Assignment Submission- Session 12

Task:

Create a flume agent that streams data from Twitter and stores in the HDFS.

Solution:

- Creating a configuration file and placing inside "conf" folder where flume is installed.

"Configuration file"

 # Define a memory channel called ch1 on agent1

agent1.channels.ch1.type = memory

agent1.sources.spool.type = spooldir

agent1.sources.spool.channels = ch1

agent1.sources.spool.spoolDir = /home/acadgild/install/flume/apache-flume-1.8.0-bin/source_sink

agent1.sources.spool.fileHeader = true

agent1.sinks.fr1.type = file_roll

agent1.sinks.fr1.channel = ch1

agent1.sinks.fr1.sink.directory = /home/acadgild/install/flume/apache-flume-1.8.0-bin/destination_sink

agent1.channels = ch1

agent1.sources = spool

agent1.sinks = fr1

- **Creating source_sink and destinataion_sink folders inside flume folder.**

- **Starting flume: Command used is as below**

flume-ng agent -n agent1 -f /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf/TwitterData.conf

Spool started successfully

```
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j
/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
18/10/03 22:53:59 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider starting
18/10/03 22:53:59 INFO node.PollingPropertiesFileConfigurationProvider: Reloading configuration file:/home/acadgild/install/f
lume/apache-flume-1.8.0-bin/conf/TwitterData.conf
18/10/03 22:53:59 INFO conf.FlumeConfiguration: Processing:fr1
18/10/03 22:53:59 INFO conf.FlumeConfiguration: Processing:fr1
18/10/03 22:53:59 INFO conf.FlumeConfiguration: Added sinks: fr1 Agent: agent1
18/10/03 22:53:59 INFO conf.FlumeConfiguration: Processing:fr1
18/10/03 22:53:59 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agent
1]
18/10/03 22:53:59 INFO node.AbstractConfigurationProvider: Creating channels
18/10/03 22:53:59 INFO channel.DefaultChannelFactory: Creating instance of channel ch1 type memory
18/10/03 22:53:59 INFO node.AbstractConfigurationProvider: Created channel ch1
18/10/03 22:53:59 INFO source.DefaultSourceFactory: Creating instance of source spool, type spooldir
18/10/03 22:53:59 INFO sink.DefaultSinkFactory: Creating instance of sink: fr1, type: file_roll
18/10/03 22:53:59 INFO node.AbstractConfigurationProvider: Channel ch1 connected to [spool, fr1]
18/10/03 22:53:59 INFO node.Application: Starting new configuration:{ sourceRunners:{spool=EventDrivenSourceRunner: { source:
Spool Directory source spool: { spoolDir: /home/acadgild/install/flume/apache-flume-1.8.0-bin/source_sink } }} sinkRunners:{f
r1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@7230e3f counterGroup:{ name:null counters:{} } }} channels
:{ch1=org.apache.flume.channel.MemoryChannel{name: ch1}} }
18/10/03 22:53:59 INFO node.Application: Starting Channel ch1
18/10/03 22:53:59 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: ch1: Successfu
lly registered new MBean.
18/10/03 22:53:59 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: ch1 started
18/10/03 22:53:59 INFO node.Application: Starting Sink fr1
18/10/03 22:53:59 INFO node.Application: Starting Source spool
18/10/03 22:53:59 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /home/acadgild/insta
ll/flume/apache-flume-1.8.0-bin/source_sink
18/10/03 22:53:59 INFO sink.RollingFileSink: Starting org.apache.flume.sink.RollingFileSink{name:fr1, channel:ch1}...
18/10/03 22:53:59 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: fr1: Successfully
 registered new MBean.
18/10/03 22:53:59 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: fr1 started
18/10/03 22:53:59 INFO sink.RollingFileSink: RollingFileSink fr1 started.
18/10/03 22:53:59 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: spool: Successf
ully registered new MBean.
18/10/03 22:53:59 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: spool started
```

Testing by dropping files in "source_sink" folder

Files getting created inside the "destination_sink" folder.

File   Edit   View   Places   Help

1538587439287-1   1538587439287-2   1538587439287-3   1538587439287-4   1538587439287-5   1538587439287-6

1538587439287-7   1538587439287-8   1538587439287-9   1538587439287-10   1538587439287-11