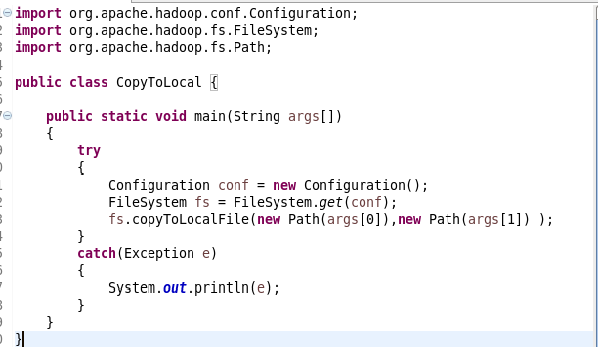
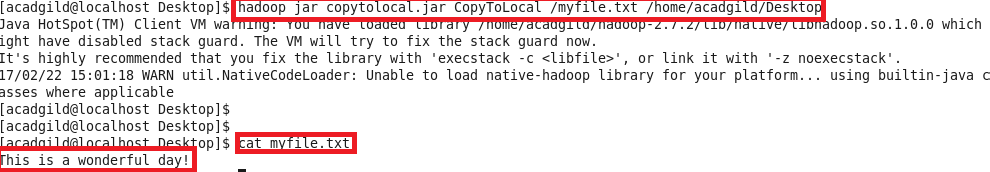
1. Write a Java program to copy a file from HDFS to LFS.

Solution:

Program:



Output:



1. Explain the importance and usage of the below terms in details

Solution:

1. **DFSInputStream**  
   -DFSInputStream is used while reading the data from HDFS by a client. When the name node returns the address of the block to the DFS, the DFS returns the DFSInputStream.

-FSDataInputStream in turn wraps a DFSInputStream which manages the datanode and namenode I/O.

-When the client calls read() operation the DFSInputStream which stores the address of the data nodes gets connected to one of the nearest datanode and then with repeated read() calls the data is read from the data node.

-When the block reaches the end of the file it closes the connection with the data node and finds next data node for the next block of data.

1. **DFSOutputStream**

- As the client writes data , DFSOutputStream splits it into packets, which it writes to an internal queue, called the data queue.

-The data queue is consumed by the Data Streamer, whose responsibility it is to ask the namenode to allocate new blocks by picking a list of suitable datanodes to store the replicas.

-DFSOutputStream also maintains an internal queue of packets that are waiting to be acknowledged by datanodes, which is called the ack queue.

1. **FSDataInputStream**

-When the client calls read() function, the DFSInputStream stores the address of first few blocks in the file, then connects to the first (closest) datanode for the first block in the file.

-Data is streamed from the datanode back to the client, which calls read() repeatedly on the stream.

-When the end of the block is reached, DFSInputStream will close the connection to the datanode, then find the best datanode for the next block.

-It will call the namenode to retrieve the datanode locations for the next batch of blocks as needed. When the client has finished reading, it calls close() on the FSDataInputStream.

-During reading, if the DFSInputStream encounters an error while communicating with a datanode, then it will try the next closest one for that block.

-It will also remember datanodes that have failed so that it doesn’t needlessly retry them for later blocks.

1. **FSDataOutputStream**

-The DFS returns an FSDataOutputStream for the client to start writing data to.

-FSDataOutputStream wraps a DFSOutputStream, which handles communication with the datanodes and namenode.