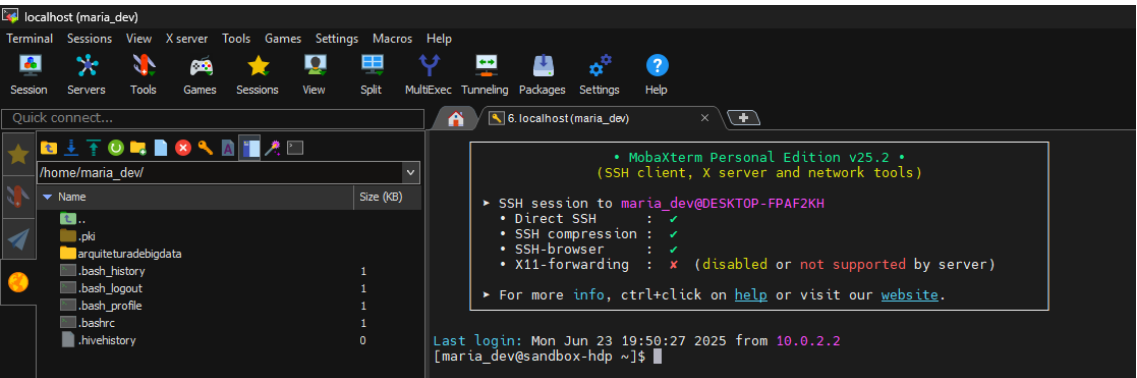
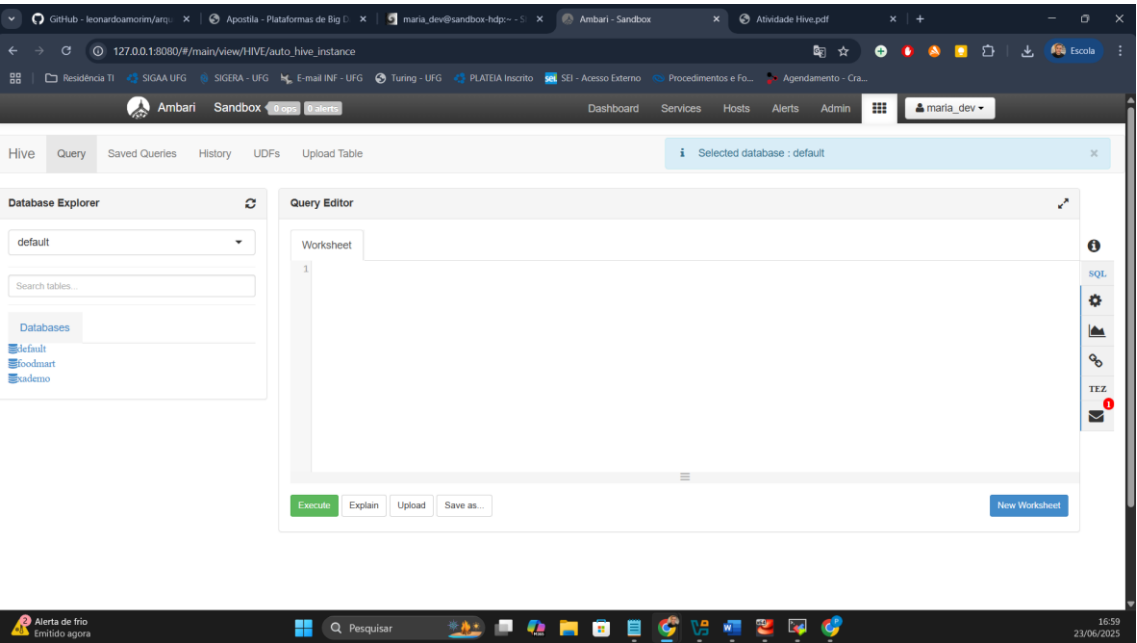
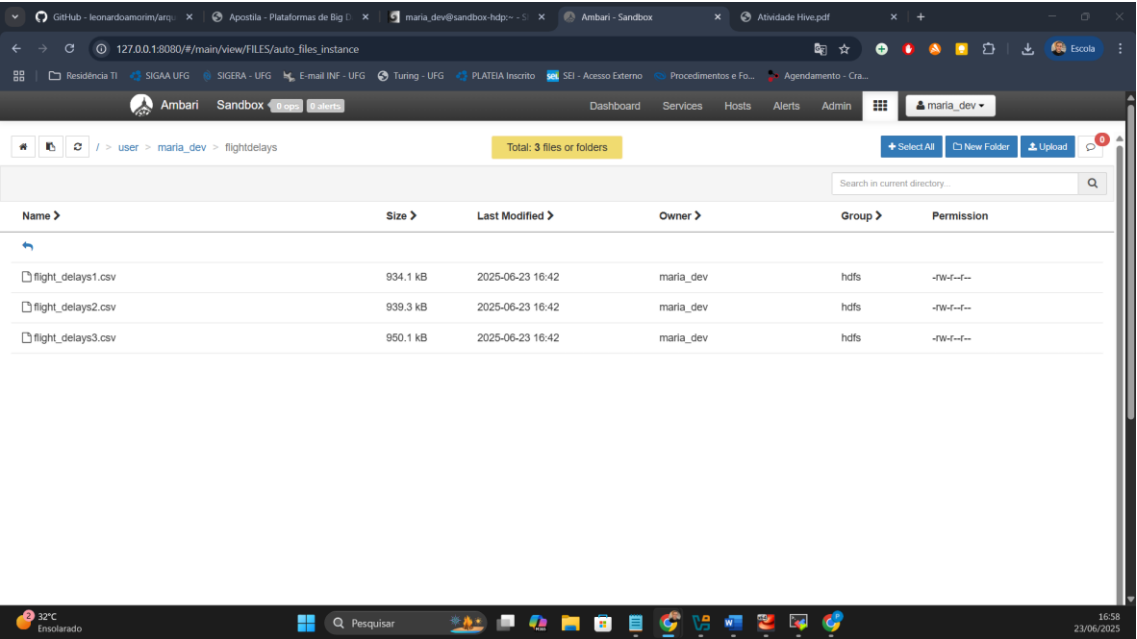


Git: <https://github.com/leonardoamorim/arquiteturadebigdata>

## Atividade 01 - Obtenção de dados e inserção de dados no HDFS



## Atividade 02 - Criação de tabela externa no Hive

```
DROP TABLE IF EXISTS flightdelays;
```

```
CREATE EXTERNAL TABLE flightdelays (
```

```
Year INT,
```

```
Month INT,
```

```
DayofMonth INT,
```

```
DayOfWeek INT,
```

```
DepTime INT,
```

```
CRSDepTime INT,
```

```
ArrTime INT,
```

```
CRSArrTime INT,
```

```
UniqueCarrier STRING,
```

```
FlightNum INT,
```

```
TailNum STRING,
```

```
ActualElapsedTime INT,
```

```
CRSElapsedTime INT,
```

```
AirTime INT,
```

```
ArrDelay INT,
```

```
DepDelay INT,
```

```
Origin STRING,
```

```
Dest STRING,
```

```
Distance INT,
```

```
TaxiIn INT,
```

```
TaxiOut INT,
```

```
Cancelled INT,
```

```
CancellationCode STRING,
```

```
Diverted INT,
```

CarrierDelay INT,

WeatherDelay INT,

NASDelay INT,

SecurityDelay INT,

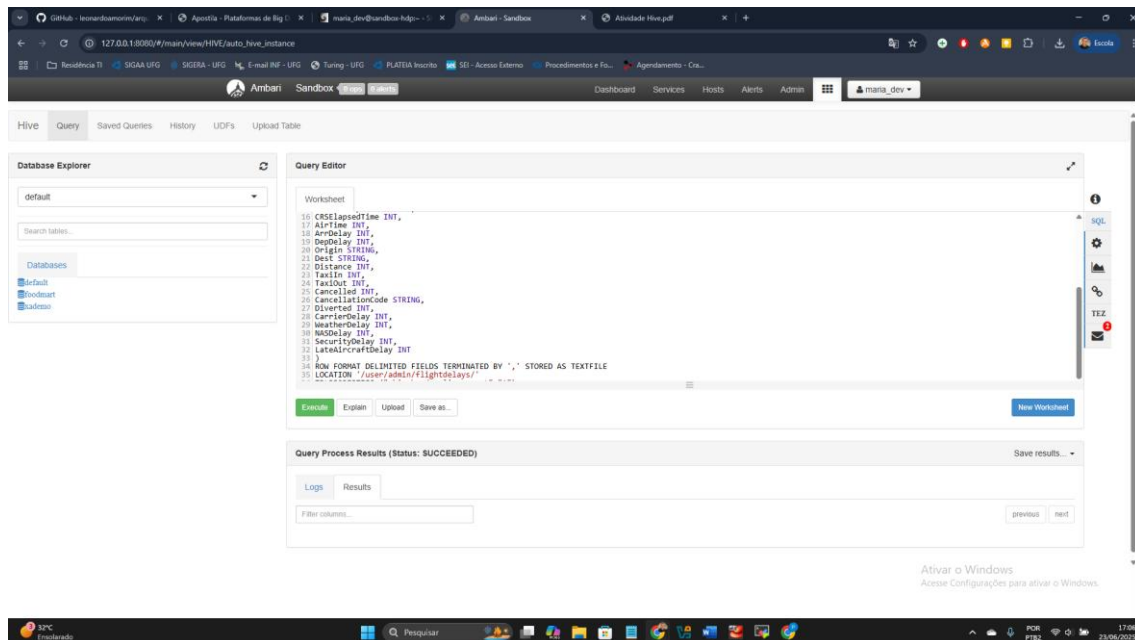
LateAircraftDelay INT

)

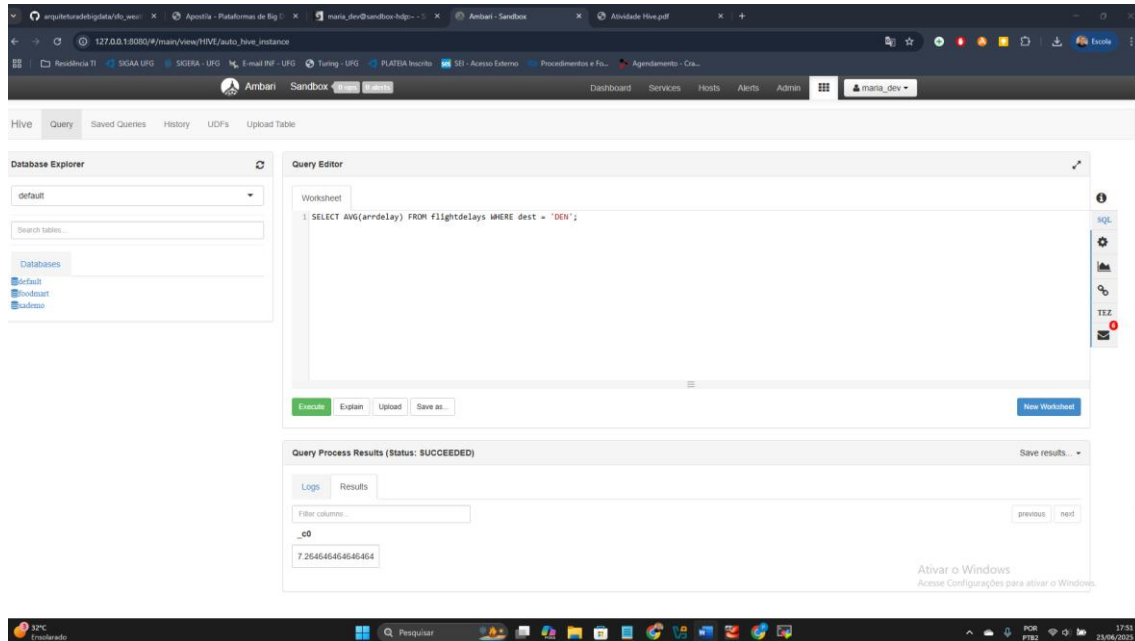
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE

LOCATION '/user/maria\_dev/flightdelays/'

TBLPROPERTIES ("skip.header.line.count"="1");



## Atividade 03 - Analisando dados com o Hive



The screenshot shows the Ambari Hive interface. The top navigation bar includes links for Dashboard, Services, Hosts, Alerts, Admin, and a user profile for 'maria\_dev'. The main interface is divided into three sections: Database Explorer, Query Editor, and Query Process Results.

**Database Explorer:** Shows a dropdown menu for 'default' and a search bar. Below, it lists databases: 'default', 'hivedb', and 'hivesite'.

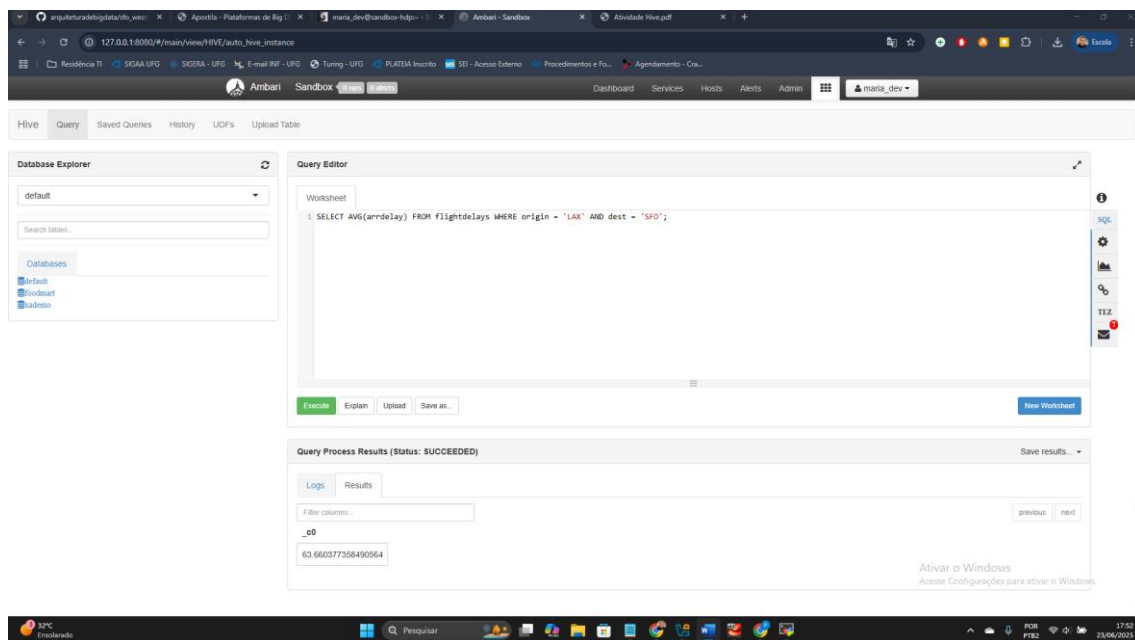
**Query Editor:** Contains a worksheet with the following SQL query:

```
1 SELECT AVG(arrdelay) FROM flightdelays WHERE dest = 'DEN';
```

Buttons for 'Execute', 'Explain', 'Upload', and 'Save as...' are visible. A 'New Worksheet' button is also present.

**Query Process Results (Status: SUCCEEDED):** Shows the results of the query execution. The 'Results' tab is selected, displaying a single row with the value '7.264646464646464'.

The bottom status bar indicates the system is '32°C Emulador' and the date is '23/06/2023'.



The screenshot shows the Ambari Hive interface. The top navigation bar includes links for Dashboard, Services, Hosts, Alerts, Admin, and a user profile for 'maria\_dev'. The main interface is divided into three sections: Database Explorer, Query Editor, and Query Process Results.

**Database Explorer:** Shows a dropdown menu for 'default' and a search bar. Below, it lists databases: 'default', 'hivedb', and 'hivesite'.

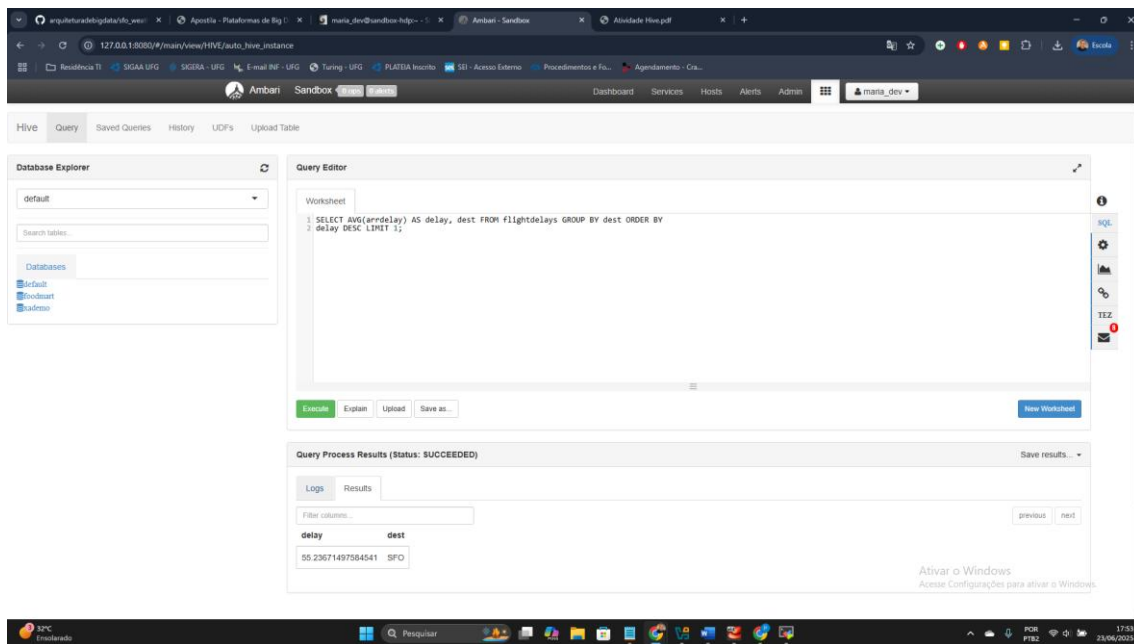
**Query Editor:** Contains a worksheet with the following SQL query:

```
1 SELECT AVG(arrdelay) FROM flightdelays WHERE origin = 'LAX' AND dest = 'SFO';
```

Buttons for 'Execute', 'Explain', 'Upload', and 'Save as...' are visible. A 'New Worksheet' button is also present.

**Query Process Results (Status: SUCCEEDED):** Shows the results of the query execution. The 'Results' tab is selected, displaying a single row with the value '63.660377358490564'.

The bottom status bar indicates the system is '32°C Emulador' and the date is '23/06/2023'.



## Atividade 04 - Definir e preencher uma tabela ORCFile

-- Cria tabela temporária em formato texto

```
DROP TABLE IF EXISTS sfo_weather_txt;
```

```
CREATE TABLE sfo_weather_txt (
```

```
    station_name STRING,
```

```
    year INT,
```

```
    month INT,
```

```
    dayofmonth INT,
```

```
    precipitation INT,
```

```
    temperature_max INT,
```

```
    temperature_min INT
```

```
)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ','
```

```
STORED AS TEXTFILE;
```

-- Carrega dados do arquivo CSV local

```
LOAD DATA LOCAL INPATH  
'/home/maria_dev/arquiteturadebigdata/sfo_weather.csv'
```

```
OVERWRITE INTO TABLE sfo_weather_txt;
```

```
-- Cria tabela definitiva em formato ORC
```

```
DROP TABLE IF EXISTS sfo_weather;
```

```
CREATE TABLE sfo_weather (
```

```
    station_name STRING,
```

```
    year INT,
```

```
    month INT,
```

```
    dayofmonth INT,
```

```
    precipitation INT,
```

```
    temperature_max INT,
```

```
    temperature_min INT
```

```
)
```

```
STORED AS ORC;
```

```
-- Insere dados da tabela texto para a tabela ORC
```

```
INSERT INTO TABLE sfo_weather
```

```
SELECT * FROM sfo_weather_txt;
```

```
-- Consulta os dados da tabela ORC
```

```
SELECT * FROM sfo_weather;
```

Database Explorer

Query Editor

```
Worksheet
8  month INT,
9  dayofmonth INT,
10 precipitation INT,
11 temperature_max INT,
12 temperature_min INT
13
14 ROW FORMAT DELIMITED
15 FIELDS TERMINATED BY ','
16 STORED AS TEXTFILE;
17
18 -- Carregar os dados do HDFS
19 LOAD DATA INPATH '/user/ariaia_dev/flightdelays/sfo_weather.csv'
20 OVERWRITE INTO TABLE sfo_weather_txt;
21
22 -- Remover tabela DBE se j   existir
23 DROP TABLE IF EXISTS sfo_weather;
24
25 -- criar tabela DBE otimizada
26 CREATE TABLE sfo_weather (
27   station_name STRING,
```

Query Process Results (Status: SUCCEEDED)

sfo_weather.station_name	sfo_weather.year	sfo_weather.month	sfo_weather.dayofmonth	sfo_weather.precipitation	sfo_weather.temperature_max	sfo_weather.temperature_min
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	1	0	122	39
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	2	0	117	39
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	3	43	150	94
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	4	533	150	100

Query Process Results (Status: SUCCEEDED)

sfo_weather.station_name	sfo_weather.year	sfo_weather.month	sfo_weather.dayofmonth	sfo_weather.precipitation	sfo_weather.temperature_max	sfo_weather.temperature_min
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	1	0	122	39
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	2	0	117	39
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	3	43	150	94
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	4	533	150	100
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	5	196	122	78
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	6	15	106	50
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	7	0	111	67
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	8	20	128	61
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	9	3	106	67
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	10	25	100	89
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	11	0	117	89
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	12	0	133	67
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	1	13	0	144	67

SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	6	3	122	83
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	7	0	139	50
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	8	0	161	78
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	9	0	189	56
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	10	0	189	72
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	11	0	189	72
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	12	0	189	72
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	13	0	156	78
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	14	0	156	83
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	15	0	156	50
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	16	0	150	56
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	17	0	122	56
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	18	0	122	83
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2008	2	19	163	122	89

## Atividade 05 - Junção de dados do Hive

SET hive.execution.engine= tez;

DROP TABLE IF EXISTS flights\_weather;

CREATE TABLE flights\_weather STORED AS TEXTFILE AS SELECT fd.\*,

sw.temperature\_max, sw.temperature\_min FROM flightdelays fd JOIN sfo\_weather  
sw

ON fd.year = sw.year AND fd.month = sw.month AND fd.dayofmonth =  
sw.dayofmonth

WHERE fd.origin = 'SFO' OR fd.dest = 'SFO';

SELECT \* FROM flights\_weather;



Database Explorer

Query Editor

```
1 SET hive.execution.engine=tez;
2
3 DROP TABLE IF EXISTS flights_weather;
4 CREATE TABLE flights_weather STORED AS TEXTFILE AS SELECT fd.*,
5   sw.temperature_max, sw.temperature_min FROM flightdelays fd JOIN sfo_weather sw
6   ON fd.year = sw.year AND fd.month = sw.month AND fd.dayofmonth = sw.dayofmonth
7 WHERE fd.origin = 'SFO' OR fd.dest = 'SFO';
8
9 SELECT * FROM flights_weather;
```

Query Process Results (Status: SUCCEEDED)

flights_weather.year	flights_weather.month	flights_weather.dayofmonth	flights_weather.dayofweek	flights_weather.deptime	flights_weather.crsdeptime	flights_weather.arrtime	flights_weather.temperature_max	flights_weather.temperature_min
2008	1	3	4	1426	1355	1605	1530	
2008	1	3	4	1009	910	1148	1045	
2008	1	3	4	2021	1700	2303	1835	
2008	1	3	4	2025	1905	2208	2040	
2008	1	3	4	603	605	729	740	
2008	1	3	4	2301	2105	59	2240	

Query Process Results (Status: SUCCEEDED)

flights_weather.year	flights_weather.month	flights_weather.dayofmonth	flights_weather.dayofweek	flights_weather.deptime	flights_weather.crsdeptime	flights_weather.arrtime	flights_weather.temperature_max	flights_weather.temperature_min
2008	1	3	4	1426	1355	1605	1530	
2008	1	3	4	1009	910	1148	1045	
2008	1	3	4	2021	1700	2303	1835	
2008	1	3	4	2025	1905	2208	2040	
2008	1	3	4	603	605	729	740	
2008	1	3	4	2301	2105	59	2240	
2008	1	3	4	1518	1215	1645	1350	
2008	1	3	4	706	710	829	835	
2008	1	3	4	null	905	null	1025	
2008	1	3	4	2321	1955	38	2115	
2008	1	3	4	null	1620	null	1740	
2008	1	3	4	2008	1805	2139	1930	
2008	1	3	4	1625	1430	1748	1550	
2008	1	3	4	1305	1050	1421	1210	
2008	1	3	4	1558	1245	1709	1405	
2008	1	3	4	2131	1915	29	2205	
2008	1	3	4	1736	1305	2031	1555	
2008	1	3	4	1319	950	1615	1240	
2008	1	3	4	1716	1440	1854	1620	
2008	1	3	4	2202	2110	2344	2250	
2008	1	3	4	1839	1720	1900	1900	
2008	1	3	4	1319	1230	1445	1410	

year	month	day	station_name	precipitation	temperature_max	temperature_min	
2008	1	3	4	1516	1440	1646	1610
2008	1	3	4	844	845	1020	1015
2008	1	3	4	2059	1620	2216	1750
2008	1	3	4	2225	2105	2350	2235
2008	1	3	4	1800	1805	1929	1930
2008	1	3	4	852	850	1009	1025
2008	1	3	4	1053	1050	1213	1220
2008	1	3	4	1801	1615	1919	1740
2008	1	3	4	1447	1235	1558	1405
2008	1	3	4	2155	1955	2308	2120
2008	1	3	4	null	710	null	840
2008	1	3	4	null	1430	null	1555
2008	1	3	4	1657	1310	2247	1920
2008	1	3	4	1911	1650	103	2300
2008	1	3	4	703	705	1316	1315
2008	1	3	4	1221	1040	1353	1210
2008	1	3	4	803	805	929	935
2008	1	3	4	619	620	742	750
2008	1	3	4	1324	1220	1447	1350
2008	1	3	4	2325	1900	null	2030
2008	1	3	4	1711	1415	1831	1545

## Atividade 06 - Tabelas particionadas do Hive

-- Remover a tabela particionada se já existir

```
DROP TABLE IF EXISTS weather_partitioned;
```

-- Criar tabela particionada por ano e mês

```
CREATE TABLE weather_partitioned (
```

```
    station_name STRING,
```

```
    dayofmonth INT,
```

```
    precipitation INT,
```

```
    temperature_max INT,
```

```
    temperature_min INT
```

```
)
```

```
PARTITIONED BY (year INT, month INT)
```

```
STORED AS ORC;
```

-- Inserir dados na partição do ano de 2008 e mês 1 (Janeiro)

```
INSERT INTO TABLE weather_partitioned
```

```
PARTITION (year=2008, month=1)
```

SELECT

station\_name,

dayofmonth,

precipitation,

temperature\_max,

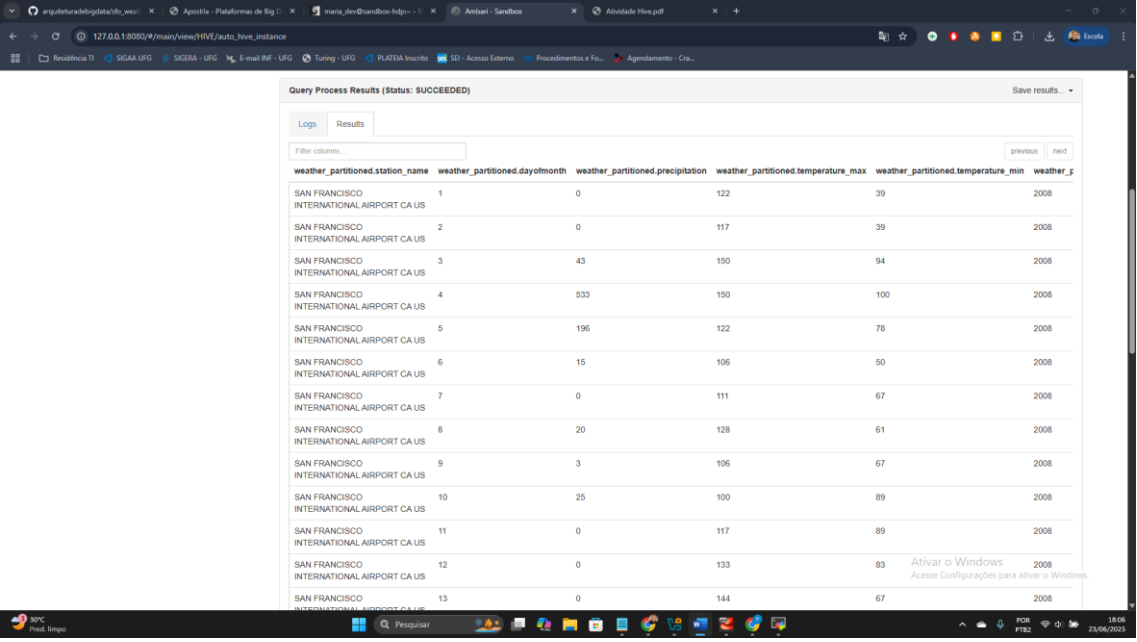
temperature\_min

FROM sfo\_weather

WHERE year = 2008 AND month = 1;

-- Consultar todos os dados da tabela particionada

SELECT \* FROM weather\_partitioned;



The screenshot shows a web browser window with a query result table. The table is titled "Query Process Results (Status: SUCCEEDED)" and has columns: weather\_partitioned.station\_name, weather\_partitioned.dayofmonth, weather\_partitioned.precipitation, weather\_partitioned.temperature\_max, weather\_partitioned.temperature\_min, and weather\_s. The table contains 13 rows of data for San Francisco International Airport in January 2008. The data is as follows:

weather_partitioned.station_name	weather_partitioned.dayofmonth	weather_partitioned.precipitation	weather_partitioned.temperature_max	weather_partitioned.temperature_min	weather_s
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	1	0	122	39	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	2	0	117	39	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	3	43	150	94	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	4	533	150	100	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	5	196	122	78	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	6	15	106	50	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	7	0	111	67	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	8	20	128	61	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	9	3	106	67	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	10	26	100	89	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	11	0	117	89	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	12	0	133	83	2008
SAN FRANCISCO INTERNATIONAL AIRPORT CA US	13	0	144	67	2008

16	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	0	120	44	2008
17	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	0	139	61	2008
18	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	0	150	33	2008
19	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	0	122	39	2008
20	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	0	111	72	2008
21	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	152	83	61	2008
22	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	25	89	44	2008
23	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	15	83	61	2008
24	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	76	78	50	2008
25	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	645	117	72	2008
26	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	58	144	94	2008
27	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	81	133	72	2008
28	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	38	100	50	2008
29	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	20	100	33	2008
30	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	28	117	50	2008
31	SAN FRANCISCO INTERNATIONAL AIRPORT CA US	13	117	72	2008

Ativar o Windows  
Acesse Configurações para ativar o Windows.