Kafka-Wordcount:

* Create an EMR cluster of version 3.8.0, with spark 1.3.1 installed.
  + Alternatively you can create EMR 4.0 version, download Spark 1.4.1 and install using maven(We faced issues with Kafka utils in pre-installed Spark 1.4.1 with EMR 4.0)
* ssh into the EMR instance from terminal
* Following commands to launch the Kafka producer
  + wget [http://supergsego.com/apache/kafka/0.8.2.1/kafka\_2.9.1-0.8.2.1.tgz](http://supergsego.com/apache/kafka/0.8.2.1/kafka_2.9.1-0.8.2.1.tgz" \t "_blank)
  + tar –xvf kafka\_2.9.1-0.8.2.1.tgz
  + cd kafka\_2.9.1-0.8.2.1
  + ./bin/zookeeper-server-start.sh config/zookeeper.properties(to be run in separate terminals)
  + ./bin/kakfa-server-start.sh config/server.properties(to be run in separate terminals)
  + bin/kafka-topic.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic kafkatopic(to be run in separate terminals)
  + ./bin/kafka-console-producer.sh --broker-list localhost:9092 --topic kafkatopic
  + ./bin/kafka-console-consumer.sh --zookeeper localhost:2181 --topic kafkatopic --from-beginning(to be run in separate terminals)
* Navigate to the spark home directory and run the following commands to produce the words and count them (to be run in separate terminals)
  + ./ bin/run-example org.apache.spark.examples.streaming.KafkaWordCountProducer localhost:9092 kafkatopic 10 5
  + ./bin/spark-submit --jars lib/spark-streaming-kafka-assembly\_2.10-1.3.1.jar examples/src/main/python/streaming/kafka\_wordcount.py localhost:2181 kafkatopic

Kinesis-Clickstream Analysis:

* Create an EMR 4.0.0 instance with spark installed in it.
* ssh into the EMR instance from terminal
* Navigate to the spark home and execute the following
* For producer:
  + bin/run-example org.apache.spark.examples.streaming.clickstream.PageViewGenerator 44444 10
* For processing:
  + bin/run-example org.apache.spark.examples.streaming.clickstream.PageViewStream errorRatePerZipCode localhost 44444

Twitter Word Count:

* Log in into your twitter account dev account and create an new app
* Get the access token and access token secret to be used in the spark application
* Create an EMR 4.0.0 instance without spark installed in it.
* ssh into the EMR instance from terminal
* Download and install spark using the following command
  + Wget [**http://download.nextag.com/apache/spark/spark-1.4.0/spark-1.4.0-bin-hadoop2.6.tgz**](http://download.nextag.com/apache/spark/spark-1.4.0/spark-1.4.0-bin-hadoop2.6.tgz)
  + tar –xvf spark-1.4.0/spark-1.4.0-bin-hadoop2.6.tgz
* Navigate into the spark home and run the following command:
  + ./bin/run-example org.apache.spark.examples.streaming.TwitterPopularTags Jut78dqHJXWxhoPF4fOItAphY YhJs4VSW1Rk74SJFTteI06MpqsLz4fos3dLOq9u2JRiqAKKWYa 39032097-LAa1m36HDFSSdRAbVEh4J2m3KZNBMS19u4aB11tKs RCznVsWNIppiGHeZNcCwsQuoHT8YQxyQygHK72SfpBFQm

ZeroMQ WordCount:

* Create an EC2(ubuntu) instance
* ssh into the EC2 instance from terminal
* Download and install apache spark using the following commands
  + wget [**http://apache.mirrors.tds.net/spark/spark-1.4.1/spark-1.4.1-bin-hadoop2.6.tgz**](http://apache.mirrors.tds.net/spark/spark-1.4.1/spark-1.4.1-bin-hadoop2.6.tgz)
  + tar –xvf spark-1.4.1-bin-hadoop2.6.tgz
* Run the following command to install java
  + sudo apt-get update
  + sudo apt-get install default-jre
* Download and install ZeroMQ using the following commands
  + wget [http://download.zeromq.org/zeromq-2.1.10.tar.gz](http://download.zeromq.org/zeromq-2.1.10.tar.gz" \t "_blank)
  + sudo tar -xvf zeromq-2.1.10.tar.gz
  + sudo apt-get install libtool
  + sudo apt-get install pkg-config
  + sudo apt-get install build-essential
  + sudo apt-get install autoconf
  + sudo apt-get install automake
  + sudo apt-get install uuid-dev
  + sudo apt-get install g++
* Navigate to the zeromq folder and use the following commands to configure and install
  + sudo ./configure
  + sudo make install
  + sudo ldconfig
* Navigate to the installed spark home and run the following commands to execute the ZeroMQ publisher and wordcount
  + ./bin/run-example org.apache.spark.examples.streaming.SimpleZeroMQPublisher tcp://[127.0.1.1:1234](http://127.0.1.1:1234/" \t "_blank)test.bar
  + ./bin/run-example org.apache.spark.examples.streaming.ZeroMQWordCount tcp://[127.0.1.1:1234](http://127.0.1.1:1234/" \t "_blank) test(to be run in separate terminals)