

Data Scientist

POSTGRADUATE STUDIES



Study program

Data analysis with R (18 hours)

- R and R Studio environment, atomic types, vectors, lists, functions, data cleaning, plotting, markdown

Database systems. SQL basics (24 hours)

- ERD diagrams, Normalization, SQL DDL, SQL DML (24 h)

Programming in python (24 hours)

- Syntax, arrays, functions, Pandas, Statistics in Python

Object oriented programming in python (12 hours)

- Attributes, classes, constructor, methods, inheritance, "magic methods"

Data analysis in java (hours)

- Basic concepts of object-oriented programming, development environment and tools, what is data analysis, data science vs. data analysis, why Java, data structures - Java Collections Framework, ETL - Extract, Transform, Load processes (16 h)
- Ways of integration with relational databases, data processing in a functional approach (16 h)

Apache kafka (6 hours)

- Apache Kafka interface: a fast start to streaming data processing (6 h)

Nosql (microsoft azure) (20 hours)

- Basic concepts of NoSQL Databases - HBase, Cassandra, Impala, Neo4j (20 h)

ADVANCED DATABASES AND DATA WHOLESALE (24 hours)

- Advanced aspects of SQL and TSQL; Concepts of data warehouse modelling (ROLAP, MOLAP, HOLAP); Technologies ETL/ELT; Elements of data presentation, e.g. Power BI

Big data tools (microsoft azure) (24 hours)

- Apache Hadoop & Apache Spark (24 h)



Project (8 hours)

- Seminar (8 h)