## The Business Intelligence Concept

(Connolly and Begg, 2015) defines Business Intelligence as *“an umbrella term that refers to the processes for collecting and analyzing data, the technologies used in these processes, and the information obtained from these processes with the purpose of facilitating corporate decision making”.*  The term Business Intelligence is a relatively new one as it appeared around mid - 90s. However, systems that provide Business Intelligence exist since early 1980s – they used to be called Enterprise Information systems (EIS) (Sharda et al., 2015). EIS were the first systems that had the capability of providing advanced data analytics such as forecasting, prediction and ad-hoc reporting. EIS replaced the Management Reporting Systems (MRS) that showed up in the 1960s and had very limited capabilities. As the EIS continued evolving, they were enhanced with additional features such as artificial intelligence and data mining. These systems are now called Business Intelligence systems (BIS) and are a very important element of a modern business in terms of supporting decision making.

## Business Intelligence Systems

(Sharda et al., 2015) defines a Business Intelligence System as system that has four components. A Data Warehouse, Business Analytics, a User Interface and Business Performance Management (BPM). The data warehouse is the central repository of data that have been collected from different internal or external sources. Business Analytics refer to tools that are used for querying, mining or analyzing data from the data warehouse such as OLAP tools or Data Mining tools. The User Interface refers to the dashboards or reports that can be produced by modern reporting tools. Finally the business performance management (BPM) is tightly connected with Business Intelligence as the latter can help BPM to identify strategy requirements, drive performance and monitor achievement. (Fidler, 2016). This project will focus on the data warehouse component, and more specifically of the methods that are used to extract data from data sources and the load the data into the data warehouse, after performing any necessary transformations.

## The Data Warehouse concept

A data warehouse is the main component of a business intelligence system. It’s another database, technically similar to the operational databases, but serves a completely different purpose. While an operational database is optimized to process transactions efficiently and assist the knowledge workers to their day to day tasks, a data warehouse is focused on analytical processes and has the goal to improve decision making. A data warehouse is a decision support system that provides clean and credible data. Businesses are using operational systems to get data in and data warehouses to get data out. (Kimball and Ross, 2013).

[Write about dimensional modelling)

## Extract, Transform, Load (ETL)

ETL stands for Extract, Transform and Load and it’s a term widely used in data warehousing. The ETL system is the backbone of a data warehouse as it is responsible for the extraction of the data from external sources, the cleaning of the data and the loading of the data into the data warehouse. The design and implementation of an ETL system is a complex project divided into many subtasks. There are many different methodologies, tools and technologies for ETL development and implementation. ETL design is a significant part of the Business Intelligence lifecycle (Moss and Atre Shaku, 2003). Before implementing any data flows, The ETL Team should take into consideration the business requirements and plan the ETL solution accordingly.

The Extract Step: The Extract Step of the ETL process should be planned based on the business needs. The business requirements set by end users define the data sources or specific entities and attributes of an operational system that need to be considered for integration into the data warehouse. Data Sources or other database objects that are not useful for analysis should not be considered.

The Transform Step: Part of the Transform step is also driven by the business needs. Specific Business rules are applied at this step. The data are cleaned, conformed and ready to be imported in the data warehouse.

The Load Step: The end users want to have easy access to information and they should be able to understand the underlying data model. The Business Intelligence Team should chose a data model that is simple, scalable and efficient and the ETL Team needs then to design the ETL processes to load the data efficiently in the data structure that has been implemented based on the chosen data model.

An additional business requirement that affects all of the steps above is the data latency requirement (Kimball Ralph, 2004). The frequency the data warehouse needs to be updated with fresh data is one of the most important aspects to consider by the design of an ETL solution. The update frequency of a data warehouse / data mart is driven by the business needs and can vary from once per month to real-time. The data latency requirement can have a huge impact in the design of an ETL solution as it can define the ETL Software/Hardware Architecture. Real-time ETL solutions require modern ETL tools that use a different architecture.

## Evolution of ETL techniques