

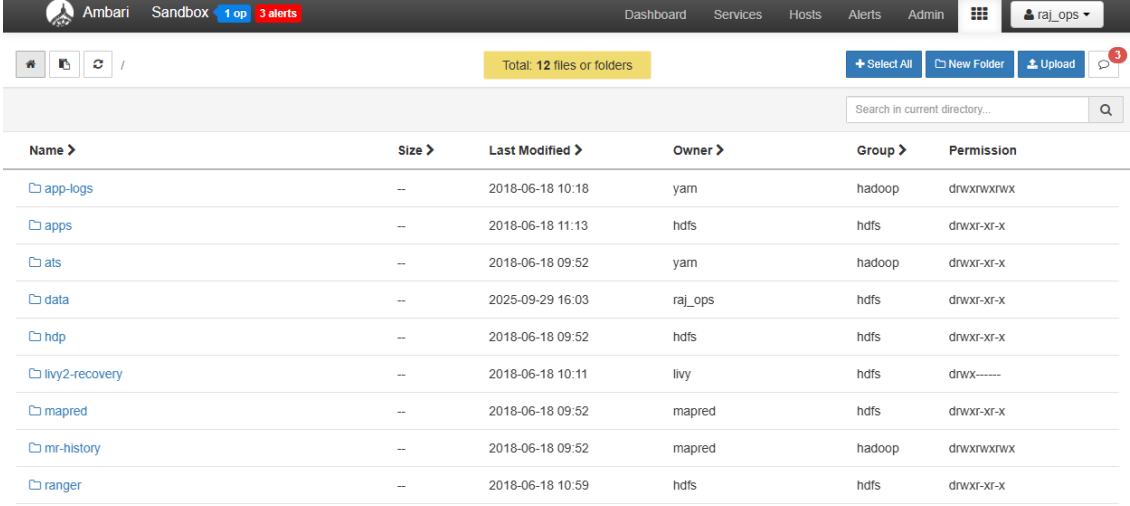
LABORATORIO 2:

Participantes:

- Cabana Cazani, Gabriel
- Larico Cruz, Diego

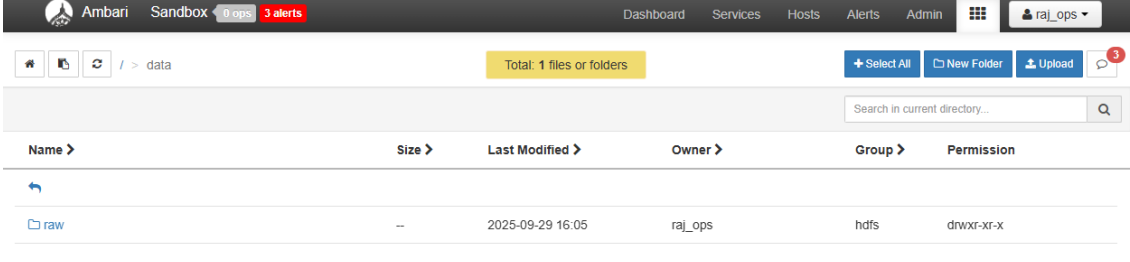
INGESTA DE DATOS

Nos localizamos en Files view para cargar los archivos csv y almacenarlos en la carpeta Data/raw



The screenshot shows the Ambari Files view for the 'Sandbox' environment. The top navigation bar includes 'Dashboard', 'Services', 'Hosts', 'Alerts', and 'Admin'. The user is logged in as 'raj_ops'. The main view displays a table of files and folders in the root directory. The table has columns for Name, Size, Last Modified, Owner, Group, and Permission. The files listed are: app-logs, apps, ats, data, hdp, livy2-recovery, mapred, mr-history, and ranger.

Name	Size	Last Modified	Owner	Group	Permission
app-logs	--	2018-06-18 10:18	yarn	hadoop	drwxrwxrwx
apps	--	2018-06-18 11:13	hdfs	hdfs	drwxr-xr-x
ats	--	2018-06-18 09:52	yarn	hadoop	drwxr-xr-x
data	--	2025-09-29 16:03	raj_ops	hdfs	drwxr-xr-x
hdp	--	2018-06-18 09:52	hdfs	hdfs	drwxr-xr-x
livy2-recovery	--	2018-06-18 10:11	livy	hdfs	drwx-----
mapred	--	2018-06-18 09:52	mapred	hdfs	drwxr-xr-x
mr-history	--	2018-06-18 09:52	mapred	hadoop	drwxrwxrwx
ranger	--	2018-06-18 10:59	hdfs	hdfs	drwxr-xr-x



The second screenshot shows the Ambari Files view for the 'data' directory. The table lists one file: 'raw'.

Name	Size	Last Modified	Owner	Group	Permission
raw	--	2025-09-29 16:05	raj_ops	hdfs	drwxr-xr-x

Usamos el siguiente comando para cargar el archivo csv al HDFS

```
[root@sandbox-hdp ~]# hdfs dfs -ls /
Found 12 items
drwxrwxrwx - yarn      hadoop      0 2018-06-18 15:18 /app-logs
drwxr-xr-x - hdfs      hdfs        0 2018-06-18 16:13 /apps
drwxr-xr-x - yarn      hadoop      0 2018-06-18 14:52 /ats
drwxr-xr-x - raj_ops   hdfs        0 2025-09-29 21:03 /data
drwxr-xr-x - hdfs      hdfs        0 2018-06-18 14:52 /hdp
drwx----- - livy      hdfs        0 2018-06-18 15:11 /livy2-recovery
drwxr-xr-x - mapred    hdfs        0 2018-06-18 14:52 /mapred
drwxrwxrwx - mapred    hadoop      0 2018-06-18 14:52 /mr-history
drwxr-xr-x - hdfs      hdfs        0 2018-06-18 15:59 /ranger
drwxrwxrwx - spark     hadoop      0 2025-09-29 22:36 /spark2-history
drwxrwxrwx - hdfs      hdfs        0 2018-06-18 16:06 /tmp
drwxr-xr-x - hdfs      hdfs        0 2018-06-18 16:08 /user
[root@sandbox-hdp ~]# hdfs dfs -ls /data
Found 1 items
drwxr-xr-x - raj_ops   hdfs        0 2025-09-29 21:05 /data/raw
[root@sandbox-hdp ~]# hdfs dfs -put dim_cliente /data/raw
```

Vemos los archivos cargados en el HDFS

Ambari Sandbox 0 ops 2 alerts Dashboard Services Hosts Alerts Admin raj_ops

/ > data > raw Total: 7 files or folders + Select All New Folder Upload 3

Search in current directory...

Name >	Size >	Last Modified >	Owner >	Group >	Permission
dim_cliente.csv	5.9 kB	2025-09-29 16:04	raj_ops	hdfs	-rw-r--r--
dim_cliente_rfm.csv	4.1 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
dim_periodo_mensual.csv	4.4 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
dim_producto.csv	6.6 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
dim_promocion_precio.csv	5.3 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
dim_tienda_canal.csv	5.6 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
tabla_salida.csv	5.2 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--

Cargamos la tabla para las consultas sql

Ambari Sandbox 0 ops 0 alerts Dashboard Services Hosts Alerts Admin raj_ops

Hive Query Saved Queries History UDFs Upload Table

Upload from Local File type CSV Database default Table name Trx_ticket Stored as ORC Contains endlines? Upload Table

cod_ticket num_secuencia sk_cliente sk_producto

STRING INT INT INT

TK0001	2	52	57
TK0002	3	47	41
TK0003	2	66	54
TK0004	1	53	24

Realizamos consulta básica para ver los 10 primeros registros

Query Editor

Worksheet * x trx_ticket sample x

```
1 SELECT * FROM trx_ticket LIMIT 10;
```

Query Process Results (Status: SUCCEEDED) Save results...

Logs Results

Filter columns... previous next

trx_ticket.cod_ticket	trx_ticket.num_secuencia	trx_ticket.sk_cliente	trx_ticket.sk_producto	trx_ticket.sk_tienda	trx_ticl
TK0001	2	52	57	99	76
TK0002	3	47	41	47	40
TK0003	2	66	54	81	68
TK0004	1	53	24	8	73
TK0005	4	89	53	2	40
TK0006	3	13	76	2	4
TK0007	5	83	69	66	74
TK0008	4	98	66	14	13
TK0009	2	79	56	8	2
TK0010	1	46	36	22	35

Usamos lo siguiente para poder crear la tabla externa

Query Editor

```

1 CREATE EXTERNAL TABLE IF NOT EXISTS tabla_salida_externa (
2   cod_ticket STRING,
3   cod_cliente INT,
4   cod_sucursal INT,
5   cod_fecha INT,
6   mto_venta_bruta DECIMAL(18,2),
7   mto_descuento DECIMAL(18,2),
8   mto_venta_neta DECIMAL(18,2),
9   mto_margen DECIMAL(18,2)
10 )
11 ROW FORMAT DELIMITED
12 FIELDS TERMINATED BY '\t'
13 STORED AS TEXTFILE
14 LOCATION '/data/raw';

```

Worksheet

SQL

TEZ

Actualizamos y si se creó la tabla

Database Explorer

default

Search tables...

Databases

- default
- tabla salida externa
- trx_ticket
- foodmart

Vemos el formato de nuestra tabla

Query Editor		
Worksheet		
1 describe formatted tabla_salida_external;		
col_name	data_type	comment
# col_name	data_type	comment
""	null	null
cod_ticket	string	""
cod_cliente	int	""
cod_sucursal	int	""
cod_fecha	int	""
mto_venta_bruta	decimal(18,2)	""
mto_descuento	decimal(18,2)	""
mto_venta_neta	decimal(18,2)	""
mto_margen	decimal(18,2)	""
""	null	null
# Detailed Table Information	null	null

Luego se puede apreciar que es una tabla externa.

Location:	hdfs://sandbox-hdp.hortonworks.com:8020/data/raw	null
Table Type:	EXTERNAL_TABLE	null
Table Parameters:	null	null
""	EXTERNAL	TRUE