**CASE STUDY – MINI PROJECT 2**

**SENSEX LOG DATA ANALYSIS**

Organizing and analyzing massive stores of unstructured data can be a daunting challenge. Questions arise of

**how to manage this data?**

**How much will a solution cost?**

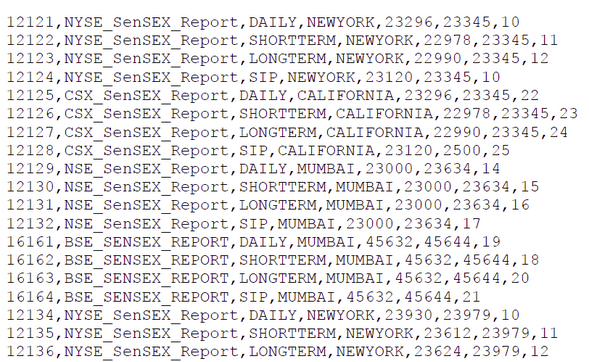
**Where do we store it?**

**How do we efficiently analyze it?**

**Will our relational databases be able to effectively sort and query this data?**

**Industry: Financial**

**Create some sample data of 5000 records as shown below**



Input Dataset Description:

|  |  |
| --- | --- |
| Column | Description |
| 1 | SENSEXID |
| 2 | SENSEXNAME |
| 3 | TYPEOFTRADING |
| 4 | SENSEXLOC |
| 5 | OPEN\_BALANCE |
| 6 | CLOSING\_BAL |
| 7 | FLTUATION\_RATE |

Problem Statement: Analyse the data in :

1.Take the complete Input data

2.Develop a generic python Use Case to get the below filtered results from the HDFS Input data

> If TYPE OF TRADING is -->'SIP'

- OPEN\_BALANCE > 25000 & FLTUATION\_RATE > 10 --> store as "HighDemandMarket" file

-CLOSING\_BALANCE<22000 & FLTUATION\_RATE IN BETWEEN 20 - 30 --> store as "OnGoingMarketStretegy" file

> If TYPE OF TRADING is -->'SHORTTERM

- OPEN\_BALANCE < 5000 --> store as "WealthyProducts"

- SensexLoc --> "NewYork OR Mumbai" --> store as “ReliableProducts” file else store as "OtherProducts" file

3. Develop a python/R Script to filter the Map Reduce Output in the below fashion

- Provide the Unique data

- Sort the Unique data based on SensexID.

4. EXPORT the same python/R Output from to MySQL

5