

# Data engineering (INFO904\_SIAY)



## En bref

- > **Langues d'enseignement:** Anglais
- > **Ouvert aux étudiants en échange:** Oui

## Présentation

### Description

The course introduces students to modern technologies used for data processing in the context of Big Data. In particular, the course aims to position the need for parallel processing architectures in the modern data environment. Through examples, students will learn how to implement simple processing pipelines and study the benefit of their usage.

### Heures d'enseignement

Data engineering - CM	Cours Magistral	12h
Data engineering - TD	Travaux Dirigés	12h

### Pré-requis obligatoires

Programming skills in Python. Notions of data processing libraries, e.g., NumPy, Pandas, etc. Basic notions of parallel computing.

### Plan du cours

- \* Data engineering 101: data stores and parallel computing
- \* Big data tools: MapReduce, Hadoop, Spark
- \* Extract, Transform, and Load pipelines

## Compétences visées

- \* Understanding of big data processing tools
- \* Ability to implement simple parallel processing pipelines
- \* Ability to display the results of data processing pipelines

## Infos pratiques

---

### Contacts

Responsable du cours

Francesco Bronzino

✉ [francesco.bronzino@univ-smb.fr](mailto:francesco.bronzino@univ-smb.fr)

---

### Lieux

› [Annecy-le-Vieux \(74\)](#)

---

### Campus

› [Annecy / campus d'Annecy-le-Vieux](#)