Intro to data engineering

Introductory Course

Goal:

learn about databases and prepare to use in data science projects

Requirements: Windows 10 or Linux Ubuntu with PyCharm Community Edition IDE with JetBrains Academy plugin installed

python 3.10 with postgres psycopg

postgres database and admin/query tools installed

Curriculum:

Class 1 Intro to data engineering

Class 2 SQL Databases – Postgres

Class 3 Table design and optimization.

3.1 type discovery

3.2. schema discovery

3.3 fk/pk discovery

3.4 type and schema validation

3.5 ER graph discovery

3.6 normalization

3.7 tall vs wide schemas

3.8 pivot and unpivot operations

Class 4 Views and Materialized views

Class 5 Constraints

Class 6 Indexes

Class 7 Explain plan analysis

Class 8 Backup and recovery

Class 9 Row and column level Security

Class 10 Data preparation

Class 11 Analytical functions

11.1 column and row functions

11.2 rolling aggegration

11.3 groupby, rollby and cubeby operators

11.4 sampling = upsampling and downsampling

11.5 aggregation through time and space

11.6 slicing and dicing

Class 12 Data quality

12.1 data quality and anomaly detections for single variable and multi-variable time series 12.2 coverage, correlation, 2d shape detection, autoencoder

Class 13 data visualization - Tableau and Seaborn