# Introduction to Data Modeling and Processing

| | | | |
|---|---|---|---|
| **Instructor:** | Tuncay Tekle | **Time:** | M 19:00-22:00+Sa 13:00-16:00 |
| **Email:** | tuncay@sabanciuniv.edu | **Place:** | Minerva Han |

**Course Page:**

1. http://tuncaytekle.com/da505.html

**Main References:** This is a restricted list of various interesting and useful books that will be touched during the course. You may find it helpful to consult them occasionally.

- Graeme Simsion and Graham Witt. 2004. *Data Modeling Essentials, Third Edition (Morgan Kaufmann Series in Data Management Systems).* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

- Toby J. Teorey. 1998. *Database Modeling and Design (3rd Ed.).* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. Foster Provost and Tom Fawcett. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking (1st ed.).* OReilly Media, Inc.

- David M. Kroenke and David J. Auer. 2013. *Database Processing: Fundamentals, Design, and Implementation (13th ed.).* Prentice Hall Press, Upper Saddle River, NJ, USA.

- Pramod J. Sadalage and Martin Fowler. 2012. *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence (1st ed.).* Addison-Wesley Professional.

**Objectives:** This course is designed for graduate students in the Data Analytics program, and will introduce an audience of varied backgrounds to the state-of-the-art in data modeling techniques for relational and nonrelational databases, and various data processing languages and systems will be introduced as well. We try to discuss as many systems as possible. We chiefly focus on SQL databases, and a variant of different NoSQL solutions.

At the end of the course, a successful student should be able to:

- model a database based on business requirements,

- implement a relational database from scratch using database modeling tools in conjunction with SQL code,

- perform data processing tasks with SQL, and understand the limitations/possibilities available with different SQL systems,

- understand why SQL may be insufficient or not desirable under certain conditions,

- design data models, and perform processing with various NoSQL systems, including document, graph and key-value models.

**Prerequisites:** None.

**Tentative Course Outline:**

- ☒ Data representation, Normal forms, normalization
- ☒ Entity-relationship modeling
- ☒ Primary keys, foreign keys, relationship types, attributes
- ☒ Object relational mapping frameworks
- ☒ UML
- ☒ Data processing primitives
- ☒ SQL
- ☒ Limitations of relational databases, alternative solutions
- ☒ Transactions, ACID, BASE
- ☒ Document model, graph model, key value model

All of these topics will be supplemented with hands-on sessions with relevant software.

**Grading Policy:** 3 Midterms (best 2 account for 30%), Final (30%), Project (40%).