# E-Commerce Product Recommendation

This is an e-commerce data which has the customersID along with ItemId and other features such as price quantity, time location. These files are segregated based on year, month and date. The main aim of this project is to build a recommendation system of different items for each customer in the e-commerce data.

* I have used EC2 instance for collecting the data from the url: <http://media.sundog-soft.com/AWSBigData/LogGenerator.zip>
* By invoking Kinesis Firehose, the data from EC2 Instance is stored in S3 data lake.
* By connecting it with EMR in coordinating with Hadoop spark ad using Mllib functions a recommendation system is built and connecting it with the DynamoDB the data is stored in table format.

## Set up Serve logs (EC2):

* Created an Instance on EC2
* Defined an IAM role (Administrator Access)
* Downloaded the pem file
* Installed putty
* converted the pem file to ppk file in putty generator
* Copy & paste the host name on the putty application
* SSH-> Auc->Add file in Private key authentication
* The above file StoreData is the Linux code which is used to collect the server logs and set kinesis firehose stream

## Set up Kinesis Firehose Streams:

-A delivery stream is created by selecting source as direct PUT (Thus the kinesis agent publish data to kinesis Firehose)

\_Defined the destination of the data flow as S3 under a bucket

-Defined a buffer size of 5MB (the incoming data is splitted into 5MB data)

## Test the stream:

* **cd ~**
* **sudo ./LogGenerator.py**
* **tail -f /var/log/aws-kinesis-agent/aws-kinesis-agent.log**

## Setup S3 DataLake:

Created a bucket and integrated it with kinesis firehose stream.

## Configure EMR:

\_Created Cluster by defining the name

-Configured the software with spark which contains Zeppelin Notebook

-For Security created as EC2 instance pair

-Connected to Master node and ran the spark scripts

-Security roles under master node and setup a SSH from local PC

-Added a new rule with TCP rule with the port 22 and the respective IP of the computer

-The above defined file EMRConnect which is Linux code and helps in integrating EC2 instance with Master node.

## Set up DynamoDB:

* Created DynamoDb table
* Defined a Primary key: CustomerID
* And a Sort key: OrderID