

# Big Data: Distributed Data Management and Scalable Analytics

## Titulaires

Dimitrios SACHARIDIS (Coordonnateur) et Gianluca BONTEMPI

## Mnémonique du cours

INFO-H515

## Crédits ECTS

5 crédits

## Langue(s) d'enseignement

Anglais

## Période du cours

Deuxième quadrimestre

## Campus

Solbosch et Plaine

## Contenu du cours

The course is subdivided into 2 parts: Big Data Management and Big Data Analytics. The part on Big Data Analytics builds on concepts introduced in the part Big Data Management.

### Part I: Big Data Management:

1. Introduction & Map/Reduce
2. Spark
3. Streaming: Spark Streaming & Storm; Big Data Architectures
4. Consistency and Availability
5. Distributed and parallel query evaluation
6. Stream Processing and Sublinear Algorithms

### Part II: Big Data Analytics:

1. Introduction
2. Batch distributed machine learning
3. Sequential machine learning and streaming
4. Recommender systems and Collaborative filtering
5. Deep learning

## Objectifs (et/ou acquis d'apprentissages spécifiques)

### Learning outcomes

After successful completion of this course, the student:

1. Understands the characteristics of big data, and the challenges these represent
2. Knows the principal architectures of Big Data Management and Analytics Systems (BDMAS), is able to explain the purpose of each their components, and is able to recognize and explain the key properties, strengths, and limitations of each type of BDMAS and their components.

3. Understands the key bottlenecks in managing and analyzing massive amounts of data and is familiar with modern algorithms for overcoming these bottlenecks using parallel and distributed computation.

4. Is able to actively use this algorithmic knowledge in the design and implementation of applications that solve common data management and analytics problems using different types of BDMAS.

5. Is able to build applications using specific instances of each type of BDMAS.

6. Is able to use established software frameworks for reproducing/sharing her/his results,

## Pré-requis et co-requis

### Connaissances et compétences pré-requis

- > Databases, SQL
- > Supervised machine learning (classification, regression, feature selection): it is highly recommended that the student followed the class INFOF422 "Statistical foundations of machine learning"
- > Notions of statistics and probability
- > Programming in Python: the student should have a high degree of autonomy in programming and installing the required computing tools, including virtualization software (Docker), version control systems (Git), and notebooks (Jupyter).

## Méthodes d'enseignement et activités d'apprentissages

Combination of Ex-Cathedra Lectures, Exercise sessions, Computer labs, Self-study, and Project Work.

### Contribution au profil d'enseignement

- > Formuler et résoudre des problèmes techniques et scientifiques, complexes, ou des questions inédites, en mobilisant des capacités d'abstraction, de modélisation, de simulation et d'analyse multidisciplinaire, en répondant aux exigences de la recherche universitaire, en intégrant les besoins, les contraintes, le contexte et les enjeux techniques, socio-économiques, éthiques et environnementaux, en vue d'apporter des solutions concrètes
- > Maîtriser et mobiliser un ensemble structuré de connaissances, tant transversales que spécialisées, et être capable de les faire évoluer avec autonomie et esprit critique
- > Définir, planifier, gérer et mener à bien des projets d'envergure compte tenu de leurs objectifs, ressources et contraintes et en assurant la cohérence et la qualité de la démarche et des livrables

- Travailler efficacement avec d'autres professionnels (en équipe, en partenariat ou en concurrence), prendre des décisions et développer son leadership, dans une variété de contextes professionnels, disciplinaires et culturels
- Communiquer et échanger des informations de manière structurée – oralement, graphiquement et par écrit, en français et dans une ou plusieurs autres langues – sur les plans scientifique, technique

## Support(s) de cours

Université virtuelle

## Autres renseignements

### Lieu(x) d'enseignement

Solbosch et Plaine

### Contact(s)

Part II: Big Data Analytics: Pr. Gianluca BONTEMPI  
Gianluca.bontempi@ulb.be

## Méthode(s) d'évaluation

Examen écrit et Projet

### Méthode(s) d'évaluation (complément)

Written exam (on UV platform) and project.

## Construction de la note (en ce compris, la pondération des notes partielles)

<sup>1</sup> Written exam (on UV): 10/20

<sup>2</sup> Project work: 10/20

## Langue(s) d'évaluation principale(s)

Anglais

## Autre(s) langue(s) d'évaluation éventuelle(s)

Français

## Programmes

### Programmes proposant ce cours à l'école polytechnique de Bruxelles

MA-IRCB | **Master : ingénieur civil biomédical** | finalité Spécialisée/bloc 2, MA-IRIF | **Master : ingénieur civil en informatique** | finalité Spécialisée/bloc 2 et MS-BGDA | **Master de spécialisation en science des données, Big data** | bloc U

### Programmes proposant ce cours à la Solvay Brussels School of Economics and Management

MS-BGDA | **Master de spécialisation en science des données, Big data** | bloc U

### Programmes proposant ce cours à la faculté des Sciences

MA-BINF | **Master en bioinformatique et modélisation** | finalité Approfondie/bloc 2, MA-INFO | **Master en sciences informatiques** | finalité Spécialisée/bloc 1 et finalité Spécialisée/bloc 2 et MS-BGDA | **Master de spécialisation en science des données, Big data** | bloc U