

COMP 333 — Week 5 Welcome

Welcome to COMP 333 Data Analytics

In Week 5 we will look at Data Warehouses that laid the foundation for *decision-support systems* (DSS) primarily in business.

So Data Warehouses were an early part of Data Analytics.

This lecture will provide you an understanding of how Data Analytics is deployed in large-scale organizations.

The lecture will discuss ETL (Extract-Transform-Load) process introduced by the field of Data Warehouses which is the fore-runner of the Data Wrangling process in Data Analytics.

We do not directly apply the content of this lecture to lab exercises nor lab assignments.

The content of this lecture does not form any significant part of your quizzes.

You should READ the supplied chapter:

Introduction to Data Warehousing, Chapter 1 of *Data Warehouse Design: Modern Principles and Methodologies*, 2009, by Matteo Golfarelli and Stefano Rizzi.

or at least the indicated sections of it below.

You may also read/watch the supplementary material.

In Week 5 we will cover

- ▶ Databases, Data Warehouses, Data Marts, Dashboards, which are the components deployed for Decision-Support Systems in business
Golfarelli and Rizzi Sections 1.2, 1.3.2, and 1.7.3
- ▶ ETL (Extract-Transform-Load) process (for Data Wrangling)
Golfarelli and Rizzi Section 1.4
- ▶ Data Cubes and OLAP (Online Analytical Processing) for data mining. OLAP introduced the slicing, aggregation, groupby operations that influenced SQL and `pandas`
Golfarelli and Rizzi Sections 1.5, and 1.7.2
- ▶ Business Intelligence, the successes of Decision-Support Systems in business, such as e-commerce, that sparked Data Analytics more broadly across society

READ the files marked READ.

You may also read/watch the supplementary material.

All the best, Greg.