scp -r -p -P 2222 nba2021_per_game.csv student@51.136.160.196:nba2021_per_game.csv

```
C:\Users\wensl\Documents\Data>scp -r -p -P 2222 nba2021_per_game.csv student@51.136.160.196:nba2021_per_game.csv
student@51.136.160.196's password:
nba2021_per_game.csv                                          100%   62KB 749.6KB/s   00:00
```

Ls afin de vérifier la présence du document

```
[student@sandbox-hdp ~]$ ls
CDR.csv  nba2021_per_game.csv
[student@sandbox-hdp ~]$
```

Uploader ce fichier de données depuis Linux de la VM vers l'espace HDFS associé à l'utilisateur

Sur votre console de la VM HDP, Taper :

Besoin de création du répertoire data/projetdata

Creation du dossier *projetdata*

hadoop fs -mkdir /user/student/data/projetdata

Upload du fichier vers l'espace HDFS

hadoop fs -put /home/student/ nba2021_per_game.csv /user/student/data/projetdata

Afficher les informations à propos du fichier de données HDFS

hadoop fs -ls data/projetdata/nba2021_per_game.csv

hadoop fs -cat /user/student/data/projetdata/nba2021_per_game.csv | head -n 10

Connexion à Hive

/usr/bin/hive shell

Création de la base de données

CREATE DATABASE db_student;

Vérification de la création de la base de données

SHOW DATABASES;

```
hive> SHOW DATABASES;
OK
db_student
default
foodmart
xademo
Time taken: 3.5 seconds, Fetched: 4 row(s)
hive>
```

Connexion a la base de données

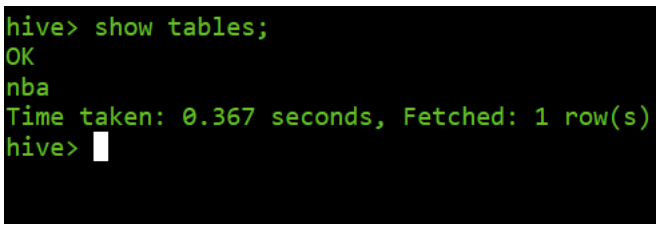
use db_student;

```
hive> use db_student;
OK
Time taken: 0.274 seconds
hive>
```

Création de la table NBA

```
CREATE TABLE NBA (
Player CHAR(50),
Pos CHAR(3),
Age STRING,
Tm VARCHAR(3),
G STRING,
GS STRING,
MP FLOAT,
FG FLOAT,
FGA FLOAT,
FG_Average FLOAT,
3P FLOAT,
3PA FLOAT,
3P_Average FLOAT,
2P FLOAT,
2PA FLOAT,
2P_Average FLOAT,
eFG_Average FLOAT,
FT FLOAT,
FTA FLOAT,
FT_Average FLOAT,
ORB FLOAT,
DRB FLOAT,
TRB FLOAT,
AST FLOAT,
STL FLOAT,
BLK FLOAT,
TOV FLOAT,
PF FLOAT,
PTS FLOAT)ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
stored as textfile;
```

Vérification de la création de la table



Vérifier le schéma de la table NBA dans Hive
describe nba;

```
hive> describe nba;
OK
player                  char(100)
pos                     char(3)
age                     string
tm                      varchar(3)
g                       string
gs                      string
mp                      float
fg                      float
fga                     float
fg_average              float
3p                      float
3pa                     float
3p_average              float
2p                      float
2pa                     float
2p_average              float
efg_average             float
ft                      float
fta                     float
ft_average              float
orb                     float
drb                     float
trb                     float
ast                     float
stl                     float
blk                     float
tov                     float
pf                      float
pts                     float
Time taken: 0.936 seconds, Fetched: 29 row(s)
```

Charger le fichier de données nba2021_per_game.csv depuis HDFS vers la table Hive NBA
LOAD DATA INPATH '/user/student/data/projetdata/ nba2021_per_game.csv ' OVERWRITE INTO TABLE nba;

Vérifier les données nba2021_per_game.csv depuis HDFS vers la table Hive NBA
select * from nba LIMIT 10;
Par la suite vous avez la main libre sur les différentes requêtes

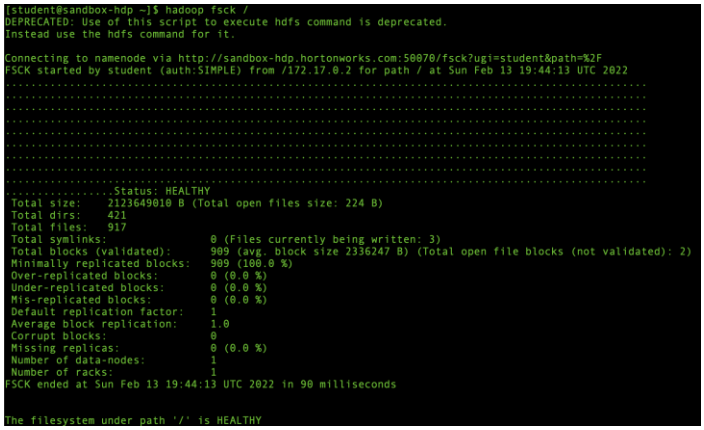
```
hive> select * from nba LIMIT 10;
OK
Player                  NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    Pos     Age     Tm      G       GS      NULL    NULL    NULL    N
Precious Achiuwa        NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    PF      21      MIA     28      2       14.6    2.6     4.4     0
.59     0.0     0.0     0.0     2.6     4.4     0.59    0.59    1.3     2.4     0.561   1.3     2.7     4.0     0.6     0.4     0.5     1.0     1.9     6.5
Jaylen Adams            NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    PG      24      MIL     6       0       2.8     0.2     1.3     0
.125    0.0     0.3     0.0     0.2     1.0     0.167   0.125   0.0     0.0     0.0     0.0     0.5     0.5     0.3     0.0     0.0     0.0     0.2     0.3
Steven Adams            NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    C       27      NOP     27      27      28.1    3.5     5.8     0
.603    0.0     0.0     0.0     3.5     5.7     0.606   0.603   1.1     2.3     0.468   4.3     4.6     8.9     2.1     1.0     0.6     1.7     1.9     8.0
Bam Adebayo             NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    C       23      MIA     26      26      33.6    7.4     12.9    0
.573    0.1     0.2     0.4     7.3     12.7    0.576   0.576   5.1     6.0     0.841   1.9     7.3     9.2     5.3     1.0     1.0     3.0     2.6     19.9
LaMarcus Aldridge       NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    C       35      SAS     18      18      26.7    5.9     12.5    0
.476    1.3     3.7     0.358   4.6     8.8     0.525   0.529   0.9     1.2     0.762   0.8     3.5     4.3     1.9     0.4     0.9     0.9     1.5     14.1
Ty-Shon Alexander       NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    SG      22      PHO     3       0       2.7     0.0     1.0     0
.0      0.0     0.3     0.0     0.0     0.7     0.0     0.0     0.0     0.0     0.0     0.0     0.3     0.3     0.3     0.0     0.0     0.0     0.3     0.0
Nickeil Alexander-Walker        NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    SG      22      NOP     23      3       19.2    3.3     8.2     0
.41     1.0     3.8     0.276   2.3     4.4     0.525   0.473   1.1     1.4     0.781   0.2     2.4     2.7     2.0     1.1     0.3     1.3     1.7     8.8
Grayson Allen           NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    SG      25      MEM     19      8       23.9    3.2     7.4     0
.429    2.3     5.3     0.436   0.8     2.1     0.41    0.586   1.7     1.9     0.892   0.4     2.5     2.9     2.1     1.0     0.2     1.1     1.3     10.4
Jarrett Allen           NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    C       22      TOT     28      10      26.2    4.4     6.8     0
.642    0.0     0.1     0.25    4.3     6.6     0.651   0.645   3.6     4.7     0.758   2.9     6.1     9.0     1.6     0.5     1.6     1.5     1.6     12.3
Time taken: 0.664 seconds, Fetched: 10 row(s)
hive>
```

```
hive> select Player, 3P, Tm
    >
    > from nba
    >
    > ORDER BY 3P DESC
    >
    > LIMIT 10;
Query ID = student_20220214214701_912baf15-cfe5-4540-81cd-a36200c087de
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1644867574857_0003)

--------------------------------------------------------------------------------
        VERTICES      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ..........   SUCCEEDED      1          1        0        0       0       0
Reducer 2 ......   SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 14.96 s
--------------------------------------------------------------------------------
OK
Stephen Curry                                                   5.0     GSW
CJ McCollum                                                     4.8     POR
Damian Lillard                                                  4.1     POR
Buddy Hield                                                     3.9     SAC
Paul George                                                     3.8     LAC
Zach LaVine                                                     3.6     CHI
Terry Rozier                                                    3.4     CHO
Donovan Mitchell                                                3.4     UTA
Malik Beasley                                                   3.4     MIN
Joe Harris                                                      3.3     BRK
```

HDFS

Exécution d'un utilitaire de vérification du système de fichiers HDFS pour savoir si le systeme de fichier est sain ou corrompus. Ici on peut voir qu'il est sain.



Player Nom du joueur Char
Pos Position Char
Age Age String
Tm Equipe Char
G Matchs joués String
GS Matchs débutés String
MP Minutes jouées Float
FG Tir réussis par match Float
FGA Tir tentés par match Float
FG% Pourcentage au tir par match Float
3P Tir à 3 points réussis par match Float
3PA Tir à 3 points tentés par match Float
3P% Pourcentage Tir à 3 points réussis par match Float
2P Tir à 2 points réussis par match Float
2PA Tir à 2 points tentés par match Float
2P% Pourcentage Tir à 2 points réussis par match Float
eFG% Effective Field Goal ➔ Pourcentage de tir total réajusté avec les 3 points Float
FT Lancé franc réussis par match Float
FTA Lancé franc tentés par match Float
FT% Pourcentage lancé franc réussis par match Float
ORB Offensive rebond par match Float
DRB Défensive rebond par match Float
TRB Total rebond par match Float
AST passe par match Float
STL Steal par match Float
BLK Blocage par match Float
TOV Perte de balle par match Float
PF Faute par match Float
PTS Points par match Float

Pas de données confidentielles utilisées lors de l'exercice. Dataset disponible sur www.kaggle.com