



FaceFinder

MÓDULO DE BÚSQUEDA DE PERSONAS DENTRO DE UNA BASE DE DATOS DE ROSTROS



Introducción

- Los algoritmos utilizados para el procesamiento de imágenes son de complejidad computacional alta.
- Por esto decidimos implementar un módulo que aprovechara los beneficios del procesamiento en paralelo.

Objetivos

- Demostrar que Hadoop es una plataforma eficiente para el procesamiento masivo de imágenes.
- Comprobar que implementar una solución de procesamiento masivo de imágenes resulta sencillo y muy poco costoso.

Marco Teórico:

Detección Facial
Reconocimiento Facial

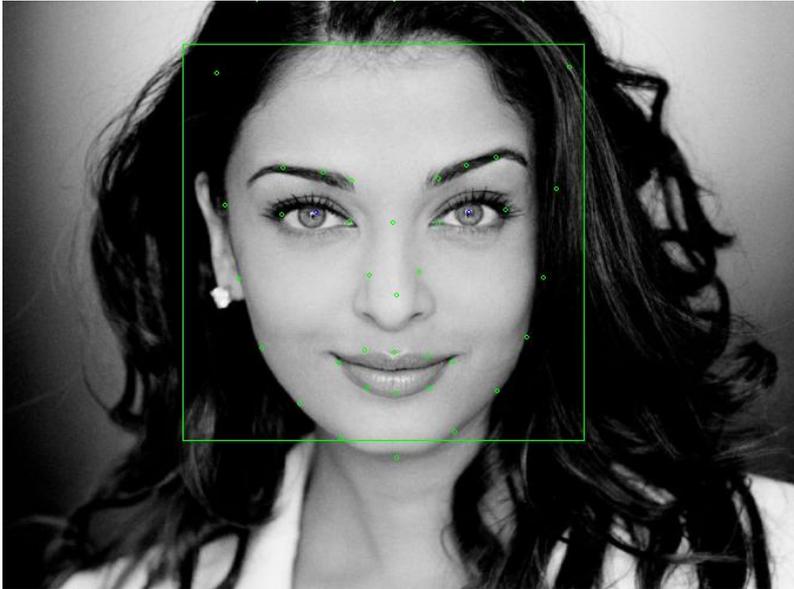


Detección Facial

- La detección facial es un método cuyo objetivo es determinar si en una imagen existen o no rostros.
- En general pueden ser agrupadas como tareas de clasificación de patrones.



Reconocimiento Facial



- Consiste en identificar o verificar una persona a partir de una imagen o un cuadro de video.
- Reconocimiento por geometría facial.
- Incorporando todos estos datos numéricos obtenidos se crea una plantilla única.

Plataforma Procesamiento Distribuido



Hadoop
MapReduce

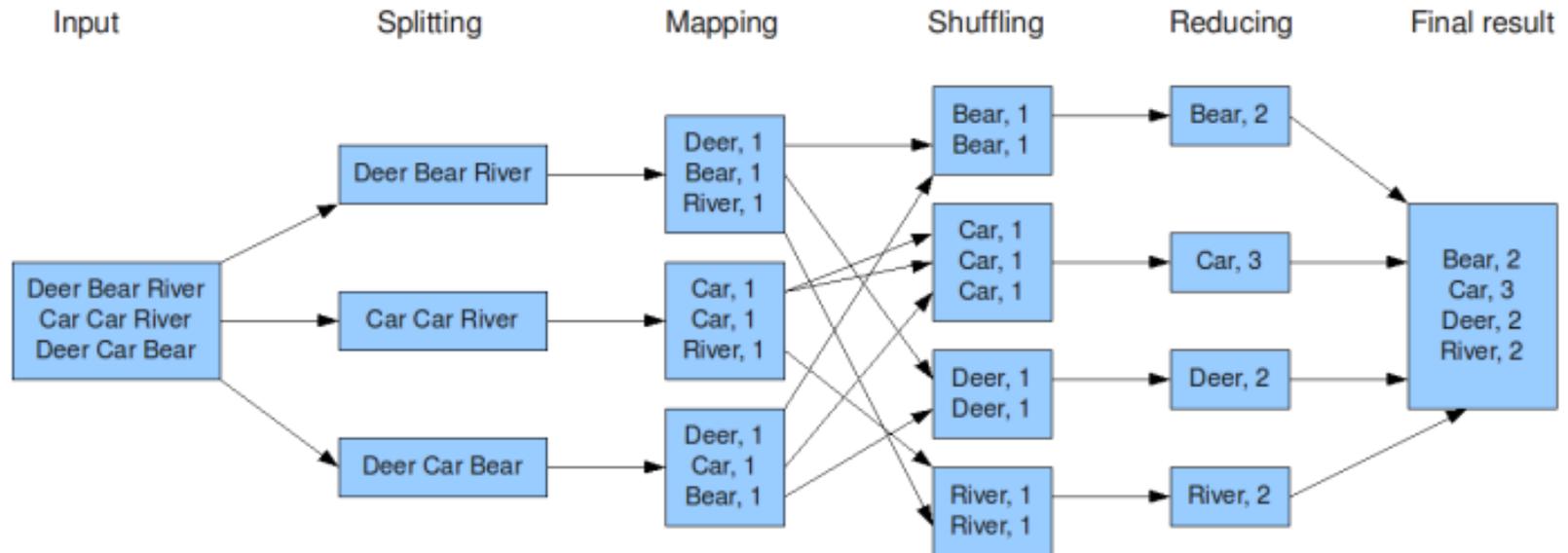
Hadoop

- Plataforma de procesamiento distribuido.
 - Proyecto de *Apache Software Foundation*.
 - Implementado en JAVA.
 - Hadoop Distributed File System: HDFS

- MapReduce: Modelo para Programación Distribuida.
 - Fase *Map*: Toma la entrada y lo divide en sub problemas
 - Fase *Reduce*: Agrupa los resultados. (clave/valor)

MapReduce

The overall MapReduce word count process



Sistema de Archivos Distribuido: Hadoop DFS

- HDFS es altamente tolerante a fallos
- Está diseñado para ser implementado en hardware de bajo costo.
- Proporciona acceso de alto rendimiento de datos de aplicación.
- Los tamaños de bloques de archivos usado por HDFS es de 64 MB.
- Compresión gzip, lzo.

Hadoop Pipes

- Librería que permite usar Hadoop DFS y poder escribir el *Mapper* y el *Reducer* en C++
- Hadoop Pipes convierte los datos a bytes y los envía vía socket.
- El trabajo puede consistir en la combinación entre C++ y Java

Amazon Web Services: AWS

➤ **Amazon Elastic Compute Cloud (EC2)**

- Se puede obtener poder de cómputo y procesamiento de manera variable.
- Flexibilidad para elegir cualquier plataforma de desarrollo o modelo de programación.

➤ **Amazon Elastic MapReduce**

- Hadoop está instalado y preconfigurado en la infraestructura de amazon.
- Interfaz web para manejar instancias.

➤ **Beneficios**

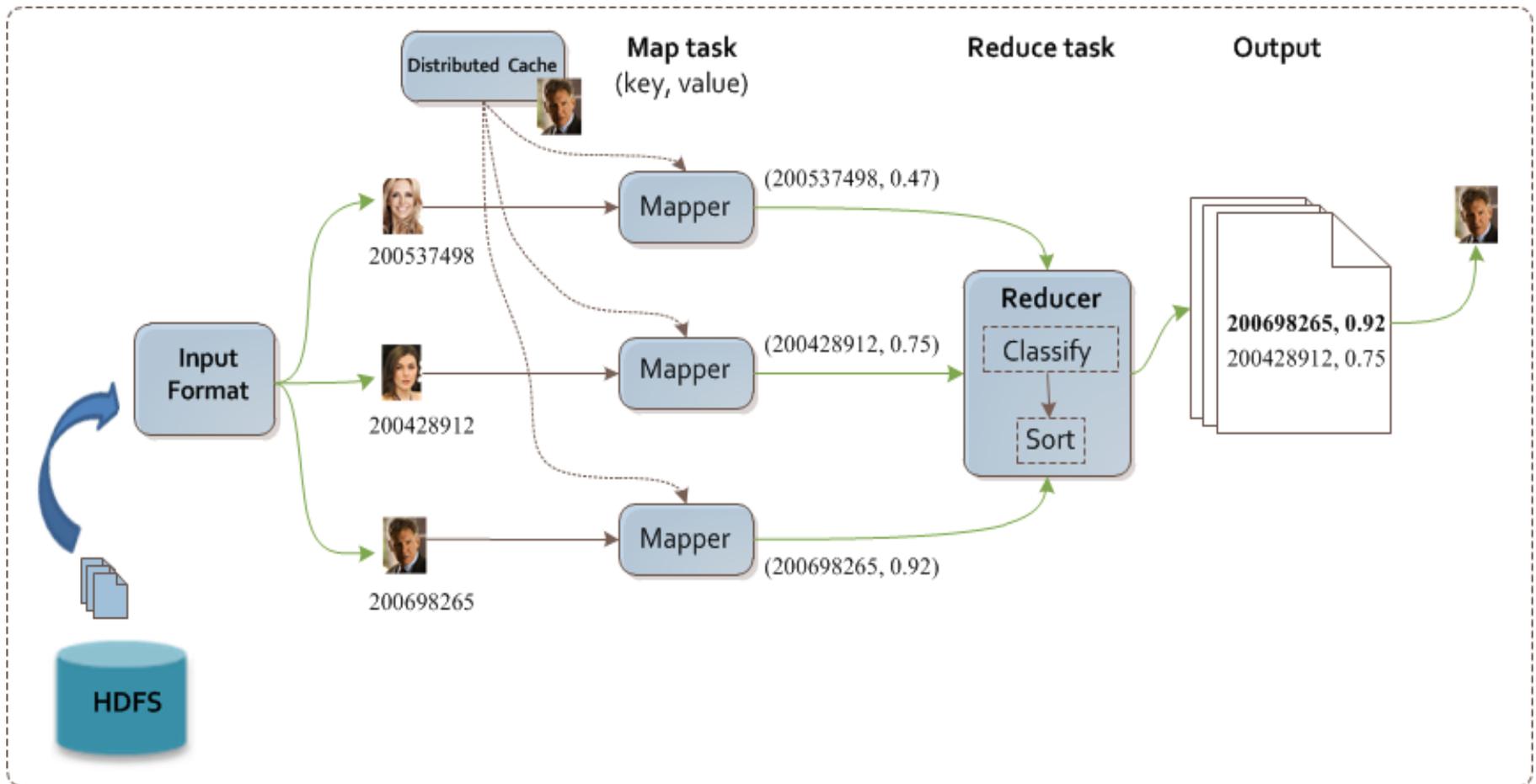
- Rentabilidad
- Confiabilidad
- Seguridad

Diseño

Diseño
Detalles de Implementación



Diseño



Implementación

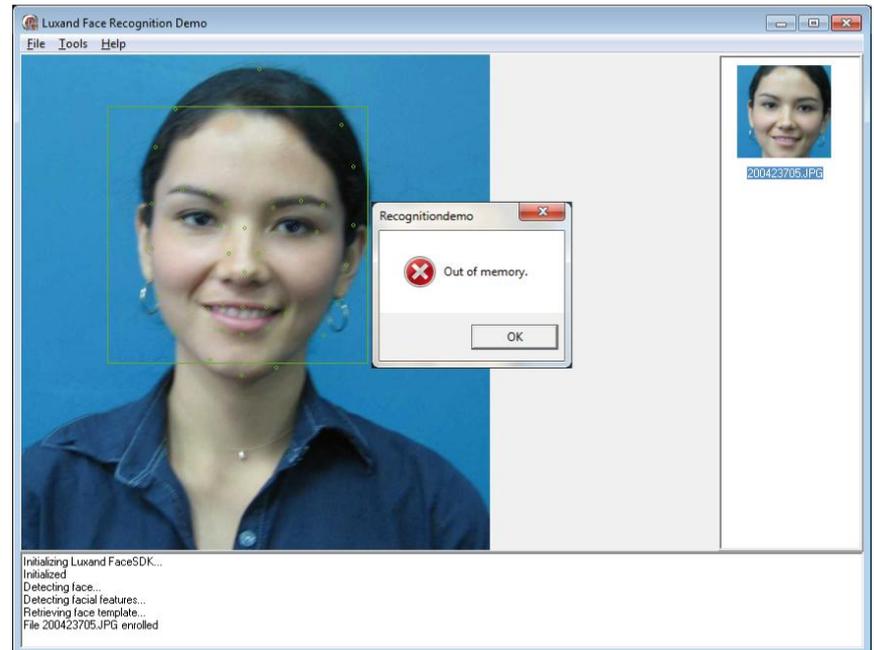
- WholeFileInputFormat
 - Lee el archivo de entrada en su totalidad.
 - Difiere de los formatos de entrada convencionales.
- Memoria Cache Distribuida de *Hadoop*.
- Obtención de Plantillas (*Templates*)
- Comparando las Plantillas
 - El resultado es el valor de similitud entre las 2.

Pruebas Realizadas



Aplicación de Escritorio

Error al tratar de encontrar un rostro entre 1000 imágenes.

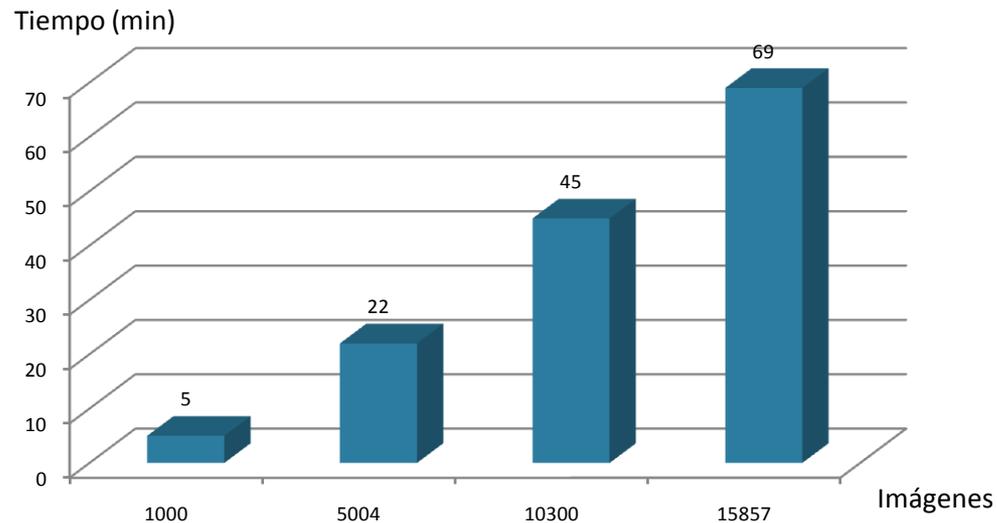


Pruebas

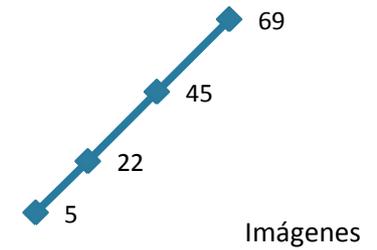
- Conjunto de Datos
 - Cerca de 15600 imágenes de estudiantes de ESPOL.
 - Aproximadamente 800 MB.

Prueba Incrementando # de Imágenes

Imágenes procesadas vs Tiempo



Tiempo (min)

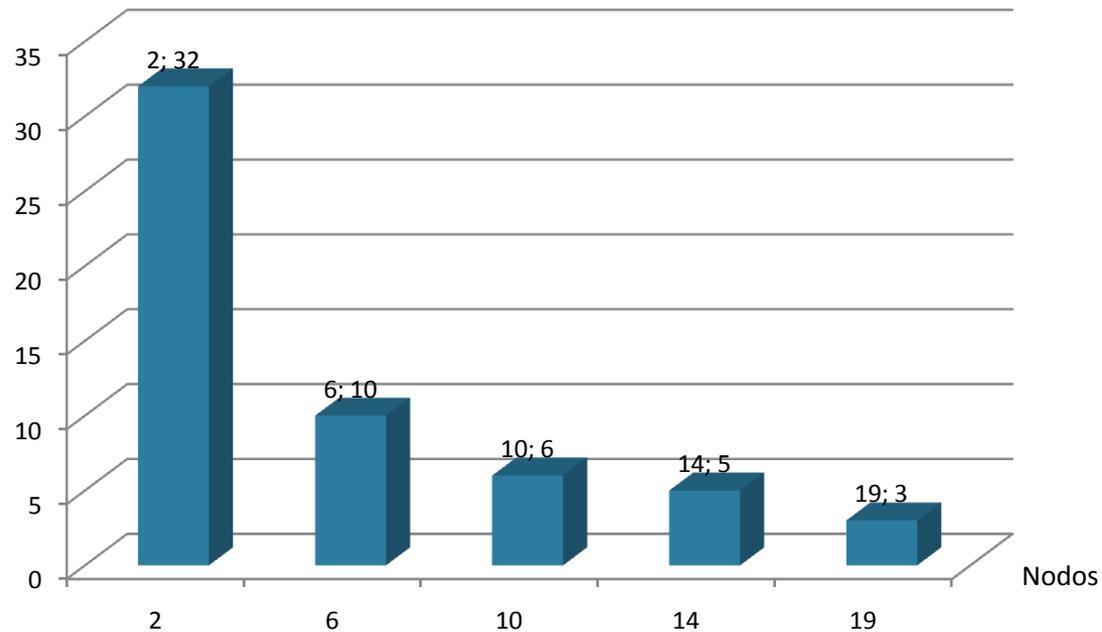


■ Prueba con 14 nodos

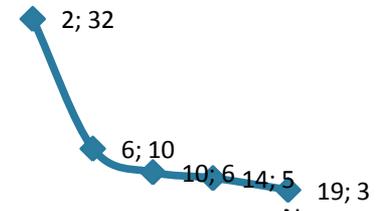
Prueba Incrementando # de Nodos

Tiempo (min)

Nodos vs Tiempo



Tiempo (min)



$R^2 = 0,9925$

■ Prueba con 1000 imágenes

Conclusiones y Recomendaciones



Conclusiones

- La plataforma Hadoop es capaz de procesar gran cantidad de imágenes eficientemente.
- El tiempo de ejecución disminuye de manera potencial con el aumento de nodos.
- La precisión de los resultados de reconocimiento facial dependen mucho de las condiciones de iluminación, rotación del rostro y calidad de la imagen.

Recomendaciones

- A partir del presente trabajo podría construirse una interfaz gráfica para este módulo.
- Se recomienda utilizar como entrada una imagen que contenga en su gran mayoría el rostro del individuo y que sea mayor a 512 x 512 píxeles.

¿Preguntas?

Gracias

