

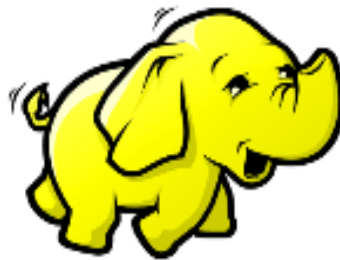


Wisdom Sprouts

We foster knowledge here

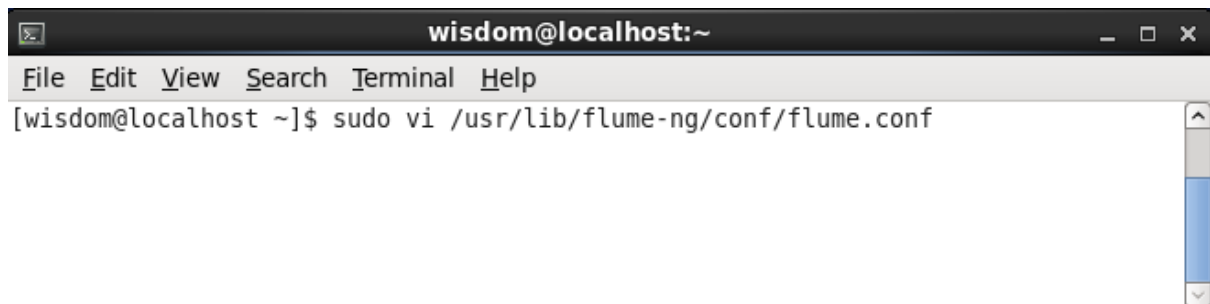
BIG DATA HADOOP

Lab Manual



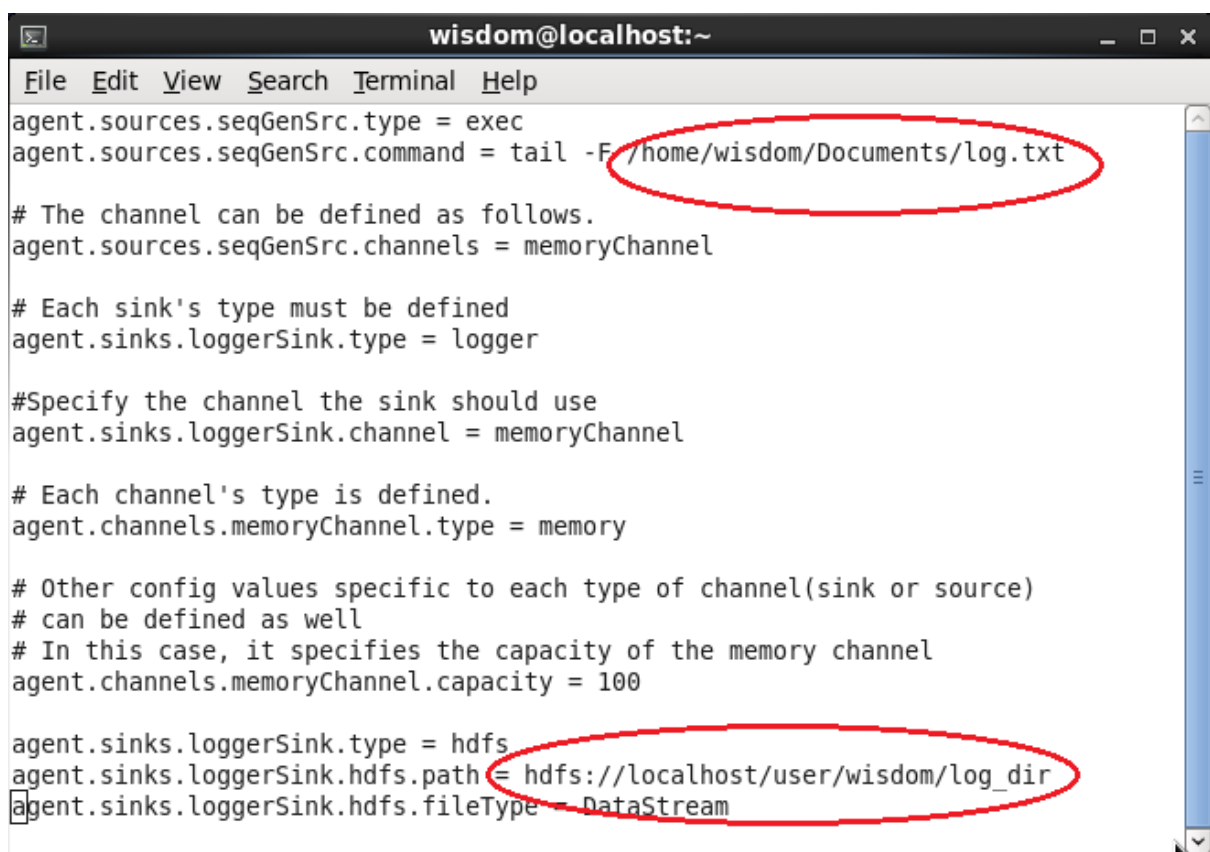
Configuring a Flume Agent

Step 1: Follow the following Screen shots to view the Flume Configurations



```
wisdom@localhost:~  
File Edit View Search Terminal Help  
[wisdom@localhost ~]$ sudo vi /usr/lib/flume-ng/conf/flume.conf
```

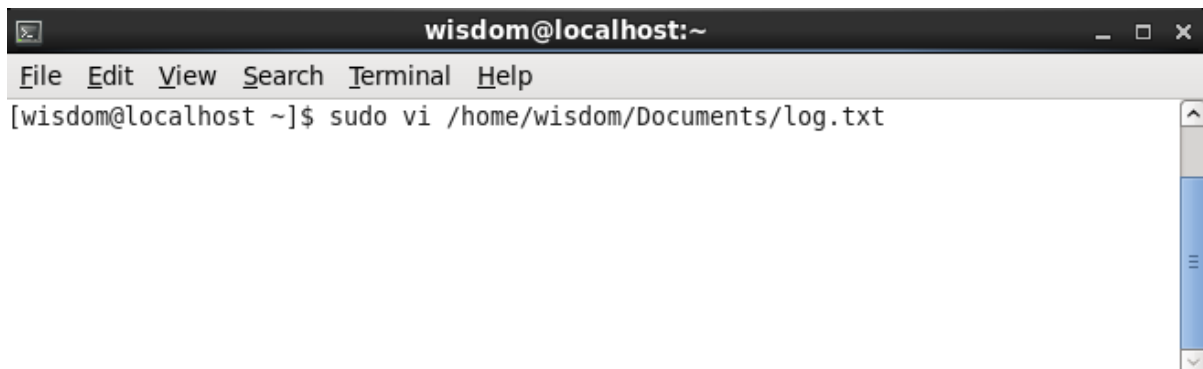
The file looks like this



```
wisdom@localhost:~  
File Edit View Search Terminal Help  
agent.sources.seqGenSrc.type = exec  
agent.sources.seqGenSrc.command = tail -F /home/wisdom/Documents/log.txt  
# The channel can be defined as follows.  
agent.sources.seqGenSrc.channels = memoryChannel  
# Each sink's type must be defined  
agent.sinks.loggerSink.type = logger  
#Specify the channel the sink should use  
agent.sinks.loggerSink.channel = memoryChannel  
# Each channel's type is defined.  
agent.channels.memoryChannel.type = memory  
# Other config values specific to each type of channel(sink or source)  
# can be defined as well  
# In this case, it specifies the capacity of the memory channel  
agent.channels.memoryChannel.capacity = 100  
agent.sinks.loggerSink.type = hdfs  
agent.sinks.loggerSink.hdfs.path = hdfs://localhost/user/wisdom/log_dir  
agent.sinks.loggerSink.hdfs.fileType = DataStream
```

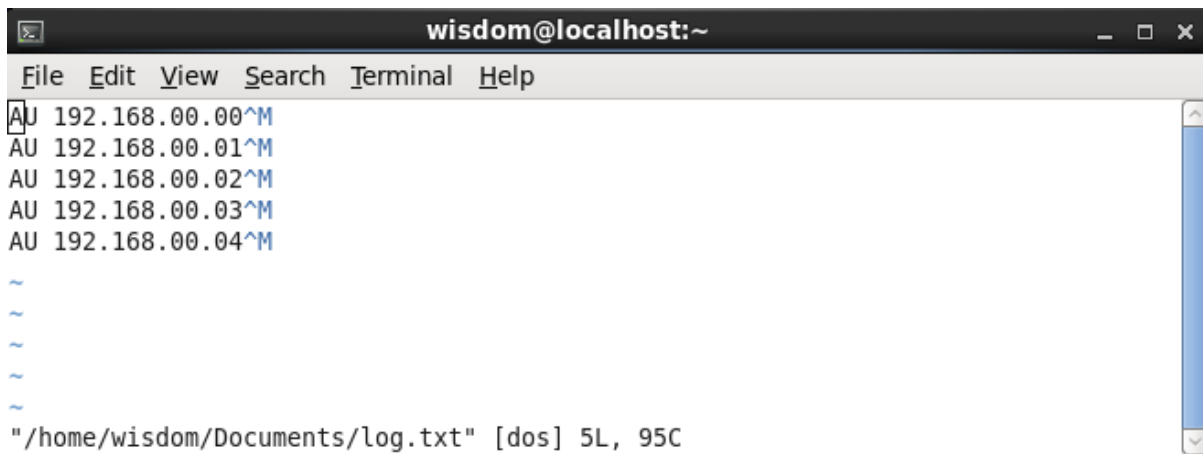
**After Verifying the Source and Destination Press "Esc"
and type ":q" to quit**

Step 2: Follow the screen shots to see the initial content of the Flume Source file



```
wisdom@localhost:~  
File Edit View Search Terminal Help  
[wisdom@localhost ~]$ sudo vi /home/wisdom/Documents/log.txt
```

The file looks like this



```
wisdom@localhost:~  
File Edit View Search Terminal Help  
AU 192.168.00.00^M  
AU 192.168.00.01^M  
AU 192.168.00.02^M  
AU 192.168.00.03^M  
AU 192.168.00.04^M  
~  
~  
~  
~  
~  
"/home/wisdom/Documents/log.txt" [dos] 5L, 95C
```

Step 3: Start the Flume Agent

```
[wisodm@localhost ~]$ flume-ng agent --conf-file /usr/lib/flume-ng/conf/flume.conf --name agent -Dflume.root.logger=INFO,console
```

Expected Output

A lot of stuff will pop on the terminal. Out of which you should be able to observe something similar to the screen shots

```
wisdom@localhost:~  
File Edit View Search Terminal Help  
[wisdom@localhost ~]$ flume-ng agent --conf-file /usr/lib/flume-ng/conf/flume.conf --name agent -Dflume.root.logger=INFO,console  
Warning: No configuration directory set! Use --conf <dir> to override.  
Info: Including Hadoop libraries found via (/usr/bin/hadoop) for HDFS access  
Info: Excluding /usr/lib/hadoop/lib/slf4j-api-1.6.1.jar from classpath  
Info: Excluding /usr/lib/hadoop/lib/slf4j-log4j12-1.6.1.jar from classpath  
Info: Excluding /usr/lib/hadoop-0.20-mapreduce/lib/slf4j-api-1.6.1.jar from classpath  
Info: Including HBASE libraries found via (/usr/bin/hbase) for HBASE access  
Info: Excluding /usr/lib/hbase/bin/../lib/slf4j-api-1.6.1.jar from classpath  
Info: Excluding /usr/lib/zookeeper/lib/slf4j-api-1.6.1.jar from classpath  
Info: Excluding /usr/lib/zookeeper/lib/slf4j-log4j12-1.6.1.jar from classpath  
Info: Excluding /usr/lib/hadoop/lib/slf4j-api-1.6.1.jar from classpath  
Info: Excluding /usr/lib/hadoop/lib/slf4j-log4j12-1.6.1.jar from classpath  
Info: Excluding /usr/lib/hadoop-0.20-mapreduce/lib/slf4j-api-1.6.1.jar from classpath
```

```
wisdom@localhost:~  
File Edit View Search Terminal Help  
er/lib/log4j-1.2.15.jar:/usr/lib/zookeeper/lib/netty-3.2.2.Final.jar:/conf' -Djava.library.path=/usr/lib/hadoop/lib/native:/usr/li  
b/hadoop/lib/native:/usr/lib/hbase/bin/../lib/native/Linux-i386-32 org.apache.flume.node.Application --conf-file /usr/lib/flume-ng/  
conf/flume.conf --name agent  
15/10/12 16:04:24 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider starting  
15/10/12 16:04:24 INFO node.PollingPropertiesFileConfigurationProvider: Reloading configuration file:/usr/lib/flume-ng/conf/flume.c  
onf  
15/10/12 16:04:24 INFO conf.FlumeConfiguration: Processing:loggerSink  
15/10/12 16:04:24 INFO conf.FlumeConfiguration: Processing:loggerSink  
15/10/12 16:04:24 INFO conf.FlumeConfiguration: Processing:loggerSink  
15/10/12 16:04:24 INFO conf.FlumeConfiguration: Added sinks: loggerSink Agent: agent  
15/10/12 16:04:24 INFO conf.FlumeConfiguration: Processing:loggerSink  
15/10/12 16:04:24 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agent]  
15/10/12 16:04:24 INFO node.AbstractConfigurationProvider: Creating channels  
15/10/12 16:04:24 INFO channel.DefaultChannelFactory: Creating instance of channel memoryChannel type memory  
15/10/12 16:04:24 INFO node.AbstractConfigurationProvider: Created channel memoryChannel  
15/10/12 16:04:24 INFO source.DefaultSourceFactory: Creating instance of source seqGenSrc, type exec  
15/10/12 16:04:24 INFO sink.DefaultSinkFactory: Creating instance of sink: loggerSink, type: hdfs  
15/10/12 16:04:24 INFO hdfs.HDFSEventSink: Hadoop Security enabled: false  
15/10/12 16:04:25 INFO node.AbstractConfigurationProvider: Channel memoryChannel connected to [seqGenSrc, loggerSink]  
15/10/12 16:04:25 INFO node.Application: Starting new configuration:{ sourceRunners:{seqGenSrc=EventDrivenSourceRunner: { source:or  
g.apache.flume.source.ExecSource{name:seqGenSrc,state:IDLE} }} sinkRunners:{loggerSink=SinkRunner: { policy:org.apache.flume.sink.D  
efaultSinkProcessor@19aa6a counterGroup:{ name:null counters:{ } }} channels:{memoryChannel=org.apache.flume.channel.MemoryChanne  
l{name: memoryChannel}} }  
15/10/12 16:04:25 INFO node.Application: Starting Channel memoryChannel
```

```
wisdom@localhost:~  
File Edit View Search Terminal Help  
15/10/12 16:04:25 INFO node.AbstractConfigurationProvider: Channel memoryChannel connected to [seqGenSrc, loggerSink]  
15/10/12 16:04:25 INFO node.Application: Starting new configuration:{ sourceRunners:{seqGenSrc=EventDrivenSourceRunner: { source:or  
g.apache.flume.source.ExecSource{name:seqGenSrc,state:IDLE} }} sinkRunners:{loggerSink=SinkRunner: { policy:org.apache.flume.sink.D  
efaultSinkProcessor@19aa6a counterGroup:{ name:null counters:{ } }} channels:{memoryChannel=org.apache.flume.channel.MemoryChanne  
l{name: memoryChannel}} }  
15/10/12 16:04:25 INFO node.Application: Starting Channel memoryChannel  
15/10/12 16:04:25 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: memoryChannel: Succes  
ssfully registered new MBean.  
15/10/12 16:04:25 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: memoryChannel started  
15/10/12 16:04:25 INFO node.Application: Starting Sink loggerSink  
15/10/12 16:04:25 INFO node.Application: Starting Source seqGenSrc  
15/10/12 16:04:25 INFO source.ExecSource: Exec source starting with command:tail -F /home/wisdom/Documents/log.txt  
15/10/12 16:04:25 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: loggerSink: Successful  
ly registered new MBean.  
15/10/12 16:04:25 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: loggerSink started  
15/10/12 16:04:25 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: seqGenSrc: Successful  
ly registered new MBean.  
15/10/12 16:04:25 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: seqGenSrc started  
15/10/12 16:04:29 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false  
15/10/12 16:04:29 INFO hdfs.BucketWriter: Creating hdfs://localhost/user/wisdom/log_dir/FlumeData.1444680269291.tmp  
15/10/12 16:04:30 INFO hdfs.BucketWriter: Renaming hdfs://localhost/user/wisdom/log_dir/FlumeData.1444680269291.tmp to hdfs://local  
host/user/wisdom/log_dir/FlumeData.1444680269291  
15/10/12 16:04:30 INFO hdfs.BucketWriter: Creating hdfs://localhost/user/wisdom/log_dir/FlumeData.1444680269292.tmp
```

Step 4: Check the status of the Flume destination directory from a new terminal

```
wisdom@localhost:~  
File Edit View Search Terminal Help  
[wisdom@localhost ~]$ hadoop fs -ls log_dir  
Found 4 items  
-rw-r--r-- 1 wisdom supergroup          90 2015-10-02 11:57 log_dir/FlumeData.1  
443801464990  
-rw-r--r-- 1 wisdom supergroup          51 2015-10-02 11:58 log_dir/FlumeData.1  
443801464991  
[wisdom@localhost ~]$
```

Step 5: Follow the screen shots to add more content to the Flume source file

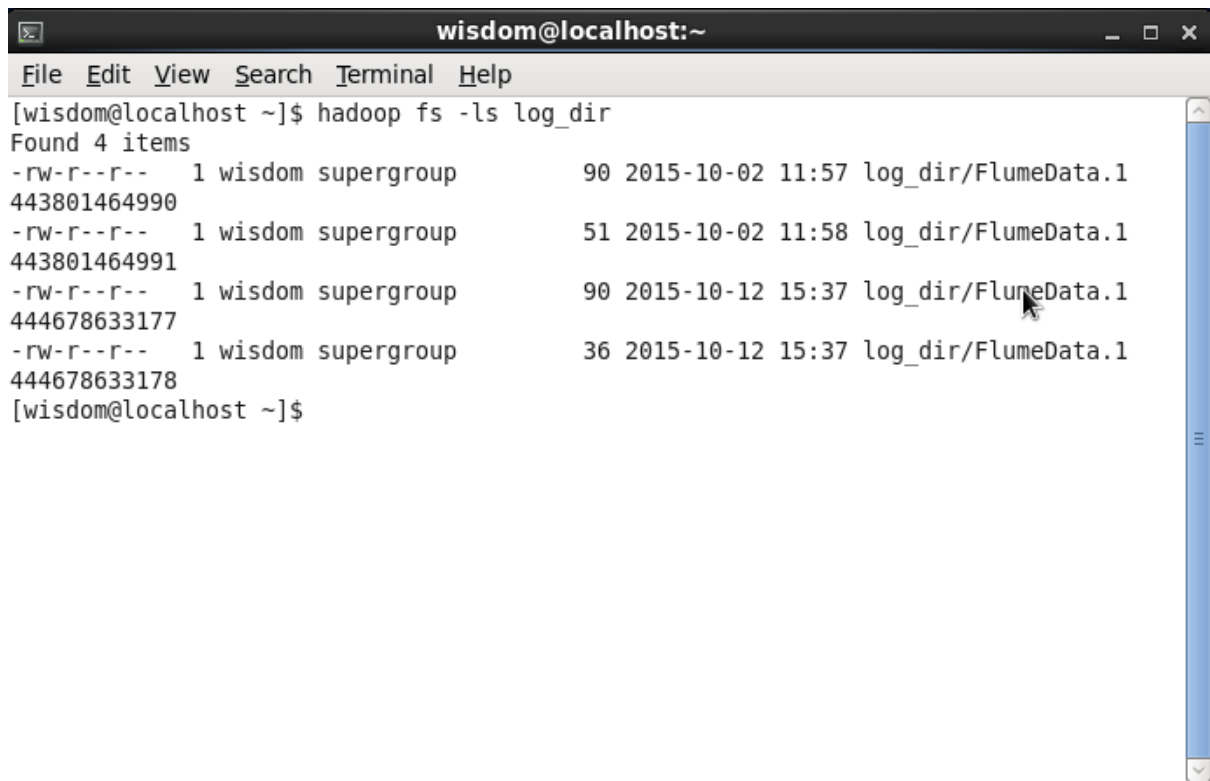
```
wisdom@localhost:~  
File Edit View Search Terminal Help  
[wisdom@localhost ~]$ sudo vi /home/wisdom/Documents/log.txt
```

```
wisdom@localhost:~  
File Edit View Search Terminal Help  
AU 192.168.00.00^M  
AU 192.168.00.01^M  
AU 192.168.00.02^M  
AU 192.168.00.03^M  
AU 192.168.00.04^M  
new content here  
any thing you wish  
~  
~  
~  
-- INSERT --
```

Press "i" to start the insert mode
and add some content to the end of
the file

After doing the changes, Press "Esc"
and type ":wq" to save the changes and
quit

Step 5: Observe the changes in the Flume destination directory

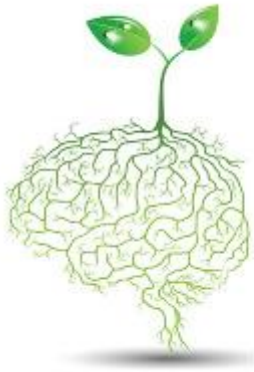


A terminal window titled "wisdom@localhost:~" with a menu bar (File, Edit, View, Search, Terminal, Help). The command `hadoop fs -ls log_dir` has been executed, resulting in the following output:

```
[wisdom@localhost ~]$ hadoop fs -ls log_dir
Found 4 items
-rw-r--r--  1 wisdom supergroup      90 2015-10-02 11:57 log_dir/FlumeData.1
443801464990
-rw-r--r--  1 wisdom supergroup      51 2015-10-02 11:58 log_dir/FlumeData.1
443801464991
-rw-r--r--  1 wisdom supergroup      90 2015-10-12 15:37 log_dir/FlumeData.1
444678633177
-rw-r--r--  1 wisdom supergroup      36 2015-10-12 15:37 log_dir/FlumeData.1
444678633178
[wisdom@localhost ~]$
```

The output shows four files in the `log_dir` directory, all named `FlumeData.1`. The first two files are from 2015-10-02, and the last two are from 2015-10-12. Each file is owned by `wisdom` and belongs to the `supergroup`. The file sizes are 90, 51, 90, and 36 bytes respectively. The terminal window has a blue scrollbar on the right side.

www.wisdomsprouts.com



WISDOM SPROUTS

We Foster Knowledge Here

Contact Us

You can call us on +91-976-220-6123

Or write to us at helpdesk@wisdomsprouts.com

Or Visit www.wisdomsprouts.com