CGC International Journal of Contemporary Technology & Research
ISSN: 2582-0486 (online)

Moving from databases to Cloud Database: Futuristic Trends

[1]Gurpreet Singh, [2]Shivangi, [3]Sumit Kumar, [4]Sunil Chawla
[1, 3, 4] Assistant Professor, Department of Computer Science, CGC-College of Engineering
[2] Student, Department of Computer Science, CGC-College of Engineering

***Abstract*: This paper provides a review of various generations of databases along with their features, advantages and limitations. We tried to cover major generations of databases and their area of utilization. Through this paper we tried to provide a review of various databases for various researchers in this area and by the end of this paper we provide the future scope in terms of next generation databases.**

***Keywords*—Object-Oriented, Database, relational, database management system, data model, Big Data, Cloud, Hadoop.**

I. INTRODUCTION

The need to store and sort data is much older than we think . The term database was first coined in 1960s but human began to store data long ago, sumerian tablet is a classic example of it ,this tablet is an index of medical prescriptions. Index card are considered to be antecessor of databases. From historic time to file system, from data models to relational database management system and object relational system there is a fascinating change in database technology.In the early 1960s, Integrated data store(IDS) was developed at General Electric Low Voltage Switchgear Department. It was designed by Charles W.Bachman. Not wanting be left out IBM developed a hierarchical model IMS (Information Management System). And later on network model came into existence which is based upon CODASYL approach, IDS forms the basis of CODASYL.Edgar Codd was unhappy with the network and hierarchical model and he came up with the concept of relational data model. There was a rapid enhancement of technology at the front end with the effect of techniques like object oriented languages, however the backend still persist with older record based techniques such as relational database based on the relational model which began to become a cumbersome of matching front end(object oriented) with backend(record based). To overcome lacking of object oriented database like high cost of transformation and less flexibility, designers tried to develop a database management system which have hybrid properties of both relational & object oriented database. Then object relational database model was developed and it allows both relational and object views of data in the same database. Over the past years massive amount of data has been produced and these voluminous data sets are termed as BIG DATA and we need BIG DATA analytics for finding the correlation and get better decisions.

II. LITERATURE SURVEY

A. Kristi Berg et.al December 2012

This paper emphasizes on the origin of databases and how the database technology has evolved and gone through various generations and changed fascinatingly. It shows us a glimpse of databases evolution different decades have witnessed and what are the future scopes of databases.

B. Hibatullah Alzahrani et. al 2016

This paper emphasizes on object -oriented database systems and highlights its importance, its merits and demerits and how it has evolved over the years.

C. Shagufta Praveen, et. al March 2017

This survey paper addresses the evolution of database from the beginning. It also introduces how database is evolved ,helped by different technologies, concepts and theories . The fact is still industries are in hope for something better in context to database technologies.

D. Akmal B. Chaudhri et. al 1993

This paper provides a detailed overview of Object Database Management Systems (ODBMSs). It also explains the limitations of existing database technology to fulfil the needs for "next-generation" applications that necessitate the reinforcement of complex structures.

E. Prateek Bhatia and Gurvinder Singh , Simplified Approach to DBMS-10th Edition

This book addresses on all aspects of databases that it can fulfill the thrust of learner of database. From history and evolution of database to NoSQL technologies this book has covered all the features, concepts and technologies.

F. Kevin Taylor-Sakyi et. al 2016

This paper reflects knowledge and understanding in the tantalizing field of big data. It aims to focus on the significance of understanding big data, reminiscing the transformation from traditional analytics into big data analytics, data storage, and the future participation.

G. Kevin Driscoll et al. August 2012

This paper uncovers the history of database technology over the periods that represent portentous changes in the accessibility and infrastructure of information processing systems. And suggests that a redemption of small scale data processing might lead to sharper popular critique in the future.

H. Antonio Badia et. al March 2004: This paper suggests that latest applications require databases to capture and exert more domain connotation than traditional applications and they also investigated ways in which revisit Entity-Relationship models could be expanded to pursue the process.

I. Meenu Dave et. al August 2012

This research paper tells about NoSQL, its background, preliminaries like ACID, BASE and CAP theorem. Also it uncovers the various types of NoSQL data stores with their examples, features, and advantages and limitations of NoSQL.

J. James P. Fry e. al, March 1976

This paper demarcates the grail of data-base management systems, scrutiny important concepts and unearth the development of data-base systems strategy and it presents some trends and issues.

K. Rahul Beakta et. al 2015

This paper gives an overview of big data, its applications in data mining, its challenges and future scope. It also discusses Hadoop framework and current researches in data mining.

III. DISCUSSION

A. Integrated Data Store

In the early 1950s , computer systems developed and focused on programming languages and algorithms[7] and they just started to become commercially available and when people started to use them for the real-world purposes, the need to store the data somewhere increases and programmers developed solutions for this. And in 1964 the first commercial database management system came into existence;Integrated Data Store designed by Charles W.Bachman. It integrates random access storage technology with high-level procedural languages[10].

B. Information Management System

The most popular data model of 1960s is IMS(Information Management System) developed by IBM and North American Rockwell Company jointly, it is the first commercial hierarchical model and originally it was written for APOLLO PROGRAM and used by NASA to manage drawings for the lunar lander[10]. This model is like a structure of binary tree with the records forming the nodes and fields forming the branches of the tree. The element present in it have a parent-child relationship, relations are only limited between parent and child records. The root of the structure is the great grand parent[5]. To access the data programmer has to write low-level calls in programs using a navigational language. The benefits or advantages of hierarchical model are numerous but there are some Limitations also.

Table I. Pros and Cons of IMS[5]

Pros	Cons
Simple Design	Implementation Complexity and Limitation
Data Security and Integrity	Lack of Structural Independence
Efficient searching	Operational anomalies

C. Network Model

In 1969, the network model was developed which was based upon CODASYL approach. CODASYL standardized IDS and released a publication that described the network model. The second publication was released in 1971[10]. The network model overcame the limitations of hierarchical model and the paradigm was also different from the hierarchical counterpart, this model replaces the hierarchical tree with a directed graph so there are more connections among the nodes. One of the main advantages of network model over hierarchical model is that network model allows a record to have more than one parent. Thus it allows to navigate on 1:1(one:one),1:M(one:many) and M:M(many:many) relationships among entities. A relationship is a set and a set is a container of pointers which identifies which set of data can be accessed from the record[5].

Table II. Pros and Cons of Network Model[5]

Pros	Cons
Data Integrity and Independence	System Complexity
Ease of data access	Operational Anomalies

D. Relational Model

Edgar Codd who worked at IBM was unhappy with the idea and concepts of network and hierarchical model and he proposed a new approach to the database and represented relational model in his paper on database construction theory, “A Relational Model of Data for Large Shared Data Banks” in 1970. The relational model allows to group data items into one or more tables that are linked by using fields common to each related table. The data can be accessed by using a high-level non-procedural language.

Table III. Pros and Cons of Relational Database Model[5]

Pros	Cons
Structural independence	Hardware overheads
Conceptual simplicity	Difficult to recover the lost data
Implementation, design and usage ease	Software is expensive
Ad hoc query capability	Structured limits

In 1973, Michael Stonebraker's team at UC Berkeley used Codd's idea to create the INGRES relational database[10].

In 1975, IBM developed SYSTEM R which used a structured query language and it was an experimental relational database[10].

E. Entity-Relationship Model

In 1976, the ER(Entity-Relationship) model was proposed by Peter Chen. In this model real world is considered as entities and relationships[5]. And only entities and relationships can have attributes and only entities can be involved in relationships. The ER modelling has top down approach. This

model provides the framework for designing of database and designers can focus on data application instead of table structure[8].

Table IV. Pros and Cons of Entity-Relationship Model[5]

Pros	Cons
Straight forward relation representation	No industry standard for notation
Easy conversion for E-R to other data models	Popular for high level design
Graphical representation for better understanding	

In the early 1980s ISO and ANSI adopted SQL as the standard language and IBM introduced DB2 which became the flagship product of IBM.

F. Object-Oriented Database Management System

With the rise of object-oriented programming languages,the requirement of applications also increased which relational model was unable to fulfill due to some shortcomings like simple data types, loss of data protection, etc. So, the concept of object-oriented databases was conceived in which data is represented as objects rather than as tables. In 1985, object-oriented database system was introduced. Won Kim headed the Orion Research Project at MCC which was the first major project initiated for the development of OODBMS. OODBMS does not have a single object-oriented paradigm, there are various object-oriented data model[2].

Table V. Pros and Cons of OODBMS[5]

Pros	Cons
Extensibility	Lack of support for views and security
Capable of handling a large variety of data types	Locking at object level may impact performance
More expressive query language	Query optimization compromises encapsulation
Support for schema evolution	Lack of universal data model
Applicability to advanced database applications	Complexity

In 2004, the first free open-source OODBMS db4o was released and it was also the first database management system to implement native queries in the programming languages like Java and C#[2].

OODBMS provided some very nice features which are not present in relational databases but it hasn't gained much acceptance like relational database management system because of some of the disadvantages like no standard query language and no universal model.

G. Object Relational Database Management System

In the 1990s, the idea of adding object-oriented principles with relational database was given by Stonebaker and his team. Object Relational Database(ORDBMS) came into existence by incorporating the good features of both types of databases forming a hybrid one which will connect relational databases and object-oriented modeling techniques used in programming languages, such as Java, C++,C#, etc[3].

Table VI. Pros and Cons of ORDBMS[5]

Pros	Cons
Reuse and sharing	Complexity
Increased Productivity	Increased costs
Large storage capacity	Not appropriate for web applications

Some of the examples of ORDBMS includes Oracle Database by Oracle Corporation,SQL Server by Microsoft,Greenplum Database by Pivotal Software,Informix by IBM.

In 1997, extensible markup language(XML) was introduced which is self-descriptive and it defines some principles for encoding documents in a format which is understandable for both human and machine. It resolves old database problems so major vendors started to integrate XML into DBMS products[1].

H. NoSQL

In 1998, Carlo Strozzi used the term NoSQL to name his Strozzi NoSQL open-source relational database that did not expose the SQL but it was relational[9]. NoSQL was reintroduced on June 11,2009 by Eric Evans, a Rackspace employee, when a meet up was organized by Johan Oskarsson, a software developer based in London. NoSQL stands for not only SQL. NoSQL databases are those databases that are non-relational and open source thus anyone can look into it, update and compile it accordingly. NoSQL data models can be categorized into four categories;Key-value data model, Document data model, Column-family model and Graph data model[9].

In 2011, database developers started to work on UnQL(Unstructured Query Language) which is claimed to be a superset of SQL. UnQL is a specification for a query language for NoSQL databases[1].

In a 2011 research paper, Matthew Aslett used the term NewSQL. It is a class of RDBMS which aims to provide the scalability of NoSQL systems for Online Transaction Process(OLTP) while still providing the ACID attributes ensured by traditional relational databases[9]. It is distributed, non-relational, open source and support semi-structured and unstructured data.

I. Most Popular Databases

According to DB Engines score, the top three databases present in the market are Oracle, MySQL and SQL server. Oracle is ranked 1st.

Table VII. Databases-Year of invention and their popular products[3]

Databases	Year	Products
Foundational databases- Hierarchical and Network database	1960s	-IMS -IDMS -Univac DMS-1100 -RDM server
Relational Database	1970	-Oracle -MySQL

		-Informix
Object-Oriented Data Model	1985	-Gemstone -GBase -VBase
Object Relational Data Model	1990s	-IBM DB2 -Postgre SQL -Illustra
NoSQL	2009	-HBase

J. Big Data

Big Data refers to enormous set of data both structured and unstructured, which is too complex to be dealt with traditional processing techniques or algorithms.

1) IBM scientist mentioned that Big Data has four dimensions:Volume,Velocity,Variety and Veracity[6].

a) Volume: The amount of data collected daily is enormous. Current data exits in petabytes and it is predicted that in few years data will exist in zettabytes which is problematic. With big data high volume of unstructured data is processed.

b) Velocity: The rate at which date is received is known as velocity. Data is too large and continuously in motion which challenges the traditional data analytics[6].

c) Variety: Data does not belong to only a specific category. Data is structured and unstructured. Traditional analytical methods manages structured data which fit well in relational database but fails to manage unstructured and semistructured data types.

d) Veracity: There are inherent discrepancies in all the data collected. And veracity refers to the noise and abnormalities within the data. Ambiguity is the primary focus[6].

By 2020, there will be approximately 50-100 billion connected devices leading to massive amount of data collection which illustrates a necessity for systematic processing and analytics[6]. Therefore, the transformation from traditional analytics to big data analytics is needed .

2) BIG DATA and Hadoop: Hadoop is an open source framework that supports the processing of big data in a computing environment. It has two main components i.e. Hadoop distributed file system and Mapreduce[12].

a). Mapreduce: It is a method which breaks a data problem into smaller data sets and run them parallely. It uses the map function to parse data and then the data is summarized by reduce function.

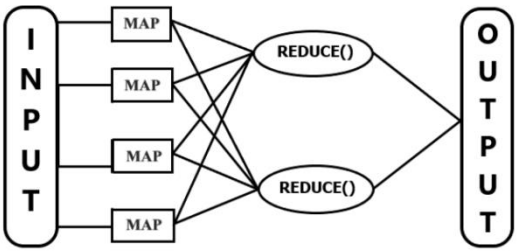

Fig.1. Mapreduce[11]

b). Hadoop distributed file system: It is virtual distributed file system. When we add a file to this system it is automatically split into small configurable fixed size blocks.

IV. ROLE OF BIG DATA IN CLOUD

In today's era, two main technologies square measure the middle of concern in IT –Big Data and Cloud Computing. Basically completely different, Big data is all concerning managing the huge scale of knowledge whereas Cloud computing is concerning infrastructure. However, the simplification offered by Big Data and Cloud technology is the main reason for their vast enterprise adoption. For instance Amazon "Elastic Map Reduce" demonstrates how the facility of Cloud Elastic Computes is leveraged for Big data processing.

The combination of each yields useful outcome for the organizations not to mention, each of these technologies square measure within the stage of evolution however their combination leverages expandable and efficient answer in big data analytics.

The Big Data can be related to cloud computing based upon various technologies which can be classified as under[12]:

1). Software as a service in hybrid cloud: Companies want to analyse massive amounts of data coming from social media for their business analysis. Cloud integrate data from various resources and big data analysis is performed for better results.

2). Platform as a Service in Private Cloud: When PaaS vendors have to deal with petabytes or terabytes of data, they don't want to deal with the management of complex softwares and hardwares. They subsume big data technologies for their services.

3). Infrastructure as a Service in Public Cloud: Unlimited amount of storage and computer power can be accessed by people by using Big Data services in IaaS. With Big Data solutions in cloud businesses can be improved with meaningful data insights.

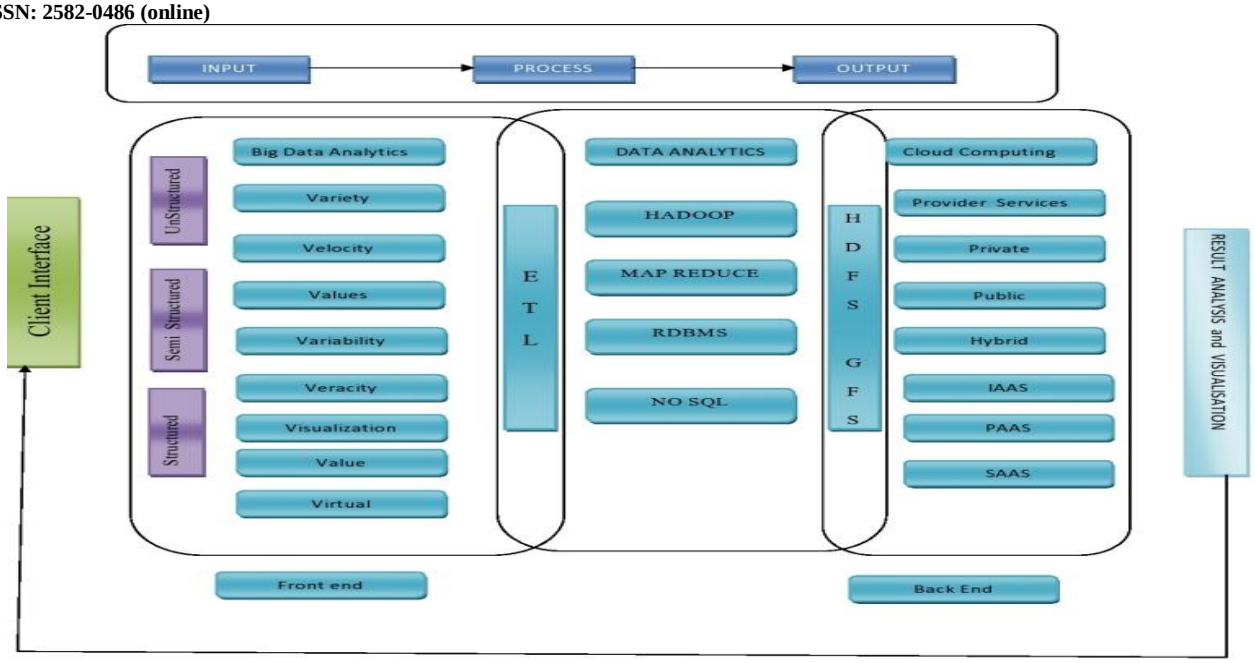

Fig. 2. Model relating Big data and Cloud Computing

After analyzing the relationship between big data and cloud computing, our findings are as below:

1). With the improvement in technology with the emergence of cloud, analysis done using big data technology is becoming more efficient as compared to previous analytical methods.

2). Virtualization technology in Big data applications gives numerous benefits which cannot be accessed by physical infrastructure. The convergence makes applications more dynamic and expandable.

3). The combination of big data and cloud computing lowers the cost of resources provided as the use can be scaled up and down, you only pay for what you use and for how much time.

4). The workloads can be managed easily due to the flexible infrastructure which handles the massive amount of data that has varying types and characteristics.

V. CONCLUSION

Finally we concluded that if technology is the body than database is the soul as no technology became successful without the support of a database. Also as the data from various sources is increasing day by day at a very rapid rate even the present era databases may not be sufficient to cope up with the requirement , hence we definitely going to require some next generation databases which can handle and respond at very high speed for some applications and will become advanced enough to store digital copies of ourselves. Hence there is bright scope for research in this field of advance or next generation databases.

REFERENCES

[1] K. Berg, T.J. Seymour, “History of Databases-”-December 2012

[2] H. Alzahrani, “Evolution of Object-Oriented Database Systems”-2016.

[3] S.Praveen, U. Chandra, A.A. Wani, “A Literature Review on Evolving Database”, 2017.

[4] A.B. Chaudhri, “Object Database Management Systems: An Overview”-1993

[5] P. Bhatia, G. Singh, Simplified Approach to DBMS-10th Edition

[6] K. Taylor-Sakyi, “Big Data: Understanding Big Data”-2016

[7] K. Driscoll, “From Punched Cards to "Big Data": A Social History of Database Populism”-2012

[8] A. Badia, “Entity-Relationship modeling revisited”-2004

[9] M.Dave, “SQL and NoSQL Databases”-2012

[10] J. P. Fry, “Evolution of Data-Base Management Systems”-1976

[11] R. Beakta, “Big Data and Hadoop: A Review Paper”-2015

[12] I.A.T. Hashem, I. Yaqoob, “The rise of “big data” on cloud computing: Review and open research issues”-2015, Journal of Information System, Elsevier.