Data Warehousing in Food Delivery Service Industry

Sentiment Analysis in Pentaho Server

Azhar Ali Shaikh[1]

Bits Pilani, Dubai, UAE
h20200006@dubai.bits-pilani.ac.in

Abstract: The food industry is one of the world's largest and most important business sectors. The field encompasses everything from producers and shipping companies to retailers and restaurants. Food is nothing less than an essential part of life and a major global economic force. This paper highlights how sentiment analysis performed with the aid of data warehousing & Pentaho supports food industry businesses with critical decision-making capabilities in the areas of pricing, product promotion, product development, and demand forecasting via statistics generation.

Keywords: Data Warehousing, Sentiment Analysis, Food Industry, Pentaho, Statistics Generation.

1. Introduction

The food industry is one of the world's largest and most important business sectors. The field encompasses everything from producers and shipping companies to retailers and restaurants. Food is nothing less than an essential part of life and a major global economic force. Analysis performed with the aid of Data Warehousing supports food industry businesses with critical decision-making capabilities in the areas of pricing, product promotion, product development, and demand forecasting. Benefits include improved product innovation, greater sales effectiveness, enhanced margins and profitability levels, extended customer reach, increased marketing ROI, and greater customer satisfaction and loyalty.

For example, Deliveroo, the British online food delivery company, and disruptor, has chosen Snowflake's cloud data warehouse to support its exponential growth and expand its Deliveroo Editions offering for customers. Deliveroo has transformed the way people think about food delivery by offering fast, transparent and reliable delivery from over 750 premium restaurants using its network of 300 freelance drivers across London. Vital to its success is

the ability to quickly respond to customer demands. By using Snowflake's data warehouse, Deliveroo can now tap into the benefits of accessing data quickly and in real-time, while also helping to launch new restaurants in areas that are seeing a greater consumer demand.

2. Literature Survey

A data warehouse is the main repository of an organization's historical data. Data warehouse is optimized for reporting and analysis [1]. A growing number of large enterprises select the data warehouse to help their decision-making analysis. With a data warehouse, enterprises can understand the information of customers, business conditions, sales channels, make timely and effective decisions, thereby reducing operating costs, improving customer satisfaction, increased operating profits and expanding market share [2].

OLAP (Online Analytical Processing) tools offer the possibility of archiving, management, analysis, and multidimensional modelling. [3]. Before applying the Business Intelligence technique, data is processed and normalized to avoid data redundancy [4].Multidimensional Data Model can be used for creation of multiple data marts and design of an ETL process for populating the data marts from the data source [5]. The use of dynamic

ETL process using meta-data ETL is required when ETL process is dealing with the operational system and to address the increased requirement for report from the users [6]. Business Intelligence tools (BI tools) can be used to analyse large amounts of data [7]. A top-down approach makes it possible to elicit and consolidate user requirements and expectations [8] The star is composed of two kinds of the basic table: a fact table and dimension tables. The fact table includes operational transactions, or the analysis values wanted, and dimension tables include the description information related to these transactions and values.

The star schema exists widely in database application systems [9]. In an article [10], it was stated that Snowflake also accommodates Deliveroo's 650% growth in 2016. Such rapid momentum prompted Deliveroo to expand its business intelligence team from two employees to 14.
Additional team members triggered the need for more access to the same data but without impacting performance. Since Snowflake is built for the cloud, an unlimited number of users can access all of an organization's data from a single repository, which is critical to Deliveroo's success. There's no replicating data, shifting queries and other workloads to non-business hours, or queuing users to preserve performance. Instead, Snowflake's true cloud elasticity means Deliveroo can automatically scale up, down and out (concurrency) to load and analyze data without
disruption.

3. Data warehouse Architecture

The Data Warehouse Architecture consists of different parts, as shown in Figure 1. Extraction Transformation Loading (ETL), Data Warehouse, Data marts, and Business Intelligence (BI) tools. The data are collected from heterogeneous sources like the Operational system, ERP, CRM, and Flat files. ETL tools are used for extracting, transforming and loading data. The data are first stored in the staging Data warehouse. Once confirmed and verified the data is pushed from staging to production Data Warehouse. The data warehouse consists of Metadata, Summary data, and Raw data, this data will be in the form of DataMart. By using OLAP tools the data are presented and visualized. The same data is used for the mining process as well for future analysis, prediction, and automation.

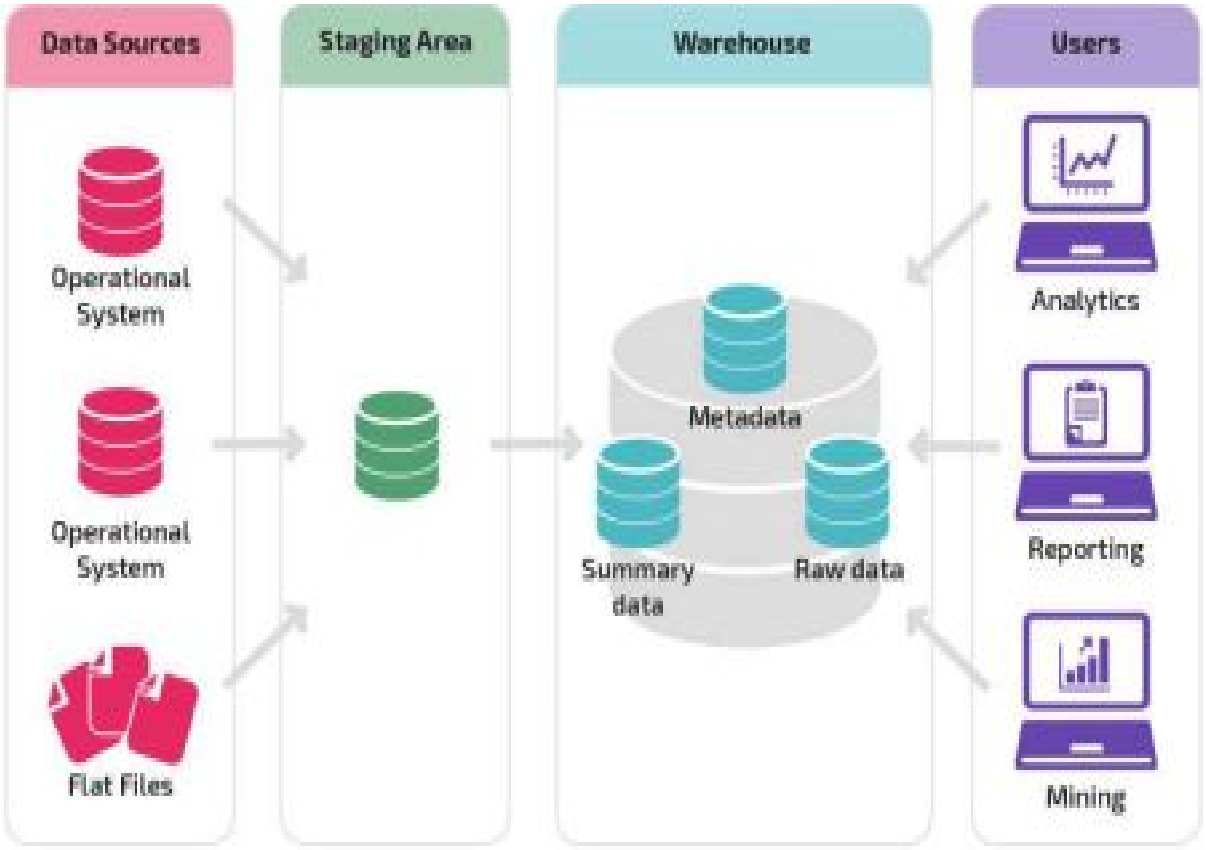


Figure 1: Data Warehouse Architecture

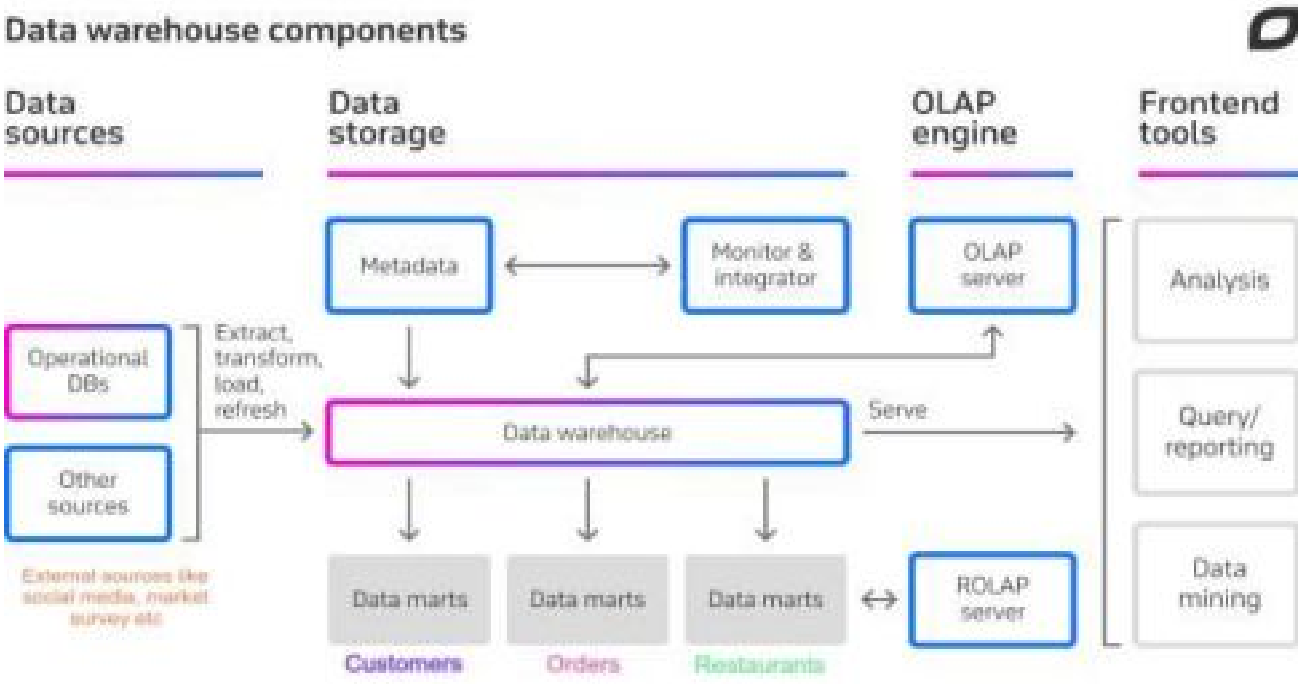


Figure 2: Data Warehouse Architecture for food delivery service

For designing a Food Service Delivery Data Warehouse, the Top-down approach for designing the data warehouse is preferred. The data acquired for this is details regarding the customer such as name, phone no., city, pin code, etc and details relating to orders as order-id, order-date,-order-amt, delivery-boy-assigned etc. The collected data once extracted is then cleansed into a format suitable for the warehouse. Various deletion, aggregation, summation techniques will be done on extracted data, and then loaded back to the data warehouse. After all rundown is finished, some more changes will apply to the data to make the information structure characterized by the data marts.

The source data from different sources are extracted and pushed to the food delivery service data warehouse. The data warehouse is built on a relational database management system (RDBMS), and star schema style is used to develop data warehouses and dimensional data marts. The source data from different sources are extracted and pushed to the food delivery service data warehouse. The data warehouse is built on a relational database management system (RDBMS), and star schema style is used to develop data warehouses and dimensional data marts. For Querying the data from the data warehouse database, the Multidimensional Expression query is used.

4. Schema Design

The Star schema has become a common term used to connote a dimensional model. Database designers have long used the term star schema to describe dimensional models because the resulting structure looks like a star and the logical diagram looks like the physical schema. The star model is the basic structure for a dimensional model. It typically has one large central table (called the fact table) and a set of smaller tables (called the dimension tables) arranged in a radial pattern around the fact table. It is a more powerful foe dealing with straightforward questions. Usually the truth tables in a star model are in the third normalized form(3NF), while dimensional tables are denormalized form.

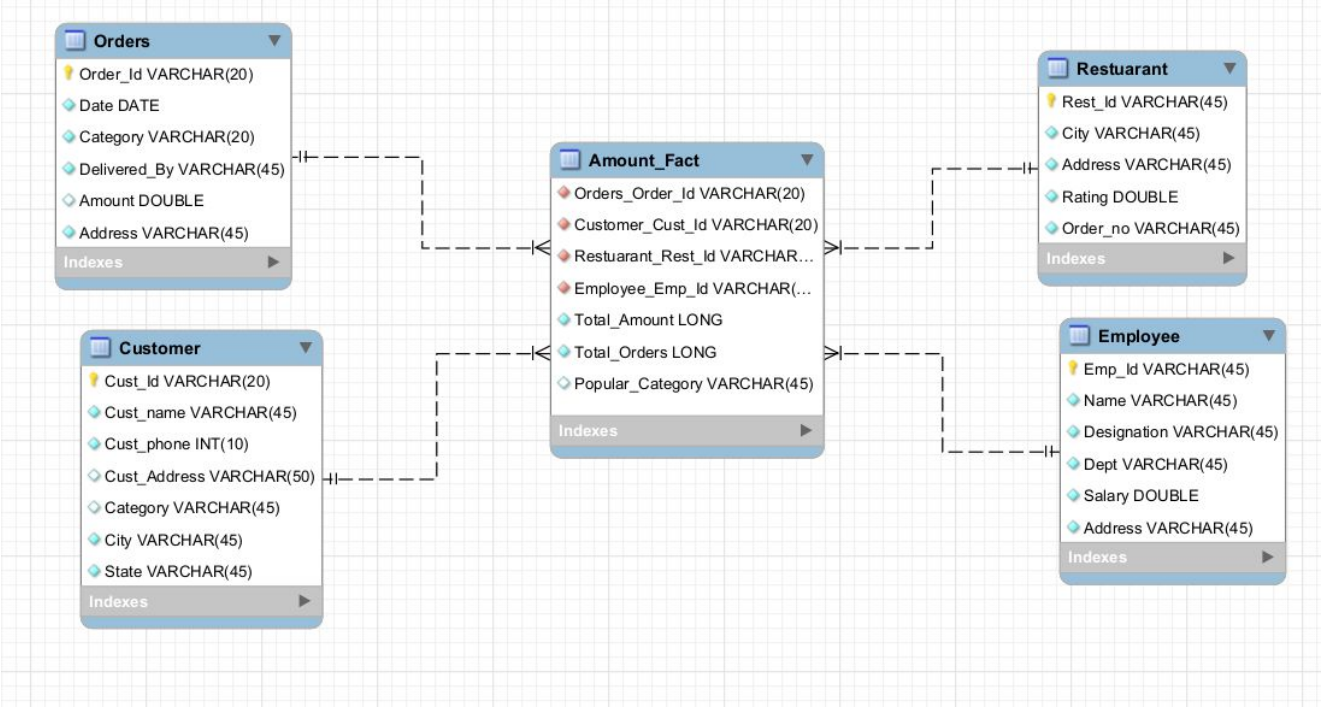


Figure 3: Star Schema for food delivery service data warehouse

The multidimensional database consists of dimensions and fact tables represented as a Star Schema. A fact table has two sorts of sections: foreign keys to measurement tables and measures those numeric attributes of a fact, representing the performance or behaviour of the food delivery service relative to the dimensions. A fact table can contain a fact's data on detail or gathered level. Dimensions are the parameters over which we want to perform Online Analytical Processing (OLAP) For example Time, Location, Customers, Salesperson etc. Figure 3 indicates one of the data marts in the food delivery service data warehouse. There's a table named Amount-fact & 4 dimension tables named Orders, Customer, Employee, and Restaurant respectively. The dimension tables hold specific data related to the subject chosen for example Customer holds customer details & the type (pro or regular customer) & the fact table references the dimension tables via foreign keys but hold

performance measure values as well like Total_Amount made by the food delivery service & Total_Orders it had for the month, quarter, year etc.

5. Experimental Results

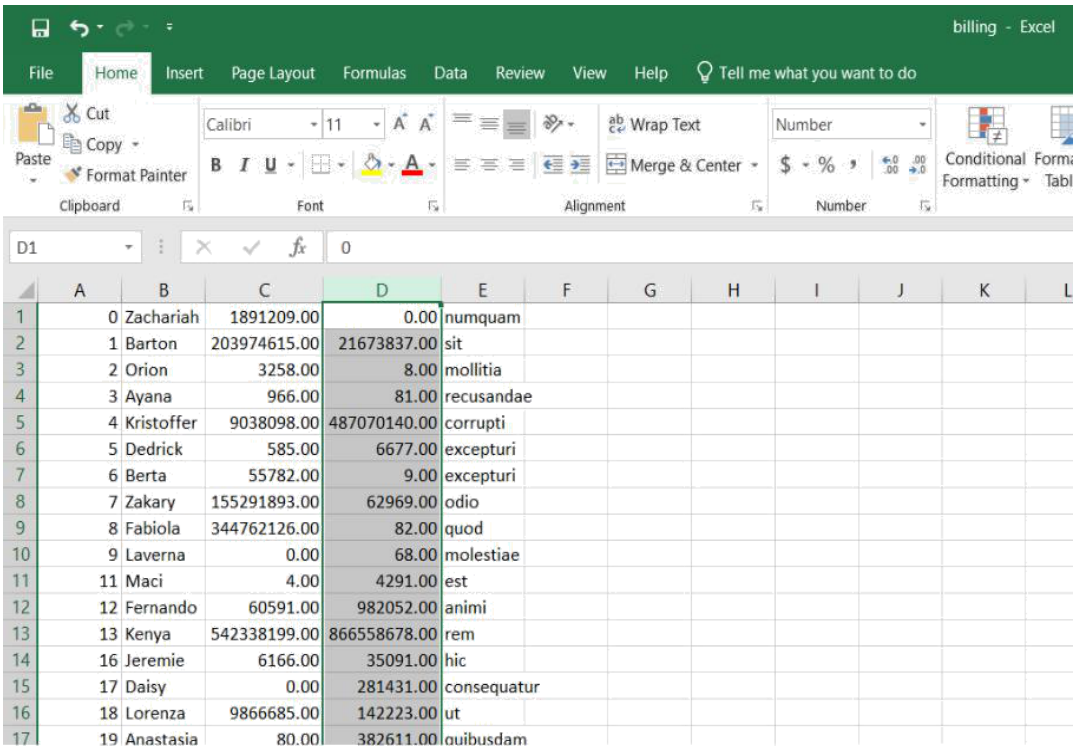
In the food delivery data warehouse each dimension table and the fact table have 1000 rows. Data is stored as a csv file as shown in figure 4. MySQL Workbench is used for loading all the data int the respective tables.

Pentaho Schema Workbench is utilized to create the data cube in figure 5. The data cube serves a purpose in Pentaho BI Server for reporting and BI's function as shown in figure 6.The figure 7 is an example of a MDX query which can be used for performing various operations in the cube based on the columns referred.

5. Conclusion & Future Scope

Initially, the restaurants had to buy ads in the papers, magazine, and/or distribute pamphlets to attract more business but only within their operating radius. With the development of food delivery apps, their reach & has grown drastically & so have the challenges. Various factors in the market lead to exponential increase in data which has to be processed & presented in such a way that management can make informed decisions as quickly as possible.

But even traditional data warehouses have some drawbacks and as stated above Deliveroo, in London has adapted to cloud based data warehouse which reduces extra hardware related costs & gives better performance in many aspects. Future cannot be predicted but with the right data & analysis tools preparations can be made to grab a bigger market share than before.



	A	B	C	D	E
1	0	Zachariah	1891209.00	0.00	numquam
2	1	Barton	203974615.00	21673837.00	sit
3	2	Orion	3258.00	8.00	mollitia
4	3	Ayana	966.00	81.00	recusandae
5	4	Kristoffer	9038098.00	487070140.00	corrupti
6	5	Dedrick	585.00	6677.00	excepturi
7	6	Berta	55782.00	9.00	excepturi
8	7	Zakary	155291893.00	62969.00	odio
9	8	Fabiola	344762126.00	82.00	quod
10	9	Laverna	0.00	68.00	molestiae
11	11	Maci	4.00	4291.00	est
12	12	Fernando	60591.00	982052.00	animi
13	13	Kenya	542338199.00	866558678.00	rem
14	16	Jeremie	6166.00	35091.00	hic
15	17	Daisy	0.00	281431.00	consequatur
16	18	Lorenza	9866685.00	142223.00	ut
17	19	Anastasia	80.00	382611.00	quibusdam

Figure 4: Customers dimension csv

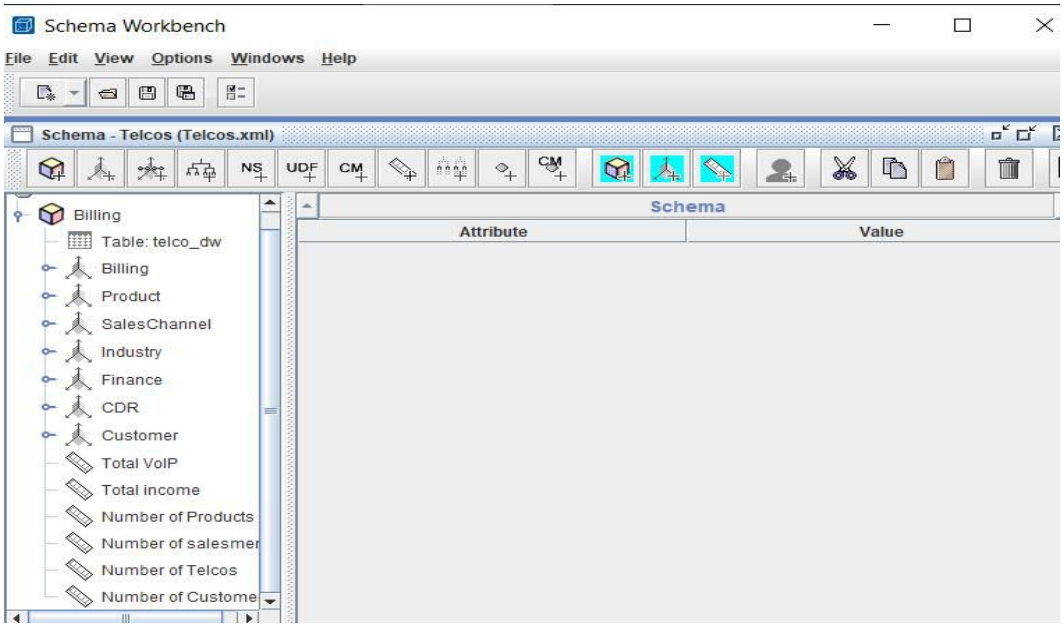


Figure 5: Schema workbench

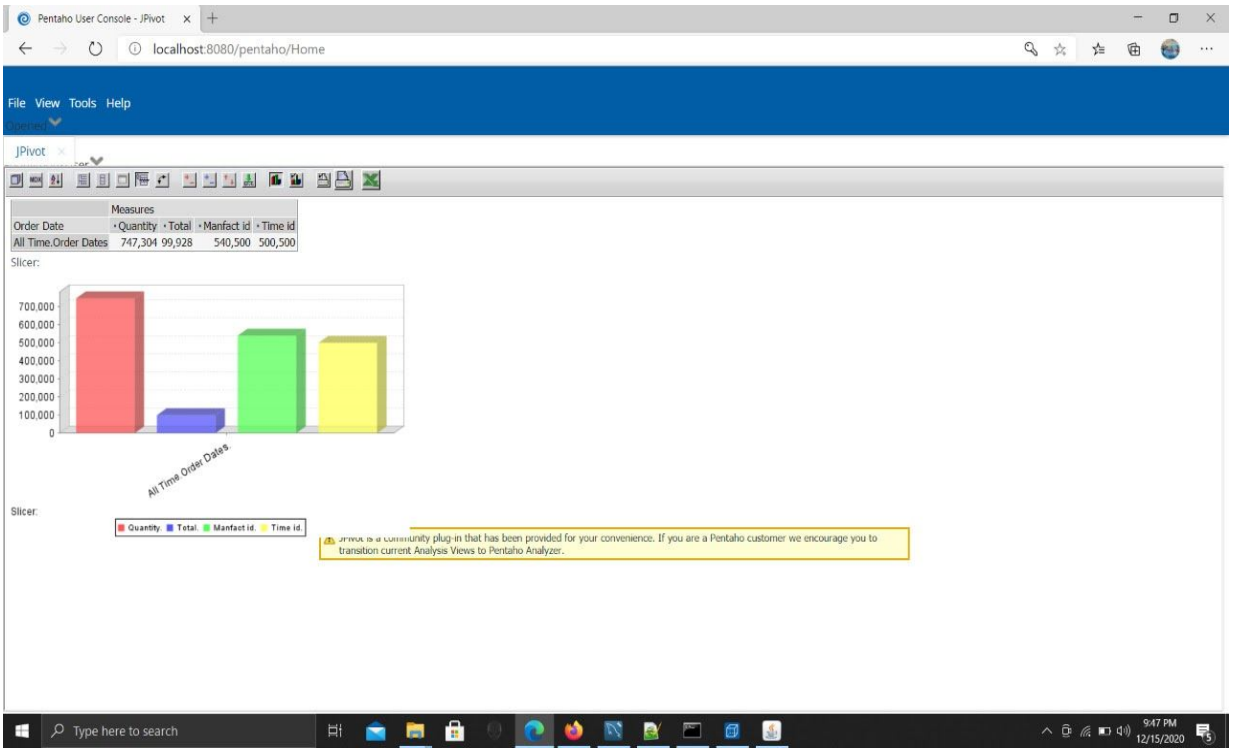


Figure 6: Total Orders

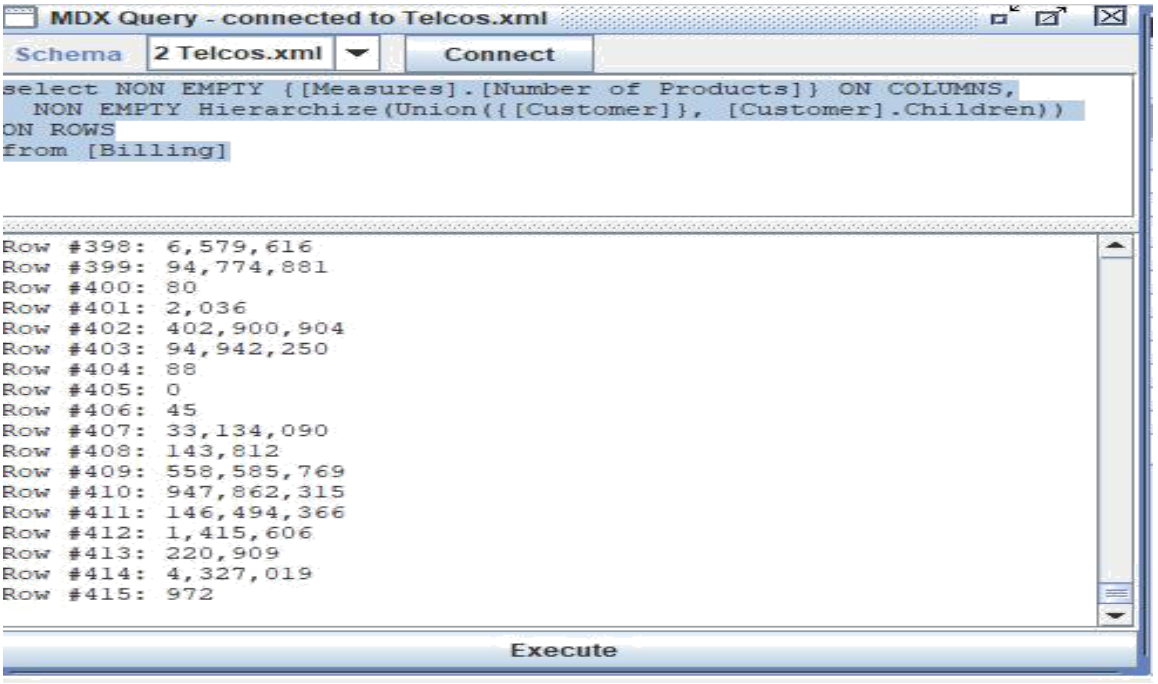


Figure 7: MDX query example

6. References

[1] M. Rifaie, K. Kianmehr, R. Alhajj and M. J. Ridley, "Data warehouse architecture and design," 2008 IEEE International Conference on Information Reuse and Integration, Las Vegas, NV, USA, 2008, pp. 58-63, doi: 10.1109/IRI.2008.4583005.

[2] Song Qiang and Lingxia Liu, "Research on the design of a new data warehouse system," 2009 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, 2009, pp. 462-465, doi: 10.1109/ICCSIT.2009.5234503.

[3] O. Boutkhoum, M. Hanine, A. Tikniouine and T. Agouti, "Integration approach of multicriteria analysis to OLAP systems: Multidimensional model," 2013 ACS International Conference on Computer Systems and Applications (AICCSA), Ifrane, 2013, pp. 1-4, doi: 10.1109/AICCSA.2013.6616464.

[4] Wang, J., & Kourik, J. L. (2015). Data Warehouse Snowflake Design and Performance Considerations in Business Analytics. Journal of Advances in Information Technology, 6(4), 212– 216. https://doi.org/10.12720/jait.6.4.212-216

[5] B. K. Seah and Nor Ezam Selan, "Design and implementation of data warehouse with data model using survey-based services data," Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014), Luton, 2014, pp. 58-64, doi:10.1109/INTECH.2014.6927748.

[6] R. Wijaya and B. Pudjoatmodjo, "An overview and implementation of extraction-transformation-loading (ETL) process in data warehouse (Case study: Department of agriculture)," 2015 3rd International Conference on Information and Communication Technology (ICoICT), Nusa Dua, 2015, pp. 70-74, doi: 10.1109/ICoICT.2015.7231399.

[7] O. Moscoso-Zea, Andres-Sampedro and S. Luján-Mora, "Datawarehouse design for educational data mining," 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET), Istanbul, 2016, pp. 1-6, doi: 10.1109/ITHET.2016.7760754.

[8] https://dl-acm-org-bpdc.new.knimbus.com/journal/tosem

[9] Zhang Chun-sheng and Zhuang Li-yan, "Association rules expand technology research based on attributes," 2013 IEEE 4th International Conference on Software Engineering and Service Science, Beijing, 2013, pp. 133-135, doi: 10.1109/ICSESS.2013.6615272.

[10] https://www.snowflake.com/blog/deliveroo-delivers-real-time-data/