

7.19: Data Warehousing and Data Mining

Learning Objectives

- Differentiate between data warehousing and data mining



All of the RIS we have discussed so far have one major thing in common—an underlying database to store their unique data. Through the process of mergers and acquisitions, most large retailers inherit duplicative systems that continue to exist independent of each other due to the large cost of consolidation. With data “everywhere,” retailers turn to the latest IT techniques.

Data warehouses (DW) are created to bring related information from disparate databases to one large database so that it can be easily analyzed.

In computing, a data warehouse (DW or DWH) is a system used for reporting and data analysis, and is considered a core component of business intelligence. DWs are central repositories of integrated data from one or more disparate sources.

Once the data has been migrated to the DW, data scientists can begin to provide retail management with meaningful information through the practice of data mining. Data mining is the process of discovering patterns in large data sets and involves methods at the intersection of machine learning, statistics, and database systems.

With the mining of information in the data warehouse, management can gain valuable insights as to how best to run the business. This is usually accomplished through queries and reporting.

Queries are business questions translated into code to bring results from the DW. What is our best-selling product line? What is the profit margin on our private brand versus the name brand products? Who are our best customers? How do our online sales affect our inventory position for our stores?

Business reporting is simply scheduling the most common or requested queries at regular intervals and pushing the information out to the organizations information consumers on a regular basis.

One of the most notable data warehouse success stories comes from the healthcare industry in the 1990’s. A large national health management company had more than nine regional centers, each operating semi-independently. Each regional center had its own management and business infrastructure, including information technology.

The company’s top medical experts noticed that the care being delivered for its diabetes patients was inconsistent across the regions. Some regions claimed that certain treatments were more effective, but came at a higher cost to the business. But the real problem was that the clinical data needed to understand what was the most effective treatment was locked up in 10 different databases, many of which were using different database software.

A data warehouse was constructed, pooling the data from all of the regional warehouses and providing access for the first time for national clinical research analysis. Out of hundreds of different treatment programs, three were found to be the most effective, and of those, two were found to be the least expensive. These data warehouse results were a win for the patients, doctors and the business.

Practice questions

<https://assessments.lumenlearning.co...sessments/9203>

Contributors and Attributions

CC licensed content, Original

- Data Warehousing and Data Mining. **Authored by:** Bob Danielson. **Provided by:** Lumen Learning. **License:** [CC BY: Attribution](#)

CC licensed content, Shared previously

- Data Mining. **Authored by:** Arbeck. **Provided by:** Wikimedia Commons. **Located at:** https://commons.wikimedia.org/wiki/File:Data_Mining.svg. **License:** [CC BY: Attribution](#)

7.19: Data Warehousing and Data Mining is shared under a [not declared](#) license and was authored, remixed, and/or curated by LibreTexts.

- **7.19: Data Warehousing and Data Mining** is licensed [CC BY 4.0](#).