



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

-

Desarrollo de una aplicación móvil de monitoreo predictivo para detecciones de sobrecarga en los recursos computacionales de un centro de datos en la ciudad de Guayaquil

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Ingeniero en Telemática

Presentado por:

Julio César Boderó Castro

César Agustín Navas Cuzme

GUAYAQUIL - ECUADOR

Año: 2020

DEDICATORIA

Este Proyecto va dedicado a Dios por sus propósitos y voluntad yo estoy culminando una etapa importante en mi vida, mi familia, mi padre Cesar Bodero Rodríguez quien fue un impulsor de que yo estudie una carrera en Espol y él que de cualquier forma me ayudaba con los materiales de estudio, además el apoyo incondicional para que yo continúe, a mi madre Patricia Castro Vásquez quien es aquella persona quien nunca dejo de levantarme los ánimos cuando yo no quería continuar y me daba esperanzas y esa fuerza para que continuara pasara lo que pasara. Mis hermanos quienes de alguna forma me ayudaban en mi formación y ellos son mi ánimo para ser su ejemplo y que ellos se formen como profesionales, y a mis demás familiares que recibí su ayuda en alguna etapa de mi vida universitaria.

Julio Cesar Bodero Castro

Dedico este proyecto de titulación principalmente a Dios por darme una familia tan unida en muchos aspectos, unos padres muy ejemplares, tanto en valores humanos y espirituales. Mi padre que se llamaba Antonio Navas Mora, quien me estará felicitando desde el cielo y alentando a seguir adelante, gracias, ¡papá! A mi mami Leonor Cuzme Villacis, quien siempre me reconfortaba cuando yo regresaba cansado de las jornadas universitarias. Mis hermanos que en su forma de ser me alentaban a seguir adelante. A mis compañero de la universidad por sus palabras de aliento y por su tiempo que desperdiciaban conmigo, en especial a mis amigos: Lino Ontano por estar junto a mi desde el inicio hasta el final en la universidad, a Carlos Cedeño por enseñarme cuando no sabía algún concepto, a Anel Martínez, Marlon Segarra y Gabriel Roque quienes me sacaron muchas sonrisas en los proyecto que hicimos , a toda la IEEE por brindarme sus instalaciones y experiencia en el desarrollo de eventos y a los amigos que conocí en niveles inferiores de mi carrera adelante que ustedes pueden!

César Agustín Navas Cuzme

AGRADECIMIENTOS

Nuestro sincero agradecimiento a la empresa TELCONET S.A y al ingeniero encargado del centro de datos, el Ing. Xavier Cabrera por brindarnos el acceso a los equipos TI del centro de datos y su disponibilidad de tiempo en las dudas presentadas en el desarrollo de este proyecto de titulación. También agradecemos a la M.Sc. Adriana Collaguazo, quien fue una mentora de quien hemos aprendido mucho durante las materias que vimos con ella, además de su ayuda con el proyecto de titulación desarrollado. Al Ing. Vladimir Sánchez Padilla, quien estuvo con nosotros guiándonos en el desarrollo y validación de este proyecto, y al Dr. José Córdova García, quien nos ofreció su apoyo, en cuanto al desarrollo de diferentes ideas implementadas en las etapas del proyecto.

TRIBUNAL DE EVALUACIÓN

.....
Vladimir Sánchez Padilla

PROFESOR DE LA MATERIA

INTEGRADORA

.....
Adriana Collaguazo Jaramillo

TUTOR ACADÉMICO

DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, me(nos) corresponde conforme al reglamento de propiedad intelectual de la institución; (nombre de los participantes) y doy(damos) mi(nuestro) consentimiento para que la ESPOI realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”



Julio Bodero castro



César Navas Cuzme

RESUMEN

Los centros de datos son una infraestructura virtual que disponen de servidores basado en sistemas hipervisores ESXi, los que se encargan de aplicar diversas técnicas de control del hardware y software virtual con el fin de utilizar diferentes sistemas operativos al mismo tiempo en una computadora o un conjunto de ellas, llamado clúster. Dentro de los clústeres, se alojan diferentes sistemas operativos virtualizados, que ofrecen recursos a sus clientes. Estos recursos están destinados al alojamiento de espacios para los equipos de cómputo, como almacenamiento de información y despliegue de servicios. Dentro de un centro de datos existen tres departamentos: los encargados del edificio, los de monitoreo y los de Tecnología de la información (TI). En este último, por lo general, existen operadores o administradores del centro de datos, los que se encargan de hacer reportes manuales de las capacidades computacionales disponibles.

Este documento dará a conocer un sistema de monitoreo móvil, mediante la utilización de las herramientas disponibles por parte de la empresa. Por ejemplo, con Vcenter es posible realizar las capturas de los recursos computacionales utilizados, con las que se procede a realizar en varias épocas el entrenamiento del algoritmo y predicciones inteligentes de los recursos. Esta predicción permitió identificar cuando los recursos de un equipo alcanzaron el umbral del 25%. Luego se procedió al diseño de la aplicación móvil encargada de visualizar los diferentes hosts y sus máquinas virtuales con los datos relevantes obtenidos en la fase anterior. Cabe resaltar que el aplicativo móvil se diseñó con la finalidad de visualizar los recursos por medio de graficas estadísticas, realizar operaciones manuales, como el apagado y el aumento de recurso en un equipo, y la generación de alertas automáticas y reportes manuales y automáticos de los recursos disponibles. De esta forma se minimiza la pérdida de clientes potenciales para la empresa.

Palabras Clave: hipervisores ESXi, Centro de datos, capacidades computacionales, predicción inteligente, host, recursos disponibles.

ABSTRACT

Data centers are a virtual infrastructure that has servers based on ESXi hypervisor systems, which are responsible for applying various techniques to control both virtual hardware and software in order to use different operating systems at the same time on a computer or a set of them, called cluster. Within the clusters, different virtualized operating systems are hosted, offering resources to their clients. These resources are intended to accommodate spaces for computer equipment, such as information storage and service deployment. Three departments are defined in a data center: one in charge of the building, other responsible for the monitoring and a third one in charge of the Information Technology (IT). In the latter, there are operators or administrators, who are responsible for making manual reports of the available computing capabilities. This document will introduce a mobile monitoring system, using the tools available from the company. For instance, with vCenter it is possible to capture the computational resources used, where the algorithm training and intelligent predictions of the resources are carried out several times. This prediction made possible to identify when the resources of a device reach the 25% threshold. Then, we proceeded to design the mobile application for displaying the different hosts and their virtual machines with the relevant data obtained in the previous phase. It should be noted that the mobile application was designed to visualize the resources through statistical graphs, perform manual operations, such as turning off and the increase in resource on a computer, the generation of automatic alerts, and both manual and automatic reports of the available resources in order to minimize the loss of potential customers for the company.

Keywords: ESXi hypervisors, Data centers, computational capabilities, smart prediction, host, available resources.

ÍNDICE GENERAL

RESUMEN.....	6
ABSTRACT.....	7
ÍNDICE GENERAL	8
ABREVIATURAS	10
SIMBOLOGÍA	11
ÍNDICE DE FIGURAS.....	12
ÍNDICE DE TABLAS.....	13
CAPÍTULO 1.....	14
1. INTRODUCCIÓN.....	14
1.1. Descripción del Problema.....	14
1.2. Justificación del Problema	15
1.3. Objetivos.....	16
1.3.1. Objetivo General.....	16
1.3.2. Objetivos Específicos	16
1.4. Marco Teórico.....	17
1.4.1. Centro de dato	18
1.4.1.1. Definición.....	18
1.4.1.2. Tendencias de centro de datos.....	18
1.4.2. Recursos computacionales.....	19
1.4.2.1. Definición.....	19
1.4.3. Teoría de Algoritmo LSTM.....	20
1.4.3.1. Redes neuronales.....	20
1.4.3.2. Redes LSTM.....	20
CAPÍTULO 2.....	22
2. METODOLOGÍA.....	22

2.1.	Diseño de la Investigación	23
2.2.	Enfoque de la Investigación.....	23
2.3.	Población, muestra y muestreo	24
2.4.	Técnicas e instrumentos de recolección de datos	25
2.5.	Técnicas e instrumentos para el procesamiento de datos.....	28
2.6.	Especificaciones Técnicas y consideraciones éticas y legales.....	28
2.6.1.	Arquitectura Backend	28
2.6.2.	Arquitectura Frontend.....	29
2.6.3.	Consideraciones éticas y legales.....	29
CAPÍTULO 3.....		30
3.	RESULTADOS Y ANÁLISIS.....	30
3.1.	Tiempos en generar reportes.	30
3.2.	Análisis de la muestra.....	31
3.3.	Apreciación de los resultados	35
3.4.	Análisis de costos del sistema.....	51
3.4.1.	Costo del software	51
3.4.2.	Costo del Hardware	54
3.4.3.	Costo del desarrollo de ALERT D.C	57
3.5.	Análisis de los resultados	57
CAPÍTULO 4.....		59
4.1.	Conclusiones	59
4.2.	Recomendaciones	60
BIBLIOGRAFÍA.....		62

ABREVIATURAS

ESXi: ESX Integrated.

RAM: Random Access Memory.

CPU: Central Processing Unit.

IT: Information Technology.

RNN: Recurrent Neural Network.

LSTM: Long short-term memory.

CPD: Data Center.

CAPEX: Capital Expenditure.

SQL: Structured Query Language

SLA: Service Level Agreement.

VPN: Virtual Private Network.

OPEX: Operational expenditures.

ARMA: Autoregressive Moving Average models

SIMBOLOGÍA



Notificación para alertas



Indicador que se realizó una selección

ÍNDICE DE FIGURAS

Figura 1 Red Neuronal LSTM (Google Colab, 2015).....	21
Figura 2 Figura 24. Red Neuronal (Duggan et al., 2018).....	21
Figura 3 Formulas Red LSTM (Google Colab, 2015).	22
Figura 4 Equipos del centro de dato, en el mes de mayo del 2020	24
Figura 5 Diagrama de despliegue.....	26
Figura 6 Diagrama de Árbol de la base de datos NoSQL.....	27
Figura 7 Distribución de máquinas virtuales con alerta y sin alertas	31
Figura 8 Cantidad de máquinas virtuales alertadas y no alertadas.	32
Figura 9 Pantalla clústeres del sistema Vcenter visualizadas en la aplicación móvil ALERT D.C.	36
Figura 10 Menú de la aplicación.	37
Figura 11 Información básica de Clúster.	38
Figura 12 Pantalla de servidores con sus alertas.	39
Figura 13 Información de servidores host.....	40
Figura 14 Pantalla de máquinas virtuales de servidores.	41
Figura 15 Información de máquinas virtuales.	42
Figura 16 Notificaciones de nuevas alertas.	43
Figura 17 Envío de reporte por email.	44
Figura 18 Correo electrónico recibido.....	45
Figura 19 Reporte parcial de todos los hosts.....	46
Figura 20 Graficas del reporte parcial de todos los hosts.....	47
Figura 21 Reporte consolidado.....	48
Figura 22 Reporte consolidado gráficas de RAM	49
Figura 23 Reporte consolidado gráficas de RAM Disponible y CPU disponible	50
Figura 24 Reporte consolidado servidores encendidos.	51

ÍNDICE DE TABLAS

Tabla 1. Comparativa de tiempos en la generación de reportes.....	30
Tabla 2. Máquinas virtuales con distribución uniforme entre alertas y no alertas.	33
Tabla 3. Máquinas virtuales con distribución “no uniforme” entre alertas y no alertas.	33
Tabla 4. Máquinas virtuales alertadas.	34
Tabla 5. Máquinas virtuales “no” alertadas.	34
Tabla 6. Gasto generador por parte de la empresa sin la aplicación móvil.....	52
Tabla 7. Gasto generado por parte de la empresa con la aplicación móvil.....	52
Tabla 8. Gasto del servicio en la nube.....	53
Tabla 9. Costo Totales en Software.....	53
Tabla 10. Costo del nuevo equipo para la predicción.	54
Tabla 11. Gasto en hardware del centro de datos.	55
Tabla 12. Depreciación de los activos.	56
Tabla 13. Valores de CAPEX y depreciaciones en el año 2020	56
Tabla 14. Costo del ALERT D.C	57

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. Descripción del Problema

Los centros de datos contienen una infraestructura tecnológica en sus servidores basada en sistemas Hipervisores ESXi. Estos se distribuyen en grandes clústeres conformados por máquinas físicas, los cuales alojan máquinas virtuales de clientes que ofrecen diferentes servicios. El monitoreo de estos recursos computacionales recae sobre seis administradores del centro de datos. Estos normalmente realizan reportes manuales de los recursos en cada clúster del centro de dato, siendo estas características físicas y lógica. A pesar de validar el correcto funcionamiento de cada servidor y los servicios alojados en cada uno, este proceso no es fiable ni factible para la empresa.

Estos reportes afectan directamente al centro de datos, debido a que se desconocen los recursos que disponen, lo cual hace perder uno o varios clientes por desconocer las capacidades computacionales disponibles en sus servidores. Además, de no saber cuándo un cliente excede los recursos asignados, lo cual podrían estar exigiéndoles mayores recursos al servidor principal, para procesar todos los requerimientos. Los factores que requiere monitorearse en los clientes son memoria RAM, almacenamiento en disco por parte del host y los recursos lógicos de los servicios virtualizados en las máquinas virtuales.

1.2. Justificación del Problema

En base al estudio realizado, surge la necesidad de monitorear el uso de los recursos de hardware como prevención ante la llegada de nuevos clientes que puedan requerir una cantidad de recursos computacionales superior a la disponible, más aún si previamente no hubo una actualización de hardware oportuna. En caso de que los recursos disponibles en el centro de datos no abastezcan los requerimientos del nuevo cliente, es evidente que una actualización de hardware es necesaria, por lo que se inicia el proceso para adquirir nuevos componentes. Considerando que la mayoría de los componentes de hardware, como procesadores y memorias RAM, son importados con un tiempo de llegada promedio de 30 a 60 días, se puede perder la posibilidad de adquirir un nuevo cliente (Roque Gabriel y Orellana Allan, 2019).

Además, mediante entrevista realizada al Gerente de IT se pudo conocer el proceso de monitoreo que realiza esta empresa en su centro de datos. Él indicó que el proceso que realizan no es el apropiado, puesto que no detalla con exactitud la cantidad de recursos físico y lógicos que dispone por medio de los reportes manuales, *“dicho reporte se lo realiza de forma mensual y no ofrece la veracidad de cuantos recursos dispone el centro de datos en tiempo real”* (Xavier Cabrera, 2019). Por otro lado, el tiempo que se demora en traer un hardware o un equipo para poder brindar servicio a nuevos clientes es aproximadamente entre 30 a 60 días, según lo indicado por el personal encargado. Los centros de datos necesitan de un proceso más automatizado, el cual ayude a los administradores del centro de datos poder realizar la generación de reportes en cualquier momento que el departamento lo necesite. De igual modo se deben realizar predicciones cuando un equipo del centro de datos exceda sus recursos computacionales.

1.3. Objetivos

1.3.1. Objetivo General

Crear una aplicación móvil de monitoreo en un centro de datos desarrollada en un entorno de programación de aplicaciones híbridas para la visualización resumida de datos más representativos requeridos por el cliente, creación y envío automático de reportes, generación de alertas predictivas de las capacidades físicas de los centros de datos.

1.3.2. Objetivos Específicos

- ❖ Desarrollar una aplicación móvil multiplataforma para el muestreo de los recursos utilizados por las máquinas físicas y las máquinas virtuales alojados en ellas.
- ❖ Generar reportes automáticos en base a la información técnica y capacidades de las maquinas físicas del centro de datos.
- ❖ Desarrollar un sistema de envío automático de reportes mediante correo electrónico.
- ❖ Implementación de un sistema de alertas automáticas para la notificación cuando un servidor se encuentra en su máxima capacidad mediante predicciones obtenidas con el algoritmo LSTM.

1.4. Marco Teórico

En el mundo, centros de datos grandes como Microsoft utilizan sistemas hipervisores desarrollados por la propia empresa. Microsoft mediante una entrevista indica *“Nuestra estrategia es considerar que la virtualización es una parte más del sistema operativo Windows Hyper-V Server”* (Lores Serrano, 2019) Así mismo muchos centros de datos como Google, AWS, utilizan sistemas de hipervisores muchos de ellos basados en VMware con costos muy elevados a diferencia de Microsoft que ofrece el sistema dentro del paquete de software de Windows Server (Lores Serrano, 2019).

La automatización y el manejo de los recursos de un centro de datos es la principal prioridad, por la cual muchas empresas están dispuestas a mejorar su infraestructura para poder estar dentro del mercado. En primer lugar, será importante definir algunos conceptos claves para el caso de estudio. Sean estos centros de datos, recursos computacionales físico y virtualizados y por último como utilizamos los datos para realizar predicciones para el correcto uso de los recursos del centro de datos.

Un centro de datos en Ecuador, utilizando 29 servidores separados en 5 clústeres, contiene aproximadamente 4 TB de memoria RAM en total, con 70% de utilización promedio. Esta cantidad de memoria está en constante aumento para satisfacer las demandas de nuevos clientes del centro de datos. Actualmente, cuando la utilización promedio sobrecarga el 75% la capacidad total, se debe realizar una mejora de hardware para de esta forma se prever que siempre haya disponibilidad suficiente de memoria para nuevas máquinas virtuales. En cuanto al procesamiento, la capacidad total es de 1.15 THz, de los cuales se utilizan en promedio 320 GHz. Esto es tan solo un 28% de utilización, por lo que las mejoras de hardware con respecto a la memoria RAM son mucho más frecuentes a las de procesadores. Tomando en cuenta que esta utilización se alcanza con 481 máquinas, el consumo promedio de memoria de cada una es de 5.7 GB y 2.39 GHz de procesamiento (Roque Gabriel y Orellana Allan, 2019).

Los centros de datos en Ecuador en base a información obtenida de entrevistas como la realizada a nuestro cliente, compra software desarrollado por otras compañías que venden su producto ofreciendo alto estándar de seguridad. Ecuador se encuentra muy atrasado en cuanto a desarrollar sistemas desde cero

y con programación en bajo nivel, por lo que no toma en cuenta el desarrollo de sistemas robustos y complejos (Proaño et al., 2012)

1.4.1. Centro de dato

1.4.1.1. Definición

Un centro de datos, también conocido como centro de procesamiento de datos (CPD), es el lugar físico o virtual donde las empresas realizan sus operaciones y levantan su infraestructura IT, además de poder alojar espacio de almacenamiento para sus clientes (Jianhua Peng ,Hui Zhou, 2020). Dentro de estos se encuentran los clústeres, los cuales alojan sus servidores y almacenan sus datos empresariales o la de sus clientes. Por lo general, estos centros de datos están bajo un control continuo de la humedad y la temperatura de tal manera que pueda garantizar un correcto funcionamiento de estos.

Es indispensable conocer las características ambientales y de infraestructura que necesitan estos clústeres los cuales ofrecen servicios virtualizados con capacidades computacionales a sus clientes. Estos servicios virtualizados por lo general son ejecutados en máquinas virtuales para sus clientes, los cuales consumen los recursos del servidor físico o también llamado host.

1.4.1.2. Tendencias de centro de datos

Al momento de pensar en tipos de centro de datos directamente pensamos en qué tipo de arquitectura y a que requisitos se rigen estas arquitecturas. La arquitectura o el tipo de centro de datos dependerá de cuál será la finalidad de este. Por ejemplo, un centro de datos para un proveedor de servicios (ISP) con servicios en la nube, requiere de una infraestructura robusta y una buena administración y despliegue de seguridad. Mientras un centro de datos privado para alguna organización se centrará en datos clasificados (Jianhua Peng ,Hui Zhou, 2020) (Shuguang Qi ; Yu Zhang ; Mengdi Wang, 2019).

Los tipos de centros de datos difieren de sus capacidades físicas e infraestructura IT, los recursos que disponga la empresa y el modelo de negocio a aplicar. Por tanto, se puede mencionar en base a el contenido de (Shuguang Qi ; Yu Zhang ; Mengdi Wang, 2019) que las tendencias más relevantes son:

- Almacenamiento de datos más eficientes y robustos.
- El nivel al que estará sometido la virtualización de los servidores.
- El tamaño del centro de datos, como pueden ser en un proxy de telefonía utilizando microcentro de datos y mega centros de datos(Aymen Abdullah Alsaffar ; Eui-Nam Huh, 2015).
- En base al modelo de capas a utilizar en base a su arquitectura empresarial.

En la actualidad las tendencias que están siendo aplicadas en los centros de datos son que su capacidad de almacenamiento será en un 88% en la nube y llegaran a tener capacidades a 1.8 ZB. Además, pueden surgir nuevos cambios a nivel tecnológico y para el uso de telefonía móvil torres de telefonía aplicando micro centro de datos (Shuguang Qi ; Yu Zhang ; Mengdi Wang, 2019).

Conocer la arquitectura y la tendencia de los centros de datos es muy cambiante y dependiente a la finalidad del centro de datos y del modelo de negocio que este por aplicar. Por ejemplo, si es un proveedor de servicios (ISP) o un centro de datos que ofrezca servicios en la nube mediante la virtualización, como podrían ser VMware, este necesitaría mejorar los ámbitos de disponibilidad, seguridad y rendimiento excepcionales para los clientes de estos centros de datos.

1.4.2. Recursos computacionales

1.4.2.1. Definición

Hablar de recursos computacionales o informáticos refiere a todo recurso físico o virtual limitado que disponga un sistema informático. Cada hardware conectado a un sistema informático es un recurso disponible por utilizar (Weichao Ding ; Fei Luo ; Chunhua Gu ; Haifeng Lu ; Qin Zhou, 2020). En el caso de los centros de datos, se denomina recursos computacionales a los

recursos que dispone el centro de datos hacia sus clientes, sean estos relacionados a mayor velocidad en tiempo de cómputo por una petición de parte del cliente o a una solicitud de almacenar datos en CPU. Los cuales llegan a limitar los demás requerimientos por parte de otros clientes hacia el servidor físico. Por tanto, las capacidades computacionales de estos centros de datos deben de ser muy basta en base a la cantidad de clientes que este tenga a su disposición.

Optimizar los recursos computacionales de un centro de datos es lo que tratan de realizar los centros de datos en la actualidad y lo que han encontrado es que la computación en la nube es la solución, es amigable con el medio ambiente puesto que reducen el consumo de energía y disminuye las emisiones de carbono. Otro factor muy importante es la optimización en el desempeño en tiempo de respuesta, la elasticidad y escalabilidad en los centros de datos (Canal Synnex Comstor, 2016). Por lo tanto, realizar virtualizaciones en la nube es la mejor forma de escalabilidad en los centros de datos.

1.4.3. Teoría de Algoritmo LSTM

1.4.3.1. Redes neuronales

Las redes neuronales son aproximadoras de funciones inspirados por las redes neuronales biológicas que constituyen el cerebro humano. Algunas de las aplicaciones de las redes neuronales incluyen: generación de energía. La Figura 1 ilustra la arquitectura para una red neuronal que está dispuesta en varias capas. La capa de entrada es responsable de tomar las entradas al modelo, la capa oculta es donde se realiza la gran mayoría de los cálculos y la capa de salida produce la salida del modelo. (Duggan et al., 2018)

1.4.3.2. Redes LSTM

Las redes de memoria a corto y largo plazo, generalmente llamadas "LSTM", son un tipo especial de RNN, capaces de aprender dependencias a largo plazo. Fueron introducidos por Hochreiter y Schmidhuber (1997), siendo refinados y popularizados por muchas personas en el siguiente trabajo(Google Colab, 2015). Funcionan tremendamente bien en una gran variedad de problemas, y ahora se usan ampliamente.

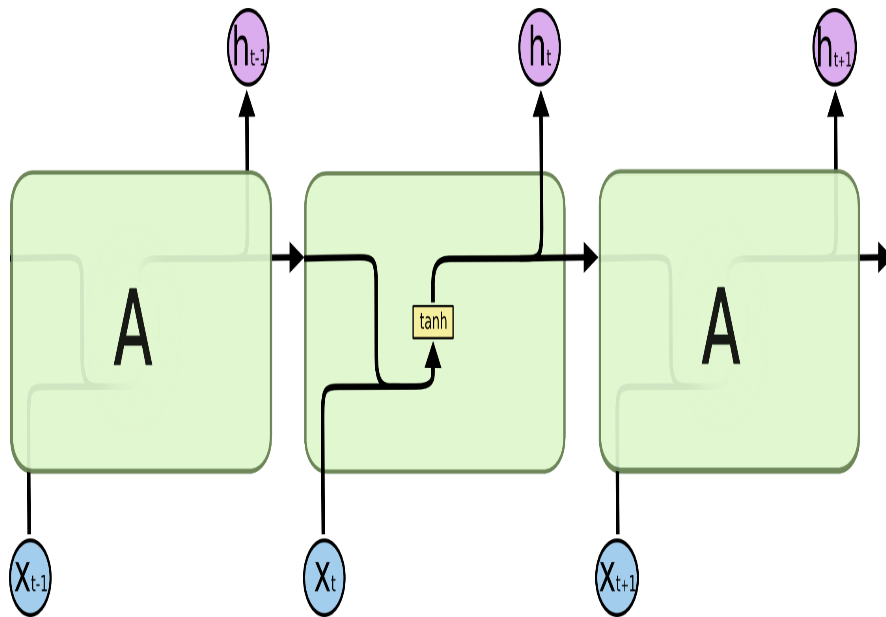


Figura 1 Red Neuronal LSTM (Google Colab, 2015)

Los LSTM están diseñados explícitamente para evitar el problema de dependencia a largo plazo. Recordar información durante largos períodos de tiempo es prácticamente su comportamiento predeterminado, sin ser catalogado como algo que les cuesta aprender.

Todas las redes neuronales recurrentes tienen la forma de una cadena de módulos repetitivos de red neuronal. En los RNN estándar, este módulo repetitivo tendrá una estructura muy simple, como una sola capa de TANH.

(Google Colab, 2015)

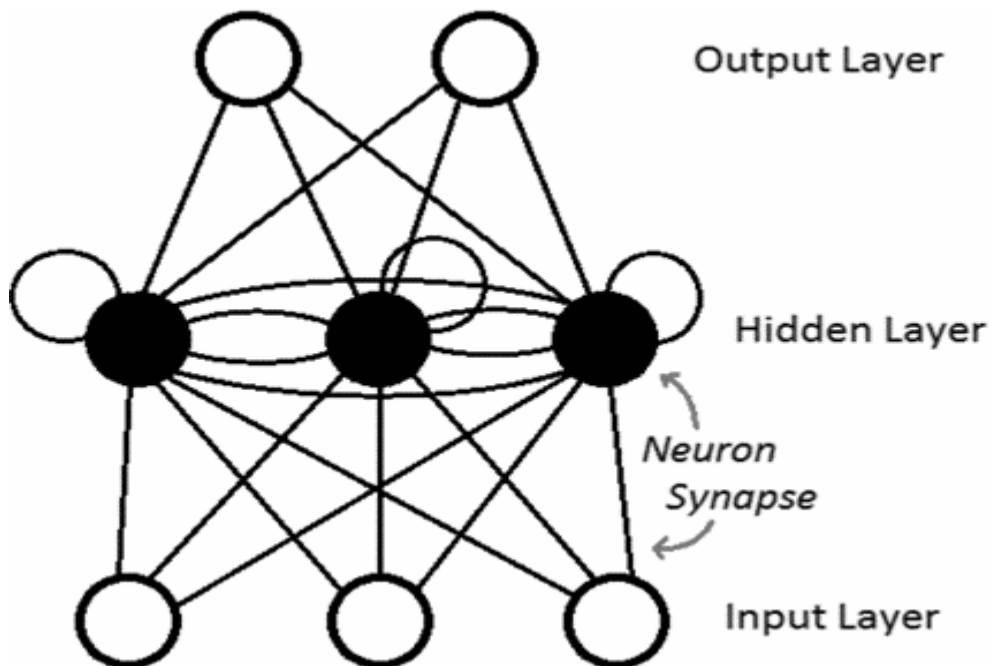


Figura 2 Figura 24. Red Neuronal (Duggan et al., 2018).

Para las fórmulas descritas en la Figura 1 tenemos que i_t, f_t, o_t representa las funciones de entrada descrita en el esquema de la Figura 2 el cual contiene 3 celdas que corresponden a las 3 funciones Gate que obtenemos en nuestra Red, tenemos σ el cual corresponde a la función de activación en la que se realiza la evaluación para obtener la respuesta de nuestro modelo, recordando que la función sigmoide es la más utilizada pues se encuentra acotada entre 0 y 1 respectivamente, w_x representa cuantas neuronas obtendrá cada Gate descrito en la Figura 2, h_{t-1} Representa la salida anterior obtenida del algoritmo LSTM, x_t Representa el tiempo de muestreo utilizado para obtener los datos, b_x Representa el bias de la ecuación para obtener que tanto la gráfica sube o baja en el eje de las Y. (Google Colab, 2015)

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i)$$
$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f)$$
$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o)$$

Figura 3 Formulas Red LSTM (Google Colab, 2015).

CAPÍTULO 2

2. METODOLOGÍA

En esta sección se explica el tipo de diseño y enfoque que cubre la implementación del proyecto, así como también los diferentes criterios de diseño y como estos afectarán en la efectividad y eficacia del sistema de alertas y reportes incluidos en la aplicación móvil, de tal forma que pueda brindar estos servicios al centro de datos.

2.1. Diseño de la Investigación

Dado que el objetivo del proyecto es monitorear el centro de datos y proporcionar al cliente la facilidad de envío automático de reportes y generación de alertas predictivas de las capacidades físicas de los centros de datos, se procedió a recurrir a un diseño no experimental y se lo enfocó de manera transversal. Considerando que el proyecto está sustentado teóricamente, se procedió a realizar un análisis de tipo descriptivo de los recursos del centro de datos, para luego dar a conocer detalladamente los recursos computacionales en los clúster, host y máquinas virtuales.

(Hernández-Sampieri, R.; Fernández Collado, C. y Baptista Lucio, 2014) definen como diseño no experimental a la capacidad de investigar algún fenómeno en un contexto natural sin la necesidad de realizar manipulación en sus variables (p.185). Estos autores mencionan la función de un diseño transversal la cual es la recolección de los datos en un instante de tiempo, para poder describir su relación e incidencia en un momento dado (p.187).

Además, de poder capturar y ubicar los contadores de rendimiento de los recursos computacionales del centro de datos, estas variables están relacionadas a la Memoria, Almacenamiento y CPU en los clústeres, hosts o máquinas virtuales. Mediante los objetos almacenados en VMware Vsphere.

2.2. Enfoque de la Investigación

Esta sección se realizó bajo el esquema del enfoque cuantitativo, puesto que el caso de estudio en este proyecto lo amerita y cumple con la características y ámbitos importantes del desarrollo del proyecto.

(Hernández-Sampieri, R.; Fernández Collado, C. y Baptista Lucio, 2014) Indica lo siguiente, lo que caracteriza al enfoque cuantitativo es la recolección de datos en base a la medición numérica y poder establecer alguna predicción en base a la información adquirida (p. 37). El enfoque cuantitativo se lo aplicó en el ámbito de la predicción de nuestro algoritmo LSTM, el cual ayudó a predecir cuándo se excedan los recursos computacionales del centro de dato, clúster, host o máquina virtual.

La técnica que se utilizó en el desarrollo de este proyecto fue la recolección de datos cuantitativos, la cual estuvo enfocada en la obtención de los recursos computacionales del centro de datos.

2.3. Población, muestra y muestreo

Una población es el conjunto de elementos a estudiar. En nuestro caso, el conjunto de elementos serían los centros de datos de la ciudad de Guayaquil. Por tanto, nuestra muestra sería los servidores físicos y virtuales alojados en el centro de datos del cliente.

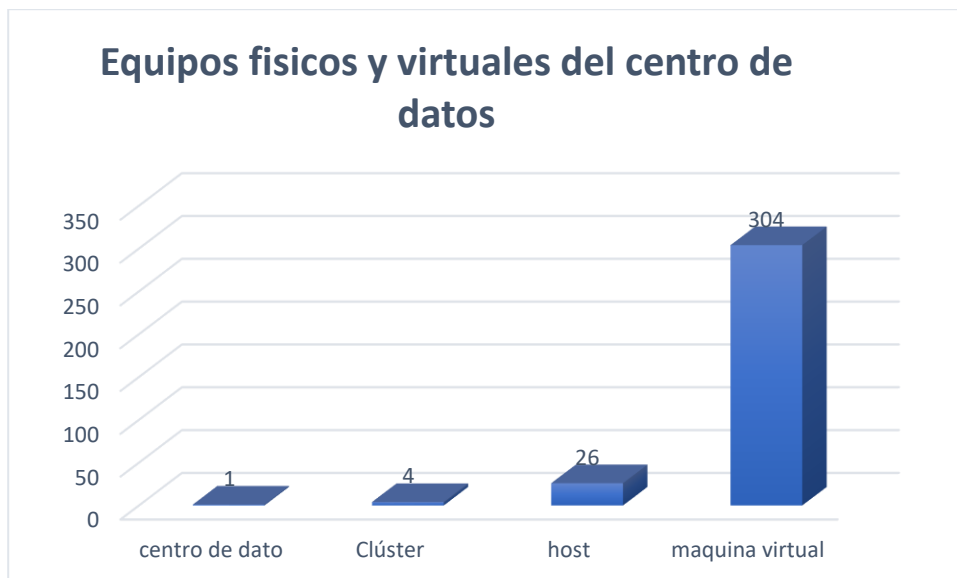


Figura 4 Equipos del centro de dato, en el mes de mayo del 2020

Este caso de estudio se centrará en una muestra no probabilística. Según (Hernández-Sampieri, R.; Fernández Collado, C. y Baptista Lucio, 2014), una muestra no probabilística es la que no depende de la probabilidad, sino de los elementos disponibles en el momento y del planteamiento del estudio (p.209). Por tanto, los recursos computacionales del centro de datos se comportan como una muestra no probabilística dependiendo de las necesidades de los clientes del centro de datos.

2.4. Técnicas e instrumentos de recolección de datos

(Falcón & Herrera, 2005) definen como técnica de recolección de datos como el procedimiento o la forma peculiar de la recolección de la información en el desarrollo de una investigación (p.12). La técnica para la recolección de datos que se va a utilizar en la implementación del proyecto es la técnica de diccionario de datos, la cual almacena detalles y descripciones de dichos elementos. En este proyecto estos elementos se llaman contadores de rendimiento los cuales almacena información de los recursos computacionales del centro de datos(VMware-vSphere 6.5, 2017).

Las encuestas realizadas al cliente indican que el centro de datos funciona por medio de la plataforma de Vcenter, perteneciente a VMware Vsphere, la cual se encarga de monitorear los objetos del inventario, sean estos centros de datos, clúster, host y máquinas virtuales (Vmware, 2019). Por lo tanto, Vcenter se encarga de supervisar los recursos computacionales del centro de datos.

Para la recolección de los recursos computacionales el medio más óptimo es por medio del API de Vsphere, puesto que permite acceso a los objetos del centro de datos por medio de VMware-Vsphere. Una posible solución alternativa podría ser el desarrollo de una plataforma basada en peticiones, pero al no disponer de mayor tiempo de desarrollo, se seleccionó el consumo de la API de Vsphere para la obtención de los recursos computacionales en el centro de datos.

Fueron proporcionadas credenciales para realizar consultas a la plataforma de Vcenter. Se crearon unos scripts en Python para capturar los contadores de rendimiento de CPU y memoria del centro de datos. Estos scripts se diferencian dependiendo a qué tipo de solicitud que se realice. Estas solicitudes pueden variar en clúster, host o máquina virtual. Python posee una librería muy útil que se llama Pyvmomi, la cual ayuda en la recolección de los contadores de rendimiento para luego almacenarla en una base de datos en MongoDB. A continuación, el diagrama de despliegue del proyecto.

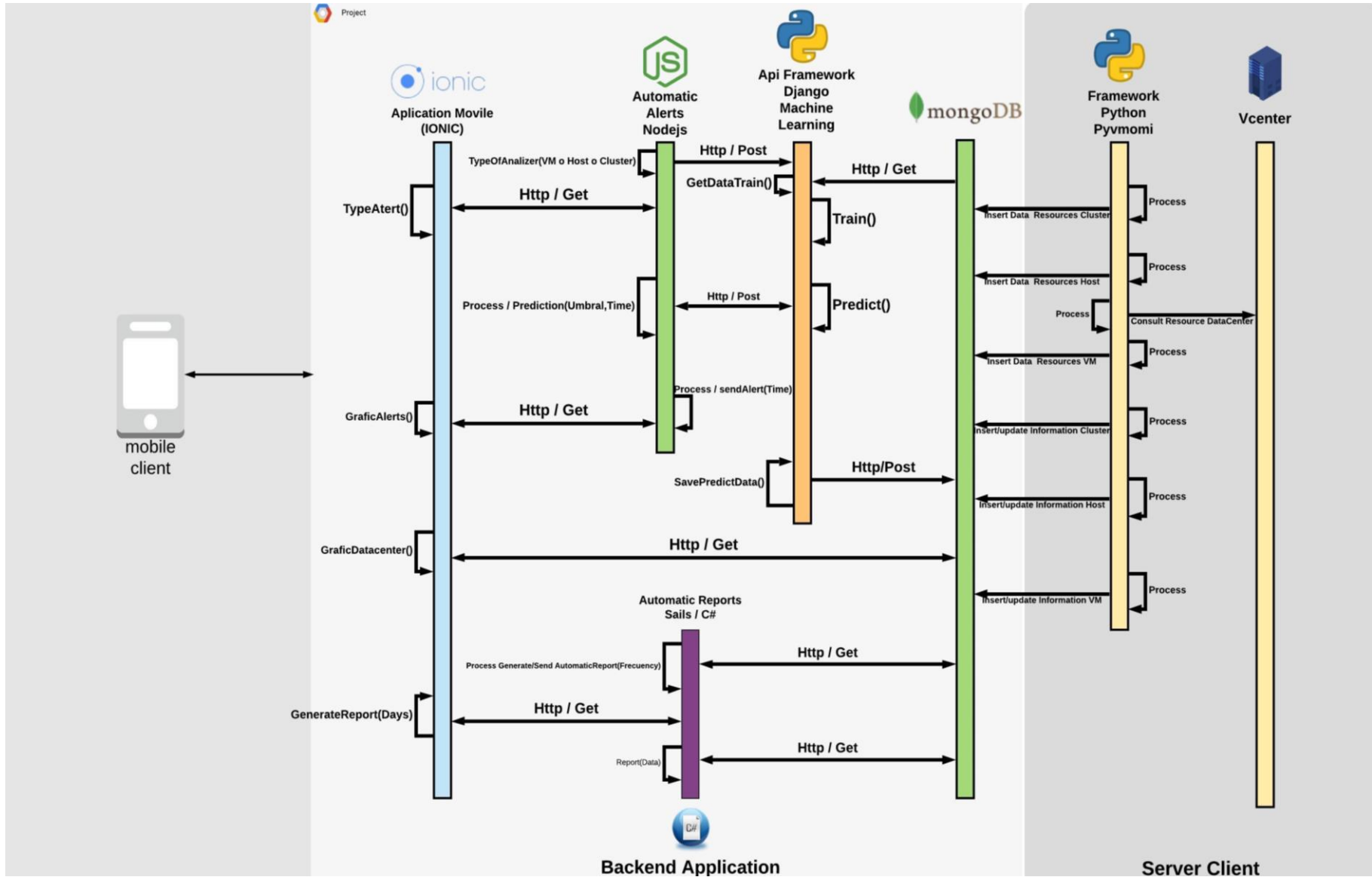


Figura 5 Diagrama de despliegue

El diagrama de despliegue del sistema completo explicará a mayor profundidad en la sección 2.6. El Framework de Pyvmomi se encarga de extraer los contadores de rendimiento alojado bajo la administración de Vcenter, para luego insertar y actualizar los registros de clúster, host y máquina virtual en la base de datos en MongoDB. En base a un estudio realizado en comparar el tiempo de inserciones y búsquedas con dos modelos de bases de datos MySQL y MongoDB (NoSQL). Se concluye que el rendimiento de la base de datos de MongoDB es mucho mayor en comparación al de una base de datos MySQL, porque el tiempo de inserción y de búsqueda de una base de datos en MongoDB es mucho más rápido y flexible a gran escala. Mientras, su contra parte aún mantiene el mapeo de objeto-relacional, lo que atrasa los tiempos de procesamiento de estos (B et al., 2016).

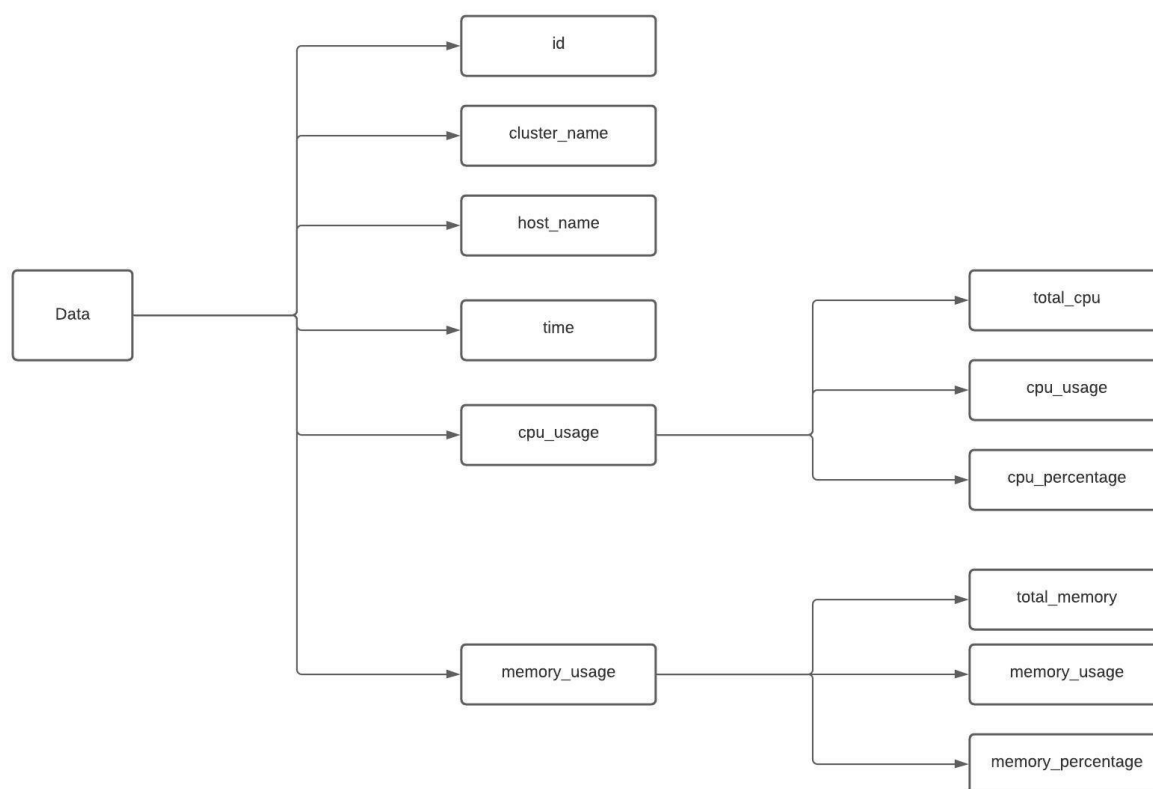


Figura 6 Diagrama de Árbol de la base de datos NoSQL

2.5. Técnicas e instrumentos para el procesamiento de datos

La técnica que se utilizó para el procesamiento de datos será la estadística descriptiva, puesto que consiste en la representación de grandes masas de datos por medio de gráficos o tablas. Por otro lado, la herramienta o instrumento que se utilizó fue el algoritmo de redes neuronales. El procedimiento que se realizó fue entrenar la red en base a datos individuales de host, máquina virtual o clúster. De esta forma se podrá realizar predicciones del tiempo en que un equipo se encuentra por encima del 25% y generar una alerta automática con su respectivo informe para validar el modelo de redes neuronales, Se utilizó el error cuadrático medio para validar que tanto se equivoca el modelo y un score de varianza. Si el valor es próximo a 1, indica que el modelo es muy bueno, entre otras métricas como el accuracy, matriz de confusión, recall y otras más.

En aplicativo móvil muestra los gráficos de los recursos computacionales de los clúster, host y máquina virtual. Se utilizaron técnicas de tratamiento de datos para eliminar valores nulos que no aportan al análisis, por otra parte, se implementan algoritmos que estimen cuando una máquina virtual ha mantenido un valor de umbral sostenido y de esta forma generar una alerta.

2.6. Especificaciones Técnicas y consideraciones éticas y legales

2.6.1. Arquitectura Backend

Dentro del servidor de la cliente se accede mediante VPN proporcionada por la empresa se ejecutan en segundo plano procesos de capturar los datos del CPU, memoria RAM, disco duro de todos los clústeres, máquina física y máquinas virtuales. Estos datos se envían la base de datos mongoDB alojada en una máquina virtual de Google Cloud. Dentro del servidor del cliente corre un proceso que actualiza la información de las máquinas agregadas en Vcenter. El framework Django implementa los métodos de consulta a la base de datos para obtener la información del centro de datos y poder utilizar los datos para mostrar al usuario final. Además, se implementaron procesos que se encargaron del entrenamiento de la red y realizaron la predicción en base a un rango de fechas ingresadas.

El sistema de alerta en Nodejs está corriendo procesos que verifican si existe alerta o no, el umbral para la alerta es proporcionado por el cliente, y el tiempo de predicción se realiza en base al tiempo de entrenamiento que ha tenido la red, en base a estos datos se generó un arreglo de tiempo y se generaron las alertas. El tipo de alerta que se generó dependerá del usuario final el cual se le muestra opciones para generar alertas, sea estos para clúster, host y máquinas virtuales o combinaciones de ellas. Cabe destacar que mientras más opciones se seleccionen, más tiempo se demore. Por tanto, si no se tiene un sistema escalable para tratamiento de grandes datos los tiempos de respuesta de la alerta, serán cada vez más largos.

EL sistema de generación de reportes automático con el Framework Sails / C# obtiene los últimos datos almacenados en la base de datos para generar el reporte de capacidades necesario para la empresa. Dicho reporte contiene graficas estadísticas para gerencia, así como datos reales de capacidad utilizada y disponible de los recursos de CPU, RAM y almacenamiento.

2.6.2. Arquitectura Frontend

Se utiliza el Framework Ionic para aplicaciones híbridas, utilizando la librería de React para el desarrollo de componentes de la aplicación en ella se programa la funcionabilidad de la vista que interactúa con el usuario y son generación de gráficas y alertas. El utilizar código JavaScript facilita el desarrollo de aplicaciones para el usuario final, además de que se creará el compilado de aplicaciones para sistemas Android y IOS. Ambos sistemas podrán utilizar la aplicación para administración del centro de datos.

2.6.3. Consideraciones éticas y legales

Consideraciones éticas y legales se refieren a métodos o estándar que utiliza Google cloud. Según el acuerdo de nivel de servicio (SLA), es cuando ocurre un problema sea por parte de Google o por parte del cliente (Google Cloud, 2020) o niveles de servicios de red (NTL) que menciona las optimizaciones que se le puede realizar en la nube a nuestro Compute Engine (Google Cloud, 2018). Por último, también no divulgar la información que brindó el cliente del centro de datos sobre sus equipos e información de índole confidencial. Este punto refiere a proteger la identidad de los nombres, dirección IP, capacidades

de CPU, RAM y almacenamiento de los equipos, utilizando etiquetados simples como pueden ser en la sección de nombre “host_1”.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

3.1. Tiempos en generar reportes.

Por medio del algoritmo de redes neuronales se logró generar reportes de los recursos computacionales del centro de datos. En el desarrollo del proyecto se enfatizó el tiempo que toma a los administradores del centro de datos generar un reporte de forma manual con un operario y la comparativa la cual es generar el reporte por medio de la propuesta desarrollada la cual es el aplicativo móvil. A continuación, se detalla la cantidad de tiempo utilizada en cada uno de los escenarios descritos

Tabla 1. Comparativa de tiempos para generar reportes.

Forma de generar Reporte	Tiempo transcurrido
Manual (Antes)	1 o 2 días
Aplicativo móvil (Después)	6 segundos

Levando el tiempo en segundos que le toma en generar el reporte de forma manual es de 86,400 segundos, realizando el respectivo porcentaje de diferencia, cuanto mejor es la generación de reporte con el aplicativo móvil que generarlo de forma manual, existe una relación de que tan bueno es un escenario respecto a otro escenario y es la siguiente:

$$\frac{\text{Diferencias}}{\text{Referencia}} \times 100 \quad (\text{Porcentaje de comparación})$$

$$\frac{86400 \text{ seg} - 6 \text{ seg}}{86400 \text{ seg}} \times 100 = 99.993 \%$$

El porcentaje refleja en cuanto tiempo ha mejorado este proceso de generación de reportes del centro de datos en casi 99.993%. Esto último indica el tiempo que

le toma al aplicativo móvil generar un reporte de los recursos, en comparación al proceso manual que utiliza el centro de datos que es muy deficiente al tiempo que le toma a la empresa generar un reporte de los recursos.

Esto es muy importante puesto que desconocer los recursos que dispones como empresa que brinda servicios y no poder adquirir nuevos clientes u optimizar los recursos que disponen sus clientes, este proceso no es para nada eficiente. Por tanto, lo óptimo y rentable para el centro de datos tanto a corto y largo plazo, es la realización de este “ALERT D.C”.

3.2. Análisis de la muestra

Esta sección explica el tipo de análisis y distribución que se dio a los datos “dataset”. Primero se ve la cantidad de equipos alertados de las 304 máquinas virtuales (MV). Al intentar realizar este análisis, observamos que habían MV de las cuales siempre pasaban alertadas. Se considera una alerta cuando el recurso del equipo alcanza o supera el 25%. En la Figura 4 se observa la distribución de alertas en las diferentes MV. Para este conjunto de datos se visualiza que hay algunas máquinas virtuales con alertas al 100% y sin alertas casi a un 0% del total de sus datos por lo que dichas MV que se encuentran siempre alertadas no se incluyeron en el análisis para el aprendizaje del Algoritmo.

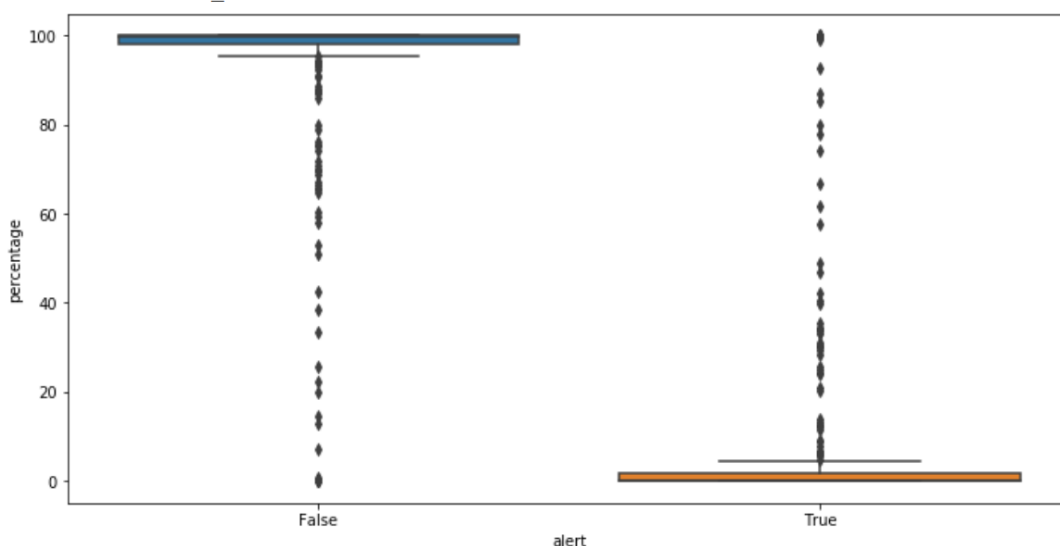


Figura 7 Distribución de máquinas virtuales con alerta y sin alertas

Una vez detectadas la cantidad de MV alertas, se pudo saber cuántos equipos están alertados y cuantos no lo estaban. La Figura 5 muestra que, de un total de 304 máquinas virtuales, existen 288 que no se encuentran alertadas a más del

25% del umbral establecido y 16 de máquinas se encuentran alertadas en la mayoría de sus puntos.

Criterios

Alerta: Puntos que sobrepasan el 25% del porcentaje de utilización del CPU.

Máquina virtual alertada: Máquina virtual donde la mayoría de sus puntos son alertas.

Máquina virtual no alertada: Máquina virtual donde la mayoría de sus puntos no son alertas.

Resultados

Maquinas virtuales:304

Máquina virtual alertada:288

Máquina virtual no alertada:16

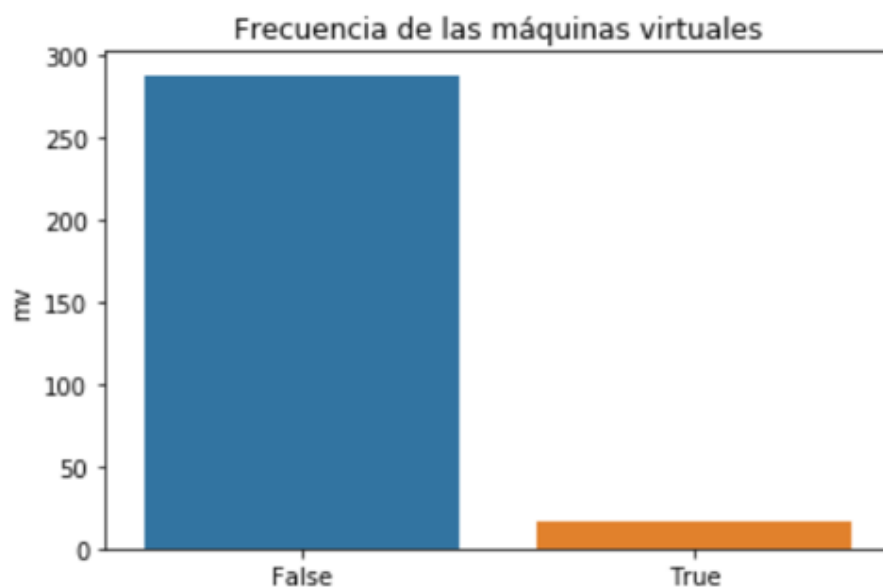


Figura 8 Cantidad de máquinas virtuales alertadas y no alertadas.

Luego se procedió a ver qué tipo de distribución seguía la muestra, sobre si era uniforme o no uniforme. La Tabla 2 describe las máquinas virtuales con puntos distribuidos de manera uniforme entre alertadas y no alertadas. En otras palabras, la cantidad de puntos por máquina virtual es equilibrada cerca de un 60% y 40% entre tener alertas y no tener alertas. La columna *alerted* indica si la máquina virtual debe tener alerta o no.

Tabla 2. Máquinas virtuales con distribución uniforme entre alertas y no alertas.

mv	alert_percentage	no_alert_percentage	alerted
V.M_260.csv	46.979	53.021	False
V.M_50.csv	42.0031	57.9969	False
V.M_88.csv	49.0299	50.9701	False
V.M_150.csv	57.5366	42.4634	True
V.M_27.csv	40.564	59.436	False

Se describe una lista de máquinas virtuales con puntos distribuidos de manera no uniforme entre alertadas y no alertadas en la Tabla 3. En otras palabras, la cantidad de puntos por máquina virtual no es equilibrada y varía entre 0 y 100% entre tener alertas y no tener alertas. La columna alerted indica si la máquina virtual debe tener alerta o no.

Tabla 3. Máquinas virtuales con distribución “no uniforme” entre alertas y no alertas.

mv	alert_percentage	no_alert_percentage	alerted
V.M_1.csv	0	100	False
V.M_155.csv	0	100	False
V.M_220.csv	1.86448	98.1355	False
V.M_150.csv	0.0349589	99.965	False
V.M_100.csv	0.0640914	99.9359	False
...
V.M_158.csv	0	100	False
V.M_3.csv	0.0524384	99.9476	False
V.M_160.csv	0	100	False
V.M_161.csv	0	100	False
V.M_8.csv	0.0524384	99.9476	False

ows × 4 columns

En lo siguiente se indica las máquinas virtuales que poseen puntos que están muy cercanos a sobrepasar el 25%, lo que nos indica que siempre están alertadas esta información se observa en la Tabla 4.

Tabla 4. Máquinas virtuales alertadas.

	mv	alert_percentage	no_alert_percentage	alerted
90	V.M_90.csv	100	0	True
111	V.M_111.csv	100	0	True
256	V.M_256.csv	100	0	True
180	V.M_180.csv	99.965	0.0349589	True
263	V.M_263.csv	99.965	0.0349589	True
51	V.M_51.csv	99.9592	0.0407854	True
171	V.M_171.csv	99.26	0.739964	True

En este apartado las máquinas virtuales no poseen puntos que sean alertas durante la toma de datos lo cual se describe en la Tabla 5.

Tabla 5. Máquinas virtuales “no” alertadas.

	mv	alert_percentage	no_alert_percentage	alerted
1	V.M_1.csv	0	100	False
155	V.M_155.csv	0	100	False
158	V.M_158.csv	0	100	False
160	V.M_160.csv	0	100	False
161	V.M_161.csv	0	100	False
...
277	V.M_277.csv	4.04358	95.9564	False
165	V.M_165.csv	4.12515	95.8748	False
227	V.M_227.csv	4.41065	95.5893	False
223	V.M_223.csv	4.42813	95.5719	False
73	V.M_73.csv	4.64371	95.3563	False

251 rows × 4 columns



Luego se consultó al cliente sobre si conocía que estas máquinas virtuales consumen gran cantidad de recursos, indicando que sí, porque pertenecen al

backbone principal de la empresa. Al conocer que estas MV siempre estarán alertados para nuestro análisis del algoritmo, se las ignora debido a que molestaría en el estudio de muestral.

Se intentó realizar un entrenamiento de aproximadamente 5 millones de datos recolectados en 12 días, pero el algoritmo LSTM no respondió como se esperaba, por ser demasiado grande la muestra. Por eso el aprendizaje se lo empleo por equipo individual del centro de datos.

3.3. **Apreciación de los resultados**

La pantalla principal de la aplicación móvil ALERT D.C donde se aprecia todos los clústeres disponibles en del centro de datos en esta pantalla se consulta la base de datos los equipos disponibles. Dentro de esta pantalla, a medida que se genere una alerta, se visualizará su respectiva notificación y número de alertas generadas. Para ello se sigue el siguiente patrón de colores:

Color		Representación
Amarillo		Alerta Tipo Memoria
Rojo		Alerta Tipo CPU
Azul		Alerta Tipo Almacenamiento

La imagen muestra cuando se realiza la selección de menú, la aplicación redirige al componente que contiene la de todos los equipos clúster del centro de datos. En la figura 9 se visualiza un menú desplegable y una campana de notificaciones.

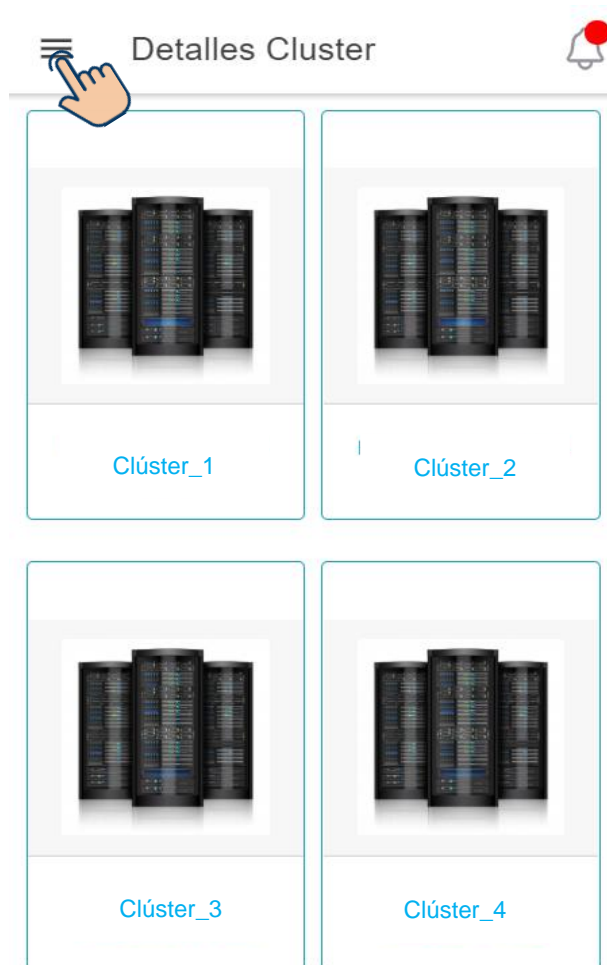


Figura 9 Pantalla clústeres del sistema Vcenter visualizadas en la aplicación móvil ALERT D.C.

En el siguiente apartado se muestra un menú desplegable que contiene las opciones disponibles para navegar en la aplicación ALERT D.C. La opción del menú clúster redireccionará a la lista de nodos clúster del centro de datos y la opción obtener reporte llevará un formulario que realiza el envío del reporte de los recursos disponibles descritos en la figura 10.

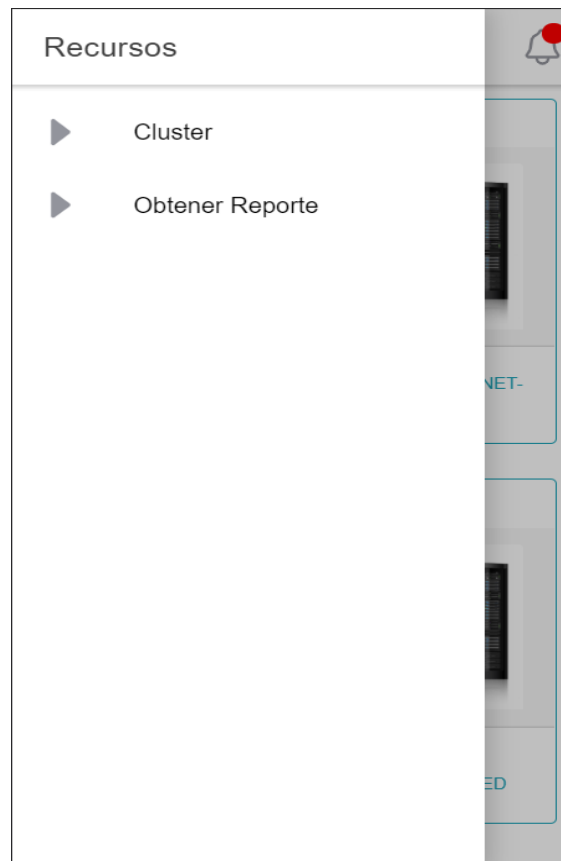


Figura 10 Menú de la aplicación.

En lo descrito a continuación se muestra una de las pantallas que detalla información básica de un clúster seleccionado: Cores, CPU en GHZ entre otra información técnica. De igual manera dentro de esta información encontramos la cantidad de alertas que se han generado debido a la predicción del algoritmo. Además, dentro de esta vista se presenta un menú informativo que por defecto se encuentra seleccionada la información del equipo “detalles”. El siguiente llamado “Lista” muestra todos los servidores host relacionados a un clúster seleccionado (Figura 11).

INFORMACION	
Nombre	Clúster_1
Cpu (GHz)	376
Cores	136
Threads	272
Memoria (GB)	992
Almac. (TB)	1423
Cantidad de host	10
Alertas CPU	0
Alertas Memoria	0

Figura 11 Información básica de Clúster.

Se puede apreciar la lista de host relacionados a un clúster, donde se visualizan los servidores con su respectivo número de alertas generadas y el tipo de alerta que se ha generado por cada servidor. Al seleccionar uno nos redirecciona a la página de información del respectivo host (Figura 12).

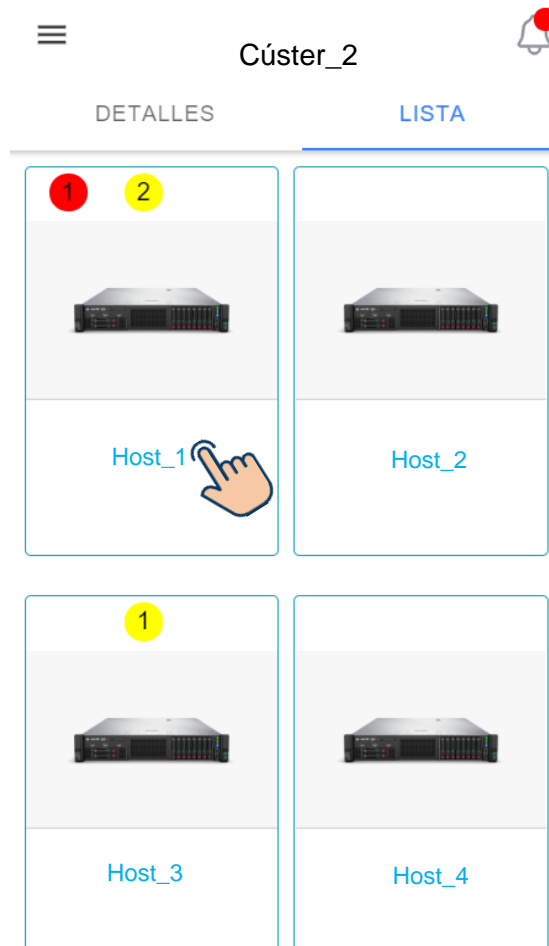


Figura 12 Pantalla de servidores con sus alertas.

Luego se presenta una pantalla de detalles de un host del centro de datos, que muestra la información de la maquina Host seleccionada y sus respectivas alertas en este ejemplo se encuentra 1 alerta CPU y 2 alertas de Memoria (Figura 13).



Host_1



DETALLES

LISTA

Nombre	Host_1 et
Cpu (GHz)	3
Cores	16
Threads	32
Memoria (GB)	128
Almac. (TB)	141
Cantidad de M. ...	13
Alertas CPU	1
Alertas Memoria	2
Alertas Almace...	0

Figura 13 Información de servidores host.

Se observa en la siguiente imagen un menú “Lista” del host el cual se puede observar sus máquinas virtuales asociadas y debido a que estas no han generado alertas estas se encuentran estables. Ahora se selecciona una máquina virtual (Figura 14).

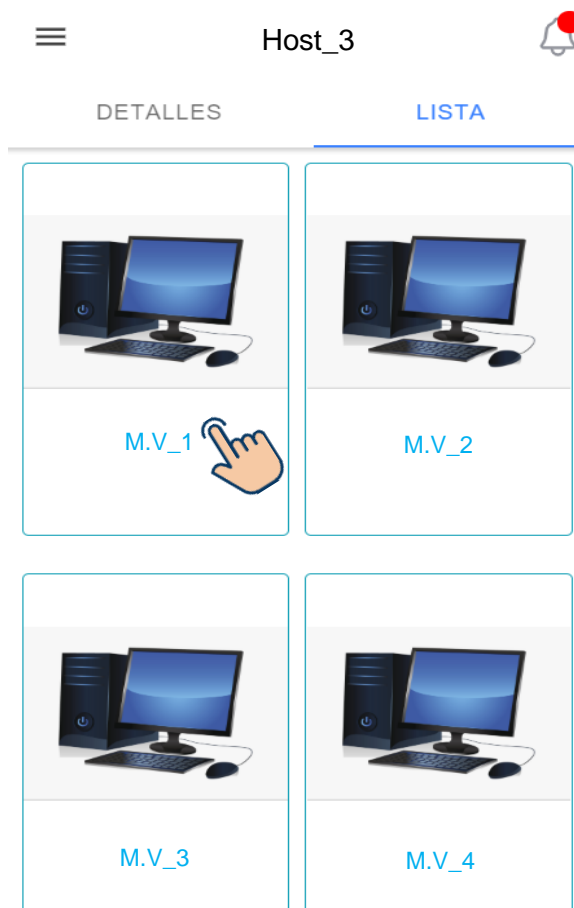
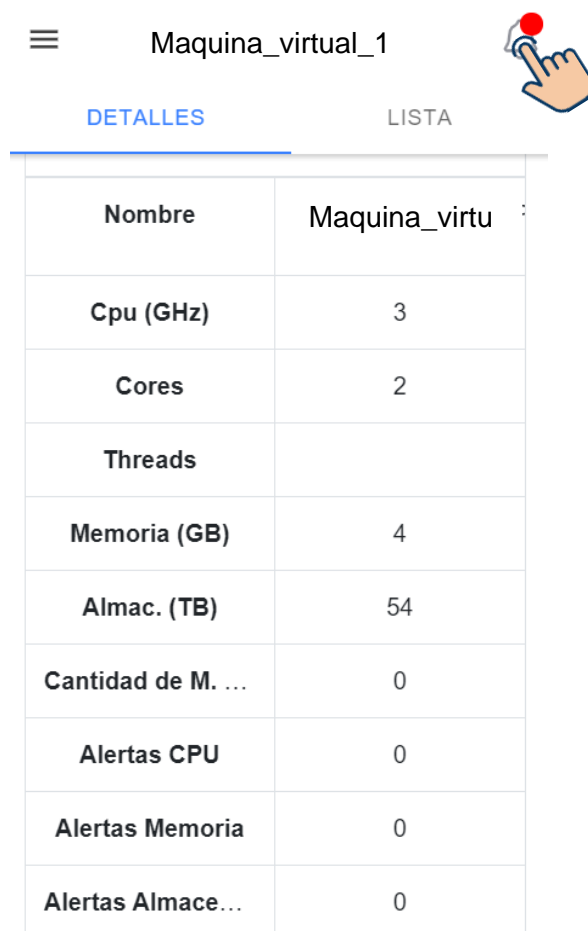


Figura 14 Pantalla de máquinas virtuales de servidores.

Se observa la información de la máquina virtual seleccionada, la cual no ha generado Alertas y por ello sus valores de Alertas de Memoria RAM, CPU y Almacenamiento no contiene alertas generadas (Figura 15).



Maquina_virtual_1

DETALLES LISTA

Nombre	Maquina_virtu
Cpu (GHz)	3
Cores	2
Threads	
Memoria (GB)	4
Almac. (TB)	54
Cantidad de M. ...	0
Alertas CPU	0
Alertas Memoria	0
Alertas Almace...	0

Figura 15 Información de máquinas virtuales.


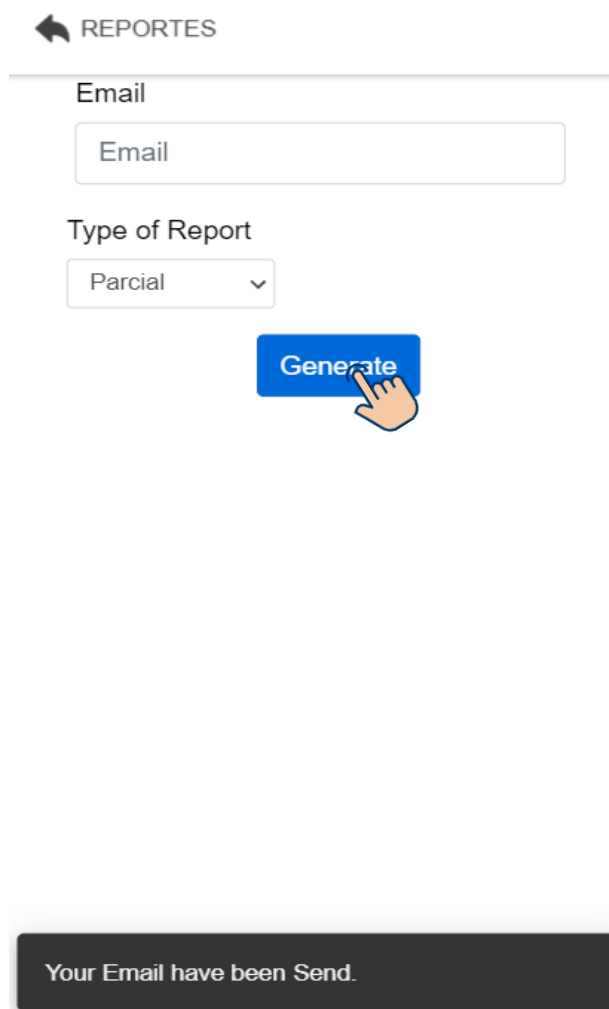
La aplicación posee un ícono de campana con un círculo rojo  que significa que hay alertas que no han sido revisadas. Al dar clic en ese ícono, se despliega la siguiente ventana, la cual contiene todas las alertas generadas en la aplicación ALERT D.C y que no han sido revisadas. Al dar clic en cualquiera de estas alertas, se marca como revisada y se redirige al respectivo recurso que ha generado la alerta. Cuando se ingresa a la campana se observan las alertas generadas. Figura 13.



Figura 16 Notificaciones de nuevas alertas.

En este apartado se observa un formulario para generar un reporte, el cual redirige a esta pantalla la cual se muestra un pequeño formulario donde se ingresa el correo al que se va a enviar la notificación y se elige el tipo de reporte: Parcial y Consolidado (Figura 17).



← REPORTES

Email

Type of Report

Parcial ▾

Generate

Your Email have been Send.

Figura 17 Envío de reporte por email.

La siguiente imagen se observa una ventana de email con el respectivo reporte de los recursos disponibles en la aplicación móvil ALERT D.C. Figura 15

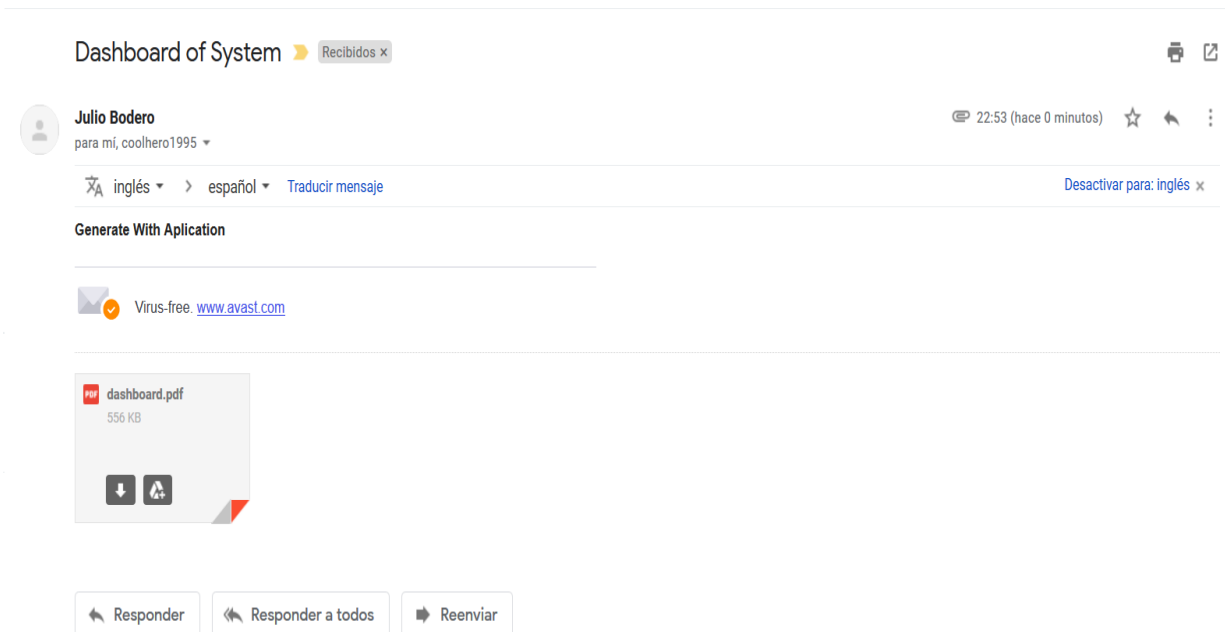


Figura 18 Correo electrónico recibido.

El reporte que se genera a través de la aplicación móvil ALERT D.C contiene información parcial de todos los hosts del cliente registrados en el centro de dato alojado en Vcenter con información relevante como el sistema operativo instalado y sus recursos usados, disponibles y totales. De igual forma, se observan gráficos que ayudan a entender mejor la información para el análisis de la compañía (Figuras 19 y 20).

Reporte HA3-EN6-TELEFONIA-GYED

Enclosure	Generacion	Servidor	Cpu Modelo	Cpu Nucleos	Cpu Threads	CPU Disponible (nucleos)	CPU Total (nucleos)	CPU Usado (nucleos)	RAM Disponible (GB)	RAM Total (GB)	RAM Usada (GB)	Sistema Operativo
HA3-EN6-TELEFONIA-GYED	HP ProLiant BL460c G6	Host_1	Intel(R) Xeon(R) CPU E5540 @ 2.53GHz	8	16	123	128	5	56	64	8	VMware ESXi 5.5.0 build-8934887
HA3-EN6-TELEFONIA-GYED	HP ProLiant BL460c G7	Host_2	Intel(R) Xeon(R) CPU E5630 @ 2.53GHz	8	16	126	128	2	60	64	4	VMware ESXi 5.5.0 build-8934887

Total Cpu: 8448
Total Disponible Cpu: 7138
Usado Cpu: 1310
Total RAM (GB): 2192
Total Disponible RAM (GB): 680
Usado RAM (GB): 1512

Figura 19 Reporte parcial de todos los hosts.

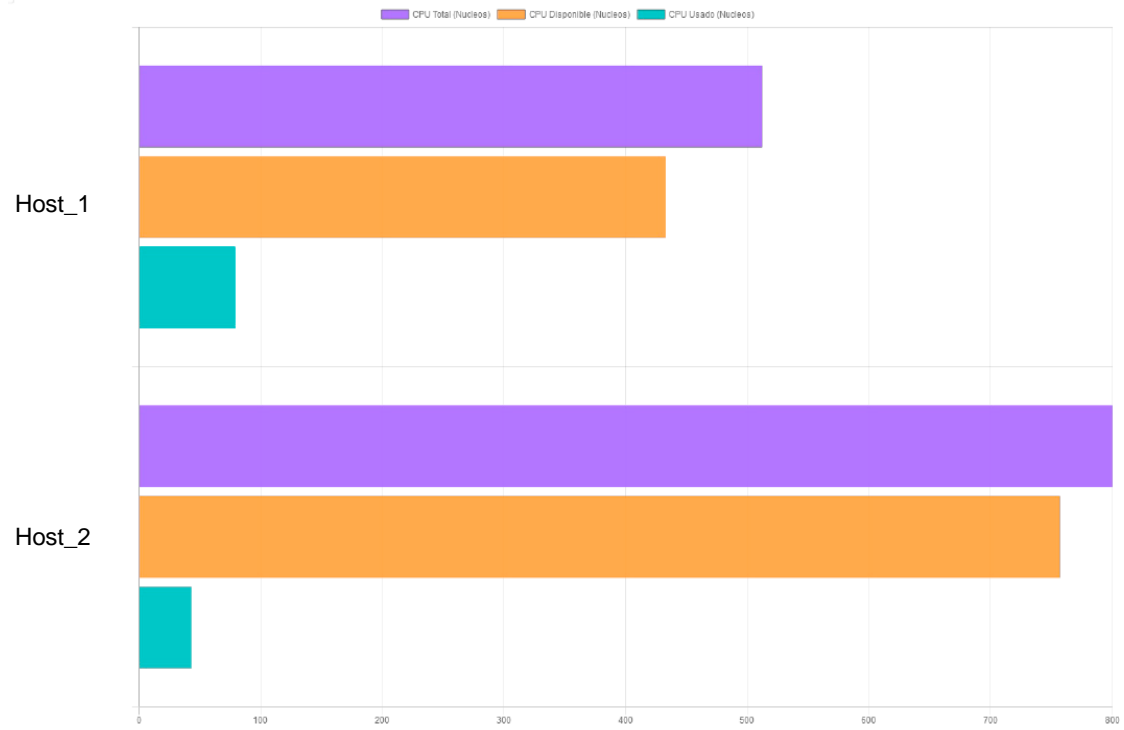


Figura 20 Graficas del reporte parcial de todos los hosts.

En la siguiente Figura se aprecia un documento consolidado el cual contiene información de recursos totales por cada Clúster, se observa información de CPU disponible, CPU usado, RAM disponible, RAM usada, además de un gráfico que ayuda a un mejor análisis de los recursos disponibles. Figura 21.

REPORTE CONSOLIDADO

Enclosure	CPU Disponible (nucleos)	CPU Usado (nucleos)	RAM Disponible (GB)	RAM Usada Total (GB)
HA1-EN7-TELCONET-GYED	3260	580	296	696
HA2-EN8-TELCONET-GYED	6889	1303	564	1499
HA3-EN6-TELEFONIA-GYED	7138	1310	680	1512
HA5-EN2FI1-TELCONET-GYED	8675	1821	960	2255

Total Cpu: 30976
Disponible Cpu: 25962
Usado Cpu: 5014
Total RAM(GB): 8462
RAM Disponible (GB): 2500
Usado RAM (GB): 5962

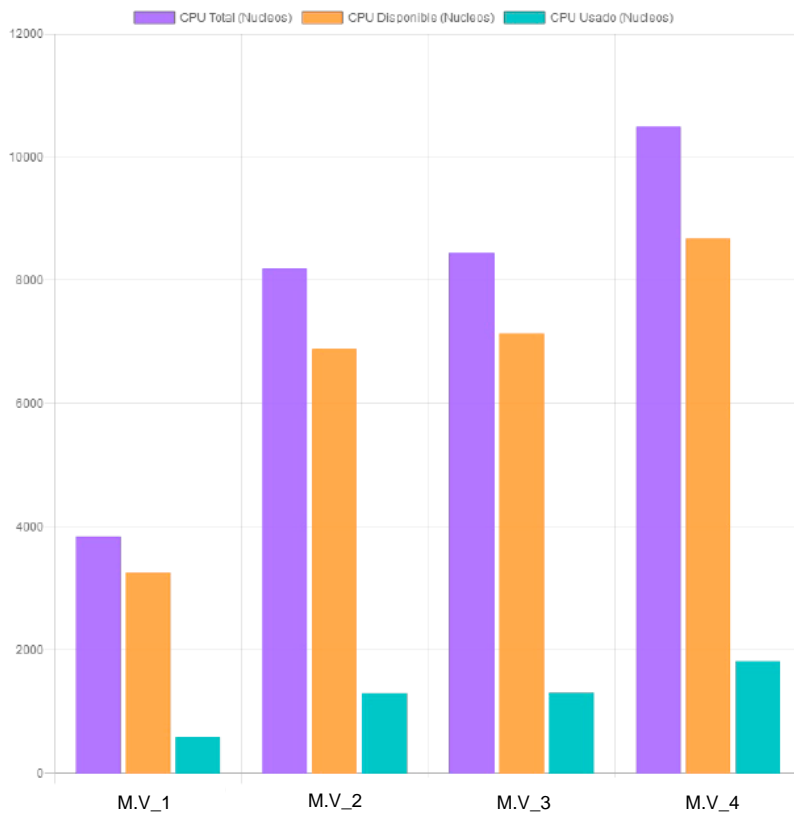


Figura 21 Reporte consolidado.

Reporte consolidado y el cual indica los totales de RAM disponibles centro de datos (Figura 22).

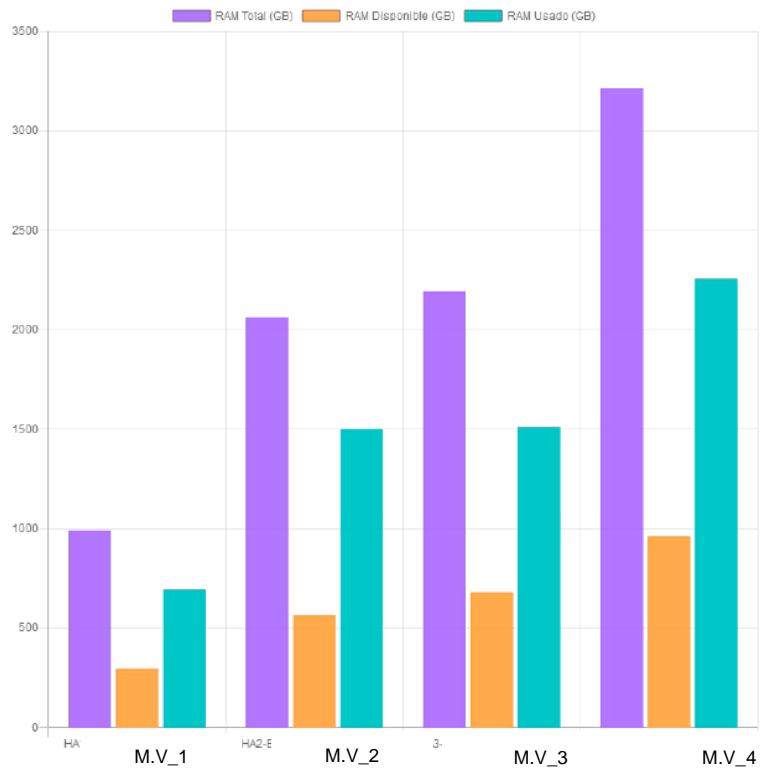


Figura 22 Reporte consolidado gráficas de RAM

Totales en porcentaje de CPU, RAM Disponible y Usado por el centro de datos. La Figura 21 muestra los porcentajes de servidores encendidos y apagados. Para este caso todos los servidores se encuentran encendidos (Ver Figuras 23 y 24).

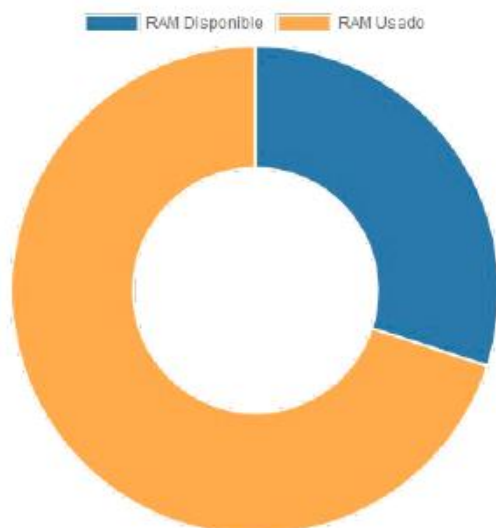
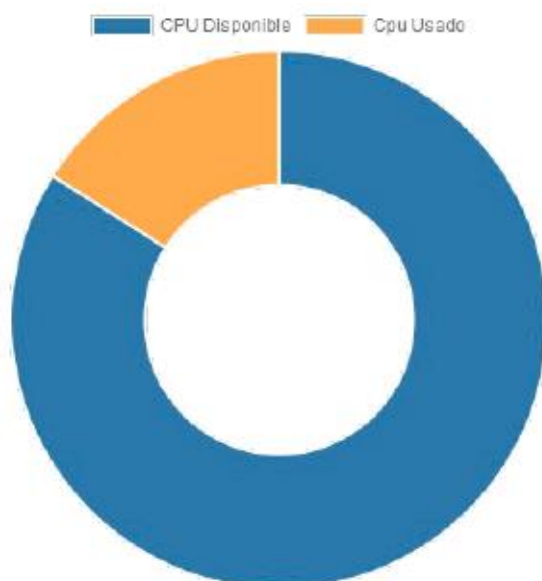


Figura 23 Reporte consolidado gráficas de RAM Disponible y CPU disponible

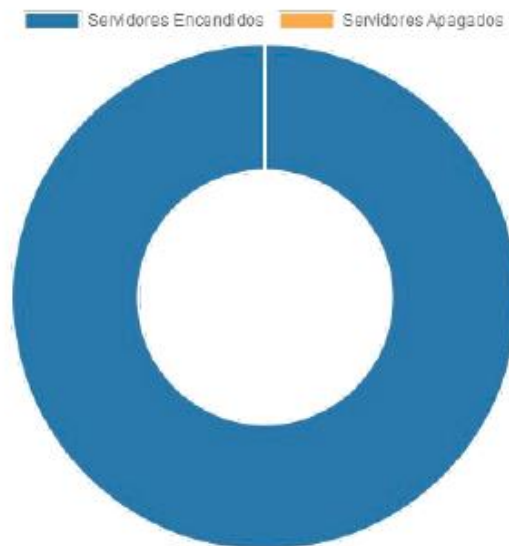


Figura 24 Reporte consolidado servidores encendidos.

Como se puede observar la cantidad de proceso y mejorar que incorpora ALERT D.C al centro de datos es muy útil en muchos aspectos, estos pueden ser: en la claridad de presentar los recursos computacionales de los equipos, la facilidad de poder generar reportes de los recursos del centro de datos, además, de poder realizar predicciones de los recursos computacionales, dependiendo del tipo de equipo y estos pueden ser: Clúster, Host y Máquina Virtual.

El desarrollo de la aplicación de monitoreo móvil ALERT D.C permite predecir cuándo un equipo del centro de datos sobrepase el 25% de los recursos es indispensable, puesto que caso contrario la empresa podría llegar a perder potenciales cliente por desconocimiento de los recursos computacionales en el centro de dato.

3.4. Análisis de costos del sistema

Esta sección posee dos instancias muy importantes para el desarrollo o implementación de la aplicación de monitoreo predictivo de los recursos computacionales del centro de datos y son costo de Software y Hardware:

3.4.1. Costo del software

Tomando en cuenta la información suministrada por el cliente, que el centro de datos posee seis administradores (ingenieros) en un año

laboral. El gasto sería de \$1,500 por mes a cada ingeniero. El gasto anual generado por estos operadores sería de \$108,000. (Escenario #1)

Tabla 6. Gasto generador por parte de la empresa sin ALERT D.C.

Tipo de empleado	Cantidad de operadores	Gasto mensual	Gasto anual
Ingeniero	6	\$9,000	\$108,000

Para la implementación del proyecto se necesitarían dos programadores full stack para el desarrollo de toda la lógica de programación y del aplicativo móvil ALERT D.C, por un aproximado de 6 meses para luego dejar a un ingeniero a cargo del monitoreo, trabajando por el resto del año. Los gastos serian de \$2,000 por mes a cada programador full stack y \$1,500 por mes a cada ingeniero. El gasto generado de forma anual sería de \$33,000. (Escenario # 2)

Tabla 7. Gasto generado por parte de la empresa con ALERT D.C.

Tipo de empleado	Cantidad de operadores	Gasto mensual	Gasto anual
Programador	2	\$4,000	\$24,000
Ingeniero	1	\$1,500	\$9,000

Parar lograr una mejor interpretación de los datos se procederá a utilizar el acrónimo de OPEX “Gastos Operativos”, el cual refleja el costo relacionado de las operaciones y servicios realizados por parte de la empresa, esto incluye el sueldo de los trabajadores de la empresa. La forma de calcular el OPEX es solo sumar los costos operativos de la empresa (M.D. Ananth ; Rinki Sharma, 2016)

Para el escenario # 1, el valor de $OPEX_1 = \$108,000$ y el valor de $OPEX_2 = \$33,000$. Ante ello surge la duda del por qué son diferentes. Una respuesta sería es porque al utilizar ALERT D.C se reduce la cantidad de tareas que realizaban los operadores (ingeniero). Por tanto, se reduce el flujo de gastos de la empresa y dichas tareas pendientes se encargaría de realizar la aplicación móvil ALERT D.C.

Además, para el escenario # 2 se compró un Compute Engine en el cual va a estar alojado una API de reportes y que a su vez se encarga de enviar los reportes de los recursos computacionales del aplicativo móvil. Dentro de las especificaciones del equipo en Compute Engine, se escogió el de segundo estándar que ofrece Google, porque ofrece dos máquinas virtuales y dentro de este servidor estará alojado la base de datos en MongoDB. El gasto de este componente y sus características se detalla a continuación.

Tabla 8. Gasto del servicio en la nube.

Tipo de Máquina	CPU Virtuales	Memoria	Precio al mes (USD)	Precio de las interrupciones (USD)	Total anual
N1-standard-2	2	7.5GB	\$48.55	\$14.60	\$757.8

Tabla 9. Costo Totales en Software.

Tipo de Escenario	Costo Total
Escenario # 1	\$108,000
Escenario # 2	\$33,757.8

Se utilizó el porcentaje de comparación para ver qué tan conveniente le resulta a la empresa implementar el escenario #2 respecto al escenario # 1. El cálculo es el siguiente:

$$\begin{aligned}
 & \text{(Porcentaje de comparación)} \quad \frac{\text{Diferencia}}{\text{Referencia}_{Esc1}} \times 100 \\
 & \frac{108,000 - 33,757.8}{108,000} \times 100 = 68.74 \%
 \end{aligned}$$

Se observa que, a pesar de añadir más costo “Compute Engine” a la empresa en el escenario # 2, es mucho más conveniente en términos de costo para la empresa en un 68.74% adquirir ALERT D.C del escenario

2. Puesto que se necesita mayor potencia de cómputo para ejecutar el algoritmo de redes neuronales y por ende realizar el proceso de predicción de los recursos computacionales, se procedió a implementar hardware adicional en el escenario # 2.

3.4.2. Costo del Hardware

Para el Escenario # 2 el nuevo hardware se encontrará alojados en el centro de datos del cliente, y las características del equipo serían las siguientes:

- ✓ Monitor Acer IPS Full-HD de 21.5", de 75 Hz.
- ✓ Teclado DELL modelo KB216.
- ✓ HPE ProLiant ML30 Gen10 Tower Server, Intel Xeon E-2124 Quad-Core 3.3GHz 8MB, 32GB DDR4 RAM, 8TB Storage.

Dentro de este equipo se alojará el API de predicciones, el cual se encargará del aprendizaje, generar la predicción y enviar las respectivas alertas de los equipos que sobrepasen el 25% de los recursos computacionales, comprar este hardware tiene un costo descrito en la siguiente tabla.

Tabla 10. Costo del nuevo equipo para la predicción.

Componente	Costo de la compra (sin envió)
HPE ProLiant ML30	\$1,799.00
Teclado DELL	\$13.51
Monitor Acer	\$89.99
Total	\$1,902.50

Para tener una mayor interpretación de los datos en el lado del software se utilizó el acrónimo llamado OPEX, que son los gastos operativos y servicios que realiza la empresa en un año. Este acrónimo es acompañado de los gastos e inversiones capitales por parte de la empresa esto se denomina CAPEX. En otras palabras, son los bienes comprados en dos o más años. Con el fin de mejorar los activos fijos de la empresa e indicar si la empresa está invirtiendo para crecer o mantenerse, donde $\text{crecer CAPEX} > \text{depreciaciones}$ y mantenerse

CAPEX= depreciaciones, el CAPEX solo se lo calcula en los activos de la empresa.(M.D. Ananth ; Rinki Sharma, 2016)

Hay que considerar de que estos acrónimos se proyectan en el sector de tecnología de la información (TI) y ayudan al análisis del retorno de los bienes de una empresa, sean estos de software y hardware. La era tecnológica que estamos viviendo hace acortar los tiempos de utilidad de los equipos de hardware. Por tanto, el acrónimo OPEX es el que mejor se adapta al departamento TI de una empresa. “CAPEX = (activos año x -activos año $x-1$) + depreciación x ”, los equipos de la empresa se degradan a 36 meses lo que equivale a “3 años”, el centro de datos tiene el objetivo depreciar los equipos en 18 meses para proyectar mejores ganancias.

Tabla 11. Gasto en hardware del centro de datos.

Tipo de escenario	Activos en 2019	Activos en 2020	Pasivos en 2019	Pasivos en 2020
Escenario # 1	\$150mil	\$160mil	\$75mil	\$85mil
Escenario # 2	\$150mil	\$165mil	\$75mil	\$86.9mil

Cabe recalcar que para el escenario # 2, el valor de los pasivos en el año 2018 es de \$86.9mil por la comprar del equipo para el proceso de predicción y esto trae ingresos a los activos de la empresa y se asumió un valor de \$3mil. Una vez obtenido los activos y pasivos fijos de la empresa, se procedió a calcular la depreciación de los activos en los años 2019 y 2020. Para observar si los activos de la empresa incrementan o se mantienen para el año 2020, el cálculo de depreciación o amortización es:

$$(Depreciación) \mathcal{D} = \frac{\text{Costo_actual} - \text{Valor_residual}}{\text{Vida_util}}$$

La siguiente tabla muestra la depreciación de los activos en los dos años, la depreciación fue calculada en años y la vida útil de los equipos en años. Además del valor residual del 5% respecto al costo actual.

Tabla 12. Depreciación de los activos.

Tipo de escenario	Depreciación 2019	Depreciación 2020
Escenario # 1	\$47,500	50,666.66
Escenario # 2	\$47,500	52,250

El cálculo del CAPEX para el escenario # 1 es:

$$CAPEX_1 = 160mil - 150mil + 50,666mil = 60,666mil$$

$$CAPEX_1 > D_{2019} = 60,666mil > 50,666mil$$

Como se puede observar en el escenario # 1 la empresa se encuentra empleando una política de expansión. Veamos el escenario # 2, el cálculo del CAPEX para el escenario # 2 es:

$$CAPEX_2 = 165mil - 150mil + 52,250mil = 67,25mil$$

$$CAPEX_2 > D_{2019} = 67,25mil > 52,250mil$$

Tabla 13. Valores de CAPEX y depreciaciones en el año 2020

Tipo de escenario	Valor del CAPEX 2020	Depreciación 2020
Escenario # 1	60,666mil	50,666mil
Escenario # 2	67,25mil	52,250mil

Se puede observar que el CAPEX₂, también se encuentra utilizando una política de expansión, pero es mucho mejor CAPEX₁. Esto indica que CAPEX₂ aumenta la base de los activos de la empresa mientras que CAPEX₁ aumenta la base de los activos de empresa, pero no tanto como CAPEX₂. Ante ello surge la duda de que tanto le favorece a la empresa. Por ello, se calcula el escenario más favorable por medio del porcentaje de comparación, tomando como referencia el CAPEX₂.

$$\frac{67,25 - 60,666}{67,25} \times 100 = 9.79 \%$$

El porcentaje de 9.79% indica que mejora el flujo de activos de la empresa en casi un 10%, tomando en cuenta el gasto que generará a la empresa en el año de su compra.

3.4.3. Costo del desarrollo de ALERT D.C

Ya expuesto los beneficios que trae ALERT D.C a la empresa se dará a conocer cuanto sería el costo total en la realización del aplicativo móvil de ALERT D.C, tendría el siguiente costo total descrito en la tabla 14.

Tabla 14.Costo del aplicativo móvil ALERT D.C.

Actividad	Costo de la actividad
Dos programadores full stack por 6 meses	\$24,000
Servidor de reportes y alertas	\$757.8
Hardware, servidor para la predicción	\$ 1,902.5
Total	\$26,660.3

La principal competencia con ALERT D.C es CenturyLink, el cual ofrece un espacio en su centro de datos y su costo de adquirir este espacio en su centro de datos es aproximadamente de \$165,24(CenturyLink, 2018). Es escalable, tiene seguridad, alto desempeño y disponibilidad en 18 centro de datos a nivel de América latina. Nuestro proyecto tiene sus diferencias a CenturyLink, dispone de la capacidad de realizar predicciones de los recursos computacionales, está desarrollada en ambiente móvil multiplataforma y tiene un esquema de clasificación del tipo de alerta.

3.5. Análisis de los resultados

Se logró aprovechar el programa administrador del centro de datos Vcenter, accediendo mediante Pymomi a las variables de administración que monitorean y controlan los recursos de la red del centro de datos y a su vez lo que realiza cada máquina asociado al gestor de administración. Mediante código utilizado en Python se capturó el contenido de las variables almacenando en una estructura de datos el valor monitoreado por el programa Hipervisor de los host y máquinas virtuales, con el fin de generar reportes análisis y aprendizaje de los datos almacenados.

Se desarrolló una aplicación móvil denominada ALERT D.C, la cual posee un proceso de generación de alertas adecuado para la medición de los recursos computacionales de los equipos del centro de datos, pudiendo fácilmente observar los diferentes tipos de alertas que se encuentra en un equipo, además de poder generar los reportes completos en base a las necesidades del centro de datos.

Una de las principales ventajas que trae al centro de datos, es el tiempo de generación de un reporte consolidado, a ALERT D.C le toma generar un reporte 6 segundos mientras que al método manual le lleva entre 1 o 2 días en generar dicho reporte, lo que acarrearía en problema a la empresa, debido a que se desconoce los recursos que disponibles al instante además de ser una pérdida indirecta de futuros clientes a la empresa.

Se procedió a realizar un estudio de cómo se encuentra distribuidos los conjuntos de datos, y de esta forma poder obtener un mejor aprendizaje del algoritmo LSTM, viendo que los datos siguen una distribución uniforme o no uniforme, se pudo observar que ciertos equipos siempre se encuentran alertados y se decidió despreciarlos por ser parte del Backbone principal de la empresa, este afectaría directamente para el análisis del algoritmo por ser equipos siempre alertados “datos aberrantes”. además, de los ajustes a los hiper-parametros del algoritmo y métricas para ver qué tan preciso es el algoritmo.

Respecto al análisis económico mediante el acrónimo OPEX se demostró que se puede realizar una reducción de un 68.74% a los gastos operativos de la empresa, los cuales anteriormente representaban un costo de \$108,000 mil y con ALERT D.C representa \$33,757.8 mil, debido que el centro de datos no contempla la automatización de alarmas. Por otro lado, el acrónimo CAPEX indicó un mayor crecimiento en los flujos activos de la empresa mejorándolos en 9.79 %, respecto al escenario que la empresa manejaba con regularidad.

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones

Con la finalidad de automatizar el proceso de creación y envío de reportes del centro de datos, y alertando cuando un equipo sobrecarga sus recursos computacionales. Se implementó una aplicación móvil “ALERT D.C”, la cual atravesó diferentes etapas en el desarrollo como el proceso metodológico, estudios de la muestra, entre otros. En base al análisis de los resultados observados se concluye que:

- Para el proceso de creación y envío de reportes ALERT D.C, logró reducir el tiempo generación de un reporte que tomaba de 1 a 2 días, a solo 6 segundos, reduciendo los tiempos de decisión al momento de adquirir un nuevo equipo, además de conocer las capacidades actuales de cada uno de los equipos en tiempo real. De forma adicional se comprobó que el consumo de los recursos en los equipos se comporta de una manera no uniforme, dando la apertura a la empresa a realizar un balance del consumo en el centro de datos.
- El modelo LSTM es muy preciso al realizar predicciones cuándo un equipo sobrecargó el umbral del 25% en sus recursos de cómputo, dando la apertura a ALERT D.C de emitir un tipo de alerta dependiendo del tipo de la anomalía. Esta anomalía puede ser sobrecarga en los recursos de CPU, RAM o storage. Sin embargo, el algoritmo LSMT presentó un problema, al realizar la experimentación e investigación se comprobó que trabaja únicamente con conjuntos de datos no tan grandes, dificultando el aprendizaje de grandes conjuntos de datos. Ante ello se realizó un entrenamiento por equipo del centro de datos, por agotamiento de los recursos para su procesamiento. Esperando que esto incentive a una futura implementación de un clúster de alto rendimiento utilizando Tensor-Flow con Apache-Spark, para poder procesar todos los equipos del centro de datos en un único entrenamiento para obtener un modelo generalizado para todo el centro de datos además de mejorar con Apache-Spark el tiempo de aprendizaje del modelo.

- Al realizar la toma de datos se observó que las variables de monitoreo del sistema Hipervisor los valores cambiaban cada 1 min, por lo que no se pudo obtener un monitoreo o reportes en tiempo real. Esto se convirtió en una desventaja, por lo que para la captura de datos se realizó un muestreo cada 1.02 minutos para garantizar que el dato no sea el anterior sino uno nuevo capturado.
- Finalmente, ALERT D.C demostró ser de gran ayuda en el ámbito económico, reduciendo en un 68.74% los gastos operativos y mejorando los activos fijos en un 9.79%, esto último le ayuda en mucho a la empresa debido a que se reduce los gastos operativos y aumenta sus activos fijos. Por tanto, producirá mayores ingresos a la empresa y posibles inversiones a futuro.

4.2. Recomendaciones

A lo largo de este experimento se visualizó ciertas limitantes, lo cual nos permite realizar las siguientes recomendaciones:

- Respecto a trabajos futuros se puede implementar lo antes mencionado con Tensor-Flow y Apache-Spark para poder entrenar todo el conjunto de datos del centro de datos obteniendo un modelo entrenado en menor tiempo y con grandes cantidades de datos para que de esta forma el algoritmo contenga en su memoria todos los diferentes escenarios que puedan ocurrir en cualquier equipo y en cualquier momento.
- Otra implementación para ALERT D.C puede basarse en Redis la cual tiene el potencial de reducir los tiempos de consultas a la base datos en MongoDB a manera de Caché. Adicionalmente, el utilizar y experimentar con otros algoritmos de predicción con machine learning, el cual puede llevar a mejores tiempos de respuesta con menor cantidad de procesamiento y de esta forma obtener resultados en menores tiempos y con mejores resultados de precisión en las métricas utilizadas para la evaluación del modelo. Más aun, no se recomienda utilizar modelos ARMA pues los datos de CPU son no estacionarios, es decir, que varían mucho en el tiempo es decir que no tienen periodos definidos.

- Para futuras investigaciones se puede implementar dentro de la aplicación móvil la parte de gráficas de los diferentes equipos del centro de datos y que sean relevantes para el administrador, además de no solo emitir alertas, sino registrarlas y notificarlas por otros medios como redes sociales o mensajes de texto. Estas alertas pueden ser del tipo equipo apagado o del tipo seguridad, esto es cuando un equipo no responda en la red o que se encuentre bajo un ataque de hackers.

BIBLIOGRAFÍA

- Aymen Abdullah Alsaffar ; Eui-Nam Huh. (2015). Multimedia delivery mechanism framework for smart devices based on mega data center and micro data center in PMIPv6 environment. *015 International Conference on Information Networking (ICOIN)*.
<https://doi.org/https://ieeexplore.ieee.org/document/7057915>
- B, D. D., Salim, S., & Vargese, S. M. (2016). MongoDB Vs MySQL: A Comparative Study of Performance in Super Market Management System. *International Journal of Computational Science and Information Technology*, 4(2), 31–38.
<https://doi.org/10.5121/ijcsity.2016.4204>
- Canal Synnex Comstor. (2016). *Blog Mexico*.
<https://blogmexico.comstor.com/nuevas-tecnologas-y-estrategias-para-optimizar-la-eficiencia-del-data-center>
- CenturyLink. (2018). *Tarifas de los Servicios de Telefonía Fija CENTURY LINK ECUADOR DIC 4 2018*.
- Duggan, M., Mason, K., Duggan, J., Howley, E., & Barrett, E. (2018). Predicting host CPU utilization in cloud computing using recurrent neural networks. *2017 12th International Conference for Internet Technology and Secured Transactions, ICITST 2017*, 67–72.
<https://doi.org/10.23919/ICITST.2017.8356348>
- Falcón, J. C., & Herrera, R. (2005). *Análisis Del Dato Estadístico Guía Didáctica*.
- Google Cloud. (2018). *NSL Google Cloud*. <https://cloud.google.com/network-tiers/docs/overview>
- Google Cloud. (2020). *SLA Google Cloud*. <https://cloud.google.com/compute/sla>
- Google Colab. (2015). *Understanding LSTM Networks*. 27 August.
<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Hernández-Sampieri, R.; Fernández Collado, C. y Baptista Lucio, M. de. P. (2014). Metodología De La Investigación. In M. G. H. Education (Ed.), *México*.
- Jianhua Peng ,Hui Zhou, Q. M. yJingli Y. (2020). Big data security access control algorithm based on memory index acceleration in WSNs. *7 de Mayo*.
<https://jwcn-urasipjournals.springeropen.com/articles/10.1186/s13638-020-01725-1>
- Lores Serrano. (2019). El objetivo de Microsoft es democratizar el software de

virtualización. *04 Diciembre 2009.*

<https://www.computing.es/infraestructuras/entrevistas/1031889001801/objetivo-microsoft-democratizar.1.html>

M.D. Ananth ; Rinki Sharma. (2016). Cloud Management Using Network Function Virtualization to Reduce CAPEX and OPEX. *2016 8th International Conference on Computational Intelligence and Communication Networks (CICN).*

Proaño, J. S., Ricaurte, D., & Sáenz Página, J. (2012). *Estudio de Mercado Servicio Desarrollo de Software en Ecuador 2012 Documento elaborado por la Oficina Comercial de ProChile en Ecuador en colaboración con la Universidad Casa Grande, Guayaquil (alumnos: Carmen.*

Roque Gabriel y Orellana Allan. (2019). *Desarrollo de un sistema de monitoreo para predicción y alertas de exceso de utilización de las capacidades computacionales de los clústeres en un centro de datos de la ciudad de Guayaquil.* ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

Shuguang Qi ; Yu Zhang ; Mengdi Wang. (2019). Study and Application on Data Center Infrastructure Management System Based on Artificial Intelligence (AI) and Big Data Technology. *2019 IEEE 4th International Future Energy Electronics Conference (IFEEEC).*

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9014987>

VMware-vSphere 6.5. (2017). *Supervisión y rendimiento de vSphere - VMware vSphere 6.5.* <https://docs.vmware.com/es/VMware-vSphere/6.5/vsphere-esxi-vcenter-server-65-monitoring-performance-guide.pdf>

Vmware. (2019). *Vmware vSphere.* 7 de Agosto de 2019.

<https://docs.vmware.com/es/VMware->

[vSphere/7.0/com.vmware.vsphere.vcenterhost.doc/GUID-4D4B3DF2-D033-4782-A030-3C3600DE5A7F.html](https://docs.vmware.com/es/VMware-vSphere/7.0/com.vmware.vsphere.vcenterhost.doc/GUID-4D4B3DF2-D033-4782-A030-3C3600DE5A7F.html)

Weichao Ding ; Fei Luo ; Chunhua Gu ; Haifeng Lu ; Qin Zhou. (2020). Performance-to-Power Ratio Aware Resource Consolidation Framework Based on Reinforcement Learning in Cloud Data Centers. *In IEEE Access.*

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8959131>