Database Systems Performance Evaluation Techniques

Subharthi Paul subharthipaul@gmail.com (A project report written under the guidance of Prof. Raj Jain)

Download 

Abstract

The last few decades has seen a huge transformation in the way businesses are conducted. There has been a paradigm shift from product portfolio based marketing strategies to customer focused marketing strategies. The growth and diversity of the market has greatly profited consumers through higher availability, better quality and lower prices. The same factors however has made it more difficult for businesses to maintain their competitive edge over one another and hence has forced them to think beyond their product portfolio and look at other means to gain higher visibility and customer satisfaction, maintaining all the while their core advantages on pricing and product through improved and more efficient methods of manufacturing and distribution. The advent and spread of computers and networking has been one of the single largest factors that has spurred and aided this enormous movement. More specifically, database management systems now form the core of almost all enterprise logic and business intelligence solutions. This survey tries to emphasize the importance of database systems in enterprise setups and looks at the methods and metrics that are used to evaluate the performance of these database systems.

Keywords

Database Systems, Transaction Processing, Performance Evaluation Techniques, Database Benchmarking

Table of Contents

---

1. Introduction

The last few decades has seen a huge transformation in the way businesses are conducted. There has been a paradigm shift from product portfolio based marketing strategies to customer focused marketing strategies. The growth and diversity of the market has greatly profited consumers through higher availability, better quality and lower prices. The same factors however has made it more difficult for businesses to maintain their competitive edge over one another and hence has forced them to think beyond their product portfolio and look at other means to gain higher visibility and customer satisfaction, maintaining all the while their core advantages on pricing and product through improved and more efficient methods of manufacturing and distribution. The advent and spread of computers and networking has been one of the single largest factors that has spurred and aided this enormous movement. More specifically, database management systems now form the core of almost all enterprise logic and business intelligence solutions.

Database Systems are one of the key enabling forces behind business transformations. Apart from supporting enterprise logic they also enable business intelligence. Information is the key to success in today's businesses. However, maintaining information in logically consistent and feasibly retrievable format is a daunting task. More so with the added complications of transaction consistency management, synchronization across multiple repositories spread geographically across the globe, failover management and redundancy management, today's database systems are truly state-of-the-art high performance software systems.

Apart from managing a plethora of complicated tasks, database management systems also need to be efficient in terms of storage and speed. Businesses have a tendency to store un-required historical data often as a result of poor data planning or less frequently owing to federal obligations or consumer law. Dynamic addition and deletion of data from the database also pose a challenge to maintaining an efficient data retrieval mechanism. Though, limited in speed by some sense due to hardware limitations, database systems nonetheless need to achieve full throttle through efficient storage and retrieval techniques. Another factor that often hurt database performance is ill-written computationally expensive queries. Often, such situations are beyond control of the database system and require external performance tuning by experts.

As is true for most systems, reliability, availability and fault-tolerance is a huge concern for database systems. Reliability of a system is generally improved through redundancy. Modern businesses cannot afford to loose data or present wrong data. Modern business activities are highly centered around and dependant on electronic data. Modern database systems thus need to build in high reliability mechanisms in their designs. Availability is another issue that concerns a lot of businesses. As an example, "Netflix" an online DVD rental business receives online requests for 1.9 million DVD's everyday. An outage for 10 minutes cost huge losses to these businesses. Similar is the case for a lot of other businesses. Amazon.com stands to loose approximately $31000 per minute for a global outage [1]. Clearly, availability is a key metric for measuring the performance of database systems.

Another metric, which is more qualitative in nature, but extremely important is security. It is generally difficult to measure the level of security of any system unless its security vulnerabilities are exposed.

Database systems often store sensitive enterprise and customer data that if inappropriately used may cause huge monetary and business losses for the organization. Hence, though difficult to quantify through standard testing procedures, database systems strive for continual upgrade of their security features.

Performance evaluation of database systems is thus an important concern. However, easier said than done, performance evaluation of database system is a non-trivial activity, made more complicated by the existence of different flavors of database systems fine tuned for serving specific requirements. However performance analysts try to identify certain key aspects generally desired of all database systems and try to define benchmarks for them. In the rest of this survey, we shall provide a formal definition of database systems followed by a few methods to categorize or classify database systems. This shall be followed by a look at the various performance evaluation techniques that are employed to benchmark database systems, some of the key benchmarking techniques used in practice in the industry and some open source benchmarking schemes available for use in the public domain.

---

2. Database Management Systems: Formal Definition and Classification

A Database Management System (DBMS) is a complex set of software programs that controls the organization, storage, management, and retrieval of data in a database. DBMS' are categorized according to their data structures or types. It is a set of prewritten programs that are used to store, update and retrieve a Database [2].

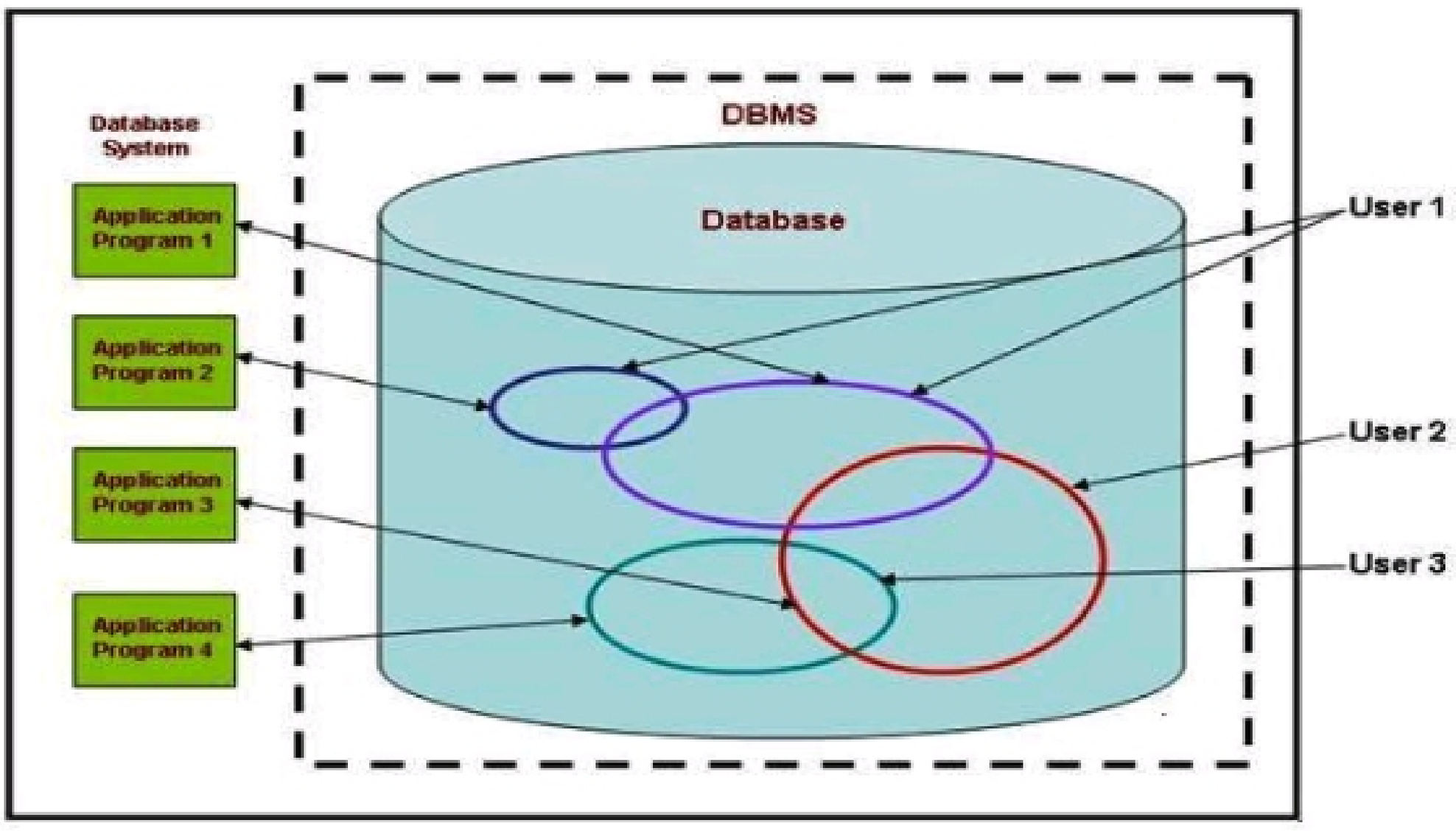

Figure 1 - A generic high level view of a Database Management System

Figure 1 presents the generic high level view of a database management system. The "Database Management System is a collection of software programs that allow multiple users to access, create, update and retrieve data from and to the database and the "Databae" is a shared resource that is at the centre of such a system. The database's functionality is optimal storage and retrieval of data, maintain correctness of the data, and maintaining consistency of the system at all times.