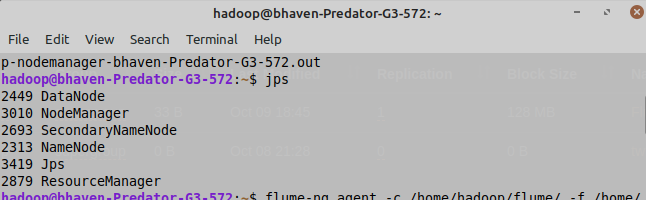
Experiment 3: Design Flume agent and load data into Hive

Steps:

1. Start Hadoop daemons:



1. Create a folder in hdfs to store flume data.

hdfs dfs -mkdir -p /flume/

1. 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/**

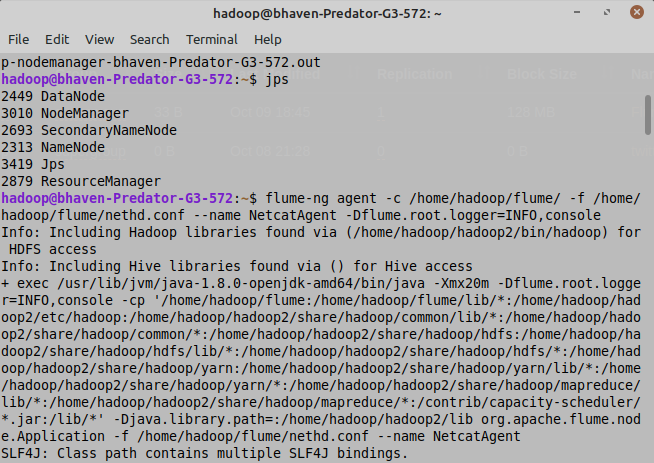
NetcatAgent.sinks.hdfsâsink.hdfs.fileType = DataStream

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

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

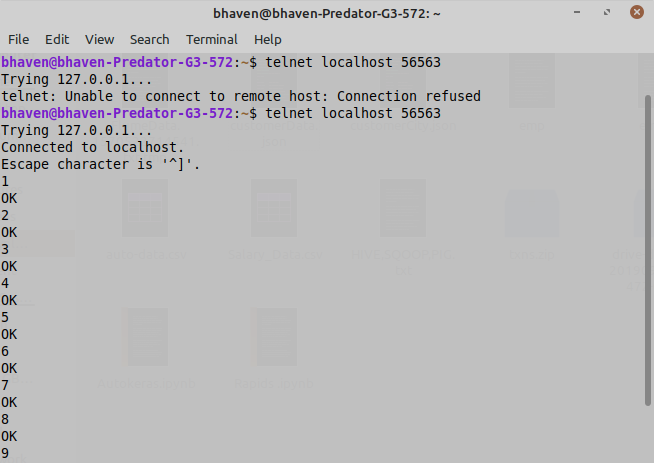
NetcatAgent.sinks.hdfsâsink.hdfs.fileSuffix=.txt

1. Execute the file using flume-ng agent:



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

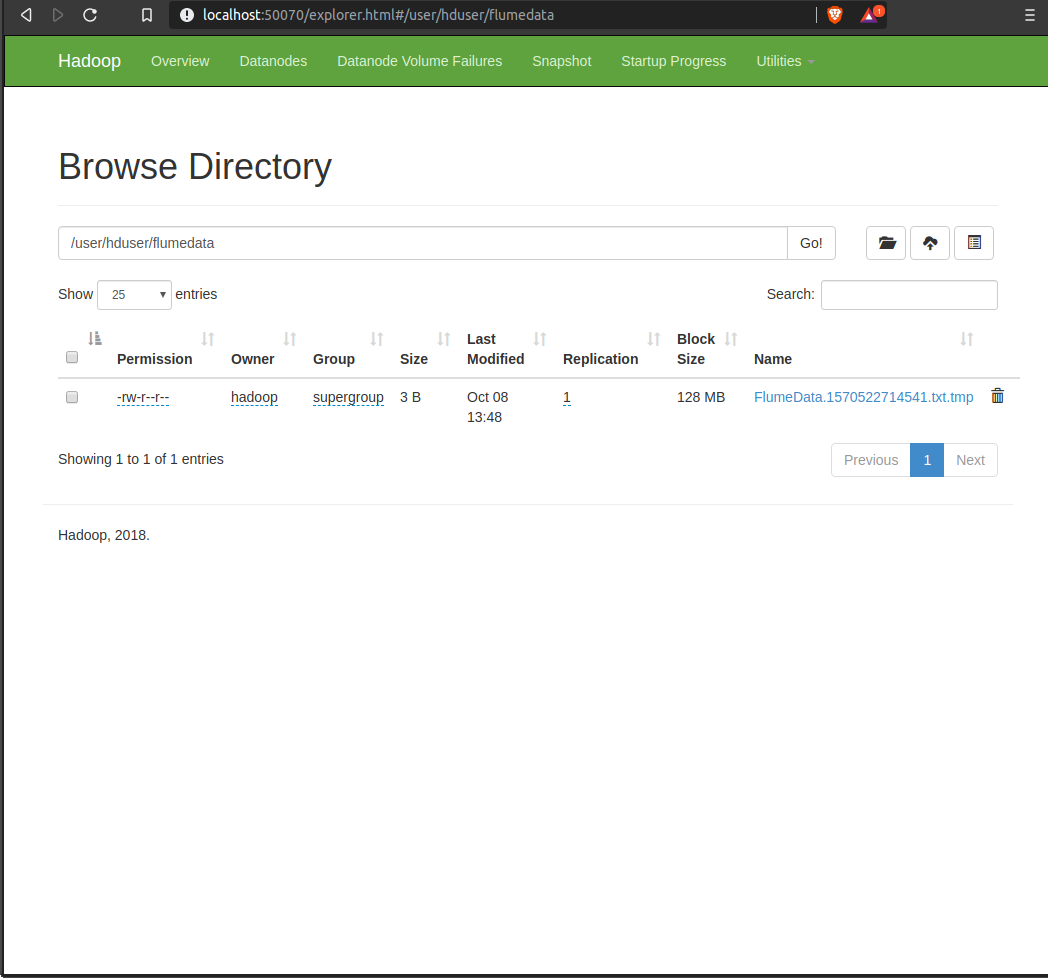
telnet localhost 56563



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.

1. You can see the FlumeData file in hdfs as show below:

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



1. Now, we have to load this file into hive. So, to do that first create a table in hive with the inputs entered in telnet localhost 56563. In this case, SSN and Name.

