A Data Warehouse Approach for Business Intelligence

Georgia Garani
Dept. of Exact Sciences
University of Thessaly
Larissa, Greece
garani@teilar.gr

Andrey V. Chernov
Dept. of Computers and Automated Control Systems
Rostov State Transport University
Rostov-on-Don, Russia
avcher@rgups.ru

Ilias K. Savvas
Dept. of Exact Sciences,
University of Thessaly
Larissa, Greece
savvas@teilar.gr

Maria A. Butakova
Dept. of Computers and Automated Control Systems
Rostov State Transport University
Rostov-on-Don, Russia
butakova@rgups.ru

***Abstract*— In a cloud based data warehouse (DW), business users can access and query data from multiple sources and geographically distributed places. Business analysts and decision makers are counting on DWs especially for data analysis and reporting. Temporal and spatial data are two factors that affect seriously decision-making and marketing strategies and many applications require modelling and special treatment of these kinds of data since they cannot be treated efficiently within a conventional multidimensional database. One main application domain of spatiotemporal data warehousing is telecommunication industry, which is rapidly dominated by massive volume of data. In this paper, a DW schema modelling approach is proposed which integrate in a unified manner temporal and spatial data in a general data warehousing framework. Temporal and spatial data integration becomes more important as the volume and sharing of data grows. The aim of this research work is to facilitate the understanding, querying and management of spatiotemporal data for on-line analytical processing (OLAP). The proposed new spatiotemporal DW schema extends OLAP queries for supporting spatial and temporal queries. A case study is developed and implemented for the telecommunication industry.**

Keywords*— Business Intelligence, Data Integration, Spatiotemporal Data, Data Warehouse, Cloud Based Data Warehouse.

I. INTRODUCTION

An increasing number of Cloud Computing (CC) platforms provide facilities for big Data Warehouse (DW) storage and manipulation. Having all the DW functionalities over the Internet simplifies the access on it and storage is no longer an issue since clouds offer almost limitless storage capacity. The Apache Hive Data Warehouse [1] manages large distributed data sets using SQL, while Microsoft with Azure SQL Data Warehouse [2] can fully manage a cloud DW providing a single holistic DW solution. Amazon offers also cloud DW capabilities over Amazon Redshift Cluster using standard SQL [3]. Google isn't out of this with Google BiqQuerry to antagonize the other big vendors [4]. Nowadays, almost all big and smaller cloud providers like IMB [5], Oracle [6], Teradata [7], CoolaData [8] etc. have already include DW services in their cloud environments.

Business intelligence (BI) is a technology-driven process for the collection, integration, analysis, and presentation of business information. It includes a wide variety of tools, applications and methodologies that permit organizations to collect data from internal and external sources for analysis and decision-making.

A component of BI is online analytical processing (OLAP). OLAP creates a multidimensional view of data for the user to do the analysis. The approach for OLAP is classified into three categories, MOLAP, ROLAP and HOLAP. In MOLAP (multi-dimensional online analytical processing) the data used for analysis is stored in specialized multidimensional databases. ROLAP works directly with relational databases. HOLAP approach is a hybrid OLAP approach which combines MOLAP with ROLAP by allowing the designer to decide which portion of the data will be stored in MOLAP and which portion in ROLAP. BI data is typically stored in a DW. The purpose of data warehousing is to construct a huge repository of integrated data, which is optimized for analysis purposes. Nowadays, the big data challenge moves the tradition DWs to cloud based DWs with limitless storage resources and internet based secure access from anywhere to everywhere.

DWs can be considered as multidimensional databases containing large amounts of data where data are organized as a set of dimensions and fact tables. Facts are events taking place in a company and for which decision-making process is required. Dimensions are properties that describe a specific fact. They are the relatively static data of the DW.

The telecommunications industry is rapidly dominated by massive volume of data. It is true that telecom companies need many dozens of relational tables to represent all relevant data concerning calls, billings and customers. For every user event on the telecommunication network, a call detail record (CDR) is generated which contains a broad spectrum of relevant information about each call, including the phone numbers of both the calling and receiving parties, the tower location, the start time of a call, and its duration, as well as more specific information such as the phone number that is charged for the call, the amount billed for each call, the total usage time in the billing period, the total free time remaining in the billing period, and the running total charged during the billing period, the call type, i.e. voice, SMS, the identification

978-1-7281-0676-2/19/$31.00 ©2019 IEEE
DOI 10.1109/WETICE.2019.00022

of the telephone equipment writing the record and many others. All these variety of information that CDRs store can be mined in order to discover patterns related to both calling behavior and service feature usage [9].

Undoubtedly, call detail analysis is one of the hottest areas of exploration in the telecommunication industry. Temporal and spatial data are essential components for this investigation. Telecommunication companies need to know when people like to talk longer or when they call more often, for example it seems that people generally like to talk longer in the evenings but they make more phone calls during the day. Geospatial features of CDR are also important for analyzing social behavior or urban planning of a geographic region [10]. Using this kind of information, companies can be more competitive by promoting products attractive to customers' needs. DWs can be designed, generated and used towards this direction.

In the present approach, a ROLAP system is employed. ROLAP systems deal with large volumes of data, using functionalities inherent in the relational database. Also, they provide better support for frequent updates.

The remainder of the paper is organized as follows. Related research work is discussed in section II. The methodology for the DW design is presented in section III. The telecommunication DW case study is given in section IV, followed by the results of the implementation in section V and a relevant discussion in section VI. Finally, conclusions and future research directions are outlined in the last section.

II. RELATED WORK

A book is dedicated to BI for telecommunications [11] where the impact of business intelligence developments on the telecommunications industry is presented.

Different solutions have been proposed so far for DWs dealing with telecommunication data, most of them based on specific companies and countries. A representative example is [12] where a methodology for DW design and its application within the main Italian telecommunication services provider, the Telecom Italia information system, is discussed based on the conceptual level. Prototype software tools have been used for the application of this methodology. The development of a DW appropriate for the analysis of customer and phone calls data is also discussed.

Another example is given in [13] where data provided by one of the largest telecommunication companies in the United States, Pacific Bell, is used for building a DW for decision-making and business intelligence by decision makers about telephone calls, lines and fraud as well as access to both summary and detail information.

A CDR data mart model is described in [14] as the basis of the star schema used in the case study presented. The OLAP data cube which has been created from the dimensional model in conjunction with OLAP tools have been used for analysis purposes by answering many different queries relevant to call detail and demographic data, such as calls by day of the week and by hour of the day, gender and age analysis and rate plans analysis.

A spatiotemporal analysis of cell-phone traces for identifying the origins of people attending specific events is presented in [10]. The results show that there is a strong relationship among events and attendees, since people who live nearby to an event are preferably attracted by it. For this study, trajectories, i.e. sequences of chronological locations visited by a user, from the individual location measurements have been used. This study assists the decision-making process about events management as well as the reduction of traffic congestion with serious implications in city management.

An approach of near real-time processing of CDRs implemented in the Telcordia mobile virtual network operators (MVNO) hosted prepaid service is presented in usage [9]. The main functionality of the built DW is to store CDRs. It is claimed that the solution is able to handle introduction of new CDRs, versioning of CDRs for different MVNOs, and different MVNO formatting requests with only changes in the configuration files that drive the ETL flows.

A data analytical method used to identify various kinds of characteristics for BI interested in analyzing customer and behavioral data to improve their understanding of customer loyalty and preference in telecommunication industry is introduced in [15] where a DW has been built for this purpose.

A DW schema is also proposed in [16] for analyzing CDR data in order to generate valuable information. In particular, the OLAP processing has been designed to analyze the sales figure of mobile handset in a specific region for a given time period.

In [17] DW schemas which combine telecom data from different sources such as CDR and customer relationship management (CRM) data are presented and then, an integration architecture is proposed for combining data from different operational databases of different telecom operators into a central DW for OLAP processing and business analysis.

A DW is built in [18] for a telecommunication company in Indonesia based on four schemas, which represent telephone usage, internet usage, invoice and payment. Customer segmentation is used for profiling of telecommunication customers according to customer's usage of services, customer invoice and customer payment by applying K-Means clustering algorithm.

All these various approaches do not consider time and space dimensions, although they are both strongly related to telecommunication data. Therefore, the motivation of this paper is to consider an integrated treatment of these two dimensions for the efficient modelling of temporal and spatial data to a telecommunication DW for business analysis.

III. METHODOLOGY FOR DATA WAREHOUSE DESIGN

A. The Starnest DW Logical Schema

The combination of some of the characteristics of the star and snowflake schemas has produced the starnest schema [19] where hierarchy levels are expressed naturally by the clustering of data in nested tables, resulting in describing the aggregation levels for a dimension in a natural way. Dimension tables are nested containing subattributes.

The fact table is linked to dimension tables with one-to-many relationships by foreign key attributes with a reference to the most detailed hierarchical attribute of each dimension. Hierarchical attributes and dimensional attributes are stored in the same dimension table. Hierarchical attributes which are inserted in the table, form a classification hierarchy. These attributes can be added in a nested way where more detailed attributes are embedded inside less detailed attributes; therefore, the definition of the hierarchy can be expressed explicitly in a nested form. The nested approach can represent precisely the hierarchical nature of data; as a result, hierarchies can be modelled adequately. In addition, attributes can easily be associated within their corresponding levels.

The dimension tables may also contain one or more dimensional attributes. The dimensional attributes cannot participate in the dimension hierarchy. However, they are always functionally dependent on one or more hierarchical attributes.

In the starnest schema, data are physically clustered in the dimension nested tables according to the existent hierarchies and so evaluation of queries is very efficient. In addition, the number of join operations between the fact table and the dimension tables is limited significantly.

B. *Spatiotemporal Support*

DWs that include data changing over time are called Temporal DWs (TDWs) similarly to temporal databases. TDWs deal mainly with aggregation relating to time-varying data. Likewise, Spatial DWs (SDWs) store and manipulate spatial data similar to spatial databases for spatial data analysis. The combination of TDWs and SDWs forms Spatiotemporal DWs (STDWs) which are dealing with both spatial and temporal data. The model presented in this approach is a STDW.

Therefore, the starnest schema presented in the preceding section is extended to support spatial and temporal data expressed in a unified manner by a set of subtuples. For the spatial dimension, each subtuple represents a point. The number of subtuples each spatial object contains, represents the points that define this specific object. The number of attributes represents the number of space dimensions, i.e. for a two-dimensional object two attributes are needed, for a three-dimensional object three attributes are needed [20]. For the telecommunication company case study only two-dimensional spatial points are required.

Accordingly, time attributes are inserted in a dimension in a nested way where more detailed attributes are nested inside less detailed attributes; therefore, the definition of the hierarchy can be expressed explicitly in a nested form. The STDW presented in this approach is bitemporal, SBTDW, since it supports both valid time and transaction time in order to keep track of the history of modifications. Valid time represents the time period during which a fact is valid in the modelled reality, while transaction time indicates the time period during which a fact is stored in the DW. By doing so, the system actually models our knowledge of the changing real world and hence, associates values with facts, and specifies when the facts were current in the DW. That is, the system allows storing retroactive and postactive changes in the DW. Both fact and dimension tables are bitemporal since they include bitemporal data.

Objects defined over both space and time (valid-time and transaction-time) dimensions are called spatio-bitemporal objects and are expressed by a set of subtuples. The spatial dimension of the object is embedded in the time dimension. This represents the general case where the spatial object represents a geometry. In the special case where the spatial object is a point in space, the spatial dimensions are at the same nesting level as the time dimensions. The spatio-temporal object is defined below as an extension of the spatial object defined in [20].

Definition 1. *Spatio-bitemporal object, two-dimensional* SBTO(SBTOId, da_1, …, da_k, SBTDim(SDim(X, Y), TDimStart$_{Vt}$, TDimStop$_{Vt}$, TDimStart$_{Tt}$, TDimStop$_{Tt}$)), where SBTOId stands for the spatiotemporal object identifier, X, Y are the spatial two-dimensional coordinates of the spatio-bitemporal object and TDimStart$_{Vt}$, TDimStop$_{Vt}$, TDimStart$_{Tt}$, and TDimStop$_{Tt}$ are the start and stop time points of the valid time and transaction time intervals respectively.

Definition 1 can be extended to support three-dimensional spatio-bitemporal objects by adding the z-coordinate to the SDim attribute.

Dimension tables containing spatial data are called spatial dimensions, those containing (bi)temporal data are called (bi)temporal dimensions and when both spatial and (bi)temporal data are supported in the dimension, this dimension is called spatio-(bi)temporal dimension.

The DW model proposed in this work is a bitemporal STDW model. In the next section, a case study is presented for a telecommunication company DW where all different kinds of dimension tables are depicted and explained.

IV. TELECOMMUNICATION DW CASE STUDY

The proposed spatio-bitemporal starnest schema for a telecommunication company DW is shown in Fig. 1. The fact table stores data about the contracts that customers signed with the company. The DW consists of four dimension tables, i.e. Customer, Employment, Call and Product. Customer and Employment are spatio-bitemporal dimensions, they both contain a spatio-bitemporal object, i.e. address which is represented as a subtuple using a X-coordinate and Y-coordinate value-pair as well as valid and transaction times. Attributes of dimension and fact tables are indicative. Many more attributes can be included in this specific telecommunication DW, containing data appropriate for business analysis and decision-making. They are not included in this work for simplicity reason.

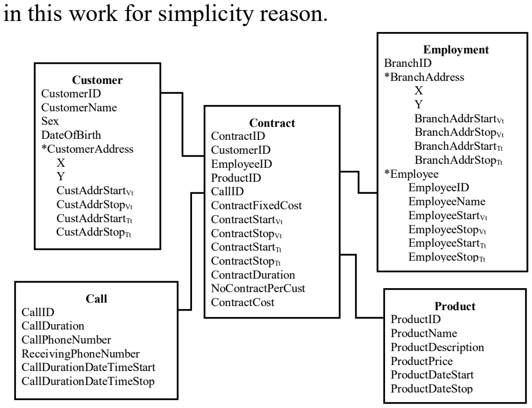

Fig. 1. Telecommunication company schema of SBTDW

A. *SQL Queries*

Ten different queries are demonstrated below based on Fig. 1 and expressed in SQL where the functionality of the SBTDW is depicted.

Query 1. How many customers signed new contracts in the year 2017 and were recorded in the DW the following year?

```
SELECT COUNT(Co.ContractID)
FROM Contract Co
WHERE Co.ContractStartVt BETWEEN ´01/01/2017´ AND ´31/12/2017´
AND Co.ContractStartTt BETWEEN ´01/01/2018´ AND ´31/12/2018´
```

Query 2. How many contracts have exactly the same duration as the corresponding products they have bought?

```
SELECT COUNT(Co.ContractID)
FROM Contract Co, Product Pro
WHERE Co.ContractID = Pro.ProductID
AND Co.ContractStartVt = Pro.ProductDateStart
AND Co.ContractStopVt = Pro.ProductDateStop
```

Query 3. What is the product description of the customer who made the longest duration phone call during Christmas holidays of the year 2018 (24-26/12)?

```
SELECT Pro.ProductDescription
FROM Contract Co, Customer Cu, Call Ca, Product Pro
WHERE Co.CustomerID = Cu.CustomerID
AND Co.ProductID = Pro.ProductID
AND Co.CallID = Ca.CallID
AND Ca.CallDurationDateTimeStart BETWEEN '24/12/2018' AND
    '26/12/2018'
GROUP BY Pro.ProductDescription
HAVING SUM(DATEDIFF(mi, Ca.CallDurationDateTimeStart,
Ca.CallDurationDateTimeStop)) =
    (SELECT MAX(T.totalCallDurationperCust)
    FROM (SELECT Cu.CustomerID, SUM(DATEDIFF(mi,
    Ca.CallDurationDateTimeStart, Ca.CallDurationDateTimeStop))
    totalCallDurationperCust
     FROM Call Ca, Contract Co
    WHERE Ca.CallDurationDateTimeStart BETWEEN '24/12/2018' AND
    '26/12/2018'
    AND Co.CallID = Ca.CallID
    GROUP BY Cu.CustomerID) T);
```

Query 4. Which customers received the most phone calls and how many are these phone calls during New Year's Day 2019?

```
SELECT Co.ContractID, COUNT(Ca.ReceivingPhoneNumber)
FROM Contract Co, Call Ca
WHERE Co.CallID = Ca.CallID
AND Ca.CallDurationDateTimeStart = '01/01/2019'
GROUP BY Co.ContractID
```

Query 5. How much time men and women spent talking (separately)?

```
SELECT Co.Sex, SUM(Ca.CallDuration) TotalCallDuration
FROM Call Ca, Customer Cu, Contract Co
WHERE Co.CallID = Ca.CallID
AND Co.CustomerID = Cu.CustomerID
GROUP BY Co.Sex
```

Query 6. In which branch was employee Tim Taylor when he sold a product to customer John Evans?

```
SELECT Emp.BranchID
FROM Contract Co, Employment Emp, Customer Cu
WHERE Co.CustomerID=Cu.CustomerID
AND Co.EmployeeID=Emp.Employee.EmployeeID
AND Emp.Employee.EmployeeName = ´Tim Taylor´
AND Cu.CustomerName = ´John Evans´
AND ((Co.ContractStartVt <= Emp.Employee.EmployeeStartVt
        AND Co.ContractStopVt >= Emp.Employee.EmployeeStartVt)
OR (Emp.Employee.EmployeeStartVt <= Co.ContractStartVt
        AND Emp.Employee.EmployeeStopVt >= Co.ContractStartVt))
GROUP BY Emp.BranchID
```

Query 7. How many contracts were signed at the branch located at (79.39504333, 94.40322917)?

```
SELECT COUNT(Co.ContractID)
FROM Contract Co, Employment Emp
WHERE Co.EmployeeID=Emp.Employee.EmployeeID
AND Emp.BranchAddress.X = 79.39504333
AND Emp.BranchAddress.Y = 94.40322917
```

Query 8. In which branch has been signed the maximum number of contracts in the first semester of the year 2018?

```
SELECT COUNT(Co.ContractID), Emp.BranchID
FROM  Contract Co, Employment Emp
WHERE Co.EmployeeID = Emp.Employee.EmployeeID
AND Co.ContractStartVt BETWEEN '01/01/2018' AND '30/06/2018'
GROUP BY Emp.BranchID
HAVING COUNT(Co.ContractID) =
    (SELECT MAX(T.CoNumber)
    FROM (SELECT COUNT(Co.ContractID) AS CoNumber,
        Emp.BranchID
    FROM Contract Co, Employment Emp
    WHERE Co.EmployeeID = Emp.Employee.EmployeeID
    AND Co.ContractStartVt BETWEEN '01/01/2018' AND '30/06/2018'
     GROUP BY Emp.BranchID) T);
```

Query 9. How many customers chose a branch for signing a contract situated less than 1 kilometer from their home address?

```
SELECT count(Co.CustomerID)
FROM Customer Cu, Contract Co, Employment Emp
WHERE Co.EmployeeID = Emp.Employee.EmployeeID
AND Cu.CustomerID = Co.CustomerID
SQRT(POWER((Cu.CustomerAddress.X - Emp.BranchAddress.X), 2) +
   POWER((Cu.CustomerAddress.Y - Emp.BranchAddress.Y), 2)) < 1,000
```

Query 10. What was the average duration of phone calls the first month of 2019 for those customers who were in a distance less than 50 kilometers from the place with coordinates (52.92480315, 76.52313444)?

```
SELECT Cu.CustomerID, AVG(Ca.CallDuration)
FROM Contract Co, Call Ca, Customer Cu
WHERE Co.CallID = Ca.CallID
AND Cu.CustomerID = Co.CustomerID
AND Ca.CallDurationDateTimeStart BETWEEN '01/01/2019' AND
    '31/01/2019'
AND SQRT(POWER((Cu.CustomerAddress.X - 52.92480315), 2) +
   POWER((Cu.CustomerAddress.Y - 76.52313444), 2)) < 50,000
GROUP BY Cu.CustomerID
```

V. Implementation

An Intel(R) Core(TM) i5-3210M processor CPU with SSD was used for the implementation running at 2.5 GHz with 16 GB ram memory, under Windows 10 (64bit) OS. The proposed spatio-bitemporal starnest schema for a telecommunication company DW was built in Oracle Data Warehouse builder 11g Release 2 (11.2.0.1.0) and Oracle SQL Developer 18.3.0.277 was used.

The telecommunication company DW consists of one fact table and 4 dimension tables, two of them contain nested tables. Table 1 presents the size and number of rows for each table of the DW. The total size of the DW (i.e. files' space physically consumes on disk) is 1.76 GB.

Table 1. The telecommunication company DW features

Table Name	Size (KB)	Number of Rows
Contract	2048	20,000

Product	64	100
Call	51200	1,000,000
Customer	832	10,000
Employment	64	500

The queries presented in the previous section can be divided into two categories; the first five queries use the three flat tables, Contract, Call and Product, while the last five queries use all five tables of the DW, flat and nested. Each query executed was run 10 times. The queries execution times are demonstrated in tables 2 and 3 and the corresponding charts are given in figures 2 and 3.

Table 2. Queries 1-5 execution times

	Q1	Q2	Q3	Q4	Q5
T1	0,016	0,016	0,078	0,094	0,125
T2	0,015	0,016	0,109	0,063	0,125
T3	0,015	0,015	0,094	0,079	0,125
T4	0,016	0,016	0,078	0,109	0,109
T5	0,015	0,015	0,109	0,078	0,125
T6	0,016	0,016	0,062	0,094	0,125
T7	0,016	0,016	0,078	0,079	0,125
T8	0,015	0,016	0,062	0,078	0,141
T9	0,015	0,015	0,063	0,079	0,125
T10	0,016	0,016	0,078	0,11	0,125
AVG(secs)	0,0155	0,0157	0,0811	0,0863	0,125
AVG(msecs)	15,5	15,7	81,1	86,3	125

Table 3. Queries 6-10 execution times

	Q6	Q7	Q8	Q9	Q10
T1	0,016	0,015	0,032	0,093	0,125
T2	0,015	0,015	0,031	0,078	0,078
T3	0,005	0,016	0,016	0,094	0,078
T4	0,016	0,016	0,031	0,078	0,078
T5	0,016	0,016	0,016	0,062	0,094
T6	0,007	0,016	0,015	0,063	0,078
T7	0,016	0,015	0,016	0,063	0,078
T8	0,016	0,016	0,015	0,078	0,078
T9	0,015	0,015	0,016	0,078	0,078
T10	0,016	0,016	0,015	0,109	0,078
AVG(secs)	0,0138	0,0156	0,0203	0,0796	0,0843
AVG(msecs)	13,8	15,6	20,3	79,6	84,3

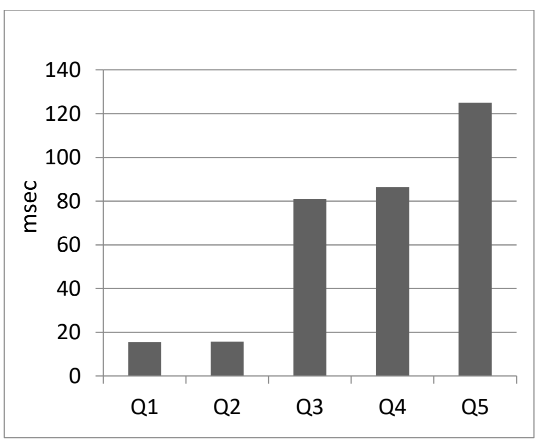

Fig. 2. Queries 1-5 average execution times chart

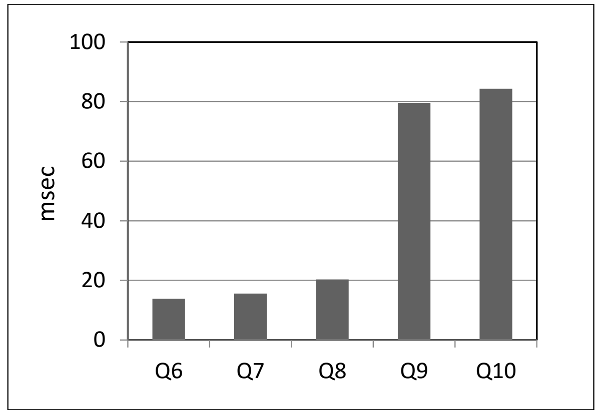

Fig. 3. Queries 6-10 average execution times chart

The two groups of queries are compared in figure 4 where their average execution time comparison chart is presented. It is apparent that the queries that use only flat tables are slower compared to those where all tables are used, nested as well as flat, since the number of rows accessed in nested tables is smaller compared to flat tables.

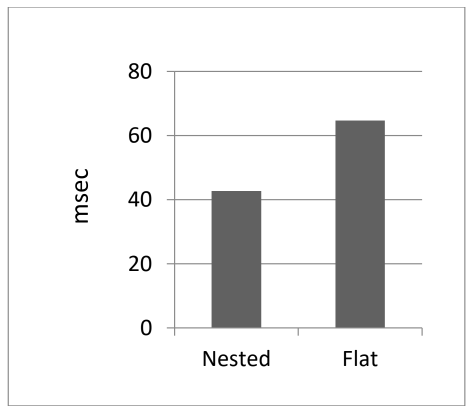

Fig. 4. Queries' average execution time comparison chart

VI. DISCUSSION

The logical DW model proposed in this research work, oriented to satisfy the needs of a telecommunication company, fulfills a number of requirements as it is explained below.

The model is capable of easily representing the temporal and spatial dimensions exclusively. New data types have not been introduced in the model; only generic data types are used.

The spatial dimension is represented as multivalued attributes since they are sets of points and are treated in a uniform manner as any other multivalued attribute. The model is also capable of querying on specific spatial constraints, i.e. comparing distances of spatial objects.

Modelling and querying the DW is possible for any given temporal or spatial condition. Both dimensions of time, valid time and transaction time are supported to the fact as well as to the dimension tables, so the designed DW is fully bitemporal.

Overall, it is apparent that the expressiveness of the model is increased, since the representation is user friendly even for non-expertise business users.

Companies invest a lot of money to BI for better business decisions and DWs consist an essential management tool towards this direction. DWs must convert operational data to a more understandable and user-friendly format for analysis and decision-making. The analysis of the data using DW improves significantly by using also, new operations such as roll-up and drill-down for retrieving data at multiple levels of aggregations. Data mining techniques can be used subsequently, for finding hiding patterns for predicting customers' behaviors and future trends, allowing businesses to make proactive and knowledge driven decisions.

The usefulness of a DW for a telecommunication company is multifaceted, it can facilitate data analysis by providing the main source of information for creating reports, and users can drill down into data details, by analyzing data faster and generating reports more easily; also, they can slice-and-dice in ways they could never do before and submitting multi-dimensional queries against spatial and historical data. The results of the analysis can lead to better decisions for promoting products attractive to customers' needs, for example people that haven't changed country or city in their whole life are not good candidates for a long-distance plan or offering different plans to different age groups at different geographical regions.

VII. CONCLUSION

In this paper, for the first time a bitemporal spatio-temporal nested DW has been developed and implemented. For the case study a telecommunication company was used. For a successful company, good decisions are required based on the efficient management of relevant data. A well-designed DW is facilitating to this direction. The area of DWs for the telecommunication business constitutes a prominent research stream. It can aid to the generation of phone bills, the recording of how funding is spent, and log usage of the telephone system, the managing of the network, call reporting, the resolving of disputes and in general, better management of the business.

A prototype implementation with real data for the verification of the argument is under its way. Future work includes solving related challenges encountered. The solution proposed in this study can be easily generalized to support services in other business domains. In addition, testing our model to several cloud vendors is the final step. The Apache Hive DW has been setup and primary testing has already been conducted with very promising results concerning quality and time complexity.

ACKNOWLEDGMENT

The reported study was funded by RFBR according to the research project 19-01-246-a, 19-07-00329-a, 18-01-00402-a, 18-08-00549-a. The authors would like to thank Christos Siopis, undergraduate student from the Department of Exact Sciences of the University of Thessaly, who helped with the implementation of this research work during his diploma thesis.

REFERENCES

[1] Apache Hive Data Warehouse, online: https://hive.apache.org/, accessed: February 2019.

[2] Microsoft Azure, "SQL Data Warehouse", online: https://azure.microsoft.com/en-us/services/sql-data-warehouse/, accessed February 2019.

[3] Amazon Web Services (AWS), "Data Warehouse Concepts" online: https://aws.amazon.com/data-warehouse/, accessed February 2019

[4] Google Cloud, "Google BigQwerry", online: https://cloud.google.com/bigquery/, accessed: February 2019.

[5] IBM, "Data warehouse", online: https://www.ibm.com/analytics/data-warehouse, accessed: February 2019.

[6] ORACLE, "Oracle Autonomous Data Warehouse", online: https://www.oracle.com/database/data-warehouse.html, accessed: February 2019.

[7] Teradata, "Teradata Integrated Warehouses", online: https://www.teradata.com/Products/Integrated-Data-Warehouses, accessed: February 2019

[8] Cooladata, "Boulding a Data Warehouse", online: https://www.cooladata.com/cost-of-building-a-data-warehouse/, accessed: February 2019.

[9] K. Srikanth, N. V. E. S. Murthy, and J. Anitha, "Wireless Telecommunications Call Records Data Warehouse ETL Flows", International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, pp. 4439-4442.

[10] F. Calabrese, F. C. Pereira, G. Di Lorenzo, L. Liu, and C. Ratti "The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events", Pervasive Computing. Ed. Patrik Floréen, Antonio Krüger, & Mirjana Spasojevic. LNCS Vol. 6030, Springer Berlin Heidelberg, 2010, pp. 22–37.

[11] D. Pareek, "Business Intelligence in Telecommunications", Auerbach Publications, 2007.

[12] S. M. Trisolini, M. Lenzerini, and D. Nardi, "Enterprise Modeling and Data Warehousing in Telecom Italia", 2004.

[13] D. Papaiacovou, L. D. Bramblett, and J. Burgess, "Data Warehouse. A Telecommunications Business Solution", online, url: https://support.sas.com/resources/papers/proceedings/proceedings/sugi22/DATAWARE/PAPER135.PDF, accessed February 2019 .

[14] D. Ćamilović, D. Bečejski-Vujaklija, and N. Gospić, "A Call Detail Records Data Mart: Data Modelling and OLAP Analysis", Computer Science and Information Systems, Vol. 6, Iss. 2, 2009, pp. 88-110.

[15] C. Olaru, "Business Intelligence in Telecommunications Industry", International Journal of Economic Practices and Theories, Vol. 4, No. 1, 2014, pp. 89-100.

[16] G. Maji, and S. Sen, "A Data Warehouse Based Analysis on CDR to Depict Market Share of Different Mobile Brands", Annual IEEE India Conference (INDICON), 2015, pp. 1-6.

[17] S. Mandal, and G. Maji, "Integrating Telecom CDR and Customer Data from Different Operational databases and Data warehouses into a Central Data Warehouse for Business Analysis", International Journal of Engineering and Technical Research, Vol. 5, Iss. 02, 2016, pp. 516-523.

[18] R. Insani, and H. L. Soemitro, "Business Intelligence for Profiling of Telecommunication Customers", Asian Pacific Journal of Contemporary Education and Communication Technology, Vol. 2 (2), 2016, pp. 151-161.

[19] G. Garani, and S. Helmer, "Integrating Star and Snowflake Schemas in Data Warehouses", International Journal of Data Warehousing and Mining, 2012, Vol. 8, No. 4, pp. 22-40.

[20] G. Garani, G. "Representing Spatial Objects in Data Warehouses: A Logical Solution", International Journal of Spatial, Temporal and Multimedia Information Systems (accepted).