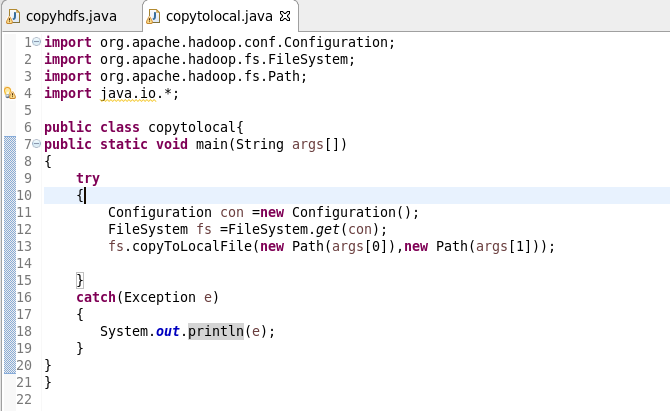
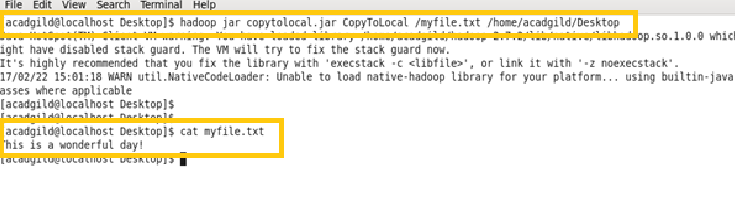
1. Write a Java program to copy a file from HDFS to LFS

PROGRAM



RUNNING JAR AND OUTPUT



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

**DFSInputStream:**

It is responsible for storing the data node adresses of the first few blocks of the file that is needed by the client. 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 of the file. This happens transparently to the client, which from its point of view is just reading a continuous stream.

**DFSOutputStream:**

DFSOutputStream splits the data which the client writes into packets, which it writes to an internal queue, called the data queue. The data queue is consumed by the DataStreamer, whose responsibility it is to ask the namenode to allocate new blocks by picking a list of suitable datanodes to store the replicas.

**FSDataInputStream:**

The open() method on FileSystem actually returns a FSDataInputStream rather than a standard java.io class. This class is a specialization of java.io.DataInputStream with support for random access, so you can read from any part of the stream

**FSDataOutputStream:**

The client creates the file by calling create() method on DistributedFileSystem. DistributedFileSystem makes an RPC call to the namenode to create a new file in the filesystem’s namespace, with no blocks associated with it.