Master the data

The need to bring order into the chaos of information

CONTENT

https://zeus.vision/

ZEUS

Companies are absolutely full of data. From demographic to employment information, the software or the social media you use... all that data is found in tens of hundreds of different sources.

If we want to take advantage of the power of data, it is essential that we unify them so they can be exploited in the right way.

This way we will have a unique and homogeneous place to consult the information and we will be able to take business decisions in a more agile and accurate way thanks to the intermingling of data from different sources, and the simplicity in their representation.

The use of data is invaluable to improve the results of companies and achieve their objectives, but to do so, the first thing is to organise the chaos that is usually produced by the multiplicity and variety of information sources.

Objective

Turn the data into information, and from the information be able to extract valuable knowledge for decision making; all through the use of technologies and methodologies of Business Intelligence.

DATAWAREHOUSE

the importance of centralising data

Generally information in companies is found in work tools such as a CRM, ERP, accounting software, management software or even Excel, which are perfectly designed to insert data, but are not created to show information in real time, for example.

Their databases are optimal for data insertion but not for data extraction.

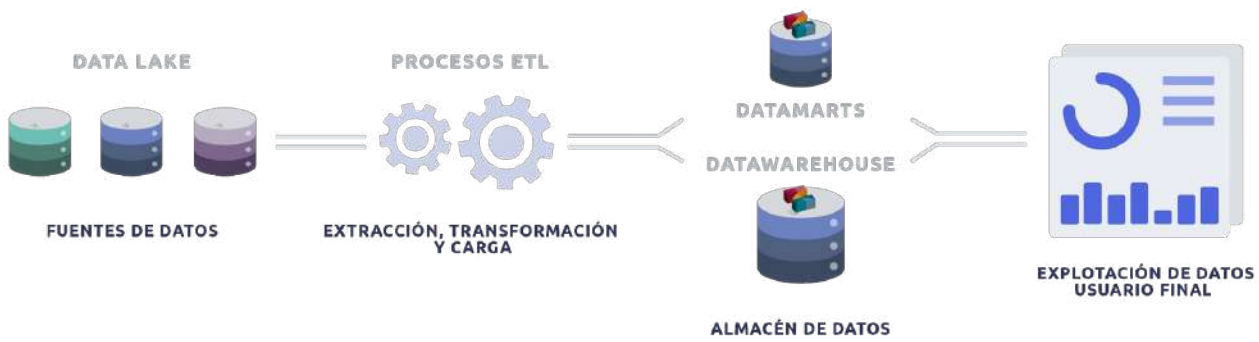
This is why one of the first jobs to be done when a company decides to take the leap and exploit its data to the fullest is **to locate all these sources of information and extract the data to be dumped into a suitable database for their transformation,** extraction and visualisation.

This type of database is known as a Data Warehouse.

Differences between **Datawarehouse, Data lake and Data Mart**

The biggest challenge in all BI projects is the quality of the source data, its location, extraction and standardisation.

This is a simple process map of a typical BI project:



Traditionally, the procedure for exploitation involves inserting the raw data, "in raw form", into a database, which is known as **Data lake**. Data lakes **are not organised or structured,** they are simply, as their name suggests, a data lake.

It is in this lake that information is "fished" to be processed and organised in a Data Warehouse. **The data in the Data Warehouse have been previously worked on through programming processes known as ETL.**

Once the data is transformed and organised it is loaded into another database, or several, smaller ones than the data warehouse, which are known as Data Mart(s). Each Data Mart usually contains the data of a specific area of the company: finance, sales, marketing...

What is ETL?

Extract, Transform, Load. *These are lines of programming code that allow the combination of the data we have extracted from different sources and that may have different formats, so they can work together. ETL projects can be written in multiple programming languages, such as Java, Python, SQL...*

Is it necessary to have these three types of databases? No.

It will depend on the BI project that needs to be carried out, the point in which the company is located, the volume of data, the loading speed required, the immediacy of the data required...

Which technology should I use **to process my company's data ?**

To make this decision the most important thing is to have the **help of a professional** who knows all the possibilities. Today, Big Data technology is very extensive and it is essential to know which tools to use. In order to not to go into technical detail, we will summarise the main points in this section.

The first thing you should know is that **the BI project** you carry out can work in **traditional hosting platform** (contracting physical servers to host the project) or in the **Cloud**, whose main characteristic is that the information is replicated between different network nodes and not in a single physical server.

The advantages of using **cloud**

For a Big Data project that, by its very nature, will tend to grow steadily, Cloud services are usually the most appropriate as they **are flexible and increase according to the project needs**. To make it simple, they work like electricity at home: you hire the Cloud service and you pay depending on the consumption (storage, consultations, speed...)

This feature is one of the reasons why it is essential to have experts when using cloud services such as Amazon, Azure (Microsoft) or Google Cloud, as costs can shoot up if resources are not correctly optimised.

If well used, the advantages of the Cloud are obvious. Lower cost and greater profitability, greater consistency (if one node fails, the information is automatically replicated to another), greater security or availability of multiple features on each of the available platforms.

How do I choose the **database** I need?

Another important point about technology in BI projects is **the type of database to be used, or the tools used to carry out the ETLs.**

With respect to databases, the most obvious difference is the amount of data to be processed and the speed required for its representation. Here again, having professionals who know and control all the options is the key.

Relational databases are those that have been used in software and business applications over the last 40 years. They can be used to process Big Data, but they may well fall short of what is required. Precisely **because of the need to easily handle huge amounts of data**, other types of databases have emerged that do not use tables, fields and rows like relational ones and that are known as **NoSQL**.

Just to name a few, the most commonly used databases for BI projects are Oracle, MongoDB, Amazon's Redfshit, or Kassandra among others.