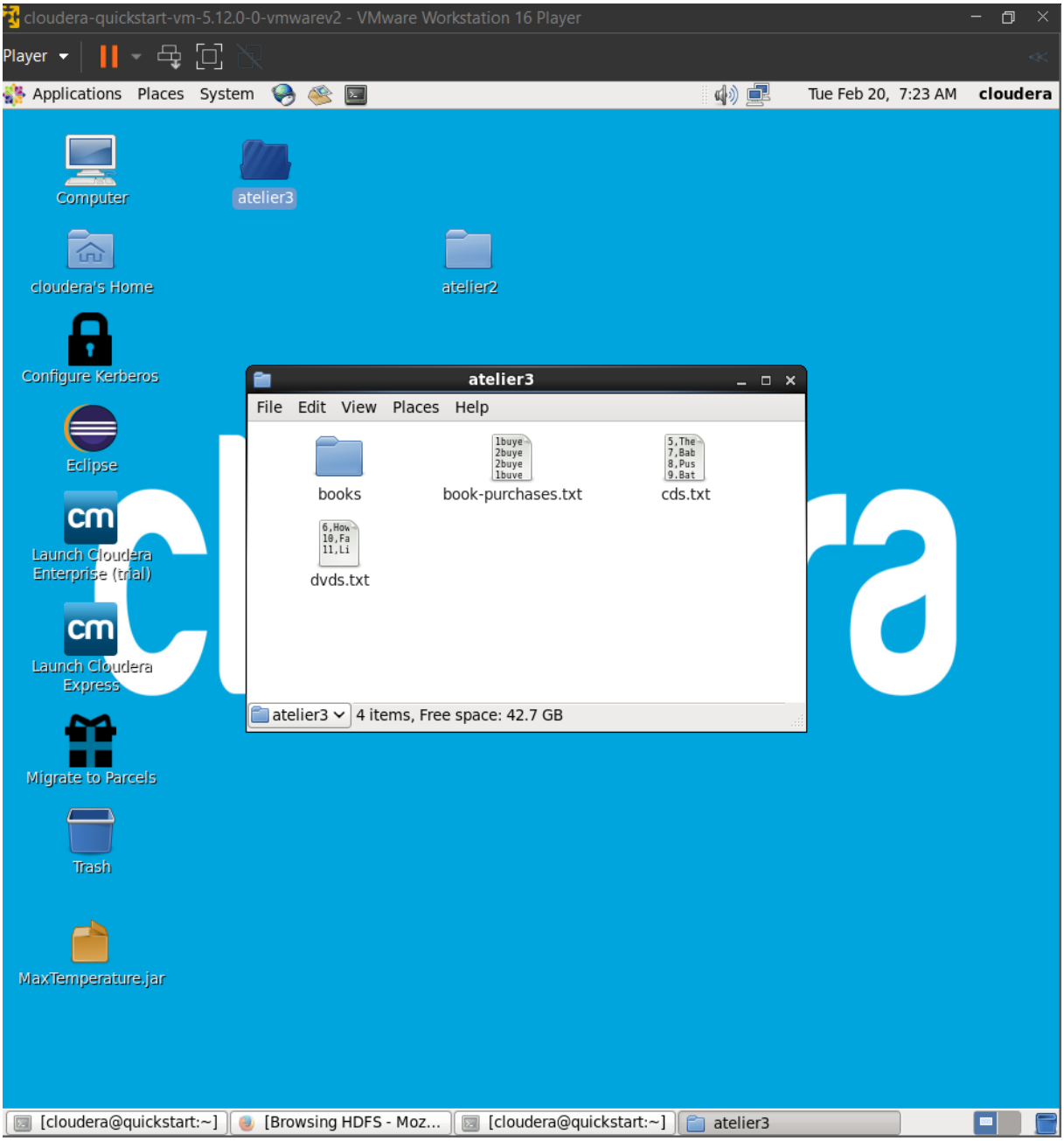
Rendu Hive

Réalisé par :Smida Molka
4éme ERP-BI9

Excercice1

1) On a créer le dossier books contenant le fichier text books.txt et ensuite nous avons créer la table livres




```
hive> create table livres (id int, title string, publishdate string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
```

2) Nous avons ensuite importé les données a partir du fichier /atelier3/books/books.txt qui sont stockés localement dans la table livres

```
hive> load data local inpath '/home/cloudera/Desktop/atelier3/books/books.txt'
    > overwrite into table livres;
Loading data to table test.livres
Table test.livres stats: [numFiles=1, numRows=0, totalSize=130, rawDataSize=0]
OK
Time taken: 0.531 seconds
```

3) Affichage des enregistrements dont l'id est égale à 2

```
Time taken: 0.531 seconds
hive> select * from livres where id=2;
OK
2       HBase: The Definitive Guide     2011/09
Time taken: 0.194 seconds, Fetched: 1 row(s)
```

4) Nous avons créer une table externe ventee sous '/user/cloudera/atelier3' :

```
hive> create external table ventee (id int, buyer string , purchaseDate string)
    > row format delimited
    > fields terminated by '\t'
    > stored as textfile
    > location '/user/cloudera/atelier3';
OK
Time taken: 0.112 seconds
```

5)-Nous avons afficher 5 lignes de la table ventee

```
hive> select * from ventee
    > limit 5;
OK
1       buyer1  2012/01/20
2       buyer2  2010/02/15
2       buyer1  2012/03/01
1       buyer3  2009/06/09
3       buyer4  2012/01/15
Time taken: 0.067 seconds, Fetched: 5 row(s)
```

6)-Nous avons créer la table vente_livre :

```
hive> create table vente_livre( id int , title string , buyer string ,purchaseDa
te string);
OK
Time taken: 0.076 seconds
```

7)- Nous avons fait une jointure entre les table livre et vente sauf que ça nous a affiché une erreur car chaque requête SQL est considérée comme étant un Job MapReduce , et au niveau des jointures on utilise le Job du MapReduce et le MapSide pour éviter la traduction et le gaspillage de mémoire on a défini le hive.auto.convert.join à une valeur false pour qu'il désactive la traduction et pour qu'on puisse effectuer la jointure des deux tables

```
hive> insert overwrite table livre_vente select b.int , b.title , p.buyer , p.purchaseDate
    > from livres b JOIN ventee p ON (b.id=p.id);
FAILED: SemanticException [Error 10001]: Line 1:23 Table not found 'livre_vente'
hive> SET hive.auto.convert.join=false;
hive> insert overwrite table livre_vente select b.int , b.title , p.buyer , p.purchaseDate
    > SET hive.auto.convert.join=false;
FAILED: ParseException line 2:4 missing EOF at 'hive' near 'SET'
hive> insert overwrite table livre_vente select b.int , b.title , p.buyer , p.purchaseDate
    > from livres b JOIN ventee p ON (b.id=p.id);
FAILED: SemanticException [Error 10001]: Line 1:23 Table not found 'livre_vente'
hive> insert overwrite table vente_livre select b.int , b.title , p.buyer , p.purchaseDate
    > from livres b JOIN ventee p ON (b.id=p.id);
FAILED: SemanticException Line 0:-1 Invalid column reference 'int'
hive> insert overwrite table vente_livre select b.id , b.title , p.buyer , p.purchaseDate
    > from livres b JOIN ventee p ON (b.id=p.id);
Query ID = cloudera_20240220073131_5e8cce63-630c-435f-91c4-ee4cef8c7d56
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1708433344543_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1708433344543
_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1708433344543_0002
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2024-02-20 07:31:29,751 Stage-1 map = 0%,  reduce = 0%
2024-02-20 07:31:55,652 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.66 sec
2024-02-20 07:32:07,909 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.69 sec
MapReduce Total cumulative CPU time: 7 seconds 690 msec
Ended Job = job_1708433344543_0002
Loading data to table test.vente_livre
Table test.vente_livre stats: [numFiles=1, numRows=13, totalSize=571, rawDataSize=558]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1   Cumulative CPU: 7.69 sec   HDFS Read: 13738 HDFS Write: 644 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 690 msec
OK
Time taken: 57.886 seconds
```

8)- L'affichage des 10 enregistrements de la table vente_livre se fait par :

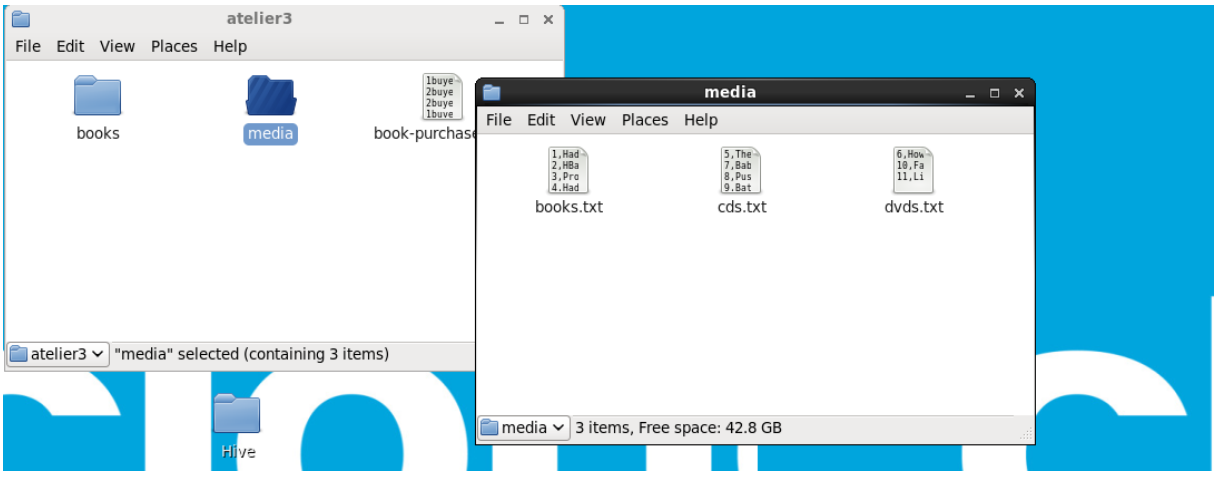
Select * from vente_livre limit 10 ;

9)-(Dû aux erreurs que j'ai commise précédemment j'ai eu l'affichage des tables « érronés que j'ai crée)

```
hive> drop table ventee
    > ;
OK
Time taken: 0.776 seconds
hive> drop table livres;
OK
Time taken: 0.151 seconds
hive> show tables
    > ;
OK
postpartion
vente
vente_livre
vente_livres
ventes
Time taken: 0.019 seconds, Fetched: 5 row(s)
hive> drop table vente_livre;
OK
Time taken: 0.193 seconds
hive> show tables;
OK
postpartion
vente
vente_livres
ventes
Time taken: 0.018 seconds, Fetched: 4 row(s)
hive> drop table vente_livres;
OK
Time taken: 0.189 seconds
hive> show tables;
OK
postpartion
vente
ventes
```

Excercice2 :

Dans un premier temps nous avons créer le dossier media où sont contenus les fichiers text comme c'est indiqué au niveau de l'énnoncé :



On a lancé la commande hive puis on a utilisé notre base de données créer précédemment qui est le test ensuite nous avons crée une table books et nous l'avons « Load »au niveau de notre base de données local et on a répété les mêmes

étapes pour les tables cds, dvds

```
hive> create table book(id int, title string, releaseDate string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTa
sk. AlreadyExistsException(message:Table book already exists)
hive> create table books(id int, title string, releaseDate string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
Time taken: 0.162 seconds
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/books.txt'
    > overwrite into table book;
Loading data to table test.book
Table test.book stats: [numFiles=1, numRows=0, totalSize=130, rawDataSize=0]
OK
Time taken: 1.812 seconds
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/books.txt'
    > overwrite into table books;
Loading data to table test.books
Table test.books stats: [numFiles=1, numRows=0, totalSize=130, rawDataSize=0]
OK
Time taken: 1.049 seconds
```

Pour la table cds :

```
hive> create table cds(id int, title string, releaseDate string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
Time taken: 0.178 seconds
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/cds.txt'
    > overwrite into table cds;
Loading data to table test.cds
Table test.cds stats: [numFiles=1, numRows=0, totalSize=85, rawDataSize=0]
OK
```

Pour la table dvds :

```
hive> create table dvds(id int, title string, releaseDate string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
Time taken: 0.21 seconds
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/dvds.txt'
    > overwrite into table dvds;
Loading data to table test.dvds
Table test.dvds stats: [numFiles=1, numRows=0, totalSize=123, rawDataSize=0]
OK
Time taken: 0.966 seconds
```

Ensuite nous avons créer la table partitionnée qui est la table média :

```
hive> create table media (id int, title string, releaseDate string)
    > partitioned by (type string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
Time taken: 0.219 seconds
```

On a chargé la partition du cd :

```
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/cds.txt'
    > overwrite into table media
    > partition (type='CD');
Loading data to table test.media partition (type=CD)
Partition test.media{type=CD} stats: [numFiles=1, numRows=0, totalSize=85, rawDa
taSize=0]
OK
Time taken: 1.793 seconds
```

Ensuite celle du dvd :

```
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/dvds.txt'
    > overwrite into table media
    > partition (type='DVD');
Loading data to table test.media partition (type=DVD)
Partition test.media{type=DVD} stats: [numFiles=1, numRows=0, totalSize=123, raw
DataSize=0]
OK
Time taken: 1.132 seconds
```

Et finalement celle des books :

```
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/books.txt'
    > overwrite into table media
    > partition (type='BOOK');
Loading data to table test.media partition (type=BOOK)
Partition test.media{type=BOOK} stats: [numFiles=1, numRows=0, totalSize=130, ra
wDataSize=0]
OK
Time taken: 1.16 seconds
```

1)-Pour afficher les enregistrements de type cd :

```
hive> select * from media where type='CD';
OK
5       The 2nd Law     2012/07 CD
7       Babel   2012/08 CD
8       Push And Shove  2012/09 CD
9       Battle Born     2012/09 CD
Time taken: 1.927 seconds, Fetched: 4 row(s)
```

2)-Pour afficher les enregistrements de type dvd :

```
hive> select * from media where type='DVD';
OK
6       How I Met Your Mother: The Complete Seventh Season      2012/10 DVD
10      Family Guy Season 10    2011/08 DVD
11      Lie To Me Season 1      2009/01 DVD
Time taken: 0.319 seconds, Fetched: 3 row(s)
```

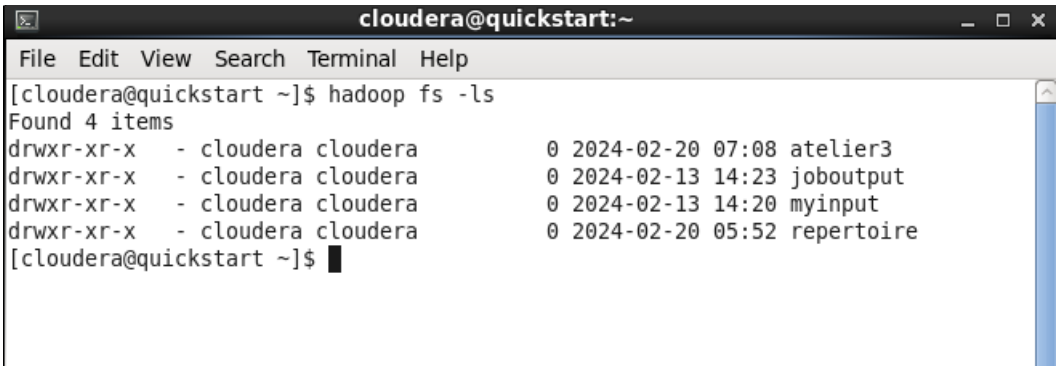
3)-Et pour afficher les enregistrements de type livres «Books » :

```
hive> select * from media where type='BOOK';
OK
1       Hadoop: The Definitive Guide    2012/06 BOOK
2       HBase: The Definitive Guide     2011/09 BOOK
3       Programming Pig 2011/10 BOOK
4       Hadoop in Action        2010/12 BOOK
Time taken: 0.254 seconds, Fetched: 4 row(s)
```

4)-Finalement pour afficher les partitions de la table media :

```
hive> show partitions media;
OK
type=BOOK
type=CD
type=DVD
Time taken: 0.206 seconds, Fetched: 3 row(s)
hive>
```

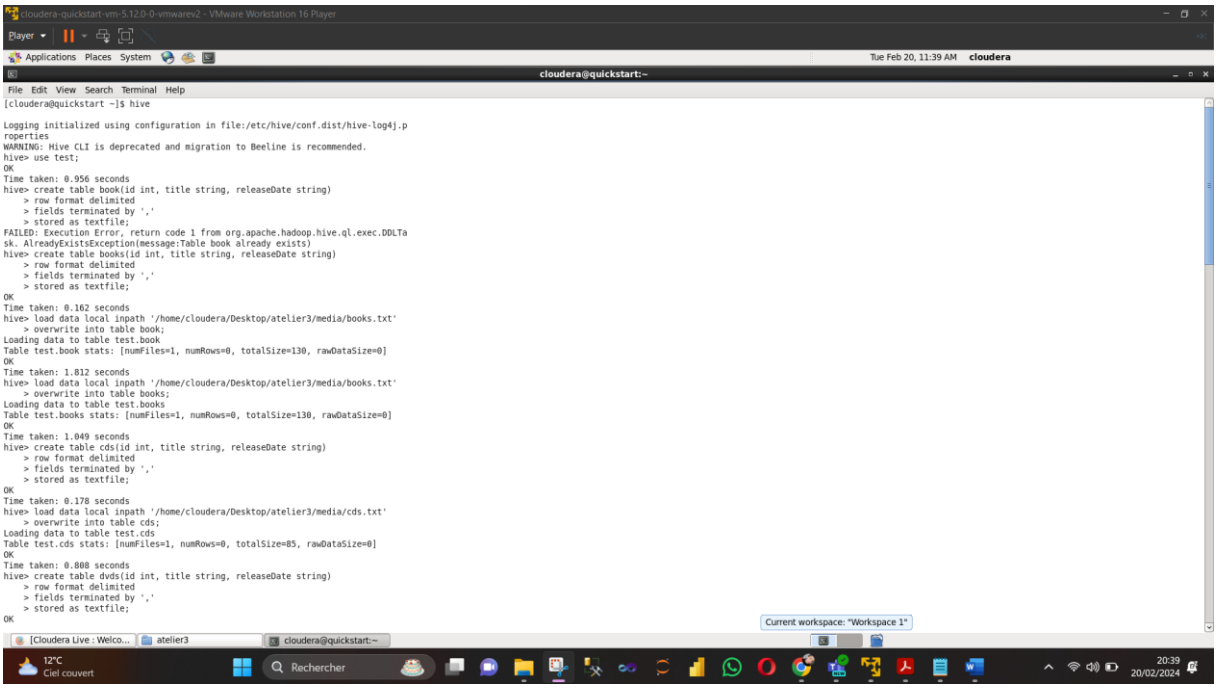
5)-Au niveau d’un autre terminal :

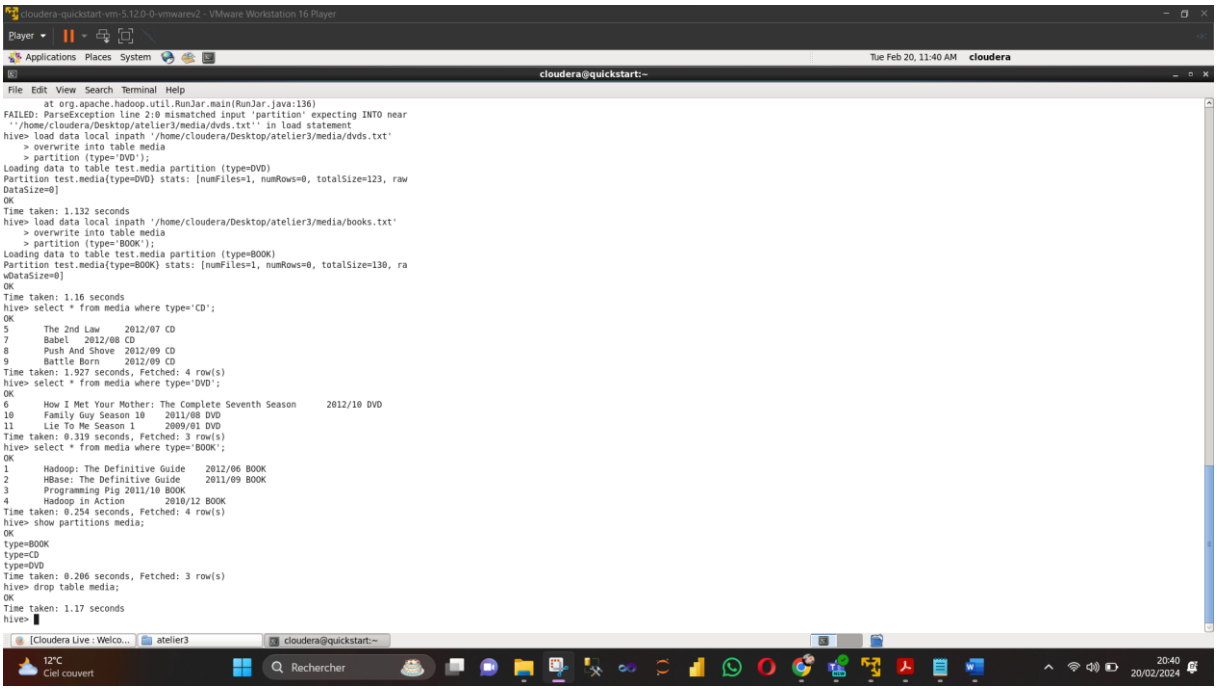


6)-Pour supprimer la table media :

```
hive> drop table media;
OK
Time taken: 1.17 seconds
hive>
```

En ci-joint, vous trouverez une capture de toute l'exécution




cloudera-quickstart-vm-5.12.0-0-vmwarev2 - VMware Workstation 16 Player
Player
Applications Places System
Tue Feb 20, 11:40 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 2:0 mismatched input 'partition' expecting INTO near
''/home/cloudera/Desktop/atelier3/media/dvds.txt'' in load statement
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/dvds.txt'
> overwrite into table media
> partition (type='DVD');
Loading data to table test.media partition (type=DVD)
Partition test.media{type=DVD} stats: [numFiles=1, numRows=0, totalSize=123, raw
DataSize=0]
OK
Time taken: 1.132 seconds
hive> load data local inpath '/home/cloudera/Desktop/atelier3/media/books.txt'
> overwrite into table media
> partition (type='BOOK');
Loading data to table test.media partition (type=BOOK)
Partition test.media{type=BOOK} stats: [numFiles=1, numRows=0, totalSize=130, ra
wDataSize=0]
OK
Time taken: 1.16 seconds
hive> select * from media where type='CD';
OK
5 The 2nd Law 2012/07 CD
7 Babel 2012/08 CD
8 Push And Shove 2012/09 CD
9 Battle Born 2012/09 CD
Time taken: 1.927 seconds, Fetched: 4 row(s)
hive> select * from media where type='DVD';
OK
6 How I Met Your Mother: The Complete Seventh Season 2012/10 DVD
10 Family Guy Season 10 2011/08 DVD
11 Lie To Me Season 1 2009/01 DVD
Time taken: 0.319 seconds, Fetched: 3 row(s)
hive> select * from media where type='BOOK';
OK
1 Hadoop: The Definitive Guide 2012/06 BOOK
2 HBase: The Definitive Guide 2011/09 BOOK
3 Programming Pig 2011/10 BOOK
4 Hadoop in Action 2010/12 BOOK
Time taken: 0.254 seconds, Fetched: 4 row(s)
hive> show partitions media;
OK
type=BOOK
type=CD
type=DVD
Time taken: 0.206 seconds, Fetched: 3 row(s)
hive> drop table media;
OK
Time taken: 1.17 seconds
hive>
[Cloudera Live : Welco...
atelier3
cloudera@quickstart:~
12°C
Ciel couvert
Rechercher
20:40
20/02/2024