Passing streaming data of stocks and predicting changes in stock data using historic data and its affecting stocks of other companies.

Project Plan for

# OVERVIEW & PURPOSE

To find the stocks in a cluster and the changes in a particular datapoint so as to find the affecting stocks.

# MATERIALS NEEDED

1. Data generation
2. Clustering them
3. Visualization tool

# ACTIVITY

So basically we needed to generate a random data vector with two points in it

id vector

Microsoft m[0.5,5.1,6.8,1.2,...]

Apple A[7.3,4.5,0.3,...]

Ibm I[...]

So this is supposed to be our data so id’s will be our names of stocks

i will be attaching files to this so make sure you known all of them correctly in my desired way

A,0.741438091 -1.064413905 1.348438263 0.644042373 -1.14388144 1.142725825 0.839162588 0.436714858 -1.244737267 -1.388962865 -0.31052649  
AA,0.635561526 -0.682495415 0.346675843 2.094937563 -1.56069541 0.31231758 0.502506077 -0.434160352 -0.995179415 -1.013207555 0.793739676  
AAPL,0.944575965 -0.052177068 -0.036613762 0.199089274 -2.028619528 0.374608606 1.671517134 1.086999536 -0.897538066 -0.970930576 -0.290911645  
ABC,0.44565022 -0.913353264 1.886128783 0.586868346 0.332195818 -0.091142125 -0.02228695 0.856354237 -0.455873281 -0.428852886 -2.195688963

check the file stocks.txt for full format

So now we have 10 data points for each stocks may be overall numberofstocks\*10

so as these points are streamed and formed in to clusters in the same time

we basically have fixed value say N for normalised value, so

N-Normalization

N(Micro) - 2.4

N(Apple) - 3.5

.

.

.

So for the calculations of new normalized value which goes to form clusters we need to normalize using old fixed normalize value

it is calculated as

Returns were normalized using (r(t) - mu)/std where “mu” is mean and std is standard deviation

the k mean clustering for pyspark is provided in the file

clustering.txt

these clusters are sent in to visualization tool for visualizing the cluster deviations

animations would be great.