```
Lab:Load Log data into HDFS using Flume ===================================================================================================================
CopyRight - Big Data Trunk LLC
www.BigDataTrunk.com
Twitter - @BigDataTrunk


====================================================================================================================


Use this command reference file to copy and paste text for your lab.


====================================================================================================================
Instructions:


1. Take Product_Ratings  files in cvs format
2.  Use Flume to move the cvs files into HDFS
3. Verify the output on HDFS
====================================================================================================================


#Hands on Lab for moving Log datas to HDFS using Flume
====================================================================================================================
We have cvs files that contains Product Ratings logs.We have to load those files into HDFS using spooling concept in FLUME
====================================================================================================================
# Goto Places -> Home Folder and create a new folder named spoolDir
# copy and paste the Products_Ratings.cvs file into previously created spoolDir folder


# Move to conf folder of flume-ng
#Command:[cloudera@quickstart ~]$ cd /etc/flume-ng/conf
cd /etc/flume-ng/conf


#Switch user to root to edit flume.conf
#Command: [cloudera@quickstart conf]$ su root
su root
#Note: Password: cloudera (password will not visible but we have to type)




# Open flume.conf file to edit the file
#Command: [root@quickstart conf]# gedit flume.conf
gedit flume.conf


#Paste the following code in flume.conf file




    agent.sources = SpoolExampleDir
    agent.channels = memoryChannel
    agent.sinks = flumeHDFS


    # Setting the source to spool directory where the file exists
    agent.sources.SpoolExampleDir.type = spooldir
    agent.sources.SpoolExampleDir.spoolDir = /home/cloudera/spoolDir/


    # Setting the channel to memory
    agent.channels.memoryChannel.type = memory


    # Max number of events stored in the memory channel
    agent.channels.memoryChannel.capacity = 100000
    agent.channels.memoryChannel.transactioncapacity = 10000


    # Setting the sink to HDFS
    agent.sinks.flumeHDFS.type = hdfs
    agent.sinks.flumeHDFS.hdfs.path = hdfs://quickstart.cloudera:8020/user/flume/system.log/
    agent.sinks.flumeHDFS.hdfs.fileType = DataStream


    # Write format can be text or writable
    agent.sinks.flumeHDFS.hdfs.writeFormat = Text


    # use a single file at a time
    agent.sinks.flumeHDFS.hdfs.maxOpenFiles = 1


    # rollover file based on maximum size of 10 MB
    agent.sinks.flumeHDFS.hdfs.rollSize = 10485760


    # never rollover based on the number of events
    agent.sinks.flumeHDFS.hdfs.rollCount = 0


    # rollover file based on max time of 1 min
    agent.sinks.flumeHDFS.hdfs.rollInterval = 60


    # Connect source and sink with channel
    agent.sources.SpoolExampleDir.channels = memoryChannel
    agent.sinks.flumeHDFS.channel = memoryChannel

# Start the flume agent using following command
 flume-ng agent --conf conf --conf-file ./flume.conf --name agent -Dflume.root.logger=INFO,console


# Goto Places -> Home Folder -> spoolDir.You will notice Product_Ratings.cvs had become  Product_Ratings.cvs.COMPLETED
# Copy Product_Ratings1.cvs  into the spoolDir folder.
# As soon as you copy the file,the file name changes toProduct_Ratings1.cvs.COMPLETED as the flume agent is running.
# Goto Hue in Firefox,Click File Browser and go to user/flume/system.log. You will see the flume data that has been downloaded
```