# Systems Design and Implementation a new paradigm of an Enterprise Data Warehouse Prototype for Business Intelligence Platforms

Bachchu Singh[1], Dr. Pushpneel Verma[2]
[1]*Research Scholar, Deptt. of Computer Science & Application, Bhagwant University, Rajasthan, India*
[2]*Associate Professor, Bhagwant University, Rajasthan, India*

## ABSTRACT

*In this paper also includes a technical piece discussing database requirements and technologies used to build and refresh the data warehouse. In this paper discusses how one can integrate data from databases and other data warehouses. In addition, there is a discussion of specific data warts within the warehouse to meet a specific requirement. Finally, there are explanations for how users will consume data in the enterprise data warehouse, such as through reporting and other business intelligence. This discussion also includes topics in system architecture on how data from different departments' databases and other data warehouses can integrate. An enterprise data warehouse prototype will be developed that will show how a pair of different databases undergo the Extract, Transform and Load (ETL) process and load into the actual set of star schemas and then make reporting easier. Separately, an important piece of this paper takes a real example of data and compares performance between them by running similar queries against different databases, a transaction, and a data warehouse. As queries increase in difficulty, the difference between the actual recorded time of running the same query in a larger environment increases.*

**Keyword:** - *ETL, Warehouse, Data, , environment & architecture etc.*

## 1. INTRODUCTION

Demand for business intelligence (BI) applications is also increasing at a time when demand for most information technology (IT) products is soft. Nevertheless, the research of information systems (IS) in this area is to make it rare, characteristically. While the term business intelligence is relatively new, computer-based business intelligence systems appear in one guise or the other, closer to forty years ago. BI replaced decision support, executive information systems and management information systems. With each new iteration, capabilities increased as enterprises became more sophisticated in their computational and analytical needs and computer hardware and software matured. The BI system in this study is defined as follows: The idea contained in this definition is (perhaps ideally) that business intelligence systems deliver information delivered at the right time, in the right place, and in the right form to help decision-makers. The objective is to improve the timeliness and quality of inputs for the decision process, hence facilitating managerial work. Sometimes business intelligence refers to on-line decision making, that is, quick response. Most of the time, this refers to narrowing the time frame so that intelligence is still useful to the decision-maker when the time comes. In all cases, the use of business intelligence is seen as being proactive. The essential components of Active BI are:

- Real-time data warehousing,
- Data mining,
- Automated anomaly and exception detection,
- Proactive alerting with automatic recipient determination,
- Seamless follow-through workflow,
- Automatic learning and refinement,
- Geographic information systems
- Data visualization

## 2. LITERATURE REVIEW

SOA has been an evolving sector since the term SOA was coined in the year of 1998 as Service-Based Architectures (SBAs). Since then SOA has been evolving in terms of approaches, designing and implementations. Some of the studies and implementations are described further in this section. Mohsen Mohammadi [2018] and Muriati Mukhtar [2018] have reviewed the basic concepts of Service Oriented Architecture (SOA), its approaches and comparison with other traditional approaches, and has described how the service communication takes place in SOA. Authors also state that how SOA can be incorporated in Business Process within an organization. Mohsen Moghaddam[2017] and other authors[2017] has done research to identify the technical and technological gap on business architectures and provides the recommendations for developing new reference models or architectures for Next-Gen (Generation) enterprises. Authors provide the study based on Industry 4.0 paradigms and architectures like IIRA, RAMI4.0, their characterization, limitations and mainly the evolution towards Modular Service Oriented Architectures. Liya Wu[2007] and other authors[2007] has described the Service Oriented Architecture for Business Intelligence domain which results in the seamless integration of technologies in a comprehensible BI environment. Authors provide the comparison of their ServiceOriented approach with other traditional BI architectures. Also, authors have implemented the prototype of their proposed solution which is SOA-ITPA (Service Oriented Architecture for IT Performance Analytic) and has elaborated some use-cases like adding data, delivering web application for multiple BI tools, and closed-loop as feedback for predicted information. Ramos Somya[2018] and other authors[2018] has implemented a reference model for Service-Oriented Business Intelligence (SoBI) and claims that if SoBI model is applied instead of traditional BI models, the ETL (Extract-Transform-Load) process which much more time-consuming process especially when data integration can be simplified with the help of web services. Authors have used academic and financial data from Satya Wacana Christian University (SWCU) for their case-study implementation. Martin Mikuf[2016] and other authors[2016] has implemented a case-study to set up web portals and create reports. Authors have used BI tools in Visual Studio from Microsoft and Cognos from IBM.

## 3. RESEARCH METHODOLOGY

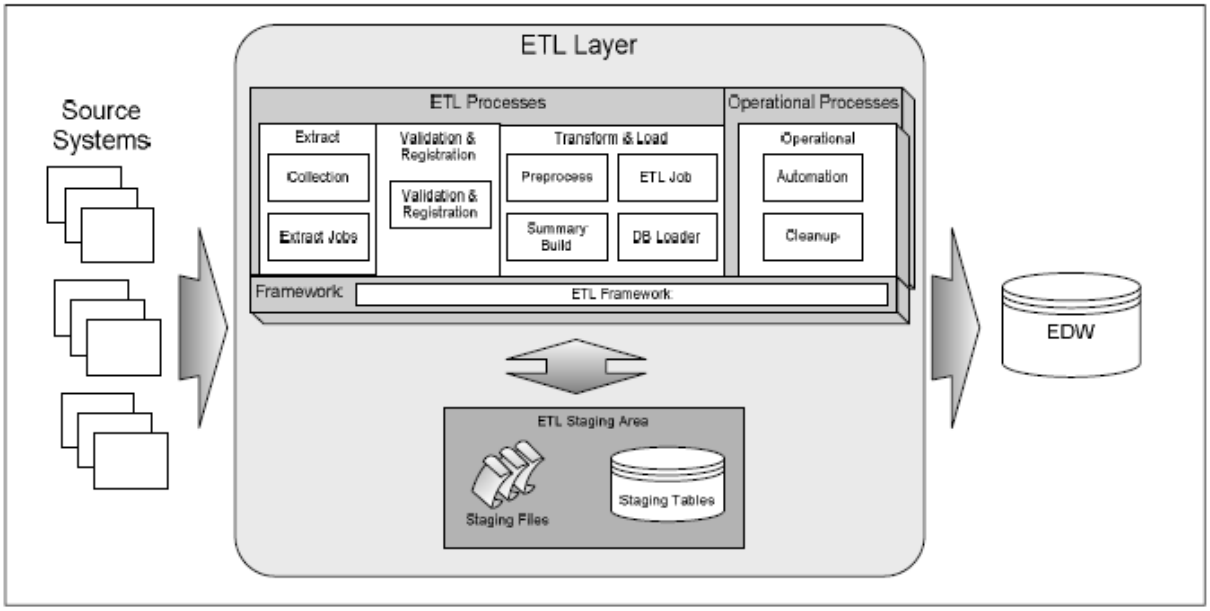
The following methods were used;

**Secondary Research:** Due to time constraints, it allows us to move closer to the target by examining the existing data by the company.

**Field-Based Research:** In order to better understand the nucleus of the synopsis we did very little field research as a question which is anonymous.

**Case Study:** Examining the purpose of the research synopsis to formulate a strategy. We see the industry as a case study.

**ETL architecture**



## Research Objective

1. To highlighting an investigation into the importance of data warehouses and business intelligence systems in the industry.
2. To design and develop data warehouse and business intelligence systems in the industry.
3. To evaluate how decision tools will assist the decision-maker in making better decisions about the company.

## 4. RESULT ANALYSIS & DISCUSSION

The expectation in this test set is that data returned from a data warehouse will be less time consuming than running the same query against a transactional database. The difference in query times is likely due to the essence of the organization of the data warehouse. The source system, or transactional database, is a typically large "normalized" database, which is "designed to organize data into stable structures, and thereby minimize update anomalies" (Sanders et al. al. 1). In the star schema used for this exercise, the data appears in an unusual state, which is typical in the design and implementation of the star schema organized to optimize the query. In this example, one of the star schema dimension tables produced by ETL uses a denormalization technique, which collapses the tables through a one-to-one join relationship, rather than requiring additional joins to complete the dimensions. Adds only required columns to the table. runtime. For example, one of the main dimension tables used for the Warehouse test set collapses five tables from the transactional system into one, using only the required columns from each for the query. Having data from one or more source tables organized in a logical dimension table will improve query results in the warehouse, with the columns readily available, completing the ETL table join after processing. By using a view in the source system that was developed to query the data for the report, it attempts to represent the data for the purpose of the report, but "since most DBMS products look up the definitions at run time, A view does not solve performance issues" (Sanders et al. 7). The desire to move away from the complexity of this or a similar view to be used as part of this test set may provide the ultimate motivation for users to use a data warehouse to accomplish reporting. It's worth noting that the ETL for this star schema runs for about two minutes as it executes the SQL that does the view, but it runs in separate sections for the dimension tables and then for the fact tables Is. There are two inherent disadvantages to having this data warehouse star schema, the time taken to load the data and the fact that the data will only be up-to-date as of the last refresh of the ETL. Because it takes an initial time investment to load the data warehouse, this is because running the refresh of the data warehouse requires the same amount of time as running the query, there's no real reason to do so. However, on a larger scale, if this data is to be queried frequently and reported throughout the day, an overall better performing system is a logical choice.

### Test set 1: Selecting All and Top

This set of queries selects all from the database, queries all records, then the top 200 to 200,000 records increasing by a factor of 20 each time. These sets of queries will test the database engines to tax them by selecting all, which is the most intensive query. After reviewing all the results in Figure 1.1, it is interesting to see the large difference in time between the view and the querying of the data warehouse.

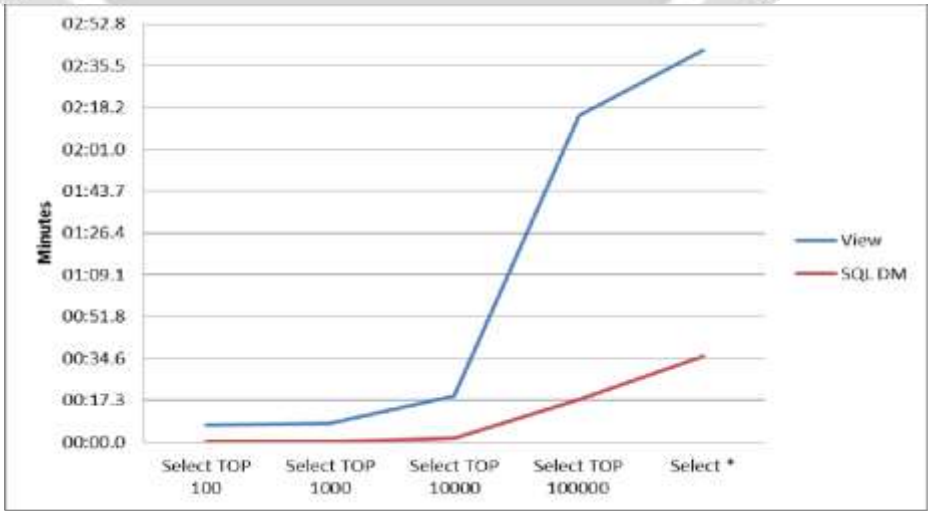



**Figure 1.1: Graphical results of selection of all questions**

**Test set 2: Selecting All Random**

This next set of queries selects all from the database, queries random records, then 1 to 200,000 records increasing by a factor of 10 each time. This time, the idea is to select random records instead of data vertex numbers. After reviewing all the results in Figure 1.2, it is once again clear that there remains a large time difference between querying the view and the data warehouse. One thing to note is that the numbers come a little closer in under two minutes, but that's not enough to say that the warehouse query isn't an improvement.

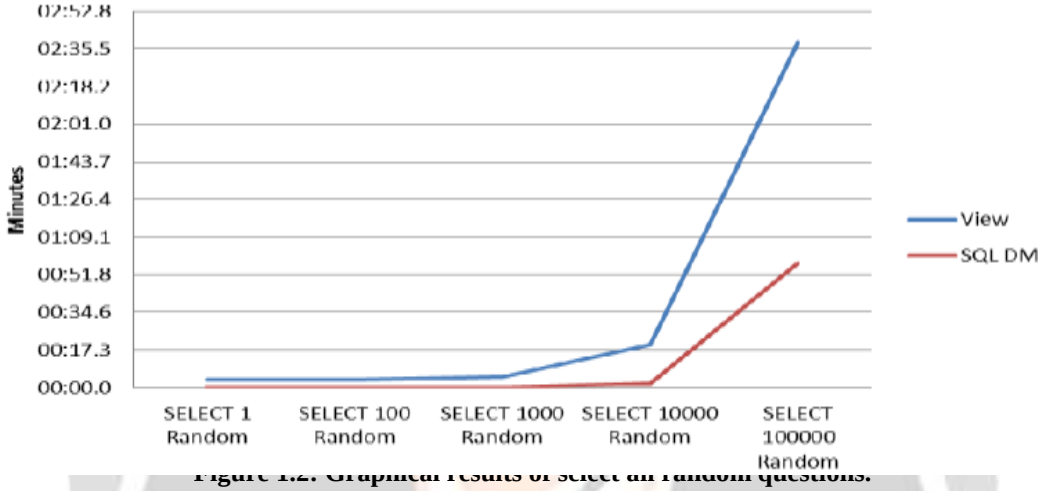



**Figure 1.2: Graphical results of select all random questions.**

**Test set 3: choose different**

The last set of queries select all but this time from the database using a where clause instead of all records. Although it was easier to query from the view because of all the data complexity, the related query on the data warehouse side would have a where clause using fields from the dimension table. Whatever the question, there is still a huge time gap between the questions, as shown in Figure 1.3.

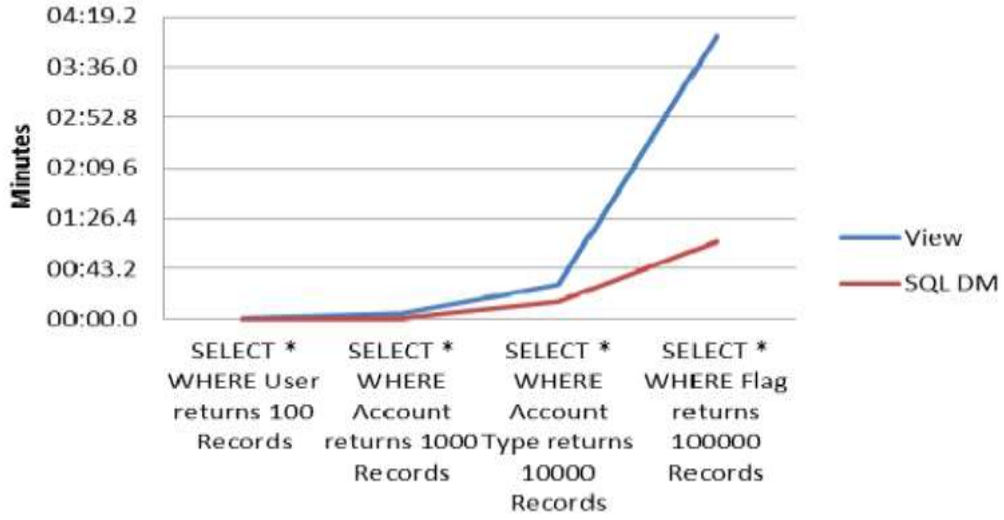



**Figure 1.3: Graphical results of Select All Where clause queries**

**CONCLUSION & FUTURE WORK**

Business users (decision-makers) lack the ability to access data easily when needed. In an effort to address this shortcoming, many departments within the retail company find their own resources, use the various data available and hire consultants to address their individual short-term data needs. In many cases, the same data was extracted

from the same source system by different departments to be accessed without a strategic overall information-delivery strategy. Management realized the negative impact on various sources of data as lack of integration on reports submitted by managers. Given the importance of information for the retail company, management was motivated to tackle the problem of data inconsistency by starting a central data warehouse and to ensure that data is available to all users regardless of their department. The need for data cohesion and consistent and quality reports gave rise to the data warehouse in the company.

Business needs and business strategies are needed for building data warehouse and business intelligence. A data warehouse is the foundation for powerful data analysis, it supports business decisions by encouraging company managers and other users to examine and better analyze data. Because the data is gathered in a repository, it can facilitate measuring the impact of various combinations such as supply chain, customer choice, demographics, geographies, and assist the analyst in customer retention processes and trend analysis Can do. Levers that a retailer can use to optimize performance include: price, promotion, markdown, classification, space, allocation and replenishment. Making data driven decisions is critical to successful decisions about all these levers. Competition in the retail sector is becoming increasingly fierce for global expansion, rapid product cycles, currency fluctuations and changing customer preferences. Therefore, Crystal Entertainment should be able to make strategic decisions that will affect the organization's position in the entertainment industry and stay on top of the game. The paper ends by studying some real studies, including building an enterprise data warehouse example, as well as studying a real case of improving queries. By designing and implementing a small-scale, small-sized enterprise data warehouse, it was easy to see how star schemas are assembled from multiple sources from a single transactional database to an organized and efficient ETL tool. As with practical exercise, we looked at running the same query against the same data, but in very different systems. When comparing the results of running the queries, it is clear in each example that data from a star schema organized in a data warehouse is returned dramatically faster than from a transactional database.

Such practical examples, as well as studies of data warehouse architecture and reporting capabilities, show the advantages of implementing an enterprise data warehouse for a better experience for users who are querying and reporting on data for better informed decisions.

**REFERENCES**


[1]. R. Van der Meulen and C. Pettey, "Gartner Says Self-Service Analytics and BI Users Will Produce More Analysis Than Data Scientists Will by 2019,"

[2]. Gartner, Inc, [Online]. Available: service-analytics-and-bi-users-will-producemore-analysis-than-data-scientists-will-by-2019. [Accessed 15 June 2019].

[3]. K. Božič and V. Dimovski, "Business intelligence and analytics for value creation: The role of absorptive capacity," International Journal of Information Management, vol. 46, pp. 93-103, 2019.

[4]. Microsoft, "Power BI | Interactive Data Visualization BI Tools," [Online]. Available: https://powerbi.microsoft.com/en-us/. [Accessed 12 6 2019].

[5]. [Tableau Software, "Business Intelligence and Analytics Software," [Online]. Available: https://www.tableau.com/. [Accessed 15 June 2019].

[6]. SAP SE, "SAP Lumira Discovery," [Online]. Available: https://saplumira.com/. [Accessed 15 June 2019].

[7]. SAP SE, "SAP Analytics Cloud | End-to-end Analytics for the Intelligent Enterprise | SAP," [Online]. Available: https://www.sapanalytics.cloud/. [Accessed 15 June 2019].

[8]. Qlik, "A Complete Business Intelligence (BI) Platform | Qlik," [Online]. Available: https://www.qlik.com/us/products. [Accessed 15 June 2019].

[9]. Gartner, Inc, "Best Business Intelligence and Analytics Software of [Online]. Available: https://www.gartner.com/reviews/customers-choice/analytics-business intelligenceplatforms. [Accessed 15 June 2019].

[10]. G2, Inc, "Best Business Intelligence Software," [Online]. Available: https://www.g2.com/categories/business-intelligence. [Accessed 15 June 2019].

[11]. Microsoft, "Power BI Desktop--Interactive Reports | Microsoft Power BI," [Online]. Available: https://powerbi.microsoft.com/en-us/desktop/. [Accessed 13 6 2019].

[12]. Miškuf, M. and Zolotová, I., 2015, January. Application of business intelligence solutions on manufacturing data. In Applied Machine Intelligence and Informatics (SAMI), 2015 IEEE 13th International Symposium on (pp. 193-197). IEEE.

[13]. Helland P. Data on the Outside Versus Data on the Inside. InCIDR 2005 Jan 4 (pp. 144-153).

[14]. Zadeh, A.T., Sahran, S. and Mukhtar, M., 2016. Automated Service Identification Methods: A Review. International Journal on Advanced Science, Engineering and Information Technology, 6(6), pp.1053-1059.
[15]. Mohammadi, M., 2017. SERVICE ORIENTED ARCHITECTURE: A REVIEW AND EVALUATION OF REFERENCE MODELS. International Journal on Information Technologies & Security, 9(4).
[16]. Florea, A.M.I., Diaconita, V. and Bologa, R., 2016. Data integration approaches using ETL. Database Systems Journal, 6(3), pp.19-27.