|  |
| --- |
| Sistemas de almacenamiento NoSQL |
| Práctica 2 |
| HBASE |

|  |
| --- |
| Alejandro del Rosal Carmona  Juan Ortega Ortega  23-3-2023 |

Logotipo, nombre de la empresa

Descripción generada automáticamente

Contenido

[2 Setup 3](#_Toc130756816)

[3 Detalle de Aeropuertos 3](#_Toc130756817)

[4 Detalle de Vuelos 4](#_Toc130756818)

[5 Consulta de Rutas 4](#_Toc130756819)

# Setup

Toda la práctica se ha realizado con scripts de Python desde el Edge accediendo a través de la IDE de VSC. Los datos de las tablas se han obtenido del siguiente directorio:

/tmp/nosql/airData/

Para ejecutar los scripts se ha empleado la versión de Python disponible en el Edge a través de:

/opt/miniconda3/bin/python3.8

# Detalle de Aeropuertos

Para el primer ejercicio, crearemos la table con los detalles de los aeropuertos, de forma que dado un código de aeropuerto obtengamos el resto de los atributos del aeropuerto buscado.

Para esto crearemos un script con las funciones auxiliares ya proporcionadas en la guía, una para realizar la conexión al database de Hbase y otra para convertir un valor dado a bytes usando el encoding utf-8, finalmente otras dos funciones principales de creación de tabla y carga de datos.

El script tomará dos argumentos a la hora de ejecutarse, –namespace y –table, que corresponden al namespace que vamos a usar y el nombre que vamos a darle a la tabla. Se emplea una única columna family ‘airport’ y el id de aeropuerto como rowKey.

Para lidiar con los elementos que contengan comas, al ser escasos se ha optado por su eliminación, por lo que al recorrer las rows si detectamos más columnas de la cuenta se eliminara el registro.

Texto

Descripción generada automáticamente

Figura 1: Ejemplo de ejecución

Captura de pantalla de computadora

Descripción generada automáticamente

Figura 2: Acceso desde el shell

# Detalle de Vuelos

A continuación, crearemos la tabla con los detalles de los vuelos, de forma que podamos consultar los vuelos de un mes o un día específico dentro de los disponibles en las tablas 2007 y 2008.

Para ello vamos a crear un script que recorrerá los csv disponibles (2 en nuestro caso) uno a uno e ira añadiendo las columnas que nos interesen. Los argumentos que requiere el script son idénticos a los del ejercicio anterior.

En cuanto a la tabla, la column family volvió a ser una ‘info’ y la rowKey una combinación de los campos: Year, Month, DayofMonth, DepTime y un contador único para cada registro. Esta combinación nos permitirá filtrar por día del mes como se pide en el ejercicio. Las consultas por fecha se realizarán a través de un scan con el siguiente filtro:

scan ‘mbd10:airData’, {STARTROW=>’20080210’}

Las columnas escogidas para incluir en nuestra tabla son: Year, Month,DayofMonth, DepTime, ArrTime, UniqueCarrier, FlightNum, TailNum, Origin, Dest, Distance y Cancelled.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Figura 3: Acceso desde el shell

# Consulta de Rutas

Para realizar el siguiente ejercicio en el que vamos a realizar consultas a través de parejas origen-destino. El primer paso va a ser la creación de una tabla con el procedimiento del ejercicio anterior, pero escogiendo las columnas que nos interesan para este caso y empleando una rowKey formada por: origen, destino y un identificador único para cada registro.

Texto

Descripción generada automáticamente

Figura 4: Acceso a origen-destino desde el shell

## Duración Promedio

Nuestro script accederá a la tabla origen-destino y buscara por la pareja que el usuario haya introducido, tras comprobar que existen vuelos para dicha combinación calculará el promedio descartando los registros que aparezcan nulos.

Texto

Descripción generada automáticamente

Figura 5: Ejemplo de ejecución

## Datos Extra

.

## Orden de Aparición

Nuestro script accederá a la tabla origen-destino y buscara por la pareja que el usuario haya introducido, tras comprobar que existen vuelos para dicha combinación calculará el promedio descartando los registros que aparezcan nulos.

Texto

Descripción generada automáticamente

## Modelo más usado

Nuestro script accederá a la tabla origen-destino y buscara por la pareja que el usuario haya introducido, tras comprobar que existen vuelos para dicha combinación calculará el promedio descartando los registros que aparezcan nulos.

Texto

Descripción generada automáticamente