**My Hadoop Workouts - Flume**

# Flume:

## Capture the real-time incremental logs generated by devices in the corresponding device servers.

**Run the below command to start the flume agent:**

flume-ng agent -n a1 -c conf -f $FLUME\_HOME/conf/tail\_exec.conf

**tail\_exec.conf**

a1.sources = r1

a1.sinks = k1

a1.channels = c1

a1.sources.r1.type = exec

a1.sources.r1.command = tail -F /home/hduser/device1/log.txt

a1.channels.c1.type = memory

a1.channels.c1.capacity = 1000

a1.channels.c1.transactionCapacity = 100

a1.sinks.k1.type=hdfs

a1.sinks.k1.hdfs.path = hdfs://localhost:54310/flume/tail\_output/

a1.sinks.k1.hdfs.fileType=DataStream

a1.sinks.k1.hdfs.writeformat=Text

a1.sinks.k1.hdfs.batchSize=1080

a1.sinks.k1.hdfs.rollSize=25

a1.sinks.k1.hdfs.rollCount=10800

a1.sinks.k1.hdfs.rollInterval=60

a1.sources.r1.channels = c1

a1.sinks.k1.channel = c1

## Capture the console log and save in HDFS instantaneously.

**Run the below command to start the flume agent a1 in one terminal:**

flume-ng agent -n a1 -c conf -f $FLUME\_HOME/conf/hdfs.conf

**hdfs.conf**

a1.sources = r1

a1.sinks = k1

a1.channels = c1

a1.sources.r1.type = netcat

a1.sources.r1.bind = localhost

a1.sources.r1.port = 44444

a1.sources.r1.interceptors = i1

a1.sources.r1.interceptors.i1.type = host

a1.sources.r1.interceptors.i1.hostheader = hostname

a1.channels.c1.type = memory

a1.channels.c1.capacity = 1000

a1.channels.c1.transactionCapacity = 100

a1.sinks.k1.type=hdfs

a1.sinks.k1.hdfs.path = hdfs://localhost:54310/flume/hdfs\_output/

a1.sinks.k1.hdfs.fileType=DataStream

a1.sinks.k1.hdfs.writeformat=Text

a1.sinks.k1.hdfs.batchSize=1080

a1.sinks.k1.hdfs.rollSize=800

a1.sinks.k1.hdfs.rollCount=10800

a1.sinks.k1.hdfs.rollInterval=600

a1.sources.r1.channels = c1

a1.sinks.k1.channel = c1

**Run the below command in other terminal to input events.**

curl telnet://localhost:44444

## Listen a terminal for console logs

**Run the below command to start the flume agent a1 in one terminal:**

flume-ng agent -n a1 -c conf -f $FLUME\_HOME/conf/netcat\_flume.conf

**netcat\_flume.conf**

# example.conf: A single-node Flume configuration

# Name the components on this agent

a1.sources = r1

a1.sinks = k1

a1.channels = c1

# Describe/configure the source

a1.sources.r1.type = netcat

a1.sources.r1.bind = localhost

a1.sources.r1.port = 44444

a1.sources.r1.interceptors = i1

a1.sources.r1.interceptors.i1.type = host

a1.sources.r1.interceptors.i1.hostheader = hostname

# Describe the sink

a1.sinks.k1.type = logger

# Use a channel which buffers events in memory

a1.channels.c1.type = memory

a1.channels.c1.capacity = 1000

a1.channels.c1.transactionCapacity = 100

# Bind the source and sink to the channel

a1.sources.r1.channels = c1

a1.sinks.k1.channel = c1

**Run the below command in other terminal to input events.**

curl telnet://localhost:44444

## Workout Log for Exercise 1, 2 & 3

[E:\Chicago\_E\_Drive\Deepak DOCs\Study\Hadoop\WO\WO Logs\Flume\_Work-00.txt](WO%20Logs/Flume_Work-00.txt)

## Retrieve tweets from twitter

* 1. **Create a App in twitter & note down the below details**

[E:\Chicago\_E\_Drive\Deepak DOCs\Study\Hadoop\WO\WO Logs\Flume\_twitter-01.txt](WO%20Logs/Flume_twitter-01.txt)

Consumer key:

Z6ZNPyDA6deV8KYMskIhEM4Ze

Consumer secret:

SNtknAEpzEnYv4wMebQF1PseaGlqUq58kUHvdOc8IZN45miP93

Remember this should not be shared.

Access token:

145580641-rpfUsxXoSfnadFdTdX9BNTcJ9YHthVwumkGuaDR2

Access token secret:

VPV2gzkgN4xXqsutBj0WPuILazG9b3GTkqviuQXfI6SlU

* 1. **twitter.conf**

# Naming the components on the current agent.

TwitterAgent.sources = Twitter

TwitterAgent.channels = MemChannel

TwitterAgent.sinks = HDFS

# Describing/Configuring the source.

#TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource

TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource

TwitterAgent.sources.Twitter.consumerKey = Z6ZNPyDA6deV8KYMskIhEM4Ze

TwitterAgent.sources.Twitter.consumerSecret = SNtknAEpzEnYv4wMebQF1PseaGlqUq58kUHvdOc8IZN45miP93

TwitterAgent.sources.Twitter.accessToken = 145580641-rpfUsxXoSfnadFdTdX9BNTcJ9YHthVwumkGuaDR2

TwitterAgent.sources.Twitter.accessTokenSecret = VPV2gzkgN4xXqsutBj0WPuILazG9b3GTkqviuQXfI6SlU

TwitterAgent.sources.Twitter.keywords = Deepak, Data, Scientist, Hadoop, Manchester United, Sachin, Nasa, Messi, Federer

# Describing/Configuring the sink.

TwitterAgent.sinks.HDFS.type=hdfs

TwitterAgent.sinks.HDFS.hdfs.path = hdfs://localhost:54310/user/hduser/twitter\_data/

TwitterAgent.sinks.HDFS.hdfs.fileType=DataStream

TwitterAgent.sinks.HDFS.hdfs.writeformat=Text

TwitterAgent.sinks.HDFS.hdfs.batchSize=10

TwitterAgent.sinks.HDFS.hdfs.rollSize=0

TwitterAgent.sinks.HDFS.hdfs.rollCount=10

TwitterAgent.sinks.HDFS.hdfs.rollInterval=10

TwitterAgent.sinks.HDFS.hdfs.filePrefix = DKM\_Twitter\_App01-V2

# Describing/Configuring the channel.

TwitterAgent.channels.MemChannel.type = memory

TwitterAgent.channels.MemChannel.capacity = 10

TwitterAgent.channels.MemChannel.transactionCapacity = 10

# Binding the source and sink to the channel.

TwitterAgent.sources.Twitter.channels = MemChannel

TwitterAgent.sinks.HDFS.channel = MemChannel

* 1. **Copy the necessary jars to the flume lib**

[hduser@Inceptez lib]$ pwd

/usr/local/flume/lib

[hduser@Inceptez lib]$ cp /usr/local/nifi/work/nar/extensions/nifi-flume-nar-0.6.0.nar-unpacked/META-INF/bundled-dependencies/twitter4j-\* $PWD

[hduser@Inceptez ~]$ cp /usr/local/nifi/work/nar/extensions/nifi-flume-nar-0.6.0.nar-unpacked/META-INF/bundled-dependencies/flume-twitter-source-1.6.0.jar /usr/local/flume/lib/

[hduser@Inceptez ~]$ ls /usr/local/flume/lib/\*twi\*

/usr/local/flume/lib/cdk-morphlines-twitter-0.4.0.jar /usr/local/flume/lib/twitter4j-media-support-3.0.3.jar

/usr/local/flume/lib/flume-twitter-source-1.6.0.jar /usr/local/flume/lib/twitter4j-stream-3.0.3.jar

/usr/local/flume/lib/twitter4j-core-3.0.3.jar

* 1. **Run the below command to start the flume agent a1 in one terminal:**

flume-ng agent -n TwitterAgent -c conf -f $FLUME\_HOME/conf/twitter.conf

The tweets will get saved in the HDFS path specified in the conf file.

[hduser@Inceptez ~]$ hdfs dfs -mkdir twitter\_data\_hive\_01

[hduser@Inceptez ~]$ hdfs dfs -cp twitter\_data/\*V2\* twitter\_data\_hive\_01

[hduser@Inceptez ~]$ hdfs dfs -ls twitter\_data\_hive\_01

Found 24 items

-rw-r--r-- 1 hduser supergroup 37355 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063122

-rw-r--r-- 1 hduser supergroup 43993 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063123

-rw-r--r-- 1 hduser supergroup 51640 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063124

-rw-r--r-- 1 hduser supergroup 53015 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063125

-rw-r--r-- 1 hduser supergroup 56410 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063126

-rw-r--r-- 1 hduser supergroup 47722 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063127

-rw-r--r-- 1 hduser supergroup 56345 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063128

-rw-r--r-- 1 hduser supergroup 43941 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063129

-rw-r--r-- 1 hduser supergroup 53330 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063130

-rw-r--r-- 1 hduser supergroup 40426 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063131

-rw-r--r-- 1 hduser supergroup 45516 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063132

-rw-r--r-- 1 hduser supergroup 51093 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063133

-rw-r--r-- 1 hduser supergroup 51476 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063134

-rw-r--r-- 1 hduser supergroup 48466 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063135

-rw-r--r-- 1 hduser supergroup 49794 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063136

-rw-r--r-- 1 hduser supergroup 43317 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063137

-rw-r--r-- 1 hduser supergroup 55714 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063138

-rw-r--r-- 1 hduser supergroup 45590 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063139

-rw-r--r-- 1 hduser supergroup 51376 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063140

-rw-r--r-- 1 hduser supergroup 57721 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063141

-rw-r--r-- 1 hduser supergroup 40796 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063142

-rw-r--r-- 1 hduser supergroup 50561 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063143

-rw-r--r-- 1 hduser supergroup 52226 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063144

-rw-r--r-- 1 hduser supergroup 17439 2016-08-16 23:33 twitter\_data\_hive\_01/DKM\_Twitter\_App01-V2.1471370063145