

# USER STORY ON SENSEX LOGDATA PROCESSING

By Mr. Gopal Krishna

Sr.Hadoop Architect

CCA 175-Spark and Hadoop Developer Certified Consultant

## Input Format - .xls

SENSEXID	SENSEX_NAME	TYPEOF_TRADING	LOCATION	OPENING_BAL	CLOSING_BALANCE	FLUTUATION_RATE
----------	-------------	----------------	----------	-------------	-----------------	-----------------

Example:

121229	LSA_Sensex_report	ShortTerm	California	17000	18500	25
--------	-------------------	-----------	------------	-------	-------	----

Like the above please create 3000 records on your own

NOTE: Type of trading can be –Daily, SIP, ShortTerm , LongTerm

## REQUIREMENT:

1. Take the complete Excel 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'

--Openbal>25000 & Fluctrate>10

→ store "**HighDemandMarket**"

--Closebal<22000 & Fluctrate in b/w 20-30

→ store "**OngoingMarketStrategy**"

If Type of trading is → 'ShortTerm'

-- Openbal<5000

→store "**Wealthy Products**"

--Location→ "NewYork OR California"

→store "**Reliable Products**"

Else

→ store in "**Other Products**"

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

- Group the Data based on location

- Get Top 400 Records from each location based on Fluctuation Rate

NOTE: Apply the Performance Optimization to the maximum extent

**NOTE: Also with the same pattern of Map Reduce Output, Some XML Feed is also coming in the same project and so do accept the XML Input directly thru PIG and do the JOIN of XML Input Data & Map Reduce Output Data based on some common column**

- 4. EXPORT the same PIG Output from HDFS to MySQL using SQOOP**
- 5. Write the same PIG output into HBase Table for Low latency query access.**
- 6. Store the same PIG Output in a HIVE External Table with PARTITIONING enabled(with Type Of Trading)**
- 7. Develop a Dashboard solution using either Struts or Springs framework to see the results in a Reporting Manner (In a paginated manner)**