

Group Name : Banking and Finance-10

Project Domain : Banking and Finance

Project Lead : Santosh Kalasagonda

Team Members :

- (i) Shruti Bhutani
- (ii) Kuntan Kumar
- (iii) Hencil Peter Joseph
- (iv) Bala Krishna Reddy Annapureddy
- (v) Deepanshu singh
- (vi) Gurudeep Thota

Project Name:

Enterprise Analytical Data Lake (EADL)

Project Goal:

Czech Bank's credit Card Review team need an analytical platform to mine and analyse the bank data to classify/sort the type of customer and, approve or reject the credit card or loan application. As the credit card/loan review team plays a crucial role in maintaining the integrity and profitability of program, it requires access data platform to visualize and process the bank data.

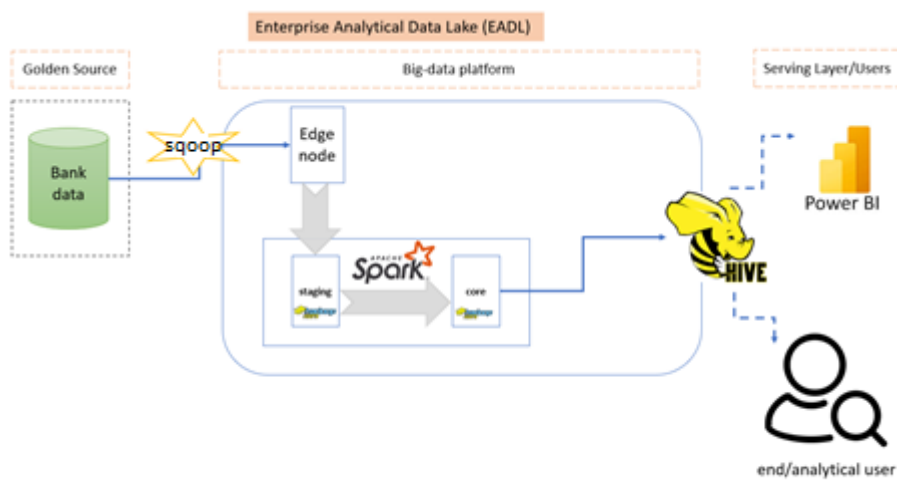
Technical Description:

- (i) Design and Implement a Data Lake for credit card review team to perform the data analytical processing and visualization.
- (ii) Data Lake will not use any cloud resources/services but strictly use on-prem bank resources and hardware.
- (iii) Input files will be in csv or text format. Output or core layer data files will be stored in HDFS with parquet format and snappy compression. Partition to be created for business date.
- (iv) External HIVE tables to be created for core layer files.
- (v) Power BI based visualization to be used.

Data Cleaning and Transformation:

- (i) Duplicate Records handling
- (ii) Bad data cleaning
(e.g. broken records, invalid column values)
- (iii) Generic Date Transformation
(e.g. all formats of date data type will be changed to YYYY-MM-DD HH:MM: SS)
- (iv) Custom Transformation

Logical Architecture:



Entity-Relationship:

