14D003

Data Warehousing and Business Intelligence

6 ECTS

Prof. Nandan Rao

Guest Industry Professionals

Prerequisites to Enroll

None

Overview and Objectives

The purpose of this course is to both give students an overview of modern solutions to storing and analyzing big data, as well as hands-on practice working with databases, building systems to collect data from the internet, and creating live web dashboards. Students will complete exercises to: A) become proficient in basic SQL syntax and queries in a MySql environment, B) build and deploy a system that can collect data from the Twitter streaming API and store it in a data warehouse, and C) scrape data from websites. Students will complete a final project in groups of 2-3 that will consist of building a web dashboard to display the data of their choosing. Students are encouraged to think creatively and use knowledge from other courses during the first term to come up with an informative display of data that they can create with the dashboard tools taught in class. In addition to the projects and hands-on exercises, lectures will give an overview of current technologies used in data acquisition and warehousing so that students have an understanding of how and where to use which system. Additionally, we will have visits from industry professionals to present case studies on how they use business intelligence in the field.

Course Outline

SQL

* Relational Algebra

* SQL Syntax and Semantics

* Users and Permissions in SQL Databases

* SQL Exercises

SQL, NoSQL, and Flat Files: Overview of Modern Warehousing Strategies

14D003

Data Warehousing and Business Intelligence

* NoSQL Databases (Mongo, Cassandra, Elasticsearch, Redis)

* The Hadoop Ecosystem (HDFS, MapReduce, Spark)

* Flat files and object storage (JSON, csv, Parquet, ORC)

The State of IAAS

* AWS/Google/Azure

* Hosted Databases, Queues, Object Storage

* Container Orchestration

Web Data: API's and Scraping

* Introduction to the World Wide Web

* API's and JSON data

* Web scraping basics

Web Dashboards

* Understanding Web Applications

* Dashboard libraries in R and Python

* HTML/CSS Basics

Business Intelligence Case Studies

* Case studies in modern business intelligence solutions

14D003

Data Warehousing and Business Intelligence

Evaluation

Assignments (60%) and Final Project (40%)

Competences

☐ Apply the knowledge of programming languages, computer programs and advanced services in the Cloud to solve the problems that are presented to the data scientist.

☐ Visualize and interact with high-dimensional data in order to contextualize the information and facilitate subsequent decision-making.

☐ Own and understand knowledge that provides a basis or opportunity to be original in the development and / or application of ideas, often in a research context.

☐ That the students know to communicate their conclusions and the knowledge and last reasons that sustain them to specialized and non-specialized publics in a clear and unambiguous way.

☐ That students have the learning skills that allow them to continue studying in a way that will be largely self-directed or autonomous.

Learning Outcomes

☐ Work with databases and cloud computing.

☐ Work with Big Data information using data mining techniques.

☐ Treat high-dimensional data environments knowing their limitations and how to present the results.

☐ Present information visually and in an orderly manner to improve decision making.

☐ Collaborate in a computing environment that requires structuring and planning.

Materials

Links to online documentation of every technology will be provided.

14D003

Data Warehousing and Business Intelligence

6 ECTS

Prof. Nandan Rao

Guest Industry Professionals