

DATA HARVESTING

powered by **CSC** - IT Center for Science

Introduction to data science (1 day)

- Harvesting data
- Data cleaning and preprocessing
- Exploring data numerically and visually
- Common analysis methods

Cloud computing, Hadoop and Spark (2 days)

- Basic of cloud computing and IaaS
- Hadoop file system and map-reduce principle
- Spark framework basics
- Using PySpark for data processing and analysis
- Handling large datasets

Data processing with Python (2 days)

- Python basics and Numpy
- Pandas and visualisation libraries

Data warehousing with SparkSQL (1-2 days)

- From SQL databases to data warehousing
- SparkSQL basics
- Exploring and analysing data with SparkSQL and R/Python

Data analysis and visualization (1-2 days)

- Basic of data analysis
- Explorative methods
- Predictive methods and statistical modelling
- Visualisation and visual analytics



<http://www.kajakdc.fi>

