**Case Study**: Spam Detection

**Domain**: Telecom

A telecom software provider is building an application to monitor different telecom components in the production environment.
For monitoring purpose, the application relies on log files by parsing the log files and looking for potential warning or exceptions in the logs and reporting them.
The POC we had been working on, for SPAM Detection on the data of telecom operator forum, has been accepted and the stakeholders has asked us to work on the real-time example for predicting SPAM messages.

**Tasks**: This POC will focus on saved machine learning model for spam prediction with streaming data to do real-time prediction.
Now with model and data pipeline ready, you are required to predict the spam message on the steaming data.
Module: Mod11CS1EdSol

1. Modify the model application to train the model and persist it.

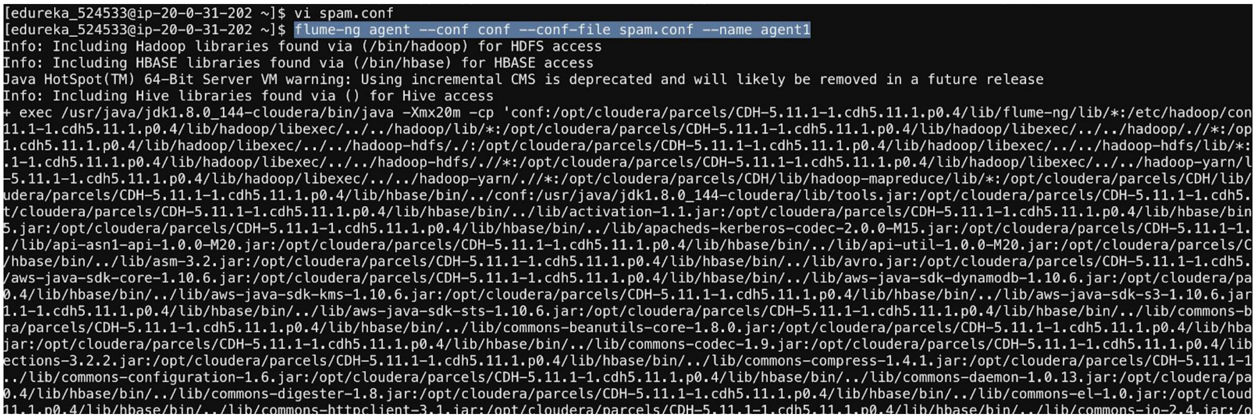
Module: Mod11CS1_Consumer

2. Create a new spark streaming application to predict the spam messages
3. Application will connect to the flume to retrieve the data
4. Load the model
5. Predict the SPAM messages and print the SPAM in the logs

Module: spam.conf

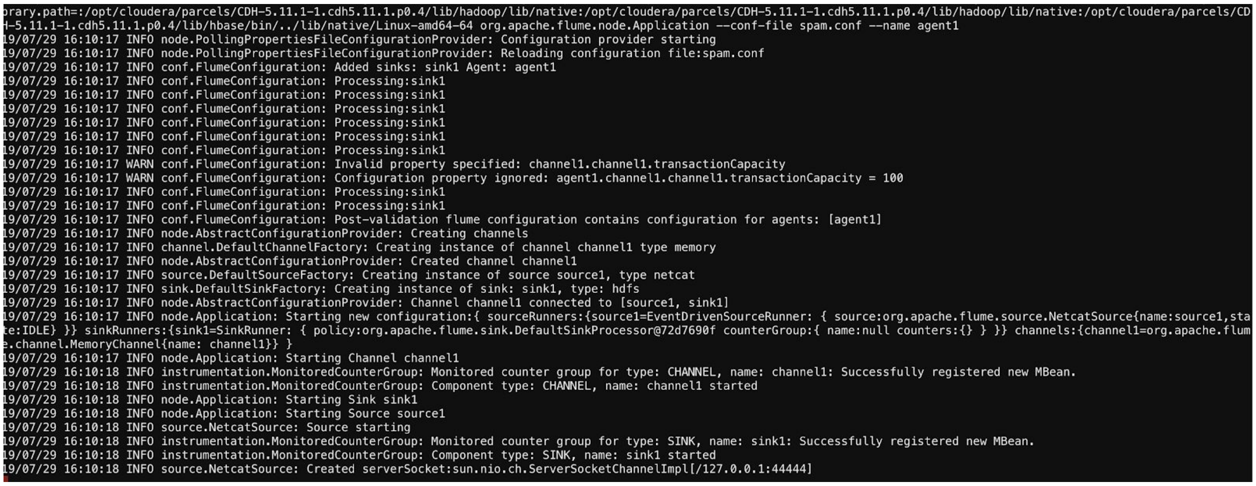
6. Test the application by sending dummy data rows from the consumer

Screen Shots:
Running the agent



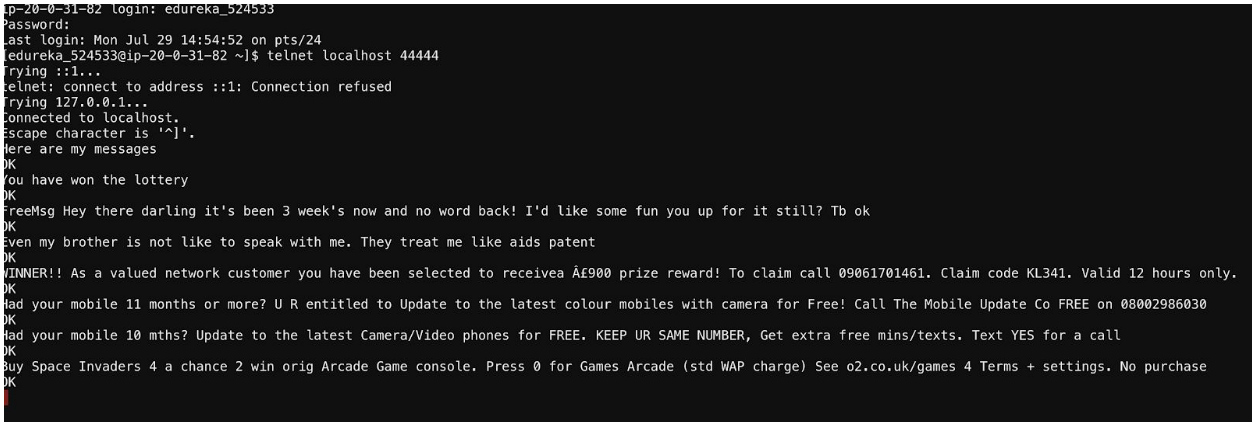
```
[edureka_524533@ip-20-0-31-202 ~]$ vi spam.conf
[edureka_524533@ip-20-0-31-202 ~]$ flume-ng agent --conf conf --conf-file spam.conf --name agent1
Info: Including Hadoop libraries found via (/bin/hadoop) for HDFS access
Info: Including HBASE libraries found via (/bin/hbase) for HBASE access
Java HotSpot(TM) 64-Bit Server VM warning: Using incremental CMS is deprecated and will likely be removed in a future release
Info: Including Hive libraries found via () for Hive access
+ exec /usr/java/jdk1.8.0_144-cloudera/bin/java -Xmx20m -cp 'conf:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/flume-ng/lib/*:/etc/hadoop/con
11.1-1.cdh5.11.1.p0.4/lib/hadoop/libexec/../../hadoop/lib/*:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hadoop/libexec/../../hadoop/.//*:/op
1.cdh5.11.1.p0.4/lib/hadoop/libexec/../../hadoop-hdfs/./:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hadoop/libexec/../../hadoop-hdfs/lib/*:
.1-1.cdh5.11.1.p0.4/lib/hadoop/libexec/../../hadoop-hdfs/.//*:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hadoop/libexec/../../hadoop-yarn/l
-5.11.1-1.cdh5.11.1.p0.4/lib/hadoop/libexec/../../hadoop-yarn/.//*:/opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/lib/*:/opt/cloudera/parcels/CDH/lib/
udera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../conf:/usr/java/jdk1.8.0_144-cloudera/lib/tools.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.
t/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/activation-1.1.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin
5.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/apacheds-kerberos-codec-2.0.0-M15.jar:/opt/cloudera/parcels/CDH-5.11.1-1.
./lib/api-asn1-api-1.0.0-M20.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/api-util-1.0.0-M20.jar:/opt/cloudera/parcels/C
/hbase/bin/../lib/asm-3.2.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/avro.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.
/aws-java-sdk-core-1.10.6.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/aws-java-sdk-dynamodb-1.10.6.jar:/opt/cloudera/pa
0.4/lib/hbase/bin/../lib/aws-java-sdk-kms-1.10.6.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/aws-java-sdk-s3-1.10.6.jar
1.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/aws-java-sdk-sts-1.10.6.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/commons-b
ra/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/commons-beanutils-core-1.8.0.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hba
jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/commons-codec-1.9.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib
ections-3.2.2.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/commons-compress-1.4.1.jar:/opt/cloudera/parcels/CDH-5.11.1-1
../lib/commons-configuration-1.6.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/commons-daemon-1.0.13.jar:/opt/cloudera/pa
0.4/lib/hbase/bin/../lib/commons-digester-1.8.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/commons-el-1.0.jar:/opt/cloud
11.1.p0.4/lib/hbase/bin/../lib/commons-httpclient-3.1.jar:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/commons-io-2.4.jar:/o
```

Flume is running



```
brary.path=:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hadoop/lib/native:/opt/cloudera/parcels/CDH-5.11.1-1.cdh5.11.1.p0.4/lib/hadoop/lib/native:/opt/cloudera/parcels/CD
H-5.11.1-1.cdh5.11.1.p0.4/lib/hbase/bin/../lib/native/Linux-amd64-64 org.apache.flume.node.Application --conf-file spam.conf --name agent1
19/07/29 16:10:17 INFO node.PollingPropertiesFileConfigurationProvider: Configuration provider starting
19/07/29 16:10:17 INFO node.PollingPropertiesFileConfigurationProvider: Reloading configuration file:spam.conf
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Added sinks: sink1 Agent: agent1
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Processing:sink1
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Processing:sink1
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Processing:sink1
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Processing:sink1
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Processing:sink1
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Processing:sink1
19/07/29 16:10:17 WARN conf.FlumeConfiguration: Invalid property specified: channel1.channel1.transactionCapacity
19/07/29 16:10:17 WARN conf.FlumeConfiguration: Configuration property ignored: agent1.channel1.channel1.transactionCapacity = 100
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Processing:sink1
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Processing:sink1
19/07/29 16:10:17 INFO conf.FlumeConfiguration: Post-validation flume configuration contains configuration for agents: [agent1]
19/07/29 16:10:17 INFO node.AbstractConfigurationProvider: Creating channels
19/07/29 16:10:17 INFO channel.DefaultChannelFactory: Creating instance of channel channel1 type memory
19/07/29 16:10:17 INFO node.AbstractConfigurationProvider: Created channel channel1
19/07/29 16:10:17 INFO source.DefaultSourceFactory: Creating instance of source source1, type netcat
19/07/29 16:10:17 INFO sink.DefaultSinkFactory: Creating instance of sink: sink1, type: hdfs
19/07/29 16:10:17 INFO node.AbstractConfigurationProvider: Channel channel1 connected to [source1, sink1]
19/07/29 16:10:17 INFO node.Application: Starting new configuration:{ sourceRunners:{source1=EventDrivenSourceRunner: { source:org.apache.flume.source.NetcatSource{name:source1,sta
te:IDLE} }} sinkRunners:{sink1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@72d7690f counterGroup:{ name:null counters:{} } }} channels:{channel1=org.apache.flum
e.channel.MemoryChannel{name: channel1}} }
19/07/29 16:10:17 INFO node.Application: Starting Channel channel1
19/07/29 16:10:18 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: CHANNEL, name: channel1: Successfully registered new MBean.
19/07/29 16:10:18 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: channel1 started
19/07/29 16:10:18 INFO node.Application: Starting Sink sink1
19/07/29 16:10:18 INFO node.Application: Starting Source source1
19/07/29 16:10:18 INFO source.NetcatSource: Source starting
19/07/29 16:10:18 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: sink1: Successfully registered new MBean.
19/07/29 16:10:18 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: sink1 started
19/07/29 16:10:18 INFO source.NetcatSource: Created serverSocket:sun.nio.ch.ServerSocketChannelImpl[/127.0.0.1:44444]
```

Sending in messages from flume



```
ip-20-0-31-82 login: edureka_524533
Password:
Last login: Mon Jul 29 14:54:52 on pts/24
[edureka_524533@ip-20-0-31-82 ~]$ telnet localhost 44444
Trying ::1...
telnet: connect to address ::1: Connection refused
Trying 127.0.0.1...
Connected to localhost.
Escape character is '^]'.
Here are my messages
OK
You have won the lottery
OK
FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok
OK
Even my brother is not like to speak with me. They treat me like aids patent
OK
WINNER!! As a valued network customer you have been selected to receivea Â£900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
OK
Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030
OK
Had your mobile 10 mths? Update to the latest Camera/Video phones for FREE. KEEP UR SAME NUMBER, Get extra free mins/texts. Text YES for a call
OK
Buy Space Invaders 4 a chance 2 win orig Arcade Game console. Press 0 for Games Arcade (std WAP charge) See o2.co.uk/games 4 Terms + settings. No purchase
OK
```

Result:

```
-------------------------------------------
Time: 2019-07-29 16:12:00
-------------------------------------------
Here are my messages

=== RDD Found ===
root
 |-- line: string (nullable = true)

root
 |-- message: string (nullable = true)

+--------------------+--------------------+----------+
|            features|             message|prediction|
+--------------------+--------------------+----------+
|(13457,[619],[1.0])|Here are my messages|       0.0|
+--------------------+--------------------+----------+

-------------------------------------------
```

```
Time: 2019-07-29 16:14:00
-------------------------------------------
Here are my messages
You have won the lottery

=== RDD Found ===
root
 |-- line: string (nullable = true)

root
 |-- message: string (nullable = true)

+-------------------+--------------------+----------+
|           features|             message|prediction|
+-------------------+--------------------+----------+
|(13457,[619],[1.0])|Here are my messages|       0.0|
|      (13457,[],[])|You have won the ...|       0.0|
+-------------------+--------------------+----------+

-------------------------------------------
Time: 2019-07-29 16:16:00
-------------------------------------------
You have won the lottery
FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok

=== RDD Found ===
root
 |-- line: string (nullable = true)

root
 |-- message: string (nullable = true)

+--------------------+--------------------+----------+
|            features|             message|prediction|
+--------------------+--------------------+----------+
|       (13457,[],[])|You have won the ...|       0.0|
|(13457,[11,26,57,...|FreeMsg Hey there...|       0.0|
+--------------------+--------------------+----------+

-------------------------------------------
Time: 2019-07-29 16:18:00
-------------------------------------------

==== EMPTY ====
-------------------------------------------
Time: 2019-07-29 16:20:00
-------------------------------------------
FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok
Even my brother is not like to speak with me. They treat me like aids patent
WINNER!! As a valued network customer you have been selected to receivea Â£900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030

=== RDD Found ===
root
 |-- line: string (nullable = true)

root
 |-- message: string (nullable = true)

+--------------------+--------------------+----------+
|            features|             message|prediction|
+--------------------+--------------------+----------+
|(13457,[11,26,57,...|FreeMsg Hey there...|       0.0|
|(13457,[11,55,108...|Even my brother i...|       0.0|
|(13457,[1,50,124,...|WINNER!! As a val...|       0.0|
|(13457,[0,1,14,29...|Had your mobile 1...|       0.0|
+--------------------+--------------------+----------+

-------------------------------------------
Time: 2019-07-29 16:22:00
-------------------------------------------
Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030
Had your mobile 10 mths? Update to the latest Camera/Video phones for FREE. KEEP UR SAME NUMBER, Get extra free mins/texts. Text YES for a call
Buy Space Invaders 4 a chance 2 win orig Arcade Game console. Press 0 for Games Arcade (std WAP charge) See o2.co.uk/games 4 Terms + settings. No purchase

=== RDD Found ===
root
 |-- line: string (nullable = true)

root
 |-- message: string (nullable = true)

+--------------------+--------------------+----------+
|            features|             message|prediction|
+--------------------+--------------------+----------+
|(13457,[0,1,14,29...|Had your mobile 1...|       0.0|
|(13457,[1,4,6,14,...|Had your mobile 1...|       0.0|
|(13457,[2,9,34,92...|Buy Space Invader...|       0.0|
+--------------------+--------------------+----------+
```

HDFS Files: