

# Data Engineering

*(Laurent Brisson et Gregory Smits)*

This course covers the fundamental concepts and practical skills required in data engineering, from traditional relational databases to modern NoSQL paradigms and big data technologies. It emphasizes hands-on experience with essential tools and frameworks, preparing students to design, implement, and maintain robust data infrastructure in real-world scenarios. The course also incorporates best practices in data governance, security, and environmental impact considerations.

Key concepts covered:

- Advanced relational database concepts (indexing, security, rights management, limitations, distributed relational models)
- NoSQL data paradigms (graph, document, semantic data)
- Data querying tools and techniques
- Externalized indexing systems (ELK stack: Elasticsearch, Logstash, Kibana)
- Data pipelines: automation, performance, quality, and validation (ETL)
- Big data technologies (HDFS, Hadoop)
- Cloud computing for data engineering
- Data governance, rights management, security, and environmental impact

By the end of this course, students will be able to:

- Design and implement scalable database solutions using both relational and NoSQL paradigms
- Develop efficient data querying and indexing strategies for various data formats
- Create and maintain robust data pipelines for ETL processes
- Apply best practices in data governance, security, and environmental considerations to data engineering projects
- Leverage cloud computing platforms for data engineering tasks
- Communicate effectively about data engineering concepts and solutions

Prerequisites:

- Proficiency in SQL and database management
- Familiarity with Linux command-line interface
- Programming experience (preferably in Python or Java)