# Paradigm Shift of Database

Prof. A. A. Govande,
Assistant Professor,
Computer Science and Applications,
V. P. Institute of Management Studies and Research, Sangli

***Abstract*— Now a day's most of the organizations are facing drastic increase in processing of huge amount of data. So it leads to the problem of capturing, storing, managing and analyzing huge amount of data, stored in multiple formats, present in different internal and external sources. They also needs to process data in real time. Also now a day's data present in various formats such like emails, text data, Audio, Video, Images, and 3D Images etc. are referred to as unstructured data. So to increases processing capacity, to give away desired results, to accommodate unstructured data kind of Database systems are essential which will manage huge amount of data and gives speedy real time results. From file system to relational, object relational, distributed, cloud systems up to Big Data, database has evolved through these several generations. From last few years lots of changes have seen in way IT is being used and viewed Database. This paper explores how Database has evolved as per the changing requirements of the organizations.**

***Index Terms*—***File System, Database Management System, Cloud Databases, Big Data*

## I INTRODUCTION:-

IT department of any organization is responsible for providing reliable transactions, storage, back up and network facilities at the lowest feasible cost. Now a day's some organizations like Google, Yahoo, Amazon, Facebook etc. are facing drastic increase in processing huge amount of data, for which real time processing database and speedy database systems are essential instead of traditional databases.

Database is having history of more than 45 years. As per the demands and needs of that era, the database and database management systems were evolved. Stand- alone applications have been replaced with web-based applications, dedicated servers with multiple distributed servers and dedicated storage with network storage. Due to the advantages like more efficiency, scalability, lesser cost, reliability Distributed and Cloud Databases are becoming very much popular now days, and these databases states the limitations of file system and relational, object relational kind of databases.

### 1. File System (1968 File-Based: predecessor of database, Data was maintained in a flat file.)

A file is nothing but collection of data. A file system is the methods and data structures that an operating system uses to keep track of files on a disk or partition; that is, the way the files are organized on the disk. The word is also used to refer to a partition or disk that is used to store the files or the type of the file system. But the Flat File brought in a lot of redundant data at every entry.

Earlier, punched cards technology was used to store data – later, files. But the files have no as such advantage, rather have several limitations. In earlier days data was stored in the flat file and there was no structure in the flat file. There was no possibility of retrieving the data efficiently and data integrity has been just a term discussed without any modeling or structure around.

#### 1.1 Advantages:-

1. Multiple access methods
2. Easy to create and store data.

#### 1.2 Disadvantages:-

1. Duplication of data – same data is held by different programs, thus, wastes space and resources.
2. High maintenance costs such as ensuing data consistency and controlling access
3. Weak Security
4. Could not consume large amount of data

## 2. Database Management System (DBMS):-

DBMS is one computerized system which manages the data present in database.

### 2.1 Advantages of Database vs. File System:-

1. **Transaction Support–** Atomic transactions guarantee complete failure or success of an operation. This includes automatic recovery of the database to a transaction consistent point in the event of an abnormal termination of the application (crash, power loss, etc.).
2. **Concurrent Access–** The ability to share data by controlling access to data items, many users (process or threads) can access data concurrently.
3. **Data Normalization–** A well designed database schema can reduce storage requirements on the target storage media by reducing duplicate data.
4. **Expandability, Flexibility, Scalability–** A database system can scale easily to larger data sets.
5. **Standards Enforcement–** One example of this advantage would be to use the DBMS for all data storage requirements for the application. Multiple data structures can be manipulated using the same API functions. The can lead to reduced application development times and reduced maintenance costs in the future.
6. **Fast Query Access–** Databases allow indexing based on any attribute or data-property (i.e. SQL columns). This helps fast retrieval of data, based on the indexed attribute. This is an importance advantage as data-sets begin to grow large as it provides a more predictable query response time.
7. **Interoperability–** Connectivity through industry standard protocols allowing third-party tools to access and analyze data

### 2.1 Database Model:

The Flat File brought in a lot of redundant data at every entry. For instance, if I want to make a single data-set with all products purchased at a grocery store with all information of the customer and product, we will have every single row consisting of all customer and product information. Wherever we have a repeat product or customer, we have repeat data. So to access appropriate data Database Models were evolved A Database model defines the logical design of data. The model describes the relationships between different parts of the data. Relational databases ruled the Information Technology (IT) industry for almost 40 years. But last few years have seen changes in the way IT is being used and viewed database.

Historically, in database design, these models are commonly used. They are,

1. **Hierarchical Model:** In this model, files are related in a parent/child manner, with each child file having at most one parent file
2. **Network Model:** In Network data model, files are related as owners and members, similar to the common network model except that each member file can have more than one owner.
3. **Relational Model:** In this model, data is organized in two-dimensional tables called relations. The tables or relation are related to each other.
4. **Object Oriented Model:** It supports the modeling and creation of the data as objects.
5. **Object Relational Model:** It is essentially a relational model that allows users to integrate object-oriented features into it.

Edgar Frank Codd was a British computer scientist who, while working for IBM, invented the relational model for database management, the theoretical basis for relational databases. He presented 12 rules for the Relational Database. Relational Database was a promising land for all the unstructured database users. Relational Database brought into the relationship between data as well improved the performance of the data retrieval. Database world had immediately seen a major transformation. The Entity Relationship model was also evolved at the same time. Entity–relationship model (ER model) is a data model for describing a database in an abstract way.

Due to development of new applications and major use of social media took the world by the storm. Every organization was feeling pressure to provide the best experience for their users based the data they had with them. While this was all going on at the same time data was growing pretty much every organization and application. Now a day's data came from a variety of sources, such as emails, text documents, videos, photos, audio files, and social media posts, are referred to as unstructured data. And traditional databases can only serve the structured data, which can be entered, stored, queried, and analyzed in a simple and straightforward manner. Stand- alone applications have been replaced with web-based applications, dedicated servers with multiple distributed servers and dedicated storage with network storage. Due to the advantages like more efficiency, scalability, lesser cost, reliability, support structured and unstructured data Distributed ,Cloud and Big Data

Databases are becoming very much popular concepts now a days, and these databases overcomes the limitations of file systems and traditional databases.

## 3. Parallel Processing and Parallel Database:

**3.1 Parallel Processing:** Parallel processing divides a large task into many smaller tasks, and executes the smaller tasks concurrently on several nodes. As a result, the larger task completes more quickly. A node is a separate processor, often on a separate machine. Multiple processors, however, can reside on a single machine.

For example, in a bank with only one teller, all customers must form a single queue to be served. With two tellers, the task can be effectively split so that customers form two queues and are served twice as fast-or they can form a single queue to provide fairness. This is an instance in which parallel processing is an effective solution.

**3.2 Parallel Database:** A parallel database system seeks to improve performance through parallelization of various operations, such as loading data, building indexes and evaluating queries. Parallel databases improve processing and input/output speeds by using multiple CPUs and disks in parallel

In Parallel processing rather memory is shared between all the CPU and present at each and every CPU. In such system all CPU executes amount of given task in parallel.Following are some possible architectures of Parallel Processing System:

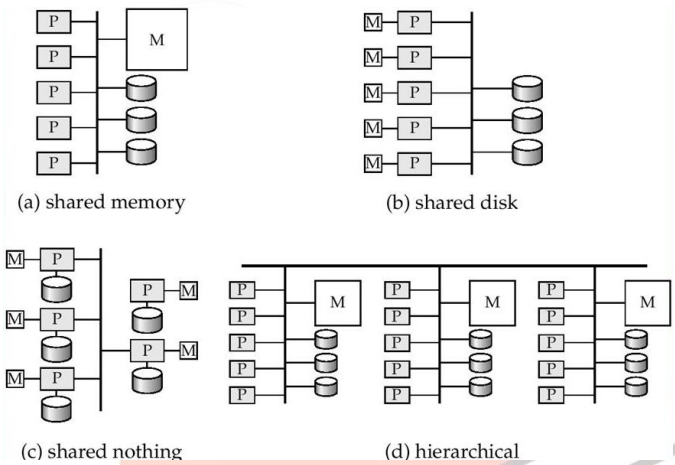


**Fig. Possible Architectures of Parallel Processing System**

**3.3 Features of Parallel Processing:**

1. Each processor in a system can perform tasks concurrently.
2. Tasks may need to be synchronized.
3. Nodes usually share resources, such as data, disks, and other devices Distributed Processing and Distributed Database

## 4. Distributed Processing:

Distributed processing is a phrase used to refer to a variety of computer systems that use more than one computer (or processor) to run an application. This includes parallel processing in which a single computer uses more than one CPU to execute programs. Shared nothing kind of Architecture is appropriate for Distributed Processing.

**4.1 Distributed Database:**

A distributed database is a database in which portions of the database are stored in multiple physical locations and processing is distributed among multiple database nodes

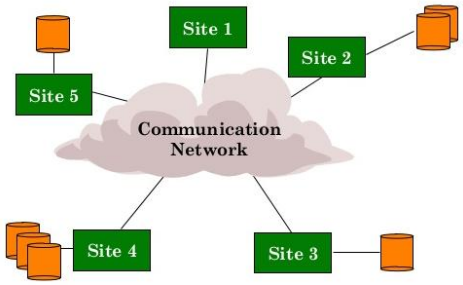


**Fig. Distributed Database**

A centralized distributed database management system (DDBMS) integrates the data logically so it can be managed as if it were all stored in the same location. The DDBMS synchronizes all the data periodically and ensures that updates and deletes performed on the data at one location will be automatically reflected in the data stored elsewhere.

Distributed databases can be homogenous or heterogeneous. In a homogenous distributed database system, all the physical locations have the same underlying hardware and run the same operating systems and database applications. In a heterogeneous distributed database, the hardware, operating systems or database applications may be different at each of the locations.

**4.2 Advantages:**

1. Easy to share resources.
2. More Reliable
3. Scalable
4. Efficient
5. Database can be accessed over different networks.
6. A user doesn't know where the data is located physically. Database presents the data to the user as if it were located locally.

**4.3 Disadvantages:**

1. Network traffic is increased in a distributed database.
2. Network traffic is increased in a distributed database.
3. Managing distributed deadlock is a difficult task.
4. Managing system catalog is a difficult task.

## 5. Cloud Database:

A cloud database is a database that typically runs on a cloud computing platform, access to it is provided as a service. Database services take care of scalability and high availability of the database.

Cloud computing has brought a paradigm shift not in the technology landscape, but also in the database landscape. With more usage of Cloud computing, demand for provisioning of database services has raised.

**5.1 Advantages:**

1. **Data Centralization:** Another key benefit of cloud services is the centralized data. The information for multiple projects and different branch offices are stored in one location that can be accessed from remote places.
2. **Data Recovery:** Cloud computing providers enables automatic data backup on the cloud system. The recovery of data when a hard drive crash is either not possible or may cost a huge amount of dollars or wastage of valuable time.
3. **Sharing Capabilities: All precious documents and files can be emailed, and shared whenever** required.
4. **Cloud Security:** Cloud service vendor chooses only the highest secure data centers for your information. Moreover, for sensitive information in the cloud there are proper auditing, passwords, and encryptions.

**5.2 Disadvantages:**

1. **Net Connection:** For cloud computing, an internet connection is a must to access your precious data.
2. **Security Issues**

3. **Minimal flexibility:** The application and services run on a remote server. Due to this, enterprises using cloud computing have minimal control over the functions of the software as well as hardware.

## 6. Big Data Analytics:

Big data is a term that describes the large volume of data – both structured and unstructured. Big data can be analyzed for insights that lead to better decisions and strategic business moves.

The new benefits that big data analytics brings to the table, however, are speed and efficiency. Whereas a few years ago a business would have gathered information, run analytics and unearthed information that could be used for future decisions, today that business can identify insights for immediate decisions. The ability to work faster – and stay agile – gives organizations a competitive edge they didn't have before. Big data analytics is mainly Non- Relational Database Tool storing huge amount of data and scaling them easily.

### 6.1 Non-relational databases:

The concept of non-relational databases came into picture to handle rapid growth of unstructured data and scale them out easily. This provides flexible schema so there is no such thing called "Referential Integrity" as we see in Relational databases. The data are highly de-normalized and do not require JOINs between objects. This relaxes ACID property of relational databases and supports CAP (Consistency, Availability and Partitioning).

### 6.2 Benefits of Big Data Analytics:

1. **Cost reduction:** Big data technologies such as Hadoop and cloud-based analytics bring significant cost advantages when it comes to storing large amounts of data – plus they can identify more efficient ways of doing business.
2. **Faster decision making:** With the speed of Hadoop and in-memory analytics, combined with the ability to analyze new sources of data, businesses are able to analyze information immediately – and make decisions based on what they've learned.
3. **New products and services:** With the ability to fulfill customer needs and satisfaction through analytics comes the power to give customers what they want. Davenport points out that with big data analytics, more companies are creating new products to meet customers' needs.

## II CONCLUSION:

IT industry are facing drastic increase in processing huge amount of data, for which real time processing database and speedy database systems are essential. Also now a day's data present in various formats such like emails, text data, Audio, Video, Image, 3D Images etc. are referred to as unstructured data. As traditional database could only handle structured data the modern database techniques like Parallel, Distributed, Cloud databases and Big Data Analytics are appear to be the good solutions to fulfill the current IT needs in more flexible way than File and Traditional Database Systems. As per the needs of that era Database has gets evolved, so that now a day's we can handle huge amount of data which may be present in structured, unstructured form. To conclude, paradigm shift in Database clearly indicates that Database processing has moved from Data Analysis to Data Analytics.

## III REFERENCES:

[1] Rajkumar Buyya et al., "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012 ISSN (Online): 1694-0814 www.IJCSI.org 82 Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved. computing as the 5th utility", Future Generation Computer Systems, Vol. 25, Issue 6, June 2009, pp. 599-616.

[2] http://xeround.com last accessed on May 25, 2012.

[3] Mike Hogan, "Cloud Computing & Databases", November 14, 2008.

[4] Arpita Mathur et al., "Cloud Based Distributed Databases: The Future Ahead", International Journal on Computer Science and Engineering (IJCSE) Vol. 3, No. 6, 2011.

[5] Bo Peng, "Implementation Issues of A Cloud Computing Platform", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.

[6] Maggiani, R. "Cloud computing is changing how we communicate", IPCC 2009, 2009, pp. 1-4.

[7] F. Chang et al., "Bigtable: A Distributed Storage System for Structured Data", in 7th Usenix Symp. Operating Systems Design and Implementation (OSDI 06), Usenix Assoc., 2006, pp. 205–218.

[8] https://en.wikipedia.org/wiki/Distributed_database

[9] Cloud Databases: A Paradigm Shift in Databases Indu Arora1 and Dr. Anu Gupta2 1 Department of Computer Science and Application, MCM DAV College for Women, Chandigarh 2 Department of Computer Science and Application, Panjab University, Chandigarh IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012 ISSN (Online): 1694-0814 www.IJCSI.org

[10] http://bigdata-madesimple.com/relational-vs-non-relational-databases-part-1