como:

La figura 17 refleja la metodología genérica, que es la seleccionada para la implementación del Datawarehouse en la

ESPOL. Las etapas presentadas son iterativas; cada etapa provee más detalle así como puede ser repetitiva. Cada cuadro representa una etapa significativa que debería suceder durante un proyecto de tecnología de Datawarehouse.

La figura 17 ilustra la mayoría de los componentes de un proyecto de Datawarehouse.

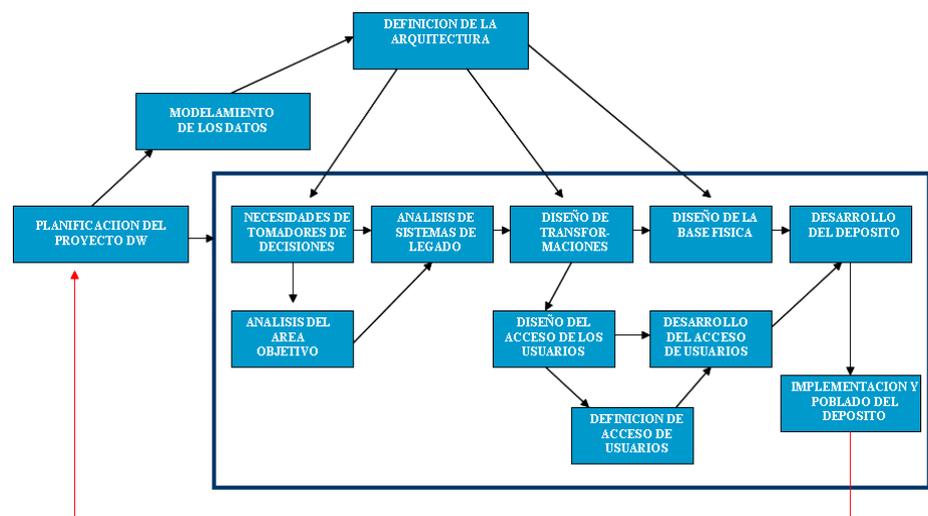


Figura 17: Metodología Genérica para el Desarrollo de un Datawarehouse

Fuente: Figure 9.1 Chapter 9 Datawarehouse: Practical Advice from the Experts

En el capítulo 4 se ha descrito en detalle la metodología aplicada para el proyecto de la ESPOL, que va desde la

planificación, definición de la arquitectura, necesidades de tomadores de decisiones, análisis del área objetivo, análisis de las fuentes de datos, diseño de las transformaciones, diseño de la base de datos física, diseño de accesos de los usuarios, desarrollo del repositorio y poblado e implementación del Datawarehouse hasta la obtención de un producto que mejorará el proceso de toma de decisiones dentro de la organización.

En esencia, Ladley (1997) manifiesta que la planificación y desarrollo de un proyecto de Datawarehouse requiere lo siguiente:

- Los proyectos de Datawarehouse deben ser iterativos.
- Las necesidades del negocio deben ser constantemente reflejadas.
- Herramientas específicas deben habilitar al equipo de desarrollo en la identificación del progreso y enfoque de cada etapa mientras se añade al “gran conjunto”.

Este capítulo debe proveer una guía en el desarrollo de una metodología para un proyecto de Datawarehouse particular. Es obvio que las características individuales de un proyecto crean las necesidades de desarrollar una planificación y análisis que

refleje la unicidad del mismo. Adicionalmente, los proyectos de Datawarehouse requieren que el departamento de Tecnología de Información que se adapte a diferentes flujos de trabajo, productos entregables y pasos en su desarrollo.

5.2.2. Implementación de los modelos del negocio

George⁷⁷ (2001) indica que un modelo de negocio es un sistema que define cómo una empresa debe planificar, construir y emplear sus recursos para ofrecer a sus clientes un mejor beneficio o valor agregado superior.

Un modelo de negocio debe satisfacer importantes interrogantes tales como:

1. ¿Quiénes son los proveedores o facilitadores de los bienes/servicios del negocio?
2. ¿Qué servicios o bienes ofrece el negocio?
3. ¿Por qué el cliente escoge los bienes/servicios del negocio?
4. ¿Quiénes serán los consumidores o clientes del negocio?

Vale indicar que responder a estas preguntas debe permitir formar un modelo de negocio, pero no siempre, ya que no es estático el proceso.

⁷⁷ Francis Stevens George es un consultor independiente de tecnología de información. Más información en <http://www.krooman.com>.

La implementación del modelo de negocio de la ESPOL abarca elementos desarrollados en la tabla 14, según los lineamientos dados por George (2001).

COMPONENTE	CARACTERISTICAS DEL MODELO
Valor del Cliente ⁷⁸	El desarrollo del proyecto permitirá dar a los usuarios un trato diferenciado por cuanto contarán con apoyo en la toma de decisiones.
Alcance	Los usuarios de la ESPOL (Directores, Analistas, Investigadores) serán los beneficiarios del valor agregado que se obtendrá.
Costos	Dado que la implementación ha sido realizada in-house, con plataforma establecida y herramientas de libre uso, el costo del proyecto es mínimo.
Beneficios	Vanguardia tecnológica Soporte a la toma de decisiones Calidad de información Satisfacción de los usuarios del sistema
Actividades relacionadas	Plataforma para herramientas de inteligencia de negocios. Soporte a sistemas de información ejecutivos. Bases para nuevos proyectos o tópicos de investigación como Datamining
Implementación	Desarrollo por parte del equipo de TI de ESPOL. Empleo de componente de hardware, software y comunicaciones propios.
Capacidades	Soporte al flujo transaccional del negocio. Posibilidad de crecimiento iterativo en cobertura. Facilidad de distribución
Vigencia	Independencia de herramientas comerciales brinda escalabilidad y flexibilidad de crecimiento. Conocimiento propio de la implementación garantiza seguridad. Ampliación de cobertura se traduce en usabilidad.

Tabla 16: Modelo de Negocio del Datawarehouse de la ESPOL

Fuente : José Rodríguez Rojas

⁷⁸ Usuarios del datawarehouse de la ESPOL

5.2.3. Implementación de las interfaces

La Interfaz de Usuario, en adelante IU, de un programa es un conjunto de elementos hardware y software de una computadora que presentan información al usuario y le permiten interactuar con la información y con el computadora. También se puede considerar parte de la IU la documentación (manuales, ayuda, referencia, tutoriales) que acompaña al hardware y al software.

Si la IU está bien diseñada, el usuario encontrará la respuesta que espera a su acción. Si no es así puede ser frustrante su operación, ya que el usuario habitualmente tiende a culparse a sí mismo por no saber usar el objeto.

Los programas son usados por usuarios con distintos niveles de conocimientos, desde principiantes hasta expertos. Es por ello que no existe una interfaz válida para todos los usuarios y todas las tareas. Debe permitirse libertad al usuario para que elija el modo de interacción que más se adecúe a sus objetivos en cada momento. La mayoría de los programas y sistemas operativos ofrecen varias formas de interacción al usuario.

Una interfaz debe facilitar el proceso de crear un modelo mental efectivo.

En la implementación de las interfaces del Datawarehouse de la ESPOl vale diferenciarlas de acuerdo a la naturaleza de las mismas, en función del componente al que pertenecen, esto es, se dará una muestra de las interfaces de administración, procesamiento, control, verificación, seguridades y consulta del proyecto.

INTERFAZ DE ADMINISTRACIÓN

La administración comprende tareas de manejo y control de la base de datos a través del DBMS y de las operaciones propias del datawarehouse con el EDW-Security.

La herramienta que administra la base de datos física (SIDW) es el Control Center de DB2. Permite realizar operaciones como crear bases de datos, columnas, triggers⁷⁹, procedimientos almacenados y gestionar parámetros de la configuración del motor.

⁷⁹ Disparadores. Son pequeños programas o rutinas que ejecuta el DBMS de manera automática ante acciones específicas sobre los datos.

Una vista de esta herramienta es mostrada en la siguiente figura que representa un cambio en una de las columnas.

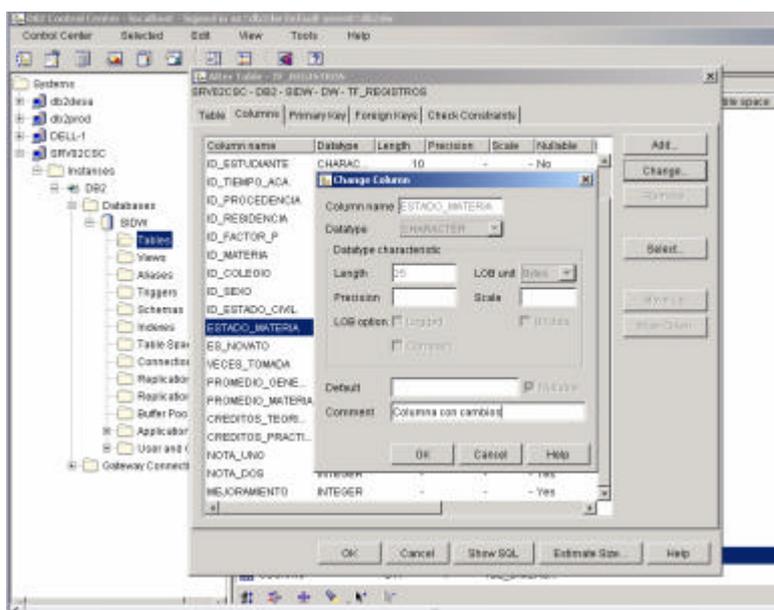


Figura 18: Vista del Control Center de DB2. Herramienta de administración de la base de datos del EDW

Autor: Juan Carlos Bustamante

La herramienta EDW-Security permite administrar usuarios, perfiles, acceso a reportes o consultas tratadas como operaciones, generación masiva de reportes con el procesamiento respectivo de extracción, resumizado y presentación en formato XML para su consulta a través del EDW-Consult.

La siguiente vista ilustra la interfaz de administración (creación, modificación y eliminación de usuarios del EDW. A este componente se denomina EDW-Security.

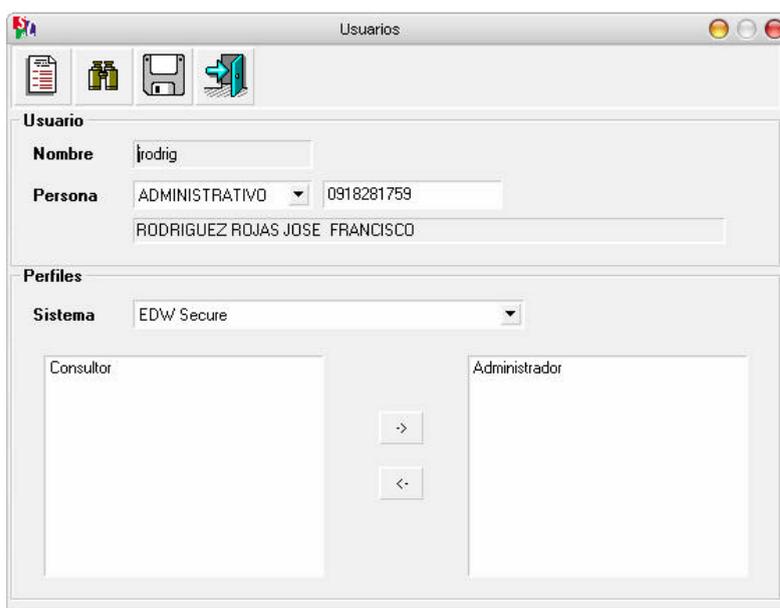


Figura 19: Administración de Usuarios a través del EDW-Security

Autor: José Francisco Rodríguez

La gestión de accesos a los diferentes niveles de información se realiza a través del componente antes descrito, pero en la modalidad de Sistemas, Perfiles y Operaciones. El manejo de los Perfiles se muestra a continuación:

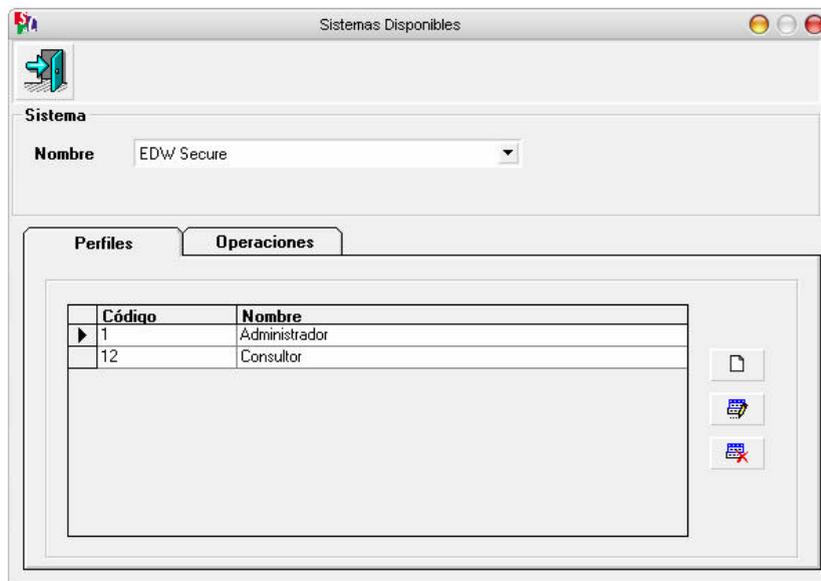


Figura 20: Administración de Perfiles del EDW-Security

Autor: José Francisco Rodríguez

Las operaciones o acceso a consultas son controladas por la siguiente interfaz.

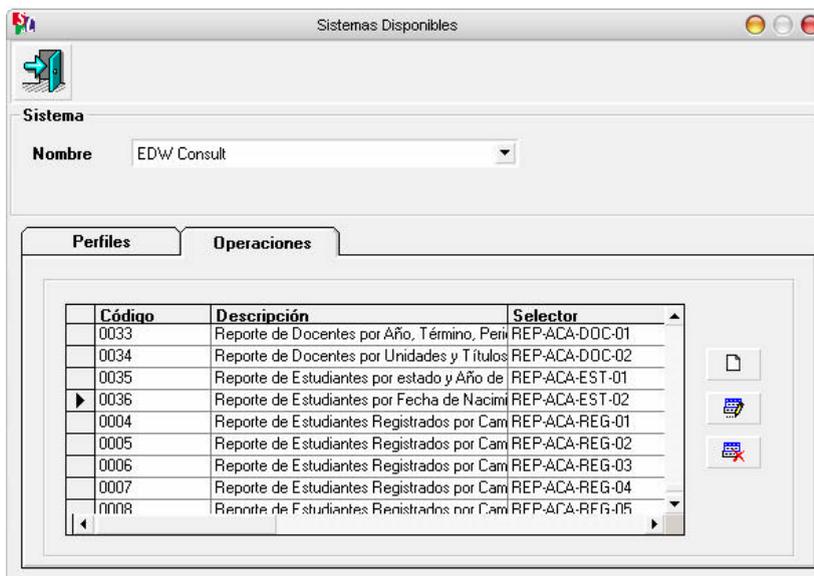


Figura 21: Interfaz de Control de Operaciones del EDW-Security

Autor: José Francisco Rodríguez

El ingreso, modificación y eliminación de elementos del datawarehouse como tablas de hechos y dimensiones son realizados por medio del componente EDW-OLAP, controlador del cubo de datos, a través de la siguiente pantalla:

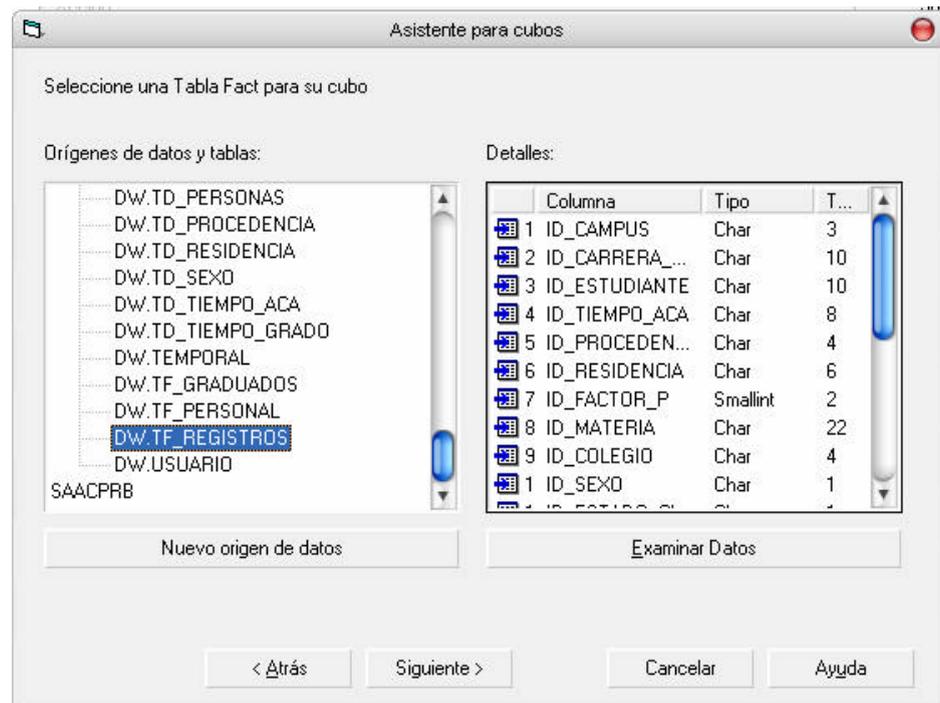


Figura 22: Vista del componente EDW-OLAP

Autor: José Rodríguez Rojas

Esta herramienta, el EDW-OLAP, permite conectarse con cualquier repositorio de datos a través de una sentencia UDL⁸⁰, otorgándole a la aplicación soporte multiplataforma.

⁸⁰ UDL: Universal Data Link o Enlace Universal de Datos, es un estándar de acceso a repositorios de datos.

Este hecho se lo hace notar en la siguiente interfaz.

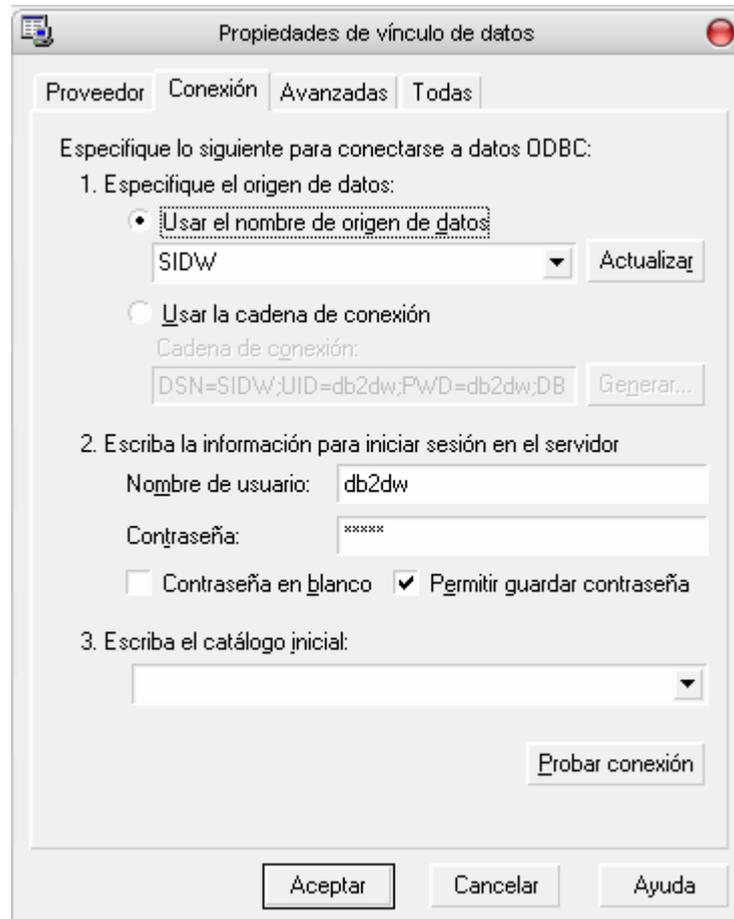


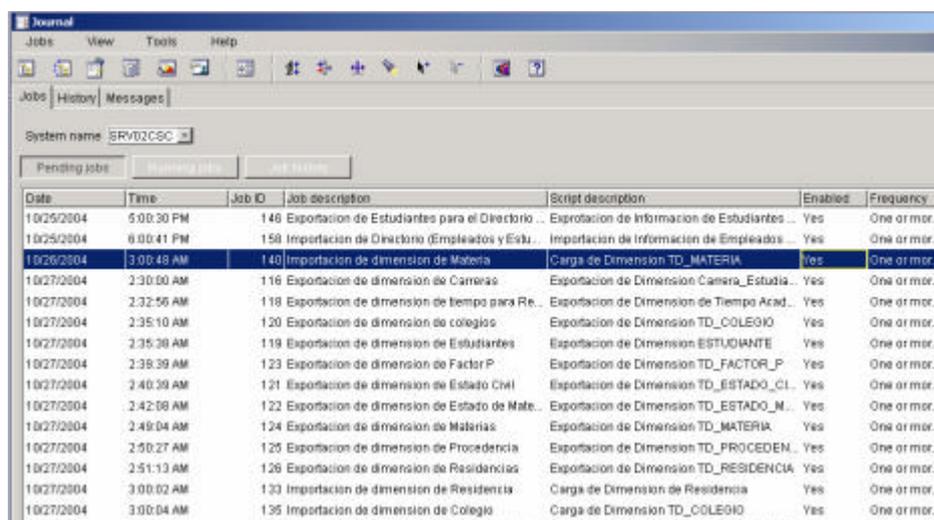
Figura 23: Interfaz de conexión de datos con el EDW-OLAP

Autor: José Rodríguez Rojas

INTERFAZ DE PROCESAMIENTO Y PROGRAMACIÓN

El procesamiento y ejecución de las tareas está a cargo del Journal Center de DB2, una herramienta que administra los trabajos y permite monitorear su planificación, ejecución e historia de manera natural.

En la figura 23 se nota el conjunto de trabajos programados para su ejecución con detalles tales como fecha y hora de ejecución, descripción, estado y demás atributos.



Date	Time	Job ID	Job description	Script description	Enabled	Frequency
10/25/2004	5:00:30 PM	146	Exportacion de Estudiantes para el Directorio	Exportacion de Informacion de Estudiantes...	Yes	One or mor...
10/25/2004	6:00:41 PM	158	Importacion de Directorio (Empleados y Estu...	Importacion de Informacion de Empleados...	Yes	One or mor...
10/26/2004	3:00:48 AM	140	Importacion de dimension de Materia	Carga de Dimension TD_MATERIA	Yes	One or mor...
10/27/2004	2:30:00 AM	116	Exportacion de dimension de Carrera	Exportacion de Dimension Carrera_Estudia...	Yes	One or mor...
10/27/2004	2:32:56 AM	118	Exportacion de dimension de tiempo para Re...	Exportacion de Dimension de Tiempo Acad...	Yes	One or mor...
10/27/2004	2:35:10 AM	120	Exportacion de dimension de colegios	Exportacion de Dimension TD_COLEGIO	Yes	One or mor...
10/27/2004	2:35:38 AM	119	Exportacion de dimension de Estudiantes	Exportacion de Dimension ESTUDIANTE	Yes	One or mor...
10/27/2004	2:38:39 AM	123	Exportacion de dimension de Factor P	Exportacion de Dimension TD_FACTOR_P	Yes	One or mor...
10/27/2004	2:40:39 AM	121	Exportacion de dimension de Estado Civil	Exportacion de Dimension TD_ESTADO_CL	Yes	One or mor...
10/27/2004	2:42:08 AM	122	Exportacion de dimension de Estado de Mate...	Exportacion de Dimension TD_ESTADO_M...	Yes	One or mor...
10/27/2004	2:48:04 AM	124	Exportacion de dimension de Materias	Exportacion de Dimension TD_MATERIA	Yes	One or mor...
10/27/2004	2:50:27 AM	125	Exportacion de dimension de Procedencia	Exportacion de Dimension TD_PROCEDEN...	Yes	One or mor...
10/27/2004	2:51:13 AM	126	Exportacion de dimension de Residencias	Exportacion de Dimension TD_RESIDENCIA	Yes	One or mor...
10/27/2004	3:00:02 AM	133	Importacion de dimension de Residencia	Carga de Dimension de Residencia	Yes	One or mor...
10/27/2004	3:00:04 AM	135	Importacion de dimension de Colegio	Carga de Dimension TD_COLEGIO	Yes	One or mor...

Figura 24: Vista del Journal Center de DB2. Ejecutor de tareas de limpieza, exportación y cargado de datos para el EDW

Autor: Juan Carlos Bustamante

La programación de ejecución de procesos de importación y exportación de datos se realiza a través de un scheduler.

En esta herramienta se determina la frecuencia, acciones, inicio y fin de cada tarea, que representa la ejecución de un archivo bien sea de exportación o importación.

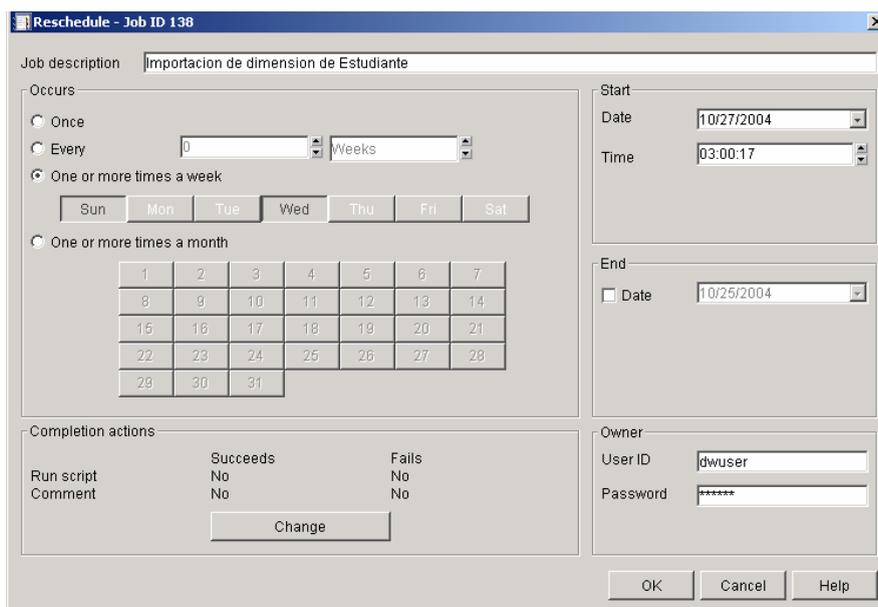


Figura 25: Vista del Scheduler. Programador de ejecución del EDW

Autor: Juan Carlos Bustamante

INTERFAZ DE CONSULTAS

En este campo entran tanto el acceso web que tiene el datawarehouse a través del EDW-Consult, así como a través de herramientas de terceros.

Una vista del sitio web creado para las consultas predefinidas del e-datawarehouse se muestra en la figura siguiente:



Figura 26: Interfaz de consulta del EDW

Fuente: <http://www.edw.csi.espol.edu.ec>

Desde esta herramienta es posible realizar las cuatro principales operaciones de lo que un modelo conceptual de datawarehouse recomienda según Kimball (1996), esto es:

Pivoting: Se rota el cubo para ver una cara en particular. Por Ej. : analizar registros referidas a los diferentes campus.

e-Data Warehouse
"Transformando datos en información."

Inicio Mapa del Sitio Contáctenos Terminología

Rep. Registros por Campus, Carrera, Año y Sectores

EDW :: ESPOL

Coloque campos de filtro aquí

Coloque campos de columna aquí

CAMPUS	Cantidad de ESTUDIANTES
CAMPUS DAULE	1543
CAMPUS PEÑAS	65112
CAMPUS PLAYAS	97
CAMPUS PROSPERINA	236616
CAMPUS SAMBORONDÓN	576
CAMPUS SANTA ELENA	5722
Total general	300686

©ESPOL 2004. Todos los derechos reservados.

Figura 27: Muestra de Pivoting del EDW-Consult

Fuente: <http://www.edw.csi.espol.edu.ec>

Slicing Dicing: Se selecciona algún subconjunto del cubo. Analizar el cubo de graduados restringiéndolo para algunos campus, carreras, años y niveles.

The screenshot shows the 'e-Data Warehouse' interface. The main content area displays a table with the following data:

AÑO	GRADO	MES	GRADO	ESPECIALIZACION	Cantidad de ESTUDIANTES
2003	INGENIERIA LICENCIATURA	ENERO			4
		FEBRERO			72
		MARZO			2
		ABRIL			1
		JUNIO			1
		JULIO			4
		AGOSTO			134
		SEPTIEMBRE			3
		OCTUBRE			2
		NOVIEMBRE			6
		DICIEMBRE			1
				Total	
2004				127	
		Total general		357	

Figura 28: Muestra de Slicing Dicing del EDW-Consult

Fuente: <http://www.edw.csi.espol.edu.ec>

Roll Up: Se agrupa por alguna dimensión determinada. Por Ej. :
 Analizar los profesores de una Unidad por años.

The screenshot shows the EDW interface with a roll-up report for the year 2004. The report is titled 'EDW :: ESPOL' and shows the number of professors by unit (UNIDAD DICTA) for the year 2004. The interface includes a navigation menu on the left and a header with the ESPOL logo and 'e-Data Warehouse' branding.

UNIDAD DICTA	ANIO PARALELO	
	2003	2004
CELEX		5
FICT	73	37
FIEC	232	104
FIMP	223	112
FMAR	121	55
ICF	53	23
ICHE	405	187
ICM	204	81
ICQ	32	14
PRTAL	24	12
PRTCO	403	80
PRTTEL	45	28
PRTWE	25	14
PRTTP	21	10
Total general	1852	752

Figura 29: Muestra de Roll-up del EDW-Consult

Fuente: <http://www.edw.csi.espol.edu.ec>

Drill Down: Operación inversa: muestra información detallada de cada agrupamiento. Por Ej. Analizar los registros por Unidad que dicta la materia y la carrera a la que fue destinada.

e-Data Warehouse
"Transformando datos en información"

Inicio Mapa del Sitio Contáctenos Terminología

Coloque campos de columna

AÑO	UNIDAD	CARRERA DESTINO	Cantidad de ESTUDIANTES
2004	CELEX		172
	FICT		304
	FIEC	FAC ING. ELECTRIC. Y COMPUTAC.	34
		ING. EN ELECTRIC. ELECTRÓNICA INDI	40
		ING. MECÁNICA	48
		INGENIERIA EN ESTADISTICA INFORM	25
		LICENCIATURA SISTEMA INFORMACI	111
		NÓ DEFINIDA	2063
		Total	2322
	FIMP		
	FIMAR		
	ICF		
	ICHE		
	ICM		
	ICQ		
	PRTAL		
	PRTCO		1506
	PRTTEL		318

Valor: 2322
Total: Cantidad de
Elemento de Filtr:
Filtro: CAMPUS =
Filtro: PERIODO =
Filtro: TERMINO =

© ESPOL 2004. Todos los derechos reservados

Figura 30: Muestra de Drill down del EDW-Consult

Fuente: <http://www.edw.csi.espol.edu.ec>

5.2.4. Implementación de las dimensiones en la Base de Datos

La implementación de las dimensiones del Datawarehouse se basa en un proceso de diseño con estructura multidimensional.

La visualización conceptual de requerimientos no tiene que ser a nivel atómico pero debe representar la vista del mundo del analista de negocios.

Las dimensiones involucradas en esta implementación corresponden a todas las detalladas en capítulos anteriores y serán explicadas en detalle en este apartado.

Se considera en primer lugar el modelo de consultas de **Registros** con las dimensiones involucradas. Se ilustra en el gráfico a continuación:

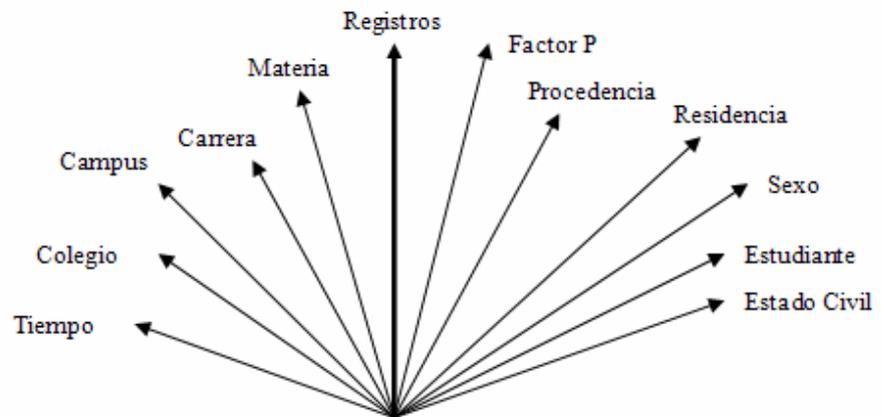


Figura 31: Modelo de consultas de Registros

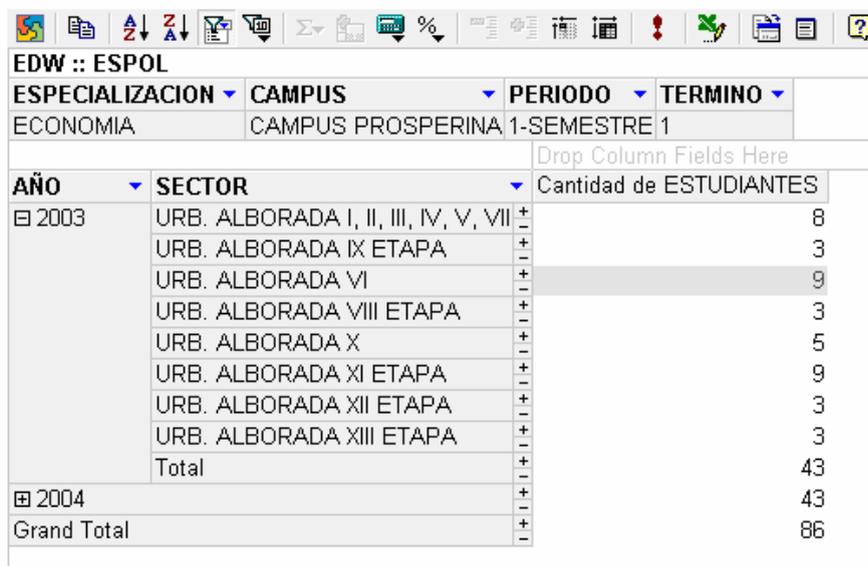
Autor : Juan Carlos Bustamante

Cada extremo del modelo representa una “cara” o criterio del cubo de datos a través del cual éste puede ser manipulado, es decir, se pueden responder requerimientos tales como:

- ¿Cuántos registrados del sexo femenino hay por campus, colegio de procedencia y año de ingreso?
- Comparación entre cantidades de graduados por carrera, factor p, ciudad de varios periodos de tiempo.
- ¿Cuál será la proyección de crecimiento (o decrecimiento) en ingreso de estudiantes para el próximo ciclo académico de una especialización determinada?
- Lista los 5 principales colegios por ciudad que aportan más estudiantes a la Escuela.

En general, los modelos de consultas para los diferentes ámbitos de este proyecto coinciden y comparten ciertos criterios (dimensiones) dada la naturaleza del área escogida, como por ejemplo, las dimensiones de sexo, estado civil, materias, procedencia, residencia, entre otras, participan en la estructura de más de un modelo de consulta por ser generales en este ámbito. A manera de ilustración se muestra el resultado de una consulta de estudiantes registrados en el campus Prosperina, pertenecientes a la carrera de Economía del año 2003 término 1, cuyo sector de residencia es La Alborada.

Rep. Registros por Campus, Carrera, Año y Sectores



The screenshot shows a report interface with a toolbar at the top and a table below. The table is titled 'EDW :: ESPOL' and has columns for 'ESPECIALIZACION', 'CAMPUS', 'PERIODO', and 'TERMINO'. The data is filtered to 'ECONOMIA', 'CAMPUS PROSPERINA', and '1-SEMESTRE 1'. The main table has columns for 'AÑO', 'SECTOR', and 'Cantidad de ESTUDIANTES'. The data is grouped by year (2003 and 2004) and then by sector (URB. ALBORADA I, II, III, IV, V, VII, IX ETAPA, VI, VIII ETAPA, X, XI ETAPA, XII ETAPA, XIII ETAPA, Total). The total number of students is 86.

EDW :: ESPOL			
ESPECIALIZACION	CAMPUS	PERIODO	TERMINO
ECONOMIA	CAMPUS PROSPERINA	1-SEMESTRE 1	
AÑO	SECTOR	Cantidad de ESTUDIANTES	
2003	URB. ALBORADA I, II, III, IV, V, VII	+	8
	URB. ALBORADA IX ETAPA	+	3
	URB. ALBORADA VI	+	9
	URB. ALBORADA VIII ETAPA	+	3
	URB. ALBORADA X	+	5
	URB. ALBORADA XI ETAPA	+	9
	URB. ALBORADA XII ETAPA	+	3
	URB. ALBORADA XIII ETAPA	+	3
	Total	+	43
2004		+	43
Grand Total		+	86

Figura 32: Reporte de Registros por Campus, Carrera, Año y Sectores

Fuente: <http://www.edw.csi.espol.edu.ec>

Se puede observar en la gráfica que existe mayor concentración de estudiantes en los sectores URB. ALBORADA VI y URB. ALBORADA XI ETAPA, con un total de 9 estudiantes que es superior en conjunto al sector URB. ALBORADA I, II, III, IV, V, VII.

El modelo de consulta de **Graduados** abarca las siguientes dimensiones mostradas en el gráfico:

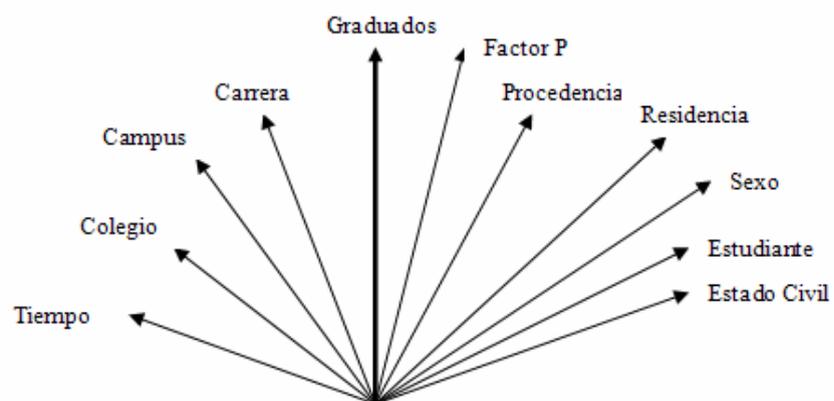


Figura 33: Modelo de consultas de Graduados

Autor: Juan Carlos Bustamante

Haciendo uso del modelo de Graduados se puede satisfacer una necesidad de información de los Graduados de las siguientes características:

- Tipo de Sangre: A+
- Nivel: Ingeniería/Licenciatura
- Status: Graduado

- Campus Prosperina
- Año de Grado 2002 y 2003
- Tipo de Grado: Tesis y Tópico
- Desglosado por sexo.

Rep. Graduados y Egresados Completo (Varios tópicos)

EDW :: ESPOL			
TIPO SANGRE	NIVEL	STATUS	CAMPUS
A+	INGENIERIA/LICENCIATURA	GRADUADO	CAMPUS PROSPERINA
Drop Column Fields Here			
ANIO GRADO	TIPO GRADO	SEXO	Cantidad de ESTUDIANTES
2002	TESIS	Femenino	7
		Masculino	14
		Total	21
	TOPICO	Femenino	1
		Masculino	5
		Total	6
Total		27	
2003	TESIS	Femenino	7
		Masculino	18
		Total	25
	TOPICO	Masculino	1
		Total	1
Total		26	
Grand Total			53

Figura 34: Reporte de Graduados y Egresados Completo

Fuente: <http://www.edw.csi.espol.edu.ec>

Con los resultados mostrados por esta consulta se puede observar que en el año 2002 el número de graduados en la modalidad de TOPICO (seis) disminuyó en el 2003 (uno), siendo compensado con los estudiantes que se graduaron en la

modalidad TESIS, que incrementaron del 2002 de 21 a 25 en el año 2003.

El modelo de consulta de **Estudiantes** abarca las siguientes dimensiones mostradas en el gráfico:

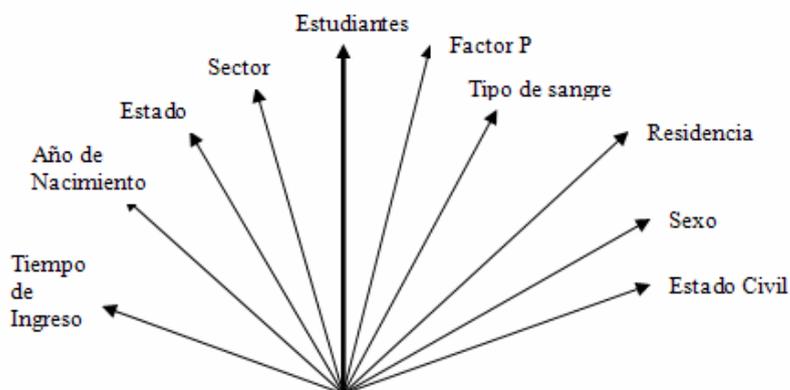


Figura 35: Modelo de consulta de Estudiantes

Autor: Juan Carlos Bustamante

Finalmente se muestra el resultado de una consulta de ingreso de estudiantes a la ESPOL con un crecimiento estimado del 12% con respecto al año anterior. Este tipo de consulta es muy útil para proyecciones financieras, de planificación de cursos y de distribución de espacio físico (aulas).

Rep. de Estudiantes por Ingreso y Datos Personales

EDW :: ESPOL				
ESTADO ESTUDIANTE		FACTOR P	TIPO SANGRE	
Activo		All	All	
Drop Column Fields Here				
AÑO INGRESO	SEXO	Cantidad de ESTUDIANTE	Proyeccion 1997	
1996	Femenino	25	28	
	Masculino	68	76.16	
	Total	93	104.16	
Grand Total		93	104.16	

Figura 36: Reporte de Estudiantes por Ingreso y Datos Personales

Fuente: <http://www.edw.csi.espol.edu.ec>

El modelo de consulta de **Profesores** abarca las siguientes dimensiones mostradas en el gráfico:

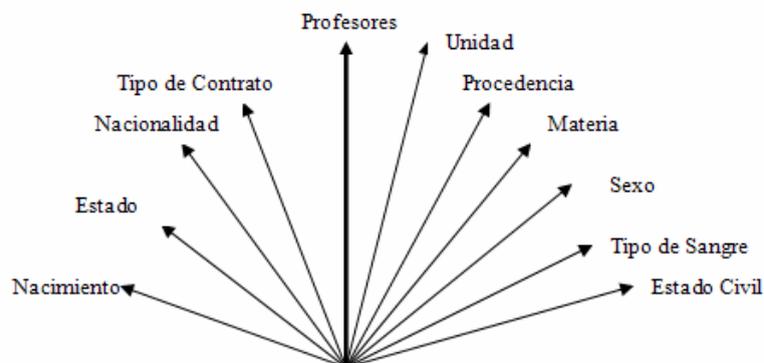


Figura 37: Modelo de consultas de Profesores

Autor: Juan Carlos Bustamante

Estos modelos permiten dar una idea más clara de lo que un diseño multidimensional persigue, que no es otra cosa que ofrecer una vista más entendible y familiar al analista del negocio (usuarios del área escogida).

Su manipulación es natural y poderosa a la vez ya que abstrae complicados esquemas relacionales y de datos en simples estructuras de negocios con fuerte orientación a su actividad.

5.2.5. Migración de datos

La migración de datos abarca tanto las fases de diseño de las transformaciones, sumarización, conversión histórica, controles, comprobación de datos, corrida del poblado y verificación auditada de información, como ya se ha mencionado en los apartados 4.6 y 5.1.

La tabla 17 lista las principales actividades realizadas en la migración de datos del datawarehouse de la ESPOL.

ACTIVIDAD	DESCRIPCIÓN
Diseño de transformaciones	Definición de requerimientos del datawarehouse Creación de procesos de limpieza, filtrado, conversión y estandarización de datos Definición de proceso de extracción de datos ya transformados.
Sumarización	Generación de campos calculados a partir de datos individuales
Conversión histórica	Se abarcó la transformación de los períodos académicos que habían sido llevados a lo largo de la historia de la ESPOL con dos componentes que eran año y término, hacia un nuevo esquema en el que se incorpora la entidad del período que representa la duración de un ciclo dependiendo de la naturaleza de la carrera bajo la que este sustentada.
Controles de transformación	Se basan en obtener totales tanto en el datawarehouse como en los sistemas OLTP de los cuales se extrajeron; dicha comprobación no es sino la verificación de estos totales en cada una de la entidades que compongan el conjunto de información
Comprobación de datos	Existe un log que detalla la cantidad de datos tanto exportados (obtenidos de las fuentes de legado) como importados (cargados en el repositorio del datawarehouse).
Corrida del poblado	Se realiza a través de procesos programados según los requerimientos de los usuarios. La programación del poblado es realizada a través de una herramienta propia del DBMS
Verificación de información	Debe ser realizada con usuarios representativos de las áreas involucradas. Requiere verificaciones de totales por rango de tiempo, comprobación de cálculos y demás auditorias de datos.

Tabla 17: Actividades realizadas en la migración de datos

Autor: Juan Carlos Bustamante

5.3. Poblado del Datawarehouse

5.3.1. *Corrida del Poblado del Repositorio*

El paso final en este esquema de implementación incluye las actividades requeridas para dar al Datawarehouse un flujo fuertemente completo, transformado, convertido, aceptado y operacional.

Varios de los pasos anteriores y de entrenamiento son completados y el usuario deberá estar contento con la utilización de la nueva aplicación.

La corrida del poblado del repositorio según Kimball (1996), “es una actividad clave en este momento”. Esta puede incluir una conversión histórica o simplemente una carga inicial al nuevo Datawarehouse.

En el Datawarehouse de la ESPOL el poblado del repositorio se lo realiza a través de procesos programados según los requerimientos de los usuarios en lo referente a tiempos de actualización, conversión histórica y cobertura del contenido.

La programación del poblado es realizada a través de una herramienta propia del DBMS (para este caso DB2), que brinda una funcionalidad completa, facilidad de uso y potencia por ser multiplataforma, este componente se conoce como *“Journal Center”*.

5.3.2. Implementación del Software en Usuarios del Sistema

La implementación del software aparece en todos los proyectos, pero el aspecto importante con el datawarehouse es asegurar que este ocurra desde muchos entornos de usuario final con actualizaciones periódicas hechas por requerimientos o desde el escritorio de un analista, antes que por la ejecución de una tarea programada regularmente. Es primordial que la herramienta o software sea implementado uniformemente a través de las estaciones de trabajo de los usuarios.

Para el caso de la ESPOL y dado que a la fecha de escribir este documento solo existe acceso vía web o por medio de un software propietario, la implementación del software en los usuarios del sistema es realizada de dos maneras:

- DW-Consult, permite el acceso vía web a la información contenida en el datawarehouse con el único requerimiento de una cuenta de acceso.
 - Si el acceso es por medio de una aplicación propietaria, deben solicitarse las respectivas autorizaciones tanto de instalación como de acceso.
-

5.3.3. Entrenamiento de Acceso y de Datos

El entrenamiento sobre los datos y el entrenamiento del acceso son a menudo englobados bajo una etiqueta: entrenamiento.

El desarrollo de un buen entrenamiento para los usuarios de este sistema puede consumir algo de tiempo. Hay que asegurar que exista énfasis sobre las maneras de encontrar un dato a través de un entrenamiento sólido.

Para garantizar el éxito del datawarehouse de la ESPOL, el usuario fue capacitado en la navegación a través de las estructuras de información implementadas. Entrenamientos específicos fueron realizados enfocando hacia donde están los datos, como utilizar las tablas y formas de consultar los datos.

La facilidad de uso de las herramientas de acceso dependerá de la familiaridad que tenga el usuario con cierto tipo de software estándar o que le sea común; como lo son navegadores de Internet, hojas electrónicas, bases de datos personales.

La flexibilidad del datawarehouse de la ESPOL en cuanto al acceso a los datos es grande ya que la información puede ser consultada a través de un browser (DW-Consult), software de terceros (Gx-Query de Genexus) y hasta manipulación directa (solo lectura) vía ODBC.

5.3.4. Prueba de Aceptación de los Usuarios del Sistema

El profesionalismo y el juego limpio imponen que existan criterios de evaluación normalizados para todos los posibles usuarios del sistema. Esto no solamente genera un carácter equitativo haciendo a todos los usuarios iguales y objetos de una consideración integral y desprovista de prejuicios, sino que también mitiga la probabilidad de comentarios y acusaciones desagradables por parte de usuarios inconformes una vez tomada la evaluación.

Los criterios de evaluación obligan a los usuarios a declarar sus puntos de vista acerca del sistema y brindan una vista de lo que ellos esperan realmente. Un ejemplo de criterio de evaluación es: “la capacidad del sistema para proveer de información sumariada y confiable en la elaboración de un informe estadístico”

En la ESPOL la prueba de aceptación por parte del usuario fue implícita ya que antes del datawarehouse el usuario analista de negocios (Programadores de CRECE) estaba en permanente dependencia del CSI para la confección de sus reportes. Una vez

instalado y capacitado el software de consulta la cantidad de requerimientos de información disminuyó radicalmente como muestra del uso continuo y aceptación por parte del usuario.

5.3.5. Control y Verificación Auditada

El control y la verificación auditada tienen lugar en el aseguramiento de que los controles internos construidos dentro del Datawarehouse sean adecuados.

La verificación de estos procesos debe ser realizada con usuarios representativos de las áreas involucradas, como el caso del Coordinador Académico de la ESPOL encargado de la auditoria interna de los datos de esta área o su equivalente.

Para el datawarehouse de la ESPOL el control de las transacciones que se realizaron en el poblado del repositorio involucran verificaciones de totales por rango de tiempo (registrados por año), comprobación de cálculos (promedios por materia) que son requeridos como aspecto crítico de este proyecto.

5.4. Seguridad Implementada

Los datawarehouses almacenan mercancías valiosas, en la forma de información acerca de los clientes, los productos patentados, las previsiones de gustos o los ingresos. Tanta información no debería estar al alcance de cualquiera, lo que explica la importancia que tiene la seguridad en datawarehousing. Según Hubel⁸¹ (1997) hay tres niveles a los que se puede implementar la seguridad:

- ✓ A nivel de las bases de datos: solamente ciertos usuarios pueden acceder a ciertas bases de datos o tablas en el datawarehouse.
- ✓ A nivel de la aplicación: ciertos usuarios pueden acceder a ciertas aplicaciones (a su vez, cada una de ellas tiene sus propios permisos de acceso a datos).
- ✓ Una combinación de seguridad de base de datos y de aplicación: este nivel puede ser personalizado a cada usuario.

⁸¹ Martin Hubel es un consultor independiente que ha trabajado extensivamente con bases de datos relacionales desde 1985, ha sido instructor de cursos avanzados de DB2 y es una autoridad reconocida en ese campo.

Existen varias formas automatizadas para imponer medidas de seguridad al datawarehouse, de las que merece mención especial el uso de middleware para que realice ciertas funciones de portería antes de permitirle a un usuario final enviar una consulta.

Para el datawarehouse de la ESPOL se ha seleccionado la tercera opción de seguridad, esto es, manejar los accesos controlando las operaciones (a nivel de aplicación) que pueden ejecutar los usuarios y los permisos en la base de datos viene dado en función de perfiles genéricos. Con esto la seguridad queda fuertemente garantizada y la administración de la misma es menos compleja.

5.5. Hardware recomendado

No es posible indicar con certeza el uso de un determinado componente de hardware debido a que cada proyecto es diferente por los requerimientos, orientación del negocio, recursos disponibles, capacidad técnica instalada, y demás condicionantes que caracterizan a cada entidad.

Como tendencia de mercado y experiencias en implementaciones exitosas previas, el empleo de procesadores con procesamiento paralelo sobre plataformas UNIX permite alta disponibilidad para los datawarehouses y un alto nivel de procesamiento transaccional con protección de integridad de datos. Esto implica, por consiguiente, añadir un nivel de complejidad al diseño y obviamente a la arquitectura de interconexión de redes.

Disponiendo de alta disponibilidad de hardware con calidad de procesamiento es posible orientar la implementación de cualquiera de las tecnologías SMP o MPP.

En un entorno multinodo o cluster, el uso de tecnología RAID con tolerancia de fallas minimiza las caídas de sistema procesamiento de transacciones para un ambiente de datawarehouse.

Es esencial que todas las piezas del datawarehouse sean probadas por medio de prototipos desde el cliente al servidor para determinar la carga del sistema a través de la red.

En la implementación del datawarehouse realizada en la ESPOL se utilizó un equipo con las características descritas en la sección 4.2.4, donde se menciona el entorno técnico preparado. La razón fundamental por el uso de una plataforma Windows en este proyecto pasa por el hecho de contar con dicho entorno destinado por el CSI para la implementación de este proyecto.

CAPÍTULO VI

CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

La implementación de un datawarehouse para la ESPOL surge con el objetivo de cubrir necesidades de información existentes de orden estadístico y especializado. Antes de este proyecto, unidades como el CRECE empleaban los datos estáticos provistos por el Sistema Académico y cuando requería información resumida tenía que solicitar al CSI dichos datos con los consiguientes problemas de dependencia en actividades propias del departamento.

El datawarehouse de la ESPOL inicialmente se creó con la extracción de información desde las bases de datos académica y personal. Los datos extraídos fueron transformados para eliminar inconsistencias y resumidos en caso que fuese necesario y luego, fueron cargados al datawarehouse. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudaron a crear un nuevo ambiente para el acceso a la información. Este nuevo enfoque ayudará a las personas individuales (analistas del negocio, usuarios, investigadores) en todos los niveles de la ESPOL, a respaldar la toma de decisiones con un fundamento histórico.

La implementación de este proyecto evolucionó la manera en que los usuarios acceden a los datos, viéndose reflejado en la disminución de carga en los sistemas fuentes al no efectuar consultas masivas que degraden su rendimiento y otorgando poder de decisión al usuario al proveerlo de información dinámica lo cual redundaba en una disminución de requerimientos de información al CSI.

Se establecieron ciertos parámetros de codificación, convenciones de nombramiento y unicidad de medidas para los datos obtenidos en los sistemas fuente, que se traduce en beneficios tangibles tales como poder contar con datos consistentes y confiables para la elaboración de reportes que sirvan de soporte en la toma de decisiones y proveer una fuente de datos histórica con datos resumidos.

El “esquema estrella” fue seleccionado para el desarrollo del modelo multidimensional por brindar ventajas en cuanto a sencillez en la manipulación de información compleja, simplicidad de implementación y facilidad de comprensión. Además garantiza flexibilidad y escalabilidad para posteriores mejoras.

Los objetivos delineados al inicio de esta investigación, en lo que concierne al desarrollo de este proyecto fueron cumplidos, incluso no fue necesaria la adquisición de software o hardware adicional ya que se emplearon recursos tecnológicos existentes en el CSI. Esto conllevó a un alto rendimiento para la inversión realizada.

Como se mencionó a lo largo de este documento la implementación del datawarehouse tiene como área objetivo la Académica. Dentro de dicha área se implementaron tres cubos de información como lo son: registros académicos, graduados y profesores. Se han creado templates de consulta que son utilizados por el DW-Consult (sistema web) para mostrar una vista de los modelos mencionados. El datawarehouse, también puede ser fácilmente consultado por herramientas de terceros, como Genexus Query al proveer una interfaz estándar de conectividad.

6.2. **Recomendaciones**

Debido a que este tipo de proyectos se realizan de manera iterativa para que su crecimiento sea paulatino e incremental, se recomienda proseguir con la cobertura de otras áreas que también pueden obtener los beneficios tangibles e intangibles que un datawarehouse puede brindar. Un área a ser considerada es la Financiera que puede ser alimentada no sólo con los datos del sistema transaccional, sino también con archivos de Excel, archivos planos y demás documentos que manejan en esa área.

Con el constante incremento de operaciones transaccionales del área académica y de ser viable de la financiera, los procesos de depuración, extracción y carga pueden afectar el rendimiento del servidor del datawarehouse actual que está soportado en una plataforma Windows, es aconsejable la migración de este servidor a uno con características más avanzadas en cuanto a capacidad de procesamiento y memoria de trabajo. De preferencia con soporte a procesamiento paralelo y bajo un sistema operativo UNIX.

La plataforma ya establecida debe dar lugar a implementar nuevos proyectos orientados a explotar y sacar el mayor provecho de esta

información, como lo son los sistemas de información ejecutivos y herramientas que brinden inteligencia de negocios a la ESPOL en la toma de decisiones. Al momento de escribir esta tesis, está planificada una investigación (tópico) en el área de datamining, la cual tiene como requisito el contar con un repositorio de datos limpio e integrado.

La difusión de este proyecto es fundamental para abrir nuevos campos de investigación a nivel académico, especialmente en aquellas unidades que realizan análisis y cálculos estadísticos, como la carrera de Ingeniería en Estadística Informática, cuyos proyectos y tesis están directamente relacionados con tratamiento de información.

DataWarehousing es un proceso, no un producto. Se confirma y recomienda que estos sistemas de información no se puedan comprar, hay que construirlos, pero en su diseño y construcción es necesario buscar estrategias tecnológicas de garantía.

CAPÍTULO VII

ANEXOS Y APÉNDICES

7.1. Apéndice A - Plan de Trabajo del Proyecto

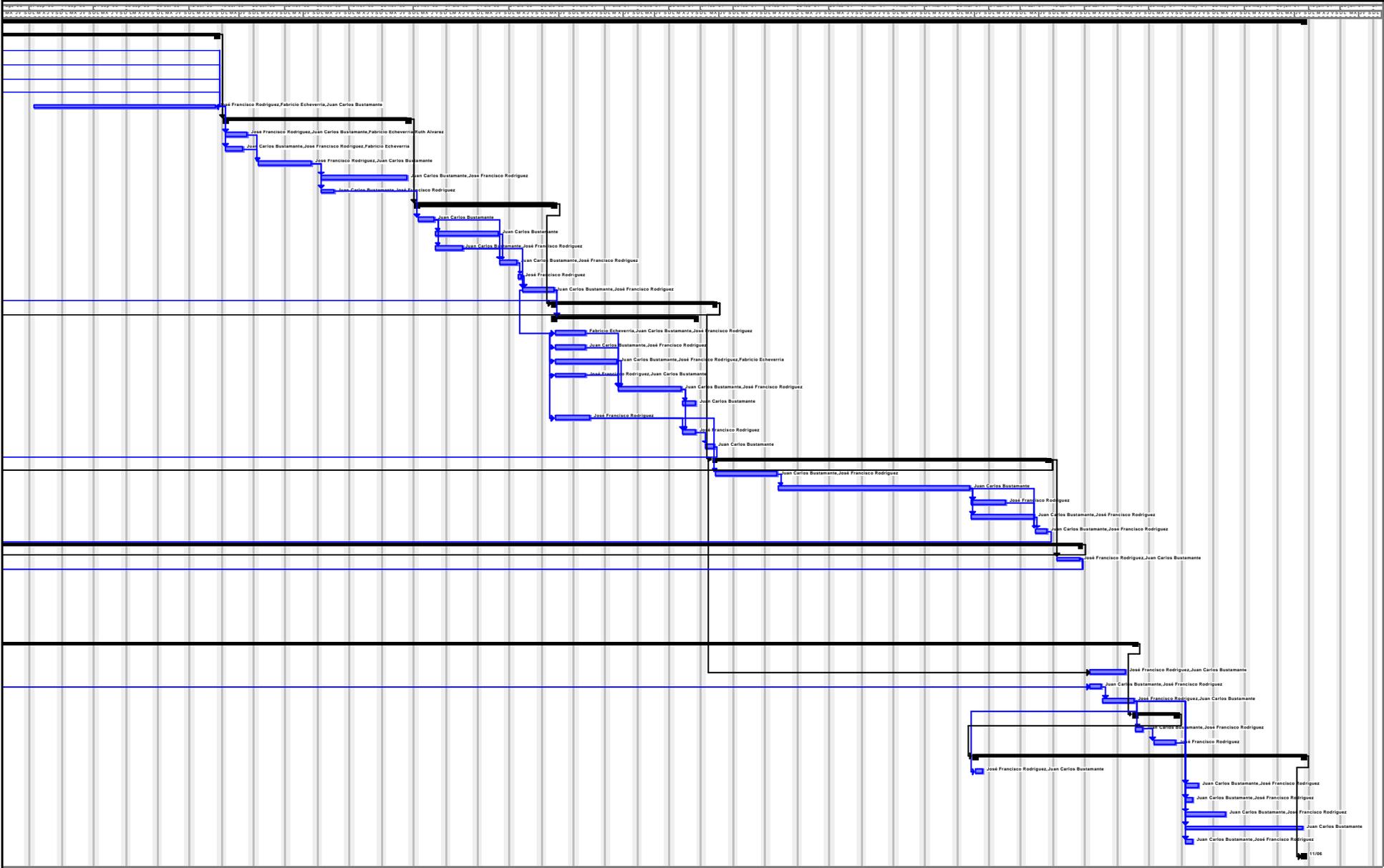
El plan de trabajo del proyecto de datawarehouse se encuentra modelado a través de un Diagrama de Gantt.

En este se muestran las tareas realizadas junto con los tiempos de entrega. Los responsables de cada tarea también son incluidos.

Estas tareas cubren la metodología aplicada en la implementación del datawarehouse para la ESPOL.

A continuación se ilustra el esquema de trabajo.

ID	Nombre de tarea	Duración	Comienzo	Fin	Precedentes	Nombre de los recursos
1	EDAT WAREHOUSE	355 días	lun 18/02/03	vie 11/06/04		
2	RECAMADO DE REQUERIMIENTOS	100 días	lun 03/06/03	vie 17/10/03		
3	NECESIDADES DE INFORMACION	5 días	lun 02/06/03	vie 06/06/03		Ruth Alvarez, Juan Carlos Bustamante
4	REQUERIMIENTOS ORGANIZACIONALES	5 días	lun 02/06/03	lun 09/06/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
5	DEFINIR LA REGION DEL PROYECTO	7 días	lun 02/06/03	mié 03/06/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
6	REQUERIMIENTOS DE ACCESO	5 días	mié 04/06/03	mié 10/06/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
7	PROTOTIPOS	30 días	lun 08/09/03	vie 17/10/03		José Francisco Rodríguez, Ruth Alvarez, Juan Carlos Bustamante
8	ANÁLISIS DEL ÁREA ESCOGIDA	30 días	lun 20/10/03	vie 28/11/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
9	NIVEL DE DETALLE	5 días	lun 20/10/03	vie 24/10/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
10	ANÁLISIS DEL ÁREA	5 días	lun 20/10/03	mié 22/10/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
11	REPORTES PRELIMINARES	10 días	lun 27/10/03	vie 07/11/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
12	MODELO DE DATOS DE LA DW	15 días	lun 10/11/03	vie 28/11/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
13	ELEMENTOS ESPECÍFICOS	3 días	lun 10/11/03	mié 12/11/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
14	FUENTES DE DATOS	22 días	lun 01/12/03	mié 20/12/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
15	FUENTES POTENCIALES	4 días	lun 01/12/03	vie 04/12/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
16	INTEGRIDAD Y PROBLEMAS OPERACIONALES	10 días	vie 05/12/03	vie 18/12/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
17	EVALUACIÓN DE CALIDAD E INTEGRIDAD	4 días	vie 05/12/03	mié 10/12/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
18	NECESIDADES DE INFORMACION	2 días	vie 19/12/03	lun 22/12/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
19	TEMAS NACIONALES/USUARIOS COBRIEROS	1 día	mié 23/12/03	mié 23/12/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
20	REQUERIMIENTOS DE METADATA	5 días	mié 24/12/03	mié 30/12/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
21	DISEÑO DE LA BASE DE DATOS	25 días	mié 31/12/03	mié 03/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
22	DISEÑO DE LAS TABLAS MULTIDIMENSIONALES	23 días	mié 31/12/03	vie 30/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
23	IDENTIFICAR LOS REQUERIMIENTOS	5 días	mié 31/12/03	mié 06/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
24	IDENTIFICAR LOS HECHOS EN LOS REQUERIMIENTOS	5 días	mié 31/12/03	mié 06/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
25	IDENTIFICAR LAS DIMENSIONES	10 días	mié 31/12/03	mié 13/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
26	IDENTIFICAR NIVELES DE ROLL-UP	5 días	mié 31/12/03	mié 06/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
27	DISEÑO DE LAS TABLAS DE HECHOS Y DIMENSIONES	10 días	mié 14/01/04	mié 27/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
28	VALIDAR EL DISEÑO CONTRA EL DBMS	3 días	mié 28/01/04	vie 30/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
29	TABLAS DE SUMARIZACIÓN	8 días	mié 31/12/03	mié 07/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
30	INDICES	3 días	mié 28/01/04	vie 30/01/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
31	BACKUPS Y GUARDES DE RECUPERACION	2 días	lun 02/02/04	mié 03/02/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
32	DISEÑO DE LAS TRANSFORMACIONES	55 días	mié 04/02/04	vie 16/04/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
33	ESPECIFICACIONES DE TRANSFORMACION	10 días	mié 04/02/04	mié 17/02/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
34	PROCESOS DE TRANSFORMACION	30 días	mié 18/02/04	mié 30/03/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
35	DISEÑO DE CONTROLES	6 días	mié 31/03/04	mié 07/04/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
36	PROCESOS DE SUMARIZACIÓN	10 días	mié 31/03/04	mié 13/04/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
37	EVALUACION DE LOS DATOS	3 días	mié 14/04/04	vie 16/04/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
38	DESARROLLO DE HERRAMIENTAS	315 días	lun 18/02/03	vie 23/04/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
39	DEFINICION DE LA ARQUITECTURA	5 días	lun 19/04/04	vie 23/04/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
40	DISEÑO E IMPLEMENTACION DE OLAP	40 días	vie 11/04/03	vie 05/06/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
41	DISEÑO E IMPLEMENTACION DE SEGURIDADES	30 días	lun 10/02/03	vie 21/03/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
42	DISEÑO E IMPLEMENTACION DE REPORTEROS	15 días	mié 15/04/03	lun 05/05/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
43	DESARROLLO DE SCHEDULER	10 días	lun 28/04/03	vie 09/05/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
44	PRUEBA DE SOFTWARE Y CORRECCIONES	5 días	vie 06/06/03	lun 12/06/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
45	POBLACION E IMPLEMENTACION	254 días	vie 13/06/03	mié 05/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
46	CORRIDA DE LA POBLACION	5 días	vie 13/06/03	vie 19/06/03		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
47	IMPLEMENTACION DEL SOFTWARE	6 días	lun 26/04/04	lun 03/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
48	PRUEBA DE SATISFACCION DE LOS USUARIOS	3 días	lun 26/04/04	mié 28/04/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
49	VERIFICACIONES DE CONTROL Y AUDITORIA	5 días	vie 23/04/04	mié 05/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
50	DISTRIBUCION DE LOS DATOS	7 días	vie 06/05/04	vie 14/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
51	HERRAMIENTAS	2 días	vie 06/05/04	vie 07/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
52	USUARIOS FINALES	5 días	lun 10/05/04	vie 14/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
53	ADMINISTRACION Y APOYO	52 días	vie 01/04/04	vie 11/06/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
54	DESARROLLO	2 días	vie 01/04/04	vie 02/04/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
55	MONTAJE	3 días	lun 17/05/04	mié 19/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
56	PRELIMINAR	2 días	lun 17/05/04	mié 19/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
57	CAPACITACION	7 días	lun 17/05/04	mié 25/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
58	SOPORTE	20 días	lun 17/05/04	vie 11/06/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
59	SEGURIDADES	2 días	lun 17/05/04	mié 19/05/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez
60	FIN	0 días	vie 11/06/04	vie 11/06/04		José Francisco Rodríguez, Juan Carlos Bustamante, Ruth Alvarez



7.2. Apéndice B - Gráfico del Proyecto Network-PERT

El gráfico del proyecto de datawarehouse se encuentra modelado también en un Diagrama de Red o de Pert.

En este se muestran las tareas realizadas junto con la secuencia de ejecución, los antecesores y predecesores, así como requisitos para su cumplimiento.

Este diagrama describe los hitos del proyecto tratados en el apartado 2.2.3 para la implementación del datawarehouse de la ESPOL.

A continuación se ilustra el diagrama de Pert.

Gráfico del Proyecto Network-PERT

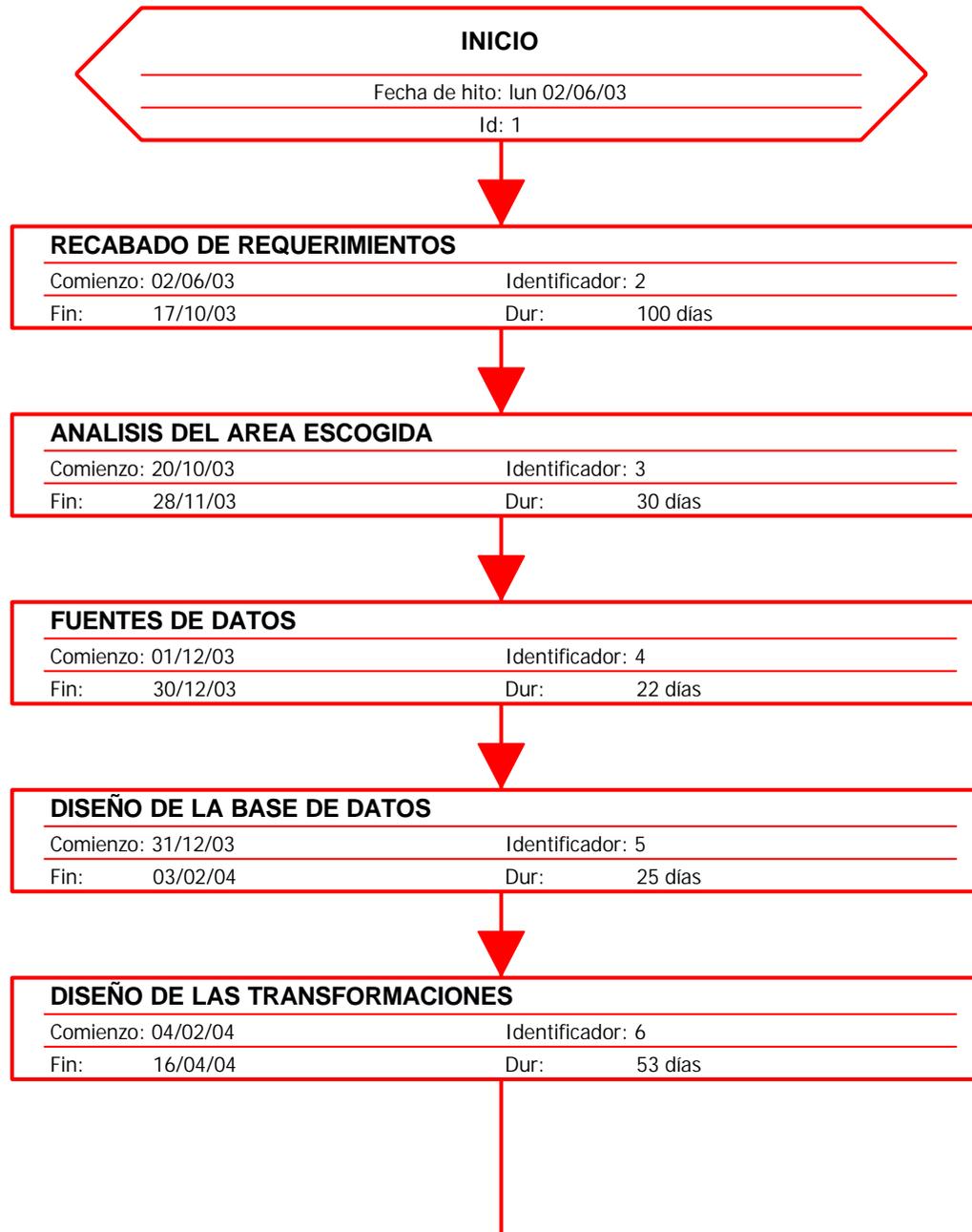


Gráfico del Proyecto Network-PERT

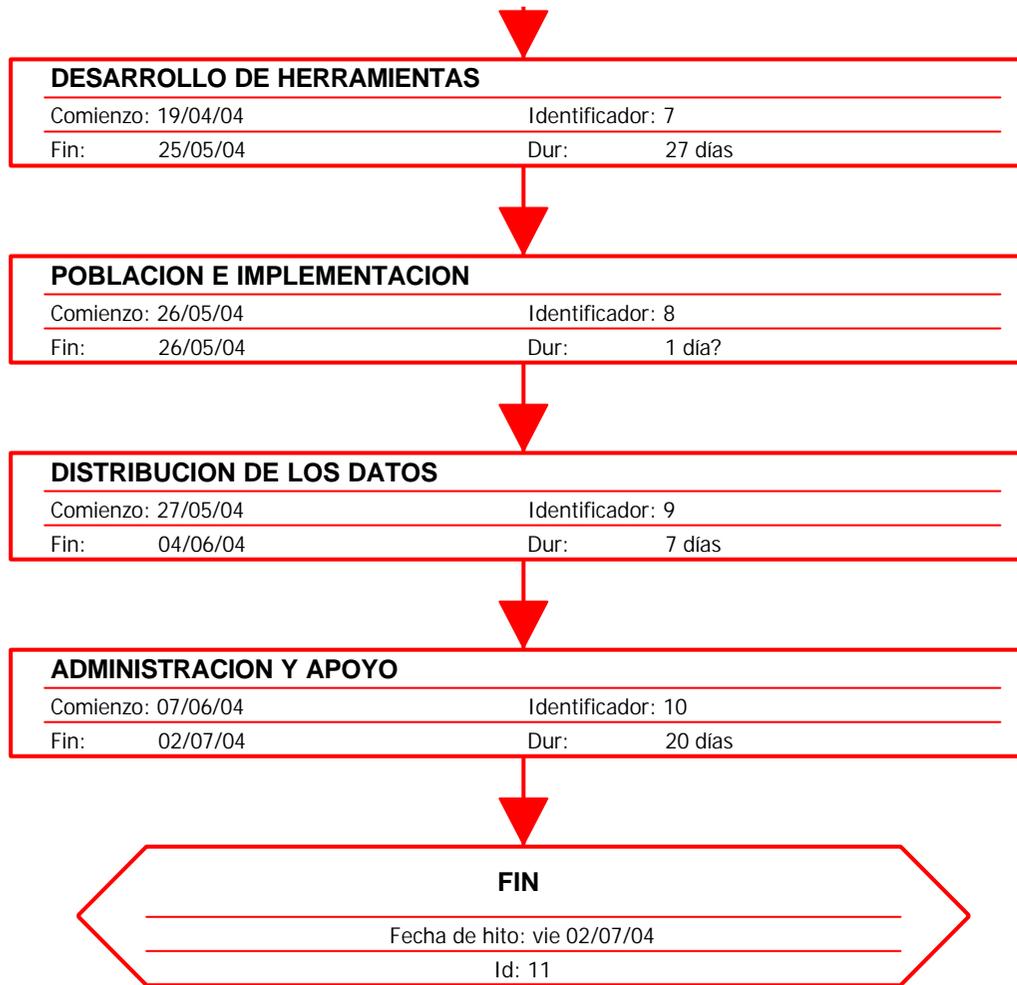
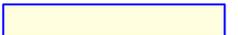


Gráfico del Proyecto Network-PERT

Proyecto: PLAN DE TRABAJO PERT.n
Fecha: jue 20/01/05

Tareas críticas	
Tareas no críticas	
Hitos críticos	
Hito	
Tareas de resumen críticas	
Tareas de resumen	
Tareas críticas insertadas	
Tareas insertadas	
Tareas críticas y marcadas	
Tareas marcadas	
Tareas externas críticas	
Externas	
Resumen del proyecto	
Tareas críticas resaltadas	
Tareas no críticas resaltadas	

7.3. Apéndice C - Directorio de Contacto del Equipo del Proyecto

Los roles de los integrantes en la implementación del EDW en la ESPOLE tomando en cuenta sus características organizacionales internas y funcionales son:

- Padrino/Sponsor – Msc. Ruth Alvarez de Estrella - Directora del Centro de Servicios Informáticos (CSI)
 - Administrador del Proyecto – Ing. Fabricio Echeverría Briones – Director de Tesis
 - Modelador de Datos – Juan Carlos Bustamante - Analista Programador y DBA
 - Analistas del Negocio – Mae. Glenda Ruíz – Coordinadora de Sistemas del CSI; Ing. Washington Medina – Coordinador del CRECE; Jorge Magallanes Borbor – Analista académico del CSI; Humberto Freire – Consultor académico de CRECE
 - Analistas de Sistemas – José Rodríguez Rojas y Juan Carlos Bustamante
 - Desarrolladores – José Rodríguez Rojas
 - Usuarios – Personal de áreas académicas y administrativas.
-

7.4. Apéndice F - Reporte de Impacto del Proyecto

El éxito de Datawarehouse no está en su construcción, sino en usarlo para mejorar procesos empresariales, operaciones y decisiones. Posicionar un Datawarehouse para que sea usado efectivamente, requiere entender los impactos de implementación en los siguientes ámbitos:

Impactos Humanos.

Efectos sobre la gente de la empresa:

- **Construcción del Datawarehouse:** Construir un Datawarehouse requiere la participación activa de quienes usarán el Datawarehouse. A diferencia del desarrollo de aplicaciones, donde los requerimientos de la empresa logran ser relativamente bien definidos producto de la estabilidad de las reglas de negocio a través del tiempo, construir un Datawarehouse depende de la realidad de la empresa como de las condiciones que en ese momento existan, las cuales determinan qué debe contener el Datawarehouse. La gente de negocios debe participar activamente durante el desarrollo del Datawarehouse, desde una perspectiva de construcción y creación.
-

- Accesando el Datawarehouse: El Datawarehouse intenta proveer los datos que posibilitan a los usuarios acceder su propia información cuando ellos la necesitan. Esta aproximación para entregar información tiene varias implicancias :
 - a) La gente de la empresa puede necesitar aprender nuevas destrezas.
 - b) Análisis extensos y demoras de programación para obtener información será eliminada. Como la información estará lista para ser accesada, las expectativas probablemente aumentarán.
 - c) Nuevas oportunidades pueden existir en la comunidad empresarial para los especialistas de información.
 - d) La gran cantidad de reportes en papel serán reducidas o eliminadas.
 - e) La madurez del Datawarehouse dependerá del uso activo y retroalimentación de sus usuarios.
 - Usando aplicaciones DSS/EIS: usuarios de aplicaciones DSS y EIS necesitarán menos experiencia para construir su propia información y desarrollar nuevas destrezas.
-

Impactos Empresariales.

- **Procesos Empresariales y Decisiones Empresariales.**
Se deben considerar los beneficios empresariales potenciales de los siguientes impactos:

a) Los Procesos de Toma de Decisiones pueden ser mejorados mediante la disponibilidad de información. Decisiones empresariales se hacen más rápidas por gente más informada.

b) Los procesos empresariales pueden ser optimizados. El tiempo perdido esperando por información que finalmente es incorrecta o no encontrada, es eliminado.

c) Conexiones y dependencias entre procesos empresariales se vuelven más claros y entendibles. Secuencias de procesos empresariales pueden ser optimizadas para ganar eficiencia y reducir costos.

d) Procesos y datos de los sistemas operacionales, así como los datos en el Datawarehouse, son usados y examinados. Cuando los datos son organizados y estructurados para tener significado empresarial, la gente aprende mucho de los sistemas de información. Pueden quedar expuestos posibles defectos en aplicaciones

actuales, siendo posible entonces mejorar la calidad de nuevas aplicaciones.

- Comunicación e Impactos Organizacionales.

Apenas el Datawarehouse comienza a ser fuente primaria de información empresarial consistente, los siguientes impactos pueden comenzar a presentarse:

- a) La gente tiene mayor confianza en las decisiones empresariales que se toman. Ambos, quienes toman las decisiones como los afectados conocen que está basada en buena información.
 - b) Las organizaciones empresariales y la gente de la cual ella se compone queda determinada por el acceso a la información. De esta manera, la gente queda mejor habilitada para entender su propio rol y responsabilidades como también los efectos de sus contribuciones; a la vez, desarrollan un mejor entendimiento y apreciación con las contribuciones de otros.
 - c) La información compartida conduce a un lenguaje común, conocimiento común, y mejoramiento de la comunicación en la empresa. Se mejora la confianza y cooperación entre
-

distintos sectores de la ESPOL, viéndose reducida la sectorización de funciones.

- d) Visibilidad, accesibilidad, y conocimiento de los datos producen mayor confianza en los sistemas operacionales.

Impactos Técnicos De Datawarehouse.

Considerando las etapas de construcción, soporte del Datawarehouse y soporte de sistemas operacionales, se tienen los siguientes impactos técnicos:

- Nuevas destrezas de desarrollo: cuando se construye el Datawarehouse, el impacto más grande sobre la gente técnica está dada por la curva de aprendizaje, muchas destrezas nuevas se deben aprender, incluyendo:
 - a) Conceptos y estructura Datawarehouse.
 - b) El Datawarehouse introduce muchas tecnologías nuevas (ETT, Carga, Acceso de Datos, Catálogo de Metadatos, Implementación de DSS/EIS), y cambia la manera que se utiliza la tecnología existente. Nuevas responsabilidades de soporte, nuevas demandas de recursos y nuevas expectativas, son los efectos de estos cambios.
-

- c) Destrezas de diseño y análisis donde los requerimientos empresariales no son posibles de definir de una forma estable a través del tiempo.
 - d) Técnicas de desarrollo incremental y evolutivo.
 - e) Trabajo en equipo cooperativo con gente de negocios como participantes activos en el desarrollo del proyecto.
- Nuevas responsabilidades de operación: Cambios sobre los sistemas y datos operacionales deben ser examinados más cuidadosamente para determinar el impacto que estos cambios tienen sobre ellos, y sobre el Datawarehouse.
-

7.5. Apéndice G – Glosario de Términos

Adición

Combinación de más de una medida en un solo valor. Algunos ejemplos son el conteo, suma, valor máximo y valor mínimo.

Agregación

Actividad de combinar datos desde múltiples tablas para formar una unidad de información más compleja, necesitada frecuentemente para responder consultas del datawarehouse en forma más rápida y fácil.

Backbone

Red principal de una red de comunicaciones.

Base de Datos (DATA BASE)

Conjunto de datos no redundantes, almacenados en un soporte informático, organizados de forma independiente de su utilización y accesibles simultáneamente por distintos usuarios y aplicaciones. La diferencia de una BD respecto a otro sistema de almacenamiento de datos es que éstos se almacenan en la BD de forma que cumplen tres requisitos básicos: no redundancia, independencia y concurrencia.

BLOB (Binary Large Object)

Objeto binario grande. Entre los tipos de datos que contienen los campos BLOB están: binarios, memo, memo con formato, de imagen, de sonido y OLE.

Cliente/Servidor

Arquitectura de sistemas de información en la que los procesos de una aplicación se dividen en componentes que se pueden ejecutar en máquinas diferentes. Modo de funcionamiento de una aplicación en la que se diferencian dos tipos de procesos y su soporte se asigna a plataformas diferentes.

Codificación

- a) Transformación de un mensaje en forma codificada, es decir, especificación para la asignación unívoca de los caracteres de un repertorio (alfabeto, juego de caracteres) a los de otro repertorio.
- b) Conversión de un valor analógico en una señal digital según un código prefijado.

Cubo

Estructura de almacenamiento para datos DW y adiciones de datos, que no está limitado a tres dimensiones.

Cubo virtual

Cubo que se deriva de uno o más cubos físicos

Datawarehouse

Base de datos que almacena una gran cantidad de datos transaccionales integrados para ser usados para análisis gestionales por usuarios especializados (tomadores de decisión de la empresa).

Datamart

Conjunto de hechos y datos organizados para soporte decisonal basados en la necesidad de un área o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un

departamento dado teniendo sólo sentido para el personal de ese departamento y sus datos no tienen porque tener las mismas fuentes que los de otro datamart.

Dataminig

Análisis de los datos para descubrir relaciones, patrones, o asociaciones desconocidas.

Diccionario de Dato

Un compendio de definiciones y especificaciones para las categorías de datos y sus relaciones. Descripción lógica de los datos para el usuario. Reúne la información sobre los datos almacenados en la BD (descripciones, significado, estructuras, consideraciones de seguridad, edición y uso de las aplicaciones, etc.).

Dimensión

Entidad independiente dentro del modelo multidimensional de una organización, que sirve como llave de búsqueda (actuando como índice), o como mecanismo de elección de datos.

Directorio de Datos

Es un subsistema del sistema de gestión de base de datos que describe dónde y cómo se almacenan los datos en la BD (modo de acceso y características físicas de los mismos).

Drill Down

Exponer progresivamente más detalle (dentro de un reporte o consulta), mediante selecciones de ítems sucesivamente.

Drill-Up

Es el efecto contrario a drill -down. Significa ver menos nivel de detalle, sobre la jerarquía significa generalizar o sumarizar, es decir, subir en el árbol jerárquico.

DSS

Sistema de Soporte de Decisiones. Sistema de aplicaciones automatizadas que asiste a la organización en la toma de decisiones mediante un análisis estratégico de la información histórica.

Esquema de estrella

Arreglo de almacenamiento DW de las tablas de Hechos y de Dimensiones en que cada tabla de Dimensión se une directamente a la tabla de Hechos. Las tablas de Dimensión no tienen tablas de claves externas. También se llama Esquema plano.

Esquema de copo de nieve (Snowflake)

Se define igual que el Esquema de estrella, a excepción que algunas tablas de Dimensión tienen anexadas tablas de Claves externas. También se llama Dimensión de tablas múltiples.

ETT (Extracción, Transformación y Transporte de datos)

Pasos por los que atraviesan los datos para ir desde el sistema OLTP (o la fuente de datos utilizada) a la bodega dimensional. Extracción, se refiere al mecanismo por medio del cual los datos son leídos desde su fuente original. Transformación (también conocida como limpieza) es la etapa por la que puede atravesar una base de datos para estandarizar los datos de las distintas fuentes, normalizando y fijando una estructura para los datos.

Finalmente está el Transporte, que consiste básicamente en llevar los datos leídos y estandarizados a la bodega dimensional (puede ser remota o localmente). Generalmente, para un Datamart no es necesario atravesar por todos estos pasos, pues al ser información localizada, sus datos suelen estar naturalmente estandarizados (hay una sola fuente).

Extranet

Constituye un servicio de comunicación orientado a un público focalizado sobre el formato de los sistemas Web, operando sobre la red Internet. Ejemplo: Una casa de ventas de productos varios, implementa un sistema de Ofertas, Consulta a Catálogos, Bancos de Datos y Compras a sus clientes preferenciales.

In-House

Aplicable a la realización de un servicio de outsourcing en las instalaciones de la organización que contrata el servicio.

Inconsistencia

El contenido de una base de datos es inconsistente si dos datos que deberían ser iguales no lo son. Por ejemplo, un empleado aparece en una tabla como activo y en otra como jubilado.

Integridad

Condición de seguridad que garantiza que la información es modificada, incluyendo su creación y borrado, sólo por el personal autorizado.

Internet

Término usado para referirse a la red más grande del mundo, que conecta miles de redes con alcance mundial. Está creando una cultura que basándose en la simplicidad, investigación y estandarización fundamentada en usos de la vida real, está cambiando la forma de ver y hacer muchas de las tareas actuales. Mucha de la tecnología de punta en redes está proviniendo de la comunidad Internet.

Intranet

Constituye un servicio de comunicación de los sistemas de información corporativos orientados a su personal, sobre el formato de los sistemas Web, operando sobre la red Internet. Ejemplo: El sistema contable de una empresa de ventas de productos de ferretería, tipo Home Center.

Jerarquía

Es un conjunto de atributos descriptivos que permite que a medida que se tenga una relación de muchos a uno se ascienda en la jerarquía. Por ejemplo: los Centros de Responsabilidad están asociados a un Tipo de Unidad, el cual pueden corresponder a una gerencia, subgerencia, superintendencia, etc.; por otro parte, cada CR está asociado a otro CR a nivel administrativo y, también existe una clasificación a nivel funcional.

MAP

- a) Conjunto de datos
 - b) Lista de datos u objetos, tal como actualmente están almacenados en memoria o en disco.
-

c) Transferir un conjunto de objetos de un lugar a otro. Por ejemplo, los módulos de programas en el disco son proyectados ("mapeados") en la memoria. Una imagen gráfica en memoria es proyectada en la pantalla.

d) Relacionar un conjunto de objetos con otro. Por ejemplo, una estructura de base de datos lógica se proyecta sobre la base de datos física.

Mapping

Proyección, correspondencia, transformación

Medición

Tipo de dato que tiene un valor a analizar.

Olap (On-line Analytical Processing)

Conjunto de principios que proveen un ambiente de trabajo dimensional para soporte decisional.

Oltip (On-line Transaction Processing)

Sistema transaccional diario (o en detalle) que mantiene los datos operacionales del negocio.

Redundancia

Repetición de los mismos datos en varios lugares.

Repositorio

Base de datos central en herramientas de ayuda al desarrollo. El repositorio amplía el concepto de diccionario de datos para incluir toda la información que se va generando a lo largo del ciclo de vida del sistema, como por ejemplo:

Componentes de análisis y diseño (diagramas de flujo de datos, diagramas entidad-relación, esquemas de bases de datos, diseños de pantallas, etc.), estructuras de programas, algoritmos, etc. En algunas referencias se le denomina Diccionario de recursos de información.

Rollup

Comando propio *del SQL mejorado que* simboliza las sumas agregadas de una variable a través de los niveles jerárquicos de las dimensiones que la sustentan.

Sistema de Gestión de Base de Datos

Software que controla la organización, almacenamiento, recuperación, seguridad e integridad de los datos en una base de datos. Acepta pedidos de datos desde un programa de aplicación y le ordena al sistema operativo transferir los datos apropiados.

Cuando se usa un sistema de gestión de base de datos, SGDB, (en inglés DBMS), los sistemas de información pueden ser cambiados más fácilmente a medida que cambien los requerimientos de la organización. Nuevas categorías de datos pueden agregarse a la base de datos sin dañar el sistema existente.

Sistema de Información (SI)

Conjunto de elementos físicos, lógicos, de comunicación, datos y personal que, interrelacionados, permiten el almacenamiento, transmisión y proceso de la información.

Snapshot

Imagen instantánea de los datos en un tiempo dado.

SQL (Structured Query Language)

Lenguaje de interrogación normalizado para bases de datos relacionales. El SQL es un lenguaje de alto nivel, no procedural, normalizado, que permite la consulta y actualización de los datos de BD relacionales. Se ha convertido en el estándar para acceder a BD relacionales. La primera versión se aprobó como norma ISO en 1987 y la segunda, conocida como SQL2 y vigente actualmente, en 1992. Actualmente se trabaja en la norma SQL3 que soportará bases de datos orientadas a objeto y bases de datos activas. El SQL facilita un lenguaje de definición de datos y un lenguaje de manipulación de datos. Además, incluye una interfase que permite el acceso y manipulación de la BD a usuarios finales.

Sumarización

Actividad de incremento de la granularidad de la información en una base de datos. La sumarización reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.

Tabla Dimensional

Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central a través de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad a esta estrategia.

Terabyte (TB)

Unidad de medida que equivale a 1024 GB.

UNIX

Sistema operativo multiproceso, multiprograma y multiusuario.

Software diseñado por AT&T para ingeniería de telecomunicación. Ha sido el primer sistema operativo concebido con independencia de los fabricantes. Posee una gran facilidad para adaptarse a ordenadores con diferentes arquitecturas, siendo ampliamente autónomo respecto del hardware. Está escrito en lenguaje de alto nivel C.

ANEXOS

Anexo 1-A

HERRAMIENTAS DE CONSULTA Y REPORTE

PRODUCTO	EMPRESA DISTRIBUIDORA
Access	<u>Microsoft</u>
Access+	<u>Sonetics</u>
Actuate Reporting System	<u>Actuate Software Corporation</u>
AMIS Information Server	<u>Hoskyns Group plc</u>
Application System	<u>IBM</u>
Approach	<u>Lotus Corporation</u>
ARPEGGIO	<u>Wall Data Inc.</u>
APTuser	<u>International Software Group</u>
AS/Access for Microsoft Access	<u>Martin Spencer & Associates</u>
ASK Joe	<u>Information Management Services</u>
aXcess/400	<u>Glenbrook Software</u>
BrioQuery	<u>Brio Technology</u>
Business Objects	<u>Business Objects, Inc.</u>
Clear: Access	<u>Sterling Software</u>
Crystal Reports, Crystal Info	<u>Seagate Software</u>
d.b. Express	<u>Computer Concepts Corp.</u>
Databoard, Dataread	<u>SLP Infoware</u>
DataDirect Explorer	<u>Intersolv</u>
DataSite	<u>NetScheme Solutions, Inc.</u>
DB Publisher	<u>Xense Technology Inc.</u>
DbPower	<u>Db-Tech Inc.</u>
Decision Analyzer	<u>Decisión Technology</u>
DECquery, DECdecision	<u>Touch Technologies, Inc.</u>
Discoverer, Discoverer/2000	<u>Oracle Corporation</u>
DS Server, DS Modeler	<u>Interweave</u>
EasyReporter	<u>Speedware Corporation</u>
Eclipse Query/Report	<u>Cornut Informatique</u>
ELF	<u>ELF Software</u>
English Wizard	<u>English Wizard</u>
EnQuiry	<u>Progress Software</u>
Esperant	<u>Speedware</u>

FOCUS Six	Information Builders, Inc.
4S-Report	Four Seasons Software, Inc
Freequery	Dimension Software Systems
Front & Center for Reporting, Nomad	Thomson Software Products
GQL	Andyne
HarborLight	Harbor Software
HP Information Access	Hewlett-Packard
if...	Leep Technology, Inc.
Impress, SqlBuddy	Objective Technologies, Inc.
Impromptu	Cognos Corporation
InfoAssistant	Asymetrix
InfoMaker	Powersoft Corporation
InfoQuery	Platinum Technology, Inc.
InfoReports	Platinum Technology, Inc.
InformEnt Warehouse Desktop	Fiserv
Internet DataSpot	DTL Data Technologies Ltd.
inSight	Williams & Partner
Interactive Query	New Generation software
IQ/Objects, IQ/SmartServer	IQ Software Corporation
Iridon Panorama	The Great Elk Company Limited
Kinetix	Hilco Technologies
LANSA/Client	LANSA USA
MARKIS/400	AS Software
Nirvana	Synergy Technologies
OR-REPORTER II	Output Reporting, Inc.
Oracle Reports, Browser	Oracle Corporation
Paradox	Borland
Platinum Report Facility	Platinum Technology, Inc
ProBit	System Builder
Productivity Series Reports	michaels. ross & cole
QBE Vision	Sysdeco
QMF	IBM
QueryObject	Cross/Z International, Inc.
Quest	Centura Software Corporation
R&R Report Writer	Concentric Data Systems
Report Writer	Raima
Reportoire	Synergistic Systems, Inc.
Reports	Nine to Five software Co.

ReporTool	Zen Software
ReportSmith	Borland
Rocket Shuttle	Rocket Software, Inc.
Safari ReportWriter	Interactive Software Systems
Sagent Datamart Solution	Sagent Technology, Inc.
SAS System	SAS Institute
Second Wind	Anju Technologies
Select!	Attachmate
SEQUEL	Advanced Systems Concepts
Snow Report Writer	Snow International Corporation
Spectrum Writer	Pacific Systems Group
SQLPRO Agent	Beacon Ware, Inc.
SQR Workbench	MITI
Star Tracker	Leep Technology, Inc.
Strategy	ShowCase Corporation
The Reporter	Sea Change Systems, Inc.
Unique XTRA	Unique AS
URSA InfoSuite	Decision Support Inc.
ViewPoint	Informix
ViewPoint	Soliton Associates
Viper	Brann Software
VisPro/Reports	Hock Ware
Visual Cyberquery	Cyberscience Corporation
Visual Dbase	Borland
Visual Express	Computer Associates International
Visual FoxPro	Microsoft Corporation
Visual Net	CNet Svenska AB
Visualizer Query, Charts	IBM
Voyant	Brossco Systems
WebBiz	Cybercom Partners
WebSeQueL	InfoSpace Inc.
WinQL	Data Access Corporation
Xentis	GrayMatter Software Corporation

Anexo 1-B
**HERRAMIENTAS DE BASE DE DATOS
 MULTIDIMENSIONAL/OLAP**

PRODUCTO	EMPRESA DISTRIBUIDORA	TIPO
Acuity ES	<u>Acuity Management Systems Ltd.</u>	MDDDB
Acumate ES	<u>Kenan Systems Corporation</u>	MDDDB
Advance For Windows	<u>Lighten, Inc.</u>	MDDDB
AMIS OLAP Server	<u>Hoskyns Group plc</u>	MDDDB
BrioQuery	<u>Brio Technology</u>	MDDDB
Business Objects	<u>Business Objects, Inc.</u>	Relacional
Commander OLAP, Decision, Prism	<u>Comshare Inc.</u>	MDDDB
Control	<u>KCI Computing</u>	Relacional
CrossTarget	<u>Dimensional Insight</u>	MDDDB
Cube-It	<u>FICS Group</u>	MDDDB
Dataman	<u>SLP Infoware</u>	MDDDB
DataTracker	<u>Silvon Software, Inc.</u>	Relacional
DecisionSuite	<u>Information Advantage, Inc.</u>	Relacional
Delta Solutions	<u>MIS AG</u>	MDDDB
Demon for Windows	<u>Data Command Limited</u>	MDDDB
DSS Agent	<u>MicroStrategy</u>	Relacional
DynamicCube.OCX	<u>Data Dynamics, Ltd.</u>	Relacional
EKS/Empower	<u>Metapraxis, Inc.</u>	MDDDB
Essbase Analysis Server	<u>Arbor Software Corporation</u>	MDDDB
Essbase/400	<u>ShowCase Corporation</u>	MDDDB
Express Server, Objects	<u>Oracle</u>	MDDDB
Fiscal	<u>Lingo Computer Design, Inc.</u>	Relacional
Fusion	<u>Information Builders, Inc.</u>	MDDDB
FYI Planner	<u>Think Systems</u>	MDDDB
Gentia	<u>Planning Sciences</u>	MDDDB
Helm	<u>Codeworks</u>	MDDDB
Holos	<u>Holistic Systems</u>	MDDDB
Hyperion OLAP	<u>Hyperion Software</u>	MDDDB
InfoBeacon	<u>Platinum technology, Inc.</u>	Relacional
Informer	<u>Reportech</u>	MDDDB/Relacional

Intelligent Decision Server	<u>IBM</u>	Relacional
IQ/Vision	<u>IQ Software Corporation</u>	Relacional
Khalix	<u>Longview Solutions, Inc.</u>	Relacional
Lightship	<u>Pilot Software, Inc.</u>	MDDB
Matryx	<u>Stone, Timber, River</u>	MDDB
MDDB Server	<u>SAS</u>	Relacional
Media	<u>Speedware Corporation</u>	MDDB
Metacube	<u>Informix</u>	Relacional
MIKSolution	<u>MIK</u>	MDDB
MIT/400	<u>SAMAC, Inc</u>	MDDB
MSM	<u>Micronetics Design Corporation</u>	MDDB
Muse	<u>OCCAM Research Corp.</u>	MDDB
OLAP Office	<u>Graphitti Software GmbH</u>	MDDB
OpenOLAP	<u>Inphase Software Limited</u>	Relacional
Pablo	<u>Andyne</u>	MDDB/Relacional
ParaScope	<u>DataVista</u>	Relacional
PowerPlay	<u>Cognos Corporation</u>	MDDB/Relacional
StarTrieve	<u>SelectStar</u>	Relacional
The Ant Colony	<u>Geppetto's Workshop LLC</u>	Relacional
TM/1	<u>Applix</u>	MDDB
Toto	<u>Ambit Research Ltd.</u>	MDDB
Track for OLAP	<u>Track Business Solutions</u>	MDDB
Visualizer Plans for OS/2	<u>IBM</u>	MDDB

MDDB: Multidimensional Data Base.

Anexo 1-C
SISTEMAS DE INFORMACION EJECUTIVOS

PRODUCTO	EMPRESA DISTRIBUIDORA	TIPO
Acuity/ES	<u>Acuity Management Systems Limited</u>	1
Applixware	<u>Applix</u>	1
BusinessMetrics	<u>Valstar Systems Ltd.</u>	1
BOARD	<u>Pragma Inform</u>	1
COINS	<u>Russell Consulting Limited</u>	1
ColumbusEIS	<u>Jitcons YO</u>	1
Commander EIS	<u>Comshare Inc.</u>	1
Corporate Management/ Financial Executive Information System	<u>Strategic Information Associates, Inc.</u>	1
CorVu	<u>CorVu Pty Ltd.</u>	1
Decision Suite	<u>Softkit</u>	1
Discovery EIS	<u>Atlantic Information Systems Ltd.</u>	1
EIS	<u>Inphase Software Limited</u>	1
Electronic Balanced Scorecard	<u>ASI Financial Services</u>	1
Enterprise Periscope	<u>Everyware Development Corp.</u>	1
Eureka	<u>European Management Systems</u>	1
ExecuSense	<u>TLG Corporation</u>	1
FOCUS EIS	<u>Information Builders, Inc.</u>	1

Forest & Trees	<u>Platinum Technologies, Inc.</u>	1
iMonitor	<u>BayStone Software</u>	1
InfoManager	<u>Ferguson Information Systems</u>	1
Iridon Almanac	<u>The Great Elk Company Limited</u>	1
inSight	<u>Arcplan Information Services</u>	2
LEADER	<u>Sterling Strategic Solutions</u>	1
MagnaFORUM	<u>Forum Systems, Inc.</u>	1
Merit	<u>GIST, s.r.o.</u>	1
Open EIS Pak	<u>Microsoft</u>	1
Panorama Business Views	<u>Panorama Business Views Inc.</u>	1
Perspectives	<u>Syntell</u>	1
Qbit	<u>Zenia Software, Inc.</u>	1
Reveal	<u>CSD Software Inc.</u>	1
SAS System	<u>SAS Institute</u>	1
Show Business EIS	<u>Show Business Software</u>	1
Tiler EIS++	<u>Avoca Systems Limited</u>	1
Track	<u>Track Business Solutions</u>	1
Traffic Control EIS	<u>Research & Planning, Inc.</u>	3
VentoMap, VentoSales	<u>Vento Software Inc.</u>	1
Virtual Headquarters Management System	<u>vHQ LLC</u>	1

Visual EIS	<u>Synergistic Software</u>	1
Visual Publisher	<u>KMA Associates International, Inc</u>	1
VITAL	<u>Braintec Corporation</u>	1
Wingz	<u>Investment Intelligence Systems Group</u>	1
Wired for OLAP	<u>AppSource Corporation</u>	1
Xecutive Pulse EIS	<u>Megatrend Systems, Ltd.</u>	1

1. Proporciona un sistema de información ejecutivo con capacidades analíticas.
 2. Proporciona un sistema de información ejecutivo con capacidades analíticas para usuarios SAP R/3.
 3. Proporciona un sistema de información ejecutivo con capacidades analíticas para usuarios SAP R/2 y R/3.
-

Anexo 1-D
**BASES DE DATOS USADAS PARA
 DATAWAREHOUSE**

PRODUCTO	EMPRESA DISTRIBUIDORA
Adabas D	<u>Software AG</u>
Advanced Pick	<u>Pick Systems</u>
DB2	<u>IBM</u>
Fast-Count DBMS	<u>MegaPlex Software</u>
HOPS	<u>HOPS International</u>
Microsoft SQL Server	<u>Microsoft</u>
Model 204	<u>Computer Corporation of America</u>
NonStop SQL	<u>Tandem</u>
Nucleus Server	<u>Sand Technology Systems</u>
OnLine Dynamic Server, Extended Parallel Server	<u>Informix</u>
OpenIngres	<u>Computer Associates</u>
Oracle Server	<u>Oracle</u>
Rdb	<u>Oracle</u>
Red Brick Warehouse	<u>Red Brick Systems</u>
SAS System	<u>SAS</u>
Sybase IQ	<u>Sybase</u>
Sybase SQL Server, SQL Server MPP	<u>Sybase</u>
SymfoWARE	<u>Fujitsu</u>
Teradata DBS	<u>NCR</u>
THOR	<u>Hitachi</u>
Time Machine	<u>Data Management Technologies, Inc.</u>
Titanium	<u>Micro Data Base Systems, Inc.</u>
Unidata	<u>Unidata, Inc.</u>
UniVerse	<u>VMARK</u>
Vision	<u>Innovative Systems Techniques, Inc.</u>
WX9000	<u>White Cross Systems Inc.</u>
XDB Server	<u>XDB Systems, Inc.</u>

REFERENCIAS BIBLIOGRÁFICAS

Libros de Administración y Toma de Decisiones:

“Administración”, James A. Stoner y R. Edgard Freeman. Editorial Prentice Hall, 1994, página 266.

“Administración”, Koontz, H., O'Donnell. Editorial Mc. Graw Hill, México 1985.

“Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer Based Marketing Program”, Arthur M. Hughes, Editorial McGraw-Hill, 1994.

Libros de introducción al DW:

“Datawarehouse: Practical Advice from the Experts”, Joyce Bischoff y Ted Alexander [1997], capítulos 1, 3, 4, 5, 9, 12.

“E-data. Transformando datos en información”, Dyché, Prentice Hall, [1996]. Capítulos 1, 2, 5, 8 y 9.

“The Datawarehouse Toolkit” Ralph Kimball, Ed. John Wiley & Sons.[1996]

“Data Warehousing”, Harjinder S. Gill y Prakash C. Rao, Prentice Hall Hispanoamericana, 1996. Capítulos 1,2, 4. Pag.5, 6, 21, 305,306

“Building the Datawarehouse” de Bill Inmon, Ed. John Wiley & Sons, Inc[1996]

“Building the Operational Data Store” Bill Inmon, Ed. John Wiley & Sons, Inc [1997]

“Designing the Database for a Datawarehouse, Chapter 9 of Building a Datawarehouse for Decision Support, Second Edition” Prentice-Hall, Poe, Vidette, Klauer, Patricia and Brobst, Stephen [1998]

“Scoping your Datamart Implementation” DBMS, Ed. J. Dyché[1998]

“Datawarehouse Administration and Management”, Information Systems Management, Benander, Alan, Fadlalla, Adam, and James, Gregory [2000]

“OLAP Solutions: Building Multidimensional Information Systems”, Eric Thomsen, John Wiley, [1997].

“Planning and Designing the Datawarehouse”, Ramon Barquin and Herb Edelstein, chapter 10, Prentice Hall PTR, [1996].

“Building a Datawarehouse for Decision Support”, Vidette Poe, Prentice Hall, [1995].

“Introducción a los Sistemas de Bases de Datos”, Ullman, Jeffrey y Widom, Jennifer. Editorial Prentice Hall, México 1999.

Artículos

"Getting set for OLAP" Michael L. González , *DB2 Magazine (spring 1998)*

http://www.db2mag.com/db_area/archives/1998/q1/98spGonz.shtml

"A Dimensional Modeling Manifesto"

<http://www.dbmsmag.com/9708d15.html>

"Bringing Up SuperMarts"

<http://www.dbmsmag.com/9801d14.html>

"Backward in Time "

<http://www.intelligententerprise.com/000929/webhouse.shtml>

"Can Datamarts Grow"

http://www.cio.com/archive/070197_gartner_content.html

"Datawarehousing "

<http://www.programacion.com/tutorial/warehouse/14/>

"Responding to Ralph : A Rebuttal to de DM Manifesto"

"Datawarehousing: Clearing the confusion"

"Datawarehousing: the fallacy of Datamart Centric Strategies (Short term Gain, Long term Pain)"

<http://www.sdw.bull.com/ve/papers.htm>

"Datamart - Datawarehouse"

http://www.dmreview.com/editorial/dmreview/print_action.cfm?EdID=1675

"Datamarts: Low Cost, High Appeal "

<http://www.techweb/cmp.com/iw/571/71iudat.html>

"La página oficial del XML"

<http://www.w3.org/XML/>

"Procesamiento Analítico en Línea (OLAP)"

<http://webepcc.unex.es/barrena/investigacion.htm#OLAP>

"Datawarehousing"

http://apuntes.rincondelvago.com/bases-de-datos_20.html

Referencias on-line:

www.teradatareview.com
www.db2mag.com
www.sqlmag.com
www.oracle.com/oramag
www.dmreview.com
www.dbmsmag.com/index.shtml
www.intelligententerprise.com
www.cio.com
www.dmreview.com
www.computerworld.com
www.metagroup.com
www.starnetinc.com/larryg
www.datawarehousing.com
www.datawarehouse.com
www.baseline-consulting.com
www.billinmon.com
www.ralphkimball.com
<http://www.intelligententerprise.com>
