Managament of Workload Plannig with Big Data by Using ETL Method

Yakup DELIBAS[1], Prof. Dr. Ali GÜNEŞ[2]

[1]Yakup DELİBAŞ, , Migros Ticaret A. S. General Headquarter, Information Technology and Business Development, Atatürk Mah. Turgut Ozal Bulvari, No: 7, 34758 Atasehir, Istanbul, Turkey İstanbul, Turkey

[2]Prof. Dr. Ali GÜNEŞ, , İstanbul Aydın University, Department of Computer Engineering, Florya Campus İstanbul, Turkey

Abstract: *In this study, first of all, information about large data and ETL (Extract, Transform, Load) system is given. Then a workforce planning program was introduced. Finally, the current troubles of the workforce program have been mentioned and the situations that have followed since the start of ETL system. The value of the big data has increased every day. The program, which is consistently made with big data, loses its update after a while and does not work at its first day's worth. For this reason, these programs need to be renewed in with the developing technology. For this reason, the workforce planning program has been combined with the ETL system so that it can keep up with both technology and better use.*

Keywords: *Big Data, Data Warehouse, ETL*

1. INTRODUCTION

The amount of data recorded with the progressive technology has increased. Technology development is like a device produced at first glance, but it is the technology that is important in the device which is actually important. All of these technologies is actually a collection of data. These datas are the ones that keep technology alive. Technologists feed on these data. For all these reasons, technologies should be developed in the section how to use such data.

To be able to use the accumulated data in the pool increases the authority of the firms so that firms take it out on other companies. All of these proficiency necessary to use resources effectively to win. In order to use resources effectively depends on the use of the accumulated data in the pool. If this data is used well, it would be effectively used at sources.

Large data methods are available to use the data. With this method, pooled data are better used and time saved. Companies can put themselves in the forefront by combining the technologies used with other technologies. Systems need to be able to make companies flexible and to gain features such as revenue growth, productivity and communication. We have begun to develop a planning program that works with big data logic by using the available means to add all these features to the company.

2. BIG DATA

The past information has been moved to the future. However, information from the past is limited. After 2000, data flow increased more. But all of the incremental data is not useful. This information is conceptualized as redundant information. Large data systems have been produced to use this obtained data correctly. Since the resulting data pool could not be managed with existing systems over time, projects started to solve this issue in 1970s (Mustafa Doğan, 2014).

In more technical terms, we should be able to analyze the data from different sources if we explain the bigger data.

There are data sources that make up this pool. These data sources are as follows.

a. Public data

The State holds is personal information of citizens.

b. Custom data

They are instructions that are kept for use by companies or individual persons. For instance, banks that hold transactions of customers in their database system could be a good example.

c. Environmental data

Every process that people use while using technology is described as environmental data. When we search the internet site, the search data is registered to the system and appropriate advertisements may accrue.

d. Community data

It is a collection of societies emerging from social environments. The ones written on social media are examples.

e. Personal data

It can happen with people's own movements. For example, the personal data that the stepper we have attached to our arm.

Big data consists of 5 groups. Big data components are as follows (Url-1). Figure 1 also illustrates the components way (Url-2)
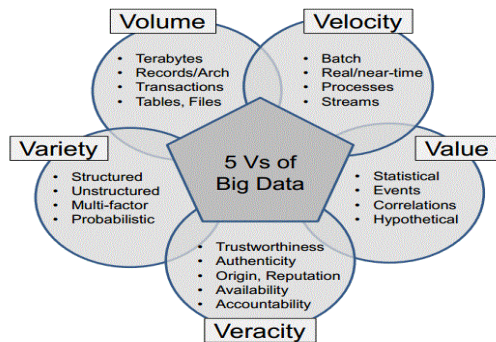

Figure 1: *The components of Big data*

a. Variety

Are the main sublayers. Is that a variety of incoming data. If the incoming data is varied, it will be much better analyzed and interpreted. For example, the richer the content of a customer's shopping in a store, the easier it will be to comment on it.

b. Volume

Data coming from the pool is increasing as technology improves. As the number of data coming into the pool increases, the size also increases. The amount of data in terabyte level in the past has increased and emerged into zettabyte levels nowadays. While the number of data in the world in 2000 was 800,000 petabytes, it is estimated that this data size will rise to 35 zetabytes in 2035. Figure 2 shows the amount of data in 1 minute on social media (Url-3)
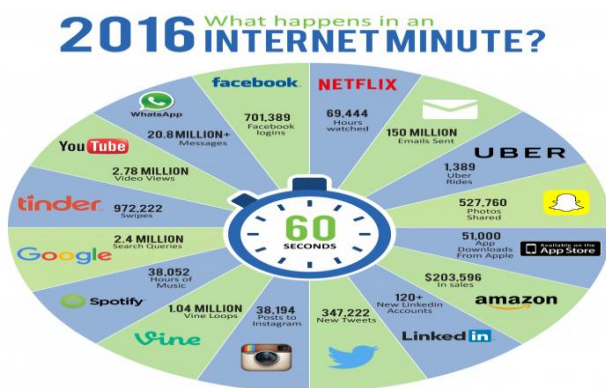

Figure 2: *What happens in 1 minute*

c. Velocity

In technology, speed is important. In large data systems, speed is the most valuable unit.

d. Value

Data in the data repository must be correct. If the incoming data are correct, the analysis would be valuable. The fact that these datas are right is important in company reliability. When we share this data with other companies, the reliability of the company decreases.

e. Veracity

Companies can archive their own data. Big data shed light on the future using this data. For this reason, every incoming data must be considered valuable.

Big data have both advantages and challenges. (McKinsey Global Institute, 2011)

Advantages

1. Making information more valuable by making it transparent and usable.

2. Companies can store more data and store it digitally so that they can reach every minute of the products and thus get more accurate and detailed information.

3. Personalized products and services are on the rise as the segment of the customer continues to decline.

4. Sophisticated analyzes make it easier to make decisions.

5. Used in new generation product development.

6. The use of large-volume data in different data types results in an increase in productivity.

Challenges

1. It should be done in such a way that companies understand the big data analysis. Companies will conduct their business according to these analyzes.

2. New technologies are constantly being developed. For this reason companies have to learn how to exploit them.

3. The need for IT specialists is high. In the US, there are 190,000 more analytical experts and more than 1.5 million data readers in need. This need can be decreased with either hiring new stuff or with the training of staff that they have.

4. Privacy and security are the most challenging part of large data. Obtained data must be protected.

Big data are used by the Health, Banking, Mobile, Traffic, Agriculture, Education, Government, Retail, and other sectors.

3. WORK FORCE MANAGEMENT SYSTEM

Technology emerges day by day in different fields. After the technology entered the retail sector, groceries and local markets were replaced by chain supermarkets. The customer should be satisfied in that there are too many customers in the chain markets. One of the most important solutions of this satisfaction is to be planned in stores. In the old system, store managers were planning the stores. However, in this method the store manager could not do enough analysis because he could not bring all the factors together, so the planning was only saving the day. For all these reasons, the workforce planning system has been put into effect.

The main purpose of this program is to calculate the density of the store by evaluating the daily, monthly sales of the Migros stores and then to calculate how the staff will use their permits to work at the time of day and hour and to make the Migros store use the work more efficiently. Stores can enter day-offs and work hours into this program.

The operational logic of the program is shown in figure 3. The program needs some information to plan. Planning can not be done without personnel data. Personnel data are also made up of other units. These are resources, needs and constraints. These needs in the general sense constitute past sales data.

Figure 3: *Operational Logic Diagram*

The stores first run the forecasting part by using past sales data. After the forecasting process is over, the information that needs to reach the target is displayed in front of the store. At the same time, it takes place when the number of personnel to work with is required to reach this goal. In this way, a list of needs will be

revealed in order to reach the target. After this phase, various constraints are entered into the system. These restrictions are legal, company rules, employee wishes, store restrictions. After all these transactions are transferred, it is necessary to transfer them to their sources at the end. These resources form the knowledge of the staff. Human resources and information from the store are collected at the source and the transfer is completed with various means. All these transactions are collected in the main data section of the personnel. Qualifications of the personnel are determined in the main data section of the personnel. It also tells you how many hours a day the staff will work each day and weekly.

After the above steps have been completed, the planning phase has been put in the process. The plan will be generated after the necessary operations are done through the program. If store wishes to change this plan, it can be done so.

Figure 4: *Personalized Planning View*

In Figure 4\ we see the final state of the Plan. It appears that the days of the day and the days on which the contacts need to work and at what time of day.

Figure 5: *Number of Actual Cashiers That Are Targeted At 15 Minute İntervals*

Figure 5 shows the time intervals with the 15-minute sections of the corresponding day. There is information on how many cashiers should work in the interval in the hours planned by the system to achieve the target.

The system allowed flexible working, therefore increased the productivity of the employees. Because of shifts according to store intensity measurement using

technical data, the case density is reduced. In addition to the increase in turnover and number of customers, customer satisfaction has also increased. Personnel are easier to track and time savings have been achieved.

4. DATA WAREHOUSE AND ETL

A data warehouse is a relational database created to analyze and query data. Data is coming from various sources. Data warehouses combine information from sources and enable their analysis to be separated from the rest of the system. Data warehouse is not called an application or product (Inmon, W. , H. , 2002). From the very first moment in the data warehouse, multiple technologies and applications are used. The data presented to the data warehouse has subject-based and homogeneous structure. The data are integral, persistent. Datas contain time information (Lane.P,)
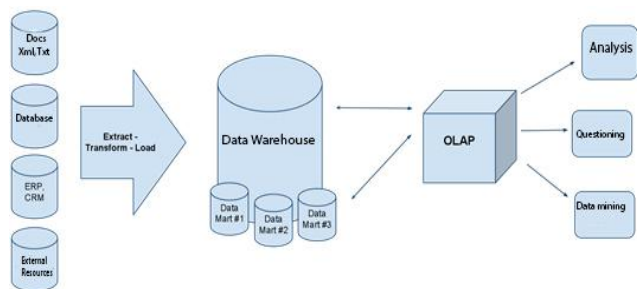

Figure 6: *Data Warehouse System Architecture*

The components present in the data warehouse are shown in Fig 6 In the figure the existing layers are as follows (Url-4)

Data sources: the main source of the data warehouse data. Each of these arriving data has its own resources and forms. The systems used by the largest companies of the data.

Data Interlayer (ETL): All data from sources is collected in this middleware. This middleware has its own tasks. These are Extract, Transform, Load.

Data Warehouse: (Warehouse): Data coming from the source is processed in the second layer and stored into third layer. This layer contains metadata, raw data and summary data. Users access this layer directly with the programs they use.

Data Presentation Layer: The layer in which the processed data is displayed. Users access data through this layer.

Data Access and Processing Applications Layer: Reporting, analysis and data mining studies are performed in this layer.
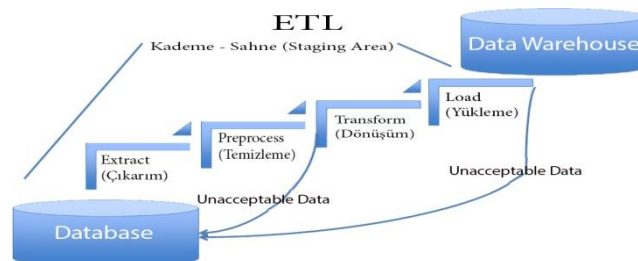

Figure 7: *ETL Processes*

Figure 7 shows the ETL processes (YBS Encyclopedia,12.2015) The data in the database are passed to the data warehouse after passing through Extract, Preprocess, Transformation and Load respectively. By examining all these processes in more detail (Ponniah.P)

i. Extract

The system tables necessary for the analysis of the data and the resources that are given are determined. The source data from the databases is moved to the data warehouse at this stage.

ii. Preprocess

It is the continuation of data extraction. In this section, the data is extracted. Tissue the data from the source and separate the damaged ones before sending them. Spelling mistakes, date mistakes, incomplete data, impossible data and repeated data are corrected or removed during the Pre Processing phase.

iii. Transformation

Once the data extraction is complete, the raw data is imported into the data warehouse. However, the raw data would not be clean and pure. For this reason, it is necessary to sort out and correct the data first. Also, the quality of the data must be increased before the data is transferred, and the data must be created in the data warehouse. In short, data is developed at this stage.

iv. Load

After the data from the source has been moved and converted later, it is ready to be transferred to the data warehouse. Ready data is loaded into the warehouse.

5. THE RELATIONSHIP BETWEEN WORK FORCE MANAGEMENT SYSTEM WITH ETL

By introducing a workforce management system, planning has started to be faster in stores. As a result, the number of customers and sales increased. There has also been a decline in the business intensity of the

store. Big data gave us an application in this context. However, it is necessary for such an application to make improvements in order not to have an error in the system. Over time, mistakes in the system have begun to give problems and the positive aspects have been wiped out and the problematic sides have begun to come to the forefront. Due to these reasons, a new development is needed through this system. The labor management system has suffered from the old conditions. These are as follows.

- The system error report feature is very limited. There is no explanatory information in case of error in the registration sent in the system. The existing report is for information purposes only and the information given is very limited. For this reason, the system needs a new reporting feature.

- Each running application leaves a log on the system. These logs cause disk to be full after a while. The system often filled the entire disc in a night so that all applications stopped when there was no room to write. For this reason, transfers can not be made.

- The speed of the workforce management system has recently begun to be questioned. The program needs to be faster. Transfers take several hours at certain times, which causes late changes to the system to be reflected in the system.

- For the system to be consistent, the data must be accurate and high quality. In addition, the data must be complete. Sometimes the data coming from the system is wrong. For this reason, the system falls into the fault as well.

Due to the problems that have arisen, it has been linked to ETL. As a result of the developments made, the structure has been changed. At the development of the ETL, the data coming from SAP started to be checked. Incorrect data was sent back when it occured. All processes are considered when forming the ETL structure. Various controls have been added to this structure to ensure maximum gain and best tracking. Since ETL controls the transfer, the error margin is very low. Certain gains have been obtained since the system has been running on ETL. With these achievements, the system became consistent. This consistency makes the system faster and more stable. In addition, the store's perspective on the system was positive. Constant systemic problems are reduced to minimal level. The benefits of transition to the ETL system is as follows.

- The applications running on the backplane were inflating the discs because they continuously leave logs on the system. Together with the ETL system, the delivery became more uniform. The unnecessary information does not flow into the system. Thus, the number of incoming data and the size of the resulting files are reduced.

- With ETL, file sizes are reduced and transfer times are shortened. In addition, the transfer has become ordered and organized since it is incorporated into a system. Another situation was that before the ETL system, background applications were working on an hourly basis. Every application was working when their time arrives. It was starting to work immediately without checking that if the other program had finished. With ETL, this problem is solved. The ETL system has restructured the system and fixed the issue of working according to hourly schedule. Instead, it uses the run-by-run sequence. The ETL system does not start the other programs until the end according to the order of the respective programs.

- In the old system, the latest state of the programs could not be tracked. Along with ETL, all the fields in the ETL system are visible.

- The ETL system follows the consistency of the transfer. The ETL system works all day except the transfer. A kind of monitoring is done by working all day. If the transfer was blocked in some way, the system will resume where it left off.

- The quality of the data stored in the system from large data programs means that the system works more smoothly. For this reason, the data in the system should be accurate and high quality. The old system did not have the necessary controls. All the data coming from the SAP was taken in. With the ETL, the checkpoints have been structured to control the incoming data from the SAP.

6. CONCLUSIONS

As technology develops, companies keep pace with it. Companies want to take full advantage of technology. However, it is important to be able to use this technology, not just to acquire it. They have to adapt the ready technology to their own companies and make improvements according to the situation. Due to the rescue features of the big data system, the density of the companies is reduced. When these systems are used correctly, companies save both time and profit

from the data. The transport structure along with the ETL system has come out of pieces and has begun to work as a whole. A detailed control system has been introduced. It started working in a sequential manner. It started working without human intervention. Data has begun to filter. With all these features, the program has added the following innovations.

- It has become faster and time-saving.

- Performance has improved.

- A detailed review of the transmission process has begun.

- The possibility of quality and correct data has been enabled.

- It started to be managed and intervened in a single structure.

- Excessive strain applied to the system by the program has been reduced and the system has been able to work more stable.

- Problems caused by human errors have been minimized.

- Warning system developed.

- The control system has developed and forced the system to comply with the rules. In short, it was putting the traffic lights on the system and opening the road according to the situation, the transfer was made progress.

- The number of errors has decreased and the returns of stores have decreased. Thus, the reliability of the program has increased.

- Continuous control shows that the transmission is in a state of failure and that the sending personnel have received errors during transmission.

REFERENCES

Doğan Mustafa, 2014, Büyük Veri'nin Kişiler Ve Kurumlar Üzerindeki Etkileri

Inmon, W. , H. , 2002 Building the Data Warehouse, s. 31-34, 113, Wiley Computer

Publishing, USA,.

Lane, P., Oracle9i Data Warehousing Guide Release 2 (9. 2)

McKinsey Global Institute, 2011, Big data: The next frontier for innovation, competition, and productivity

Ponniah, P, 2001 Data Warehouse Fundamentals / A Comprehensive Guide

for IT Professionals, s. 36, 64, 67, 70, 118, 193, 196, 262-263, 272-273, 279, 281, Wiley Publications, USA.

YBS Ansiklopedi , 12. 2015 Cilt 2, Sayı 4,S. 6-13

INTERNET REFERENCES

Url-1<www.sas.com/en_us/insights/big-data/what-is-big-data.html> Last access date: 02.2017

Url2<http://andressilvaa.tumblr.com/post/87206443764/big-data-refers-to-5vs-volume> Last access date: 02.2017

Url3<http://www.excelacom.com/resources/blog/2016-update-what-happens-in-one-internet-minute> Last access date: 02.2017

Url4<http://www.kadirbekiroglu.com/component/content/article/1-son-haberler/173-ETL-extract-transform-load-nedir html> Last access date: 02.2017