Vedant Jolly
BE Comps, 2019130026
BDA, Batch A

EXPERIMENT NO. 5 FLUME

AIM: Insert live streaming data into HDFS.

Starting Hadoop Services:

```
D:\hadoop-env\hadoop-3.2.1\sbin>.\start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

D:\hadoop-env\hadoop-3.2.1\sbin>jps
24288 ResourceManager
2836 NodeManager
6872 DataNode
9320 Jps
12028 NameNode

D:\hadoop-env\hadoop-3.2.1\sbin>
```

Check the Flume version if it is installed properly:

```
C:\apache-flume-1.10.1-bin\bin>flume-ng version

C:\apache-flume-1.10.1-bin\bin>powershell.exe -NoProfile -InputFormat none -ExecutionPolicy unrestricted -File C:\apache
-flume-1.10.1-bin\bin\flume-ng.ps1 version

WARN: Config directory not set. Defaulting to C:\apache-flume-1.10.1-bin\conf
Sourcing environment configuration script C:\apache-flume-1.10.1-bin\conf\flume-env.ps1
Including Hadoop libraries found in (D:\hadoop-env\hadoop-3.2.1) for DFS access
WARN: HBASE_HOME not found
WARN: HIVE_HOME not found

  Running FLUME version :
    class: org.apache.flume.tools.VersionInfo
    arguments:

Flume 1.10.1
Source code repository: https://git.apache.org/repos/asf/flume.git
Revision: 047516d4bd5574c3e67a5d98ca2cfe025886df7c
Compiled by rgoers on Sat Aug 13 11:16:08 MST 2022
From source with checksum de1cf990338c759d311522e65597e457
C:\apache-flume-1.10.1-bin\bin>
```

seq.conf

Naming the components on the current agent

SeqGenAgent.sources = SeqSource
SeqGenAgent.channels = MemChannel
SeqGenAgent.sinks = HDFS

# Describing/Configuring the source
SeqGenAgent.sources.SeqSource.type = seq

# Describing/Configuring the sink
SeqGenAgent.sinks.HDFS.type = hdfs
SeqGenAgent.sinks.HDFS.hdfs.path = hdfs://localhost:9820/user/Hadoop/seqgen_data/
SeqGenAgent.sinks.HDFS.hdfs.filePrefix = log
SeqGenAgent.sinks.HDFS.hdfs.rollInterval = 0
SeqGenAgent.sinks.HDFS.hdfs.rollCount = 10000
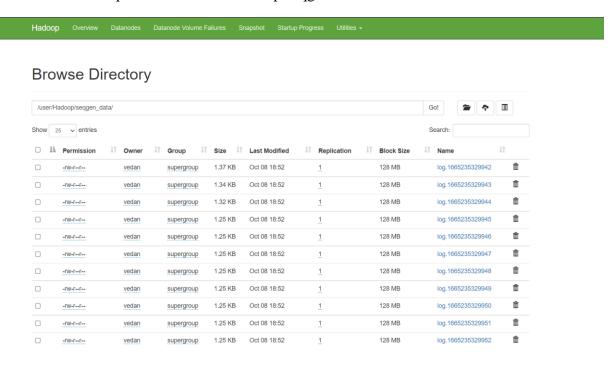SeqGenAgent.sinks.HDFS.hdfs.fileType = DataStream

# Describing/Configuring the channel
SeqGenAgent.channels.MemChannel.type = memory
SeqGenAgent.channels.MemChannel.capacity = 1000
SeqGenAgent.channels.MemChannel.transactionCapacity = 100
agent.channels.memoryChannel.keep-alive = 60

# Binding the source and sink to the channel
SeqGenAgent.sources.SeqSource.channels = MemChannel
SeqGenAgent.sinks.HDFS.channel = MemChannel

#Modify the maximum memory size of java vi bin/flume-ng
JAVA_OPTS ="-Xmx2048m"

flume-ng agent -n SeqGenAgent -c conf -f
C:\apache-flume-1.9.0-bin\conf\seq.conf -property flume.root.logger=DEBUG,console

http://localhost:9870/explorer.html#/user/Hadoop/seqgen_data/

CONCLUSION:

I learnt in this experiment how to stream data live using in HDFS using Flume in order to
analyze the current data.