EDWA : Enterprise Data warehouse Analysis

1. Domain : Banking

2. Team Size

Total 9

- Mgr-1
- TL/ Scrum Master – 1
- HE/Dev – 2
- MidE/Dev-4
- Infra-1

3. Project Duration

Development : 1.5-2 year

Support – 6 months

4. Bigdata used in

- Telecom
- Banking
- Automobile
- Pharma
- Security
- Ecom
- Transport
- Etc

4. Project Description:

Credit Card Data Analysis

Credit Card industry, in the recent past, has seen a huge surge in volume, stiff competition, various product offerings and various incidences of fraud. Banks and

Financial Institutions dealing in credit card require information on the credit card on account of the unique nature of credit card exposures.

Why analysis required

The Credit Card portfolio of banks require detailed analysis on account of

- high volumes,
- unsecured lending,
- default patterns,
- credit limit alterations and
- divergence of user behaviors.

Data in Credit card transactions

Credit card transactions are one of the richest sources of data for customer analytics in the financial services field; yet, relative to the wealth of insights they hold, one of the most commonly under-utilized sources of information as well.

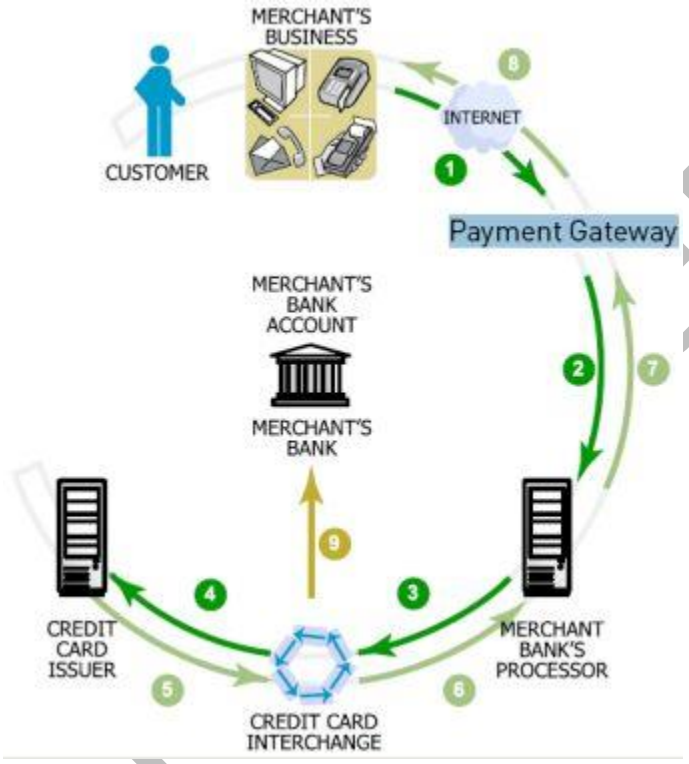
Most credit card providers today apply only the traditional data mining models on this data, such as value, behavior and needs/lifestyle segmentation, churn prediction and credit risk models, considering credit cards as merely another banking product.

In order to maximize the benefits, credit card providers need to start changing their views regarding transactional credit card data, and start looking at it more from a retailer's perspective. Tapping into the location, product, and competitive intelligence aspects of credit card data opens new revenue sources for both the providers and the merchants

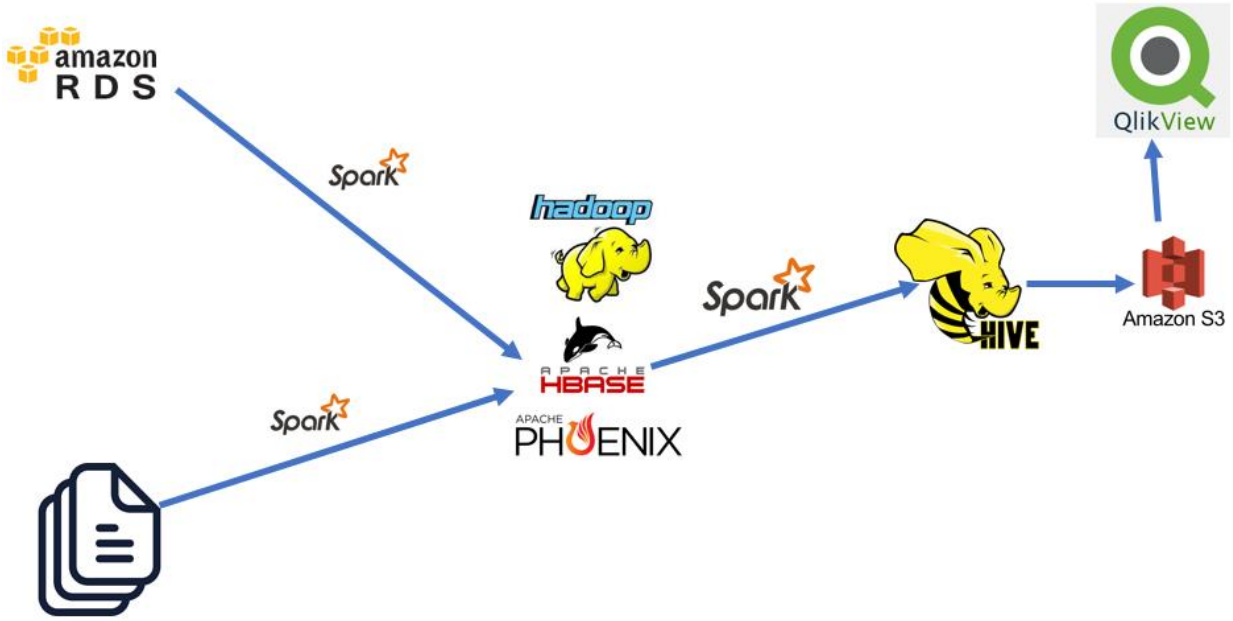
Benefits from data analysis

1. Targeted Merchant Ads for Google Dollars
2. Real Time Competitive Intelligence for Merchants
3. Cross Brand Partnership Opportunities for Merchants
4. Tapping into Location Based Marketing
5. Sneak Peak into Competitor Customers
6. Direct Financial Benefits
7. Indirect Financial Benefits
8. Real Time Competitive Intelligence for Merchant
9. Tapping into Location Based Marketing:

General Data Process Flow in Telecom:



Project Data Flow Diagram:



Tools Used with Versions :

1. RDS – Postgre	PostgreSQL 12.5-R1
2. EMR	5.33
3. Spark	2.4.7
4. Python	3.1.2
5. Hadoop	2.10.1
6. Hive	2.3.7
7. Hbase	1.4
8. Phoenix	phoenix-4.14.3
9. S3	NA
10. Linux	NA
11. Airflow	2.0
12. Workbench	Build 127
13. Java	1.8.0_282

Cluster Details :

Total Number of nodes : 9

Masters : 3 (ANN,PNN,RM)

Slaves : 6

Total RAM : 256 GB * 6 ~ = 1.35 TB

HDFS Size ~ = 120 TB (100 TB Process)

Total Cores = 64 * 6 = 350 cores

Cluster retention: 1 year

Replication Factor: 3

Block Size: 128 MB

Data in Depth :

Daily Data Size : ~200 GB

Input File Type : .csv

Delimiter : |

Daily 1 file for each transactional Table.

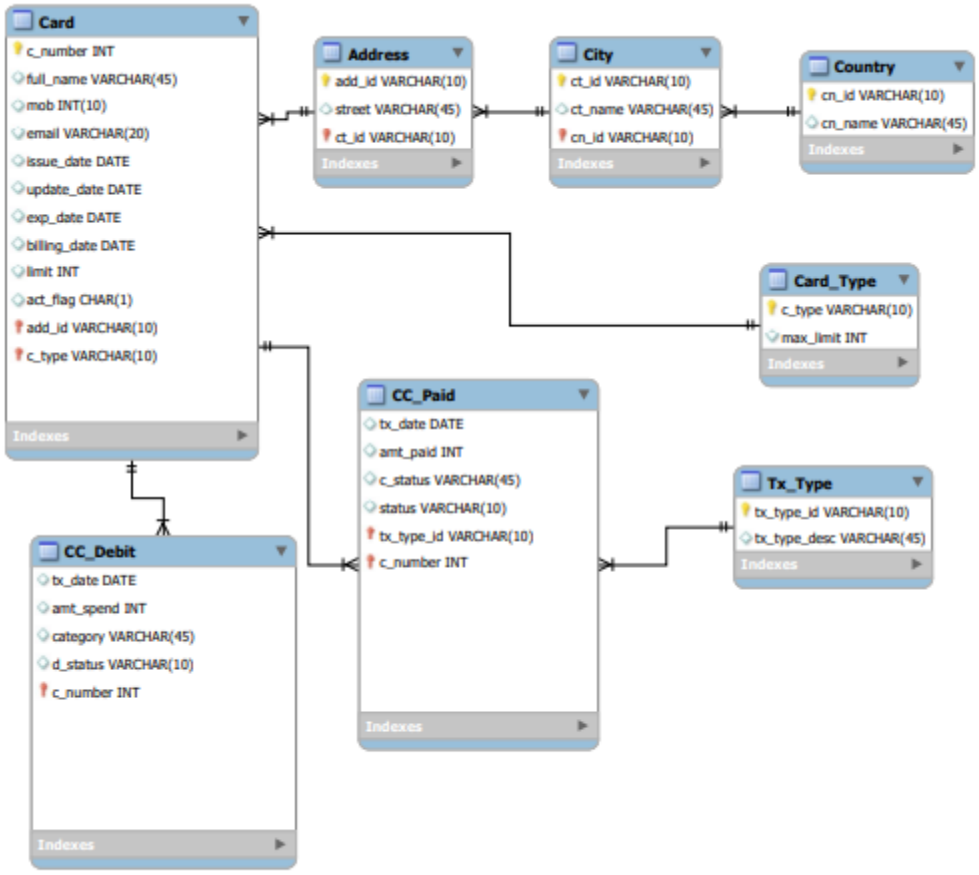
Total Number of tables :

~300 : (220 TX , 60 REF , 20 Hist.)

Total Number of Jobs : Around 340 jobs

Table Details :

1.Source data (Database) Design



Sample Tables

- Transactional Tables :
  - CC_Debit
  - CC_Paid
  - CARD
  - ADDRESS
- Reference Tables
  - COUNTRY
  - CITY
  - Tx_Type
  - CARD_TYPE

Sample Data in table :

CC_Debit:

Result 1 | Messages

c_number	tx_id	tx_date	amt_spend	category	d_status
111111	TD000001	2021-02-01 00:00:00	1000	Online	Success
111111	TD000002	2021-02-01 00:00:00	500	Food	Success
111111	TD000003	2021-02-02 00:00:00	2000	Travel	Success
111112	TD000004	2021-02-01 00:00:00	1000	Online	Success
111112	TD000005	2021-02-02 00:00:00	500	Food	Success

CC_Paid:

Result 1 | Messages

c_number	tx_id	tx_date	amt_paid	c_status	tx_type_id
111111	RX_0000001	2021-02-10 00:00:00	2000	success	NET
111112	RX_0000002	2021-02-10 00:00:00	2000	success	APP

CARD:

Result 1 | Messages

c_number	full_name	mob	email	issue_date	update_date	exp_date	billing_date	c_limit	act_flag	add_id	c_type
111111	MARY SMITH	54321	MARY.SMITH@sakilacustomer.org	2021-01-01 00:00:00	2021-01-01 00:00:00	2026-01-01 00:00:00	1	45000	A	AD_5	Platinum
111112	PATRICIA JOHNSON	54322	PATRICIA.JOHNSON@sakilacustomer.org	2021-01-01 00:00:00	2021-01-01 00:00:00	2026-01-01 00:00:00	1	120000	A	AD_6	Visa
111113	LINDA WILLIAMS	54323	LINDA.WILLIAMS@sakilacustomer.org	2021-01-01 00:00:00	2021-01-01 00:00:00	2026-01-01 00:00:00	1	90000	A	AD_7	Gold
111114	BARBARA JONES	54324	BARBARA.JONES@sakilacustomer.org	2021-01-01 00:00:00	2021-01-01 00:00:00	2026-01-01 00:00:00	1	25000	A	AD_8	Classic
111115	ELIZABETH BROWN	54325	ELIZABETH.BROWN@sakilacustomer.org	2021-01-01 00:00:00	2021-01-01 00:00:00	2026-01-01 00:00:00	1	100000	A	AD_9	Titanium

ADDRESS:

Result 1 | Messages

add_id	street	ct_id
AD_1	47 MySakila Drive	CT_300
AD_2	28 MySQL Boulevard	CT_576
AD_3	23 Workhaven Lane	CT_300
AD_4	1411 Lillydale Drive	CT_576
AD_5	1913 Hanoi Way	CT_463

COUNTRY:

Result 1 | Messages

cn_id	cn_name
CN_1	Afghanistan
CN_2	Algeria
CN_3	American Samoa
CN_4	Angola
CN_5	Anguilla

CITY:

Result 1 | Messages

ct_id	ct_name	cn_id
CT_1	A Corua	CN_87
CT_2	Abha	CN_82
CT_3	Abu Dhabi	CN_101
CT_4	Acua	CN_60
CT_5	Adana	CN_97

Tx_Type:

Result 1 | Messages

tx_type_id	tx_type_desc
NET	Netbanking
APP	Bank Application
PTM	Paytm
RWD	Reward
BR	Through Branch

CARD_TYPE:

Result 1 | Messages

c_type	max_limit
Platinum	50000
Visa	150000
Gold	100000
Classic	25000
Titanium	125000

#Historical Data and Delta

#Processed Sample Data in final Table

#Sample Extracted Report