International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

Cloud Technology As A Support For The ETL Process And Its Influence On Decision Making.

Frank Escobedo-Bailon[1], Antonio Arqque-Pantigozo[2], Carlos Alzamora-Aragon[3], Blanca Pasco-Barriga[4], Soledad Olivares-Zegarra[5], Katherin Rodriguez-Zevallos[6]

[1,2,5]National Technological University of South Lima, Lima, Perú
[3,4]Private University of the North, Lima, Perú
[6]National University of Huancavelica, Huancavelica, Perú

[1]fescobedo@untels.edu.pe, https://orcid.org/0000-0002-2058-0976
[2]aarqque@untels.edu.pe, https://orcid.org/0000-0003-1107-434X
[3]carlos.alzamora@upn.pe, https://orcid.org/0000-0003-0335-259X
[4]blanca.pasco@upn.edu.pe, https://orcid.org/0000-0001-7264-2393
[5]solivares@untels.edu.pe, https://orcid.org/0000-0002-9932-1036
[6]katherin.rodriguez@unh.edu.pe, https://orcid.org/0000-0001-8350-720X

Abstract : There are many important points in the digital transformation of organizations today that decisions must be accurate, and the data warehouse ETL process and its underlying concepts help interpret them and enhance corporate interpretive strategic planning. Purpose: The objective is to understand the importance of the impact of the ETL process on decision making. Materials and methods: The methodological cut-off points between two methods (Kimball, Immon) by different methods depend on the requirements and benefits of system recovery. A literature review on the subject was conducted to search and reference scientific articles. Development: The ETL process works in the same way as information is reconstructed and made accessible. For this reason, new data sources and models also require specific validation based on the business rules of each organization. Considering the ETL process associated with modern business intelligence, we can observe the different applications of these concepts in a variety of disciplines, from mathematical and scientific rules to organizational business rules. The end result: you can get a solid partnership that makes decisions more realistic and transformational when running BI, and gives you new metrics, such as analyzed data. keyword. ETL, Kimball's method, Power BI, data. Different applications of these concepts can be seen in fields as diverse as mathematical sciences to organization-specific business rules. Conclusion: business intelligence makes decisions more realistic and offers strong partnerships, including new metrics that can be acquired, such as data transformation and data analysis. keywords. ETL, Kimball Methodology, Power BI, Data. Different applications of these concepts can be observed in fields as diverse as mathematical sciences or organization-specific business rules. The bottom line: Doing your business intelligently makes your decisions more realistic and gives you solid associations as new metrics you can retrieve, such as transformed and analyzed data. keyword. ETL, Kimball methods, Power BI, data.

Keywords: ETL, Kimball Methodology, Power BI, Data

1. INTRODUCTION

We are currently facing the process of a digital revolution in the field of business and individuals. The creation, storage and transmission of information is fundamental to the functioning of the financial sector. The information element records the customer's monetary rights and obligations. The flow of information enables the payment and exchange of assets. Information analysis will help you assess the risk of a potential loan. These are just a few examples of the extent to which the financial industry has depended on information processing since its inception. This explains why the financial industry has historically been one of the largest investors in information and communication technologies (ICT). (Fernandes Derris, 2017)

In the latter case, we learn more and more through the use of skills and tools that make our lives easier and, little by little, we use our empirical knowledge appropriately. Organizationally, it is a bit complicated because people with a culture of change talk about groups of people coming to the same place, company goals, missions and visions.

This is the digital transformation of companies like devices, computers, Whether you use servers, social networks or transactional systems, create a digital culture in your organization that influences the development of your business strategy. The power of an organization is decision making. For example, depending on the business area we are talking about, you could be an administrator. So how can these decisions be strengthened beyond the use of real company data and information, real numbers and indexes? It sounds simple, but in reality every business is the world, i.e., every system in every business is the world of data. This is where the ETL process is generated and helps us a lot by looking at the data and information in the business system from another angle. In addition to helping with these steps, there are also tools to aggregate visual interpretations from multiple sources. This article describes a bit about transaction modeling and alternative decision-making approaches that use the ETL process as an independent variable.

2. METHODOLOGIES

2.1. Kimball Methodology

Ralph Kimball pioneered these approaches, and Kimball's data model implements a bottom-up approach to DW (Data Warehouse Architecture) design, where the data mart was first conceived, according to business needs. Then evaluate the underlying data source and use the ETL (Extract Transform, Load) tool to take different types of data types from different sources and load them into the relational data preparation area of the machine. Following Kimball's bottom-up approach, after the data is loaded into the data warehouse staging area, the next step is to load the data into a data warehouse model with a denormalized baseline size. This model divides the data into either a fact table, which is numeric transaction data, or a dimension table, which is the reference information that supports the data. (Tehreem Naeem, 2021)

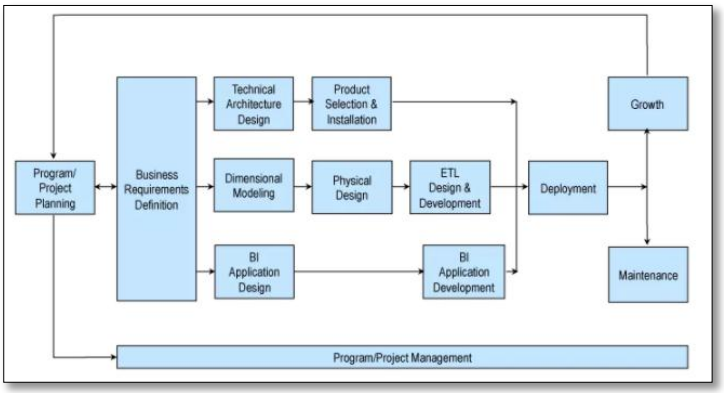

Figure 1: Kimball's methodology for the data warehouse life cycle. Source: (Tehreem Naeem, 2021)

2.2. Immon's methodology

Bill Inmon, the father of data warehousing, introduced the concept of data warehouse development. He started by designing a data model for a company's data warehouse that defines the key areas and entities of the associated companies, such as customers, products and suppliers. Thus. Bill Inmon's definition of a data warehouse is "thematic, integrated, non-volatile data collection to support management decision making." The model then creates a complete logical model for each key entity. For example, a logical model is created for a product that has all the attributes associated with that entity. This logic model can contain 10 different entities in a product, including all the details, such as business drivers, aspects, relationships, dependencies, associations and more. (Tehreem Naeem, 2021)

The Kimball option is the most successful because you are trying to organize information about your personal information requirements. The art of cloud allows you to implement or deploy the service provider. In this case, all methods of secure growth are carried out with the help of the cloud. The instrument provided varied and at the same time gives it once with the roots of the other. Clouds are a straight way to send computing resources and some software is how to save saved data, how to save saved data how to save data to save data and function. Users will continue to update monthly updates for maximum employment and security. (Tehreem Naeem, 2021)

3. DATA MODELING

Database design is not a simple process. The complexity of the information and the number of information system requirements complicate the information. For this reason, it is very useful to adopt a known partitioning strategy when designing your database. Therefore, the design process should be divided into several phases. In any case, the intermediate results obtained serve as a starting point for the next step, and the desired results are obtained in the final step. As you can see, you do not have to solve all design problems, but only face one type of subproblem at each step. This simplifies the process when the problems are split up. Most data modeling methods provide a way to graphically represent features and relationships. (Dominguez, 2018)

3.1. Transactional data modeling.

This data model is a collection of information that implements the search, insert, update, and delete procedures as needed in an organization's business rules-based system. Optimization is a very good pillar of this model because the entire system is model-based and we are constantly looking for ways to optimize our research resources.

Should this analysis be done in an orderly and systematic way, first detailing what you are trying to analyze? Regardless of the nature of the information contained in a database, it can always be subdivided or classified into "information domains". When we talk about information domains, we mean that not all information has the same purpose, but can be divided into business domains that can be analyzed independently. (Troche, 2014)

3.2. Dimensional data modeling

Dimensional modeling aims to derive an interpretation based on the dimensions and measures defined by the organization. Therefore, it may not need to go through the standardization process. Implementations of this model are not intended to provide or operate APIs, mobile applications, ERP or CRM systems. It is limited to studying the information and collecting indicators from it.

4. PREVIOUS CONCEPTS

4.1. Microsoft Power BI

Integrated program for business intelligence, which allows the integration of multiple data sources of different origins and formats.

It allows the use of remote and local databases, text files, .xlsx files, among others.

It also has many tools to perform ETL processes and also data transformation, to later generate reports in a very simple way.

Among the most relevant are:

- Power Query.
- Data flows.
- Data sets.

4.2. SQL Language

It is a structured query language based on calculations and allows insertion, deletion, update and retrieval operations.

It also allows us to create entities within our manager or database engine in use, of which the most relevant are:

- SQL Server
- Oracle
- PostgreSQL
- Mysql

4.3. ETL Process

This is an acronym that includes the extraction, conversion and downloading of information for a variety of purposes. As you can see, this is a fundamental part of business intelligence. This is essentially a bridge to other data models, because it can be better interpreted or already targeted. There are many ways to create it, including cloud desktop tools from various vendors such as Azure, Amazon and Google Cloud Platform.

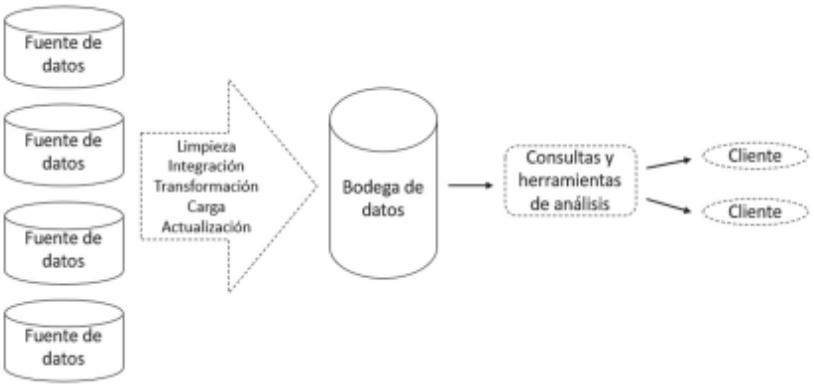

Figure 2: Flow of an ETL process associated with a datawarehouse

4.4. Cloud computing

Cloud computing is an efficient way of providing computing resources and is a form of outsourcing, such as data storage and processing. Users who want to purchase it pay a monthly fee, the software is managed by the provider and is constantly updated for maximum performance and security. (Salesforce, 2019)

4.5. Platform as a service (PaaS).

Platform-as-a-Service (PaaS) is a complete cloud-based development and deployment environment that can deliver everything from simple cloud-based applications to complex cloud-based enterprise applications. When you buy the resources you need from your cloud service provider, you access them through a secure Internet connection, but you only pay for what you use. (Azure, 2021)

4.6. Infrastructure as a Service (IaaS)

IaaS providers provide virtualization, storage, networking and servers. As a result, users do not have to have an on-site data center and do not have to worry about physically upgrading or maintaining these elements. In most cases, IaaS users have full control over their infrastructure through an application programming interface (API) or a control panel. IaaS is the most flexible cloud-as-a-service model that allows you to easily scale, upgrade and integrate resources such as cloud storage without having to anticipate or cope with new demands. This can generate upfront costs in the future. (RedHat, n.d.)

4.7. Software as a Service (SaaS)

This is a cloud computing service that allows users to access software in the provider's cloud. Users do not install applications on their local devices. Instead, the application resides on a remote cloud network that can be accessed via the web or API. The application allows users to store and analyze data and collaborate on projects. (IBM, 2020)

4.8. Data loss or gain?

When we talk about data transformation, we are referring to generating more information for purposes other than revenue generation. For example, a single commercial campaign or product over a period of time. You can create factors to know what kind of product you are offering to make decisions about the work.

This may be interpreted as data gain, but it is not. Data or other sources are not changed for this analysis. Maybe in terms of commercial systems you generate a lot of data every day. But in terms of dimensional modeling, there is not. As mentioned earlier, you get information in your hands. Don't forget that this is data interpreted by information.

Therefore, the amount of data continues to grow and become more accurate and continues to be understood as you can get more information in different situations. The reliability of processes and programs is not just based on words, but there are many ways to perform verification.

In this case, it is information and the source and destination are readily available. Duque (2016) shows that to confirm the behavior of the proposed ETL model, it is necessary to perform a management of measured data sets in each of the different stations. In this case, he shows the difference.

The source of the data source. Kimball's methodology is a benchmark for many smart business enterprises. Medina (2018) states that four steps were used to define the methodology used in the study: analysis, multidimensional design, ETL procedures and analytical processing.

(Mojica, 2020) teaches that they have configured a Microsoft Power BI business intelligence tool to explain virus growth through four steps, including the creation of an impact dashboard and subsequent analysis system requirements. Visualize and analyze your data as a case study to design the architecture, data integration and final steps to use.

When talking about improving an organization, it is necessary to take into account the business rules that means when information is interpreted and analyzed according to the products and services it provides.

This is already a very important factor and has a significant impact on business decision making.

In reality, metrics or reports do not always satisfy an organization. Instructions on the action to be taken.

(Enriquez, 2019) contributes to the implementation of business intelligence applications for organizational planning and development departments to generate analytical and strategic information based on the requirements needed to perform the required analytics and make decisions. I added. Better management.

(Castañeda, 2015) argues that implementing business intelligence can help reduce reporting time. The implementation of business intelligence using the Ralph Kimball method has been shown to improve decision making in the field of admissions to the Autonomous University of Peru. Similarly, using the data set (Caldas, 2020), we found that sales increased from 65.96% to 80.25%, achieving a growth of 14.29%.

Finally (Quintanilla, 2018) reports are now generated much faster, which undoubtedly helps to make timely decisions in the management area.

4.9. Post implementation validations

The confirmation process for information manipulation and transformation can be performed in several ways, as described above. Any source used to move, transform or reorganize data is available and can perform end-to-end validation. It also refers to the adaptation of information systems connected to new models when computational techniques, regressions or parallel developments are being implemented or are being implemented by applying business rules. It also guides you through understanding test-driven development (TDD), supporting development based on previously developed test cases and the ability to refine. An important concept in understanding the similarities to test-driven development

(Torres, 2017) is that unit tests are small units (methods) of modular, well-organized software code. This allows you to test function with minimal formulas. Among the functions that the system must have. Unit tests are part of the software design and help to specify the code requirements and validate the results. The process associated with unit test development starts with defining the system requirements (user stories) and the next step focuses on coding the unit test. All validation and validation is performed here, solving the proposed customer requirements, the final problem presented and eliminating duplication, duplication and feature group code error characteristics in a series of unit tests. Similarly, you can use this method to create a new schema or data flow according to the procedure you choose for ETL.

4.10. Database Migration

Information transformation also includes the movement of certain information. Database migration is shifting data from one system to another. This could be due to the arrival of other applications, memory changes or mode changes. Simply put, this is the process by which a large number of databases are migrated. From the old system to the new system. To clean, recover and migrate various data to your new system, you must perform a few steps. (Regi Zamon, 2017)

4.11. Data corruption

Problems arise when detecting changes in data content and format between the legacy system and the target system. Data corrupted during migration may now contain duplicates or duplicates. Ultimately, this is one of the migration risks that has the greatest impact on integrity and, therefore, operational and business efficiency. (PowerData, 2017)

4.12. Data loss.

Sometimes, once the data transfer is completed, the data cannot be used as a destination. This phenomenon is known as data loss and is the most serious possible migration risk. The cost of fixing this problem can range from data loss and image loss to loss of business due to missing data and reputational behavior as it solves data loss problems. (PowerData 2017)

4.13. Semantic risk.

This is another hot spot faced by those who are aware of the dangers of data migration. The traditional thermal column has the same meaning, but it occurs when the units of measurement are used differently in the two cases. These changes in the data completely change the meaning of the data. (PowerData, 2017)

4.14. Risk of interference.

This is a very common situation that occurs when multiple stakeholders use the source application at the same time during a transfer. Crashing, access problems and the inability to move certain information are risks that lead to migration. (PowerData, 2017)

5. CONCLUSIONS

Operations and other users can try various information solutions, including those mentioned in this user story. After that, we will learn our new model in our new model and express a photographer with new information we have. Then, you must decide according to the strategy by changing and interpreting this information to any area of the group. In addition,

with the importance of romantic testing for the migration of information. Each group has a trademark, each of the information needs to determine the metrics.

6. REFERENCES

[1] Tehreem, N. (2021). Concepts and approaches to datawarehouse. https://www.astera.com/es/type/blog/data-warehouse-concepts/
[2] Martinez, J (2016). Conceptual modeling and implementation of a genomic information loading process based on ETL technologies. Polytechnic University of Valencia, Valencia, Spain.
[3] Sanchez, P (2020). Data modeling and automation of the ETL process to support the NAUTIA methodology developed by the UPM platform on refugees. Polytechnic University of Madrid, Madrid, Spain.
[4] Enriquez, J (2019). Business Intelligence in the monitoring and evaluation processes of the annual operating plan of UPEC. Universidad Regional Autónoma de los Andes, Ibarra, Ecuador.
[5] Sandoval, A (2018). Analysis of existing data cleansing methods and techniques and application in a CRM system for a Lima educational institution. Pontificia Universidad Católica del Perú, Lima, Peru.
[6] Castañeda, A (2015). Development of Business Intelligence, based on Ralph Kimball's methodology, to improve the decision-making process in the admissions area of the Universidad Autónoma del Perú. Universidad Autónoma del Perú, Lima, Peru.
[7] Leguizamon, A (2017). Guidelines for a correct database migration. Universidad Libre, Bogotá, Colombia.
[8] Duque, D (2016). Model for the extraction, transformation and loading process in data warehouses. An application with environmental data. Universidad Nacional de Colombia, Medellín - Colombia.
[9] Monja, E (2020). Comparison of Naive Bayes and time series in the prediction of delinquency of social quotas of the College of Engineers of Peru Lambayeque departmental council. Señor de Sipan University. Lambayeque, Peru.
[10] Troche, A (2020). Application of transactional database mining. Research Institute of Science and Technology, La Salle - Bolivia.
[11] Fernández de Lis, A (2017). Digital transformation and competition in the financial sector.
[12] Torres, O (2017). Application and evaluation of the test-driven development (TDD) methodology, case study: spot.co. International University of La Rioja. Duitama - Colombia.
[13] Caldas, R (2020). Datamart for the evaluation of sales in the company Dismac Peru S.A.C. Universidad Cesar Vallejo, Lima- Peru.
[14] Domínguez, J. (2018). Introduction to data modeling. https://www.researchgate.net/publication/327872261_Introduccion_al_Modelado_de_datos
[15] Help Salesforce (2017). Glossary. https://help.salesforce.com/articleView?id=glossary.htm&type=5
[16] PowerData (2017). Risks in data migration. https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/riesgos-de-la-migracion-de-datos-y-aplicaciones-y-como-mitigarlos
[17] Azure (2021). PaaS. https://azure.microsoft.com/es-es/overview/what-is-paas/

[18] RedHat (n.d). PaaS. https://www.redhat.com/es/topics/cloud-computing/what-is-iaas
[19] IBM (n.d). Service models. https://www.ibm.com/pees/cloud/learn/iaas-paas-saas
[20] Talend (n.d). What is a service ETL process. https://www.talend.com/es/resources/what-is-e

Authors

Frank Escobedo Bailón,
Bachelor in systems engineering, systems and computer engineer, master in systems engineering, Doctor in education and Doctor in Systems Engineering, Principal Professor at the UNTELS and Associate Professor at the UNMSM, with experience in software architect.

Antonio Arqque Pantigozo
Systems and computer engineer, Bachelor's degree in education with a major in mathematics and computer science, Master's degree in systems engineering with a major in information technology management, Master's degree in strategic business administration and doctorate in systems engineering, with extensive knowledge in project management and information technology in the education sector.

Carlos Alzamora Aragon
Chemical Engineer with a Master's Degree in Public Management, I have developed professionally in industrial and service operations, both in management positions that together cover my years of work experience.
The academic and consulting side I have developed in the teaching at university level of Chemistry courses, organic and inorganic, Chemical Analysis and environmental courses.

Blanca Pasco Barriga
Teacher with 30 years of experience in university teaching in private and state universities, with a Master's degree in management and administration of education and doctorate in education with thesis to sustain, expert in virtual education in the management of the Moodle platform, blackboard and classroom as well as in the use of various virtual tools and simulators; former scholarship holder of the Israeli government by the Mashav-Israel center in 2013.

Soledad Olivares Zegarra
Doctor in Business Administration and Master's Degree in Marketing and International Business from Alas Peruanas University with a Master's Degree in Constitutional Law from José Carlos Mariátegui University and Post Doctoral studies at Hermilio Valdizán National University of Huánuco. She has a degree in administration from the Universidad Nacional Federico Villarreal, a degree in psychology from the Universidad Nacional Mayor de San Marcos and a law degree from the Universidad Particular Garcilaso de la Vega with experience in teaching classes at the undergraduate and

postgraduate levels.

Katherin Rodriguez Zevallos
Bachelor's Degree in Nursing, with Second Specialty in Nephrological Care, with a Master's Degree in Education Sciences. With professional experience in the Javier Prado Clinic. With a Diploma in: (1) Educational problems and development. (2) Postgraduate Diploma in university didactics, (3) Diploma in Family and Community Health in Nursing, (4) Diploma in Surgical Nursing, (5) Diploma in Nursing Care in Nephrology (6) Diploma in Ethics and Deontology. Currently, he is pursuing his doctoral studies in Public Health at the National University of Callao.