

# BI Implementation in Telecommunication

Bibek Subedi

Yomari Information Services

[bibek.subedi@yomari.com.np](mailto:bibek.subedi@yomari.com.np)

*This report aims at discovering the business intelligence technologies that can be applied to the telecommunication industries to facilitate the business process. I have used the data warehouse approach to build the warehouse of telecom data from which various reporting tools can be used to visualize these data and perform various types of analysis to support the decision making process. Furthermore, I have pointed out some data mining methods that can be added to make it more intelligent as its name “Business Intelligence”. This report shows all the steps from data collection to report generation through warehouse ETL process with related technological background and tools and technologies wherever necessary.*

## Data Set

Data is the core of any business intelligence analysis. The main source of data in telecommunication is the customers’ Call Detail Record (CDR). CDR is actually a log event by the carrier’s network equipment. Many hidden patterns can be obtained from those raw CDR data. The CDR data sample is given in APPENDIX I. It contains the record of each call generated by calling parties. For example customer’s MSIDN number, IMISI, IMEI, Call plan, Call type, MSIDN of corresponding party, Duration, Time and Date of call. Another source of data is Customer Data Records. It contains the Telecom and Business Data of each customer. Telecom data consists of customer’s MSIDN, IMEI, IMISI, Subscription plan whereas Business Data consists of Customer’s Name, Address, Gender, Age, Marital Status etc. This information is needed for geographic and demographic analysis. The sample Customer Data is given in APPENDIX I. The CDR and Customer Data Records are almost sufficient for building business intelligence system.

## Fact and Dimension Identification

The first step in building the data warehouse system is identification of fact and dimension tables. The number of type of dimension and fact tables designs the final analytics and dashboards. Here I have identified the following dimension tables

- ⌚ Time Dimension
- ⌚ Date Dimension
- ⌚ Location Dimension
- ⌚ Customer Dimension
- ⌚ Services (Plan) Dimension

The **Time Dimension** associates the time-stamp of each call. Information held by time dimension includes the start and ending of events occurring during call. **Date Dimension** is used in conjunction with time dimension in order to specify the exact occurrence of the call. **Location Dimension** is used for specifying the location of the calling parties. Location is specified through area codes. The higher level tables contain the zone and region specific data corresponding to the individual district obtained by the cross reference of area codes and the district information. **Customer Dimension** is center to all analytics in the data and is related to other dimensions in a manner that suits the purpose of the analysis.

The simultaneous correlation of the customer dimension with other dimensions assist us in generating facts associated with the spatio-temporal events recorded by the CDR switches. **Services Dimension** provides the functionality that could add up to various other system functionalities and services. It helps us analyze the various aspects of interactions of a user to a telecommunication industry and its services.

I have identified a single **Fact Table** for the analysis. The table consists of all the primary keys of various base dimension tables along with the measurement facts. Fact table is the final table generated after the ETL process on various sources and dimension tables. Fact tables are loaded in the data warehouse and are the source to perform mining operations and acquiring useful information. The fact table holds two main measurement metrics which are duration and cost. Duration field provides information about the total time-span of the execution of a call whereas the cost field provides information regarding the expenses incurred during the call. The warehouse schema using Snowflake mode is given in APPENDIX II.

### **Extract, Load and Transform**

The raw data obtained from the CDR switches and customer form is not organized. Only limited data fields are required for the analysis and various cleaning operation must be done in order to make it more structured for the analysis. Similarly data may be in different format and should be transformed to a single format before loading to the warehouse. ETL process starts with the Extract operation. The **extraction** step is conceptually the simplest task of all, with the goal of identifying the correct subset of source data that has to be submitted to the ETL

workflow for further processing. For extraction, there needs to be a logical data map before the physical data can be transformed. The logical data map describes the relationship between the extreme starting points and the extreme ending points of the ETL system usually presented in a table or spreadsheet.

**Transformation** is the main step where the ETL adds value. Data can actually have different issues which can be classified into: schema-level problems, record-level problems or value-level problems. To deal with such issues, the integration and transformation tasks involve a wide variety of functions, such as normalizing, denormalizing, reformatting, recalculating, summarizing, merging data from multiple sources, modifying key structures, adding an element of time, identifying default values, supplying decision commands to choose between multiple sources, and so forth.

The end of the source records' journey through the ETL workflow comes with their **Loading** to the appropriate table. The source data is first loaded into staging table; data transformation is performed and finally loaded into the target table.

The whole ETL process is carried using UNIX scripting. It may be BASH or Korn shell scripting. As file operation and manipulation is easy using UNIX shell scripting, it's the most preferable for ETL operation. The data loading process (to Oracle DB) can be done using control file. The control file maps the source data with database table and provides optional attributes that facilitates the loading process. After ETL, the data is loaded into the data warehouse in suitable format.

### **Business Intelligence and Reporting**

The core part of the BI is analysis of the warehoused data. There are several tools available for data visualization and analysis as well as data mining. Some of the tools are listed below

**OBIEE:** Oracle Business Intelligence Enterprise Edition Makes corporate data easier for business users to access. It provides a common infrastructure for producing and delivering enterprise reports, scorecards, dashboards, ad-hoc analysis, and OLAP analysis. It includes rich visualization, interactive dashboards, a vast range of animated charting options, OLAP-style interactions and innovative search, and actionable collaboration capabilities to increase user adoption and reduces cost with a proven Web-based service-oriented architecture that integrates with existing IT infrastructure

**MicroStrategy:** MicroStrategy, Inc., is a business intelligence (BI) software vendor whose software enables organizations to analyze vast amounts of data to make strategic business decisions. The software delivers actionable information to business users via the web and mobile devices, including Apple, Android and BlackBerry smart phones and tablets. The software includes technologies to capitalize on the big data, mobile, cloud, and social media trends. MicroStrategy Report Services generates dashboards, scorecards, and production reports.

**JasperReports Library:** The JasperReports Library is the world's most popular open source reporting engine. It is entirely written in Java and it is able to use data coming from any kind of data source and produce pixel-perfect documents that can be viewed, printed or exported in a variety of document formats including HTML, PDF, Excel, OpenOffice and word. iReport is a visual

designer for JasperReports. iReport is free, open source report designer for JasperReports and JasperReports Server. Create very sophisticated layouts containing charts, Images, sub reports, crosstabs and much more. Access the data through JDBC, TableModels, JavaBeans, XML, Hibernate, CSV, and custom sources. Then publish the reports as PDF, RTF, XML, XLS, CSV, HTML, XHTML, text, DOCS, or OpenOffice.

**BIRT:** BIRT is an Eclipse-based open source reporting system for web applications, especially those based on Java and Java EE. BIRT has two main components: a report designer based on Eclipse, and a runtime component that you can add to your app server. BIRT also offers a charting engine that lets you add charts to your own application.

**SpagoBI:** SpagoBI is an integration platform focused on business intelligence needs at the enterprise level. It's a full open source solution. SpagoBI offers a complete analytical layer (e.g. reporting dimensional analysis, OLAP, data mining, dashboards, free and visual data inquiring) with many solutions for every area.

**RapidMiner:** RapidMiner is the world-wide leading open source data mining solution due to the combination of its leading-edge technologies and its functional range. Applications of RapidMiner cover a wide range of real-world data mining tasks.

**JPivot:** JPivot is a JSP custom tag library that renders an OLAP table and let users perform typical OLAP navigations like slice and dice, drill down and roll up. It uses Mondrian as its OLAP Server. JPivot also supports XMLA datasource access.

In the analysis part following analysis is mandatory in telecommunications.

### **OLAP Analysis:**

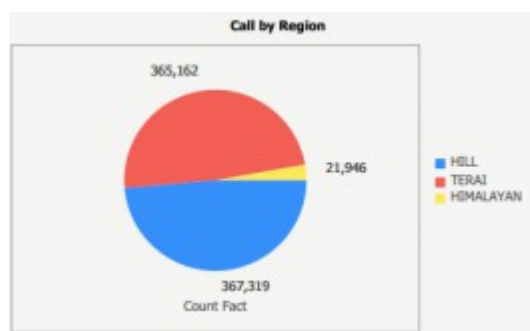
In this layout, data from the intelligent cube (OLAP Cube) is taken and presented to create different reports. The main benefit of OLAP analysis is that it provides the drilling capabilities while viewing the reports as well as provides caching of data for faster response in dashboard navigation. Under OLAP analysis, following analysis can be done

**Call Analysis:** Call duration is analyzed in the basis of number of calls. What call duration do most users prefer? This question is answered. Sample result may look like



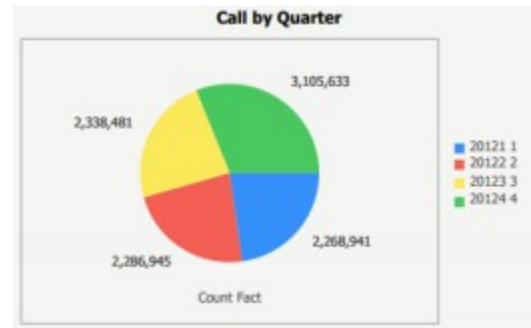
*Fig: Call Count by Duration*

**Geographic Analysis:** This includes call patterns on the basis of geographical regions like district, zones and development regions. Sample result may look like



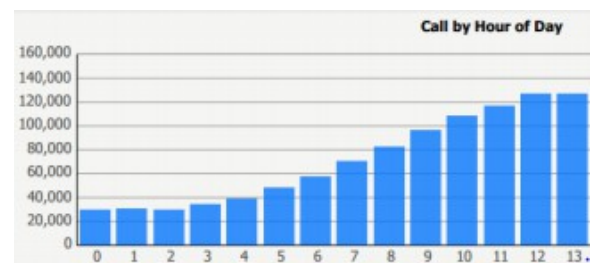
*Fig: Call by Region*

**Datewise Analysis:** This includes analysis of call pattern on the basis of time span. For example call by day, call by quarter, call by month etc. The sample report looks like



*Fig: Call by Quarter*

**Timewise Analysis:** Similar to the datewise analysis, the timewise analysis presents the call pattern of a day. This includes call by hour of a day, call by hour of a day of the week etc. Sample report may look like.



*Fig: Call by Hour of Day*

**Revenue Analysis:** Revenue is by far the most important thing for a business enterprise and telecommunication enterprise is not an exception. Their objective is generally to earn profits by giving quality service to customers. Sample report may look like.

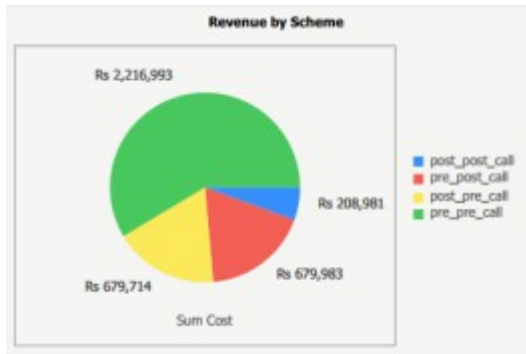


Fig: Revenue by Scheme

**User Analysis:** User analysis shows the call patterns based on user information like gender, age, age-range, occupation etc. The sample report may look like

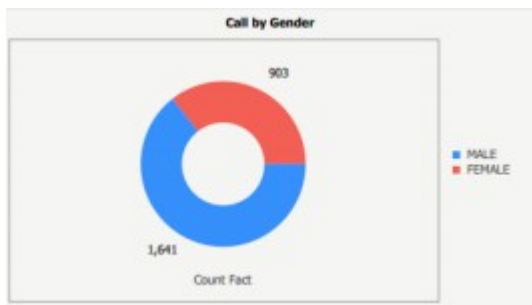


Fig: Call by Gender

**Loss Analysis:** Call dropping is a serious issue in telecommunication industry. Many calls don't get connected because they can't reach the receiving server. The layout showing comparison of dropped calls and succeeded calls on the basis of time of day can be created visualize the loss analysis.

## Use of data mining in Knowledge Discovery

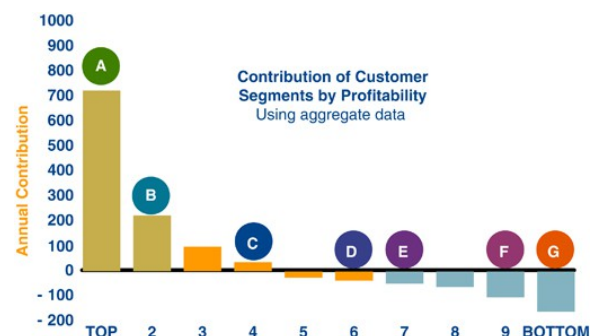
The above analysis is just visualizing the data from warehouse. We used different reporting tool to generate the various kinds of reports. In today's competitive market,

without knowledge discovery from the data, simple visualization cannot do much. The hidden pattern that exists in the tons of data can be discovered by the use of various data mining and machine learning techniques. I have compiled some of the area where telecommunication needs the data mining techniques.

**Customer Segmentation:** Customer segmentation allows a company to target specific groups of customers effectively and allocate marketing resources to best effect. Most of the telecommunications use the following three kinds of segmentations

### 1. Customer Value Segmentation:

This calculates a value measure for each customer, sorts the customer base into descending order by value and then splits and base into 10 equal segments. The first or top decile is the top 10 percent of the base. The second decile is the next 10 percent, and so on. The number of segments may vary based on the number of customers. A sample report may look like



### 2. Customer Behavior Segmentation

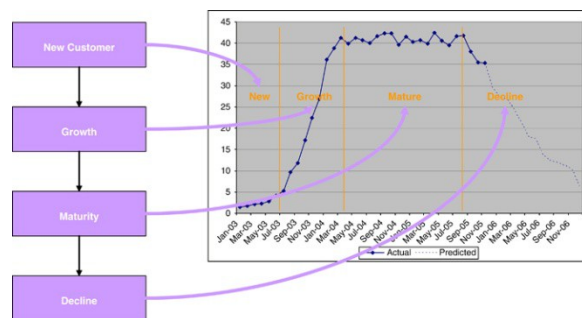
The second form of segmentations that most marketers will be familiar with is segmentation according to customer calling

behavior.

Segment	Percentage high value
Quick talkers	9
SMS but will talk	43
Valuable roamers	4
SMS but high top-ups	19
Outbound voice sociables	25

### 3. Customer Lifecycle Segmentation

Life cycle segmentation is understanding and predicting what a customer is likely to do at each stage in the life cycle: its purpose is to move the customer profitably to the next stage. The life cycle event includes the point at which an individual becomes a new customer, a growth stage at which they opt for an upgrade, a maturity stage during which phone usage is stable, and, possibly, another growth stage when they obtain a second phone. Finally, there is decline stage. The sample report may look like

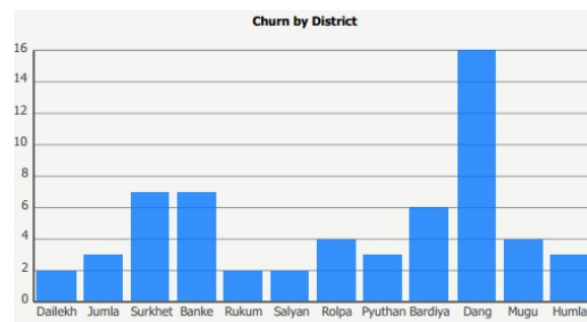


Customer can be stimulated at maturity into growth through new services and appropriate offers.

Various clustering algorithm like K-Means can be implemented for the customer segmentation. RFM (Recency, Frequency, and Monitory) method is very popular for finding out the best customers by examining how recently a customer has purchased, how often they purchase, and how much the

customer spends. RFM analysis is based on the marketing axiom that “80% of your business comes from 20% of your customers.”

**Customer Churn Prediction:** Customer churn prediction allows telecommunication to predict the customers that are ceasing the business with the company. The ability to predict that a particular customer is at high risk of churning while there is still time to do something about it represents the huge additional potential revenue for a telecommunication company. Besides, the direct loss results from customer abandoning the business, the most of acquiring that customer many not have already been fulfilled by his/her spending to date. So churn prediction is another important factor to be considered in today’s competitive world. A sample report of churn prediction may look like



**Network anomalies detection:** The analysis of behavior of different cells in the network is another problem that we encountered. Each cell in the network is associated with different parameters. These parameters are divided into a few subgroups. First of them are configurations parameters for every cell and other attributes are

parameters gathered during some period of time and describe the behavior of certain network elements assigned to this cell.

For example parameter Attempts gives the total number of requests to allocate a channel and the parameter Blocks gives the number of unsuccessful requests. If Blocks/Attempts achieve value above 2% it means that the quality of service is not good enough and network optimization is needed. There are several types of errors that indicate cell anomalies such as channel congestion, blocking, call drops etc on levels higher than acceptable for well designed network.

## **Conclusion**

This research tried to explore the different methods that can be applied in business intelligence for telecommunication

industries. The main source of data was described and how those data can be processed using ETL process is stated. Finally techniques for reporting those data using interactive and attractive graphs and UI were described with added data mining intelligence.

## **References**

A. Acharya, S. Ghimire, S. Kafle, S.R.Shrestha.

"Corporate Decision Support Sytem for Telecommunicaion." 2013.

B. Timilsina, B. Bhattarai, N. Kandel, N. Karki.

"Subscriber Data Mining for Business Reporting and Decision Making in Telecommunications." 2013.

Weiss, Gary M. "DATA MINING IN TELECOMMUNICATIONS." Business Intelligence Journal (2011).



## APPENDIX I

### CDR Sample data

```
# MSIDN:IMSI:IMEI:PLAN:CALL_TYPE:CORRESP_TYPE:CORRESP_ISDN:DURATION:TIME:DATE
068373748102;208100167682477;351905149071;PLAN1;MOC;CUST1;0612287077;247;12:07:12;01/01/2012
068373748102;208100167682477;351905149071;PLAN1;MTC;CUST2;0600000001;300;12:15:09;01/01/2012
068373748102;208100167682477;351905149071;PLAN1;SMS-MO;CUST1;0613637193;0;12:18:18;01/01/2012
068373748102;208100167682477;351905149071;PLAN1;SMS-MT;CUST1;0612899062;0;12:21:07;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;MOC;CUST1;0612283725;90;12:00:00;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;MOC;CUST1;0613069656;82;12:02:27;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;MOC;CUST1;0613481951;78;12:04:41;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;MTC;CUST2;0600000001;92;12:07:13;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;MTC;CUST2;0600000002;94;12:09:40;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;MTC;CUST1;0612063352;114;12:12:40;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;SMS-MO;CUST1;0613103364;0;12:13:42;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;SMS-MO;CUST1;0613751973;0;12:14:44;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;SMS-MO;CUST1;0613672843;0;12:15:44;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;SMS-MT;CUST1;0612769488;0;12:16:42;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;SMS-MT;CUST1;0613164676;0;12:17:39;01/01/2012
065978198280;208100310191699;356008289837;PLAN3;SMS-MT;CUST1;0613399901;0;12:18:39;01/01/2012
067599860569;208120276653317;353297808290;PLAN2;MOC;CUST1;0612089847;116;12:00:00;01/01/2012
```

### Customer Sample Data

```
0653012098;208210235495782;351039632010;Bouygues Telecom;Mirth;Mcbride;PLAN2
0653160284;208210264675080;359064852839;Bouygues Telecom;Morel;Owen;PLAN2
0653295417;208210197752919;355331128596;Bouygues Telecom;Myer;Clarke;PLAN1
0604767913;208110388482814;354649633893;SFR;Neddy;Burnett;PLAN3
0633803544;208010296185031;358072528291;Orange F;Nickey;Kirby;PLAN2
0600769719;208100296170083;356013977714;SFR;Nonnah;Keith;PLAN2
0653184421;208200128404116;359285256060;Bouygues Telecom;Ofelia;Beard;PLAN1
0600822336;208020280655249;350819452482;Orange F;Onfre;Prince;PLAN2
0600316735;208200282005384;359162609083;Bouygues Telecom;Otha;House;PLAN2
0600751278;208100255930038;359924218130;SFR;Paolo;Velez;PLAN2
0653081466;208200347359444;352074627130;Bouygues Telecom;Peace;Pennington;PLAN3
0600721021;208100264663048;356723967873;SFR;Petronella;Mcknight;PLAN2
0600901060;208210192285827;354027503078;Bouygues Telecom;Pietra;Berg;PLAN1
0653011432;208210318520371;350097985139;Bouygues Telecom;Quillan;Gay;PLAN3
0653013856;208210153984675;359608198247;Bouygues Telecom;Rashidi;Mcneil;PLAN1
0601176246;208110211867314;351690053785;SFR;Renwick;Wilder;PLAN2
0600910421;208210220301469;358028233413;Bouygues Telecom;Rina;Levy;PLAN2
0653150842;208210255678299;356263608021;Bouygues Telecom;Romelle;Cleveland;PLAN2
0600252590;208100289645308;358550712887;SFR;Rowney;Bright;PLAN2
0601630896;208100228718912;356732541503;SFR;Sal;Slater;PLAN2
```



## APPENDIX II

### Data warehouse snowflake schema

