RESEARCHERID
THOMSON REUTERS
[ICESTM-2018]
ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

MODELING AND OPTIMIZATION OF DATA IN DATA ANALYTICS

Yallamanda Challa[1], K Purna Prakash[2], P Ramaiah Chowdary[3], V Gopinath[4] & Dr S Krishna Rao[5]
Department of Information Technology, Sir C R Reddy College of Engineering, Eluru

ABSTRACT
Now a day's large amount of data is going to be collected from the sensors and other data received devices. The forms of data are image, video or any raw data. Such that we have a large amount of data to be stored at the terminal end. So that the data should be converting from one from to another form. There the conversion of data should be understood by the machine. The large voluminous data sets by central processor and storage units seem infeasible. Statistical learning tools have to be re-examined in today's high-dimensional data regimes. Such collection of large data we have organized and manage it in proper dimensional to access for the future use. The proposed technique elaborates the generic -tool of projections to form a sequence of estimates in Reproducing Kernel Hilbert Spaces.

Keywords*- data, infeasible, re-examined, optimization, reproducing kernel Hilbert's spaces, modeling of data.*

I. INTRODUCTION

1.1 Big Data:
The information explosion propelled by the arrival of online social media, Internet, and global-scale communications has rendered data-driven statistical learning increasingly important. At any time around the globe, large volumes of data are generated by today's ubiquitous communication, imaging, and mobile devices such as cell phones, surveillance cameras and drones, medical and e-commerce platforms, as well as social networking sites.

The emergence of data rich domains has led to an exponential growth in the size and number of data repositories, offering exciting opportunities to learn from the data using machine learning algorithms. In particular, sequence data is being made available at a rapid rate. In many applications, the learning algorithm may not have direct access to the entire dataset because of a variety of reasons such as massive data size or bandwidth limitation.

1.2 Modeling Data:
One effect of the NoSQL side of big data development has been to delay schema creation. The early definition of the data schema was once a group of data quality practices, and a prerequisite for just getting a project going. "It's not that we don't care about quality. It's that we are not caring about the schema upfront. This doesn't mean designs become "schema-less." Instead, they come to support something akin to "schema-on-read" model, "Now, as far as data modeling is concerned, things are more descriptive. Instead of trying to plan everything out ahead of time, you see use cases developed sort of on the fly," she said. Instead of trying to plan everything out ahead of time, you see use cases developed sort of on the fly. The NoSQL graph database has the ability to capture information on the many interactions that occur in, for example, Web and customer relationship systems, she said. In this way, it can be helpful in creating a descriptive model of wide-ranging application. The trend bears the name of agile methodology and many of its principles.

1.3 Data Modeling:
Data modeling is a process used to define and analyze data requirements needed to support the business processes within the scope of corresponding information systems in organizations. Therefore, the process of data modeling involves professional data modelers working closely with business stakeholders, as well as potential users of the information system. There are three different types of data models produced while progressing from requirements to the actual database to be used for the information system. The data requirements are initially recorded as a conceptual data model which is essentially a set of technology independent specifications about the data and is used

to discuss initial requirements with the business stakeholders. The conceptual model is then translated into a logical data model, which documents structures of the data that can be implemented in databases. Implementation of one conceptual data model may require multiple logical data models. The last step in data modeling is transforming the logical data model to a physical data model that organizes the data into tables, and accounts for access, performance and storage details. Data modeling defines not just data elements, but also their structures and the relationships between them. Data modeling techniques and methodologies are used to model data in a standard, consistent, predictable manner in order to manage it as a resource. The use of data modeling standards is strongly recommended for all projects requiring a standard means of defining and analyzing data within an organization.

1.4 Example:

1. To assist business analysts, programmers, testers, manual writers, IT package selectors, engineers, managers, related organizations and clients to understand and use an agreed semi-formal model the concepts of the organization and how they relate to one another.
2. To manage data as a resource.
3. For the integration of information systems.
4. For designing databases/data warehouses (aka data repositories).

II. RELATED WORK

2.1 Data Models Performance:

Data modeling may be performed during various types of projects and in multiple phases of projects. Data models are progressive; there is no such thing as the final data model for a business or application. Instead a data model should be considered a living document that will change in response to a changing business. The data models should ideally be stored in a repository so that they can be retrieved, expanded, and edited over time.

Strategic data modeling: This is part of the creation of an information systems strategy, which defines an overall vision and architecture for information systems. Information engineering is a methodology that embraces this approach.

Data modeling during systems analysis: In systems analysis logical data models are created as part of the development of new databases.

Data modeling is also used as a technique for detailing business requirements for specific databases. It is sometimes called database modeling because a data model is eventually implemented in a database. Data models provide a framework for data to be used within information systems by providing specific definition and format. If a data model is used consistently across systems then compatibility of data can be achieved. If the same data structures are used to store and access data then different applications can share data seamlessly. The results of this are indicated in the diagram. However, systems and interfaces are often expensive to build, operate, and maintain. They may also constrain the business rather than support it. This may occur when the quality of the data models implemented in systems and interfaces is poor. Business rules, specific to how things are done in a particular place, are often fixed in the structure of a data model. This means that small changes in the way business is conducted lead to large changes in computer systems and interfaces. So, business rules need to be implemented in a flexible way that does not result in complicated dependencies, rather the data model should be flexible enough so that changes in the business can be implemented within the data model in a relatively quick and efficient way.

2.2 Modeling data types

2.2.1 Conceptual schema:

Describes the semantics of a domain (the scope of the model). For example, it may be a model of the interest area of an organization or of an industry. This consists of entity classes, representing kinds of things of significance in the domain, and relationships assertions about associations between pairs of entity classes. A conceptual schema specifies the kinds of facts or propositions that can be expressed using the model. In that sense, it defines the

ISSN 2348 – 8034
Impact Factor- 5.070

allowed expressions in an artificial "language" with a scope that is limited by the scope of the model. Simply described, a conceptual schema is the first step in organizing the data requirements.

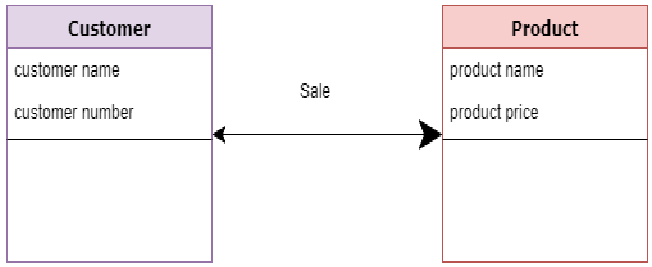

Figure 2.1 Conceptual data model

Characteristics of a conceptual data model:

i) Offers Organization-wide coverage of the business concepts.
ii) This type of Data Models are designed and developed for a business audience.
iii) The conceptual model is developed independently of hardware specifications like data storage capacity, location or software specifications like DBMS vendor and technology. The focus is to represent data as a user will see it in the "real world."

“Conceptual data models known as Domain models create a common vocabulary for all stakeholders by establishing basic concepts and scope.”

2.2.2 Logical schema:

Describes the structure of some domain of information. This consists of descriptions of (for example) tables, columns, object-oriented classes, and XML tags. The logical schema and conceptual schema are sometimes implemented as one and the same.

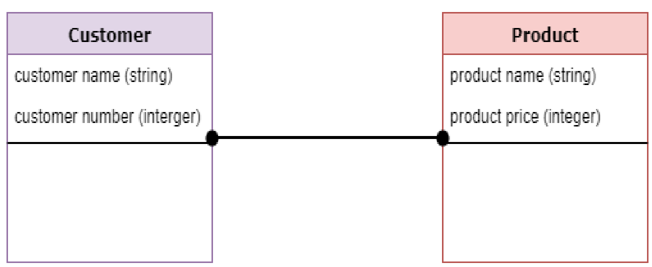

Figure 2.2 Logical data model

Characteristics of a Logical data model:

i) Describes data needs for a single project but could integrate with other logical data models based on the scope of the project.
ii) Designed and developed independently from the DBMS.
iii) Data attributes will have data types with exact precisions and length.
iv) Normalization processes to the model is applied typically till 3NF.

2.2.3 Physical schema:

Describes the physical means used to store data. This is concerned with partitions, CPUs, table spaces, and the like.

Data Optimization is a process that prepares the logical schema from the data view schema. It is the counterpart of data de-optimization. Data optimization is an important aspect in database management in particular and in data warehouse management in general. Data optimizations is most commonly known to be a non-specific technique used by several applications in fetching data from a data sources so that the data could used in data view tools and applications such as those used in statistical reporting.

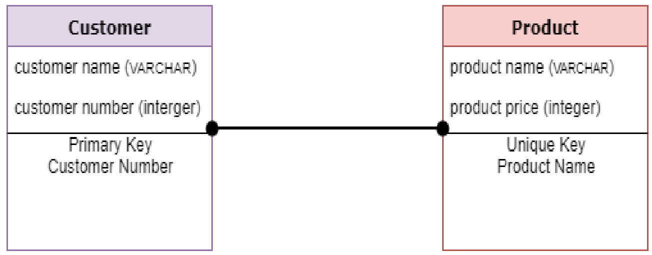

Figure 2.3 Physical data model

Characteristics of a physical data model:

i) The physical data model describes data need for a single project or application though it may be integrated with other physical data models based on project scope.
ii) Data Model contains relationships between tables that which addresses cardinality and null ability of the relationships.
iii) Developed for a specific version of a DBMS, location, data storage or technology to be used in the project.
iv) Columns should have exact data types, lengths assigned and default values.
v) Primary and Foreign keys, views, indexes, access profiles, and authorizations, etc. are defined.

III. OPTIMIZATION OF DATA

A logical schema is also a non-physical dependent method of defining a data model of a specific domain in terms of a particular data management technology without being specific to a particular database management vendor. In more simple terms, the logical schema refers to the semantics describing a particular data manipulation technology and these descriptions could be in terms of tables, columns, XML tags and object oriented classes.

3.1 Data Views

Data views are tools for creating effective reports based on accurate queries. To have a data view, the database management system need to retrieve the desired data and display the expected output. Since the database, especially those databases dealing with high volumes such as those used in data warehouses, need to retrieve large bulks of data, getting a data view may be a slow and complex process. Employing data optimization can reduce the complexity of the process while trying to optimize the needed resources by reducing physical processing needs.

In some database applications, the database management system itself is loaded with features to make querying data views easy by directly executing the query and immediately generating views. Some database applications have its own flexible language for mediating between peer schemas extending from known integration formalisms to more complex architecture.

3.2 Data optimization

Data optimization can be achieved by data mapping, an essential aspect in data integration. This process of data optimization includes data transformation or data mediation between a data source and its destination, and in this case, the data sources could refer to the logical schema and the destination the data view schema. Data mapping as a means of data optimization could translate data between various kinds of data types and presentation formats into a unified format used in different reporting tools.

Some software applications offer a graphical user interface (GUI) based tool used in designing and generating XML based queries and for data views. Since data can come from a variety of sources of from a heterogeneous data source, running queries with this tool can be an effective means of generating a data view. Using graphical data view can free a data consumer from having to focus on the intricate nature of query languages as they tool can provide pictorial and drag and drop mapping approach.

Being free from all the intricacies associated with query languages means that one can focus more on the information design and conceptual synthesis information which could come from many different disparate sources. Since high level tools need to shield end users from the back end intricacies, it needs to manage the data from the back end efficiently.

Having a graphical tool may have its benefits but its downside is that the graphics could add load to the computer memory. So, graphical tools need so much data optimization in order to balance the load toll from the graphical components.

There are several modules available designed for data optimization. These modules can be easily "plugged" to existing software and the integration may be seamless. Having these pluggable data optimization modules can definitely make database related applications focus more on the development of graphical reporting tool for non technical data consumers.

IV. PROPOSED WORKS

In proposed work we are going to modeling and optimizing of large amount of data. We are getting the amount of data from the various devices at the terminal end we are storing all data. While we are go through the data for finding small amount of data from the large amount of data. When we apply the optimizing technique we can find the exact data in a small amount of time.

4.1 Algorithm

Reproducing kernel Hilbert space (RKHS): Reproducing kernel Hilbert spaces are elucidated without assuming prior familiarity with Hilbert spaces. Compared with extant pedagogic material, greater care is placed on motivating the definition of reproducing kernel Hilbert spaces and explaining when and why these spaces are efficacious. The novel viewpoint is that reproducing kernel Hilbert space theory studies extrinsic geometry, associating with each geometric configuration a canonical over determined coordinate system. This coordinate system varies continuously with changing geometric configurations, making it well-suited for studying problems whose solutions also vary continuously with changing geometry. This primer can also serve as an introduction to infinite-dimensional linear algebra because reproducing kernel Hilbert spaces have more properties in common with Euclidean spaces than do more general Hilbert spaces.

V. CONCLUSION

Data modeling is the process of developing data model for the data to be stored in a Database. Data Models ensure consistency in naming conventions, default values, semantics, and security while ensuring quality of the data. Data Model structure helps to define the relational tables, primary and foreign keys and stored procedures. There are three types of conceptual, logical, and physical. The main aim of conceptual model is to establish the entities, their attributes, and their relationships. Logical data model defines the structure of the data elements and set the relationships between them. A Physical Data Model describes the database specific implementation of the data model. The main goal of a designing data model is to make certain that data objects offered by the functional team are represented accurately. The biggest drawback is that even smaller change made in structure requires modification in the entire application.

REFERENCES

[1] *A Primer on Reproducing Kernel Hilbert Spaces Jonathan H. Manton, Pierre-Olivier Amblard, submitted 5 Aug 2014 (v1), last revised 5 Nov 2015.*

[2] *Lee C, Saccomanno F, Hellinga B (2002) Analysis of crash precursors on instrumented freeways. Transp Res Rec. doi: 10.3141/1784-01.*

[3] *Barai S (2003) Data mining application in transportation engineering Transport 18:216–223. doi:10.1080/16483840.2003. 10414100.*

[4] *Kumar S, Toshniwal D (2016) A data mining approach to characterize road accident locations. J Mod Transp 24(1):62–72.*

[5] *Yallamanda Challa,V.Gopinath, K purna prakash,dr s Krishna rao, "a reinforce parallel k means clustering using openMp:OM K means",international of scientific research &engineering research,ISSN NO:2229-5518,volume 7, issue*

[6] *Yallamanda Challa, V.Gopinath, dr s Krishna rao ,k purna prakash "a journey of big data:3V's to 3^2 V's"international journel of research in computer technology,ISSN No:2278-5841,volume 5 ,issue 3.*

[7] *Yallamanda Challa, V.Gopinath, dr s Krishna rao, k purna prakash "cloud sharing for economic benifits"international journel of computer science and information securityISSN No:1947- 5500,volume 14.*

[8] *Gil Press, "A Very Short History Of Big Data", Forbes Tech Magazine, May 9. URL: http://www.forbes.com/sites/gilpress/2013/05/09/avery- short-history-of-big-data/*

[9] *Frank Ohlhorst, Big Data Analytics, Turning Big Data into Big Money, John Wiley & Sons, Inc., 2013, pp 2-pp 171.*

[10] *Yuri Demchenko, Defining architecture components of the Big Data Ecosystem, IEEE Collaboration Technologies and System (CTS), 2014, pp 104-112.*

[11] *Aggarwal G, Feder T, Kenthapadi K, Motwani R, Panigrahy R, Thomas D, Zhu A (2005) Anonymizing tables. In: Eiter T, Libkin L (eds) Database theory— ICDT 2005, vol 3,363., Lecture Notes in Computer Science. Springer, Berlin, p 246–258.*

[12] *Improving Decision Making in the World of Big Data http://www.forbes.com/ sites/christopherfrank/2012/03/25/improving-decision-making-in-the-world-of-big- data/*

[13] *World's data will grow by 50X in next decade, IDC study predicts http://www.computerworld.com/s/article/9217988/World-s-data-will-grow-by-50X-in- next-decade-IDC-study-predicts*

[14] *Mohammad Raji ; John Duggan ; Blaise DeCotes ; Jian Huang ; Bradley T. Vander Zanden "Modeling and Visualizing Student Flow" may 2018.*