

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"
```

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

```
C:\apache-flume-1.9.0-bin> flume-ng agent -n SeqGenAgent -c conf -f C:\apache-flume-1.9.0-bin\conf\seq.conf -property flume.root.logger=DEBUG,console
C:\apache-flume-1.9.0-bin> powershell.exe -NoProfile -InputFormat none -ExecutionPolicy unrestricted -File C:\apache-flume-1.9.0-bin\bin\flume-ng.ps1 agent -n SeqGenAgent -c
conf -f C:\apache-flume-1.9.0-bin\conf\seq.conf -property flume.root.logger=DEBUG,console

Security warning
Run only scripts that you trust. While scripts from the internet can be useful, this script can potentially harm your computer. If you trust this script, use the
Unblock-File cmdlet to allow the script to run without this warning message. Do you want to run C:\apache-flume-1.9.0-bin\bin\flume-ng.ps1?
[D] Do not run [R] Run once [S] Suspend [?] Help (default is "D"): R

WARN: Config directory not set. Defaulting to C:\apache-flume-1.9.0-bin\conf
Sourcing environment configuration script C:\apache-flume-1.9.0-bin\conf\flume-env.ps1

Security warning
Run only scripts that you trust. While scripts from the internet can be useful, this script can potentially harm your computer. If you trust this script, use the
Unblock-File cmdlet to allow the script to run without this warning message. Do you want to run C:\apache-flume-1.9.0-bin\conf\flume-env.ps1?
[D] Do not run [R] Run once [S] Suspend [?] Help (default is "D"): R
Including Hadoop libraries found in (F:\Hadoop\hadoop-3.2.2) for DFS access
WARN: HBASE_HOME not found
WARN: HIVE_HOME not found

Running FLUME agent :
  class: org.apache.flume.node.Application
  arguments: -n SeqGenAgent -f "C:\apache-flume-1.9.0-bin\conf\seq.conf"

SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/apache-flume-1.9.0-bin/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/F:/Hadoop/hadoop-3.2.2/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2022-10-05 19:37:17,456 (main) [DEBUG] - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95) No global SSL keystore path specified.
2022-10-05 19:37:17,460 (main) [DEBUG] - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95) No global SSL keystore password specified.
2022-10-05 19:37:17,461 (main) [DEBUG] - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95) No global SSL keystore type specified.
2022-10-05 19:37:17,462 (main) [DEBUG] - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95) No global SSL truststore path specified.
2022-10-05 19:37:17,463 (main) [DEBUG] - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95) No global SSL truststore password specified.
2022-10-05 19:37:17,464 (main) [DEBUG] - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95) No global SSL truststore type specified.
2022-10-05 19:37:17,468 (main) [DEBUG] - org.apache.flume.util.SSLUtil.initSysPropFromEnvVar(SSLUtil.java:95) No global SSL include protocols specified.
```

[http://localhost:9870/explorer.html#/user/Hadoop/seqgen\\_data/](http://localhost:9870/explorer.html#/user/Hadoop/seqgen_data/)

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot




Startup Progress

Utilities

# Browse Directory





















/user/Hadoop/seqgen\_data/

Go!



Show25entries

Search:

| <input type="checkbox"/> |  Permission |  Owner |  Group |  Size |  Last Modified |  Replication |  Block Size |  Name |  |
|--------------------------|--|---|---|--|---|---|---|--|---|
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.37 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329942</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.34 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329943</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.32 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329944</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.25 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329945</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.25 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329946</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.25 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329947</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.25 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329948</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.25 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329949</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.25 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329950</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.25 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329951</a>  |  |
| <input type="checkbox"/> | <a href="#">-rw-r--r--</a>   | <a href="#">vedan</a>   | <a href="#">supergroup</a>  | 1.25 KB  | Oct 08 18:52  | <a href="#">1</a>   | 128 MB  | <a href="#">log.1665235329952</a>  |  |

## CONCLUSION:

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