

# Opleiding : Spark Java, toepassingen ontwikkelen voor Big Data

*Praktijkcursus - 3d - 21u00 - Ref. SPK*

*Prijs : 2010 € V.B.*



4,7 / 5

Spark wordt vaak voorgesteld als de opvolger van Hadoop. Spark vereenvoudigt het programmeren van big data processing en maakt het gebruik van Scala, Python of Java mogelijk. Deze cursus leert programmeurs hoe ze een gegevensstroom in realtime kunnen verwerken en batchverwerking kunnen uitvoeren (van SQL tot machine learning).



## Pedagogische doelstellingen

Aan het einde van de training is de deelnemer in staat om:

- ✓ Beheers de fundamentele concepten van Spark
- ✓ Toepassingen ontwikkelen met Spark Streaming
- ✓ Een Spark-cluster implementeren
- ✓ Gegevens exploiteren met Spark SQL
- ✓ Een eerste benadering van machinaal leren

## Doelgroep

Projectmanagers, datawetenschappers, ontwikkelaars, architecten.

## Voorafgaande vereisten

Goede kennis van Java. Kennis van big data.

## Praktische modaliteiten

### Praktisch werk

Praktische toepassing van de concepten die in de cursus worden behandeld met behulp van de Java-taal.

## Opleidingsprogramma

### DEELNEMERS

Projectmanagers, datawetenschappers, ontwikkelaars, architecten.

### VOORAFGAANDE VEREISTEN

Goede kennis van Java. Kennis van big data.

### VAARDIGHEDEN VAN DE CURSUSLEIDER

De deskundigen die de cursus leiden zijn specialisten op het betreffende vakgebied. Zij werden geselecteerd door onze pedagogische teams zowel om hun vakkennis als hun pedagogische vaardigheden voor elke cursus die zij geven. Zij hebben minstens vijf tot tien jaar ervaring in hun vakgebied en oefenen of oefenden verantwoordelijke bedrijfsfuncties uit.

### BEOORDELINGSMODALITEITEN

De cursusleider beoordeelt de pedagogische vooruitgang van de deelnemer gedurende de gehele cursus aan de hand van meerkeuzevragen, praktijksituaties, praktische opdrachten, ... De deelnemer legt ook van tevoren en naderhand een test af ter bevestiging van de verworven kennis.

## 1 Inleiding tot Apache Spark

- Geschiedenis van het kader.
- De verschillende versies van Spark (Scala, Python en Java).
- Vergelijking met de Apache Hadoop-omgeving.
- De verschillende Spark modules.

### Praktisch werk

Spark installeren en configureren. Een eerste voorbeeld met woordentelling uitvoeren.

## 2 Programmeren met veerkrachtige gedistribueerde datasets (RDD)

- Presentatie RDD.
- RDD's maken, verwerken en hergebruiken.
- Accumulatoren en uitzendvariabelen.
- Gebruik partities.

### Praktisch werk

Het manipuleren van verschillende datasets met behulp van RDD's en het gebruik van de API van Spark.

## 3 Gestructureerde gegevens verwerken met Spark SQL

- SQL, DataFrames en datasets.
- De verschillende soorten gegevensbronnen.
- Interoperabiliteit met RDD's.
- Spark SQL prestaties.
- JDBC/ODBC-server en Spark SQL CLI.

### Praktisch werk

Dataset manipulatie via SQL queries. Verbinding met een externe database via Java DataBase Connectivity (JDBC) of Open Database Connectivity (ODBC).

## 4 Spark op een cluster

- De verschillende soorten architectuur: standalone, Apache Mesos of Hadoop YARN.
- Een cluster configureren in standalone modus.
- Een applicatie verpakken met zijn afhankelijkheden.
- Toepassingen implementeren met Spark-submit.
- Een cluster dimensioneren.

### Praktisch werk

Een Spark-cluster opzetten.

## 5 Real-time analyse met Spark Streaming

- Hoe het werkt.
- Presentatie van gediscrètiseerde streams (DStreams).
- De verschillende soorten bronnen.
- Omgaan met de API.
- Vergelijking met Apache Storm.

### Praktisch werk

Loggebruik met Spark Streaming.

### PEDAGOGISCHE EN TECHNISCHE MIDDELEN

- De gebruikte pedagogische middelen en cursusmethoden zijn voornamelijk: audiovisuele hulpmiddelen, documentatie en cursusmateriaal, praktische oefeningen en correcties van de oefeningen voor praktijkstages, casestudies of reële voorbeelden voor de seminars.
- Na afloop van de stages of seminars verstrekt ORSYS de deelnemers een evaluatievragenlijst over de cursus die vervolgens door onze pedagogische teams wordt geanalyseerd.
- Na afloop van de cursus wordt een presentielijst per halve dag verstrekt, evenals een verklaring van de afronding van de cursus indien de stagiair alle sessies heeft bijgewoond.

### TOEGANGSMODALITEITEN EN TERMIJNEN

De inschrijving dient 24 uur voor aanvang van de cursus plaatsgevonden te hebben.

### TOEGANKELIJKHEID VOOR MINDERVALIDEN

Is voor u speciale toegankelijkheid vereist? Neem contact op met mevr. FOSSE, contactpersoon voor mindervaliden, via het adres psh-accueil@ORSYS.fr om uw verzoek en de haalbaarheid daarvan zo goed mogelijk te bestuderen.

## 6 Grafieken verwerken met GraphX

- Maak kennis met GraphX.
- De verschillende operaties.
- Grafieken maken.
- Vertex en Edge RDD.
- Presentatie van verschillende algoritmen.

### Praktisch werk

Omgaan met de GraphX API aan de hand van verschillende voorbeelden.

## 7 Machinaal leren met Spark

- Inleiding tot machinaal leren.
- De verschillende klassen van algoritmen.
- Inleiding tot SparkML en MLlib.
- Implementaties van de verschillende algoritmen in MLlib.

### Praktisch werk

Gebruik van SparkML en MLlib.

## Data en plaats

### KLAS OP AFSTAND

2026 : 15 juni, 14 sep., 23 nov.

### PARIS LA DÉFENSE

2026 : 15 juni, 14 sep., 23 nov.