HDFS\_MANUAL\_TASK

1. Create HDFS directory /user/student and nested directories airports, airlines, routes, planes, countries
2. 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)

Text

Description automatically generated

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

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?

Text

Description automatically generated

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

Для изменения фактора репликации для уже существующего файла была использована команда:



(изначально было 1, не сделала скрин)

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?

Text

Description automatically generated

Разный размер папок, так как для файла airports.dat установлен фактор репликации 4

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

Text

Description automatically generated

HDFS\_TASK\_AUTOMATION

