

File Management Tasks.

Aim:

Implement the following file management tasks in Hadoop :

- i) Adding files and directories
- ii) Retriving files
- iii) Deleting files

Algorithm:

step 1: Start the process

step 2: Open command prompt (Administrator)

step 3: Run the following command to start

Hadoop :

- i) start-all
- ii) dfs

step 4: Open the browser and enable "localhost:9870" and "localhost:8085".

step 5: Go to command prompt (Administrator) using the "hdfs dfs -mkdir" to make a directory

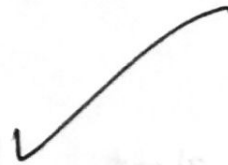
step 6: After making the directory using the "hdfs dfs -put" command to make a file.

step 7: Go to the browser and check the directory and file has been created in the "localhost:9870".

step 8: Go to command prompt (administrator),
use the "hdfs dfs -get" command to
retrieve a file from HDFS to the
local system, and use the "hdfs dfs -cat"
command to view the data in a file.

step 9: Run the "hdfs dfs -rm" command to
delete a existing file in the directory

step 10: stop the process



MapReduce Paradigm

Aim:

To Run a basic word count Map Reduce program to understand MapReduce ~~Paradigm~~ Paradigm.

Algorithm:

step 1: Start the process

step 2: open command prompt (Administrator)

step 3: Run the following command to start Hadoop.

i) start-all

ii) jps

step 4: open the browser and enable "localhost:9870" and "localhost:8088".

step 5: Open text document to create a document in the name of "example.txt" and save it in some particular directory.

step 6: Go to command prompt using the "hdfs dfs -mkdir" and "hdfs dfs -put" command insert the "example.txt" document in Hadoop environment.

step 7: Copy the Hadoop MapReduce jar file in hadoop directory in local system.

step 8: Go to command prompt using the "hadoop jar" command to execute the jar file with "example.tex".

step 9: After execution, verify the output directory for the status of program (Failure or success).

step 10: If we got 'success' as a output file, ~~the~~ program has executed successfully.

step 11: Stop the process.

weather Report using mapReduce.

Aim :

To write a Map Reduce program that mines weather data.

Algorithm :

step 1 : Start the process.

step 2 : open command prompt (Administrator) and do all necessary works. to start the Hadoop daemons.

step 3 : create a text file in the name of "weather_data.txt"

step 4 : And create python file in the names of "weather_mapper.py" and "weather_reducer.py".

step 5 : Go to command prompt and create a directory with the help of "-mkdir" keyword. and put these files in the directory with the help of "-put" keyword.

step 6 : copy the "Hadoop-streaming-3.2.4.jar" file path in the hadoop directory in local system.

step 7 : Go to command prompt using the "hadoop jar" command to execute the jar file with "weatherdata.txt", "weather-mapper" and "weather-reducer.py".

step 8 : After execution verify the output directory for the status of the program.

step 9 : Stop the process.



Matrix multiplication using MapReduce.

Aim:

to implement matrix multiplication with Hadoop Map Reduce.

Algorithm:

Step 1: start the process

Step 2: open command prompt (Administrator) and do all necessary works to start the Hadoop daemons.

Step 3: Create a text files in the name of 'MatrixA.txt' and 'MatrixB.txt'.

Step 4: And create python files in the names of "matrix_mapper.py" and "matrix_reducer.py"

Step 5: Go to command prompt and create a directory with the help of "-mkdir" keyword. and put the files in the directory with the help of "-put" keyword.

Step 6: copy the "hadoop-streaming-3.2.4.jar" file path in the hadoop directory in the local system.

Step 7: Go to command prompt using the "hadoop jar" command to execute the jar file with 'MatrixA.txt', 'MatrixB.txt', 'matrix_mapper.py' and 'matrix_reducer.py'.

step 8 : After execution verify the output
directory for the status of the program.

step 9 : stop the process.

Sales Data Report using MapReduce

Aim:

Develop a Map Reduce program to find the number of product sold in each country by considering sales data.

Algorithm:

- step 1: start the process
- step 2: open command prompt (administrator) and do all the necessary work to start the hadoop daemons.
- step 3: create a text file in the name of "sales_data.txt".
- step 4: And create python file in the name of "sales_mapper.py" and "sales_reducer.py".
- step 5: Go to command prompt and create a directory with the help of "mkdir" keyword and put the files in the directory with the help of "-put" keyword.
- step 6: copy the "hadoop-streaming-3.2.4.jar" file path in the hadoop directory in the local system.
- step 7: Go to command prompt using the "hadoop jar" command to create the jar file with "sales_data.txt", "sales_mapper.py" and "sales_reducer.py".

step 8: After execution verify the output,
directly for the status of the program.

step 9: stop the process.

Electrical Consumption Report using MapReduce.

Aim :

Develop a MapReduce program to find the Maximum electrical consumption in each year

Algorithm:

Step 1: start the process

step 2: open command prompt (Administrator) and do all necessary works to start the Hadoop daemons.

Step 3: create a text file in the name of "electricity.txt"

step 4: And create Python files in the name of "electricity_mapper.py" and "electricity_reducer.py".

step 5: Go to command prompt and create a directory with the help of "-mkdir" keyword. and put the file in the directory with the help of "-put" keyword.

step 6: Copy the "hadoop-streaming-3.2.4.jar" file path in the hadoop directory in the local system.

step 7: Go to command prompt using the "hadoop jar" command to execute the jar file with "electricity.txt", "electricity_mapper.py" and "electricity_reducer.py".

step 8: After execution verify the output directory for the status of the program.

step 9: stop the process.

Real-time data analysis using MapReduce.

Aim:

Develop a MapReduce program to analyze any real-time data.

Algorithm:

- Step 1: Start the process
- Step 2: Open command prompt (administrator) and do all necessary works to start the Hadoop daemons.
- Step 3: Create a text file in the name of 'twitter-data.txt'.
- Step 4: And create python file in the name of 'twitter-mapper.py' and 'twitter-reducer.py'.
- Step 5: Go to command prompt and create a directory with the help of "-mkdir" keyword. and put the files in the directory with the help of "-put" keyword.
- Step 6: Copy the 'hadoop-streaming-3.2.4.jar' file path in the hadoop directory in the local system.
- Step 7: Go to command prompt using the 'hadoop jar' command to execute the jar file with 'twitter-data.txt', 'twitter-mapper.py' and 'twitter-reducer.py'.

Step 8 : After execution verify the output directory for the status of the program.

Step 9 : Stop the process.