Store Analytics using ETL tool

[1]Prof. Dr.Amol Kadam , [2]Vaishanvi Kulkarni, [3]Shruti Chandge, [4]Ravi Yadav, [5]Pooja Yadav

[1]Assistant Professor, [2345]Final Year Students
Department of Computer Science and Business Systems,
Bharati Vidyapeeth (Deemed to be University) College of Engineering,
Pune, India.

Abstract: Business intelligence (BI) is the collection of processes, technologies, skills, and applications used to make informed, data-driven business decisions. BI includes data collection, data aggregation, analysis, and meaningful presentation that facilitates decision-making. Data-driven organizations use a variety of BI tools to access historical and real-time data in a data repository to perform queries, generate customized reports, and predict future trends. These tools include advanced analytics performed by trained data scientists as well as insights generated autonomously by machine learning algorithms. In this project we will be using Talend Open Studio which will provide us the power to Extract huge amount of data from different sources and integrate and Transform it to provide solutions to various business problems related to analytics. This will help various business organizations to know the market trends and make correct data driven decisions which will lead to a successful and growing business. In this project we will be performing analytics for retail stores which will also support e-commerce analytics.

Index Terms—ETL, Data Warehouse, Business Intelligence

INTRODUCTION

Business intelligence (BI) is a collection of processes, technologies, skills, and applications used to make informed, data-driven business decisions**.** Data-Driven Organizations use a variety of BI tools to access historical and real-time data in a data repository to perform queries, generate customized reports, and predict future trends. In this project, we will be using the Talend Open Studio Tool that will provide us the power to Extract a large amount of data from different sources and integrate and Transform it to give solutions to various business problems related to analytics. Most retailers use a front-end system for creating an invoice, maintaining daily data, customer information, etc. However, the information is lost on a daily/weekly basis as the data is not updated in the EDW. As historical data is lost, various analytics cannot be performed to support business decisions. In a few cases, front-end applications can provide you daily aggregated data, however, weekly/monthly aggregation is done manually which leads to errors in calculations. If a store has multiple invoice systems, it is hard to aggregate data across all the invoice systems in one go.

I. LITRETURE REVIEW

In computing, **extract, transform, load** (**ETL**) is the general procedure of copying data from one or more sources into a destination system which represents the data differently from the source(s) or in a different context than the source(s). The ETL process became a popular concept in the 1970s and is often used in data warehousing. Data extraction involves extracting data from homogeneous or heterogeneous sources; data transformation processes data by data cleaning and transforming it into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target database such as an operational data store, a data mart, data lake or a data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions. Since the data extraction takes time, it is common to execute the three phases in pipeline. While the data is being extracted, another transformation process executes while processing the data already received and prepares it for loading while the data loading begins without waiting for the completion of the previous phases.

ETL systems commonly integrate data from multiple applications (systems), typically developed and supported by different vendors or hosted on separate computer hardware. The separate systems containing the original data are frequently managed and operated by different employees. For example, a cost accounting system may combine data from payroll, sales, and purchasing.

Sr.no	Name of Author	Title of paper	Method used	Number of feature point extracted	Feature expression	advantages	Future work/ Limitations
1.	Sagar Bhujbal	A New Database For Retail Domain Using ETL System	. 1.Conceptual and logical modeling of ETL processes 2. Examining Problems that fall under ETL steps. 3. Research Prototype for inhouse projects.	3	1.Data Extraction , Transformation and Loading from heterogenous Sources. 2. Recurring Process to add new data in Warehouse. 3. Ease of regular ETL Modifications.	. 1.Availability of Historical Data 2. Data for Reporting that supports Business Decisions.	1.Multiple future predictions using historical data. 2. Evolution of Data Warehouse as per the market influence
2.	Saeed Rouhani	Business Intelligence Concepts and Approaches	1.Gathering Relevant Research Papers. 2. Extractions Of Classifications And Definitions. 3. Approach Classifications. 4. Analyzing Results.	4	1.Aid in controlling vast flow of bussiness information in the organization. 2. Managerial approach with focus on improving management decision making. 3. Technical approach by focusing on tools supporting the processes associated with intelligence in management approach. 4. Enabling approach by focusing on value-added capabilities in support of information.	1.Faster analysis. 2. Increased organizational efficiency. 3. Data-driven business decisions. 4. Increased competitive advantage.	1.Automation of solution approach for a business problem. 2. Improvements in BI algorithm.

II. METHODOLOGY

Agile is a project management methodology that uses short development cycles to prioritize continual improvement in the development of a product or service where requirements and solutions evolve through collaboration between self-organizing cross-functional teams.
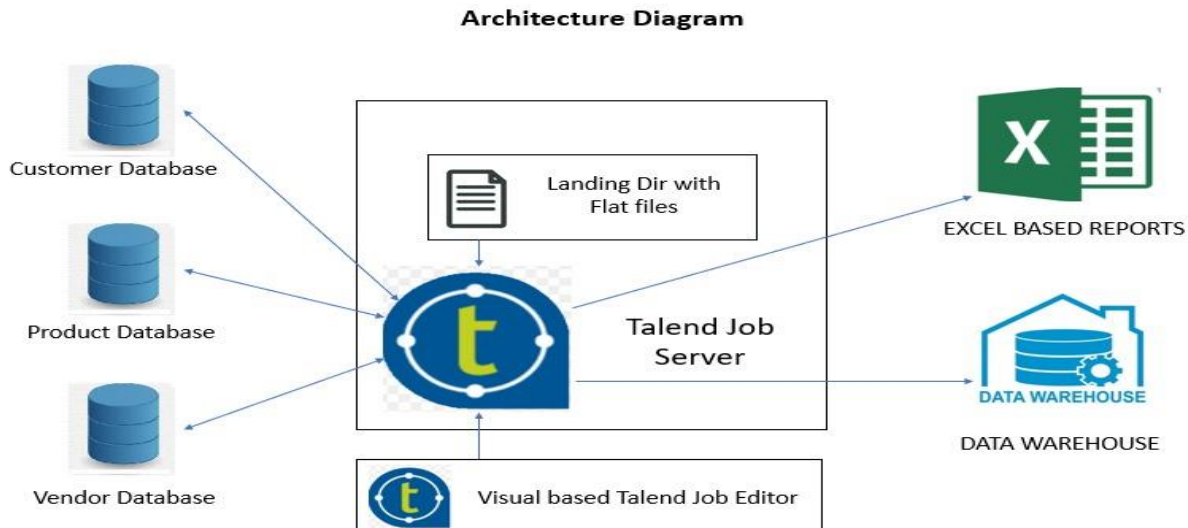
The fig 1 shows the system Architecture diagram.

Architecture Diagram

Fig 1. System Architecture Daigram

III. PROPOSED SYSTEM

To demonstrate an ETL project working we will be using 2 laptops out of which the first laptop will be used as a Talend job server and the second laptop will have the Talend open studio, SQL Developer, and Putty.

We have to make sure that both the laptops are connected through a network so that we can create jobs in Talend open studio, build it, and place the build on the Talend job server where the ETL job will be executed. In addition to that, we have installed MySQL Database on the Unix box (Talend job Server) from where we will load mock data for store management, Talend ETL job will read the data from the MySQL Database and Transform it As per business requirement.

We have taken ten different basic problem statements which can be easily solved through data analytics and ETL. Below are the scenarios:-

1. Aggregation of Transactional Data daily:-

Calculation of sales of all products in a large store is again a big problem however this can be achieved through aggregating the data of individual invoice machines and then summarizing them. However, this approach has many drawbacks:-

- For various analytics purposes, invoice data is not usable.
- There would be no central repository from where we can fetch data for all the invoices.
- This system is very time-consuming and could lead to several manual works.

So these drawbacks could be covered if the invoice machine is capable of writing data on Enterprise Data Warehouse and further could be aggregated through ETL Tools. For aggregation purposes, we can also use SQL. However, many Store procedures are not capable of writing these data onto Excel Sheets. Also, it is hard to read and maintain the store procedure.

2. Maintaining the Number of products and reordering them.

One of the fundamental requirements is to maintain and organize the product inventory. Usually, this is done by either manual effort or through a billing system. However, an automatic notification is not possible with this approach. Using our system, we can build an ETL process that generates alerts when a product's stock is running low, along with contact information for reordering purposes.

3. Check for products that are expiring.

With the help of ETL, we can scan the inventory database daily and generate a list of products that are approaching their expiration date.

We can also generate an excel based report daily which can help store managers to add promotions/discounts on specific products or return them to the vendor.

4. Setting Scope for promotional items.

There are various types of promotions applied on different products either by supplier/vendor or store manager. With ETL jobs we can mark those products as promotional products in the inventory database so that the invoice machine can catch the reduced price/quantity change.

5. Maintaining Customer Databases along with Loyal Customers.

We will maintain the customer database along with their contact number/ address with the Slowly Changing Dimension Technique(SCD). Some of the stores also offer certain promotional cards to their loyal customers using which they can redeem cheaper products or get cashback offers on their purchases. We can also maintain EDW for such customers.

6. Calculating Hourly wages of the store employees.

It is very tough to keep track of all the employees in a huge retail store. And to do it manually is a very time-consuming task and not an optimal solution. We can do this easily by creating and running an ETL job that would help us keep track of all the employees in the store.

7. Maintaining Pricing for loyal customers.

Commonly, small retailers keep the product price the same for everyone because they cannot differentiate between your regular customer and your once-in-a-lifetime customer. However, we can easily manage the price for your frequent customer, in addition, to providing them with occasional offers and discounts.

8. Maintaining Transactional Data from Online Payments.

Nowadays all major stores accept online payments however, due to some issues sometimes the payment gets delayed/duplicated, or canceled. So to maintain the history of online payments is important to cross verify and support better customer relationships. This problem can be resolved by creating a few ETL jobs which can maintain/aggregate online transactional data and load it into the tables. We can also create an application to fetch online transactional data for required customers to support them.

9. Maintaining Supplier's order summary/payments.

For a large store which maintains a huge number of products, they also have to keep track of the payments which they have to provide to the supplier. On several occasions, these things are maintained manually. However, with help of ETL jobs, we can create daily reports on various payments that are to be provided to the supplier.

10. Generating Excel Based Reports for Store Manager.

For all the above problems we can generate an excel based reports that provide the summary of all the given problems to the store manager. Such reports will help him to make data-driven decisions to properly assess and make the correct decisions.

IV. FUTURE SCOPE

A good ETL tool always helps the data processing with enhanced data connectivity and scalability. The tool should connect with different relational databases. Also, this tool should enable end-users to customize the data modules as per their business requirements.

Future ETL will be providing a data management framework – comprehensive and hybrid approach for managing big data. ETL solutions will encompass not only data integration but also data governance, data quality, and data security.

V. CONCLUSION

Under this project, we will practice various ETL components/codes which can read data from heterogeneous sources (Flat files, Tables) process them in a meaningful manner, and maintain them into EDW.

ETL model is being used by many companies for more than 30 years. Many companies read data from various sources, transform this extracted data using different techniques and then load it into the destination sources/systems. Though, some challenges to be faced while using/testing ETL tools, the ETL Tools are in use for many years. Companies use ETL to safely move their data from one system to another.

REFERENCES

1. 1.https://www.talend.com/resources/what-is-business-intelligence/https://www.datapine.com/blog/analytics-and-bi-projects-examples-templates/
2. https://www.lightsondata.com/6-must-have-goals-of-your-business-intelligence-program/
3. 3.https://www.projectpro.io/article/how-big-data-analysis-helped-increase-walmarts-sales-turnover/109
4. 4.https://www.retailsupplychaininsights.com/doc/kroger-wins-for-food-temperature-innovation-0001
5. 5.https://synoptek.com/insights/it-blogs/future-business-intelligence-trends/
6. https://medium.com/@sudarhtc/agile-project-management-methodology-manifesto-frameworks-and-process-f4c332ddb779