

**NOM : ACHIR**  
**Prénom : Ounissa**

1/Première partie :

```
hdfs dfs -mkdir /TPhive
```

```
hdfs dfs -cp /home/cloudera/Desktop/data/visits.txt /Tphive/visits.txt
```

```
create database if not exists TPhive;
```

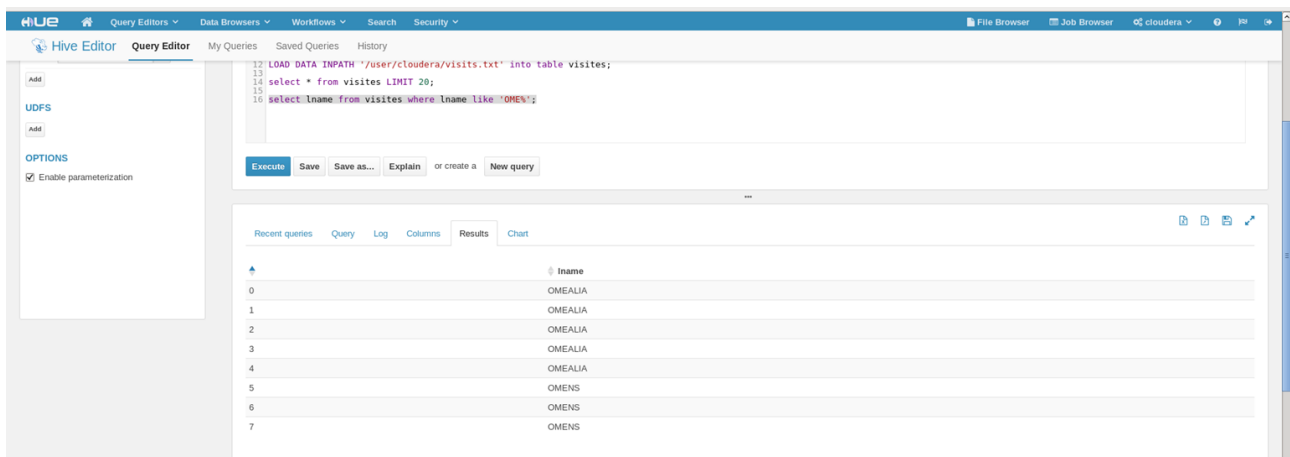
```
use TPhive;
```

```
create table visites  
( lname string,  
  fname string,  
  time_of_arrival timestamp,  
  appt_scheduled_time timestamp,  
  meeting_location string,  
  info_comment string  
)ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';
```

```
load data inpath '/Tphive/visits.txt' into table visites;
```

```
select * from visites limit 20;
```

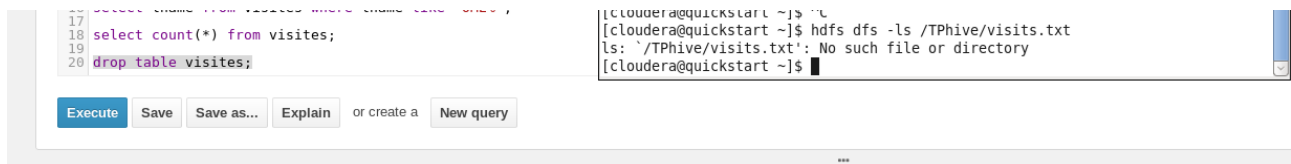
```
select lname from visites where lname like 'OME%';
```



```
select count(*) from visites;
```

```
drop table visites;
```

```
hdfs dfs -ls /Tphive/visits.txt
```



```
hdfs dfs -mkdir /demohive;
```

```
create external table visites
```

```
( lname string,
```

```
  fname string,
```

```
  time_of_arrival string,
```

```
  appt_scheduled_time string,
```

```
  meeting_location string,
```

```
  info_comment string
```

```
)ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
```

```
location '/demohive';
```

```
load data inpath '/TPhive/visits.txt' into table visites;
```

```
hdfs dfs -ls /TPhive/ => vide
```

hdfs dfs -ls /demohive/ => visits.txt existe dans cette directory puisque on a spécifié cette directory lors de la création de la table.

2/Deuxième partie :

```
create external table visites
```

```
( lname string,
```

```
  fname string,
```

```
  time_of_arrival string,
```

```
  appt_scheduled_time string,
```

```
  meeting_location string,
```

```
  info_comment string
```

```
)ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' ;
```

```
load data inpath '/demohive/visits.txt' into table visites;
```

```
select * from visites where time_of_arrival=" limit 20 ;
```

```
select * from visites where time_of_arrival='' limit 20;
```

Execute Save Save as... Explain or create a New query

Recent queries Query Log Columns Results Chart

visites.lname	visites.fname	visites.time_of_arrival	visites.appt_scheduled_time	visites.meeting_location	visites.info_comment
LANIER	JAZMIN		10/13/2010 13:00	WH	BILL SIGNING/
MOUZON	LARINA		10/13/2010 13:00	WH	BILL SIGNING/
SEIBERT	PAMELA		10/13/2010 13:00	WH	BILL SIGNING/
SEIBERT	SABRINA		10/13/2010 13:00	WH	BILL SIGNING/
GARVER	LORI		10/28/2010 14:10	WH	CHILEAN MINER RESCUERS
NEUBAUER	JOSEPH		10/28/2010 14:10	WH	CHILEAN MINER RESCUERS
BARTUSKA	ANN		10/5/2010 11:00	WH	COMUNNITY COLLEGE EVENT **EAST ROOM**

```
select * from visites limit 10;
```

```
select info_comment, count(info_comment) as nb
from visites
group by info_comment
order by nb desc;
```

The screenshot shows the Hue interface with a Hive query editor. The query editor contains the following SQL script:

```
20 drop table visites;
21 select * from visites where time_of_arrival='' limit 20;
22
23 select * from visites limit 10;
24
25 select info_comment , count(info_comment) as nb from visites group by info_comment order by nb desc;
```

The results pane shows the following data:

info_comment	nb
	18072
HOLIDAY BALL ATTENDEES/	2506
WHO EOP RECEP 2	1788
WHO EOP 1 RECEPTION/	1400
RESIDENCE STAFF HOLIDAY RECEPTION/	1202
PRESS RECEPTION ONE (1)/	1172
GENERAL RECEPTION 1	1160
HANUKKAH RECEPTION/	1080
GEN RECEP 5/	1080
GENERAL RECEPTION 3	1032
GEN RECEP 6/	996

```
select info_comment, count(info_comment) as nb
from visites
where info_comment != ''
group by info_comment
order by nb desc ;
```

The screenshot shows the Hue Hive Editor interface. The top navigation bar includes links for Cloudera, Hue, Hadoop, HBase, Impala, Spark, Solr, Oozie, Cloudera Manager, and Getting Started. The main area is divided into a left sidebar with 'SETTINGS', 'FILE RESOURCES', 'UDFS', and 'OPTIONS', and a central workspace. The workspace contains a HiveQL query and its execution results.

```

10 meeting_location string,
11 info_comment string
12 } row format delimited fields terminated by '\t';
13 LOAD DATA INPATH '/user/cloudera/visits.txt' into table visites;
14
15 select * from visites LIMIT 20;
16
17 select lname from visites where lname like 'OMEN';
18
19 select count(*) from visites;
20
21 drop table visites;
22
23 select * from visites where time_of_arrival='' limit 20;
24
25
26 select * from visites limit 10;
27
28 select info_comment , count(info_comment) as nb from visites where info_comment !='' group by info_comment order by nb desc;
  
```

Below the query editor, the 'Results' tab is active, displaying a table with two columns: 'info\_comment' and 'nb'.

info_comment	nb
HOLIDAY BALL ATTENDEES/	2506
WHO EOP RECEP 2	1788
WHO EOP 1 RECEPTION/	1400
RESIDENCE STAFF HOLIDAY RECEPTION/	1202
PRESS RECEPTION ONE (1/	1172

```

select info_comment, count(info_comment) as nb
from visites
group by info_comment
order by nb asc;
  
```

The screenshot shows the Hue Hive Editor interface with a different query executed. The query editor displays the following HiveQL:

```

20 drop table visites;
21
22 select * from visites where time_of_arrival='' limit 20;
23
24
25
26 select * from visites limit 10;
27
28 select info_comment , count(info_comment) as nb from visites group by info_comment order by nb asc;
  
```

The 'Results' tab shows a table with two columns: 'info\_comment' and 'nb'.

info_comment	nb
merged to u59031	2
WHO EOP/	2
WHO EOP RECLEAR	2
WAITING FOR SUPERMAN VISIT	2
ST. PATRICK'S RECEPTION GUESTS	2
SCIENCE FAIR	2
RES PARTY/	2
PRIVATE MEETING	2
PRIVATE LUNCH	2
POTUS PHOTO W/ US ATTORNEYS/	2
POLITICAL RECORDINGS	2

```

select info_comment, count(*) as nb_visite
from visites
where info_comment like 'GENERAL RECEPTION%'
or info_comment like 'GENERAL RECEP%'
group by info_comment;
  
```

The screenshot shows the Hue Hive Editor interface. The query editor contains the following SQL code:

```

25 select * from visites limit 10;
26
27 select info_comment , count(info_comment) as nb from visites group by info_comment order by nb asc;
28
29 select info_comment, count(*) as nb_visite from visites where info_comment like 'GENERAL RECEPTION%' or info_comment like 'GENERAL RECEP%' group by info_comment;
30
31
32
33

```

The results tab displays a table with two columns: `info_comment` and `nb_visite`. The data is as follows:

info_comment	nb_visite
GENERAL RECEP 5	16
GENERAL RECEPTION 1	26
GENERAL RECEPTION 1	1160
GENERAL RECEPTION 1 /	2
GENERAL RECEPTION 1/	2
GENERAL RECEPTION 2	62
GENERAL RECEPTION 2/	4
GENERAL RECEPTION 3	46
GENERAL RECEPTION 3	1032
GENERAL RECEPTION 5	40
GENERAL RECEPTION 6	40

```

select lname, fname, count(*) as nb
from visites
where meeting_location='WH'
group by lname, fname
order by nb desc
limit 1;

```

The screenshot shows the Hue Hive Editor interface. The query editor contains the following SQL code:

```

24 select * from visites where time_of_arrival="" limit 20;
25
26 select * from visites limit 10;
27
28 select info_comment , count(info_comment) as nb from visites group by info_comment order by nb asc;
29
30 select info_comment, count(*) as nb_visite from visites where info_comment like 'GENERAL RECEPTION%' or info_comment like 'GENERAL RECEP%' group by info_comment;
31
32 select lname,fname, count(*) as nb
33 from visites
34 where meeting_location='WH'
35 group by lname,fname
36 order by nb desc
37 limit 1;
38
39
40
41
42

```

The results tab displays a table with three columns: `lname`, `fname`, and `nb`. The data is as follows:

lname	fname	nb
PRATHER	ALAN	32