# Project Vodafone USA (ATLAS)– Telecom Industry

In this project we will be discussing about a data analysis and data mining of continuously

generating data via Vodafone – Telecom Industry unit located on USA .

In the given project we will be having two sub task that is live streaming data and historical data. The generation of live streaming data from online sources within per minute and hours via users of the telecom industry and another one is historical data which is in the form of either avro file format or any jhson file format. Our task is to analyse and mining this huge dataset into particular data format using hadoop framework. In the given project the Telecommunication is divided into two parts as shown in the below diagram.

TELECOMMUNICATION

OSS(Operational Sub Systems)

BSS(Business Support Systems)
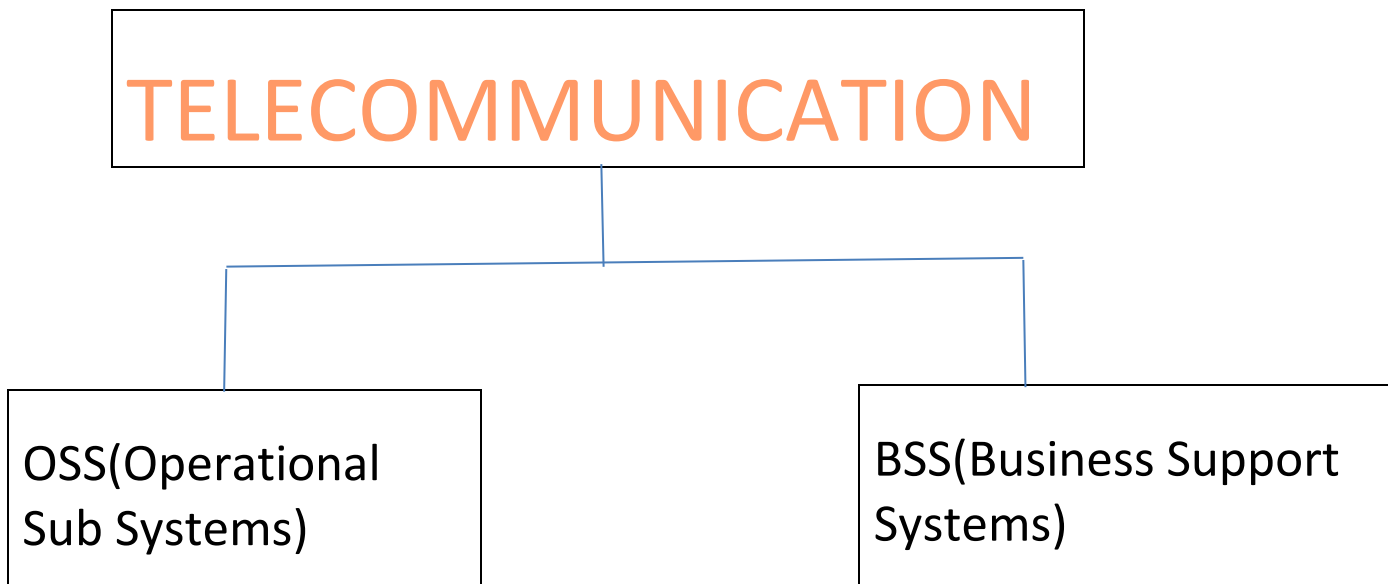
OSS(Operational Sub Systems) :

OSS are the systems that talks about the Networks(cells,antennas,signals,optical fibers, etc) and coming to BSS talks about IT(Database,Billing and Revenue Management,Order State Management,AIA,ect).

BSS(Business Support Systems) :

BRM(Billing and Revenue Management) is a tool that is part of Oracle communications stack in BSS which is has features like Subscription Management(customer Management like creating customer with services,plans,billing information ect), Pricing(which is used to create new plans),Rating(which is used to rate the services used by customer by using plan details as rate called as Usage in BRM),Billing(bill the customer every customer bill end date(lets say customer is created on July 15 every month bill end date will be 15 of every month)),Invoicing(sending Invoices to customer),AR(Accounts Receivables)(handling payments,Adjustments,Write off).

Problem described in a project with solution :

Copied from project description

Initially there is no Hadoop Layer in the project but when the customers started using the services data stated increasing day by day and it became really though to Analyze the data to improve the business.Then architects decided to add hadoop layer to analyze data for improving their marketing techniques.That is the first move towards the hadoop and later there are the serious signal drops in USA.Then customer started deactivate their accounts so there is Customer Churn(It is defined as the percentage of subscribers moving from a specific service or a service provider to another in a given period of time.).To retain back the customers Vodafone came up with concept call Customer Churn Analysis.as part of this analysis Vodafone had given a facility to customers to register their complaints through Customer Portal.Once the complaint is registered they will be returned with a complaint or registration id.In the same way sensors will be keeping sending the CRDS through online Mediation Controller when call is initiated by the customer of every region these logs will be moved to hadoop cluster using flume for analysis.By analyzing the customer complaint and sensor logs .request will be raised to network team to fix the issue. This is the second bigger move towards the Hadoop.

Before starting the task let us first discuss the phases of the project:

There are three phases in this project that is :

1) Data Migration	2) Data Processing	3) Data Analysis

Data Migration : **Data migration** is the process of transferring data between computer storage types or file formats. It is a key consideration for any system implementation, upgrade, or consolidation. Data migration is performed to achieve an *automated migration*, freeing up human resources from tedious tasks. Data migration occurs for a variety of reasons, including server or storage equipment replacements, maintenance or upgrades, application migration, website consolidation and data center relocation.

To achieve an effective data migration procedure, data on the old system is mapped to the new system utilising a design for data extraction and data loading. The design relates old data formats

to the new system's formats and requirements. Programmatic data migration may involve many phases but it minimally includes *data extraction* where data is read from the old system and *data loading* where data is written to the new system.

After loading into the new system, results are subjected to data verification to determine whether data was accurately translated, is complete, and supports processes in the new system. During verification, there may be a need for a parallel run of both systems to identify areas of disparity and forestall erroneous data loss.

Automated and manual data cleaning is commonly performed in migration to improve data quality, eliminate redundant or obsolete information, and match the requirements of the new system.

Data migration phases (design, extraction, cleansing, load, verification) for applications of moderate to high complexity are commonly repeated several times before the new system is deployed.

Data Processing :

Data processing is, generally, "the collection and manipulation of items of data to produce meaningful information."[1] In this sense it can be considered a subset of *information processing*, "the change (processing) of information in any manner detectable by an observer." [note 1]

The term Data Processing (DP) has also been used to refer to a department within an organization responsible for the operation of data processing applications.

Data processing functions

Data processing may involve various processes, including:

- Validation – Ensuring that supplied data is correct and relevant.
- Sorting – "arranging items in some sequence and/or in different sets."
- Summarization – reducing detail data to its main points.
- Aggregation – combining multiple pieces of data.
- Analysis – the "collection, organization, analysis, interpretation and presentation of data."
- Reporting – list detail or summary data or computed information.
- Classification – separation of data into various categories