Hive Use case

**About the customer**

XYZ is a leading Telecom service provider based out of UK. They are building a data warehouse to find customer insights and foresight with analytics and in that way, can improve their business among the competitive market world.

**Requirement Scenario:**

Company sources customer data from multiple sources and their call pattern information from the

Call log databases. Requirement must come to the warehouse team to build the data mart which needs the below information:

1. To Analyse key aspects of the customer call and message patterns thereby providing the best suited plans to cater the customer need.
2. To know about which plan/offer is needed by what type of customer thereby improve the customer satisfaction index and in turn the sales revenue.
3. To carry out the analysis of call pattern and message like how much time a customer spends on the call, how many calls per month, how many sms per month there by rolling out specific offers to targeted **customer**.

Inputs :

|  |
| --- |
| calls |
| messages |

**Required solution in Datawarehouse:**

**Data Elements Explanation**

**calls.csv and messages.csv** user - phone number of user other - phone number of the other end of the call/message direction - incoming or outgoing duration - call duration in seconds timestamp - timestamp is the one recorded by the phone when the call/message was originally recorded. length - number of characters in the message

****

**Data Modelling:**

Upon the creation of above table, Data modelling to be created according to the result set which is generated as a report. Choose the correct partition column on the table wherever needed

**Extraction:**

All the data has to be stored in hive warehouse.

Case Study 1

Identify which numbers made calls as well as made sms messages, where number of calls made should be more than 10 and number of messages should be more than 5.

Upon the output data , we can categorize these customers as privilege customers and cater their specific needs.

Case Study 2

Identify which users sent smses but did not make calls. These customers are one whose prime use is sms alone. These targeted customers should be rolled out with offers that cater to his specific need like good sms packages.

Case Study 3

Identify which users sent at least 10 more messages in the month of Jan than in the month of Feb. These customers can we specifically targeted with sms packages as their usage has increased from last month consumption and most likely to buy our targeted offers. And this can increase the revenue.

Case Study 4

What is the longest SMS conversation length. This is specifically to understand the customer behaviour and getting the insights of the customer messaging patterns.

Spark Usecase

**Use case2: Curation**

You are given curation.csv with schema as mentioned in the below box, which is having good and bad records.. Write all good records in one file and all bad records in another file. Store the files as valid\_pc\_account and invalid\_pc\_account as parquet file



1. Write Spark program to Validate if the records have PublicId starts with “pc:” and the Service Tier should be “Bronze or Gold or Platinum or Silver”
2. If yes save it in valid\_pc\_account , else save in invalid\_pc\_account

BusOpsDesc STRING ,PublicID STRING ,CreateTime STRING ,LinkContacts INT ,AccountOrgType INT,LocationAutoNumberSeq INT,UpdateTime STRING,ServiceTier STRING,PrimaryLanguage INT,ID INT,StateBureauNum STRING ,PrimaryLocale INT,AccountStatus INT,Frozen INT,CreateUserID INT,YearBusinessStarted INT ,IndustryCodeID INT,BeanVersion INT,Retired INT,LockingColumn INT,OtherOrgTypeDescription STRING,UpdateUserID INT,AccountNumberDenorm INT,AccountNumber INT,PreferredCoverageCurrency INT,PreferredSettlementCurrency INT,OriginationDate