Distributed Database Selection for Vital Registration System of Addis Ababa

Mesfin Debebe
HiLCoE School of Computer Science & Technology
mesfindebebe@gmail.com

Abstract

This paper explores existing literatures on Distributed Databases and how they are implemented by different Database Management Systems. It will focus on identifying the basic properties of Distributed Databases and how they are addressed by two commonly used database management systems: Oracle and SQL Server.

The Vital Registration Agency of Addis Ababa has implemented a centralized database that is accessed by applications running at the different Woreda offices. The users at these offices are facing frequent problems in using the system due to network connection problems.

We propose that a distributed database is an ideal solution for such a problem and the aim of this research paper is to recommend the right technology to be used for the vital registration system in Addis Ababa. Hence, we will attempt to make a comparative analysis on the implementations of a distributed database using Oracle and SQL Server in order to recommend the better option based on the requirements and problems of the Vital Registration System of Addis Ababa.

Keywords: Distributed Databases; Distributed Database Management Systems; Oracle; SQL Server; Distributed Query; Distributed Transaction; Vital Registration System;

1 Introduction

The definition of a Distributed Database (DDB) as given by Özsu and Valduriez [1] is a collection of many databases that are connected by a computer network and a Distributed Database Management System (DDBMS) as a software system that manages these databases that are connected over the network. The different databases need to be logically interrelated and the DDBMS has the responsibility of making the distribution of data over multiple databases transparent to the user or applications. The DDBMS should also have the capability of using the processing and storage capacity of all the different database management systems that maybe located at different location

The different Database Management Systems in a DDBMS may be similar (homogeneous) or different (heterogeneous), but the DDBMS has to have the ability to manage them so that they appear as one database to the user. The databases are

interconnected by a computer network, and they cooperate in performing certain tasks. The Distributed Database Management System coordinates these tasks are coordinates them so that the distribution of the databases appears to be transparent to the user, i.e., the user is not aware of the existence of multipole databases.

Elmasri and Navathe [2] pointed out that a full- scale, dedicated and complete DDBMS that meets the techniques proposed in DDB researches are not yet been developed, and so they are not commercially available. Many Relational Database Management Systems (RDBMS) however, have tried to support the basic features of a DDBMS by including some functionalities to their products, e.g., Oracle and SQL Server.

Vital Registration Systems are systems where a government records the vital events of its citizens and residents. These systems are particularly involved in capturing and maintaining the basic information about each and every citizen and resident which may be residing in different parts of a country. The current Vital Registration System of the city of Addis Ababa uses a centralized database which is accessed from multiple application located at the Woreda sites.

The requirements of such a system can surely benefit from a DDBSs. The benefits of DDBSs becomes even more obvious when there are major problems of connectivity and the fact that data of residents is most likely to be processed at the location of their current residency.

We strongly believe that the existing problems of connectivity and availability can be greatly reduced by using a distributed database. We propose that a distributed database is an ideal solution for such a problem and the aim of this research paper is to recommend the right technology to be used for the vital registration system in Addis Ababa.

The remainder of this paper is divided into four sections. Section 2 briefly describes the general and specific objectives. Section 3 reviews relevant literatures related to Distributed Databases, their implementations and the Vital Registration system of Addis Ababa. Section 4 provides comparative analysis of the implementations of Distributed Databases using Oracle and SQL Server. Finally, section 5 will provide our conclusion.

Thus, the objective of this work is to examine problems of the current Vital Registration System of Addis Ababa and recommend a Distributed Database implementation by making a comparison between different implementations of a Distributed Database by different vendors, namely Oracle and SQL Server.

2 Related Work

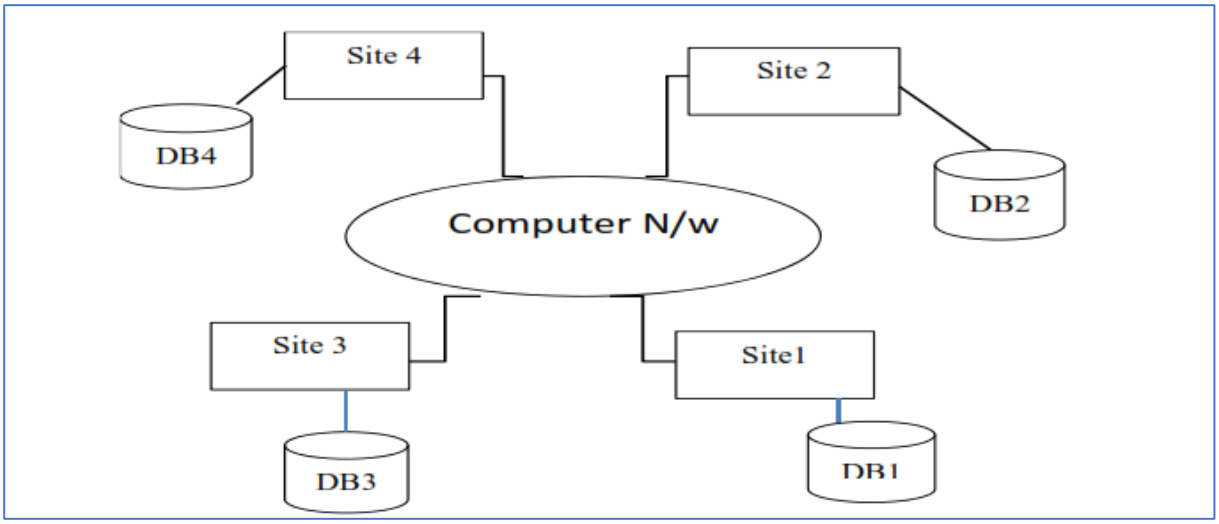
In this section selected research papers, reviews and articles on Distributed Database are briefly discussed. Focus will be made on the basic features of Distributed Database, different implementations of Distributed Database and the particular requirements of the Vital Registration System of Addis Ababa.

Although we have referred several literatures related to the Vital Registration System of Addis Ababa and Distributed Database and their implementations, the most relevant ones are summarized and cited.

2.1 Basic Features of Distributed Databases

Kaur and Singh [3] describe a distributed database management system (DDBMS) as a system that consists of a single logical database. The logical database in composed of a number of fragments and each fragment resides on its own database that is distributed over a computer network. Thus, we can think of a DDB technology as the result of a merger of two technologies: database technology, and network and data communication technology.

The figure given below, extracted from Tomar and Megha [4], illustrates the architecture of a distributed database management system.



In a Distributed Database, a user or an application that is located at one site is allowed to access data that is not found at the local site without actually knowing that the data is physically in another location. This is called distribution transparency. Another important aspect of a DDBMS mentioned by Grech [5], is that a user or an application that is located at one site will be able to process distributed queries; that is, a query should be able to reference more than one site.

We can say that one of the major advantages of a distributed computing systems is the partitioning of a big and unmanageable problem into smaller pieces and solve it efficiently in a coordinated manner. Elmasri and Navathe [2] site this as an additional factor to the economic feasibility of such approach. They claim that this advantage comes from two reasons: more computing power is used to solve a complex task, and each autonomous processing element can be managed independently.

Özsu and Valduriez [1] infer that the benefits of a DDBS and therefore the potentials of the technology can be summarized into four basic features. These are the transparent management of the data that is distributed over multiple databases, the consistent access to

the distributed data using distributed transactions, the improved performance that one gets from accessing local databases, and easier system expansion.

Elmasri and Navathe [2] summarize the basic features of Distributed Databases and conclude that they have to satisfy the following minimum conditions:

- There are different sites (computers) connected over a computer network.
- There are multiple databases and Database Management systems at the different sites
- The databases at different sites have to be logical interrelated.
- The database managements systems at the different sites communicate by sending data and queries (commands) over the network.
- The databases at the different sites may not be the same - in terms of data, hardware, and software.

Elmasri and Navathe [2] also outline some of the important reasons why organizations resort to distributed database management are the following [2]:

- Improved ease and flexibility of application development.
- Increased reliability and availability:
    - When a failure is encountered at one of the sites, other sites continue to operate.
    - Only the data and software that exist at the failed site cannot be accessed.
    - Therefore, in a distributed database, some of the data may be unreachable, but users may still be able to access other parts of the database.
- Improved performance:
    - A distributed DBMS fragments the database by keeping the data closer to where it is needed most. This reduces the delays encountered when accessing data in wide area networks.
    - Queries and transactions executed at a site have better performance not only because the data is not accessed over the network but also the because of the queries and transactions target a smaller local database.
- Easier expansion. In a distributed environment, expansion of the system is easier. Expansion could be in terms of
    - Adding more data
    - Increasing database sizes, or
    - Adding more sites is much easier

2.2 Implementation of Distributed Databases

Many Relational Database Management Systems (RDBMS) have tried to support the basic features of a DDBMS by including some functionalities to their products. This paper will review the implementation of DDBMS in Oracle and SQL Server.

Several DBMS solutions from various vendors exist on the market. In this paper we will compare and evaluate two of them that share major portion of the market: Oracle and Microsoft SQL Server. Oracle and Microsoft offer different versions and editions which can be used for different business needs. In this section, we will compare Oracle 11g and Microsoft SQL server 2005.

In order to make the evaluation of the products, we will start with the basic criteria for a distributed database system and filter these criteria using the particular case of the Vital Registration system of Addis Ababa.

Grech [5] states that the a distributes system provides the following transparencies [5]:

- Fragmentation Transparency: The user is not aware how the data is fragmented
- Location Transparency: The user is not aware of the physical location of the database
- Distributed Transaction Transparency: Transactions ensure the integrity and consistency of the database by concurrency and failure recovery
- Replication Transparency: The user is not aware of the replication of the data
- Support Heterogeneous System: Ability to connect with other databases which have different DBMS product
- Operating System: The Operating System on which the DBMS is capable of operating
- Cost: The cost of the solution

	Oracle	**Microsoft SQL Server**
Product Version / Edition	Oracle 11g Standard Edition	SQL Server 2005 Standard Edition
Fragmentation Transparency	NO	NO
Location Transparency	YES	Yes
Distributed Transaction Transparency	YES	YES (MSDTC)
Replication Transparency	YES	YES
Support Heterogeneous	YES	YES
Support Homogeneous	YES	YES
Operating System Support	Windows, UNIX, Linux	Windows

Functionality Comparison between Oracle and SQL Server

The above table shows that both Oracle 11g and Microsoft SQL server 2005 can be used

for the implementation of the Vital Registration system of Addis Ababa.

2.3 The Vital Registration System of Addis Ababa

The authors in [6] describe Vital Registration, also referred to as Civil Registration, as the process of recording the different important events about the residents of a nation as they happen. Some of the events that need to be recorded are the births, deaths, marriage, divorce, fetal death, annulment, judicial separation and adoption. The registration of these events makes them legitimate and accessible for statistical purposes.

Nielsen and Sahay [7] further enforce this concept and infer that the records resulting from Vital Registration systems have two major benefits. The first one is that they serve as personal legal documents that are required by each resident as proof of facts. Examples are a person's age, identity, children, etc. The second benefit is that the information derived from these records form the basis of a country's vital statistics system.

Hence, vital registration systems are used to derive the fundamental demographic and epidemiological measures that are needed in national planning across multiple sectors such as education, labor and health. They are also critical for a wide range of government activities (e.g., population registers and other administrative registers) and commercial enterprises (e.g., life insurance, marketing of products).

The World Health Organization (WHO) [8] has recognized a well-functioning Civil Registration and Vital Statistics (CRVS) information system as a key ingredient in strengthening CRVS systems in general. Further, the United Nations Commission on Information Accountability (COIA) for Women and Child health have especially mandated that countries would need to strengthen their CRVS information systems.

Furthermore, Setel, Macfarlane et al. [9] state that Civil Registration (CR) as being defined by the United Nations as the universal, continuous, permanent and compulsory recording of vital events. They also state that this registration process needs to be enforced through decree or regulation in accordance with the legal requirements of each country.

In the UN Resolution [10], the United Nations has adopted a resolution "Fundamental Principles of Official Statistics" on January 2014. This resolution states that member countries are required to keep official statistics. Principle 8 of this declaration states that "Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system".

Braa and Sahay [11] emphasis that the task managing Civil Registration and Vital Statistics systems is fundamental and that it requires a multisectoral approach. This should include al the minimum, the Ministries of Health, Justice and Finance. This approach calls for a collaboration between the different sectors and would help the different systems talk to each other and share data. With this regard, the World Health Organization [12] explicitly states that a holistic approach to the development of health information systems is required.

In line with this, the authors in [13] acknowledge that the Government of Ethiopia has given great emphasis to vital events registration across the country. Accordingly, it has put in place appropriate policies and allocated the required resources. Hence a system has been put in place up to the lowest administrative level. The effort included two Proclamations for the enforcement vital registration. The first is Proclamation No.760/2012: Registration of Vital Events and National Identity Card Proclamation. The second is Proclamation no.1049/2017: Vital Events Registration and National Identity Card proclamation (amendment) proclamation.

The Addis Ababa Vital Registration system was launched even before any legislative framework was formulated. This was done in 2008 under the shared responsibility of the Ministry of Science and Technology and the Vital Registration Agency of Addis Ababa. The responsibilities of the Ministry of Science and Technology has now moved to the Ministry of Innovation and Technology. Before 2008, there was no centralized database, and major problems in security and redundancy.

The authors in [6] indicated that the vital registration system of the city of Addis Ababa has put in place a centralized database where all the centers (woredas) of registrations are linked. They [6] further stare that all these centers of registration along with the central database are connected over the woreda.net network. This woreda.net network is also being used by so many government agencies and hence is loaded with a lot of traffic.

The Woreda network is being administered by the Ministry of Innovation and Technology and the Vital Registration Agency has to report any network related problem of the system (which is frequent) to the Ministry of Innovation and Technology. This is also another major factor that contributes to the inefficiency of the system.

The existing system works with different type of data. Besides the textual, numeric and date information captured by the system, photographs and biometric identification like figure prints and iris scans are used. Currently, the system is being used for preparing the newly designed Digital ID. This new digital ID is now being printed at one location only, the central office.

3 Comparative Analysis

In this section a comparative analysis of the implementations of a Distributed Database using Oracle and SQL Server for the Vital Registration System of Addis Ababa will be made.

Both Microsoft SQL Server and Oracle Database are popular relational database management systems but can also be used for the implementation of distributed databases. Hence the comparison will be based on how well suited these technologies will be for the Vital Registration system for the city of Addis Ababa.

We will explore the basic features of Distributed Databases and select the ones that are more relevant to the Vital Registration system of Addis Ababa. Hence, our comparison will

be based on this set of parameters.

Based on the requirements and problems of the Vital Registration system of Addis Ababa, the following parameters have been selected to make the selection of a DDBMS for its implementation:

- Integration of the system with other systems is fundamental.
- This implementation should handle huge amount of data over the network.
- We should be able to send and retrieve data and commands among different Woredas/Kebeles easily and quickly.

3.1 Design Considerations of a Distributed Database

There are different factors that need to be considered when designing a system. The design has to take into account the particular need of the system. The same is true in designing Distributed Database system. We should not only consider how the system is to be distributed, but also how the specific issues of the problems are to be addressed.

We have identified the following main features as major factors in the design of the Distributed Database system for the Vital Registration System of Addis Ababa. These are: Connectivity, Fragmentation, Distributed Transaction and Replication.

Connectivity: The Distributed Database system needs to connect to both homogenous and heterogeneous databases.

Fragmentation: Fragmentation is the partitioning of the database data structures into smaller portions. Three types of fragmentation exist: horizontal fragmentation, vertical fragmentation and mixed fragmentation. In the case of the Vital Registration System of Addis Ababa, Horizontal fragmentation will be used to allocate data into their respective sites (Woreda/Kebele). Meaning that the data for each resident will be stored according to the place of residence i.e., it will be located at the sites where it would be mostly used.

Distributed Transaction: It is believed that when data is stored horizontal fragmentation according to the place of residence, it will significantly reduce the concurrency problem that may be encountered in a distributed database implementation. Thus, the system can use serializability which means that transactions are executed serially, without any major conflict.

Replication: Replication is the process of maintaining multiple copies of data items at different locations. When changes occur, they are captured at the local database. These changes are then sent to the other databases or to the main database at a convenient time. Hence, since several copies exist, if a site becomes unavailable, the DDS can still respond to queries using the replicas.

Though replication is not a main feature of a Distributed Database system, it is a major consideration in its design in order to provide high availability and fault-tolerance.

Support to Standard Operating Systems: It is preferable for the Distributed Database Management System to support the Standard Operating Systems.

According to the design consideration stated in the previous section Connectivity, Fragmentation, Distributed Transaction, Replication and Support to Standard Operating Systems were identified as the critical factors for the implementation of the Vital Registration System of Addis Ababa.

Connectivity with homogeneous database systems is supported by both Oracle and SQL Server. Oracle makes use of database links while SQL Server uses linked server links to connect with other databases. In case of heterogeneous database connectivity, generic connectivity such as ODBC is provided by both Oracle and SQL Server. Oracle provides a larger number of possible data sources when compared to SQL Server. The Vital Registration system may involve only homogeneous databases, but it is anticipated that integration with other systems may be required in the future. Both Oracle and SQL Server provide connectivity to heterogeneous systems but Oracle provides a larger number of possible connections.

Though both Oracle and SQL Server do not have built-in features to distribute the fragments, the distribution can be included in the application logic. But they still provide location transparency which can be achieved by the use of synonyms and views in both Oracle and SQL Server.

Distributed Transactions are supported by both solutions by using the two phase-commit mechanisms. In the case of SQL Server, an additional component (Microsoft distributed Transaction Coordinator - MS DTC) has to be installed and configured.

Oracle uses its Oracle Streams and materialized views to support Replication while SQL Server provides three types of replications which are Transactional, Merge and Snapshot replication.

Kaur and Singh [3] report that Oracle supports Windows, UNIX and Linux Operating Systems while SQL Server support only the Windows Operating System. Microsoft's MSDN [14] claims that, beginning with SQL Server 2017, the platform is now available on Linux as well.

Another factor that is considered in the implementation of databases is the size of the data that needs to be stored. In this case the size of the data that need to be captured for each resident and the total population of the city needs to be considered.

In regards to the size of the data that need to be captured for each resident; we have made an estimation based on the data structure of the current system. The authors in [13] state that the basic registrations are done on the events of birth, mirage and death. Most of the data capture during these events are alpha-numeric and do not have an extensive need of massive storage capacity.

The National Central Statistical Agency [15] reports that though there are different estimations of the population of Addis Ababa, some taking it as high as close to 5 million. However, based on the last census on 2015, the population of Addis Ababa is estimated to be 3,273, 000 (CSA, 2015). Furthermore, this population is distributed into more than 10 Sub-cities.

With this amount of population and the data that needs to be captured for each individual, the total amount of capacity required for the databases at each Sub-city or Woredas will not be an issue for both Oracle and SQL Server.

Another factor is the price of the products. The Enterprise edition of Oracle costs about 20,000USD per year while Enterprise edition of SQL Server costs about 12,000USD per year [16], Definitely, SQL Server is significantly cheaper than Oracle.

4 Conclusion

Based on our findings, both Oracle and SQL Server have their own features for implementing a Distributed Database system for the implementation of the Vital Registration System of Addis Ababa. We can say both provide a feasible solution, even when we consider the particular needs and problems of the Vital Registration System of Addis Ababa. The ultimate breaking point boils down to cost, and SQL Server is significantly cheaper than Oracle.

Thus, our recommendation for the choice of a Database Management system for implementing a Distributed Database system for the Vital Registration System of Addis Ababa is Microsoft's SQL Server.

It should be noted that our recommendation is solely based on the features of Oracle and SQL Server in relation to the requirements of the Vital Registration System of Addis Ababa, Recent versions of Oracle and SQL Server may also have improved or new feature that complement the implementation of Distributed database systems.

References

[1] M. Tamer Özsu and Patrick Valduriez, *Principles of Distributed Database Systems*, Springer Science & Business Media, LLC 2011.

[2] Ramez Elmasri & Shamkant B. Navathe, FUNDAMENTALS OF Database Systems, 6th Edition, 2011.

[3] Kirandeep Kaur and Harwant Singh, "Distributed database system on web server: A Review" *International Journal of Computer Techniques* -– Volume 3, Issue 6, Nov - Dec 2016.

[4] Parul Tomar and Megha, "An Overview of Distributed Databases" Volume 4, Number 2, 2014.

[5] Joseph Grech, "Designing and Implementing a Distributed Database for a Small Multi-Outlet Business", Regis University, Summer 2009, Available at: https://epublications.regis.edu/theses/5

[6] "Rapid assessment of national civil registration and vital statistics systems", Technical document, 9 May 2010, Available at: https://www.who.int/publications/i/item/rapid-assessment-of-national-civil-registration-and-vital-statistics-systems

[7] Peter Nielsen and Sundeep Sahay , "CIVIL REGISTRATION AND VITAL STATISTICS: A NEGLECTED AREA IN INFORMATION SYSTEMS RESEARCH", University of Oslo, Norway, Murodillo Latifov, EPOS Health Management, Tajikistan, Proceedings of the 13th International Conference on Social Implications of Computers in Developing Countries, Negombo, Sri Lanka, May 2015

[8] World Health Organization WHO, "Keeping promises, measuring results: Commission on information and accountability for Women's and Children's Health" , Report, Geneva, 2011

[9] Setel, P.W., Macfarlane, S.B., Szreter, S., Mikkelsen, L., Jha, P., Stout, S., and AbouZahr, C. 2007. "A scandal of invisibility: Making everyone count by counting everyone", The Lancet. 370, 9598, 1569-1577.

[10] United Nations, General Assembly Resolution, "Fundamental Principles of Official Statistics", A/RES/68/261, 73rd plenary meeting, 29 January 2014

[11] Braa, J. and Sahay, S, "Integrated health information architecture: Power to the users", Matrix Publishers, Delhi, India, 2012

[12] World Health Organization, "World health statistics 2007", Geneva, 2007.

[13] UNICEF Website, Available at: https://data.unicef.org/crvs/ethiopia/

[14] Microsoft MSDN, https://docs.microsoft.com

[15] National Central Statistical Agency, 2015), Available at: https://www.statsethiopia.gov.et

[16] Oracle Database vs. Microsoft SQL Server, Available at: https://www.datamation.com/