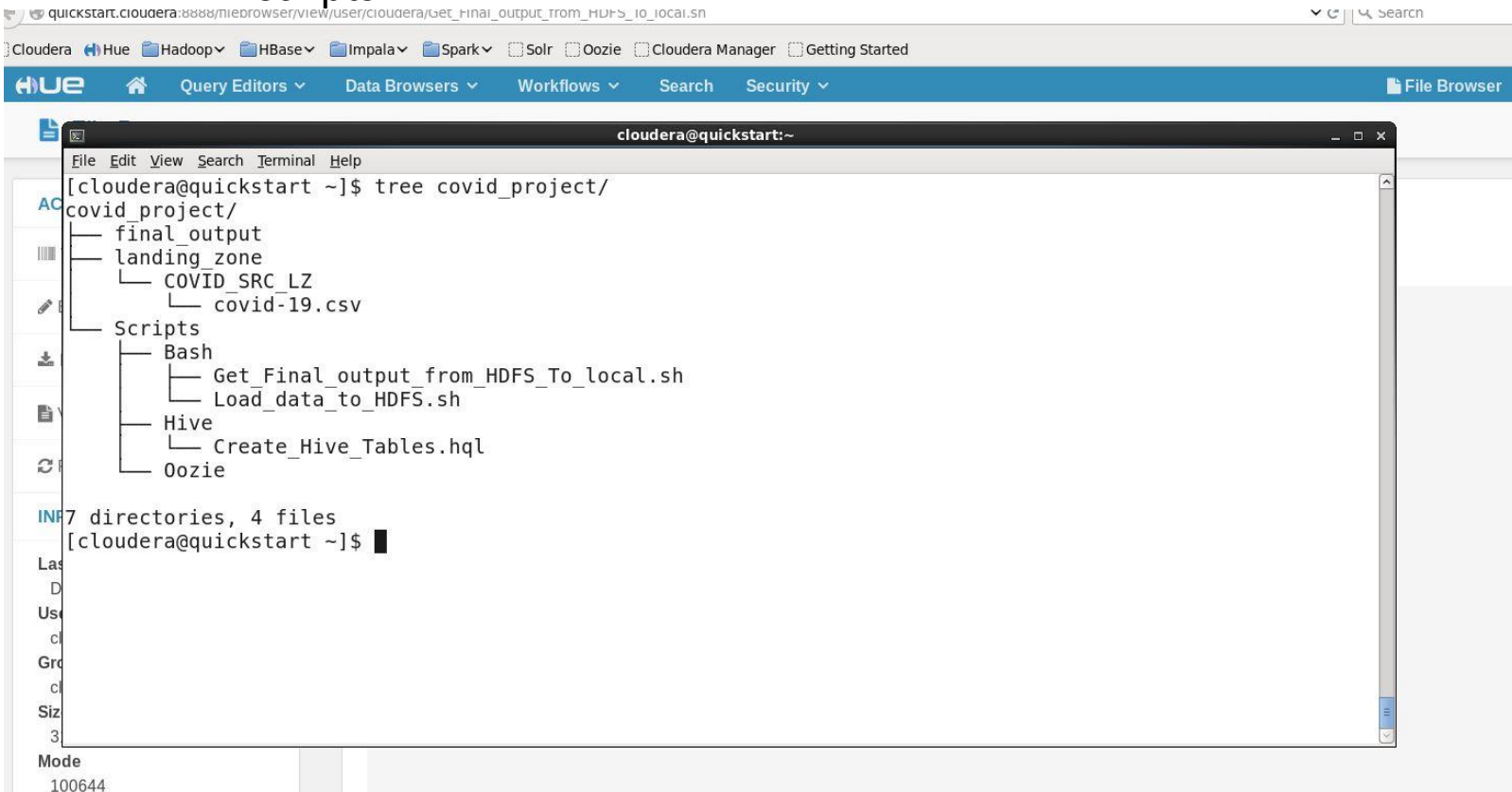


1. run (tree) command to see folders before running the scripts .

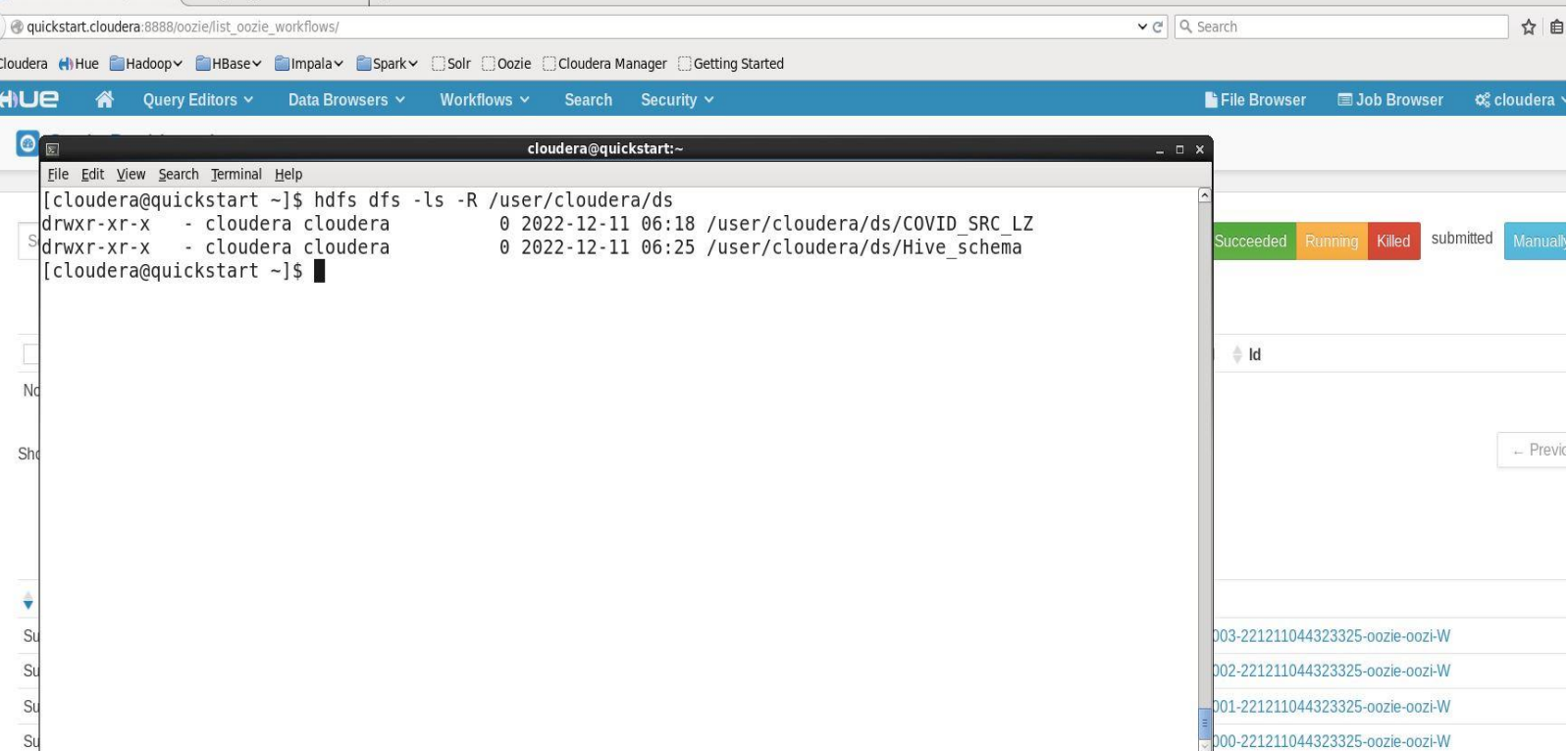


The screenshot shows the Hue web interface with a terminal window open. The terminal displays the output of the 'tree' command for the 'covid\_project' directory. The output shows a directory structure with 'final\_output', 'landing\_zone', and 'Scripts' folders. The 'landing\_zone' folder contains 'COVID\_SRC\_LZ' and 'covid-19.csv'. The 'Scripts' folder contains 'Bash', 'Hive', and 'Oozie' subfolders. The 'Bash' folder contains 'Get\_Final\_output\_from\_HDFS\_To\_local.sh' and 'Load\_data\_to\_HDFS.sh'. The 'Hive' folder contains 'Create\_Hive\_Tables.hql'. The terminal also shows the output of the 'find' command, indicating 7 directories and 4 files.

```
[cloudera@quickstart ~]$ tree covid_project/
covid_project/
├── final_output
├── landing_zone
│   ├── COVID_SRC_LZ
│   └── covid-19.csv
└── Scripts
    ├── Bash
    │   ├── Get_Final_output_from_HDFS_To_local.sh
    │   └── Load_data_to_HDFS.sh
    ├── Hive
    │   └── Create_Hive_Tables.hql
    └── Oozie

INF7 directories, 4 files
[cloudera@quickstart ~]$
```

2. list all of folders and files in ds folders before running the scripts .



The screenshot shows the Hue web interface with a terminal window open. The terminal displays the output of the 'hdfs dfs -ls -R' command for the '/user/cloudera/ds' directory. The output shows two directories: 'COVID\_SRC\_LZ' and 'Hive\_schema', both created on 2022-12-11 at 06:18 and 06:25 respectively. The terminal also shows the output of the 'find' command, indicating 2 directories and 0 files.

```
[cloudera@quickstart ~]$ hdfs dfs -ls -R /user/cloudera/ds
drwxr-xr-x - cloudera cloudera      0 2022-12-11 06:18 /user/cloudera/ds/COVID_SRC_LZ
drwxr-xr-x - cloudera cloudera      0 2022-12-11 06:25 /user/cloudera/ds/Hive_schema
[cloudera@quickstart ~]$
```

### 3. Run the Hive and Bash Scripts through Hue .

The screenshot displays the Oozie Dashboard interface within the Hue web application. The top navigation bar includes links for Cloudera, Hue, Hadoop, HBase, Impala, Spark, Solr, Oozie, Cloudera Manager, and Getting Started. The main header shows 'Oozie Dashboard' with tabs for Workflows, Coordinators, Bundles, SLA, and Oozie. The left sidebar contains sections for Workflow (Running Scripts, SUBMITTER, STATUS, PROGRESS, ID, VARIABLES, MANAGE) and a 'Rerun' button. The main content area shows the 'Workflow Running\_Scripts' in a 'SUCCEEDED' state with a 100% progress bar. The workflow graph shows a sequence of actions: a start node, a 'Shell' action (Load\_data\_to\_HDFS.sh), and a 'HiveServer2 Script' action (Create\_Hive\_Tables.hql), followed by an end node.

### 4. Here after running the scripts we found that the output folders and files were created.

The screenshot shows a terminal window on a Cloudera node. The user runs the following commands and receives the following output:

```
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/ds/
Found 4 items
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:55 /user/cloudera/ds/COVID_FINAL_OUTPUT1
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:55 /user/cloudera/ds/COVID_FINAL_OUTPUT2
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:53 /user/cloudera/ds/COVID_SRC_LZ
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:53 /user/cloudera/ds/Hive_schema
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/ds/COVID_FINAL_OUTPUT1/*
-rwxr-xr-x 1 cloudera cloudera 162 2022-12-11 07:55 /user/cloudera/ds/COVID_FINAL_OUTPUT1/000000_0
-rwxr-xr-x 1 cloudera cloudera 138 2022-12-11 07:55 /user/cloudera/ds/COVID_FINAL_OUTPUT2/000000_0
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/ds/COVID_FINAL_OUTPUT1/*
China0.041E7
USA0.3868332E7
Russia0.35E7
India0.2661252E7
UK0.5177265E7
Brazil0.3748152E7
Germany0.10197366E7
Spain0.955615.0
Italy0.790596.0
UAE0.265918.0
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/ds/COVID_FINAL_OUTPUT2/*
USA0.177424.0
Brazil0.112423.0
Mexico0.58481.0
India0.54975.0
UK0.1403.0
Italy0.35418.0
France0.0480.0
Spain0.28813.0
Peru0.27034.0
Iran0.20264.0
[cloudera@quickstart ~]$
```

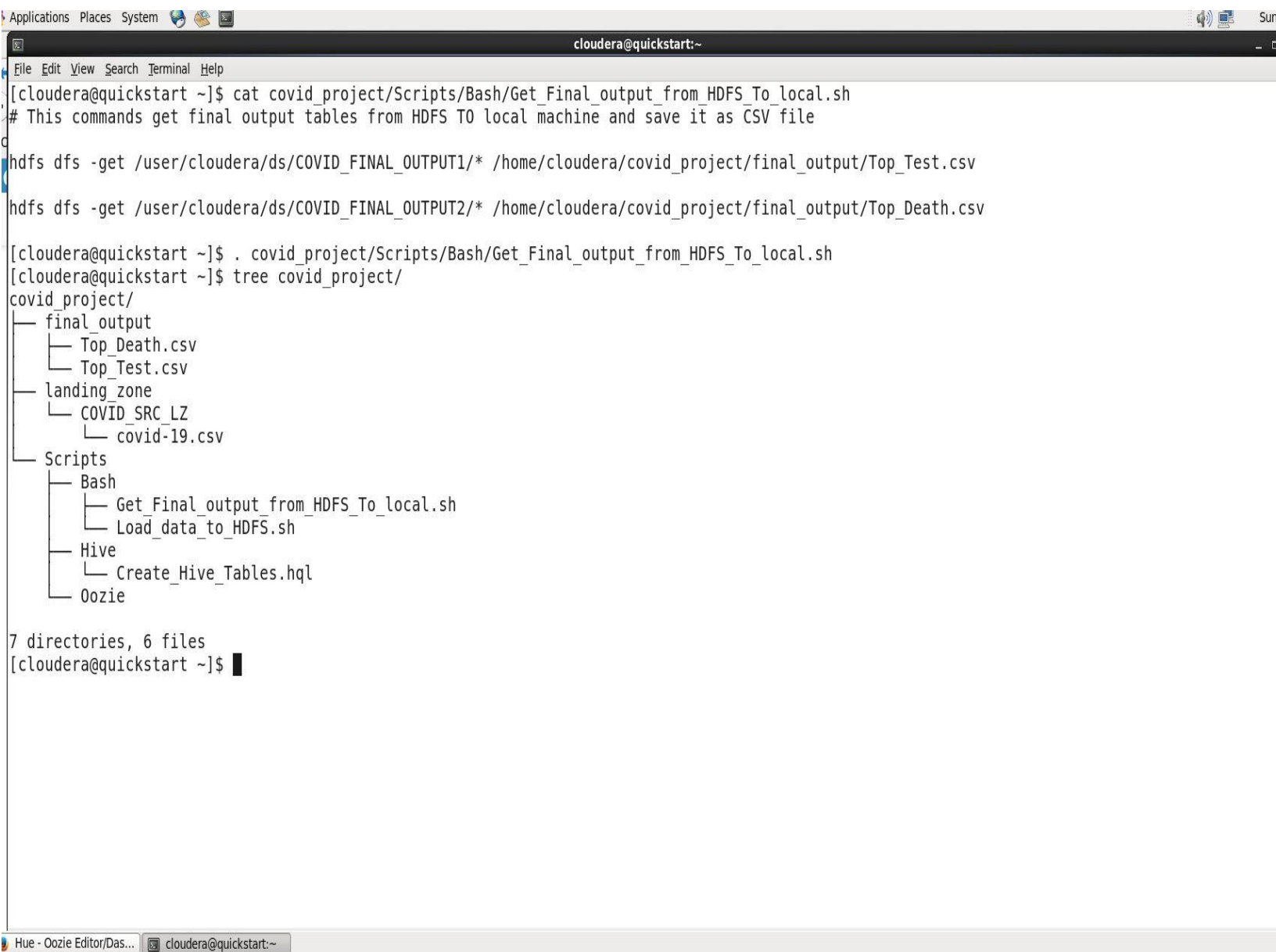
## 5. Data (covid-19.csv) was loaded in HDFS after running the scripts.

```
Applications Places System cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/ds/Hive_schema/Basic_Table  
Found 1 items  
-rw-r--r-- 1 yarn cloudera 20707 2022-12-11 07:52 /user/cloudera/ds/Hive_schema/Basic_Table/covid-19.csv  
[cloudera@quickstart ~]$
```

## 6. And the partition table that partition by (Country\_name) was created.

```
Applications Places System Sun Dec 11, 08:11 cloudera  
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/ds/Hive_schema/Partition_Table  
Found 216 items  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Afghanistan  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Albania  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Algeria  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Andorra  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Angola  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Anguilla  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Antigua and Barbuda  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Argentina  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Armenia  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Aruba  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Australia  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Austria  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Azerbaijan  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Bahamas  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Bahrain  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Bangladesh  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Barbados  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Belarus  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Belgium  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Belize  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Benin  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Bermuda  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Bhutan  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Bolivia  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Bosnia and Herzegovina  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Botswana  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Brazil  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=British Virgin Islands  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Brunei  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Bulgaria  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Burkina Faso  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Burundi  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=CAR  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Cabo Verde  
drwxr-xr-x - cloudera cloudera 0 2022-12-11 07:54 /user/cloudera/ds/Hive_schema/Partition_Table/country_name=Cambodia
```

7. Here is the bash script that gets the final output from HDFS to the local machine and after running. the output has appeared in the (final\_output ) folder.



```
Applications Places System [system icons] Sur
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ cat covid_project/Scripts/Bash/Get_Final_output_from_HDFS_To_local.sh
# This commands get final output tables from HDFS TO local machine and save it as CSV file

hdfs dfs -get /user/cloudera/ds/COVID_FINAL_OUTPUT1/* /home/cloudera/covid_project/final_output/Top_Test.csv

hdfs dfs -get /user/cloudera/ds/COVID_FINAL_OUTPUT2/* /home/cloudera/covid_project/final_output/Top_Death.csv

[cloudera@quickstart ~]$ . covid_project/Scripts/Bash/Get_Final_output_from_HDFS_To_local.sh
[cloudera@quickstart ~]$ tree covid_project/
covid_project/
├── final_output
│   ├── Top_Death.csv
│   └── Top_Test.csv
├── landing_zone
│   ├── COVID_SRC_LZ
│   └── covid-19.csv
├── Scripts
│   ├── Bash
│   │   ├── Get_Final_output_from_HDFS_To_local.sh
│   │   └── Load_data_to_HDFS.sh
│   ├── Hive
│   │   └── Create_Hive_Tables.hql
│   └── Oozie
7 directories, 6 files
[cloudera@quickstart ~]$
```

Hue - Oozie Editor/Das... cloudera@quickstart:~



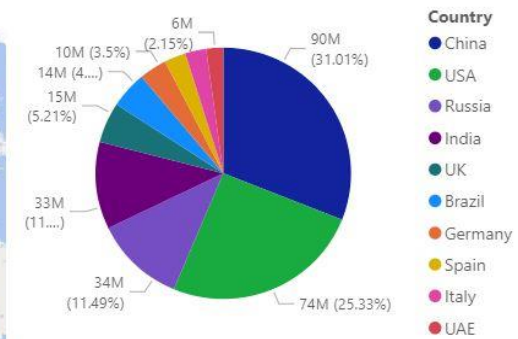
8. Finally final output was load data to Power BI to be visualized .

### Covid Analysis

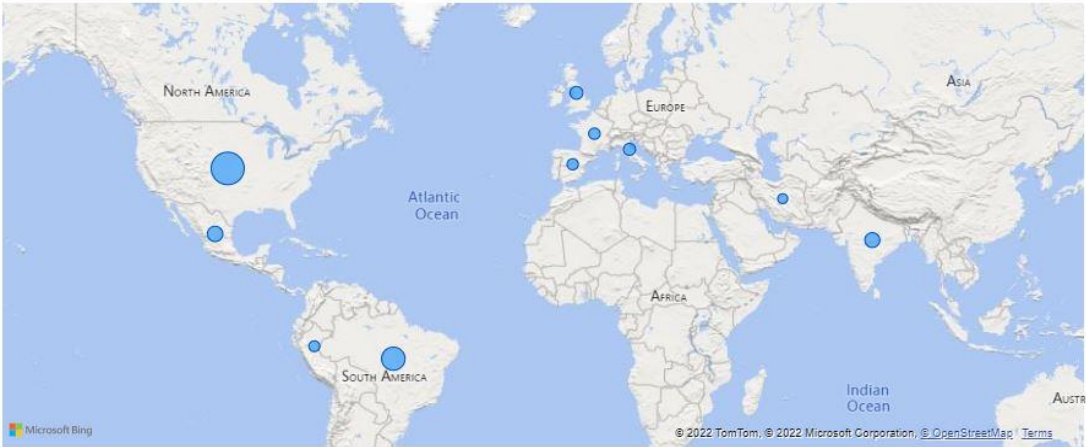
Total Number of Tests / Country



NUM\_OF\_Tests per Country



Total Number of Deaths / Country



NUM\_OF\_Deaths per Country

