

## Course Project - ETL & Batch Processing

### Problem Statement

*Ingest the India Annual Health Survey (AHS) 2012-13 data hosted on Amazon RDS into Hadoop correctly and process it to generate the following analyses:*

#### Analyses

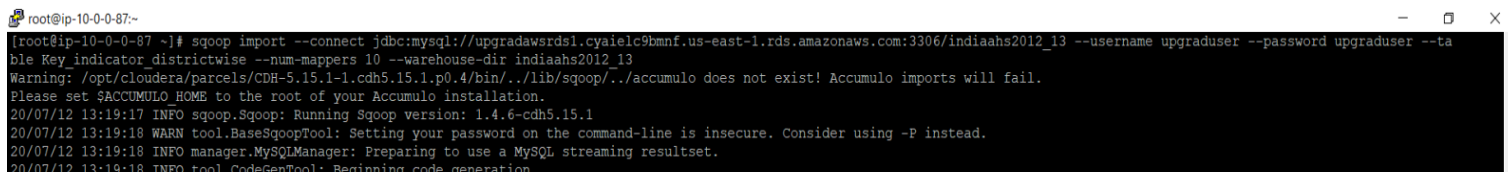
- *The child mortality rate in Uttar Pradesh*
- *The fertility rate in Bihar*
- *State-wise child mortality rate and state-wise fertility rate and does high fertility correlate with high child mortality?*
- *Find top 2 districts per state with the highest population per household*
- *Find top 2 districts per state with the lowest sex ratios*

*Such analyses would help in vivid understanding and timely monitoring of different determinants on well-being and health of population particularly Child and Reproductive Health. Based on the analyses, one can also compare India's position in Global HDI and can suggest ways that can improve it.*

### I. Data Ingestion from the RDS to HDFS using Sqoop

#### 1. Sqoop Import command

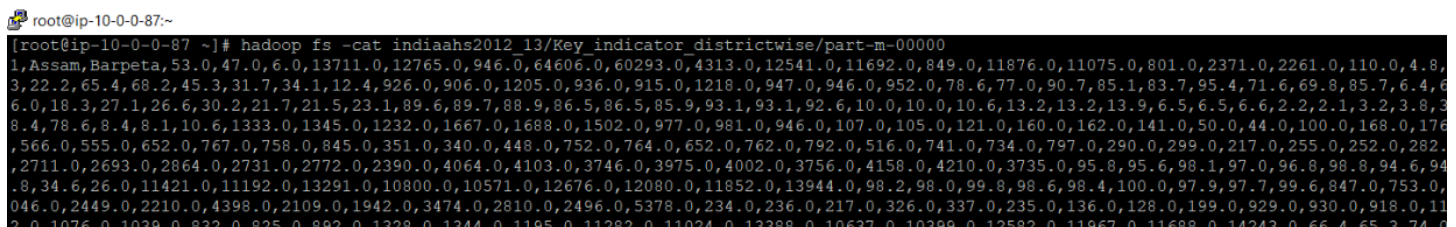
```
sqoop import --connect jdbc:mysql://upgradawsrds1.cyaieic9bmnf.us-east-1.rds.amazonaws.com:3306/indiaahs2012_13 --username upgraduser --password upgraduser --table Key_indicator_districtwise --num-mappers 10 --warehouse-dir indiaahs2012_13
```



```
root@ip-10-0-0-87:~# sqoop import --connect jdbc:mysql://upgradawsrds1.cyaieic9bmnf.us-east-1.rds.amazonaws.com:3306/indiaahs2012_13 --username upgraduser --password upgraduser --table Key_indicator_districtwise --num-mappers 10 --warehouse-dir indiaahs2012_13
Warning: /opt/cloudera/parcels/CDH-5.15.1-1.cdh5.15.1.p0.4/bin/./lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
20/07/12 13:19:17 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.15.1
20/07/12 13:19:18 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
20/07/12 13:19:18 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
20/07/12 13:19:18 INFO tool.CodeGenTool: Beginning code generation
```

#### 2. Command to see the list of imported data

```
hadoop fs -cat indiaahs2012_13/Key_indicator_districtwise/part-m-00000
```



```
root@ip-10-0-0-87:~# hadoop fs -cat indiaahs2012_13/Key_indicator_districtwise/part-m-00000
1,Assam,Barpeta,53.0,47.0,6.0,13711.0,12765.0,946.0,64606.0,60293.0,4313.0,12541.0,11692.0,849.0,11876.0,11075.0,801.0,2371.0,2261.0,110.0,4.8,
3,22.2,65.4,68.2,45.3,31.7,34.1,12.4,926.0,906.0,1205.0,936.0,915.0,1218.0,947.0,946.0,952.0,78.6,77.0,90.7,85.1,83.7,95.4,71.6,69.8,85.7,6.4,6
6.0,18.3,27.1,26.6,30.2,21.7,21.5,23.1,89.6,89.7,88.9,86.5,86.5,85.9,93.1,93.1,92.6,10.0,10.0,10.6,13.2,13.2,13.9,6.5,6.5,6.6,2.2,2.1,3.2,3.8,3
8.4,78.6,8.4,8.1,10.6,1333.0,1345.0,1232.0,1667.0,1688.0,1502.0,977.0,981.0,946.0,107.0,105.0,121.0,160.0,162.0,141.0,50.0,44.0,100.0,168.0,176
,566.0,555.0,652.0,767.0,758.0,845.0,351.0,340.0,448.0,752.0,764.0,652.0,762.0,792.0,516.0,741.0,734.0,797.0,290.0,299.0,217.0,255.0,252.0,282.
,2711.0,2693.0,2864.0,2731.0,2772.0,2390.0,4064.0,4103.0,3746.0,3975.0,4002.0,3756.0,4158.0,4210.0,3735.0,95.8,95.6,98.1,97.0,96.8,98.8,94.6,94
.8,34.6,26.0,11421.0,11192.0,13291.0,10800.0,10571.0,12676.0,12080.0,11852.0,13944.0,98.2,98.0,99.8,98.6,98.4,100.0,97.9,97.7,99.6,847.0,753.0,
046.0,2449.0,2210.0,4398.0,2109.0,1942.0,3474.0,2810.0,2496.0,5378.0,234.0,236.0,217.0,326.0,337.0,235.0,136.0,128.0,199.0,929.0,930.0,918.0,11
2.0,1076.0,1039.0,832.0,825.0,892.0,1328.0,1344.0,1195.0,11282.0,11024.0,13388.0,10637.0,10399.0,12582.0,11967.0,11688.0,14243.0,66.4,65.3,74.0
```

## II. External table creation in Hive and loading the ingested data into it. Data ingestion verification.

### 1. Command to create the external table

```
create external table Key_Indicator_ext_Full (  
  > State_Name string,  
  > State_District_Name string,  
  > YY_Under_Five_Mortality_Rate_U5MR_Total_Person int,  
  > LL_Total_Fertility_Rate_Total int,  
  > AA_Population_Total int,  
  > AA_Households_Total int,  
  > CC_Sex_Ratio_All_Ages_Total int)  
  > row format delimited  
  > fields terminated by ','  
  > location '/key_indicator';
```

*[ An external table Key\_Indicator\_ext\_Full is created with selected columns inorder to render the required data]*

### 2. Command to load the ingested data into the external table

Load data inpath '/user/root/indiaahs2012\_13/Key\_indicator\_districtwise' into table Key\_Indicator\_ext\_Full;

*[The Ingested data is loaded into external table created above]*

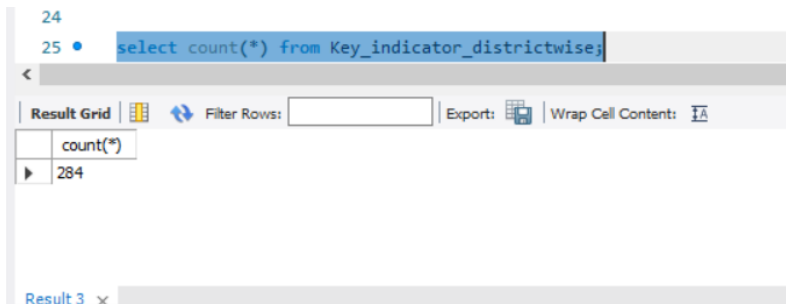
### 3. Queries to verify that the ingestion is correctly accomplished

a. Query to count the total number of rows along with the screenshots of the data fetched by the query on MySQL Workbench and Hue

*Query:*

```
Select count(*) from Key_indicator_districtwise;
```

MySQL Workbench:



*[Total row count is 284]*

Hue:



[Query run in hue and getting the row count as 284 ]

b. Query to select the top 10 rows and first 8 columns along with the screenshots of the data fetched by the query on MySQL Workbench and Hue

Query:

```
select State_Name,State_District_Name,AA_Sample_Units_Total,AA_Sample_Units_Rural,
AA_Sample_Units_Urban,AA_Households_Total,AA_Households_Rural,AA_Households_Urban
from Key_indicator_districtwise limit 10;
```

MySQL Workbench:

The screenshot shows the MySQL Workbench interface. The query editor at the top displays the SQL statement: `select State_Name,State_District_Name,AA_Sample_Units_Total,AA_Sample_Units_Rural,AA_Sample_Units_Urban,AA_Households_Total,AA_Households_Rural,AA_Households_Urban from Key_indicator_districtwise limit 10;`. Below the editor, the 'Result Grid' tab is active, showing the results of the query in a table with 8 columns and 10 rows. The columns are: State\_Name, State\_District\_Name, AA\_Sample\_Units\_Total, AA\_Sample\_Units\_Rural, AA\_Sample\_Units\_Urban, AA\_Households\_Total, AA\_Households\_Rural, and AA\_Households\_Urban. The rows represent data for various districts in Assam.

|   | State_Name | State_District_Name | AA_Sample_Units_Total | AA_Sample_Units_Rural | AA_Sample_Units_Urban | AA_Households_Total | AA_Households_Rural | AA_Households_Urban |
|---|------------|---------------------|-----------------------|-----------------------|-----------------------|---------------------|---------------------|---------------------|
| ▶ | Assam      | Barpeta             | 53                    | 47                    | 6                     | 13711               | 12755               | 946                 |
|   | Assam      | Bongaigaon          | 89                    | 73                    | 16                    | 17384               | 14904               | 2480                |
|   | Assam      | Cachar              | 105                   | 84                    | 21                    | 27488               | 24207               | 3281                |
|   | Assam      | Darrang             | 26                    | 24                    | 2                     | 5951                | 5769                | 182                 |
|   | Assam      | Dhemaj              | 121                   | 108                   | 13                    | 14481               | 12619               | 1862                |
|   | Assam      | Dhubri              | 42                    | 35                    | 7                     | 11001               | 9954                | 1047                |
|   | Assam      | Dibrugarh           | 91                    | 66                    | 25                    | 21378               | 16514               | 4864                |
|   | Assam      | Goalpara            | 64                    | 56                    | 8                     | 15891               | 14630               | 1261                |
|   | Assam      | Golaghat            | 70                    | 61                    | 9                     | 16021               | 14183               | 1838                |
|   | Assam      | Hailakandi          | 10                    | 8                     | 2                     | 2802                | 2381                | 421                 |

Hue:

```
SELECT State_Name,State_District_Name,AA_Sample_Units_Total,AA_Sample_Units_Rural,
AA_Sample_Units_Urban,AA_Households_Total,AA_Households_Rural,AA_Households_Urban
FROM Key_Indicator_ext_Full LIMIT 10;
```

| state_name | state_district_name | aa_sample_units_total | aa_sample_units_rural | aa_sample_units_urban | aa_households_total | aa_households_rural | aa_households_urban |
|------------|---------------------|-----------------------|-----------------------|-----------------------|---------------------|---------------------|---------------------|
| Assam      | Barpeta             | 53                    | 47                    | 6                     | 13711               | 12765               | 946                 |
| Assam      | Bongaigaon          | 89                    | 73                    | 16                    | 17384               | 14904               | 2480                |
| Assam      | Cachar              | 105                   | 84                    | 21                    | 27488               | 24207               | 3281                |
| Assam      | Darrang             | 26                    | 24                    | 2                     | 5951                | 5769                | 182                 |
| Assam      | Dhemaji             | 121                   | 108                   | 13                    | 14481               | 12619               | 1862                |
| Assam      | Dhubri              | 42                    | 35                    | 7                     | 11001               | 9954                | 1047                |
| Assam      | Dibrugarh           | 91                    | 66                    | 25                    | 21378               | 16514               | 4864                |
| Assam      | Goalpara            | 64                    | 56                    | 8                     | 15891               | 14630               | 1261                |
| Assam      | Golaghat            | 70                    | 61                    | 9                     | 16021               | 14183               | 1838                |
| Assam      | Hailakandi          | 10                    | 8                     | 2                     | 2802                | 2381                | 421                 |

*[Verified the outputs from MySQL Workbench and Hue to be same]*

### III. Subset schema creation in Hive to support the analyses

#### 1. Columns used in the subset schema

- >ID
- >State\_Name
- > State\_District\_Name
- > YY\_Under\_Five\_Mortality\_Rate\_U5MR\_Total\_Person
- > LL\_Total\_Fertility\_Rate\_Total
- > AA\_Population\_Total
- > AA\_Households\_Total
- > CC\_Sex\_Ratio\_All\_Ages\_Total

#### 2. Storage format used

- Default
- ORC

#### 3. Create and insert command for the default format

##### Create command

```
CREATE EXTERNAL TABLE IF NOT EXISTS Key_Indicator_ext_default(  
    ID int,  
    State_Name string,  
    State_District_Name string,  
    YY_Under_Five_Mortality_Rate_U5MR_Total_Person double,  
    LL_Total_Fertility_Rate_Total double,  
    AA_Households_Total double,  
    AA_Population_Total double,  
    CC_Sex_Ratio_All_Ages_Total double  
)  
  
LOCATION '/user/root/indiaahs2012_13/Key_indicator_districtwise';
```

### **Insert command**

```
INSERT INTO Key_indicator_districtwise
SELECT ID,State_Name,State_District_Name,
       YY_Under_Five_Mortality_Rate_U5MR_Total_Person,
       LL_Total_Fertility_Rate_Total,
       AA_Households_Total,
       AA_Population_Total,
       CC_Sex_Ratio_All_Ages_Total
FROM Key_Indicator_ext_full
```

### ***4. Create and insert command for the formats such as ORC***

#### **Create command**

```
CREATE EXTERNAL TABLE IF NOT EXISTS Key_Indicator_ext_orc(
ID int,
State_Name string,
State_District_Name    string,
YY_Under_Five_Mortality_Rate_U5MR_Total_Person double,
LL_Total_Fertility_Rate_Total double,
AA_Households_Total double,
AA_Population_Total double,
CC_Sex_Ratio_All_Ages_Total double
)
STORED AS ORC
LOCATION '/user/root/indiaahs2012_13/key_indicator_orc'
tblproperties("orc.compress"="SNAPPY")
```

### **Insert command**

```
INSERT INTO Key_Indicator_ext_orc
SELECT ID,State_Name,State_District_Name,
       YY_Under_Five_Mortality_Rate_U5MR_Total_Person,
       LL_Total_Fertility_Rate_Total,
       AA_Households_Total,
       AA_Population_Total,
       CC_Sex_Ratio_All_Ages_Total
FROM Key_Indicator_ext_full;
```

### ***5. Create and insert command for the Hive-HBase integrated table***

#### **Create command**

```
create table Key_Indicator_ext_hive(
  'ID' int,
  'State_Name' string,
  'State_District_Name' string,
  'Mortality_Rate' double,
  'Fertility_Rate' double,
  'Households' double,
  'Population' double,
  'Sex_Ratio' double
)
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,district:State_Name,district:State_District_Name,
serveydata:Mortality_Rate,serveydata:Fertility_Rate,serveydata:Households,serveydata:Population,serveydata:Sex_Ratio"
)
TBLPROPERTIES ("hbase.table.name" = "Key_Indicator_ext__hive_hbase");
```

## Insert command

```
insert overwrite table Key_Indicator_ext_hive
```

```
select Key_Indicator_ext_full.ID, Key_Indicator_ext_full.State_Name, Key_Indicator_ext_full.State_District_Name,  
       Key_Indicator_ext_full.YY_Under_Five_Mortality_Rate_U5MR_Total_Person,  
       Key_Indicator_ext_full.LL_Total_Fertility_Rate_Total,  
       Key_Indicator_ext_full.AA_Households_Total,  
       Key_Indicator_ext_full.AA_Population_Total,  
       Key_Indicator_ext_full.CC_Sex_Ratio_All_Ages_Total  
from Key_Indicator_ext_full;
```

*6. Screenshot of runtimes against each query given above for the default format, formats such as ORC format as well as Hive-Hbase integration*

*For default format:*

## Screenshot of run time for query 1

The screenshot displays a Hive query execution interface. At the top, the query command is entered: `select count(*) from Key_Indicator_ext_default;`. Below the command, the execution logs are shown, detailing the job ID, MapReduce jobs launched, stage progress, and the total time taken. The time taken, `25.351 s`, is circled in red. At the bottom, the results are displayed in a table with one column, `_c0`, and one row containing the value `284`.

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

INFO : Ended Job = job\_1571418346229\_0001  
INFO : MapReduce Jobs Launched:  
INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.9 sec HDFS Read: 24803 HDFS Write: 4 SUCCESS  
INFO : Total MapReduce CPU Time Spent: 2 seconds 900 msec  
INFO : Completed executing command(queryId=hive\_20191018171313\_11297506-c2d0-47ed-95ff-b1ad46d95afc); Time taken: 25.351 s

Query History Saved Queries Results (1)

|   | _c0 |
|---|-----|
| 1 | 284 |



## Screenshot of run time for query 2

```
select State_Name, count(*) from Key_Indicator_ext_default group by State_Name;
```

```
INFO : Ended Job = job_1571416340229_0002
INFO : MapReduce Jobs Launched:
INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.07 sec HDFS Read: 25249 HDFS Write: 120 SUCCESS
INFO : Total MapReduce CPU Time Spent: 3 seconds 70 msec
INFO : Completed executing command(queryId=hive_20191018171616_edcfef9a-93d3-444d-b65c-cc592b7ceaea); Time taken: 22.198 s
```

Query History Saved Queries Results (9)

|   | state_name   | _c1 |
|---|--------------|-----|
| 1 | Assam        | 23  |
| 2 | Bihar        | 37  |
| 3 | Chhattisgarh | 16  |

## Screenshot of run time for query 3

```
select * from Key_Indicator_ext_default where State_Name = 'Uttar Pradesh';
```

```
INFO : MapReduce Jobs Launched:
INFO : Stage-Stage-1: Map: 1 Cumulative CPU: 1.84 sec HDFS Read: 21663 HDFS Write: 4263 SUCCESS
INFO : Total MapReduce CPU Time Spent: 1 seconds 840 msec
INFO : Completed executing command(queryId=hive_20191018171818_02999246-f57f-42e7-8bda-02661fa294fa); Time taken: 13.294 s
```

Query History Saved Queries Results (70)

|   | key_indicator_ext_default.id | key_indicator_ext_default.state_name | key_indicator_ext_default.state_district_name |
|---|------------------------------|--------------------------------------|---|
| 1 | 202                          | Uttar Pradesh                        | Agra  |
| 2 | 203                          | Uttar Pradesh                        | Aligarh                                       |
| 3 | 204                          | Uttar Pradesh                        | Allahabad                                     |

For formats such as ORC:

### Screenshot of run time for query 1

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

NFO : MapReduce Jobs Launched:  
NFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.96 sec HDFS Read: 25263 HDFS Write: 4 SUCCESS  
NFO : Total MapReduce CPU Time Spent: 2 seconds 960 msec  
NFO : Completed executing command(queryId=hive\_20191018172121\_d254c959-a50f-4c24-bdca-cdbb46be732e); (Time taken: 21.825 s)  
conds

Query History Saved Queries Results (1)

|   | _c0 |
|---|-----|
| 1 | 284 |

### Screenshot of run time for query 2

```
select State_Name, count(*) from Key_Indicator_ext_orc group by State_Name;
```

NFO : MapReduce Jobs Launched:  
NFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.05 sec HDFS Read: 26286 HDFS Write: 120 SUCCESS  
NFO : Total MapReduce CPU Time Spent: 3 seconds 50 msec  
NFO : Completed executing command(queryId=hive\_20191018172323\_6a4fbdcf-71ce-456d-b981-d4306b2bd74d); (Time taken: 21.786 s)  
conds

Query History Saved Queries Results (9)

|   | state_name   | _c1 |
|---|--------------|-----|
| 1 | Assam        | 23  |
| 2 | Bihar        | 37  |
| 3 | Chhattisgarh | 16  |
| 4 | Jharkhand    | 18  |

### Screenshot of run time for query 3

```
select * from Key_Indicator_ext_orc where State_Name = 'Uttar Pradesh';
```

NFO : MapReduce Jobs Launched:  
NFO : Stage-Stage-1: Map: 1 Cumulative CPU: 1.68 sec HDFS Read: 25943 HDFS Write: 4263 SUCCESS  
NFO : Total MapReduce CPU Time Spent: 1 seconds 680 msec  
NFO : Completed executing command(queryId=hive\_20191018172424\_bef8d697-75e7-4c9e-810d-661b2104d5b9); (Time taken: 14.083 s)  
conds

Query History Saved Queries Results (70)

|   | key_indicator_ext_orc.id | key_indicator_ext_orc.state_name | key_indicator_ext_orc.state_district_name | key |
|---|--------------------------|----------------------------------|---|-----|
| 1 | 202                      | Uttar Pradesh                    | Agra                                      | 69  |
| 2 | 203                      | Uttar Pradesh                    | Aligarh                                   | 90  |
| 3 | 204                      | Uttar Pradesh                    | Allahabad                                 | 104 |
| 4 | 205                      | Uttar Pradesh                    | Ambedkar Nagar                            | 78  |

## For Hive-Hbase Integrated Table:

### Screenshot of run time for query 1

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

INFO : mapreduce jobs Launched.  
INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.86 sec HDFS Read: 16793 HDFS Write: 4 SUCCESS  
INFO : Total MapReduce CPU Time Spent: 3 seconds 860 msec  
INFO : Completed executing command(queryId=hive\_20191019111919\_ddfe5cff-dc66-481f-a16d-5968df150301); Time taken: 23.824 seconds  
INFO : OK

Query History Saved Queries Results (1)

|   | _c0 |
|---|-----|
| 1 | 284 |

### Screenshot of run time for query 2

```
select State_Name, count(*) from Key_Indicator_ext_hive group by State_Name;
```

INFO : mapreduce jobs Launched.  
INFO : Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.26 sec HDFS Read: 17132 HDFS Write: 120 SUCCESS  
INFO : Total MapReduce CPU Time Spent: 4 seconds 260 msec  
INFO : Completed executing command(queryId=hive\_20191019112020\_1d9a9115-27f7-437f-b062-85b54e5ab71c); Time taken: 25.461 seconds  
INFO : OK

Query History Saved Queries Results (9)

|   | state_name     | _c1 |
|---|----------------|-----|
| 1 | Assam          | 23  |
| 2 | Bihar          | 37  |
| 3 | Chhattisgarh   | 16  |
| 4 | Jharkhand      | 18  |
| 5 | Madhya Pradesh | 45  |

### Screenshot of run time for query 3

```
select * from Key_Indicator_ext_hive where State_Name = 'Uttar Pradesh';
```

INFO : mapreduce jobs Launched.  
INFO : Stage-Stage-1: Map: 1 Cumulative CPU: 3.38 sec HDFS Read: 13809 HDFS Write: 4263 SUCCESS  
INFO : Total MapReduce CPU Time Spent: 3 seconds 380 msec  
INFO : Completed executing command(queryId=hive\_20191019113030\_eb9c8cec-9e70-4a22-9c12-b849c281b137); Time taken: 20.25 seconds  
INFO : OK

Query History Saved Queries Results (70)

|   | key_indicator_ext_hive.id | key_indicator_ext_hive.state_name | key_indicator_ext_hive.state_district_name |
|---|---------------------------|-----------------------------------|--|
| 1 | 202                       | Uttar Pradesh                     | Agra                                       |
| 2 | 203                       | Uttar Pradesh                     | Aligarh                                    |
| 3 | 204                       | Uttar Pradesh                     | Allahabad                                  |
| 4 | 205                       | Uttar Pradesh                     | Ambedkar Nagar                             |
| 5 | 206                       | Uttar Pradesh                     | Auraiya                                    |

*6. Create and insert command for the partition table for analyses 1 & 2.*

**Create command**

```
CREATE EXTERNAL TABLE IF NOT EXISTS Key_Indicator_ext_partition(  
  ID int,  
  State_District_Name    string,  
  YY_Under_Five_Mortality_Rate_U5MR_Total_Person double,  
  LL_Total_Fertility_Rate_Total double,  
  AA_Households_Total double,  
  AA_Population_Total double,  
  CC_Sex_Ratio_All_Ages_Total double  
)  
PARTITIONED BY (State_Name string)  
STORED AS ORC  
LOCATION '/user/root/key_indicator/key_indicator_partition'  
tblproperties("orc.compress"="SNAPPY");
```

**Insert command**

```
INSERT INTO Key_Indicator_ext_partition  
PARTITION (State_Name)  
SELECT  
ID,State_District_Name,YY_Under_Five_Mortality_Rate_U5MR_Total_Person,LL_Total_Fertility_Rate_Total,AA_Househ  
olds_Total,AA_Population_Total,CC_Sex_Ratio_All_Ages_Total,State_Name  
FROM Key_Indicator_ext_orc;
```

## IV. Analysis

### 1. The child mortality rate of Uttar Pradesh

#### Query

```
SELECT State_Name,  
AVG(YY_Under_Five_Mortality_Rate_U5MR_Total_Person) AS Child_Mortality_Rate_UP  
FROM Key_Indicator_ext_partition  
WHERE State_Name='Uttar Pradesh'  
GROUP BY State_Name;
```

#### Screenshot

```
SELECT State_Name,  
AVG(YY_Under_Five_Mortality_Rate_U5MR_Total_Person) AS Child_Mortality_Rate_UP  
FROM Key_Indicator_ext_partition  
WHERE State_Name='Uttar Pradesh'  
GROUP BY State_Name;
```

\*\*\*

| Query History | Q             | 📄                       | Saved Queries | Q | Results (1) | Q | ↗ |
|---------------|---------------|-------------------------|---------------|---|-------------|---|---|
| state_name    |               | child_mortality_rate_up |               |   |             |   |   |
| 1             | Uttar Pradesh | 90.22857142857143       |               |   |             |   |   |

## 2. The fertility rate of Bihar

### Query

```
SELECT State_Name,  
AVG(LL_Total_Fertility_Rate_Total) AS Fertility_Rate_Bihar  
from Key_Indicator_ext_partition  
WHERE TRIM(State_Name) = 'Bihar'  
GROUP BY State_Name;
```

### Screenshot

```
SELECT State_Name,  
AVG(LL_Total_Fertility_Rate_Total) AS Fertility_Rate_Bihar  
from Key_Indicator_ext_partition  
WHERE TRIM(State_Name) = 'Bihar'  
GROUP BY State_Name;
```

Query History  

Saved Queries 

Results (1)  

state\_name

fertility\_rate\_bihar

1 Bihar

3.532432432432432

3. State wise child mortality rate and state wise fertility rate and does high fertility correlate with high child mortality?

- *State wise Child Mortality Rate:*

**Query**

```
SELECT State_Name ,  
AVG(YY_Under_Five_Mortality_Rate_U5MR_Total_Person) AS Child_Mortality_Rate  
from Key_Indicator_ext_partition  
GROUP BY State_Name;
```

**Screenshot**

|   | state_name     | child_mortality_rate |
|---|----------------|----------------------|
| 1 | Assam          | 71.43478260869566    |
| 2 | Bihar          | 69.62162162162163    |
| 3 | Chhattisgarh   | 62.5                 |
| 4 | Jharkhand      | 53.44444444444444    |
| 5 | Madhya Pradesh | 83.37777777777778    |
| 6 | Odisha         | 75.8                 |
| 7 | Rajasthan      | 75.0625              |
| 8 | Uttar Pradesh  | 90.22857142857143    |
| 9 | Uttarakhand    | 41.84615384615385    |

*[State wise child mortality and fertility rate is retrieved using the respective queries and there by correlation is calculated]*

- *State wise Fertility Rate:*

### Query

```
SELECT  
State_Name ,  
AVG(LL_Total_Fertility_Rate_Total) AS Fertility_Rate  
from Key_Indicator_ext_partition  
GROUP BY State_Name;
```

### Screenshot

|   | state_name     | fertility_rate     |
|---|----------------|--------------------|
| 1 | Assam          | 2.4                |
| 2 | Bihar          | 3.532432432432432  |
| 3 | Chhattisgarh   | 2.70125            |
| 4 | Jharkhand      | 2.894444444444445  |
| 5 | Madhya Pradesh | 3.031111111111111  |
| 6 | Odisha         | 2.28               |
| 7 | Rajasthan      | 3.028125           |
| 8 | Uttar Pradesh  | 3.3978571428571427 |
| 9 | Uttarakhand    | 2.022307692307692  |

- *Correlation between High Child Mortality and High Fertility:*

### Query

```
SELECT CORR(Child_Mortality_Rate,Fertility_Rate) AS Correlation  
FROM  
(  
SELECT  
State_Name ,  
AVG(YY_Under_Five_Mortality_Rate_U5MR_Total_Person) AS Child_Mortality_Rate,  
AVG(LL_Total_Fertility_Rate_Total) AS Fertility_Rate  
from Key_Indicator_ext_partition  
GROUP BY State_Name)a;
```

### Screenshot

|   | correlation        |
|---|--------------------|
| 1 | 0.5580197666452783 |



#### 4. Find top 2 districts per state with the highest population per household

##### Query on the table with the chosen format such as orc

```
SELECT State_Name AS State_Name,  
       State_District_Name AS State_District_Name,  
       pp_per_hh As Population_Per_Household  
FROM (  
       SELECT State_Name ,  
              State_District_Name ,  
              (AA_Population_Total/AA_Households_Total) as pp_per_hh,  
              RANK() OVER(PARTITION BY State_Name ORDER BY (AA_Population_Total/AA_Households_Total) DESC) AS RNK  
       FROM Key_Indicator_ext_orc) a  
WHERE RNK IN(1,2);
```

##### Screenshot of the result

|    | state_name     | state_district_name          | population_per_household |
|----|----------------|------------------------------|--------------------------|
| 1  | Assam          | Dhemaji                      | 5.2103445894620535       |
| 2  | Assam          | Marigaon                     | 4.978445126406547        |
| 3  | Bihar          | Gopalganj                    | 5.979195301761839        |
| 4  | Bihar          | Nawada                       | 5.944978455419291        |
| 5  | Chhattisgarh   | Durg                         | 4.716408016844732        |
| 6  | Chhattisgarh   | Rajnandgaon                  | 4.651162790697675        |
| 7  | Jharkhand      | Kodarma                      | 5.868167462952465        |
| 8  | Jharkhand      | Giridih                      | 5.787106964805766        |
| 9  | Madhya Pradesh | Jhabua                       | 5.5903925014645575       |
| 10 | Madhya Pradesh | Sehore                       | 5.366774132372464        |
| 11 | Odisha         | Bhadrak                      | 4.765950743055191        |
| 12 | Odisha         | Jajapur                      | 4.494145867839397        |
| 13 | Rajasthan      | Dhaulpur                     | 5.810972222222222        |
| 14 | Rajasthan      | Barmer                       | 5.629192111322455        |
| 15 | Uttar Pradesh  | Sant Ravidas Nagar (Bhadohi) | 6.210831290394473        |
| 16 | Uttar Pradesh  | Baghpat                      | 6.11956799591002         |
| 17 | Uttarakhand    | Udham Singh Nagar            | 5.1164532900989546       |
| 18 | Uttarakhand    | Nainital                     | 4.748913659550349        |


[Find top 2 districts in each state with the highest population per household is retrieved]

### 5. Find top 2 districts per state with the lowest sex ratios

**Query on the table with the chosen format such as orc**

```
SELECT State_Name,District_Name,Sex_Ratio
FROM(
SELECT
State_Name ,
State_District_Name AS District_Name,
CC_Sex_Ratio_All_Ages_Total AS Sex_Ratio,
DENSE_RANK() OVER(PARTITION BY State_Name ORDER BY CC_Sex_Ratio_All_Ages_Total) AS RNK
from Key_Indicator_ext_orc)a
WHERE a.RNK IN(1,2);
```

### **Screenshot of the result**

|   | state_name     | district_name       | sex_ratio |
|---|----------------|---------------------|-----------|
| 1   | Assam          | Kamrup              | 925       |
|  | Assam          | North Cachar Hills  | 941       |
| 3   | Bihar          | Pashchim Champaran  | 894       |
| 4   | Bihar          | Khagaria            | 900       |
| 5   | Chhattisgarh   | Koriya              | 937.3     |
| 6   | Chhattisgarh   | Bilaspur            | 948.43    |
| 7   | Jharkhand      | Dhanbad             | 913       |
| 8   | Jharkhand      | Bokaro              | 917       |
| 9   | Madhya Pradesh | Morena              | 833.13    |
| 10  | Madhya Pradesh | Datia               | 852.12    |
| 11  | Odisha         | Sonapur             | 941       |
| 12  | Odisha         | Jharsuguda          | 944       |
| 13  | Rajasthan      | Karauli             | 837       |
| 14  | Rajasthan      | Dhaulpur            | 838       |
| 15  | Uttar Pradesh  | Gautam Buddha Nagar | 836.82    |
| 16  | Uttar Pradesh  | Shahjahanpur        | 853.67    |
| 17  | Uttarakhand    | Haridwar            | 884.93    |
| 18  | Uttarakhand    | Udham Singh Nagar   | 914.31    |

*[Top 2 District from each state with lowest sex ratio is successfully retrieved]*