

# ESPECIALIZACIÓN BIG DATA

## CTIC

Grupo Paperos

Universidad Nacional de Ingeniería  
Centro de Tecnologías de Información y Comunicaciones

October 24, 2020



# Schedule

## 1 Creando Tablas En GCP

## 2 Creando Tablas En Cloudera

- Configuración inicial
- Cloudera Manager
- Creando Tablas
- Generando Reporte



## Creando Tablas En GCP



# Creando DDBB

```
0: jdbc:hive2://> create database paperos location '/user/main/bd/paperos';
OK
No rows affected (2.793 seconds)
0: jdbc:hive2://> create external table paperos.DIM_CATEGORIA(
. . . . . > ID_CATEGORIA string,
. . . . . > NOM_CATEGORIA string
. . . . . > )
. . . . . > row format delimited
. . . . . > fields terminated by ';'
. . . . . > lines terminated by '\n'
. . . . . > stored as textfile
. . . . . > location '/user/main/bd/paperos/DIM_CATEGORIA';
OK
No rows affected (0.621 seconds)
0: jdbc:hive2://> load data inpath 'dataset/Categoria.csv' into table paperos.DIM_CATEGORIA;
Loading data to table paperos.dim_categoria
OK
No rows affected (2.806 seconds)
```

Figure: Elaboración propia.



# Visualizando tabla

```
0: jdbc:hive2://> select * from paperos.DIM_CATEGORIA;
OK
+-----+-----+
| dim_categoria.id_categoria | dim_categoria.nom_categoria |
+-----+-----+
| ID_CATEGORIA              | NOM_CATEGORIA              |
| 1                          | LACTEOS                    |
| 2                          | LIMPIEZA                   |
| 3                          | ABARROTES                  |
| 4                          | LICORES                    |
| 5                          | CARNES                     |
+-----+-----+
6 rows selected (2.373 seconds)
0: jdbc:hive2://> █
```

Figure: Elaboración propia.



## Verificando tablas

```
lpfernando@cluster-a51f-m:~$ hdfs dfs -ls /user/main/bd/paperos
Found 9 items
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 20:56 /user/main/bd/paperos/DIM_CATEGORIA
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 21:05 /user/main/bd/paperos/DIM_CLIENTE
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 21:05 /user/main/bd/paperos/DIM_DEPARTAMENTO
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 21:05 /user/main/bd/paperos/DIM_PRODUCTO
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 21:02 /user/main/bd/paperos/DIM_PROVINCIA
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 21:03 /user/main/bd/paperos/DIM_SEDE
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 21:04 /user/main/bd/paperos/DIM_TIEMPO
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 21:04 /user/main/bd/paperos/DIM_VENDEDOR
drwxr-xr-x - lpfernando hadoop      0 2020-10-15 21:06 /user/main/bd/paperos/FACT_VENTA
lpfernando@cluster-a51f-m:~$
```

Figure: Elaboración propia.



## Creando Tablas En Cloudera



# Configuración Cloudera

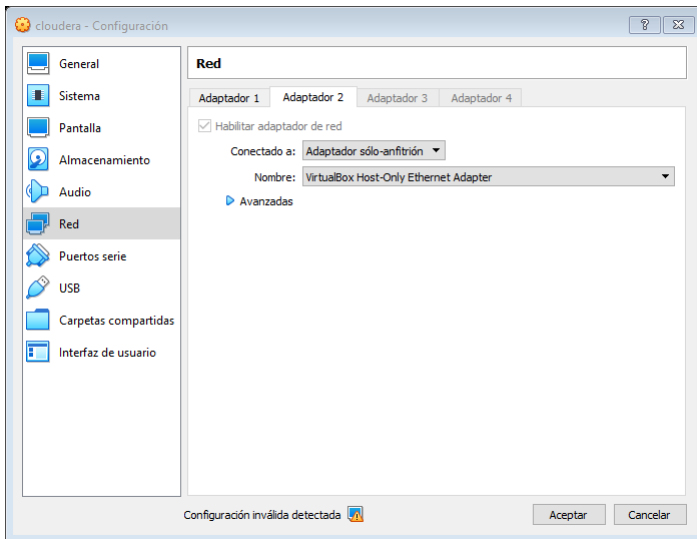


Figure: Elaboración propia.



# Transfiriendo Datasets

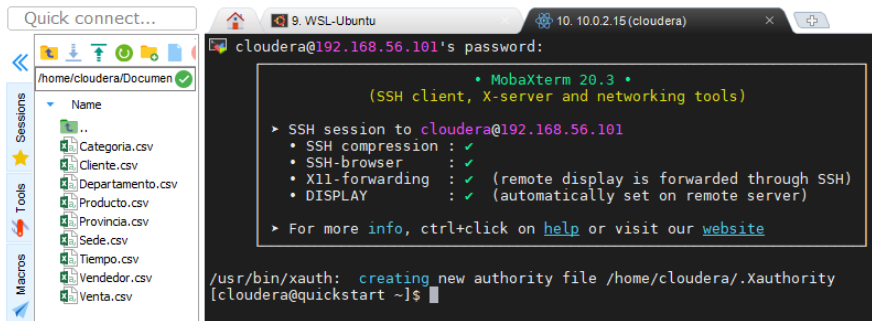
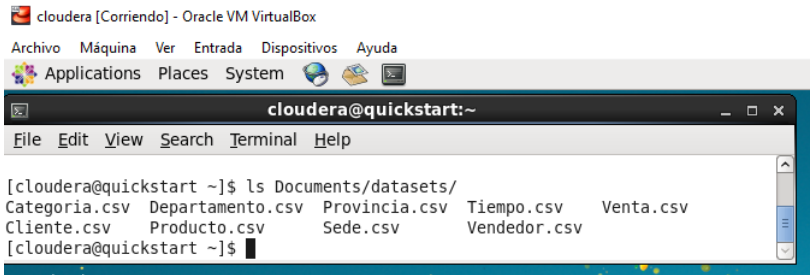


Figure: Elaboración propia.



# Comprobando Datasets



The screenshot shows a terminal window titled "cloudera [Corriendo] - Oracle VM VirtualBox". The window has a menu bar with "Archivo", "Máquina", "Ver", "Entrada", "Dispositivos", and "Ayuda". Below the menu bar is a toolbar with icons for "Applications", "Places", "System", and a terminal icon. The terminal window itself has a title bar "cloudera@quickstart:~" and a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal content shows the command `ls Documents/datasets/` being executed, resulting in a list of CSV files: `Categoria.csv`, `Departamento.csv`, `Provincia.csv`, `Tiempo.csv`, `Venta.csv`, `Cliente.csv`, `Producto.csv`, `Sede.csv`, and `Vendedor.csv`. The prompt `[cloudera@quickstart ~]$` is visible at the bottom.

```
cloudera [Corriendo] - Oracle VM VirtualBox

Archivo  Máquina  Ver  Entrada  Dispositivos  Ayuda

Applications  Places  System

cloudera@quickstart:~

File Edit View Search Terminal Help

[cloudera@quickstart ~]$ ls Documents/datasets/
Categoria.csv  Departamento.csv  Provincia.csv  Tiempo.csv  Venta.csv
Cliente.csv    Producto.csv      Sede.csv       Vendedor.csv
[cloudera@quickstart ~]$
```

Figure: Elaboración propia.



# Obteniendo Autenticación

```
[cloudera@quickstart ~]$ sudo /home/cloudera/cloudera-manager --force --express
[QuickStart] Shutting down CDH services via init scripts...
[QuickStart] Disabling CDH services on boot...
[QuickStart] Starting Cloudera Manager daemons...
[QuickStart] Waiting for Cloudera Manager API...
[QuickStart] Configuring deployment...
[QuickStart] Deploying client configuration...
[QuickStart] Starting Cloudera Management Service...
[QuickStart] Enabling Cloudera Manager daemons on boot...
```

Success! You can now log into Cloudera Manager from the QuickStart VM's browser:

`http://quickstart.cloudera:7180`

Username: cloudera

Password: cloudera

Figure: Elaboración propia.



# Activando herramientas

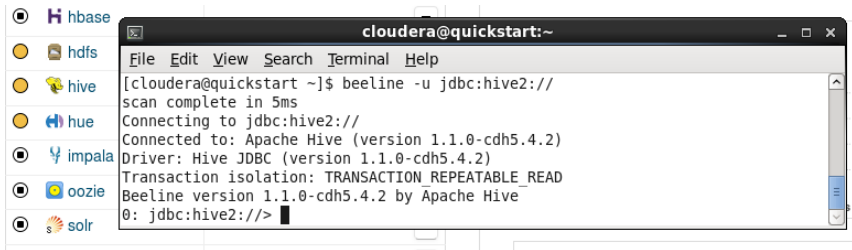


Figure: Elaboración propia.



# Consultando Tabla

```
0: jdbc:hive2://> select * from papacloud.DIM_CATEGORIA;
```

```
OK
```

dim_categoria.id_categoria	dim_categoria.nom_categoria
1	LACTEOS
2	LIMPIEZA
3	ABARROTES
4	LICORES
5	CARNES

```
5 rows selected (1.307 seconds)
```

```
0: jdbc:hive2://> █
```

Figure: Elaboración propia.



# Verificando Creación DDBB

```
0: jdbc:hive2://> show databases;
OK
+-----+
| database_name |
+-----+
| default      |
| papacloud    |
| paperos      |
+-----+
3 rows selected (0.065 seconds)
0: jdbc:hive2://> █
```

Figure: Elaboración propia.



# Mostrando tablas

```
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/papacloud
```

```
Found 9 items
```

```
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:11 /user/cloudera/papacloud/DIM_CATEGORIA
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:17 /user/cloudera/papacloud/DIM_CLIENTE
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:18 /user/cloudera/papacloud/DIM_DEPARTAMENT
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:18 /user/cloudera/papacloud/DIM_PRODUCTO
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:19 /user/cloudera/papacloud/DIM_PROVINCIA
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:19 /user/cloudera/papacloud/DIM_SEDE
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:19 /user/cloudera/papacloud/DIM_TIEMPO
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:20 /user/cloudera/papacloud/DIM_VENDEDOR
drwxr-xr-x - cloudera cloudera      0 2020-10-16 16:20 /user/cloudera/papacloud/FACT_VENTA
[cloudera@quickstart ~]$ █
```

Figure: Elaboración propia.



# Instalando ODBC

## Cloudera ODBC Driver for Apache Hive DSN Setup



Data Source Name:	<input type="text" value="HiveConnection"/>
Description:	<input type="text"/>
Hive Server Type:	<input type="text" value="Hive Server 2"/>
Service Discovery Mode:	<input type="text" value="No Service Discovery"/>
Host(s):	<input type="text" value="192.168.56.101"/>
Port:	<input type="text" value="10000"/>
Database:	<input type="text" value="default"/>
ZooKeeper Namespace:	<input type="text"/>
Authentication	
Mechanism:	<input type="text" value="User Name"/>
Realm:	<input type="text"/>
Host FQDN:	<input type="text" value="_HOST"/>
Service Name:	<input type="text" value="hive"/>
<input checked="" type="checkbox"/> Canonicalize Principal FQDN	
<input type="checkbox"/> Delegate Kerberos Credentials	
User Name:	<input type="text" value="cloudera"/>
Password:	<input type="password"/>
<input type="checkbox"/> Save Password (Encrypted)	





# Mostrando tablas en PowerBI

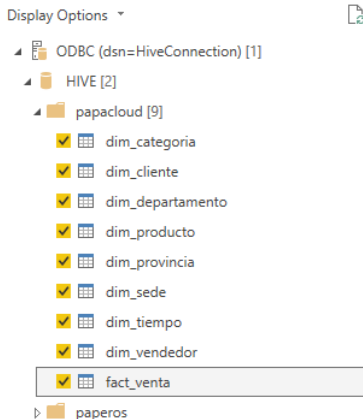


Figure: Elaboración propia.



# Visualización Reporte

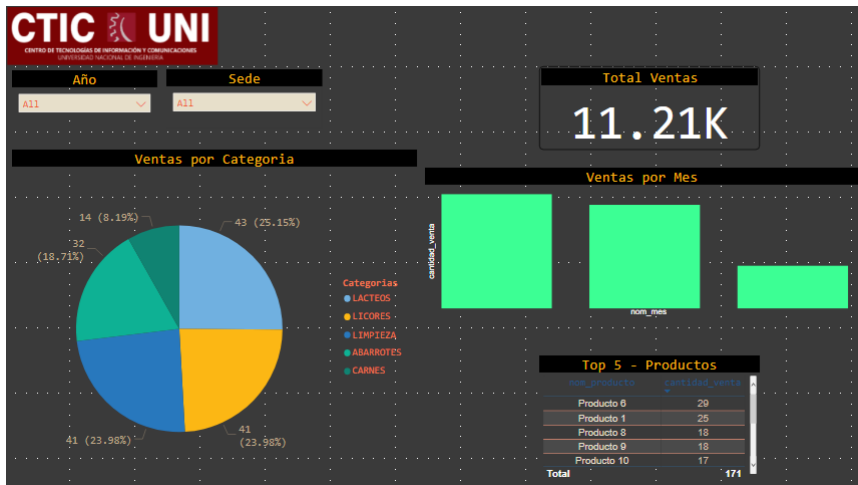


Figure: Elaboración propia.