

Course : Hadoop, developing applications for Big Data

Practical course - 4d - 28h00 - Ref. APH

Price : 2960 CHF E.T.

This hands-on course will teach you how to develop applications that enable you to process distributed data in batch mode. You'll collect, store and process data in heterogeneous formats with Apache Hadoop, to set up processing chains integrated with your information system.

Teaching objectives

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

- ✓ Building a MapReduce-based program
- ✓ Integrating Hadoop HBase into an enterprise workflow
- ✓ Travailler avec Apache Hive et Pig depuis Hadoop Distributed File System (HDFS)
- ✓ Using a task graph with Hadoop

Intended audience

Concepteurs, développeurs.

Prerequisites

Good experience in Java development. Knowledge of web architecture a plus.

Practical details

Hands-on work

Application development for Big Data.

Teaching methods

Lectures 30%, practical work 70%.

Course schedule

PARTICIPANTS

Concepteurs, développeurs.

PREREQUISITES

Good experience in Java development. Knowledge of web architecture a plus.

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.

1 Big data

- Defining the scope of big data.
- The role of the Hadoop project.
- Basic concepts of big data projects.
- Introduction to cloud computing.
- The difference between private and public cloud computing.
- Big data architectures based on the Hadoop project.

Demonstration

Use of Hadoop and GoogleApp.

2 Collecting data and applying MapReduce

- Analysis of company data flows.
- Structured and unstructured data.
- The principles of semantic analysis of enterprise data.
- MapReduce-based task graph.
- Data consistency granularity.
- Transfer data from a persistence system to Hadoop.
- Transferring data from a Cloud to Hadoop.

Hands-on work

Managing the collection of customer information using MapReduce.
Configuring the Yarn implementation. Developing a Map Reduce-based task.

3 Data storage with HBase

- Several types of XML database.
- Usage patterns and their application to the cloud.
- Application of Hadoop database within a workflow.
- Using Hive/Pig projects.
- Using the HCatalog project.
- HBase Java API.

Hands-on work

Manage modifications to a supplier data catalog.

4 Data storage on HDFS

- Usage patterns and their application to the cloud.
- Architecture and installation of an HDFS system, journal, NameNode, DataNode.
- Operations, orders and order management.
- Java HDFS API.
- Data analysis with Apache Pig.
- The Pig Latin language. Using Apache Pig with Java.
- Querying with Apache Hive.
- Data replication. Data sharing on HDFS architecture.

Hands-on work

Administering a shared client repository on Hadoop. Using the visualization console.

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.

5 Spring Data Hadoop

- Introduction to Spring and Spring Data.
- The Hadoop namespace for Spring.
- Using Spring to simplify Hadoop configuration.
- Distributed cache configuration.
- Job definition and dependencies between jobs.
- Integration of tools (Pig, Hive...).

Hands-on work

Redesign of supplier data catalog management using Spring Data.

Dates and locations

REMOTE CLASS

2026 : 26 May, 6 Oct.