Experiment 3: Design Flume agent and load data into Hive

1. **jps**
2. **hdfs dfs -mkdir -p /flume/**
3. Copy the file nethd.conf in your flume directory and add the hdfs path.

**nethd.conf:**

NetcatAgent.sources = Netcat

NetcatAgent.channels = MemChannel

NetcatAgent.sinks = hdfsâsink

NetcatAgent.sources.Netcat.type = netcat

NetcatAgent.sources.Netcat.bind = localhost

NetcatAgent.sources.Netcat.port = 56563

NetcatAgent.sources.Netcat.channels = MemChannel

NetcatAgent.channels.MemChannel.type = memory

NetcatAgent.channels.MemChannel.capacity = 1000

# Define a source on agent and connect to channel memoryChannel.

NetcatAgent.sinks.hdfsâsink.type = hdfs

NetcatAgent.sinks.hdfsâsink.channel = MemChannel

**NetcatAgent.sinks.hdfsâsink.hdfs.path = hdfs://localhost:8020/flume/**

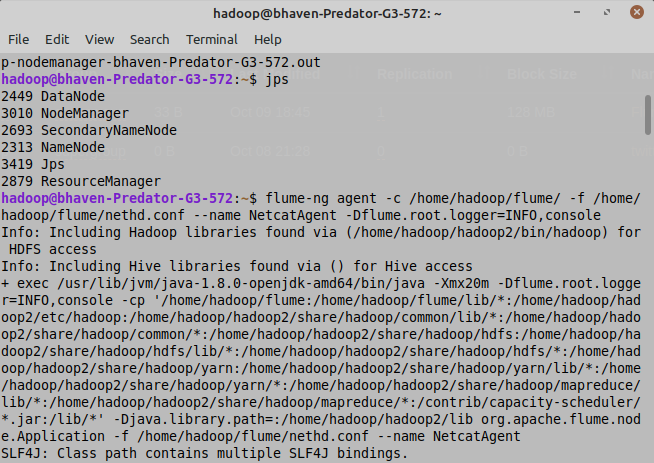
NetcatAgent.sinks.hdfsâsink.hdfs.fileType = DataStream

NetcatAgent.sinks.hdfsâsink.hdfs.writeFormat = Text

NetcatAgent.sinks.hdfsâsink.hdfs.filePrefix=

NetcatAgent.sinks.hdfsâsink.hdfs.fileSuffix=.txt

4. Execute the file using flume-ng agent:



**flume-ng agent -c /hom****e/hadoop/flume/ -f /home/hadoop/flume/nethd.conf –name NetcatAgent -Dflume.root.logger=INFO,console**

**5.** Open another Terminal and type following as mentioned in nethd.conf file:

**telnet localhost 56563**

(enter again if unable to connect)

then enter numbers eg. 1 <press enter key> 2 enter 8 enter 9 enter

After connecting to localhost, type anything you want and press enter after each line of input.

Then, terminate the process using Ctrl-C after entering desired inputs.

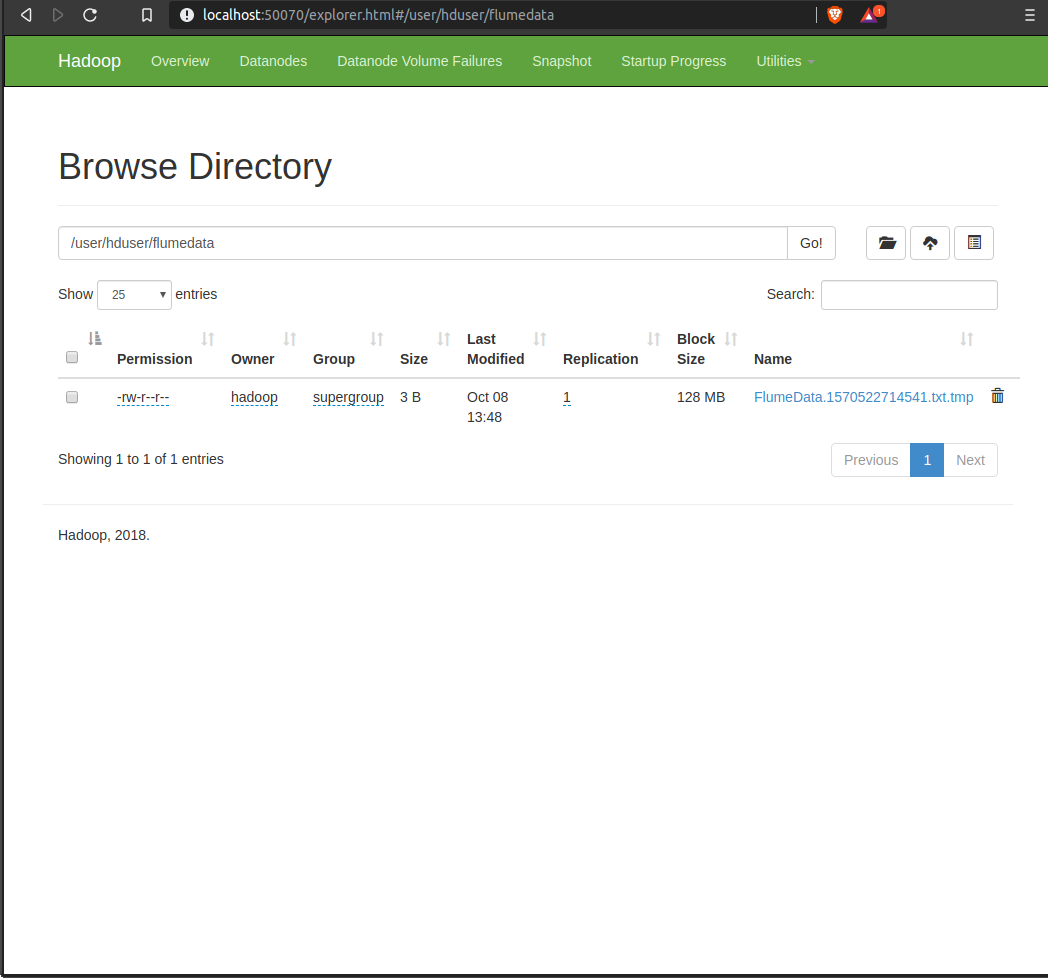
6. You can see the FlumeData file in hdfs as:

Note: Your path will be one created above, so move to that path.

**Open localhost:50070/**

**Browse Directory:**

**/user/hduser/flumedata**



7. Now, load this file into hive. So, create a table in hive with the inputs entered in telnet localhost 56563. In this case, SSN and Name.

**hive**

**(hive> opens)**

**show tables;**

**create external table data(SSN int, Name string)**

**row format delimited**

**fields terminated by ‘ ‘**

**lines terminated by ‘\n’;**

**load data inpath ‘/flume/FlumeData.\*’ into table data;**

**select \* from data;**