Table countries

CREATE TABLE countries(country STRING, population INT, area INT, currency STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '\;'
tblproperties("skip.header.line.count"="1");

Loading statement

LOAD DATA LOCAL INPATH '/home/bigdata/hive/apache-hive-1.2.1-
bin/examples/files/countries.csv'
OVERWRITE INTO TABLE countries;

Table covid-19

CREATE TABLE covid(isocode STRING, continent STRING, location STRING, dateInput STRING,
newcases INT, newdeaths INT, icupatients INT, hospitalpatients INT, newvaccinations int)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
tblproperties("skip.header.line.count"="1");

Loading statement

LOAD DATA LOCAL INPATH '/home/bigdata/hive/apache-hive-1.2.1-bin/examples/files/covid19-
detail.csv'
OVERWRITE INTO TABLE covid;

Q4.1)
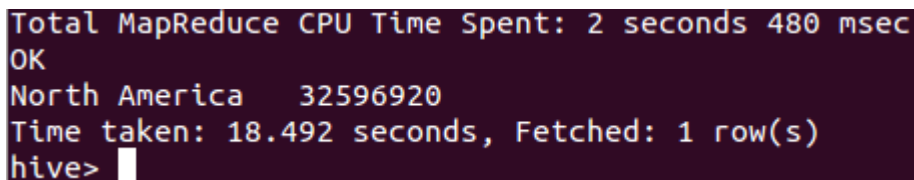QUERY
SELECT continent, SUM(newcases)-SUM(newdeaths)
FROM covid
where continent = "North America"
GROUP BY continent;

OUTPUT
North America   32596920

```
Total MapReduce CPU Time Spent: 2 seconds 480 msec
OK
North America    32596920
Time taken: 18.492 seconds, Fetched: 1 row(s)
hive>
```

Q4.2)
QUERY
SELECT location, SUM(newcases)
from covid
GROUP BY location

HAVING SUM(newcases) > 1000000;

OUTPUT
Argentina   2149636
Brazil 11019344
Colombia   2276656
Czechia   1321331
France 3964078
Germany   2508655
India   11229398
Indonesia   1379662
Iran   1689692
Italy   3067486
Mexico   2128600
Netherlands   1135258
Peru   1364964
Poland 1794914
Russia 4274263
South Africa   1521068
Spain   3149012
Turkey 1957192
Ukraine   1449741
United Kingdom   4231166
United States   28993309
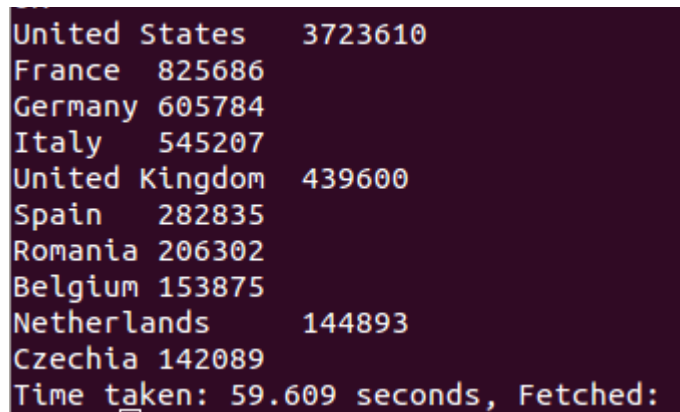
```
Argentina          2149636
Brazil   11019344
Colombia           2276656
Czechia 1321331
France   3964078
Germany 2508655
India    11229398
Indonesia          1379662
Iran     1689692
Italy    3067486
Mexico   2128600
Netherlands        1135258
Peru     1364964
Poland   1794914
Russia   4274263
South Africa       1521068
Spain    3149012
Turkey   1957192
Ukraine 1449741
United Kingdom   4231166
United States    28993309
```

Q4.3)
QUERY

SELECT location, SUM(icupatients) as sum
from covid
GROUP BY location
SORT BY sum DESC
LIMIT 10;

OUTPUT
United States 3723610
France 825686
Germany   605784
Italy   545207
United Kingdom   439600
Spain   282835
Romania   206302
Belgium   153875
Netherlands   144893
Czechia   142089

```
United States    3723610
France  825686
Germany 605784
Italy    545207
United Kingdom    439600
Spain    282835
Romania 206302
Belgium 153875
Netherlands       144893
Czechia 142089
Time taken: 59.609 seconds, Fetched:
```

Q4.4)
QUERY

SELECT location, SUM(newdeaths)/100000 AS mortalityrate
FROM covid
GROUP BY location
SORT BY mortalityrate DESC
LIMIT 10;

OUTPUT
United States 5.25031
Brazil 2.65411
Mexico   1.90604
India   1.57853

United Kingdom   1.24736
Italy   0.99785
France 0.88727
Russia 0.87614
Germany   0.71984
Spain   0.71138

```
Total MapReduce CPU Time Spent: 4 seconds 590 msec
OK
United States   5.25031
Brazil   2.65411
Mexico  1.90604
India    1.57853
United Kingdom  1.24736
Italy    0.99785
France   0.88727
Russia   0.87614
Germany 0.71984
Spain    0.71138
Time taken: 61.189 seconds, Fetched: 10 row(s)
hive>
```

Q4.5)
QUERY

SELECT cov.location, (SUM(cov.newvaccinations)/country.population)*100 AS vacinationrate,
country.population
from covid cov JOIN countries country ON (cov.location = country.country)
WHERE country.population > 1000000
GROUP BY cov.location, country.population
SORT BY vacinationrate DESC
LIMIT 3;

OUTPUT

Israel   108.69809526625498 8049314
United Arab Emirates 84.94468628455161   5779760
United Kingdom   32.23297878352749 64088222

```
Total MapReduce CPU Time Spent: 5 seconds 120 msec
OK
Israel   108.69809526625498      8049314
United Arab Emirates    84.94468628455161       5779760
United Kingdom  32.23297878352749       64088222
Time taken: 62.005 seconds, Fetched: 3 row(s)
hive>
```