

Good essay on critique of relational database technology

[Business](#), [Management](#)



Other Details

Abstract

Relational Database Systems (RDBMS) are the most popular database systems today. The relational model of database management emerged from a dynamic database model called DataBase Management System (DBMS). Dbase and FoxPro were the popular DBMS in the 1970's and 1980's. RDBMS emerged in the 1990's. Though RDBMS is a popular form of database management, few appreciate its true potential. Modern applications that allow multiple devices to access and manipulate data are overshadowing RDBMS. NoSQL models such as Hadoop and MapReduce allow document processing which is a feature that the relational model lacks. An in-depth understanding of the structure and concepts of the relational model, its strengths and weaknesses, and the challenges that it faces in the IT industry will provide an insight into the marketability and sustenance of the relational model. In this paper, we first explain the relational database model and examine its strengths and weaknesses. This is followed by a description of other models like ERP, NoSQL and cloud computing. In the concluding paragraphs, we look at the future of Relational Databases in the IT industry and in the business world.

A database is a collection of heterogeneous but related information (Stajano, 1998). Relational Database Technology, also called Relational Database Management System (RDBMS), was first proposed by E. F Codd in 1970. Codd was a computer scientist working with International Business Machines (IBM). RDBMS was his most significant contribution to computer science (Edgar F. Codd from the ACM Portal). According to Stajano (1998), RDBMS's

<https://assignbuster.com/good-essay-on-critique-of-relational-database-technology/>

are called relational because the data storage is based on mathematical relations. Structured Query Language (SQL) is used by most RDBMS's to store and retrieve data (Stajano, 1998). In this paper, we examine the relational database model and its efficacy in the context of rapidly advancing technology.

When one studies a particular subject, field, or topic, it is first necessary to familiarize oneself with the terminology of that field or subject. Listed here are some of the common terms used in relational database technology. (UHI Millennium Institute).

Row - A row or a record is a set of data values related to a common area or item.

Column - A column is data of a particular type.

Entity – An entity is an object like a table or an occurrence like a relation between tables.

Attribute - An attribute is an element in an entity that describes that entity.

Object - An object may be a table that stores data or a program or method that uses the data.

The computer stores data in files. Files are stored on hard disk, compact disk, and other storage devices. Programmers write programs to manipulate the data within a computer. This is known as data access. Data access refers to reading and writing data to and from the computer. There are mainly two modes of accessing data in a computer - sequential and random. Sequential access is one in which records are stored sequentially and must be read sequentially. For instance if one wants to access the 50th record in a file, one

must first read the preceding 49 records to do so. With the introduction of indexing, random access became possible. Random access allows programs to directly access the required record without having to read all the preceding records. Relational database systems use random access to read and write data (itec. hust. edu. cn).

A multidimensional database or a relational database is a database system designed to store data and retrieve it efficiently. RDBMS stores data in the form of tables rather than files. Each table consists of a predetermined number of columns and infinite rows. Each row in a table represents a dataset. The tables are related to one another using relations. The data within a table is homogeneous. A collection of interrelated tables makes up a database. Data is retrieved from a database by linking the tables using the relationships. This data is then presented in the required format.

A research project by the name of System R was first launched by IBM in 1974, following the introduction of relational database by E. F. Codd in 1970 (National Academies Press, 1988). The first RDBMS to be commercially launched was called SQL/DS (Structured Query Language/ Data Storage). This product was launched in 1988 (S. Sumathi, 2008). Later Oracle emerged as that largest relational database model. Other commercial models included Sybase, DB2 and Informix (Profit Magazine, 2007).

Data is stored in computers in a structured form. The structure may be sequential or random but a structure exists. Structured data can be easily accessed by computer programs and then presented in the desired format. As business expanded and markets became international, the demand for quick access to large volumes of data grew. Sequential access files of

yesteryear could not meet these demands. The concept of indexing gave birth to random access. This later evolved into database management systems (DBMS) and thence to RDBMS. The way data is physically stored in a computer is sequential, even in an RDBMS. The mode of access is, however, different. Indexing techniques allow for large volumes of diverse data to be quickly and easily accessed (Microsoft).

Indexing is a technique by which a programmer can quickly access the required data from a large file. An index is a structure that contains pointers to data. Pairs of data values and pointers are stored in a file. The value can be a tuple or row. For example, a table indexed on customer-id may have an index file consisting of customer-id and a pointer to the customer-id.

Broadly indexing can be of two types - Hash index and B-Tree index. A detailed discussion on indexing techniques is beyond the scope of this paper. (Microsoft)

Structured Query Language (SQL) is the programming language used to manage a relational database. SQL is made up of two parts - data definition language (DDL), Data manipulation language (DML). DDL is used to create and modify database and table structures. DML is used to insert data into the tables, retrieve data, and update the tables with new data (Microsoft).

Database management has evolved over the last four decades. The software industry has faced many challenges like changing requirements of the user, increase in the number of users, demands for dynamic access, and advances in hardware technology. These challenges have been met by adapting the concept of RDBMS to the new technology. Many new software concepts like big data, cloud computing, Oracle, ERP, and SAP have emerged because of

this evolution. It would not be appropriate to discuss in detail the evolution of the database systems here. Feuerlicht (2010), in his paper entitled “ Database Trends and Directions: Current Challenges and Opportunities”, presents the evolution of database and the future of Database technology. According to Feuerlicht (2010), the emerging NoSQL movement contends that RDBMS is inefficient and unable to provide solutions for non relational data. Proponents of NoSQL prefer Open Source technology. Quoting Dean (n. d.), and Borthakur (2007), Feuerlicht (2010) presents the example of MapReduce by Google and Hadoop, two new software technologies used to process text data. MapReduce uses parallel programming to facilitate data management on multiple machines. Feuerlicht (2010) opines that database concepts should not be applied to applications that involve unstructured data. He suggests that the scope of database paradigm should be clearly defined.

According to The University Oof West England (UWE, n. d.), RDBMS’s supports a simple data structure. The structure limits duplication of data. Since updates are required only in one place, the possibility of inconsistency is greatly reduced. Integrity of data is also ensured by table level constraints. The database is independent of the program that is used to manipulate data. This makes the relational database portable. The structure of the database is independent of the program that created it. It is said to be logically independent. Different users can view the data in different ways because of this independence. RDBMS supports dynamic queries. These queries can be independent one time queries or programmed to prepare periodic reports (UWE, n. d.).

Despite these advantages, the RDBMS model has its drawbacks. The University Of West England (UWE) outlines some of these limitations. The process of normalization often results in virtual entities that do not exist in the real world. The model does not clearly differentiate between entities and relationships. Since data needs to be retrieved from several entities, querying becomes a complex process. RDBMS is based on entity relationships. These relationships are linear and not hierarchical. It therefore becomes difficult to represent hierarchical data in a relational database. Sequential processing which was a drawback of earlier forms of data representation was overcome by RDBMS. The relational model however swung the pendulum to the other end, by eliminating sequential processing. This can be construed, as a major drawback where sequential processing is required (UWE, n. d.).

Over the last four decades, the software industry has faced many challenges. Changing requirements of the user, increase in the number of users, demands for dynamic access, and advances in hardware technology are some of the major challenges facing the software industry. Many new software concepts like big data, cloud computing, Oracle, ERP, and SAP have emerged because of this evolution. According to Feuerlicht (2010), the emerging NoSQL movement contends that RDBMS is inefficient and unable to provide solutions for non relational data. Proponents of NoSQL prefer Open Source technology. Quoting Dean (n. d.), and Borthakur (2007), Feuerlicht (2010) presents the example of MapReduce and Hadoop, two new software technologies used to process text data. MapReduce uses parallel programming to facilitate data management on multiple machines.

Feuerlicht (2010) suggests that the scope of relational database paradigm should be clearly defined.

For the last four decades, RDBMS have been used in nearly all applications. Web programmers use databases to store data online. Commercial packages like Oracle, ERP and SAP use this technology as a back end. As technology advances, some changes have been seen in the use of information technology. Single dedicated servers are giving way to multiple servers over geographically diverse locations. Cloud computing is the latest trend evolution in information technology. Its low cost and scalability has made it popular in the business world. Increasing volumes of data have given rise to the need for cloud computing and cloud storage (Buyya, 2009 in Arora & Gupta, 2012).

Cloud storage refers to DaaS (Data as a Service) model of cloud computing. In this model, the service provider provides data management as a service. A user may store his data on the cloud instead of his own office computer. The cloud is a virtual storage area that allows users to store data, documents and other objects. Some cloud storages like DropBox, allow storage of documents only while others restrict themselves to databases (Wu et. al. 2010 in Arora & Gupta, 2012). Database storages are mainly used for backing up data. DBaaS or Data Base as a Service offers complete database management services. It allows users to store and retrieve data at any time from any place. Examples of DBaaS are RDS from Amazon, BigTable from Google and Sherpa from Yahoo (oracle. com, 2011, in Arora & Gupta, 2012). Enterprise Resource Planning (or ERP) systems are being widely used by business around the globe. Earlier used only by big businesses, ERP is now

being used by small and medium business too. ERP system is a software used for business management. It comprises of several modules each representing a functional area of business such as sales, accounting, inventory, and HRM. The modules are integrated and information flows from one module to another within the package. Using ERP allows corporates to work on a single system and build an information base that processes data across all functional areas (Kumar & Van Hillsgrsberg, 2000, in Rashid et. al., 2002). ERP systems process the transactions of the organization and make real time information available to the management (O'Leary, 2001 in Rashid et. al., 2002).

ERP systems have a modular design. Each module comprises of one functionality of business. A common database is used to store data. The different modules exchange data through the database. This increases transparency between functionalities. Because of the multiple functionalities and large volume of data, ERP systems are costly. The system is highly flexible and an enterprise may opt for selected functionalities if desired. ERP can be tailored to the needs of the organization. Installation is a time consuming process. Few ERP systems work on the internet but efforts are being made in this direction.

A new model of database NoSQL is gaining recognition. NoSQL databases are used where the volumes of data are large and it is more important to retrieve simple data rather than complex related data. Although NoSQL is gaining popularity with large volumes of data, its non-relational structure does not allow dynamic retrieval. and therefore this model cannot be use in areas where the data is complex and required by multiple users for diverse

purposes. For companies that require high volumes of data storage, high scalability and high availability, NoSql is the ideal model. This model allows quick retrieval of data from a large database (North, 2010 in Nance et. al., 2013). An e-commerce website on the other hand has a lot of data that is not permanent for example a shopping cart. Such data is can only be supported by the relational model (Sadalage and Fowler, 2012 in Nance et. al., 2013). Decisions regarding use of RDBMS or NoSQL will depend on the problem to be resolved or the purpose of storage. A hybrid mix of the two models may be used in some cases.

A rRelational database management system (RDBMS) was first proposed by E. F. Codd in 1970 in an effort to overcome the drawbacks of sequential processing. Prior to 1970, data was stored in sequential files and retrieval required processing all preceding records before accessing the required record. The concept of RDBMS was widely accepted. It facilitated dynamic access and ensured data integrity at storage level. . Elimination of data redundancy and ensuring consistency of data were the two major strengths of this model.

In this paper, we examined the relational database model, and studied its strengths & weaknesses. RDBMS has evolved over the years, and a number of software programs based on this model have become commercially available. These include Oracle, SAP (System Applications and Products), and ERP (Enterprise Resource Planning). These programs take advantage of the structured nature of RDBMS.

One major drawback of the relational model is that it does not support sequential processing. The loop construct becomes complex when applied to

the relational database. Another major drawback is that RDBMS does not support hierarchy of data. The relational model therefore is not suited to process text data or data that requires recurrent processing, like payroll data.

NoSQL is a new model of SQL that has been designed to overcome these drawbacks of RDBMS. The perspective of SQL programmers has changed rapidly with the emergence of dotcoms and the need for internet programming. New programming languages such as java and JQuery have emerged that can manipulate data in a relational database more efficiently. The advent of cloud computing has further changed the concept of internet programming. The concept of data storage has also undergone a change. Big data, which refers to large volumes of data accessed by multiple users, has given rise to new techniques for data manipulation. ERP systems are moving to the cloud. Cloud databases are being adopted by organizations to store large volumes of data and access data it from multiple devices. This poses the challenge of security, Security of data has become a major concern and cloud databases are increasingly being used for backup. Cloud storage poses a new problem of security, consistency, and integrity of data. Efforts are being made to meet these challenges with new techniques and technology. As technology advances, new concepts emerge, needs of the users change, and as a result concepts of data storage and programming also undergo change. This change is a continuous process. The challenge for the software industry is to meet the rapidly changing needs of the users. With new devices being launched, programs need to be altered to adapt to these devices, while at the same time maintaining integrity, and security of data.

Research in the field of technology and information technology is an ongoing process. In this paper, we have examined the concept of relational data base. Further research into the prevalence, adaptation, and variations of this model as well as and its future in the IT industry, will show, whether the relational database model is becoming outdated and being replaced by NoSQL. Questions like will NoSQL and models of NoSQL such as Hadoop and MapReduce replace SQL? Will ERP move to the cloud or will it become redundant in the face of newer models? These are questions that can be answered by further research.

References

" Oracle Timeline" (PDF); Profit Magazine (Oracle) 12 (2): 26; May 2007; Retrieved 2013-05-16.

" Microsoft" Structured Query Language (SQL)". Retrieved From [http://msdn.microsoft.com/engb/library/windows/desktop/ms714670\(v=vs.85\).aspx](http://msdn.microsoft.com/engb/library/windows/desktop/ms714670(v=vs.85).aspx)

Carlos Ordonez University of Houston, Yeol Song Drexel University, and Carlos Garcia-Alvarado, University of Houston (2010); Relational versus Non-Relational Database Systems for Data Warehousing Copyright is held by the author/owner(s); DOLAP'10, October 30, 2010, Toronto, Ontario, Canada; ACM 978-1-4503-0383-5/10/10. www2.cs.uh.edu/~dbms/dolap2010/p-2010-DOLAP-relnonrel.pdf

Chapter 17; File Processing Retrieved From

CS 452 File Organization Retrieved From osm.cs.byu.edu/CS452/supplements/FileOrg.pdf

Edgar F. Codd from the ACM Portal

<https://assignbuster.com/good-essay-on-critique-of-relational-database-technology/>

Funding a Revolution: Government Support for Computing Research;
National Academies Press; 8 Jan 1999; ISBN 0309062780; System R did not
convince IBM management to abandon its existing product

George Feuerlicht (2010) Database Trends and Directions: Current
Challenges and Opportunities J. Pokorn_y, V. Sn_a_sel, K. Richta (Eds.):
Dateso 2010, pp. 163{174, ISBN 978-80-7378-116-3.

HYPERLINK " <http://itec.hust.edu>.

[cn/~liuwei/common-cpp/text_book/chapter17.pdf](http://itec.hust.edu)" <http://itec.hust.edu>.
[cn/~liuwei/common-cpp/text_book/chapter17.pdf](http://itec.hust.edu)

Indu Arora and Dr. Anu Gupta Cloud Databases: A Paradigm Shift in
Databases; IJCSI International Journal of Computer Science Issues, Vol. 9,
Issue 4, No 3, July 2012; ISSN (Online): 1694-0814; <http://www.ijcsi.org>

S. Sumathi, S. Esakkirajan (13 Feb 2008); Fundamentals of Relational
Database Management Systems; Springer; ISBN 3540483977.

Srikanth Technologies; Oracle for Beginners www.srikanthtechnologies.com

Stajano Frank (1998), A Gentle Introduction to Relational and Object
Oriented Databases Olivetti Research Limited Frank Stajano <http://www.orl.co.uk/~fms/> ORL Technical Report TR-98-2

UHI Millennium Institute Relational Database Management Systems
Database Terms the Database Terms of Reference (Terminology)

UWE (University Oof West England) (n. d.) http://www.cems.uwe.ac.uk/~pchatter/2011/dm/readings/rdb_strengths_weaknesses.html