Updated 05/30/2024

Sign up

Architectures and Infrastructures for Big Data training course

3 days (21 hours)

Presentation

Big Data plays a key role in the smooth running of your business. It helps you achieve rapid growth by analyzing large volumes of data in real time.

Traditional database management systems are no longer capable of handling vast quantities of data. As a result, it's essential to adapt your ecosystem's architecture to Big Data in order to ingest, process and analyze masses of data in record time.

Big Data technology enables you to identify weak points to correct and opportunities to seize for your company. You'll gain a competitive edge by making relevant decisions faster, based on concrete information.

Our Big Data architecture and infrastructure training course will introduce you to the main concepts of Big Data, the technological ecosystem of a Big Data project, security and legal issues. You'll learn how to determine the nature of the data and exploit its architecture.

At the end of this course, you'll be able to initiate the design of a Big Data architecture and infrastructure, with an overview of the different solutions dedicated to processing massive amounts of data.

Objectives

- Master the essential concepts of Big Data and the technological ecosystem of a Big Data project.
- Analyze the challenges of a Big Data project
- Identify the nature of the data handled
- Safety, ethics and legal issues

- Exploiting Big Data architectures
- Setting up complete technical foundations for Big Data projects

Target audience

- Project managers
- Architects
- Developers
- Data scientists
- Anyone wishing to learn how to design a Big Data architecture

Prerequisites

- Have a good general knowledge of information systems
- Basic knowledge of relational models, statistics and programming languages

Program of our Big Data architecture and infrastructure training course

Introduction to Big Data

- The main concepts of Big Data
- From data to value creation
- Adding value
- Project ecosystem
- Big Data players and their positioning
- Big Data opportunities

Big Data architectures

- Big Data calculation models
- Data management steps
- Publication and subscription systems
- Continuous processing
- Advanced architecture
- Industrial approaches
- Actor- and agent-based Big Data architectures

Big Data solutions

- Projects, applications and platforms
- Data storage

- Hardware considerations :
  - CPU
  - Network
  - Memory
  - Distributed systems
- Predictive analysis

Big Data functionalities

- Recommendation systems
- Search
- Ad-tech algorithms and Real Time Bidding (RTB)
- Inter-device graph generation
- Forecasting and prediction systems
- Social Media Big Data
- Detecting anomalies and fraud
- Control and monitoring of smart grids

Legal issues

- Ethical rules and legal issues
- Securing personal data
- Prohibition on collection: sensitive information
- Intra-country agreements

Distributed architectures

- Challenges and opportunities of a distributed architecture
- Massively parallel architecture
- Types of distributed computing paradigms
- Complex data processing
  - Machine Learning
  - Datamining
- NoSQL with distributed computing

Technical and application architecture

- The main differences
- Data management steps
- The different tools on the market
- Data injection
- Data storage with Hadoop, HDFS, NoSQL
- Restitution and visualization
- Data capture
- Data distribution
- Real-time processing
- Platform supervision (Ambari, App Dynamix)

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level on different types of technology, as well as his or her expectations and personal objectives with regard to the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.