Course Abstract

- This course is the first course in big data for the ABC courses (AI model/algorithm, **Big data**, Computing) of the **Graduate School of Data Science**. The course covers the foundation of data management for data science and related fields. It covers the following topics: - Theoretical background of data management, including data type, first-order logic, second-order logic, relational calculus and algebra, schema, and normalization. - Relational database, including the ER model, transaction, concurrency control, logging, recovery, SQL, OLTP, query optimization. - Distributed and federated database systems. - Data analytics, including OLAP, column store, ETL, operational data stores, data warehouse, data lake, and in-memory databases. - Physical design of databases with hands-on exercises to implement database system functions, such as B-tree, using Postgres or MySQL. - Data wrangling with hands-on exercises with NumPy, Pandas, and Python.