**Task 1:**

1. Create HDFS directory /user/student and nested directories airports, airlines, routes, planes, countries

hadoop fs -mkdir /user/student

hadoop fs -mkdir /user/student/airports

hadoop fs -mkdir /user/student/airlines

hadoop fs -mkdir /user/student/routes

hadoop fs -mkdir /user/student/planes

hadoop fs -mkdir /user/student/countries

Table

Description automatically generated

1. Copy each file from the local file system's airlines dataset to the corresponding HDFS directory created in step #1 (i.e. routes.dat should be stored HDFS as /user/student/routes/routes.dat)

hadoop fs -put /home/cloudera/Documents/airlines.dat user/student/airlines/airlines.dat

hadoop fs -put /home/cloudera/Documents/airports.dat user/student/airports/airports.dat

hadoop fs -put /home/cloudera/Documents/countries.dat user/student/countries/countries.dat

hadoop fs -put /home/cloudera/Documents/planes.dat user/student/planes/planes.dat

hadoop fs -put /home/cloudera/Documents/routes.dat user/student/routes/routes.dat

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generated

1. Print out to the console first 9 lines of /u/ser/student/countries/countries.dat file

Hadoop fs -cat user/student/countries/countries.dat | head -n 9

Text

Description automatically generated

1. Compare MD5 checksums between the local file airports.dat and HDFS file /user/student/airports/airports.dat. Is there any difference?



Checksums are not equal due to different algorithms.

1. How much replicas does the /user/student/airports/airports.dat file have? Set number of replicas to 4 for the file

Hadoop fs -ls user/student/airports/airports.dat



1 replica, change to 4:

Hadoop fs -setrep -w 4 /user/student/airports/airports.dat

Text

Description automatically generated

1. Print out the total size of the /user/student dir in megabytes. Why the space consumed by the directory in HDFS doesn't equal to the space consumed by the same directory in local FS?

Hadoop fs -du -s -h /user/student

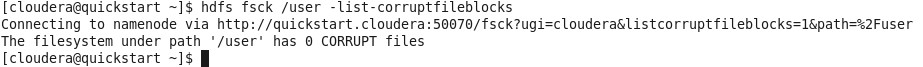


Size is different due to replication factor 4.

1. Run the filesystem checking utility for the /user directory in HDFS. Analyze utility's output in order to answer the following questions:

Hadoop fsck /user

* Are there any corrupted files and blocks?



No corrupted files or blocks.

* Are there any missed blocks?

Graphical user interface, text, email

Description automatically generated

No missed blocks. Note: file with replication factor 4 was not replicated due to 1 active DataNode.

* How HDFS is configured? How many Data Nodes and Racks are in place?

Graphical user interface, text, application, email

Description automatically generated

1 datanode, 1 rack

* Where the blocks are located? (block path)

Hdfs fsck /user/student -files -blocks -locations

**Text, letter

Description automatically generated**

Blocks are located in 10.0.2.15:50010.