USEES8 - Big Data Technologies for Connected Industries

Mis à jour le 05-07-2024

Code : USEES8

Unité spécifique de type cours

3 crédits

Responsabilité nationale : EPN05 - Informatique / Stefano SECCI

Présentation

Prérequis

- Basics of data and databases.
- Basics of programming.
- Working knowledge of Python.
- Working usage of Command Line Interface (CLI).

Objectifs pédagogiques

- In an age defined by the sheer magnitude, diversity, and speed of data production, expertise in Big Data Technologies is indispensable. Traditional data management tools are insufficient for managing this data avalanche, necessitating innovative solutions. Our advanced course, 'Big Data Technologies', is tailor-made to equip students with the knowledge and hands-on skills crucial for navigating the realm of Big Data.
- Our goal is simple: to instill a profound understanding of Big Data principles, frameworks, and state-of-the-art tools necessary for constructing resilient data systems capable of handling massive and intricate datasets. Throughout this course, students will master the basics of Big Data, recognize its pivotal role in today's data-centric world, and become proficient in employing various technologies and frameworks to design and implement scalable data solutions.
- By the end of this intensive program, students will emerge with a refined skill set, enabling them to harness Big Data technologies adeptly, analyze data on a massive scale, and architect data systems primed for real-world challenges. Graduates will be primed to meet the burgeoning industry demand for skilled Big Data professionals, positioning them as invaluable assets in our data-driven landscape.

Programme

Contenu

The STC will cover the following topics:

- The big picture: tech megatrends.
- Data modelling: Data vs data representation; Structured vs unstructured data; Relational data model; Semi-structured data models; Examples: csv, json, xml etc.; Graph data models; Data model vs data format; Data streams; Batch vs stream processing.
- Characteristics of big data: The 3 (5) Vs, Big data vs Small data; Getting value out of big data, Big data strategy.
- Big data management systems: Relational DBs; No-SQL DBs.
- Storing big data: HDFS; Data warehouse; Data lake; Object storage.
- Big data retrieval: Querying SQL; Querying JSON; SPARQL.
- Big data ingestion: Ingestion infrastructure; Message queues; Pub/Sub; MQTT; Apache Kafka.
- Batch processing: MapReduce; Apache Spark.
- Stream processing: Spark Streaming; Apache Flink.

Modalités de validation

- Contrôle continu
- Projet(s)

Description des modalités de validation

Written quiz; a project assignment to perform after the STC execution will also be evaluated.