

[L4-DE] Data Engineering: Specialization

Session 1	
Data pipeline types ETL: data extraction	Introduction Data pipeline types: ETL vs ELT File handling Web services Databases Best practices for data extraction
Session 2	
ETL: data transformation and loading Data anonymization	Basic transformations recap Best practices for data transformation Data anonymization Data storages Data loading
Session 3	
ELT, big data & Hadoop, Spark, cloud	A quick intro to Hadoop Big data connectors, IO on Hadoop Querying data on Hive Processing data on Apache Spark Performance optimization on Spark Running Hadoop applications in a cloud
Session 4	
Data pipelines orchestration Data governance	Data pipelines orchestration Data governance: data quality, security, and observability Best practices for data pipelines design and development Data presentation
Session 5	
Wrap up session	Q&A