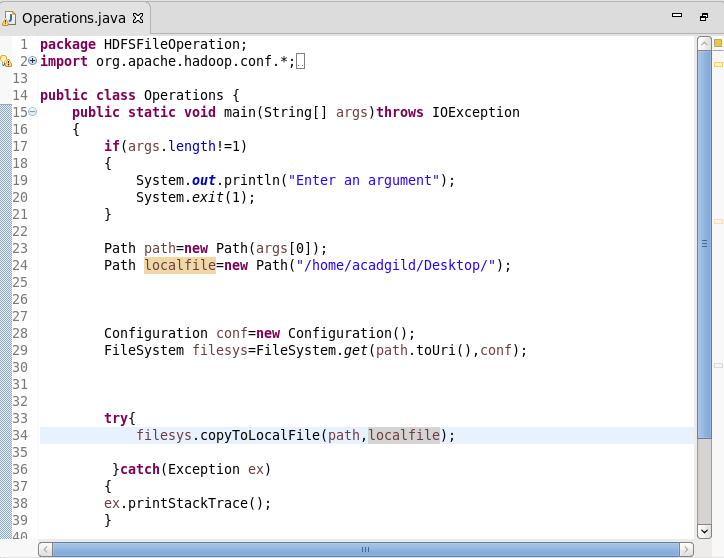
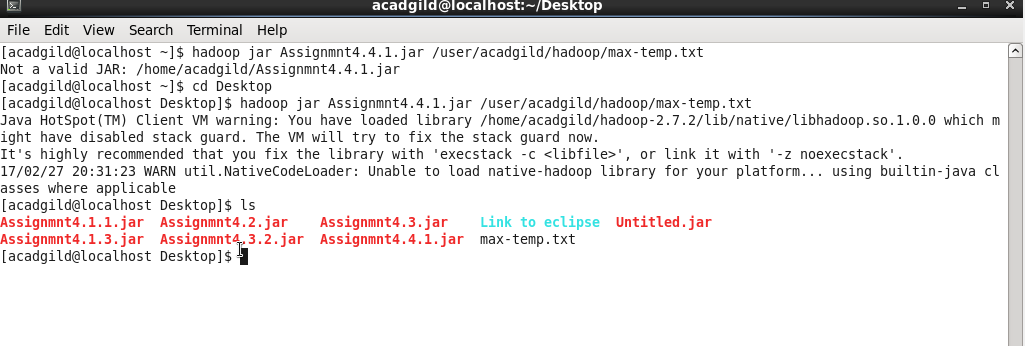
Java program to copy a file from HDFS to LFS





Question 2

|  |
| --- |
| DFSInputStream |
|  |  |
|  | DFSInputStream is the distributed file system and returns an input stream that supports file seeks, to the user to read data from. FSDataInputStream wraps a DFSInputStream, that manages the datanode and namenode I/O. |
|  | DFSInputStream has stored the datanode addresses for the first few blocks in the file, then connects to the first datanode for the first block in the file. When the end is reached ,DFSInputStreamStream will close the connection to the datanode and then find the best datanode for the next block. If the DFSInputStream encounters an error while communicating with a datanode, it will try to find the next block closest to it. |
|  |  |
|  | DFSOutputStream |
|  |  |
|  | FSOutputStream is an output stream which wraps DFSOutputStream and returns it to the client. The user uses the FSDataOutputStream object to write the data and calls the write() method. It also handles the communication with the NameNode and the datanodes. DFSOutputStream splits the files into blocks and with the NameNode identifies the DataNode and its copies. DFSOutputStream anages a queue to check if all the packets have been written by the NameNode and sent. It also handles all the failures while sending the data packets. |