

Course : Flink, developing applications for Big Data

Practical course - 3d - 21h00 - Ref. FKB

Price : 2010 € E.T.

Apache Flink is a recent big data framework. It simplifies the processing of large real-time flows as well as batch processing on huge quantities of data (on Hadoop HDFS, on Amazon S3, on MongoDB...). This course will enable you to install Flink and carry out a variety of big data processes in Java.

Teaching objectives

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

- ✓ Master the fundamental concepts of Flink
- ✓ Develop applications using DataSet and DataStream APIs
- ✓ Distributed data processing with Flink and Hadoop
- ✓ Exploiting data with Table API
- ✓ A first approach to machine learning

Intended audience

Developers, architects.

Prerequisites

Good knowledge of Java.

Practical details

Hands-on work

Practical application of the concepts covered in the course using the Java language.

Course schedule

PARTICIPANTS

Developers, architects.

PREREQUISITES

Good knowledge of Java.

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

1 Introduction to Apache Flink

- History of the framework.
- The different versions of Flink.
- Comparison with the Apache Hadoop and Apache Spark environments.
- The different Flink modules.

Hands-on work

Install and configure Flink. Run a first example with word counting.

2 Data processing with the DataStream API

- Runtime environment and data sources.
- Transformations: Map, FlatMap, Filter, KeyBy, Reduce...
- Operations on multiple flows: Union, Cogroup, Connect, Join, Iterate...
- Windows operations: Global, Tumbling, Sliding, Session...
- Customized physical partitioning, randomization, rebalancing and resizing.
- DataSink and connectors: Kafka, X (formerly Twitter), Elasticsearch...

Hands-on work

Consumption and handling of different data streams.

3 Data processing with the Batch API

- The different types of data sources.
- Transformations and aggregations.
- Data writing.
- DataSink and connectors: HDFS, S3, Avro, MongoDB.

Hands-on work

Manipulate DataSets from multiple data sources.

4 Data processing with the Table API

- Save and read saved tables.
- Operators: selection, filter, join, orderBy...
- Use SQL on the data stream.
- Handling complex events.

Hands-on work

Set up an analysis with SQL on a data stream.

5 API Flink Graph - Gelly

- What is a graph?
- The different operations.
- Create graphs.
- Graph transformations.
- Presentation of different algorithms.

Hands-on work

Handling the API through various examples.

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.

6 Deploying Flink

- Flink on YARN Configurations.
- Start and stop a cluster.
- Submit a job to Flink.
- Flink on Google Cloud.
- Flink on AWS.

Hands-on work

Configure a multi-node cluster and deploy an application.