# CREATING STRATEGIC BUSINESS VALUE FROM BIG DATA ANALYSIS - APPLICATION TELECOM NETWORK DATA AND PLANNING DOCUMENTS

Otmane Yazidi Alaoui *[1], Salaheddine Hamdoune [2], Hassan Zili [2] ,Hakim Boulassal [2], Mariem Wahbi[2], Omar El Kharki [2]

[1]Equipe de la Geomatique et de l'Aménagement du Territoire, Faculté des Sciences et Techniques de Tanger, Université de Abdelmalek Saadi , Ancienne Route de l'Aéroport, Km 10, Ziaten. BP : 416, Tanger, Marocco. - yazidiotmane@gmail.com
[2] Laboratoire d'Informatique, Systèmes et Télécommunications, Faculté des  Sciences et Techniques de Tanger, Ancienne Route de l'Aéroport, Km 10, Ziaten. BP : 416, Tanger,,-Morocco.
hakim.boulaassal@gmail.com, wahbimr@yahoo.fr, elkharki@gmail.com

**KEY WORDS:**  Business Intelligence, Data warehouse, Data Mart, OLAP, Open Source. SIG, SOLAP.

**ABSTRACT:**

Mobile networks carrier gather and accumulate  in their database system a considerable volume of data, that carries  geographic information which  is crucial  for the growth of the company. This work aimed develop a prototype called Spatial On -Line Analytic Processing (SOLAP) to carry out multidimensional analysis and to anticipate the extension of the area of radio antennas.
To this end, the researcher started by creating a Data warehouse that allows storing  Big Data received  from the Radio antennas. Then, doing the OLAP(online analytic processing) in order to perform multidimensional Analysis which  used through GIS to represent the Data in different scales in  satellite image  as a  topographic background). As a result, this prototype enables the carriers to  receive continuous reports  on different scales( Town, city, country)  and to identify the BTS that  works and performs well or shows the rate of its working( the way  behaves )  its pitfalls. By the end, it gives a clear image on the future working strategy respecting the urban planning, and the digital terrain model (DTM).

## 1. INTRODUCTION

A mobile operator network is composed of three essential parts: first, an access network of cellular nature, composed of relay radio antennas. This network  connected  to the heart of the system by a supply network or backhaul. The spatial-temporal analysis of traffic fluctuations in the access network is a key issue, both in terms of operational maintenance of the network and the decisions to be taken in the long-term prospects of extending its infrastructure.

Since the early 1970s, Information Technologies (IT) offer businesses and researchers several types of applications that favor decision support (Wang et al., 2018). Among these applications, the research distinguished 3 groups of technology:

1- Geographic Information Systems (GIS) designed to collect, store, process, analyze, manage and display on maps, all types of data related to the space infrastructure.
2- Online Analytical Processing (OLAP), enables data online analysis by producing summary reports that provide cross-sectional view of the company's activity;
3- The SOLAP is the result of the coupling of the functions proper to the two previous tools.

Big Data has become a major competitive advantage for many organizations (Skracic et al., 2017),   Mobile telephony entrepreneurs accumulate at the heart of their system and in various operational databases (Excel, Access, Oracle, etc.) a considerable mass either geo-referenced or not geo-referenced numerical and alphanumeric data. These data contain latent potential information that cannot be exploited in transactional systems.

This article aims to  propose  a geo-decisional prototype (SOLAP) to extract and store this big data in order to use the hidden information by representing them as interactive maps that are easily accessible and interpretable by decision-makers. The functional capabilities of this prototype will be tested, validated and illustrated from the data relating to the transmitting-receiving antennas of a mobile operator in Rabat City, Morocco.

*A.Generic Structure of a Mobile operator Network*
A mobile network is subdivided into 3 parts (Figure 1):
• Access  network provides modulation of the traffic,
• A network guarantees the mobility of the user,
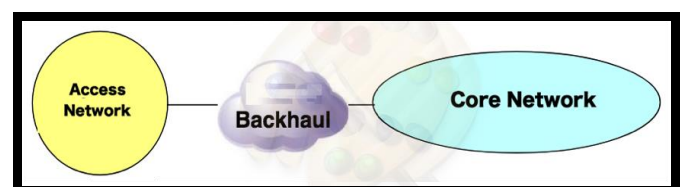• A backhaul (feeds network),



Figure 1. Three Parts of network The access network provides coverage of geographical areas called cells. These latter have a hexagonal shape. Their size can vary from thirty kilometres to a few meters depending on the intensity of the traffic, and the type of antenna used. In addition to this, the size of a cell is not fixed. In fact, it contracts or expands, depending on the weather, the presence of obstacles and the intensity of traffic. Each cell contains hardware and software needed to communicate with mobile stations (Zoha et al., 2018)

These parts  link  the access network with the core network. The links of the supply network may be optical fiber, copper or supported by radio-relay systems.

*B. Decision Support Systems*

Data Warehouse information represents intangible capital. the good management of this later is a key factor for the success of any mobile network carrier. Decision support systems or BI (for Business Intelligence) aim to support the realization of the activities of a carrier. They are built from business requirements and company-defined processes to store, process and communicate information. Information Systems (IS) companies accumulate over time a large amount of data stored in internal memory of the company or from its external (partners, Web, etc.). The problem of this task is to make effective use of these data in order to enable decision-makers to optimize and anticipate their choices. Thus, the need for an efficient use of data in decision-making perspective that has led to the development of new systems -called Decision Support Systems (DSS)- facilitating the storage and synthetic processing of large volumes of data.

These DSS make it possible to select operational information that is relevant to the company and in particular to decision-makers. They combine information and tools effectively to support decision-making.DSS bring together, standardize and coordinate databases, analysis models and visualization techniques.Decision support systems are used in all areas where decision-making is necessary, namely, the areas of trade (marketing, sales), logistics, health (medical decision support), science (eg bioinformatics), telecommunications, transport (motorway traffic) ...

In order to offer a transversal vision of the company's activity, the decision support systems collect and store data from the databases of the various business lines and from external sources (websites, emails ...). The architecture of the decision support systems has 4 levels: two levels of storage and two operational levels (ETL tools and OLAP tools) (Figure 2)
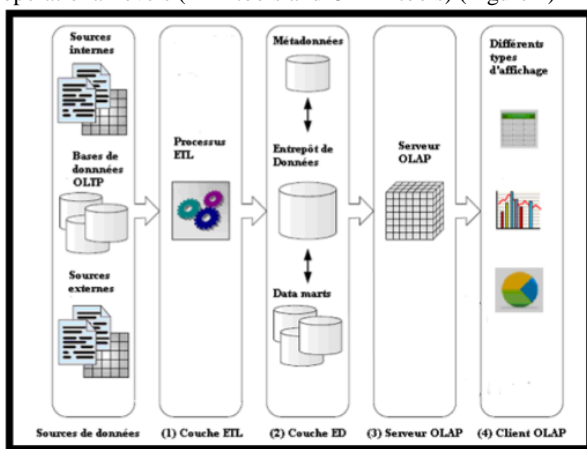


Figure 2: typical architecture of an OLAP system.
Adapted from.

(1) Extract, Transform, Load (ETL) tools are used to power and refresh data in warehouses and data mart from sources.
(2) The data warehouse contains all the decisional information. It is the central storage place of an extract of the sources. Its organization must facilitate the integration of data and the preservation of their evolution. The data mart is an extract from the data warehouse. The organization of data mart follows a specific model which facilitates decision analysis.
(3) OnLine Analytical Processing (OLAP) tools allow decision makers to query and analyse data in a suitable form for them.
(4) The OLAP Client is used to manage the different types of display and reporting.

*C. The SOLAP or OLAP Spatial Concept*

An efficient solution for spatial-multidimensional spatial data analysis, Spatial Data Warehouses (SDW) allow the integration, multidimensional organization, storage of very large volumes of spatial and non-spatial data from multiple data sources to support the decision-making process within a carriers.

SOLAP systems are a category of software tools that enable interactive exploration based on multi-level spatial-multidimensional approaches to SDW. These systems enrich OLAP system analysis and the representation of capabilities with new concepts / operators (SOLAP concepts / operators) which enhance OLAP analysis with spatial analysis. In addition to spatial-multidimensional navigation via SOLAP-specific operators (eg spatial drill-down), these systems also permit the visualization of decision-making results in the form of interactive maps, which show a better understanding of the phenomenon analyzed.

SDW and SOLAP systems are based on the spatial-multidimensional model that modifies the classical multidimensional model by adding other concepts, such as measurement and spatial dimension, which allow the integration of spatial information in multidimensional analysis(Aissi et al., 2015).

*D. Spatial dimensions*

The spatial dimension makes it possible to represent, in the axis of analysis, the spatial information related to the geographical and non-geographical locations of the facts formulated in another way: the spatial dimension consists of using spatial information as an axis of analysis in a decision-making application. Several definitions was proposed:

The spatial dimension concept supports three types of spatial dimensions: descriptive, geometric and mixed in addition to non-spatial dimensions (Mateus et al., 2016) and (Layouni et al., 2016). The "descriptive" dimension uses only the nominal spatial reference (ex: places, names) and no map representation is associated with the members of the dimension. This type of spatial dimension is commonly used in conventional OLAP tools. The "geometric" dimension is hierarchical level includes a set of geometric shapes (polygons, points, lines, etc.). The "mixed" dimension when spatial component levels are associated with purely textual levels. An example of the descriptive, geometric and mixed spatial dimensions is proposed (Figure 3).
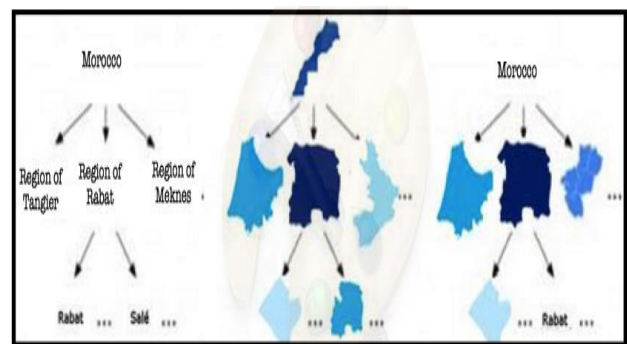


Figure 3: The three modes of representation of spatial dimensions supported by SOLAP represent the administrative subdivision of Morocco (eg country, region, city).

## 2. MATERIAL AND METHODS

### 2.1 Open Source

Source terminology (also known as free software) reflects a certain philosophy. Richard STALLMAN, the founding father of the Free Software Foundation, summarizes what Free Software is : "Freedom, Equality, Fraternity". The license of an "Open-Source" program must meet the following criteria:

The license must not prevent the sale or giving of the software as a component of a distribution of a set containing programs of various origins. The license must not require the sale to be subject to the payment of royalties.

The program must include the source code, and the distribution in source code form as compiled form must be allowed. When a product is not distributed with the corresponding source code, there must be a clear way to download the source code from the Internet at no additional (supplementary) charge. The source code is the most appropriate form for a programmer to modify the program. It is not allowed to propose a source code which is intentionally incomprehensible. It is not allowed to propose intermediate forms, like what a preprocessor or an automatic translator generates.

### 2.2 Rationale for Choosing Open Source

The researcher based development strategy on the integration of Open Source solutions, for the following reasons:
- Licenses become free, the risk of software investment, which in general represents 20% of the project budget.
- 100% of the investment is allocated to the services (personalization, training, support) which maximize the chances of the success of the project.
- The free access to the sources makes it possible to freely personalize the application to the business of the company and increases the adequacy of the solutions to the needs of the users.

The Research of free software publishers and the collaborative development model allow a fast and reliable evolution of Open Source software. Open Source solutions based on interoperable object components. Customers are masters of the evolution of their information system since there is no constraint with publishers or obligation to migrate to new versions. The logic of open source is the durability of the software that ensured by the attractiveness of the community concerning the tool and not by economic laws. The logic of free use of software brings transparency and allows users to work efficiently and quickly.

### 2.3 Structure of the Proposed Data Warehouse

The data warehouse model designed to use interactive cartographic representations, so that the geographic information that is usually highlighted by GIS can complement by geometric information and measurements that will be shown on the cards (Figure4). To do this, the proposed model supports spatial data through the "Urban Planning" dimension, which is one of the most important factors in the subsequent analysis (Batty, 2013).

From a conceptual point of view, the spatial data warehouse was created on the principle of multidimensional modelling, which is the origin of notions of facts and dimensions. In this perspective, the analyzed subject considered as a point in a

space with several dimensions. This vision corresponds to a structuring data along several axes of analysis (or dimensions) which can represent various notions ,such as time, geographical location, urban planning, etc. (Bimonte et al., 2016) and (Yeo and Lee, 2018).

Data storage based on the relational mode. This warehouse structured by following the flake diagram derived from the star schema in this case, the dimensions are divided into several tables according to their hierarchies. The flake model is recommended when dimension tables contain a very large volume of data (Putri and Sitanggang, 2016).
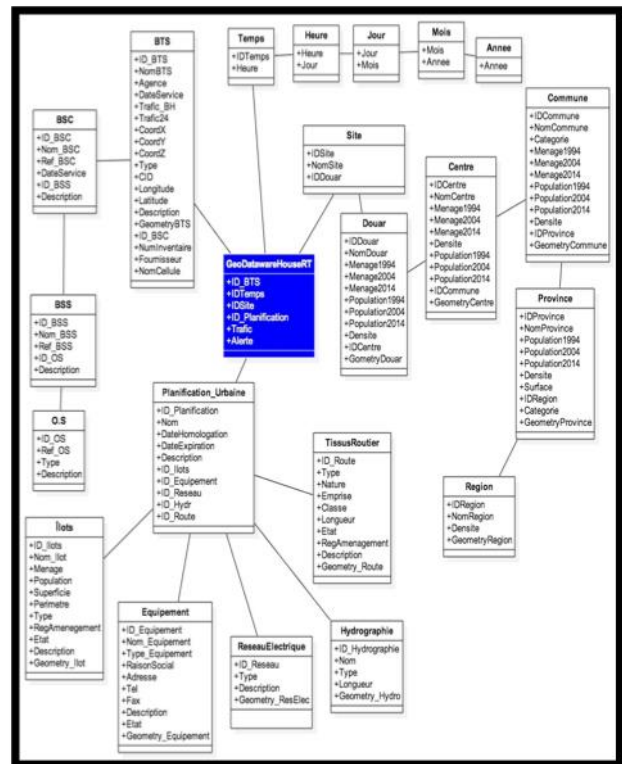


Figure 4 The model of our Data warehouse (Figure 4) is represented by the fact table (blue table) which contains the necessary information (Measure) in the description of a mobile network such as traffic.

The model has 4 Data Marts each one represents a specific subject:

Time Data mart

This data mart (Figure 4) contains time information. It represents the division of time following the hierarchy:
- Year dimension, Month Dimension, Day Dimension, Hour Dimension.

### 2.4 Technical Data Mart

This Data Mart (Figure 4) is hierarchical according to the following dimensions:
i. Dimension BTS: (Base Transceiver System): it represents the heart of spatial Data warehouse, since it constitutes the essential element in the decision-making. This dimension contains information about a BTS, namely: Name, Agency, Traffic max, X, Y, Z coordinates with a BTS geometry field of point nature that represents the spatial part of the latter.

ii. Cell Dimension: A cell is composed of a set of BTS(s).
iii. BSC Dimension: Base Station Controller. Its role is to order a number of BTS up to several hundred. It contains the following information: Name and date of commissioning.
iv. BSS Dimension: (Base Station Subsystem). This dimension contains a description of the BSS.
v. OS dimension: (Operating System). It contains all the information about the operating system.

**2.5 Urban Planning Data Mart**

This stage involves literature review to define the problem and identifying activities required to accomplish the goals. Thus, objectives and research question were being identified based on the problem identification

This store (Figure 4) includes the spatial dimension provided by development plans or PDUs (Urban Development Plan) or PDAR (Rural Development Plan) with dates of approval and expiries. They represent the different existing or projected urban entities. For a better readability, this data was hierarchical in 5 dimensions:

**"Ilot dimension":** contains the islands of populations, the number of households, a description of the urban regulations in force of this block, and its spatial part in the form of a polygon.

**"Equipment dimension":** it contains administrations, the agencies and the different existing or planned equipment;

**"Grid Dimension"**: its spatial part Poly Line represents the different entities of the HV (High Voltage), MV (Medium Voltage), LV (Low Voltage) electrical network.

**"Hydrography dimension"**: represents the watersheds and the routes of the various rivers;

**"Road Fabrics Dimension":** represents the part of the road networks and their different types (national road, secondary road, unclassified, railway, highway) and with a spatial part in the form of poly Line.

# 3. RESULTS

In order to illustrate the resilience, speed and flexibility of the performance of warehouse data, the paper provides different types of display that can be achieved.

**3.1 Displaying Data**

The data selected to feed the cube are related to the hourly traffic recorded for each BTS on one hand. On the other hand, to the different alarm signals. The table (Figure 5) shows only the average monthly traffic value for each BTS for the year 2013. The visualization, using Pantaho, of the data stored in our spatial data warehouse "GouvData" (Figures 5 to 7) shows that our Spatial Data Warehouse is fully operational and allows flexible navigation within the data used to Pantaho whose graphical power allows a display of the results under different types of graphs (Figure 5). In addition, Pantaho allows a projection on some satellite or cartographic databases at different scales. The results of the analyses

carried out for each BTS (Figure 8) are illustrating the graphical and spatial power of Pantaho.

However, Pantaho requires deep knowledge of computer science and data cube modeling with the WorkBensh tool. That's why the researcher choses to develop a SOLAP tool in the form of a plug-in. This tool designed to be adapted to the needs of users who are new to computer science. In other words, this tool aims to display results identical to those obtained using Pantaho with no previous knowledge in computer science.
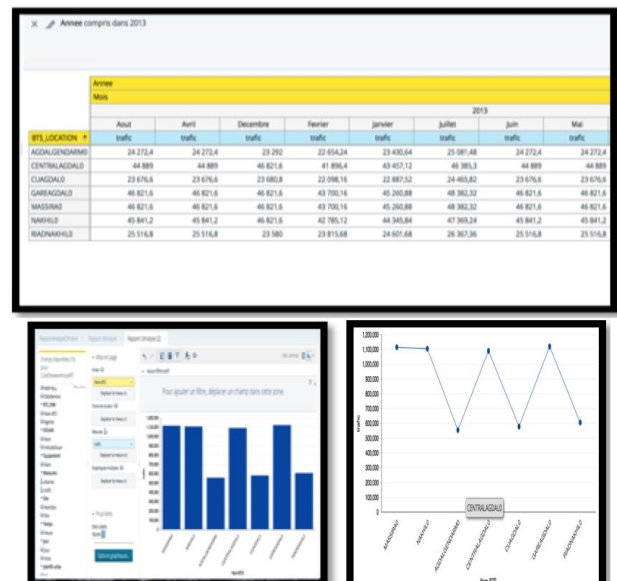Displaying Data on a Satellite Image.



Figure 5.6.7. Data display in graphs and cross-table format

The main interest of SOLAP is to represent data on a geographical background in order to offer a synthetic regional vision. (Figure 8) shows a recent satellite image of Rabat region, the dots in red represent each BTS is proportional to the annual traffic intensity recorded.



Figure 8: Image Satellite Displaying data : areas of the red dots representing each BTS is proportional to the annual traffic intensity recorded for each of the BTS

# 4. DISCUSSION

The work presented in this paper is a prototype of an operational spatial data warehouse supporting a multidimensional analysis system dedicated to the governance of mobile networks and more specifically intended for the

prospective analysis of access networks. Its purpose is to analyze and predict the development of infrastructure at the scale of neighborhoods, cities, a regions and countries. The relevance of predictions for the development of such complex networks does not stem from the multiplicity of parameters, but from their interactions that the multidimensional analysis aims to detect.

## 4.1 Data

Mobile operators accurately record a very large amount of data but in a context that is not conducive to multidimensional analysis. Moreover, these records do not account for the spatial dimension, whereas a large amount of the data is directly impacting the dimensions. This involves the implementation of specific tools to select and standardize the data supplying spatial data warehouses.

## 4.2 Medium and Long-term Management of BTS

Offering seamless service is one of the strategic objectives of mobile operators. The suggested prototype makes it possible to identify the BTS which reach a saturation threshold which makes them unfit to support a constant and regular traffic. As a result, it becomes possible to predict either:
- The increase of the capacities of the defective equipment,
- To envisage a modification of their territorial location,
- Plan the installation of new BTSs taking into consideration the geographical environment and the socio-economic context.

## 4.3 Open source

This work is a good example of what can be done using open source tools. These tools are as powerful as the proprietary tools, whether they are creating data warehouse, feeding it, and querying it. In addition, open source tools are free, while proprietary tools are extremely expensive. From this point of view, this prototype makes it possible to achieve significant economies of scale. It may be of interest to newly established operators who develop their networks in developing regions that are still under-equipped.

## 5. CONCLUSION

This paper is a guideline for creating Strategic Business Value from Big Data Analysis. One of the challenges that face mobile operators is to maintain a constant radio traffic in space and time to satisfy millions of users. Each network is composed of a very complex set of structures. Among these, the access network, made up in particular of transceiver antennas (BTS, NoteB, eNoteB), must make it possible to distribute, as efficiently as possible, a unique radio spectrum between these millions of users. It is therefore essential for the operator to set up expansion forecasts for this access network in order to maintain the operational capacity of the existing site and to increase its territorial existence.

The objective of this paper is to design and implement a geo-decisional prototype respecting the statistical analysis of the fluctuations of the radio traffic in space and time, also to characterize, according to these statistics the plan of the territory, the new sites of implantation of this network.

The data relating to the mobile access network has particularity of being geo-referenced and having spatial characteristics. As a result, SOLAP-based geo-decision analysis systems are therefore particularly well suited to support extension forecasts for this network.

To this end, the researcher has designed and developed a Geo-Decisional Prototype that includes a Spatial Data Warehouse and a SOLAP type "Integrated GIS". This SOLAP combines the capabilities of the GIS tool "QGIS" with those of the "Cube" OLAP developed on a "Python" platform.

## REFERENCES

Y. Wang, L. Kung, et T. A. Byrd, « Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations », Technol. Forecast. Soc. Change, vol. 126, p. 3‑13, janv. 2018.

K. Skracic et I. Bodrusic, A Big Data Solution for Troubleshooting Mobile Network Performance Problems. New York: Ieee, 2017.

A. Zoha, A. Saeed, H. Farooq, A. Rizwan, A. Imran, et M. A. Imran, « Leveraging Intelligence from Network CDR Data for Interference Aware Energy Consumption Minimization », Ieee Trans. Mob. Comput., vol.17, no7, p.1568‑1581, juill. 2018.

S. Aissi, M. S. Gouider, T. Sboui, et L. Ben Said, « OLAP and GIS personalisation : Comparitive study and perspectives of SOLAP personalisation », J. Decis. Syst., vol.25, no1, p. 42‑55, 2016.

R. C. Mateus, T. L. Lopes Siqueira, V. C. Times, R. R. Ciferri, et C. D. de Aguiar Ciferri, « Spatial data warehouses and spatial OLAP come towards the cloud: design and performance », Distrib. Parallel Databases, vol.34, no3, p. 425‑461, sept. 2016.

O. Layouni, F. Alahmari, et J. Akaichi, « Recommending Multidimensional Spatial OLAP Queries », in Intelligent Interactive Multimedia Systems and Services 2016, vol.55, G. DePietro, L. Gallo, R. J. Howlett, et L. C. Jain, Éd. Cham: Springer Int Publishing Ag, 2016, p. 405‑415.

M. Batty, « Big data, smart cities and city planning », Dialogues Hum. Geogr., vol.3, no3, p. 274‑279, nov. 2013.

S. Bimonte, A. Hassan, et P. Beaune, « From Design to Visualization of Spatial OLAP Applications: A First Prototyping Methodology », vol.9975, Éd. Cham: Springer Int Publishing Ag, 2016, p. 113‑123.

I.Yeo et E. Lee, « Quantitative study on environment and energy information for land use planning scenarios in eco-city planning stage », Appl. Energy, vol.230, p. 889‑911, nov. 2018.

A. I. Putri et I. S. Sitanggang, "Data Cubes Integration in Spatial OLAP for Agricultural Commodities", in 3rd International Seminar on Sciences Sciences on Precision and Sustainable Agriculture (iss-2016), vol.58, Bristol: Iop Publishing Ltd, 2017, p. UNSP 012034.

*Revised August 2019*