Data Warehouse and Data Mining

Omar Alomari, Osama Kfaween, Yousef Shaban

ABSTRACT

Data Warehouse and data mining are two main components that complement the other, in which a warehouse focuses on storing data in a structured database, meanwhile datamining focuses on extracting facts and useful information from these storages, in order to apply these results into business needs.

From medium to huge, nearly all industries and organizations can benefit gracefully from this approach!

https://www.comptia.org/content/articles/how-is-data-mining-used-in-marketing
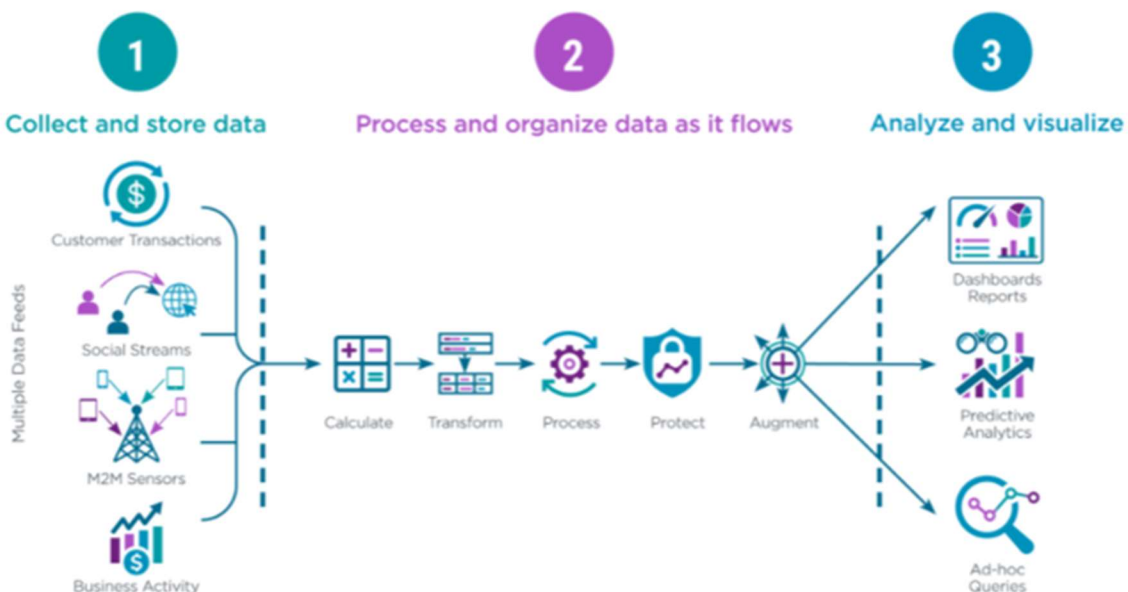
INTRODUCTION

Building up from the concept that **every successful leader must develop data-driven strategies** [1], it is essential to acquire reports and statistical analysis that is obtained through meaningful historical records.

Better yet, it is mentioned that 80% of organizations that apply business intelligence find the process of data mining important. [2]

So, what is a data warehouse and data mining? What are they complementary? And how they solve business needs?

https://www.springernature.com/gp/researchers/text-and-data-mining
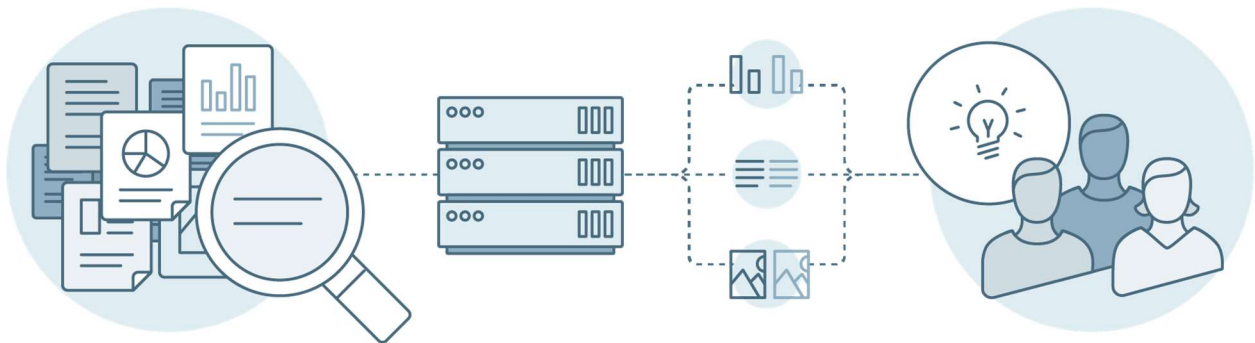
Data Warehouse

What is a Data Warehouse?

According to Oracle, a data warehouse us a sort of management system that stores historical data, in which it is designed to support business intelligence (BI) and particularly analysis. That can be useful for data scientists and business analysts.[3]

Main characteristics of a Data Warehouse:

1. Non-volatile
2. Time-Variant
3. Subject Oriented
4. Integrated

Why is Data Warehousing so important?

A data warehouse is consistent, which means it offers reliable quality data, provides historical precise data, have a huge storage amount, and plays a key role in cost reduction.

Data Mining

What is Data Mining?

Data mining is a process, in which the techniques evolve around examining massive amounts of data, in which the process could be conducted in parallel or distributed computing.

Main characteristics of Data Mining:

1. Manage Datasets
2. Predictions
3. Business Enhancement
4. Large Calculations

Why is Data Mining so important?

Data mining assist organizations making better business decisions, it may minimize risks, increase sales. It encourages not just smart decision making, but also detects trends, and different types of frauds.

Data Mining Issues

- **Noisy and Incomplete Data**

  Data mining is a method of extracting information from massive amounts of data. This knowledge about current reality is noisy, fragmentary, and varied. Massive volumes of data will be erroneous or wrong on a frequent basis. These problems might be the result of human errors, blunders, or flaws in the devices that measure the data.

- **Distributed Data**

  True data is often stored in phases in dispersed processing situations. It is possible that it is on the internet, particular systems, or databases. Carrying all the data to a consolidated data store is difficult for technological and organizational reasons.

- **Performance**

  The presentation of the data mining framework is mostly determined by the efficiency of the techniques and algorithms used. If the strategies and algorithms proposed are insufficient, it will have a negative impact on the presentation of the data mining measure.

- **Scalability and Efficiency of the Algorithms**

  The Data Mining technique should be scalable and efficient in order to extract information from massive amounts of data in the data set.

- **Data Privacy and Security**

  Data mining frequently raises serious concerns about governance, privacy, and data security. For example, when a store checks purchase details, it learns about clients' purchasing habits and preferences without their knowledge.

- **Others such as: security and social challenges, complex data, data visualization, etc.** [4]

Data Mining Tools

RapidMiner

is a free open- source data science platform that features hundreds of algorithms for data preparation, machine learning, deep learning, text mining, and predictive.

Oracle data mining

is a component of oracle advanced analytics that enables data analysts to build and implement predictive models. it contains several data mining algorithms for tasks like classification, regression, anomaly detection, prediction, and more.

Weka

is an open-source machine learning software with a vast collection of algorithms for data mining. it was developed by the university of Waikato, in New Zealand, and it's written in JavaScript.

Bibliography

[1] https://www.herzing.edu/blog/what-data-warehousing-and-why-it-important#:~:text=Data%20warehousing%20improves%20the%20speed,them%20apart%20from%20their%20competitors.

[2] https://www.iberdrola.com/innovation/data-mining-definition-examples-and-applications

[3] https://www.oracle.com/in/database/what-is-a-data-warehouse/

[4] https://www.jigsawacademy.com/blogs/data-science/data-mining-challenges