

## Big Data

Course title – Intitulé du cours	Big data – Part 2: Big Data Management with AI and Cloud Computing
Level / Semester – Niveau /semestre	M2 / second semester
School – Composante	Ecole d'Economie de Toulouse
Teacher – Enseignant responsable	Nicolas LE GALL
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Lecture Hours – Volume Horaire CM	12
TA Hours – Volume horaire TD	
TP Hours – Volume horaire TP	
Course Language – Langue du cours	English / Anglais
TA and/or TP Language – Langue des TD et/ou TP	English / Anglais

### **Teaching staff contacts – Coordonnées de l'équipe pédagogique:**

Email : nicolas.le-gall@aqzone.com

Office number: -

Office Hours: -

Preferred means of interaction: Email, or after class

### **Course's Objectives - Objectifs du cours:**

This course offers a deep dive into the strategic management of Big Data using AI and cloud computing, with a focus on practical applications within a business context. Students will develop both theoretical insights and hands-on skills, particularly in using Google Cloud Platform (GCP) for Big Data solutions.

The course starts with an exploration of Big Data's historical development and its critical role in shaping modern business strategies. Students will study how Big Data has evolved, examining its impact on various industries and the strategic benefits it offers to companies. This theoretical foundation is complemented by practical sessions on Google Cloud Platform (GCP), where students will familiarize themselves with the platform's tools and services for Big Data management.

Following this, the course will shift focus to data engineering and storage within the cloud. The theoretical component will cover the essential aspects of data storage, including the various technologies and methodologies used in cloud environments. In parallel, students will engage in practical exercises on GCP, where they will perform data ingestion, connect to ETL platforms, and

develop data engineering pipelines. These activities will solidify their understanding of how to effectively manage and process data in a cloud-based infrastructure.

The course concludes with a focus on data modelization and visualization. Students will explore theoretical concepts related to modelization and the significance of data visualization in deriving actionable insights. The practical sessions will involve working with data prepared in earlier modules, applying GCP tools to perform modelization and create visualizations that can inform business decisions. This will ensure that students can translate data into valuable insights, supporting strategic decision-making processes.

By the end of the course, students will have acquired the knowledge and skills necessary to manage Big Data strategically, using AI and cloud computing to drive business success in a data-driven world.

To summarize it, here are the learning objectives:

- Understand the Role of Big Data in Business
- Introduction to Google Cloud Platform (GCP)
- Basic Data Engineering Concepts
- Overview of AI and Machine Learning Applications
- Fundamentals of Data Visualization
- Apply Concepts to a Simple Big Data Project

#### **Course outline:**

- 1 hour to cover general concepts
- 4-5 hours to cover storage and preprocessing of the data
- 4-5 hours to cover modelization and vizualisation of the data
- 2 hours to review, and discuss the projects

Each lecture will consist of two parts. During the first part, students will interact with premade Python exercises. The second part will be a lecture covering the topics seen in each exercise. In other words, students will first learn by doing.

#### **Prerequisites - Pré requis:**

We will be covering some advanced concepts in machine learning and programming. Therefore it is expected that students have some experience with Python programming and are somewhat at ease with the basics of machine learning. Some knowledge of SQL and Cloud concepts would help too.

#### **Grading system - Modalités d'évaluation:**

Final project (80%)

Practical sessions (20%)

#### **Bibliography/references - Bibliographie/références:**

No specific references but you can have a look at everything related to Cloud for understanding the generic concept. And feel free to Google the keywords in the paragraph above.