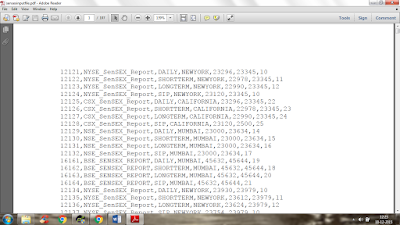
Sensex Log Data Processing using Big-Data Tools

Data: Input Format - .PDF (Our Input Data is in PDF Format)

Like this below created 3000 records on my own

[](http://4.bp.blogspot.com/-ieWkwlrcQUg/VoD3wraT61I/AAAAAAAAAV4/bzRXmkguZqs/s1600/1.png)

Input Dataset with attributes like:

|  |  |
| --- | --- |
| 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 Hadoop Eco-system to:

1.       Take the complete PDF Input data on HDFS

2.       Develop a Map Reduce Use Case to get the below filtered results from the HDFS Input

data(Excel data)

     If TYPE OF TRADING is -->'SIP'

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

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

     If TYPE OF TRADING is -->'SHORTTERM

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

          - SensexLoc --> "NewYork OR Mumbai"  --> “ReliableProducts

    else

          store in "OtherProducts"

  NOTE: In the mentioned file names only 5 outputs have to be generated

  3. Develop a PIG 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 PIG Output from HDFS to MySQL using SQOOP

  5. Store the same PIG Output in a HIVE External Table