OVERVIEW OF DATABASE EVOLUTION AND ELEVATION

Dr. A. R. Revathi[1], R. Bhuvaneswaran[2], K. Mohamed Arshath[2], U. Hari Prakash[2],

M. Loga Aravinth[2]

[1]*Associate Professor, Department of IT, SRM Valliammai Engineering College, Tamil Nadu, India.*

[2] *Student, Department of IT, SRM Valliammai Engineering College, Tamil Nadu, India.*

ABSTRACT

"INFORMATION IS THE OIL OF THE 21st CENTURY, AND ANALYTICS IS THE COMBUSTION ENGINE. "

-PETER SONDERGAARD

Nowadays, databases and managing databases is essential because as most of the applications deals with an integral part of user related data. For an organization database must be very crucial. As it deals with user current or stored records either structurally or unstructurally in order to retrieve the required records. Thus the need for a database is to manipulate data, store and identify data and finally to organize data in two ways namely schema based and schema less. This paper illustrates the need of database technology and its properties along with a transition path towards the track to origin of database with advantages and disadvantages, development and step forward to view the upcoming database with New Structured Query Language.

Keyword*: Database, Structured and unstructured data, Schema, New Structured Query Language.*

1. INTRODUCTION

"A database is a self-describing collection of integrated records. A record is a representation of some physical or conceptual object. A database is self-describing in that it contains a description of its own structure. This description is called metadata - data about the data. The database is integrated in that it includes the relationships among data items, as well as including the data items themselves" [1]. "Traditional database systems and Relational Data Base Management System (RDBMS) are used to store important data of organizations by providing back-end of a web services" [2]. It handles limited amount of data since the data's are stored in the form of table. "A **Data Base Management System** (DBMS) is a collection of programs enabling users to create and maintain a database."[3].This paper illustrates various types of databases its properties and the evolution of database with its elevation towards current technology aspects and various application that use various databases.

In section 2 properties of database such as Atomicity, Consistency, Isolation and Durability (ACID) are discussed. Section 3 describes various types of data. In section 4 evolution and elevation of databases are figured. In section 5 conclusion with future works have been demonstrated.

2. PROPERTIES OF DATABASE

A database has to maintain the properties such as Atomicity, Consistency, Isolation and Durability (ACID). In atomicity, when a user updates or alters a database, either all or none of the changes to the database becomes

available to all except the one or application performing the action. This action is called a transaction that performs either commits or aborts. In consistency any changes made to values in a particular instance, should be made consistent with changes to other values in that same instance. In isolation, when performing concurrent transactions the transactions should occur at the same time, such as shared multiple access to users via shared objects. In durability, if a system or storage media fails database should have an ability to recover the committed transactions.

3. TYPES OF DATA

A data can be classified into three types such as structured data, semi structured data, and unstructured data.

3.1 STRUCTURED DATA

Structured data concerns of dealing data that can be stored via SQL in a table with rows and columns. Helps to store and manage information in an easier way. An example of structured data is relational data.

3.2 SEMI STRUCTURED DATA

Does not concern about relational data but have an special organizational properties. An example for semi structured data is XML data

3.3 UNSTRUCTURED DATA (UnSQL)

Unstructured data are schema free, not organized in a specific manner that includes multimedia data such as text, document, email, images and many other.

4. EVOLUTION OF DATABASE

ANCIENT TO MODERN

As the computer was invented in 1943 we need a storage medium to store and retrieve data. Thus the necessity to manage and to deal with data becomes high. From ancient times, object and relational systems have gone through several generations and its development is still growing as data becomes more varied. Also, this paper distinguishes various property characteristics among different database. Initially, Electronic memory with least capacity, but with fast access to time. Later on online storage is used to store larger data which lead to slowest access. Finally, Off-line storage where data storage is based on physical storage.

4.1 NAVIGATIONAL DATABASE

Early 90's were computer becomes popular among companies where people need to improve the quality of database storage capacity where database were being introduced collide with direct storage access like disks and drums (tape based system). Removable storage media can be viewed as storage media, which requires low cost per bit. Data access to Tape media storage within cartridge is sequential. First generation database is navigational where data is accessed through pointer from on record to another. Thus, If a user needs adding information required rewriting the underlying access or modification scheme. The disadvantage is that Maximum 1.8 inch drive can store only up to a gigabyte of data which leads to lesser data storage. Relation such as insert, alter or update data between database cannot be established. In Figure 1 various types of database and its functionalities has been discussed.

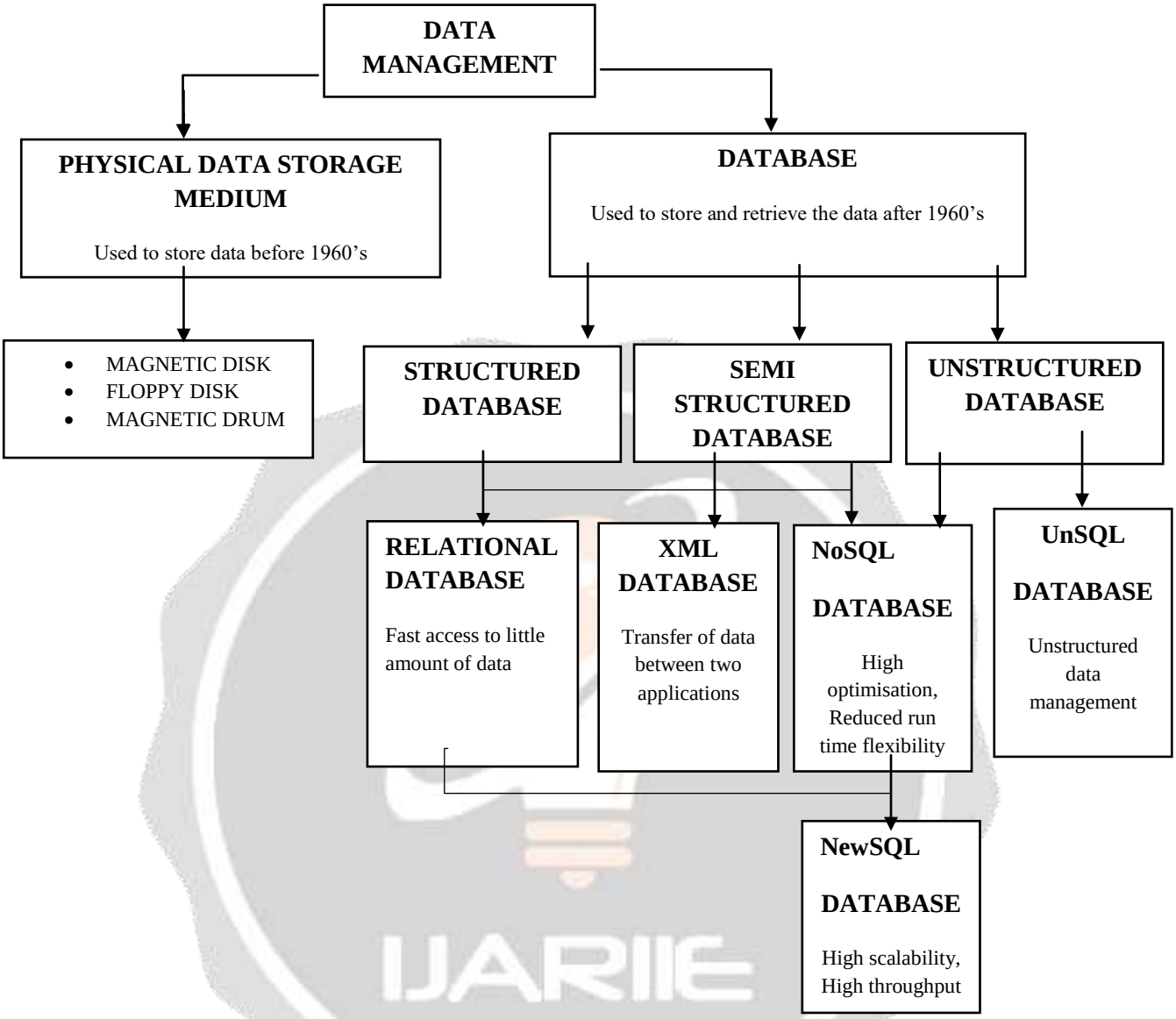

Fig. 1 Illustrates different types of databases and its functionalities

4.2 RELATIONAL DATABASE

Instead of searching for data through links relational database done through content. Stores data in the form of a table where rows specify individual data's and internal structure of operating database and columns specifies a unique collection including the name of the relation and type of column. For storing a countable amount of data traditional database and relational database are used to store important data about an organization by providing a back end web services.

The relational database is very costly and complex in terms of dealing with social related data and for Data Analytics. The user interacts with a logical view and doesn't known much about the actual structure. Uses three schemas such as conceptual, Internal and External schema. Conceptual schema describes all data in database logical, Internal database (Physical schema) specifies how data are being stored in a database, External schema (User view) describes whether the data that is stored can be used by the user or not. RDBMS is built from a set of unique tables (relational database) and basically contains a data about a particular entity where a data can be identified using primary key and association between tables can be done with the help of foreign keys also.

4.2.1 STRUCTURED QUERY LANGUAGE (SQL)

A Computer language or a standard language for organizing manipulating, inserting, and retrieving data that have been stored in a database(Relational Database). It is used for relational Database management system like MySQL, PostgreSQL, My Access and SQL Server use SQL as their base language. The purpose of SQL is to allow users to access data from an RDBMS, to describe the data, to define and manipulate the data, to embed other language using module libraries, to create and drop tables, to create a view, functions as well stored procedures and to set permission for the database.

- **Data Definition Language (DDL):** It consists of syntax that are similar to computer programming Language used to define data structure. DDL used to create and modify database objects such as tables. Commands such as Create, Alter, Drop, Rename are used.
- **Data Manipulation language (DML):** Used to insert, delete, update, retrieve data from database and it is a query language. Commands such as Select, Update, Insert, Delete are used.
- **Data Control Language (DCL):** Which control user access to data in order to safeguard database. Commands such as Grant and Revoke are used.
- **Transaction Control Language (TCL):** Commands are used to manage transactions in the database and also manage the changes made in DML statements into the database. Commands such as Rollback, Commit, Save point is used.

To store and process Big Data efficiently and demand for high performance when reading and writing makes traditional databases to possess a new challenge that too in large scale and high scalability and concurrency applications such as Google search engines with relational database systems becomes a greater loss. Therefore, transition towards NoSQL is required.

Does not use distributed indexes for highly customized and efficient storage and don't have the ability to add new attributes over records. Data storage is not limited, but accessing across multiple servers cannot be done through relational databases. Only handles efficient work loads by a limited amount of users. As when users increases workload for relational databases also increases. If want to improve the size and quality of data being accessed by user relational database requires large and expensive servers.

4.3 NOT ONLY STRUCTURED QUERY LANGUAGE (NoSQL)

To overcome the problems of Relational Data Base Management System (RDBMS) NoSQL is implemented to handle structured, semi structured, unstructured data such as documents, Email, social media and multimedia for efficient storage access for processing Big Data. NoSQL are described as schema free, no support joining operations, high scalability with simple data modelling with query language. It has high availability which fails ACID properties, also provides reliable and cost efficient commodity server for managing huge amount of data. NoSQL products with distributed architecture allows duplication of data across multiple nodes to overcome single point failure. In case of system failure data can be easily recovered from other available nodes. NoSQL can be categorized based on how it deals and stores data with the following categories

4.3.1 KEY VALUED STORE

The structure of Key value store is simple, and the query speed is higher when compared to RDBMS. In the key value store the value corresponds to a key.

4.3.2 COLUMN ORIENTED

Key valued store uses tables as the data model. It has the following characteristics: consistency and partition tolerance (CP), Availability and partition tolerance (AP).

4.3.3 BIG DATA IMPLEMENTATION

Big Data is becoming popular in Cloud service. NoSQL Databases are growing for Big Data applications. The Relational Database is used in most of the existing system. With the growth of data size in applications, Organization tends to handle big data, with NoSQL Database for analysis purpose and for the fast access of the data's.

4.3.4 DOCUMENT STORE

Document store describes schema free organization where records doesn't need uniform structure (Different records have different columns which may have more than one value called arrays) and records may have a nested structure.

4.3.5 GRAPH DATABASE

Based on the graph data model consists of vertices which defines a person, object and edges which defines the relationship between two nodes. The graph data store are schema free, allowing for flexibility of the Document and key valued store. The graph data model requires to evaluate the data mode as the situation may vary. Determines relationship types such as one-one or many-many and also determines the schema of data. Finally, when analyzing complex data and querying multiple types of data, graph database is highly achievable. NoSQL offer highly optimized to retrieve and append data with little functionality beyond record storage. Reduces run time flexibility and comparative to SQL it offers to gain scalability and performance for certain data models. Figure 2 Illustrates NoSQL association with various types of data.

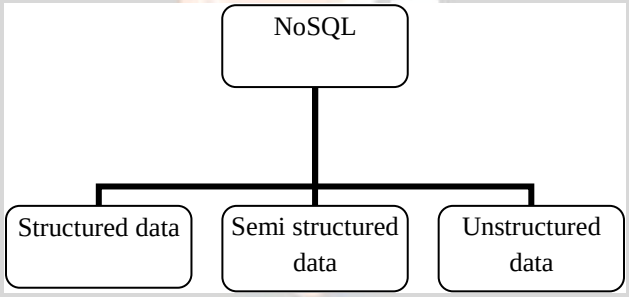

Fig.2 Illustrates NoSQL association with various types of data.

4.4 New Structured Query Language (NewSQL)

NewSQL is a kind of relational database which seeks to provide high scalability for NoSQL applications like transaction oriented processing which supports basic ACID Properties. Combines with NoSQL to provide high throughput and preserve high level language query capabilities such as SQL. Initially OnLine Transaction Processing (OLTP) manages and stores data using RDBMS. Now, NewSQL considers as an alternative to NoSQL and Old SQL for new OLTP applications. If OLTP is a big basket for next generation application, then NewSQL engines will be the best source of future. Table 1 illustrates a comparison of various databases with its advantages, disadvantages and its application.

DATABASE	APPLICATION	ADVANTAGE	DISADAVANTAGE
Navigational Database	Magnetic tapes Floppy disk Drums	**[1]** Data is accessed through a pointer from on record to another. **[2]** If a user needs adding information required rewriting the underlying access or modification scheme	**[1]** Maximum 1.8inch drive can store only up to a gigabyte of data which leads to lesser data storage. **[2]** Relation such as insert, alter or update data between database cannot be established.
Relational Database	Oracle **Database** IBM DB2.	**[1]** Instead of searching for data through links relational database does through content **[2]** For storing a countable amount of data traditional database and relational database are used to store important data about an organization by providing a back end web services.	**[1]**For storing a countable amount of data traditional database and relational database are used to store important data about an organization by providing a back end web services. **[2]** User interacts with a logical view and doesn't known much about the actual structure.
Structured Query Language	Microsoft SQL Server MySQL	**[1]** It is used for relational Database management system like MySQL, PostgreSQL, My Access and SQL Server use SQL as their base language SQL allows users to access data from the RDBMS	**[1]** It cannot store large amount of data with efficiency, **[2]** Real time applications like Google, which need high scalability and concurrency cannot use SOL , because of its inability to process large data thus moving to NoSQL.
NoSQL	Amazon DynamoDB Google Mail	**[1]** To overcome the problems of RDBMS NoSQL is implemented to handle structured, semi structured, unstructured data. **[2]** NoSQL are described as schema free, no support joining operations, high scalability with simple data modelling with query language.	**[1]** NoSQL products with distributed architecture allows duplication of data across multiple nodes to overcome single point failure. **[2]** In case of system failure data can be easily recovered from other available nodes.

Table 1. Illustrates elevation of database with pros and cons and its application

5. CONCLUSION

The evolution towards a database management system is required because there is a need for processing and object management, controlling redundancy and inconsistency, efficient memory management and indexing, concurrency control and transaction management, access control and ease in accessing data, integrity constraints. As future purely depends and manages data in a database and hence requires high security and availability of an organized database. Perhaps the day for future enhancement to databases is not too far where the database becomes more advanced to store digital information which requires a base study of various databases exists so far. This paper discussed properties of database, various types of data, evolution towards database with its advantages, disadvantages and its application.

6. REFERENCES

[1] Kroenke, D. and Auer, D. **"Database Concepts."** 3rd ed. New York, NY: Prentice, 2007.

[2] Yoon, J., Jeong, D., Kang, C. H., & Lee, S. **"Forensic investigation framework for the document store NoSQL DBMS: MongoDB as a case study. Digital Investigation",** 17, 53-65, 2016.

[3] Elmasri & Navathe , **"Databases and Database Users",** Pearson edition, 5th/6th editions**.**

[4] Kristi Berg, Dr.Tom Joseph Seymour, **"History of Databases"**, Research Gate publication, International Journal of Management & Information Systems – First Quarter 2013 Volume 17, Number 1, 2016.

[5] Dharavath Ramesh, **"Inclusion of e-Commerce Workflow with NoSQL DBMS: MongoDB Document Store"**, IEEE International Conference on Computational Intelligence and Computing Research, 2016.

[6] Jing Han, Haihong E, Guan Le,**" Survey on NoSQL Database",** IEEE, 978-1-4577-0208-2/11, 2011.

[7] Aditi Sharma, Parmeet Kaur, **"A Multitenant Data Store Using a Column Based NoSQL Database"**, IEEE, 978-1-7281-3591-5/19, 2019.

[8] Sanket Ghule, Rankrishna Vadali, **"Transformation of SQL system to NoSQL system and performing Data Analytics using SVM"**, IEEE International Conference on Trends in Electronics and Informatics, 2017.