**KIET Group Of Institutions**

**Computer Science and Engineering Department**

**Implementing Business Intelligence on Pharmaceutical Data**

**US Pharma company BMS data for case study**

**By**

**Aashi Sharma**

**Table of Contents**

|  |  |
| --- | --- |
| **Content** | **Page number** |
| **Abstract**  **s** | **3** |
| **Introduction** | **3** |
| **Aim of the study** | **4** |
| **Significance of the Study** | **4** |
| **Questions of the study** | **4** |
| **literature review** | **4-5** |
| **Data of the Study** | **5** |
| **Methodology** | **5** |
| **Analysis and discussion** | **5+6+7+8+9** |
| **Conclusion** | **9** |
| **References** | **10** |

**Abstract**

Now-a-days, pharmaceutical industry is one of the fastest growing industries in the world. With increase in demand of pharmacy, it is beneficial for pharmaceutical organization to have a basic understanding of what they are how doing and how can their stats help them. We have made a system named C-SCAN where our pharmaceutical client can see the views, sales, growth and performance in their industry. Through this architecture, organizations would also be able to view what is going on in the pharma industry, they compare their growth with other organizations etc. In this fast-growing era, nobody has enough time to go and look out and analyze something in depth. Therefore, the information provided by C-SCAN in the form of reporting to our clients would make their life easier and help them to flourish in their domain.

Raw data is taken as input, and output are the formatted extracts with implied business rules.

We are using SQL, PostgreSQL, Python scripts, shell scripts, etc. in C-SCAN. The functionalities of the system can be updated as per the client specifications.

***I. Introduction***

It takes a huge amount for a manufacturer to manufacture a drug and get ‘Patent’. Once the drug loses its Patent after 15 years it goes ‘Generic’ in the market.To help the client earn the money back invested in manufacturing the drug and to gain maximum amount of profit from it in minimum time by gathering data for marketing and sales before the drug goes generic.

**Patent approval:** For a drug to get patent, it needs to go under some clinical trials.

* **Clinical Trial 0:** In this, drugs are tested on animals like rats, rabbits in different atmospheric conditions.
* **Clinical Trial 1:** Here, drugs are tested on humans like refuges, homeless people who opts for these trials for money.
* **Clinical Trial 2:** ‘Placebo’ effect comes into picture. Actual medicine and fake medicines are given to different people. If actual medicine cures more people, then it gets the patent. This trial ensures efficacy. It is also known as **‘Double Blind’**, because even doctors don’t know that fake medicines are given to the patients.
* **Clinical Trial 3:** In this, drugs are checked for side effects.

In the end, the final report of each ‘Clinical Trial’ is submitted to **‘FDA’** which approves the drug for selling it.

On the day of approval, medicine gets a ‘NDC’ number from ‘FDA’.

features.

***II. Aim of the study***

My main goal of this study is to provide solutions to our clients so that they can flourish in their domain with minimum of efforts and investment. To analyze the data for different clients coming from different vendors and to generate meaningful information after applying business rules and tools for better understanding.

.

***III.* *Significance of the Study:***

This study propelled business rules can be applied to pick up as much data from the current information. DQM checks can be applied and changed by the client's interest. Further improvements can be put forth in defense information format is relied upon to get changed in not so distant future to stay away from information disappointment and information bungle

***IV. Questions of the study***

The present paper seeks answers to these questions: what are the business intelligence rules that can be applied to the pharmaceutical data and how can we maintain the large Datawarehouse structure ? What are the solutions for these problems? What are the strategies used by various Pharmaceutical company to understand how their drug is performing in the market?

***V. literature review***

**Data Warehouse fixers:**In the information distribution center framework, the measurements are set of components which are progressively associated. Normally, they are exacting and covering with the goal that they bolster outlines at different degrees of granularity. On the off chance that all the rollup relations are capacities, at that point the measurement is said to be exacting. In the event that each component in a class is associated with its predecessor component, at that point the measurement is said to cover. This exploration paper shows the Data Warehouse Fixer which is a framework that reestablishes consistency of measurements. It underpins the measurements that doesn't fulfill the necessary severity imperatives. The manner in which this DWF works is that it checks the consistency of the framework and registers insignificant fixes for conflicting measurements. This is finished by executing Data log programs with powerless limitations. It additionally helps in fixing conflicting measurements.

**Mobility Data Warehouse:** With the wide accessibility of gadgets that can follow the situation of moving items, the curiousty in versatility information investigation has significantly expanded. Direction information warehousing techniques can be utilized for portability investigation. It typically incorporates portions of directions that are associated with spatial and non-spatial logical measurements. This examination paper goes past this idea. It incorporates the way of moving items anytime. By this technique, we can consolidate the moving item inquiries to the online expository handling (OLAP) questions, normally including conglomeration. The let us to communicate questions like "Locate the complete number of trucks which are running at the speed under 2 km from one another in the Antwerp" in a compact and exquisite way. Existing recommendations don't bolster comparable inquiries, as they depend on division of directions or pre-collection of measures.

***VI. Data of the Study:***

As a new client was being incorporated in the system we have prepared for managing the US pharmaceutical data on the large basis, It provided us the data in relation to drug, consumer (patient), doctor. The data was provided at different level such as country level data and also the territory level data. Various sources such as online click data, survey data etc also helped us understand the scope on a particular drug in the territory.

***VII .methodology:***

To give answers for our customers so they can prosper in their space with least of endeavors and venture. To examine the information for various customers originating from various merchants and to produce significant data in the wake of applying business rules and devices for better understanding.

***VIII. Analysis and discussion***

It takes a colossal sum for a maker to make a medication and get 'Patent'. When the medication loses its Patent following 15 years it goes 'Conventional' in the market. To enable the customer to win the cash back put resources into assembling the medication and to increase greatest measure of benefit from it in least time by social affair information for showcasing and deals before the medication goes nonexclusive.

1. **Data warehouse:**

Data warehousing is an innovation which summarizes organized information from different sources that helps in looking at and dissecting the information for greater business knowledge.

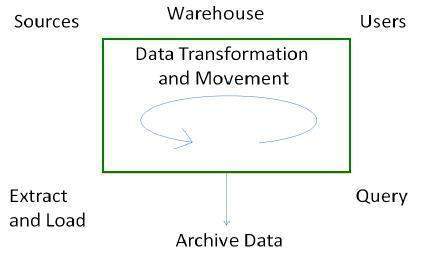
Data warehouses are somewhat unique in relation to standard operational databases as far as structure. The operational databases keep up severe exactness of information at the time by refreshing constant information quickly. Information stockrooms interestingly, are intended to give a long-extend perspective on information after some time

**1.1 Advantages:** Capacity to investigate information from different sources and to beat contrasts away composition utilizing the ETL process.

**1.2 Elements of Data Warehouse Tools and Utilities:**

Below are the functions of data warehouse tools and technologies −

* **Data Extraction** − Collection of information from numerous heterogeneous sources.
* **Data Cleaning** – It includes finding and adjusting the blunders in the data.
* **Data Transformation** – It changes over the information to distribution center arrangement from heritage design.
* **Data Loading** – This procedure includes summing up, arranging, merging, checking uprightness, and building segments and so forth.
* **Refreshing** – I It incorporates the refreshing of data into the distribution centers.



## 1.3 Extract and Load Process:

Data extraction implies taking the information from different source frameworks. Information load implies removing the information and stacking it into the data distribution center..

Prior to handling the data into the distribution center, the data must be recreated accumulated from the sources

### **1.4 ETL-process:**

**ETL (Extract, Transform and Load)** isa procedure liable for hauling the information out of all the source frameworks and embeddings it into an information distribution center. ETL includes the accompanying errands:  
  
extricating the data from source, information from different source frameworks is united into one information distribution center, which is prepared for change and handling.

**transforming the data includes:**

* applying business rules (realities and measurements),
* cleaning (mapping of NULL to 0 or 'Male' to 'M'),
* parceling a solitary segment into various sections
* joining the information originating from a few sources.
* applying any kind of information approval.

**loading the data**  stacking data into an data distribution center.

**1.5 Business Intelligence(BI**) - innovation for increasing most extreme measure of data from the accessible information for improving the business forms. It performs examination and gives sensible data that helps in making effective and high-quality business choices.

**2 AWS SERVICES USED:**

# **2.1: Amazon Redshift:**

Amazon Redshift is a completely overseen, quick and petabyte-scale information stockroom which makes it simpler and practical to investigate the whole information utilizing the current business knowledge apparatuses. It permits us to begin little without responsibilities, and scale to petabytes for less expense of conventional arrangements. Clients ordinarily experience multiple times pressure, hence altogether diminishing the expenses.

## Features and Benefits:

Amazon Redshift uses columnar storage technology that improves I/O efficiency and provides multi equal preparing over the hubs because of which it permits us to convey quick question execution. It has custom JDBC and ODBC drivers which permits us to utilize a more extensive scope of recognizable SQL customers. It likewise gives the choice of standard PostgreSQL JDBC and ODBC drivers. Speed of the information load shifts straightly with the bunch size, with reconciliations to Amazon S3, DynamoDB, Amazon Elastic MapReduce.

## Amazon Redshift's information distribution center engineering permits us to computerize the managerial assignments connected with designing, provisioning, and checking a cloud information stockroom. Reestablishes are extremely fast; we can begin questioning in a couple of seconds minutes while the information is handled down out of sight. Amazon S3 reinforcements are gradual, nonstop and programmed.

## More features and benefits includes:

### Fast and optimised

### Cheap

### SimpleFully Managed

### Automated Backups

### Secure

### **2.2 Amazon EC2:**

It is a web organization. It grants us to orchestrate figure limit with irrelevant scouring. It gives full oversight of enrolling resources and runs on Amazon's handling condition. It decreases the time taken to boot new server guides to minutes, allowing fast scale limit. It grants us to pay only for limit that we truly use.

Python scripts and shell scripts run on EC2 processes everything.

### **2.3 Amazon Relational Database Service (RDS):**

### It is a managed Relational Database Service. RDS makes it easier to set up, and operate a relational database in cloud. It is cost-efficient and provides resizable capacity in managing time-consuming administration tasks, allows us to focus on our business.

It provides access to multiple database engines, like Amazon Aurora, MySQL, PostgreSQL, Oracle etc. The code, tools, and applications that we already use can be used with Amazon RDS.

### **2.4 Amazon Aurora:**

Amazon Aurora is a MySQL and PostgreSQL perfect social database worked for the cloud. It joins the presentation of top of the line business databases with cost-adequacy of open source databases.

It is up to 5X quicker than standard MySQL databases and 3X quicker than standard PostgreSQL databases. It gives the accessibility, security and unwavering quality of business grade databases at 1/tenth of the expense. Aurora is completely overseen by Amazon Relational Database Service which mechanizes tedious assignments like equipment provisioning and reinforcements.

**AWS LAMBDA:**

AWS Lambda permits us to run the code without overseeing servers and provisioning. We pay just for the process time that we expend. Nothing is charged when the code isn't running. With the assistance of Lambda, we can run the code for any application or backend administration with no organization.

We simply need to transfer our code and Lambda takes care of everything required to execute the code. We can naturally trigger our code from different AWS administrations.

It is server less, that is, we don't need to deal with the server and is auto activated. Lambda is utilized for record development starting with one area then onto the next.

***IX. Conclusion***

The conclusion here comes out that the whole system developed is well working, The system provides solutions and also it is able to satisfy the need of the clients. The framework, engineering and the handling are tried very well for different use cases and mistakes are appropriately repaired to maintain a strategic distance from process disappointments. The principle point of the undertaking is satisfied effectively as it reports back to the clients or customers to separate significant data from the information for additional examination and making techniques

Subsequent to having various gatherings with the customer and the merchant, we went to an understanding about how the records will be conveyed to C-SCAN. We applied all the essential DQM checks and other quality checks to maintain a strategic distance from inconsistencies. Information types and the recorded limits of the documents were chosen to guarantee smooth information handling. Tables were made in C-SCAN for the equivalent with the goal that the information in the documents gets invigorated in them which can be utilized for investigation and different business prerequisites.

***X. References***

1. [www.zspace.com](http://www.zspace.com)
2. [www.amazonaws.cn](http://www.amazonaws.cn)
3. [www.datawarehouse4u.info](http://www.datawarehouse4u.info)
4. <http://zsservices.com/>
5. <https://www.researchgate.net/>