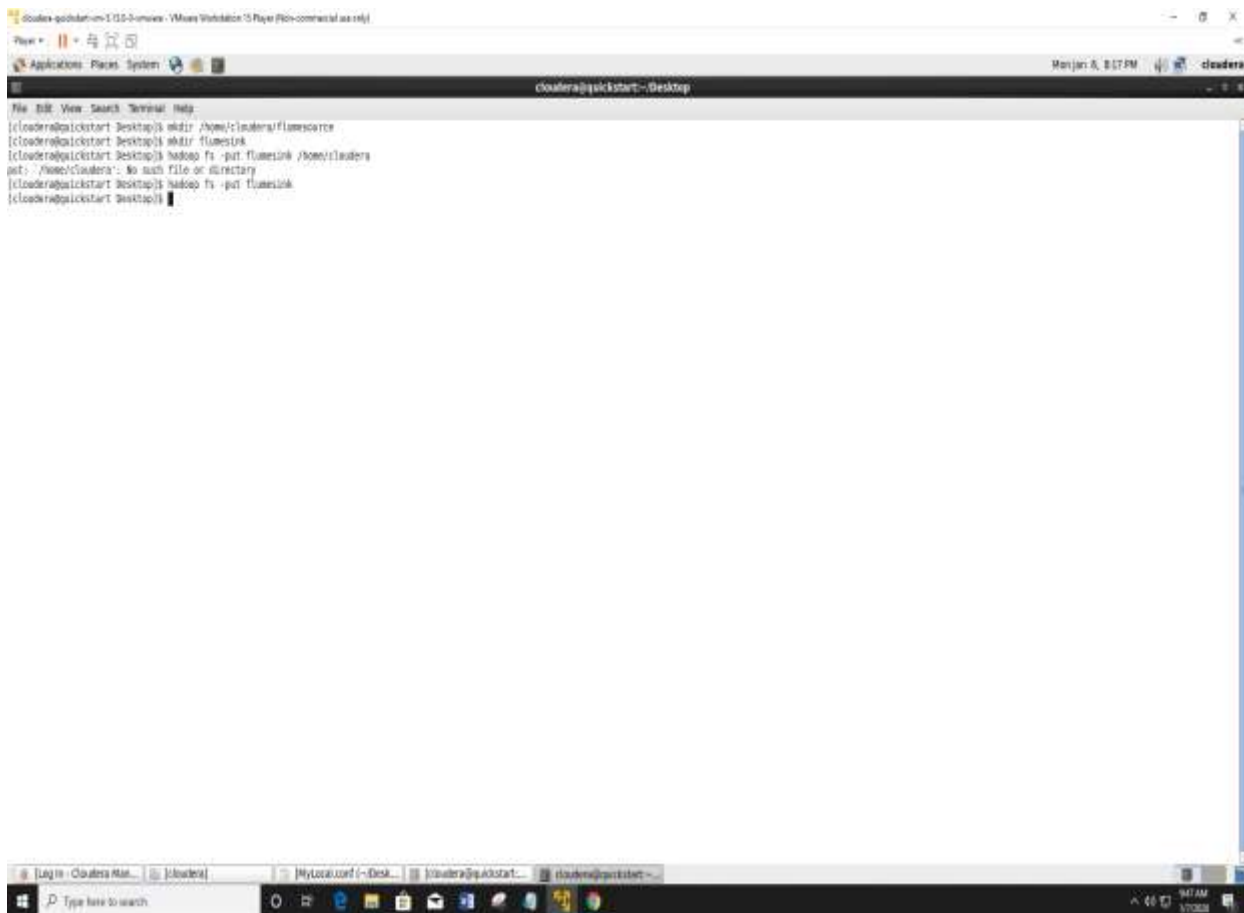


# Hadoop Assignment - 6

Name - **Nipun Garg**

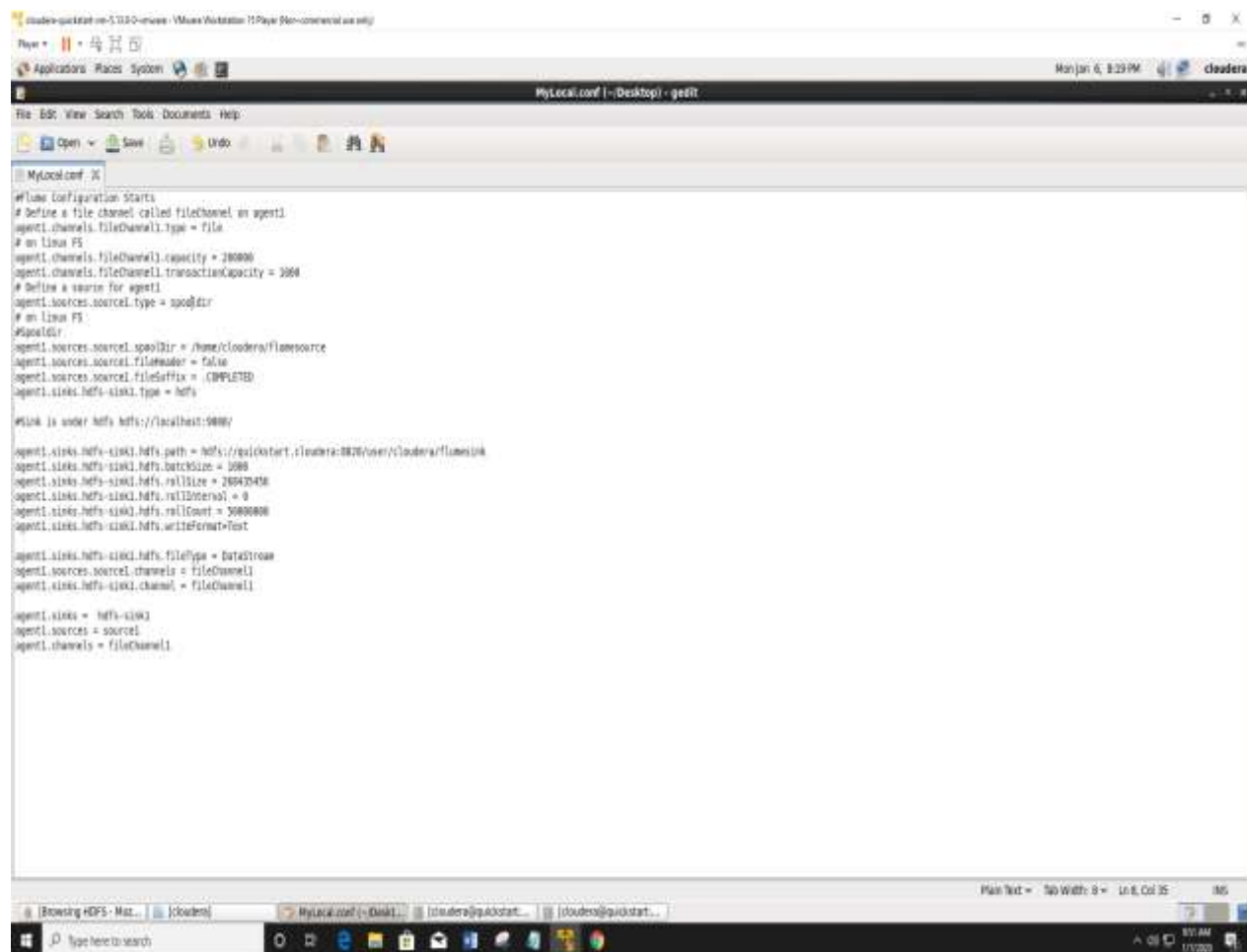
Roll num - **18**

*CREATING THE DIRECTORY FLUMESOURCE AND FLUMESINK*



```
cloudra@quickstart:~/Desktop$ mkdir /home/cloudra/flumesource
cloudra@quickstart:~/Desktop$ mkdir flumesink
cloudra@quickstart:~/Desktop$ hadoop fs -put flumesink /home/cloudra
put: /home/cloudra/: No such file or directory
cloudra@quickstart:~/Desktop$ hadoop fs -put flumesink
cloudra@quickstart:~/Desktop$
```

## CONFIGURE THE FILE CREATED



The screenshot shows a Cloudera Quickstart VM window with a terminal editor open to the file `MyLocal.conf`. The file contains configuration for a file channel and a sink. The configuration is as follows:

```
#Flume Configuration Starts
# Define a file channel called fileChannel on agent1
agent1.channels.fileChannel.type = file
# on linux FS
agent1.channels.fileChannel.capacity = 20000
agent1.channels.fileChannel.transactionCapacity = 2000
# Define a source for agent1
agent1.sources.source.type = spooldir
# on linux FS
#spooldir
agent1.sources.source.spooldir = /home/cloudera/flumesource
agent1.sources.source.fileHeader = false
agent1.sources.source.fileSuffix = .COMPLETED
agent1.sinks.hdfs.sink.type = hdfs

#Sink is under hdfs hdfs://localhost:9800/

agent1.sinks.hdfs-sink1.hdfs.path = hdfs://quickstart.cloudera:8020/user/cloudera/flumesink
agent1.sinks.hdfs-sink1.hdfs.batchSize = 2000
agent1.sinks.hdfs-sink1.hdfs.rollSize = 268435456
agent1.sinks.hdfs-sink1.hdfs.rollInterval = 0
agent1.sinks.hdfs-sink1.hdfs.rollCount = 50000000
agent1.sinks.hdfs-sink1.hdfs.writeFormat=Text

agent1.sinks.hdfs-sink1.hdfs.fileType = DataStream
agent1.sources.source.channels = fileChannel
agent1.sinks.hdfs-sink1.channel = fileChannel

agent1.sinks = hdfs-sink1
agent1.sources = source1
agent1.channels = fileChannel1
```

The bottom of the window shows a taskbar with various application icons and a system clock indicating 11:01 AM on 1/1/2020.

FLUME AGENT QUERY

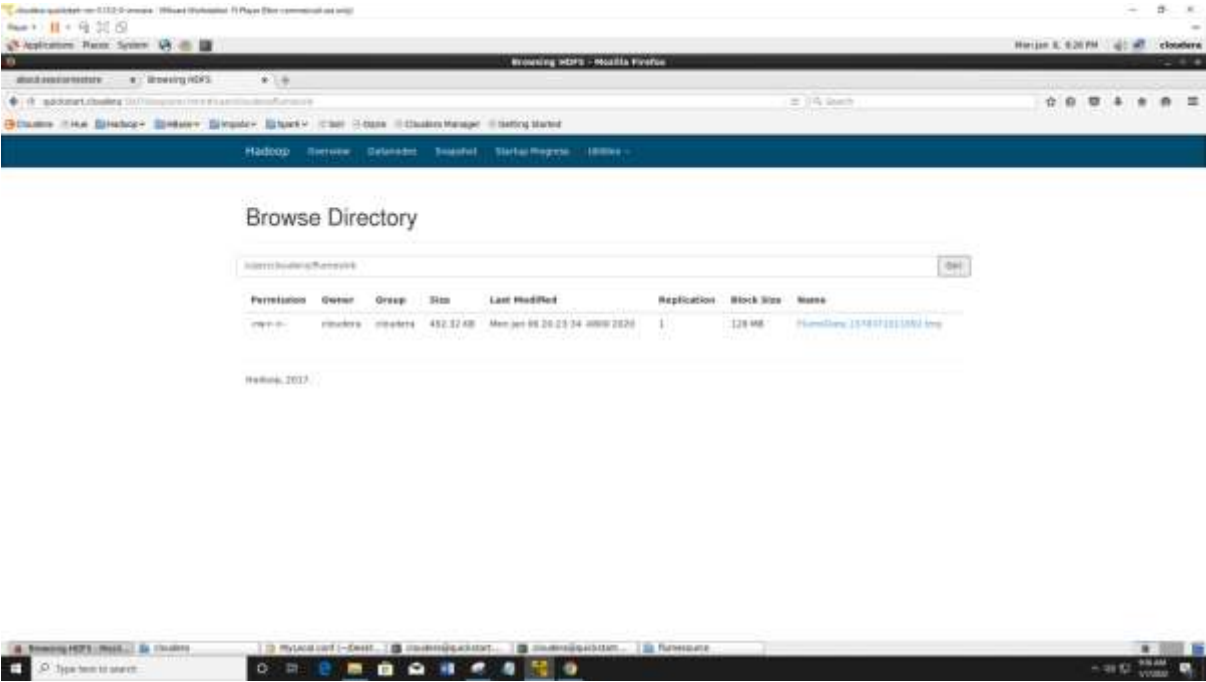
[illegible]

```
File Edit View Search Terminal Help
26/EL/06 20:25:02 INFO conf.FlumeConfiguration: Processing:hdfs-sink1
26/EL/06 20:25:02 INFO conf.FlumeConfiguration: Processing:hdfs-sink1
26/EL/06 20:25:02 INFO conf.FlumeConfiguration: Processing:hdfs-sink1
26/EL/06 20:25:02 INFO conf.FlumeConfiguration: Processing:hdfs-sink1
26/EL/06 20:25:02 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agent1]
26/EL/06 20:25:02 INFO node.AbstractConfigurationProvider: Creating channels
26/EL/06 20:25:02 INFO channel.DefaultChannelFactory: Creating instance of channel fileChannel type file
26/EL/06 20:25:02 INFO node.AbstractConfigurationProvider: Created channel fileChannel1
26/EL/06 20:25:02 INFO source.DefaultSourceFactory: Creating instance of source source1, type spoolDir
26/EL/06 20:25:02 INFO sink.DefaultSinkFactory: Creating instance of sink hdfs-sink1, type: hdfs
26/EL/06 20:25:02 INFO node.AbstractConfigurationProvider: Channel fileChannel1 connected to [source1, hdfs-sink1]
26/EL/06 20:25:02 INFO node.Application: Starting new configuration: sourceNumbers: [source1=DefaultSpoolDirectorySource source1: { spoolDir: /home/cloudera/flumesource } ] sinkNumbers: [hdfs-sink1=sinkNumber
: { policy.org.apache.flume.sink.DefaultSinkInMemorySink counterGroup: { name=roll counters: { } } } ] channels: [fileChannel1=fileChannel fileChannel1: { dataDir: /home/cloudera/flume/file-channel/data } ]
26/EL/06 20:25:02 INFO node.Application: Starting channel fileChannel1
26/EL/06 20:25:02 INFO file.FileChannel: Starting FileChannel fileChannel1: { dataDir: /home/cloudera/flume/file-channel/data } ...
26/EL/06 20:25:02 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: CHANNEL, name: fileChannel1: Successfully registered new MBean.
26/EL/06 20:25:02 INFO instrumentation.MonitoringCounterGroup: Component type: CHANNEL, name: fileChannel1 started
26/EL/06 20:25:02 INFO file.Log: Encryption is not enabled
26/EL/06 20:25:02 INFO file.Log: Replay started
26/EL/06 20:25:02 INFO file.Log: Found NextFileId 2, from [/home/cloudera/flume/file-channel/data/log-1, /home/cloudera/flume/file-channel/data/log-2]
26/EL/06 20:25:02 INFO file.EventBasedCheckpointFileV1: Starting up with /home/cloudera/flume/file-channel/checkpoint/checkpoint and /home/cloudera/flume/file-channel/checkpoint/checkpoint.meta
26/EL/06 20:25:02 INFO file.EventBasedCheckpointFileV1: Reading checkpoint metadata from /home/cloudera/flume/file-channel/checkpoint/checkpoint.meta
26/EL/06 20:25:02 INFO file.FlumeEventQueue: QueueSet population: inserting 0 task 0
26/EL/06 20:25:02 INFO file.Log: Last Checkpoint New Jan 06 20:24:00 PST 2018, queue depth = 0
26/EL/06 20:25:02 INFO file.Log: Replying logs with v2 replay logic
26/EL/06 20:25:02 INFO file.ReplayHandler: Starting replay of [/home/cloudera/flume/file-channel/data/log-1, /home/cloudera/flume/file-channel/data/log-2]
26/EL/06 20:25:02 INFO file.ReplayHandler: Replying /home/cloudera/flume/file-channel/data/log-1
26/EL/06 20:25:02 INFO tools.DirectMemoryUtils: Unable to get maxDirectMemory from VM: java.lang.reflect.InvocationTargetException: sun.misc.VM$unsafeMemoryAccess
26/EL/06 20:25:02 INFO tools.DirectMemoryUtils: Direct Memory Allocation: Allocated = 2640576, Allocated = 0, MaxDirectMemorySize = 10874368, Remaining = 10874368
26/EL/06 20:25:02 INFO file.LogFile: Checkpoint for file:/home/cloudera/flume/file-channel/data/log-1 is: 1573708200782, which is beyond the requested checkpoint time: 1573708020501 and position 0
26/EL/06 20:25:02 INFO file.ReplayHandler: Replying /home/cloudera/flume/file-channel/data/log-2
26/EL/06 20:25:02 INFO file.LogFile: Fast-forward to checkpoint position: 2052000
26/EL/06 20:25:02 INFO file.LogFile: Encountered EOF at 1052400 in /home/cloudera/flume/file-channel/data/log-2
26/EL/06 20:25:02 INFO file.ReplayHandler: read: 0, put: 0, take: 0, rollback: 0, commit: 0, skip: 0, eventCount: 0
26/EL/06 20:25:02 INFO file.FlumeEventQueue: Search Count = 0, Search Time = 0, Copy Count = 0, Copy Time = 0
26/EL/06 20:25:02 INFO file.Log: Rolling /home/cloudera/flume/file-channel/data
26/EL/06 20:25:02 INFO file.Log: Roll start /home/cloudera/flume/file-channel/data
26/EL/06 20:25:02 INFO file.LogFile: Opened /home/cloudera/flume/file-channel/data/log-3
26/EL/06 20:25:02 INFO file.Log: Roll end
26/EL/06 20:25:02 INFO file.EventBasedCheckpointFileV1: Start checkpoint for /home/cloudera/flume/file-channel/checkpoint/checkpoint, elements to sync = 0
26/EL/06 20:25:02 INFO file.EventBasedCheckpointFileV1: Updating checkpoint metadata: logFileId=10: 157371202400, queueSize: 0, queueHead: 500
26/EL/06 20:25:02 INFO file.Log: Updated checkpoint for file:/home/cloudera/flume/file-channel/data/log-3 position: 0 [logFileId=10: 157371202400]
26/EL/06 20:25:02 INFO file.FileChannel: Queue Size after replay: 0 [channel=fileChannel1]
26/EL/06 20:25:02 INFO node.Application: Starting Sink hdfs-sink1
26/EL/06 20:25:02 INFO node.Application: Starting Source source1
26/EL/06 20:25:02 INFO source.SpoolDirectorySource: SpoolDirectorySource: source starting with directory: /home/cloudera/flumesource
26/EL/06 20:25:02 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: SINK, name: hdfs-sink1: Successfully registered new MBean.
26/EL/06 20:25:02 INFO instrumentation.MonitoringCounterGroup: Component type: SINK, name: hdfs-sink1 started
26/EL/06 20:25:02 INFO instrumentation.MonitoringCounterGroup: Monitored counter group for type: SOURCE, name: source1: Successfully registered new MBean.
26/EL/06 20:25:02 INFO instrumentation.MonitoringCounterGroup: Component type: SOURCE, name: source1 started
```

*Dragging File into Flume Source*



*Copying File from File System to Hadoop by using Flume*



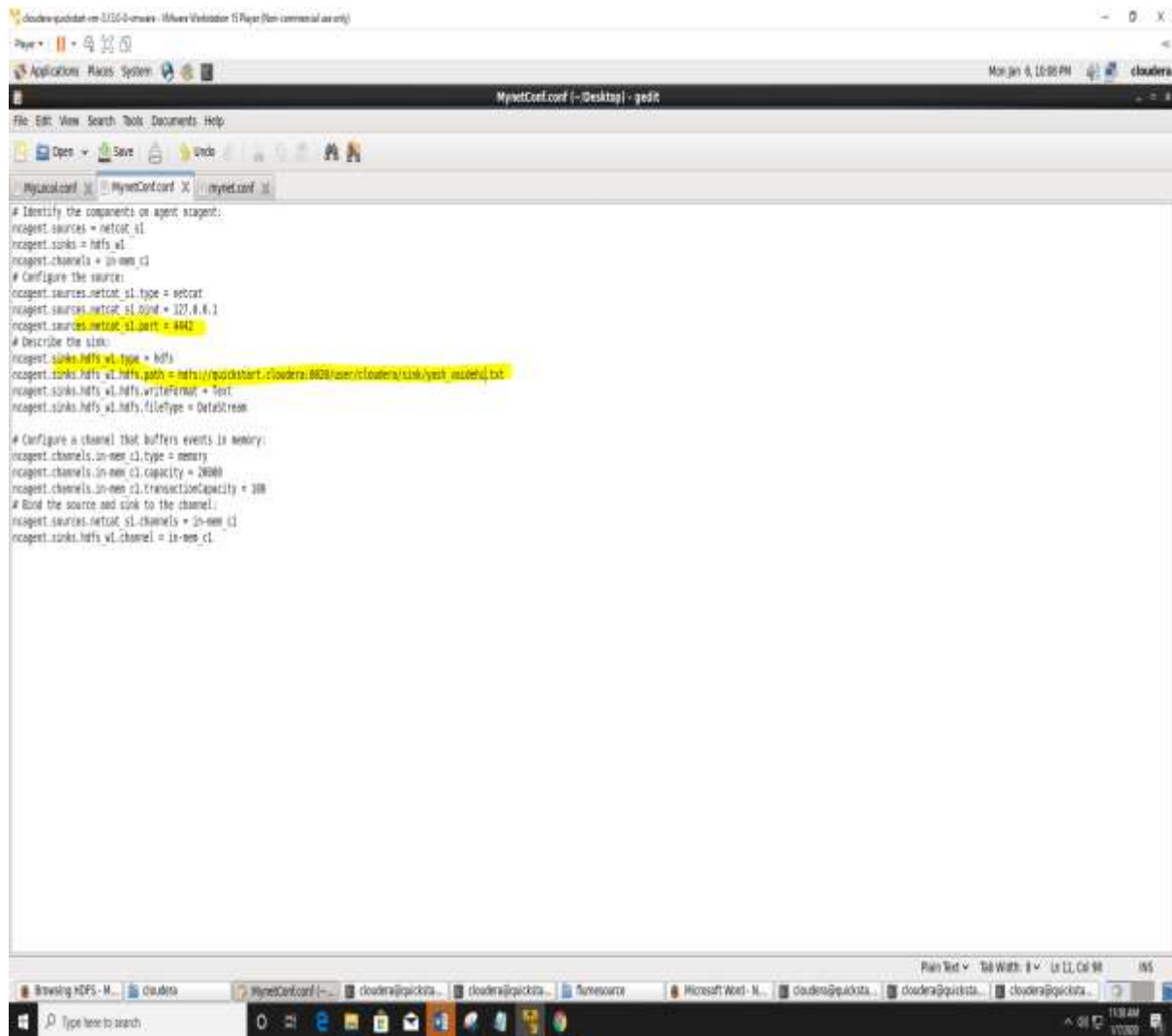
## COPYING FILE FROM FILESYSTEM TO HADOOP



## LAB ASSIGNMENT: QUESTION 2

BY USING NETCAT :

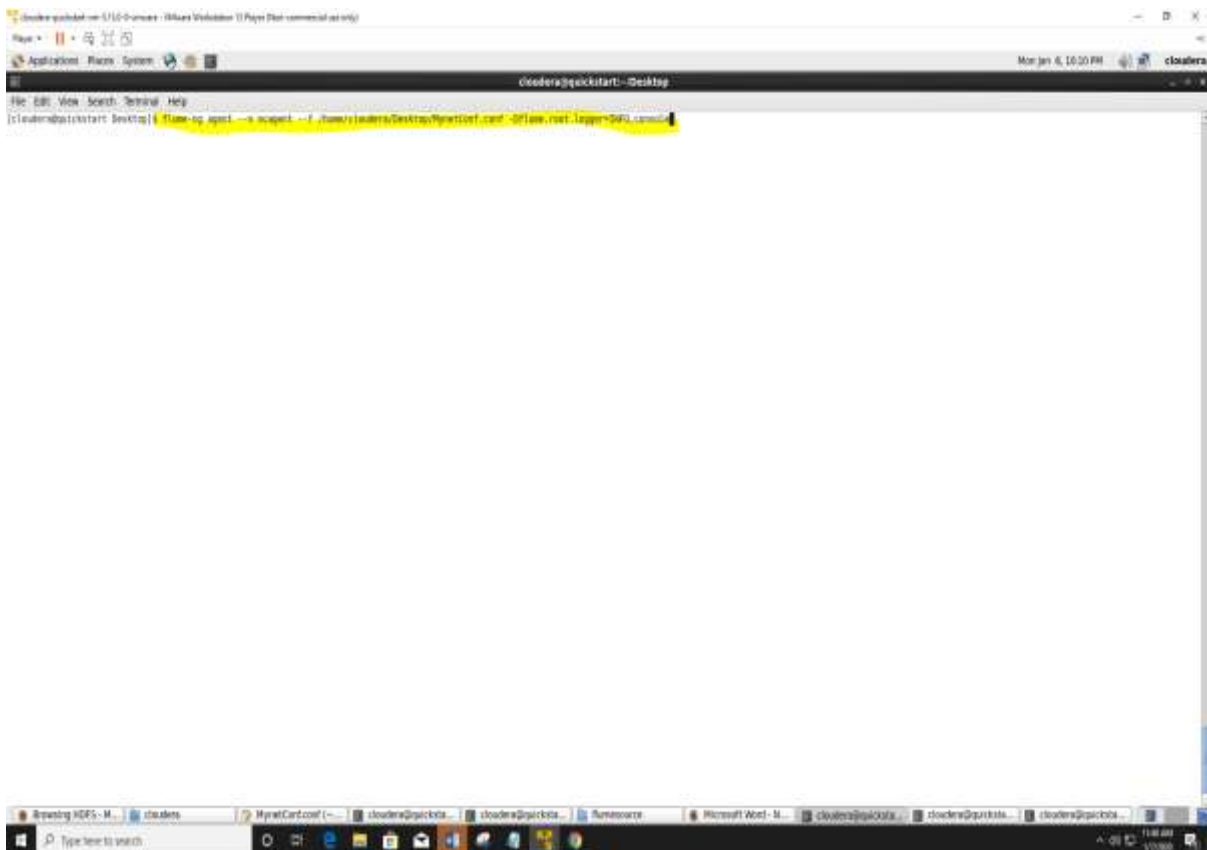
**CONIFG FILE FOR NETCAT PATH UPDATED**



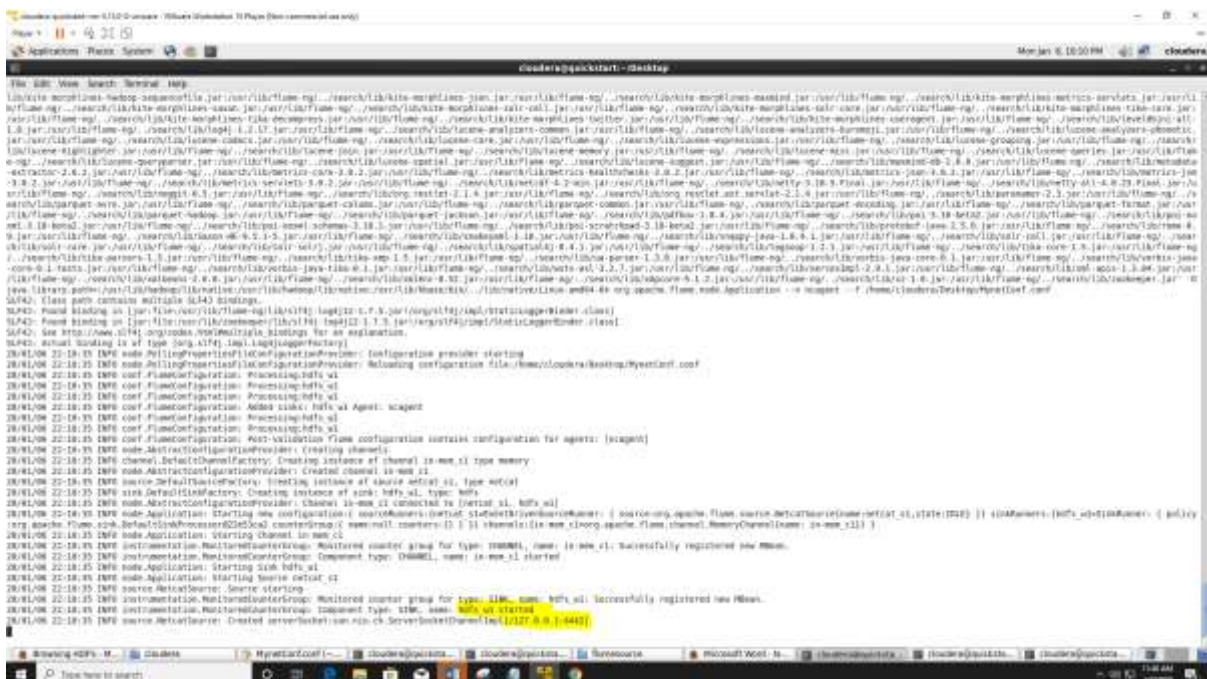
```
# Identify the components on agent:
ncagent.sources = netcat_sl
ncagent.sinks = hdfs_wl
ncagent.channels = in-mem_cl
# Configure the source:
ncagent.sources.netcat_sl.type = netcat
ncagent.sources.netcat_sl.host = 127.0.0.1
ncagent.sources.netcat_sl.port = 4444
# Describe the sink:
ncagent.sinks.hdfs_wl.type = hdfs
ncagent.sinks.hdfs_wl.hdfs.path = hdfs://quickstart.cloudera:8020/user/cloudera/sink/yes/no/echo.txt
ncagent.sinks.hdfs_wl.hdfs.writeFormat = Text
ncagent.sinks.hdfs_wl.hdfs.fileType = DataStream

# Configure a channel that buffers events in memory:
ncagent.channels.in-mem_sl.type = memory
ncagent.channels.in-mem_sl.capacity = 26840
ncagent.channels.in-mem_sl.transactionCapacity = 300
# Bind the source and sink to the channel:
ncagent.sources.netcat_sl.channels = in-mem_sl
ncagent.sinks.hdfs_wl.channel = in-mem_sl
```

## QUERY FOR AGENT



## AGENT STARTED





## CONNECTING SERVER WITH CLIENT



The screenshot shows a terminal window titled "cloudera@quickstart:~" with the following commands and output:

```
File Edit View Search Terminal Help
cloudera@quickstart ~$ curl telnet://127.0.0.1:4442
telnet from valdehl and yash
OK
welcome to telnet
OK
what are you REY
OK
```

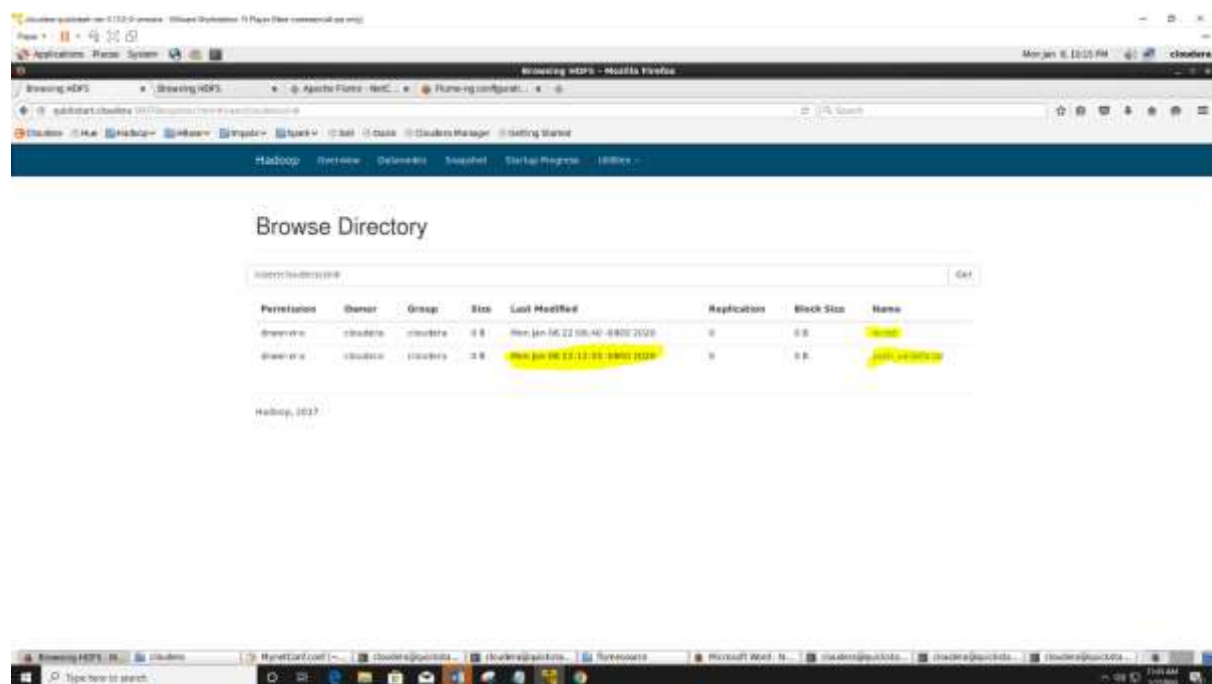
## SERVER ACCEPTED THE REQUEST



The screenshot shows a terminal window titled "cloudera@quickstart:~\$ telnet" with the following output:

```
File Edit View Search Terminal Help
cloudera@quickstart ~$ telnet
telnet from valdehl and yash
OK
welcome to telnet
OK
what are you REY
OK
```

**MESSAGE REFLECTED IN HDFS FILE**



## FILE CHECKING IN HDFS

