Delivering Business Intelligence to the Ministry of Communication and Information Technology for Enhanced ICT Capacity Planning in Egypt A Case Study

Ahmed B. Farid[1] and Khalid M. Salama[2]

[1]Faculty of Computers & Information Science, Helwan University, Egypt
[2]School of Computing, University of Kent, UK

Abstract

Due to the critical importance of the operations performed by different sectors of the government, along with the increase of their work size and complexity, analytical high-level IT systems need to be implemented effectively to provide decision makers in the Egyptian government deep insight in their domains and efficiently support their decision making process. Business Intelligence (BI) is a broad category of applications and technologies for gathering, storing, analyzing, sharing and providing access to data, and information to help executive users for making better decisions. This paper proposes a business intelligence solution for the Ministry of Communication and Information Technology (MCIT) in Egypt this is focusing on the business domain of Information & Communication Technology (ICT) infrastructure capacity planning. We describe the business domain and the user's needs, we explore the benefits of providing such a BI solution in this sector, then we go into the details of designing and implementing the various layers of the solutions; data integration, data warehouse, and OLAP cubes.

Keywords: *Business Intelligence, ETL, Data Warehousing, OLAP, Dimensional Modeling.*

1. Introduction

Business Intelligence (BI) is a set of concepts, methods, and technologies designed to pursue the elusive goal of turning all the widely separated data in an organization into useful information and eventually into knowledge [1]. This information has historically been delivered to an organization's analysts and management through reporting and analysis capabilities, but increasingly BI is being delivered to all parts of an organization by integrating smarter capabilities into the applications and tools that people use to perform their everyday jobs. The most successful BI solutions can create exceptionally valuable capabilities for an organization, such as the ability to proactively spot opportunities to improve operational processes and practices.

BI refers to providing an end-to-end data integration, analytical and reporting solution. On the other hand, a data warehouse plays a single role in the BI solution. The data warehouse is a database model that suits the analytical purposes, and is described as a subject oriented, integrated, non-volatile, and time variant collection of data in support of management's analytical and decision making process. According to W. Inmon and E. F. Codd, Online Transaction processing (OLTP) in the operational-level applications and Online Analytical Processing (OLAP) in decision support applications cannot coexist efficiently in the same database environment, mostly due to their very different orientation and transactional characteristics [1] [2].

While the analytic-ready data lives in the data warehouse, a BI solution extends the data warehouse with several layers to provide a complete decision support IT system; a pre-processing data integration layer (also referred to as ETL) that extracts the data from the OLTP source systems an populates the data warehouse, and the a post-processing data materialization, where OLAP cubes are built to store the data in multi-dimensional structure specified for efficient querying and flexible analytics. On top of that, a rich visualization layer is provided to act as the user dashboard, by which the decision maker can monitor, query and analyze his domain operation and performance. Such a complete stacked solution is the true Business Intelligence application, which can add a real value in its implemented domain.

In Egypt, providing the executives, analysts and knowledge workers in government, an integrated, consistent and summarized information about their domain is the key function of the BI systems. Moreover, due to the critical importance of the operations performed by different sectors of the government, along with the increase of their work size and complexity, the crucial need of implementing high-level decision support system increased to enable the decision maker to right in right piece of information, in the right time, from anywhere it is needed. In this paper, we discuss a case study of delivering a BI solution to the Ministry of Communication and Information Technology (MCIT). The domain of interest in this work is the capacity planning, including focus areas of fixed telephone usage, mobile phone usage and internet usage. This work covers the BI application stack from ETL process, data warehouse design, and OLAP cubes implementation. The dashboard and visualization operations are left for future work.

The rest of the paper organized as follows. Section 2 gives a background on BI basic concepts. Section 3 discusses the business domain of the MCIT capacity planning, highlighting the business need and the benefits from providing the BI solution. In Section 4, we exhibit the overall BI solution architecture and define its various parts. The data integration process, the data warehouse implementation and the OLAP cubes design are described in details in Sections 5, 6 and 7 respectively. Finally, we conclude with future work in Section 8.

2. Business Intelligence

With the quick evolution of information and communication technologies and dissemination of computer use, many organizations use automated system to implement their important processes, which are known as OLTP. A lot of data are being produced during the OLTP operations, but the data is not integrated. Such data are stored within one or more platforms and constitute the resource for the organizations. However, OLTP systems are not suitable to manage and store strategic information. They are formed by operational data needed for daily transactions. In terms of decisions, data are empty and without any transparent value for the decision process of organizations. Decisions are taken based on administrators experience and sometimes based on historical facts stored in different information systems.

For that, Decision Support Systems (DSS) [3,4,5,6], also known as Online Analytical Processing (OLAP), or recently Business Intelligence (BI) solutions, have emerged to support the managerial level users though the decision making process. To provide a complete and effective BI solution, the basic different layers in the BI solution are shown in Figure 1. These layers are defined and discussed in detail in the context of the MCIT capacity planning case study later in the paper.

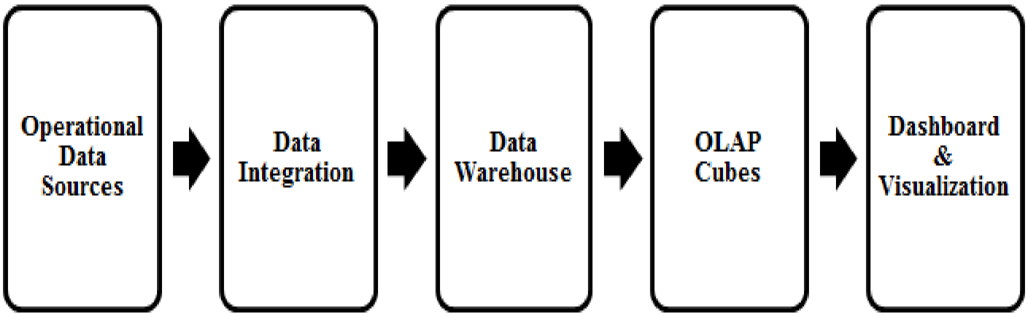


Figure 1 – The BI Solution Layers

3. The MCIT Business Case and the Need for BI

3.1 Business domain

In the last decade, the Egyptian government has made big strides towards establishing Egypt as a potential ICT hub in the region. Based on that, Ministry of ICT (MCIT) has been established on October 1999. Its Major objective was (and still) achieving a Major infrastructure developments towards updating the ICT capacities in Egypt. In addition to that, MCIT has decided to lay a modern communication infrastructure, the government is promoting the use of computers, and the Internet for economic growth – the goal is to become another India or Ireland.

Dr. Ahmed Nazief, the former minister of ICT, launched a National Plan for Communications and Information Technology, establishing projects and initiatives. One of those initiatives was the Egyptian Information Society Initiative (EISI). This initiative has been geared to support, and to empower public-private partnerships in order to develop and expand telecommunications infrastructure that aimed providing ICT access to all citizens, develop meaningful Arabic content, establish a large pool of trained ICT professionals to create and innovate, build a policy framework of supporting infrastructure to foster the growth of a powerful, and competitive ICT industry. Additionally, this infrastructure have been planned to leverage ICT in order to empower development in health, education, government, commerce, culture, Postal services, and other area [7].

Planning for developing such a huge infrastructure is not an easy or a cheap task. The decision making process in that context needs an extensive amount of sorted, cleansed, scrubbed, categorized, and classified information. This information includes but not limited to:

- Fixed line phones indicators
    - Fixed line subscriptions.
    - Fixed line subscriptions in urban areas.
    - Fixed line penetration.
- Mobile phones indicators.
    - Mobile subscriptions.
    - Mobile penetration.
    - Mobile subscriptions annual growth.
- Internet usage indicators
    - Estimated internet users.
    - Internet penetrations.
    - Internet user's annual growth.
    - Internet users through mobile.
    - USB Modem users.
    - ADSL subscribers.
    - ADSL penetration.
    - International Internet bandwidth.

3.2 Business Domain Problem

Calculating the over mentioned sets of indicators should be based on datasets. Unfortunately, these datasets are actually fragmented in a set of multiple systems. This set includes but not limited to; Egypt Telecom OLTP systems that contain the fixed phone lines data as well as the local phone exchanges data, Mobile Operators OLTP systems that includes whatever needed data regarding the mobile phones as well as the 2G/edge/3G internet usages. Additionally, Dial-up & ADSL providers' OLTPs; where the major source of data for internet usages of landlines lay. Moreover, it is important to mention that population numbers are always needed to calculate the penetration ratios of these service into the Egyptian population. Thus, the Database of Central Agency for Public Mobilization and Statistics (CAPMAS) is of a great immense. Without having all these kind of data in one repository, calculating, and comparing the growth, and development of the over mentioned indicators will not be possible.

It is important to mention that the accuracy of the calculated figures is crucial, not only for the planning of the capacity of the Egyptian ICT infrastructure, but also, for two kinds of important stakeholders. The first important stakeholder is the private sector investor. This kind of stakeholders always needs such kind of data for any rigid feasibility study that is taking place in the ICT sector. MCIT used to issue reports with these kinds of indicators for private sector [8], [9]. The Second kind of stakeholders is International Organizations (e.g. ITU, UN, UNCTAD, etc.). Those Organizations need always consolidated information about the Egyptian ICT infrastructure in order to monitor the growth of the development of this infrastructure, population penetration of this infrastructure, etc. [10], [11].

Consolidating these datasets in one data warehouse repository that could be used for calculating the previously mentioned indicators in an integrated way, is not an easy task due to the following issues:

- OLTP systems are non-integrated, nor synchronized. The same lookup definition ID may have different interpretations in different OLTP source systems.
- The quality of the imported data from the above OLTP sources is insufficient. For instance, some data sources talk about governorates that never existed before.
- There is no conformed geographical definition for all source systems. While some user their own geographical division for the transactional data, others use the formal administrative division of the Egyptian state (Governorates based).

Due to that some data are coming in a textual form that is entered in Excel sheets. Double entries are being done with different character for same objects (e.g. local exchange name, company names, governorate names, etc.)

3.3 BI Solution requirements

All work that has been be done is targeting one aim which is; to find out an accurate as well as detailed understanding levels for the ICT service's penetration inside the Egyptian society based on the available hard data which can lead at the end, for better forecasting for the future capacity of the ICT sector. Understanding the ICT infrastructure capacity can help in; understanding the required governmental future investments into ICT infrastructure. Additionally, it shows the size of the market for more private sector investments in Egypt. This may help the strategic governmental decision makers to take required actions towards increasing the capacity and the quality of the ICT sector infrastructure. In order to reach this level of understanding, it is crucial to have accurate, and consolidated statistical understanding by means of numbers and ratios of the IT sector data that could be navigated in multidimensional hierarchal manner using the, Time, Geographical location, and host. In order to fulfill this objective; an integrated business intelligence solution has to be built to plot the data over a visualization dashboard that delivers the information to the decision maker in an informative shape.

4. MCIT BI Solution Architecture

The outline of the proposed MCIT BI solution architecture is shown in Figure 2. The architecture illustrated in the figure highlights the main layers and components that cover the data integration and reporting needs of the specific the capacity planning business domain discussed in the previous section. The architecture is flexible enough to allow integrating more domains in the future, as was required by the users. The important property of flexibility and extensibility was considered in engineering a user-oriented ETL process, designing the appropriate data warehouse schema, implementing a unified dimensional modeling in the OLAP cubes and developing a rich visualization dashboard.

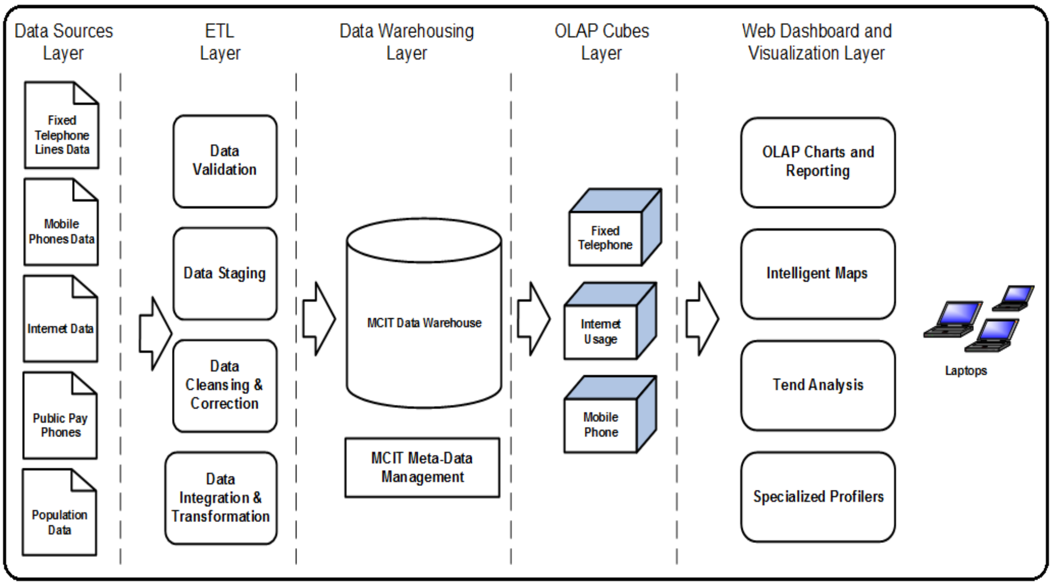


Figure 2 - MCIT BI Solution Architecture

The proposed MCIT BI solution consists of 4 layers. The main tools and characteristics used to in each layer are defined as follows.

ETL Layer: The ETL layer serves as the data integration and transformation in our BI solution. The input for this layer are the excel sheets that include data about the latest operational updates in the fixed telephone lines, the mobile phone lines and the internet usages domains, along with the up-to-date population census. The output of this layer is clean and consistent data to be loaded in the data warehouse. In our solution, we used SQL Server Integration Services (SSIS) to implement the extract, transform, and load functionalities. The ETL layer, which is described in the following section, is responsible for the following operations:

- External data validation.
- Data staging for pre-load user investigation and manipulation.
- Data cleansing and correction.
- Integrating the data from different data sources and transform the data into the appropriate the destination format.
- Meta-data update and error logging and reporting for manual user correction.

The data warehousing layer: We used SQL Server 2008 to implement the MCIT data warehouse. The MCIT data warehouse is the multi-dimensional repository for the integrated and consistent data that serves the analytical and reporting methods. Complex queries usually require a vision of the data from different perspectives. Answers to this type of questions can lead to correct or wrong decisions. The MCIT implements the Star Model, where dominant tables exist in the center of the model. The table in the center is called Facts table. This table has multiple junctions connecting with other tables called Dimension tables. Each secondary table has only one junction with a fact table. The star model has the advantage of being simple and intuitive for analytical purposes.

The facts table contains thousands (or millions) of values and measures of business, as transactions of sales or purchases. The most useful facts are numerical and additive. Facts tables represent many-to-one relationships with business dimension tables. On the other hand, dimensions table store textual descriptions of business dimensions. Each table represents one business dimension, e.g. time and product. One important factor related to facts table is that, as it represents the relationship many-to-many between dimension tables, it has as primary key, a key composed of all foreign keys of dimension tables. The dimensional model presents several advantages [8]:

- Predictable and standard architecture.
- Dimensions of the model are equivalent.
- It is flexible, because it allows the inclusion of new data elements.
- Easy alteration of facts and dimension tables.
- All the applications that existed before the changes continue operating without problems.

Metadata constitutes the nervous system of data warehouse. Without metadata, data warehouse and its associated components in the projected environment become isolated components, functioning independently and with separated objectives [2]. In the operational environment users interact with information through screens and forms, allowing users to be unaware of how information is stored in the database. Metadata is treated later in the process and normally has the same importance as system documentation. A part of the meta-data stored for the measures of the MCIT data warehouse is shown in Figure 3. More details of the MCIT data warehouse are discussed in Section 5.

```
SELECT TOP 1000 [ID]
      ,[Measure]
      ,[Perspective]
      ,[Type]
      ,[Unit]
      ,[Periodicity]
      ,[Source]
      ,[CalculationMethod]
  FROM [MCIT_DW].[dbo].[MetaData.Measures]
  where [Perspective] = 'Local Exchanges'
```

Results | Messages

	ID	Measure	Perspective	Type	Unit	Periodicity	Source	CalculationMethod
1	23	Cumulative Local Exchanges Capacity	Local Exchanges	Cumulative	Line	Month	Telecom Egypt	NULL
2	27	Cumulative Number of Fixed Lines Subscribers	Local Exchanges	Cumulative	Subscriber	Month	Telecom Egypt	NULL
3	30	Cumulative Number of Local Exchanges	Local Exchanges	Cumulative	Exchange	Month	Telecom Egypt	NULL
4	38	Cumulative Waiting List for Fixed Lines	Local Exchanges	Cumulative	Line	Month	Telecom Egypt	NULL
5	144	Local Exchanges Capacity	Local Exchanges	Regular	Line	Month	Telecom Egypt	NULL
6	145	Local Exchanges Capacity Parallel Growth	Local Exchanges	Regular	Line	Month	Telecom Egypt	Current Value - Parallel Value
7	146	Local Exchanges Capacity Parallel Growth Ratio	Local Exchanges	Regular	Percent	Month	Telecom Egypt	(Current Value - Parallel Value)\Parallel Value
8	147	Local Exchanges Capacity Sequential Growth	Local Exchanges	Regular	Line	Month	Telecom Egypt	Current Value - Previous Value
9	148	Local Exchanges Capacity Sequential Growth Ratio	Local Exchanges	Regular	Percent	Month	Telecom Egypt	(Current Value - Previous Value)\Previous Value
10	214	Number of Fixed Lines Subscribers	Local Exchanges	Regular	Subscriber	Month	Telecom Egypt	NULL
11	215	Number of Fixed Lines Subscribers Parallel Growth	Local Exchanges	Regular	Subscriber	Month	Telecom Egypt	Current Value - Parallel Value
12	216	Number of Fixed Lines Subscribers Parallel Growt...	Local Exchanges	Regular	Percent	Month	Telecom Egypt	(Current Value - Parallel Value)\Parallel Value
13	217	Number of Fixed Lines Subscribers Ratio To All	Local Exchanges	Regular	Percent	Month	Telecom Egypt	NULL
14	218	Number of Fixed Lines Subscribers Ratio To Capa...	Local Exchanges	Regular	Percent	Month	Telecom Egypt	NULL
15	219	Number of Fixed Lines Subscribers Ratio To Popu...	Local Exchanges	Regular	Percent	Month	Telecom Egypt	NULL
16	220	Number of Fixed Lines Subscribers Sequential Gro...	Local Exchanges	Regular	Subscriber	Month	Telecom Egypt	Current Value - Previous Value

Figure 3 - MCIT Measures Meta-data Table

OLAP Layer: On-Line Analytical Processing (OLAP) is a different kind of database technology designed specifically for BI. Instead of organizing information into tables with rows and columns like a relational database, an OLAP database stores data in a multidimensional format. Rather than trying to get a relational database to meet all the performance and usability needs by the MCIT executive and managerial level, we build an OLAP database that the MCIT users can query for the analytical proposes.

An OLAP cube consists of data from one or more fact tables and presents information to the users in the form of measures and dimensions. In addition to the measures, we implement complex calculations and analytical logic to the cube to meet the analysis requires of the users. Moreover, a careful aggregation design was carried out to improve the querying efficiency. For the infrastructure capacity planning domain, we implemented three subjective OLAP cubes; each is responsible representing a specific analytical theme in the domain, namely Fixed Lines, Mobile Phones, and Internet Usage. Each analytical theme has a set of measure groups, where each has a set of measures and calculations. More details are shown in Section 7. We used SQL Server Analysis Services (SSIS) 2008 to implement the OLAP database. Figure 4 shows the object browser of the MCIT_BI OLAP database, which includes the three aforementioned cubes.

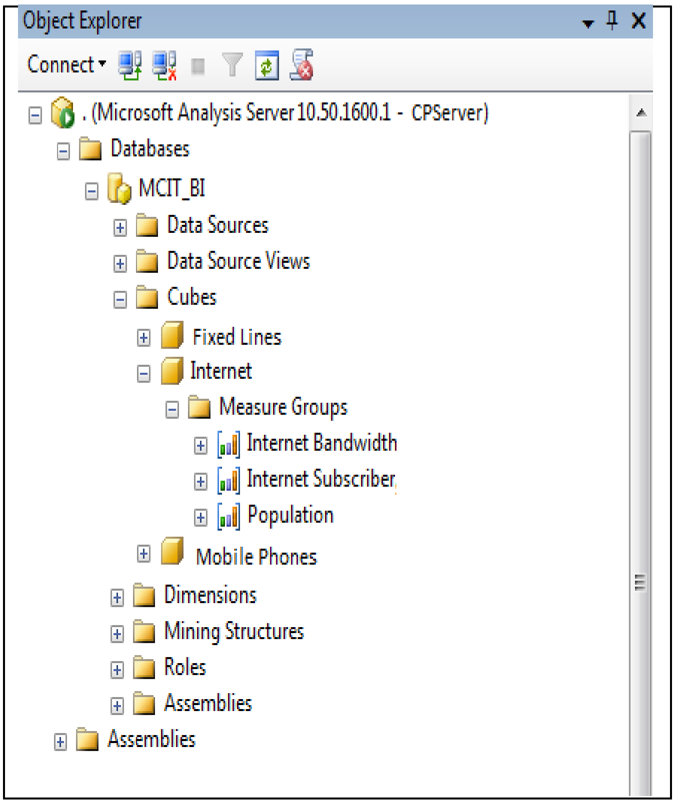


Figure 4 - MCIT_BI SSAS OLAP Cubes

Web-based Dashboard and Visualization Layer: This layer provides the MCIT executive user with tools and capabilities to easily query and the data of it domain of concern. This layer should provide applications such as OLAP charting and reporting, performance monitoring gauges, trend analysis, and specialized profilers. This part is not cover in the current paper, and left to future work.

5. The MCIT ETL Process

The following diagram shows the conceptual design of the MCIT ETL layer. The diagram exhibits the ELT various parts and components as well as the direction of the data follow through the ETL system. As shown in Figure 5, and according to the scenario of data handling in MCIT, the data integration and transformation process is performed in two steps. The first step is to stage the data into an intermediate database, where the second step is to do the necessary data cleansing and transformation to load the data to the data warehouse.

In the first step, the knowledge workers in the MCIT receives new excel sheets, from various external systems, including new data about the domain of concerns, i.e. fixed lines, internet usage and population. More precisely, for each analytical theme in the aforementioned domains, a specific excel sheet with a fixed format should be used. For each analytical theme, we fixed a schema (file format) to be the contract between the user and the ETL system that would take this file, understand the format and parse the data. A sample of the file format concerning the local exchange subscribers' analytical theme, in the fixed lines domain, is shown in Table 1.

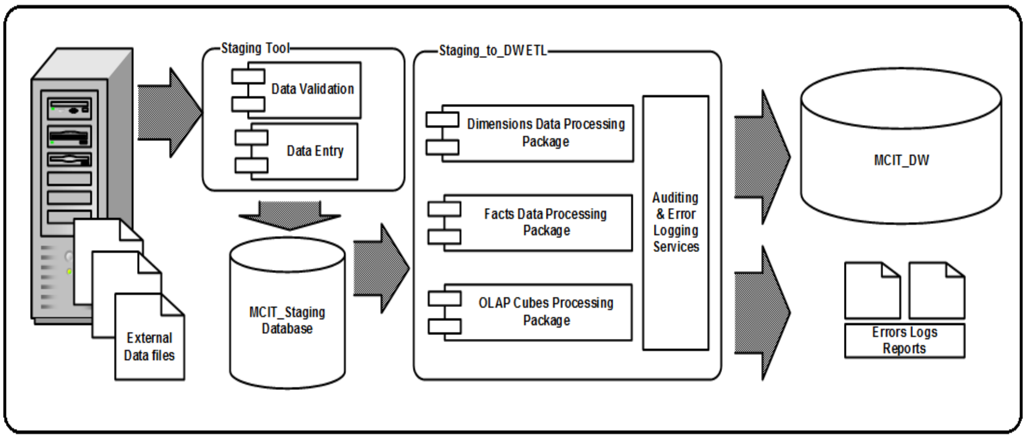


Figure 5 - MCIT Data Integration and Transformation Layer

Table 1: A sample source file format description and the destination table in the MCIT_staging Database

Local Exchange Subscriber – Source to Distention			
Excel Data Source		**Staging Table**	
File Name	Centrals Subscribers.xls	**Table Name**	Stage.CentralSubscribers
Column		**Column**	**Data Type**
Gov		Gov	nvarchar(255)
Central_Name		Central_Name	nvarchar(255)
Sub_Type		Sub_Type	nvarchar(255)
Number_of_subscribers		Number_of_subscribers	Int
conn_type		conn_type	nvarchar(255)
Year		Year	Smallint
Month		Month	Tinyint

The data in each excel sheet is mapped to a table in the MCIT_Staging database, to be stored for further processing before it is loaded to the data warehouse. For this step, we developed MCIT Staging_Tool, a Windows based UI application that enables the knowledge workers to select the wanted excel sheet and loaded to the MCIT_Staging database. The Staging_tool is responsible for performing the following functions. First, it performs basic data validation, which confirms that the selected excel sheet complies with its pre-defined format. Second, after the data is loaded to its correspondence staging table, it enables the user to explore the data in the table and perform quick modifications if needed. These two functions (data validation and data entry) are shown in Figure 5 under the staging_tool entity.

In the second step, the MCIT_ETL performs further data processing to prepare the data to be loaded in the data warehouse. The following are the main tasks performed by the MCIT staging_to_DW ETL:

- Cleansing and Data Correction:
  - Detect null values
  - Detect invalid value type and correct it if possible.
  - Detect redundant data and perform de-duplication.
  - Validate business rules and integrity with current data.
- Data Transformation:
  - Generate artificial keys.
  - Derive Calculations.
  - Perform aggregation (summarizing multiple rows of data).
  - Join and merge data from multiple tables.
  - Translate coded value into its textual representation.

Note that one of the main responsibilities of Staging_to_DW ETL is to log errors and data violation records to be fixed and reloaded. Another responsibility for this ETL is to process the OLAP cubes in order to calculate the aggregations and prepare it for effective and up-to-date analytical processing. Metadata about successful updates and loaded records count are maintained by the ETL with timestamps.

There are two types of ETL packages, fact data processing packages and dimension data processing packages. According to SSIS, In order to load a fact table, the following steps are performed. First, a data flow task is used. We select the appropriate data source component, which is an excel data source in our case. Second, we select the appropriate task as needed in order to perform the previous cleansing and data correction operations. After that, we replace the application codes that came with the source data with the dimensions keys from the data warehouse. For that, we perform a lookup task for each dimension, for each record. If the code\key map is found during the lookup, the record proceeds to the second lookup. Otherwise, it is considered an error, and is annotated with the appropriate error message to be logged in the error report.

Before the data is loaded its destination table in the data warehouse, a fact record should be tested whether it already exists in the Fact table in the data warehouse or not. If there is a fact record in the data warehouse has the same dimension keys as a coming record from the ETL stage, the measures of the fact in the data warehouse are updated by the coming record. Otherwise, it coming record is considered a new record and it is inserted in the data warehouse. Finally, the data is transformed in to the appropriate form to be loaded to the fact table in the data warehouse. Figure 6 and 7 shows a sample ETL package to load the local exchange subscribers' facts' and dimensions' data.

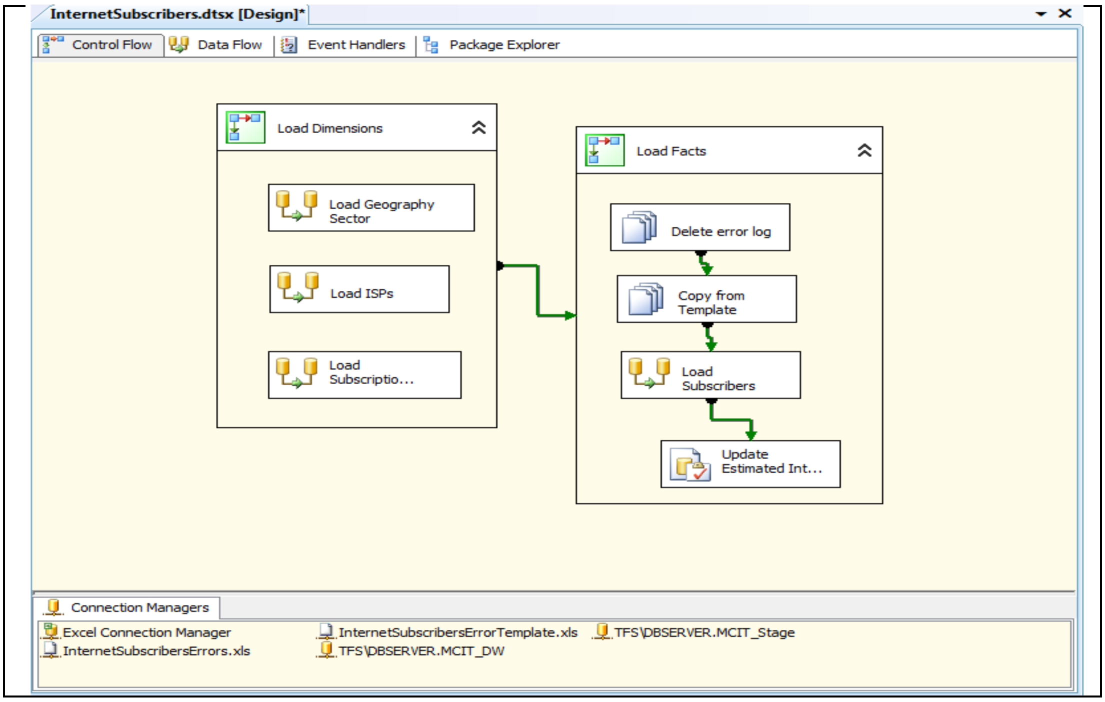


Figure 6 – MCIT_ETL Package for loading Local Exchange Subscribers' data

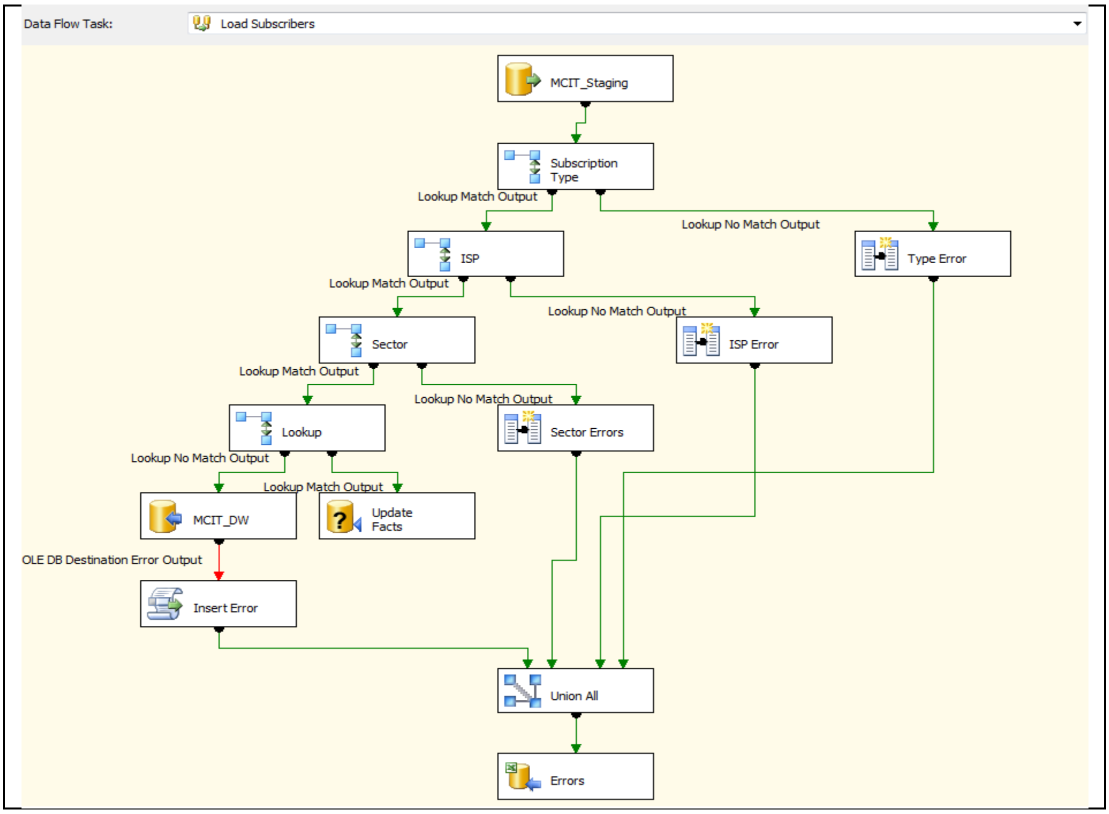


Figure 7 – The SSIS ETL process to load Local Exchange Subscribers' data

6. The MCIT Data Warehouse Design

The approach we employed in designing and developing the data warehouse tin our proposed BI solution for the MCIT is the Kimbal's bottom-up approach [8]. In other words, we started by designing and implementing a set of unified conformed business dimensions, that can serve as analytical entities across different domains in the MCIT. Although we have only started with the telecommunication infrastructure capacity planning - which includes several facts such as Fixed Telephone Lines, Mobile Phones, and Internet, more facts, corresponding to different business domains, can be incrementally added and easily integrated to the data warehouse. The result of approach is the Bus Architecture [8]. Figure 8 shows the Entity Relationship Diagram (ERD) for the data warehouse with the current MCIT domains of concerns.

The bus architecture uses a grid of business functions and dimensions to deliver a set of tightly integrated business analytical domain (also known as data marts). The data warehouse bus architecture is composed of a set of tightly integrated data marts that draw their "power" from a common set of conformed dimensions and facts. A conformed dimension is defined and implemented one time and used throughout the multiple star schemas that make up the consolidated MCIT data warehouse.

To order to deliver a set of unified conformed dimensions, every key knowledge worker and data analyst in the MCIT agreed on a common definition for the dimension, so that the dimension means the same thing no matter where it's used. For example, instead of having a time dimension whose granularity is expressed in weeks, quarters, and years for some dimensional models and in days, months, quarters, and years for other models, we designed one conformed dimension that's used wherever a time dimension is required in the MCIT data warehouse. We kept in mind that the conformed dimension's structure must reflect the finest grain that might be needed for any fact table. Conforming fact involves selecting the appropriate units of measures, and, conform and storing different measures in the same fact table with the same granularity, which we tried to be as fine grained as was available by the data at hands.

There are many reasons to why we use the bus architecture in to design our proposed MCIT data warehouse to serve as the analytical data store for the BI solutions. The following are the main reasons:

- **Efficiency**: A single copy of a dimension (e.g., time, geography) involves less maintenance.
- **Consistency**: By definition, a conformed dimension means the same thing everywhere it's used.
- **Ease of use**: From the users of the MCIT point of view, a conformed dimension can be used easily to perform flexible and consistent cross-domain queries and analysis.
- **Standards Enforcement**: Once everyone in the MCIT has come to consensus on the architecture, meaning, and data content of a conformed dimension, it will become a business standard.
- **Tightly Integrated Analytical Domains:** Conformed dimensions and facts are the backbone architecture of the MCIT data warehouse. Thus, new domain facts can be add to the "bus" when it's needed, knowing that it will integrate with all the pre-existing data marts.

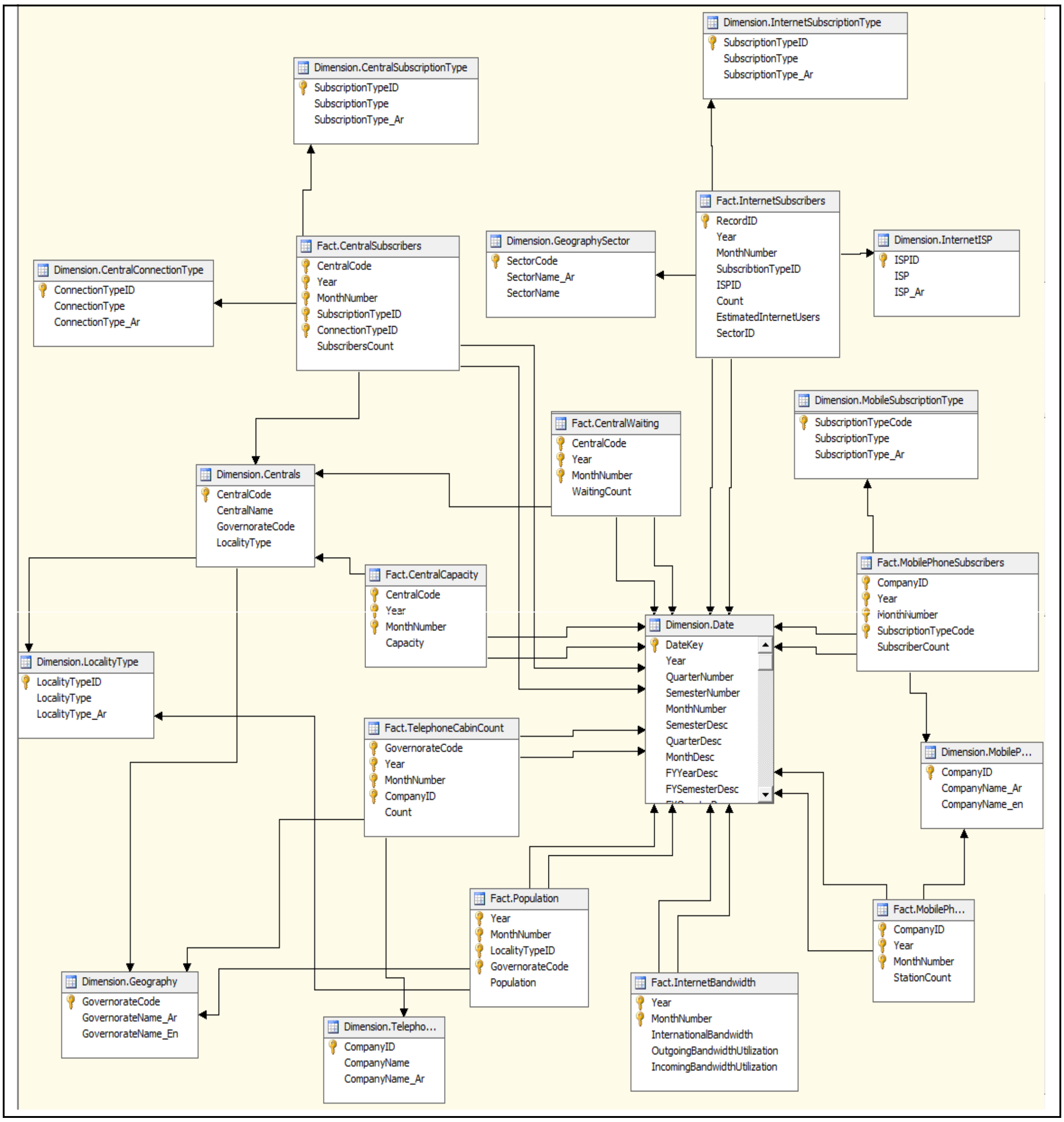


Figure 8 – The MCIT data warehouse consolidated ERD

Next, we describe two of the most important conformed dimensions that are used with (all most) all the analytical themes in the MCIT: the Time and the Geography dimensions. We present the attributes of the dimensions as well as their available hierarchies. The fact tables that correspond to the MCIT various analytical themes are described in the context of the subject-oriented OLAP cubes, next section.

Time Dimension

Time dimension (represented in the [Dimension.Date] table) enables chronological analysis for various facts and views measures in different levels of abstraction and detail. Time dimension adds intelligence to analytical theme by which complex calculation can be performed such as Year-To-Date value, Sequential Growth, Parallel Growth and many more. The Table 2 shows the logical design of time dimension in the context of OLAP cubes. Attributes are listed, values are given for illustration. Hierarchies are exhibited in Table 3.

Table 2: Time Dimension's Attributes Description

Attribute	Sample Value
Date	21/07/2001
Month	March
Calendar Quarter	Q3 CY 2001
Calendar Semester	H2 CY 2001
Calendar Year	CY 2001
Fiscal Quarter	Q1 FY 2002
Fiscal Semester	H2 FY 2001
Fiscal Year	2001

Table 3: Time Dimension Hierarchies

Hierarchy	Levels
Fiscal	Fiscal Year
	Fiscal Semester
	Fiscal Quarter
	Fiscal Month
	Date
Calendar	Calendar Year
	Calendar Semester
	Calendar Quarter
	Calendar Month
	Date

Geography Dimension

Geography dimension location intelligence the analytical theme as it enables analysis of the facts according to its geographical distribution. A conformed geographical dimension is the key element for building intelligent maps and visualizing various measures and indicators form different domains in the same location –aware dashboard. As show in the data warehouse ERD in Figure 8, the geography dimension plays a part of several star schemas for the multi-dimensional model of different analytical theme in the data warehouse. Related population fact outriggers the geography dimension to help calculating penetration indicators for different analytical themes.

Table 4 and 5 show the logical design of geography dimension in the context of the OLAP database, presenting attributes and hierarchies respectively. Ragged dimension design is considered as not all members have descendants and ancestors in each level.

Table 4: Geography Dimension Attributes' Description

Attribute	Sample Value
Governorate	Sohag
Center	Jerja
City	Jerja
Village	El- Barba
Sub-Village	El - Masaed

Table 5: Geography Dimension Hierarchy

Hierarchy	Levels
Geography	All
	Governorate
	Center
	City
	Village
	Sub-Village

7. The MCIT OLAP database Implementation

The OLAP database, as mentioned in Section 2, is built on top of the data warehouse, to provide a multi-dimensional structure to store and query the data in a flexible and efficient manner. The scope of the current MCIT data consists of four major categories of measures and performance indicators. Each category represents a separate domain of concern, in which each contains several sub-domain analytical themes. Each domain is represented in a separate cube as a discrete unit of analysis, while sub-domains are represented in the same cube as perspective s with measure groups. This design balances between cube maintainability and cross analytical theme flexible analysis. The following is the list of cubes in the MCIT BI solution covered in the current work:

- Fixed Telephone Lines.
- Mobile Phones.
- Internet Usage

7.1 Fixed Telephone lines Cube

The Fixed Telephone lines cube provides information and analytical facilities for communication and infrastructure in Egypt. This includes analytical themes of Local Exchange capacity, Fixed Telephone Lines Subscribers, and Fixed Telephone Lines Waiting. Figure 9 shows the overall view of the multi-dimensional design of the cube. Each analytical theme is describes in a separate table next to the figure. A sample of how the Fixed Telephone lines cube is browsed is using its related dimensions is shown in Figure 10.

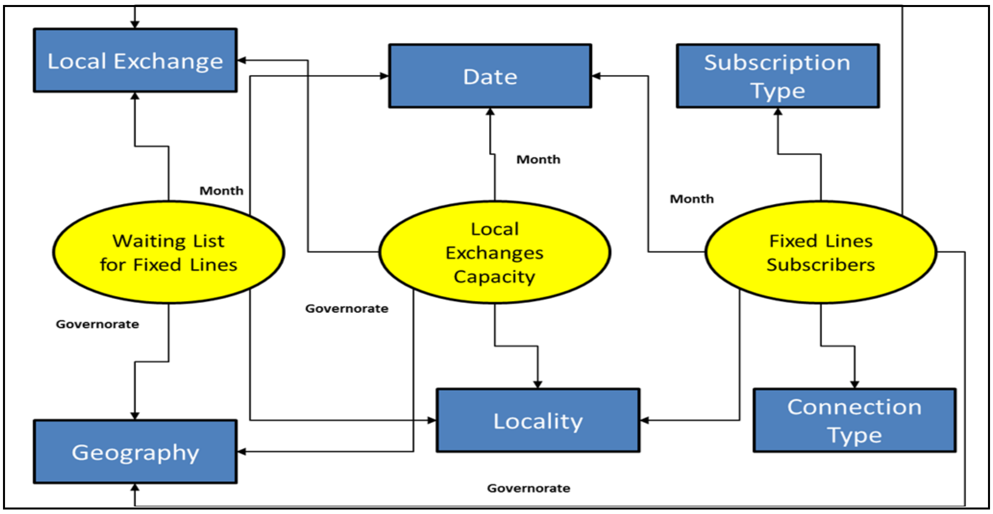


Figure 9 – The Fixed Telephone Line multi-dimensional cube design

Fixed Telephone Lines Subscribers Analytical Theme

Fact Table	**Measure\Calculation**	**Aggregation**
Fact.CentralSubscribers PK **CentralCode** PK **Year** PK **MonthNumber** PK **SubscriptionTypeID** PK **ConnectionTypeID** SubscribersCount	Number of fixed Lines Subscribers	Current - Previous
	Cumulative Number of fixed Lines Subscribers	Last None Empty
	Derived Indicators	Parallel Growth, Sequential Growth, per capita

Local Exchanges Analytical Theme

Fact Table	Measure\Calculation	Aggregation
Fact.CentralCapacity PK CentralCode PK Year PK MonthNumber Capacity	Local Exchanges Capacity	Current - Previous
	Cumulative Number of Local Exchanges Capacity	Last None Empty
	Number of Local Exchanges	Distinct Count
	Cumulative Number of local Exchanges	Sum
	Derived Indicators	Parallel Growth, Sequential Growth

Fixed Telephone Lines Waiting Analytical Theme

Fact Table	Measure\Calculation	Current - Previous
Fact.CentralWaiting PK CentralCode PK Year PK MonthNumber WaitingCount	Waiting List for Fixed Lines	Last None Empty
	Cumulative Waiting List for Fixed Lines	Parallel Growth, Sequential Growth
	Derived Indicators	Current - Previous
		Last None Empty
		Parallel Growth, Sequential Growth

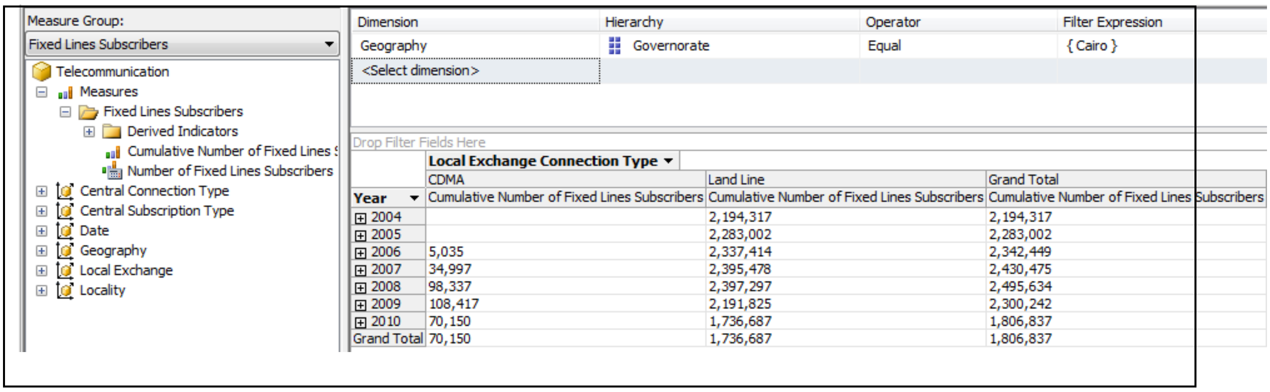


Figure 10 – A Sample output of browsing Fixed Telephone Lines cube to show the Cumulative Number of Fixed Line Subscribers' measure across the time dimension and filtered by "Cairo" member in the geography dimension. The data is diced using the local exchange dimension members.

7.2 Mobile Phones Cube

This cube provides information to the analytical themes concerning mobile phone users and the mobile stations. Figure 11 shows the overall multi-dimensional design of the cube, while each analytical theme is described in a separate table after the figure. Figure 12 shows output of browsing the Mobile Phones Cube.

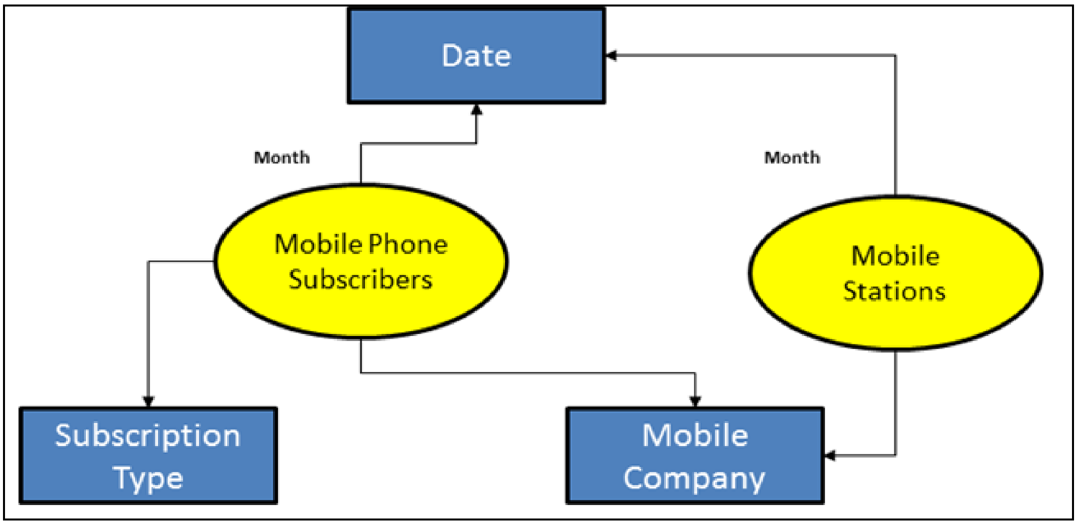


Figure 11 – The Mobile Phone multi-dimensional cube design

Mobile Phone Subscribers Analytical Theme

Fact Table	Measure\Calculation	Aggregation
Fact.MobilePhoneSubscribers PK **CompanyID** PK **Year** PK **MonthNumber** PK **SubscriptionTypeCode** SubscriberCount	Number of Mobile Line Subscribers	Current - Previous
	Cumulative Number of Mobile Line Subscribers	Last None Empty
	Derived Indicators	Parallel - Sequential Growth, Penetration

Mobile Phone Stations Analytical Theme

Fact Table	Measure\Calculation	Aggregation
Fact.MobilePhoneStations PK **CompanyID** PK **Year** PK **MonthNumber** StationCount	Number of Mobile Station	Current - Previous
	Cumulative Number of Mobile Station	Last None Empty
	Derived Indicators	Parallel - Sequential Growth, Penetration

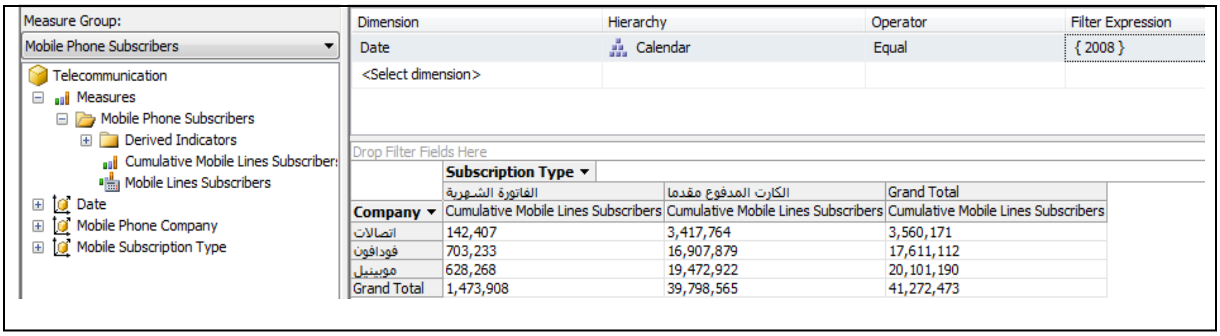


Figure 12 – A Sample output of browsing Mobile Phone cube to show the Cumulative Number of Mobile Line Subscribers' measure sliced by the "2008" member of the time dimension and diced by both subscription type and company dimensions.

7.3 Internet Usage Cube

The cube provides information about the internet includes both Internet Users and Bandwidth domain of analysis. Figure 13 shows the overall multi-dimensional design of the cube, while each analytical theme is described in a separate table after the figure. Figure 14 shows a sample output of the cube.

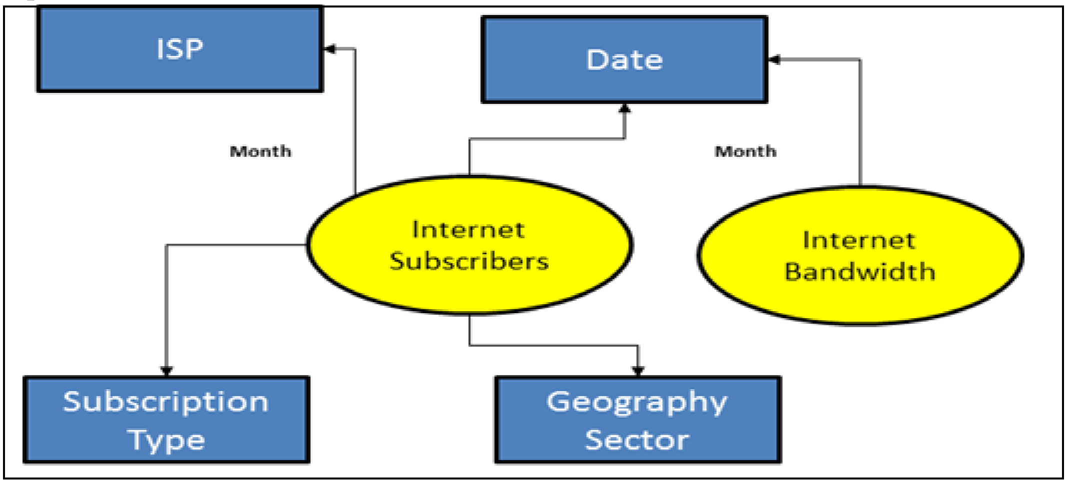


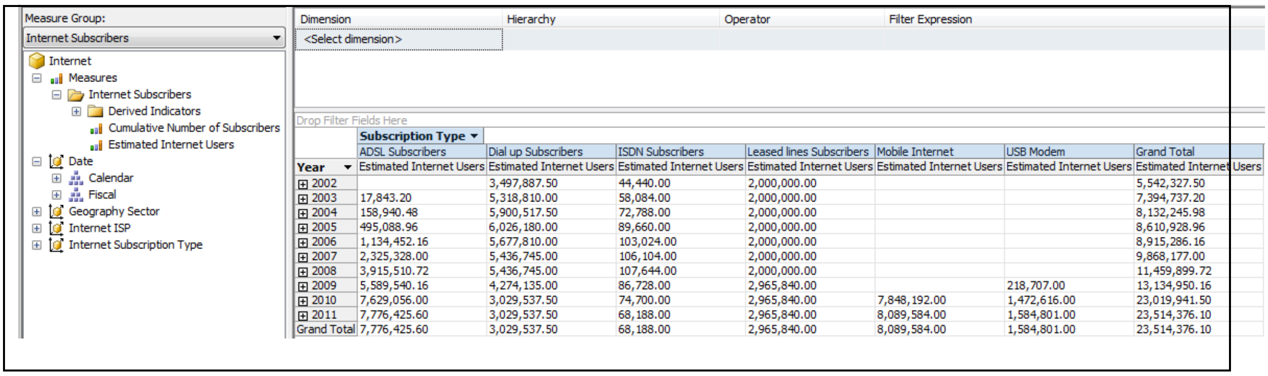
Figure 13 – The Internet Usage multi-dimensional cube design

Internet Subscribers Analytical Theme

Fact Table	Measure\Calculation	Aggregation
Fact.InternetSubscribers PK Year PK MonthNumber PK SubscribtionTypeID PK ISPID PK SectorID Count EstimatedInternetUsers	Number Internet Subscribers	Current – Previous
	Cumulative Number Internet Subscribers	Last None Empty
	Estimated Internet Users	Last None Empty
	Derived Indicators	Parallel , Sequential Growth, Penetration

Internet Bandwidth Analytical Theme

Fact Table	Measure\Calculation	Aggregation
Fact.InternetBandwidth PK Year PK MonthNumber InternationalBandwidth OutgoingBandwidthUtilization IncomingBandwidthUtilization	International Bandwidth	Average
	Outgoing Bandwidth Utilization	Average
	Incoming Bandwidth Utilization	Average
	Derived Indicators	Parallel , Sequential Growth, Penetration



Year	ADSL Subscribers Estimated Internet Users	Dial up Subscribers Estimated Internet Users	ISDN Subscribers Estimated Internet Users	Leased lines Subscribers Estimated Internet Users	Mobile Internet Estimated Internet Users	USB Modem Estimated Internet Users	Grand Total Estimated Internet Users
2002		3,497,887.50	44,440.00	2,000,000.00			5,542,327.50
2003	17,843.20	5,318,810.00	58,084.00	2,000,000.00			7,394,737.20
2004	158,940.48	5,900,517.50	72,788.00	2,000,000.00			8,132,245.98
2005	495,088.96	6,026,180.00	89,660.00	2,000,000.00			8,610,928.96
2006	1,134,452.16	5,677,810.00	103,024.00	2,000,000.00			8,915,286.16
2007	2,325,328.00	5,436,745.00	106,104.00	2,000,000.00			9,868,177.00
2008	3,915,510.72	5,436,745.00	107,644.00	2,000,000.00			11,459,899.72
2009	5,589,540.16	4,274,135.00	86,728.00	2,965,840.00		218,707.00	13,134,950.16
2010	7,629,056.00	3,029,537.50	74,700.00	2,965,840.00	7,848,192.00	1,472,616.00	23,019,941.50
2011	7,776,425.60	3,029,537.50	68,188.00	2,965,840.00	8,089,584.00	1,584,801.00	23,514,376.10
Grand Total	7,776,425.60	3,029,537.50	68,188.00	2,965,840.00	8,089,584.00	1,584,801.00	23,514,376.10

Figure 14 – A Sample output of browsing Interment cube to show the estimated internet users' measure. The measure is shown across the time dimension and categorized by the subscription type dimension.

8. Conclusion and Future work

This paper has proposed a case study of designing and implementing a Business Intelligence (BI) solution to an important government sector, the Ministry of Communication and Information Technology (MCIT). In this work, we illustrated the business domains and their need to a decision support system. After that, we described the various layers of the BI solution proposed to the MCIT, including the data integration layer (ETL), the data warehousing layer, and the OLAP cubes layer. The paper covered three domains: Fixed Telephone lines, Mobile Lines, and Internet Usage, where each has a set of analytical themes and performance indicators. We showed a sample output of the multi-dimensional analysis resulted from browsing the OLAP cubes that correspond to these domains.

In the future work, we shall tackle the final layer, which provides visualization capabilities to the users. A dashboard should be developed to provide analytical features, such as OLAP reporting, Trend Analysis, and Intelligent maps.

References

[1] John C. Hancock and Roger Toren. Practical Business Intelligence with SQL Server 2005, Addison Wesley Professional. 2006

[2] W. H. Inmon. Building the Data warehouse. John Wiley & Sons, Third edition, USA, 2002.

[3] Gorry G.A., Morton M.S.S, "A Framework for Management Information Systems", Sloan Management Review 13, 1971, pp. 55-70.

[4] Keen, P., Scott Morton, M,S., Decision Support Systems: An Organizational Perspective. Addison-Wesley, Inc., Reading, MA, 1978.

[5] Sprague R.H. Jr, Watson H.J., Decision Support Systems - Putting Theory into Practice 2nd Edition, Prentice Hall, Eaglewood Cliffs, NJ, 1980.

[6] Turban, E., Decision Support and Expert Systems: Management Support Systems, 4th Edition, Macmillan Publishing Company, NY, 1995 .

[7] Tarek Gebba, and Mohamed R. Zakaria, "E-Government in Egypt: An Analysis of Practices and Challenges", International Journal of Technology and Management ISSN 1927-9000, vol. 1 no. 1, 2012, pp. 11-25.

[8] ICT Indicators in Brief, Ministry of Communication and Information, Cairo, Egypt, Feb. 2013, http://www.egyptictindicators.gov.eg/en/Publications/PublicationsDoc/English%20Flyer%20new%20Feb%202013.pdf

[9] Information and Telecommunications Technology Indicators Bulletin, Quarterly Issue, March 2012, pp. 1-4.

[10] Brahima Sanou, "The World in 2013:ICT Facts, and Figures", ICT Data and Statistics Division-International Telecommunication Union, Geneve, Switzerland, Feb.2013

[11] United Nations E-Government Survey, United Nations, NY, USA, 2010, Chapter 4:World E-Government Rankings, pp. 59-82.

[12] A. Hendawi. A Conceptual Model for Extraction, Transformation and Loading Processes in Data Warehousing. MSc Thesis, Faculty of Computers and Information, Cairo University, Egypt, 2006.

[13] R. Kimball. The Data Warehouse Toolkit. John Wiley & Sons 2002.