

External Table without Partition

Slide # 42

```
[root@sandbox-hdp ~]# hdfs dfs -mkdir /user/anup/happinessreport
[root@sandbox-hdp ~]# hdfs dfs -put /home/hdfs/2019.csv /user/anup/happinessreport
[root@sandbox-hdp ~]#
[root@sandbox-hdp ~]# hdfs dfs -chmod 777 /user/anup/happinessreport
[root@sandbox-hdp ~]# hdfs dfs -chmod 777 /user/anup/happinessreport/2019.csv
[root@sandbox-hdp ~]#
```

```
[root@sandbox-hdp ~]# hdfs dfs -ls /user/anup/happinessreport
Found 1 items
-rwxrwxrwx   1 root hdfs      8667 2020-05-23 20:11 /user/anup/happinessreport/2019.csv
```

Slide # 43

```
[root@sandbox-hdp ~]# hive
log4j:WARN No such property [maxFileSize] in org.apache.log4j.DailyRollingFileAppender.

Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.properties
hive> create external table 2019_ext(overall_rank int, country_region string, score int, GDP_per_capita int, social_support int, healthy_life_expectancy int, freedom_to_make_life_choices int, generosity int, perceptions_of_corruption int)
  > row format delimited
  > fields terminated by ','
  > stored as textfile
  > location '/user/anup/happinessreport';
OK
Time taken: 188.944 seconds
hive> show tables;
OK
2019_ext
Time taken: 0.321 seconds, Fetched: 1 row(s)
```

Slide # 45

```
Logging initialized using configuration in file:/etc/hive/2.6.5.0-292/0/hive-log4j.properties
hive> load data inpath '/user/anup/happinessreport/2019.csv'
  > into table 2019_ext;
Loading data to table default.2019_ext
Table default.2019_ext stats: [numFiles=1, totalSize=8667]
OK
Time taken: 191.079 seconds
hive>
```

Slide # 46

```
hive> describe formatted 2019_ext;
OK
# col_name          data_type          comment
overall_rank        int
country_or_region    string
score                float
gdp_per_capita       float
social_support        float
healthy_life_expectancy float
freedom_to_make_life_choices float
generosity           float
perceptions_of_corruption float

# Detailed Table Information
Database:            default
Owner:               root
CreateTime:          Sat May 23 22:48:50 UTC 2020
LastAccessTime:      UNKNOWN
Protect Mode:        None
Retention:           0
Location:            hdfs://sandbox-hdp.hortonworks.com:8020/user/anup/happinessreport
Table Type:          EXTERNAL_TABLE
Table Parameters:
    EXTERNAL          TRUE
    numFiles           2
    totalSize          17176
    transient_lastDdlTime 1590274392
```

```
hive> select * from 2019_ext limit 5;
OK
NULL    Country_or_region    NULL    NULL    NULL    NULL    NULL    NULL    NULL
1        Finland 7.769    1.34    1.587    0.986    0.596    0.153    0.393
2        Denmark 7.6        1.383    1.573    0.996    0.592    0.252    0.41
3        Norway 7.554    1.488    1.582    1.028    0.603    0.271    0.341
4        Iceland 7.494    1.38    1.624    1.026    0.591    0.354    0.118
Time taken: 0.172 seconds, Fetched: 5 row(s)
```

Calculating the mean of values column for the top one-third of countries, middle and bottom one-third.

```
hive> select avg(score) from 2019_ext where overall_rank between 1 and 52;
Query ID = root_20200523234511_3831lead7-d46b-4cab-bbcf-c517a500886e
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_1590192907351_0022)
```

	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: 61.20 s
OK
6.644807678002578
Time taken: 80.564 seconds, Fetched: 1 row(s)
```

```
hive> select avg(score) from 2019_ext where overall_rank between 53 and 104;
Query ID = root_20200523234730_63bf4652-e60e-43d5-b591-e6086d7cf632
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1590192907351_0022)
```

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: 9.01 s
OK
5.406288467920744
Time taken: 10.208 seconds, Fetched: 1 row(s)
```

```
hive> select avg(score) from 2019_ext where overall_rank between 105 and 156;
Query ID = root_20200523234915_44b1129a-945d-4e52-b3cf-7ad40777f018
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1590192907351_0022)
```

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: 5.97 s
OK
4.170192278348482
Time taken: 6.961 seconds, Fetched: 1 row(s)
```

Rank of Country	Average Happiness Score
1-52	6.64
53-104	5.41
105-156	4.17

```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> CREATE EXTERNAL TABLE IF NOT EXISTS mozilla4_external_part('Key' int,'id' int,'start' int,'end' int,'event' int,'size' int,'state' int)
.....> COMMENT 'Mozilla class sizes'
.....> ROW FORMAT DELIMITED
.....> FIELDS TERMINATED BY ','
.....> STORED AS TEXTFILE
.....> LOCATION '/user/wangjh/part';
... ..
```