Big Data: Practical methods and solutions for data analysis

Hands-on course of 5 days - 35h
Ref.: BID - Price 2025: 3 610 (excl. taxes)
The price for the 2026 session dates may be revised

EDUCATIONAL OBJECTIVES

At the end of the training, the trainee will be able to:

Understand the concepts and benefits of Big Data with respect to business challenges

Understand the technological ecosystem needed to carry out a Big Data project

Acquire the technical skills to manage massive, unstructured, complex data flows

Implement statistical analysis models to address business needs

Learn about a data visualization tool for reporting dynamic analyses

HANDS-ON WORK

Set up a Hadoop platform and its basic components, use an ETL to manage the data, create analysis modules and dashboards.

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more.
Participants also complete a placement test before and after the course to measure the skills they've developed.

TEACHING AIDS AND TECHNICAL RESOURCES

• The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
• At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
• A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@ORSYS.fr to review your request and its feasibility.

THE PROGRAMME

last updated: 08/2024

1) Understanding the concepts and challenges of Big Data

- Origins and definition of Big Data.
- Key figures in the international and French markets.
- The challenges of Big Data: ROI, organization, data privacy.
- An example of Big Data architecture.

2) Big Data technologies

- Description of the architecture and components of the Hadoop platform.
- Storage methods (NoSQL, HDFS).
- Operating principles of MapReduce, Spark, Storm, etc.
- Most popular distributions on the market (Hortonworks, Cloudera, MapR, Elastic Map Reduce, Biginsights).
- Installing a Hadoop platform.
- Technologies for the data scientist.

Exercise : Exercise

3) Installing a Hadoop Big Data platform (via Cloudera Quickstart or other software).

- Operating principles of the Hadoop Distributed File System (HDFS).
- Importing outside data into HDFS.
- Creating SQL requests with HIVE.
- Using PIG to process the data.
- Using an ETL to industrialize the creation of massive data flows.
- Overview of Talend For Big Data.

Exercise : Operating principles of the Hadoop Distributed File System (HDFS).

4) Importing outside data into HDFS.

- Creating SQL requests with HIVE.
- Using PIG to process the data.

- The principle of ETL (Talend, etc.).
- Managing massive data streaming (NIFI, Kafka, Spark, Storm, etc.)
Exercise : Implementing massive data flows

5) Big Data Analytics techniques and methods
- Machine Learning: A component of artificial intelligence.
- Discovering the three families: Regression, Classification, and Clustering.
- Data preparation, feature engineering.
- Generating models in R or Python.
- Ensemble Learning.
Exercise : Exercise

6) Setting up analyses with the tools studied.
- Takeaways.
- Summary of best practices.
- Bibliography.
Defining the data visualization need.

DATES

REMOTE CLASS
2025 : 08 déc.

2026 : 16 mars, 18 mai, 20 juil.,
16 nov.