International Journal of Innovative Research in Science and Engineering
Vol. No.2, Issue 04, April 2016
www.ijirse.com
ISSN: 2454-9665

EVOLUTION OF DATA HANDLING APPROACHES: A REVIEW

Shreya Jain[1], Purnima Lokwani[2], Deepali Tripathi[3], Dr. N. K. Joshi[4]

[1,2,4]Student, [3]Research, Scholar, MIMT Kota, Raj,(India)

ABSTRACT

Data in today's scenario forms the base of all transactions performed throughout the world. Authenticity accuracy and speed of applications depends on the data being served to them. Data visualization, storage, security, speed are various aspects on which lots of research is being done. This paper focuses on presenting a comparative view of various data management techniques. It's a review of various forms of data management and provides a bird's eye view of their evolution.

Keywords:Data management; DBMS, RDBMS, OODBMS, Big Data.

I. INTRODUCTION

Since the era of computing, the need to manage data started. Many trends in data management were introduced such as File System, RDBMS, ODBMS and many more. Traditional approaches were cumbersome and time consuming so to eliminate their limitations database approach was introduced. To organize large bulk of data on server, concept of Big Data was introduced. This review is about the various trends of managing data that were developed from traditional approaches to latest approaches such as Big Data, XML and others. We also aim to propose the context in which certain trends should be used.

II. EXISTING APPROACHES

Here we have discussed about the following approaches such as File System, the Database, RDBMS, OODBMS, ORDBMS, Big Data which helps us to manage data, retrieve data and establish relationship between them and also to handle large amount of data.

A. File System

File System allows us to control how we can store and retrieve data. They are designed for specific applications such as the ISO 9660 is designed for optical discs. It helps us to store large amount of data which can be shared with another database users and information is stored permanently.

- **Need** – A file system is a collection of related bytes where the file can be free formed, indexed, structured, etc. There are many kinds of file system which can be used on different kinds of storage devices as each storage device uses different kind of media and common storage device. In this mostly two types of Access It may have attributes such as name, identifier, type, location, size, protection, time, date and user identification. methods are used which are Sequential Access and Direct Access. Sequential method is implemented by a file system and data is accessed one record right after the last

and this type of method is reasonable for tape and Direct method is useful for disks in which file is viewed as a numbered sequence of blocks or records and has no restrictions, blocks are read/write in any order.

- **Merits** – It provides heterogeneous operating systems including Unix, Linux and Windows operating systems. It gives multiple client machines which can be accessed by single resource simultaneously. It allows us to control how data is stored and retrieved. It helps in organizing and keep track of the files and directories.
- **Restricted Areas** – It requires the enough memory to store files in drive, if the file format is incorrect it cannot recognize the right program which is to be open. There is restriction of certain symbols in file system.

B. **The Database**

Database is a collection of interrelated files. It helps us in creating, retrieving, updating and deleting data which is organized in the form of tables. Some of the database software such as Fox Pro, MySQL Server and many more.

- **Need –** It provides us data in an organized way, access to them and protects them from any loss. It stores data in convenient way which is easy to retrieve, manipulate and helps to produce information. We need database to maintain records accuracy. It helps in maintenance of proper data and security to data.
- **Merits –** Modern database are more realistic to design architecture as it uses real world entities. It provides less redundancy as possible as it follow rules of normalization. It provides facility to retrieve and manipulate data in an efficient way. It uses the concepts of Acid Properties i.e. Atomicity, Consistency, Isolation and Durability. It provides multi-user environment and allows them to access and manipulate data in parallel. It gives security at some extinct where users are unable to access data of other users and departments.
- **Restricted Areas** – They are complex and extremely large to understand and it is also time consuming while designing. The cost of hardware and software are considerably of high, as database is affected adversely it affects all application programs and there is requirement of training for both programmers and users.

C. RDBMS (Relational Database Management System)

It is database system made up of files including data elements in two dimensional array. It defines the approach to design schema that are in appropriate normal forms. It defines various normal forms in term of functional dependencies and data dependencies.

- **Need** – It has the capability to recombine data elements to form different relations while using data in flexible way. Its main elements are based on Codd's 13 rules for relational systems. Thse relational database is recognized as a collection of tables. Each table consists of a series of row/columns intersection.
- **Merits** – It supports very large databases. It is easy to use, design, implement and manage in efficient way. We use Structure Query Language in this database. It is powerful database management system.
- **Restricted Areas** – In this database we have not enough storage area to handle data just as images, digital and audio/videos. It has restrictions that information must be in tables. It not supports adequate storage of complex objects.

C. **OODBMS**

(Object Oriented Database Management System)

This database is object oriented; an OBJECT corresponds to an entity in E-R Model. The object oriented paradigm is based on inheritance, data encapsulation, object identity and polymorphism. An Object is associated with set of variables, set of messages and set of methods.

- **Need –** Its main objective is to provide secured, controlled, data independent and consistent services so that object-oriented model may be supported. They were created to handle big and complex data that relational database could not. While joining of object-oriented programming with database technology, it provides an integrated application development system.
- **Merits –** It provides us the possibility of reusing the code. It can be formulated with small procedural differences without affecting the entire system. In this we can use enhanced modeling capabilities which is used in applications like computer-aided design (CAD), Computer-aided Software engineering(CASE), multimedia systems and many more.
- **Restricted Areas** – There is lack of experience as it is still relatively limited. It is typically based on specific language. In this when schema is updated or modified; we need to update the instance of that class.

D. **ORDBMS**

(Object Relational Database Management System)

This database model extends the relational database model by providing a richer type system including complex data types and object orientation. It provide convenient migration path for users of relational database who wish to use object oriented features.

- **Need** – Object-relational database is similar to relational database. It provides interface between both object-relational and relational database. This database is created to handle new data such as audio, video and image files as compared to relational database. When ORDBMS was development started, as a result usage of object-oriented programming languages was increased. Its main objective is to bridge the gap between relational databases and object-oriented modeling techniques used in programming languages such as Java, C++, Visual Basic .NET.
- **Merits** – It provides service to organizations to work with their current systems, without making major changes to them. It provides users and programmers to start using object-oriented systems in parallel. It is user-extensible type system in which dynamic binding of methods is used.
- **Restricted Areas** – In ORDBMS we accomplish relationships and encapsulated objects correctly than it will be regarded in disorder which is a drawback for same. This is a complex process.

F. Big Data

This technique is used to organize and handle large amount of data present on the server. and integrate them with per-existing enterprise data to be analyzed.

- **Need** - It helps us to acquire data, organize data and analyze data. If we utilize Big Data, enterprises must involve their IT infrastructures to handle these new high-volume, high-velocity, high-variety of data
- **Merits –** Big Data Analytic Tools may be expensive, but eventually it is cost saving. This data is secured as it can save the average company 1.6% of annual revenues. In this errors and fraud can be detected at the moment and precautions could be taken to limit the damage. With help of this we can achieve cost reductions, quicker and better decisions and can provide new offers to customers. In this we can access new data sources and use into different type of data to generate value from that data.
- **Restricted Areas** - This is mainly used in selected channels which are difficult to apply user data directly to execution. It can be presented through visualizations such as flow diagram as it is not accurate for all users. It requires care in transferring from machine to machine and server to server.

G. XML (Extensible Mark-up Language)

This technique is used to create information formats and share electronically structure data through public Internet in a flexible way. This can also be known as self-describing or self-defining. The XML format information can be shared by any individual or group of individuals or companies.

- **Need** – It is a textual data format which provides strong support via Unicode from different human languages. It is mostly used for the representation of random data structures such as those used in web services. It is a very important part of Web and all electronic information in our future world. It works with many other technologies to display and process information which looks like working on HTML. Its main objective is to emphasize simplicity, generality and usability across the Internet. As it is said, mark-up language in which document can be in human-readable and machine-readable format.
- **Merits -** It is easy to read and understand as it is easily processed by computers. There are no restrictions on set of tags, they can be created whenever needed. As it contains meta data, it can be stored without schemas. It allows comparing document efficiently element by element. In this document storage and processing can be formatted both offline and online. It is platform- independent so we can make changes in technology as accordance.
- **Restricted Areas –** Its data redundancy may affect application efficiency. The hierarchical structure for representing data is limited. The XML documents must be converted into HTML before presenting to user. This database is still on experimental as it has not solidified yet.

International Journal of Innovative Research in Science and Engineering
Vol. No.2, Issue 04, April 2016
www.ijirse.com
ISSN: 2454-9665

BASIS	FILE SYSTEM	RDBMS	OODBMS	ORDBMS	XML	BIG DATA
MAIN OBJECTIVE	*Store data in different formats.*	*Data independence and data integrity.*	*Data encapsulation and independence.*	*Achieve benefits of both relational and object model.*	*Mark-up language which have unlimited tags and transfer data in different formats.*	*Organise voluminous amount of data in efficient manner.*
REDUNDANCY	*Yes*	*No redundancy*	*No redundancy*	*No redundancy*	*Yes*	*Statistical data redundancy*
DATA SECURITY	*No data security*	*Support Limited security*	*Support strong security*	*Support strong security*	*Secure data approach*	*High security is there*
RELATIONSHIP BETWEEN DATA	*Data is not related*	*Data is related by references between tables.*	*Data is related*	*Data is related*	*Data is not related*	*Data is not related*
COMPLEXITY	*File system is complex as access and updation of data is not easy.*	*Structure is simple as users perceive data columns, rows/tuples and tables*	*Structure of data is complex as it involves different data types.*	*It is also a complex approach.*	*It provide simplicity , generality and usability across the internet*	*It is a complex technique as it requires an IT platform and Data Warehousing.*
STORAGE CONTENT	*Store data in random or indexed form.*	*Store data in form of tables or entity sets.*	*Store data and objects.*	*Store data and objects.*	*Set of rules for encoding documents in a format which is both human and machine readable*	*Store machine generated and social data*

III. COMPARATIVE STUDY OF DATABASE APPROACHES

We have discussed various approaches of managing data but it is a still big decision to decide what format of data storage should be selected for the current project. In the current section we provide a comparative study of different approaches which assist the developers in selecting the format of storing data.Table 1 Presents this comparative study. The main objective of File System and XML is to transfer and store data in different formats .RDBMS and OODBMS provides data independence but RDBMS also provides data integrity and OODBMS provides data encapsulation. ORDBMS and Big Data provide benefits of both relational and object model and have also organized voluminous amount of data in efficient manner respectively. File system, XML and Big Data results in statistical data redundancy as compared to RDBMS, OODBMS and ORDBMS. OODBMS, ORDBMS, XML and Big Data provide strong and high security as compared to File System and RDBMS. In

ISSN: 2454-9665

File System, XML and Big Data. File System, OODBMS, ORDBMS and Big Data are complex technique as it requires data warehousing and updating is not so easy as compared to RDBMS and XML. In OODBMS and ORDBMS we can store data in objects but in File System we store data in random or indexed form, in RDBMS we store data in form of tables and entity sets as compared to XML and Big Data ,the data stored in them is easily readable to user and machine in which document is in encoded format.

IV. CONCLUSION

In thispaper we have discussed about the approaches recently used in today's world. We come to know how data can be managed and in which manner it can be arranged. These all approaches used for managing data are efficient and effective. The approaches such as File System, Database, Rdbms, Odbms, Oodbms, Ordbms, Big data, Xml which are uniquely identified and every approach has its own importance. If some are complex to use, than some are easy for user to use and understand.

V. FUTURE SCOPE

In today's world data is playing a big role in decision making and managing big organizations in an efficient way. We presented a comparative study and further we aim to propose a generic model for data management which can be customized and adapeted according to user's need.

REFERENCES

[1] Matthew Trunnel Manager, Research Computing Broad Institute

[2] Revista Informatica Economică, nr. 4 (44)/2007 Gheorghe SABĂU, Bucharest, Romania

[3] Bennett, S., Myatt, M., Jolley, D., & Radalowicz, A. (2001). Data Management for Surveys and Trials.

[4] A Practical Primer Using EpiData. The EpiData Documentation Project. Available: www.epidata.dk/downloads/dmepidata.pdf.

[5] Calvert, William S. and J. Meimei Ma. Concepts and case studies in data management . Cary, NC: SASInstitute, c.

[6] Davidson, Fred. Principles of statistical data handling. Thousand Oaks, California, SAGE, ,266pp.

[7] Graham JW, Hofer SM, Piccinin AM. Analysis with missing data in drug prevention research. IN:

[8] Collins LM, Seitz LA (eds). Advances in data analysis for prevention intervention research.NIDA Research Monograph 142. U.S.

[9] Hulley, Stephen B. and Steven R. Cummings. Designing clinical research: an epidemiologic approach. Baltimore, Williams & Wilkins, 1988. Chapter 15: Planning for data management and analysis.