

Opleiding : Spark Gevorderd

Machine Learning en industrialisatie van analytische workflows

Praktijkcursus - 3d - 21u00 - Ref. SPN

Prijs : 2010 € V.B.

Spark is een gedistribueerd rekenraamwerk dat complexe Big Data verwerking en analyse mogelijk maakt. Als je Spark al hebt gebruikt, willen we je analyses graag een stap verder brengen met machine learning en je kennis laten maken met MLOps voor het implementeren en industrialiseren van analytische modellen.



Pedagogische doelstellingen

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

- ✓ Leer geavanceerde gegevensanalyse met Spark
- ✓ Machinaal leren (ML) verwerken met Spark
- ✓ Inzicht in Docker en hoe het kan worden gebruikt om analytische workflows te industrialiseren
- ✓ Uitwerken en implementeren van de stadia van de analytische cyclus met Spark
- ✓ Leer hoe u de analyseworkflow kunt industrialiseren
- ✓ Ontdek MLOps

Doelgroep

Professionals die Spark willen gebruiken voor batch en real-time analytics.

Voorafgaande vereisten

Connaissances des API Spark, notamment RDD et DataFrame. Connaissances des algorithmes d'apprentissage supervisés et non supervisés. Maîtrise d'un des langages suivants : Scala, Python.

Praktische modaliteiten

Praktisch werk

Afwisselend theorie en praktijk. 60% oefeningen voor meer diepgang. Praktische feedback.

Opleidingsprogramma

DEELNEMERS

Professionals die Spark willen gebruiken voor batch en real-time analytics.

VOORAFGAANDE VEREISTEN

Connaissances des API Spark, notamment RDD et DataFrame. Connaissances des algorithmes d'apprentissage supervisés et non supervisés. Maîtrise d'un des langages suivants : Scala, Python.

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 vak kennis 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

- Een herinnering aan de Spark API.
- Docker-concepten en hun gebruik in gegevensanalyse.
- Docker-containers.

Praktisch werk

De werkomgeving onder de knie krijgen, Docker-containers maken.

2 De analytische cyclus met Spark

- Invoer van gegevens.
- Verkenning.
- Voorbereiding van gegevens.
- Leerplaatsen.
- Industrialisatie.

Workshop storytelling

Presentatie van casestudies en bespreking van de verschillende stadia in de cyclus.

3 Invoer van gegevens.

- Laden van gegevens.
- Batchverwerking.
- Streamingbehandelingen.
- Gegevensindelingen: afbeeldingen, binair, gestructureerd, Graph...

Praktisch werk

Gegevens laden vanuit verschillende bronnen.

4 Datamining

- Beschrijvende statistieken.
- Uitschieters en lege gegevens identificeren.
- Ongeldige waarden en andere afwijkingen identificeren.

Praktisch werk

Anomalieën in een dataset identificeren.

5 Voorbereiding en feature engineering (datatransformatieproces)

- Opschonen van gegevens.
- Pijpleidingen.
- Transformer les valeurs numériques, catégoriques, binaires et texte.
- Création de nouvelles features.
- Réduction de dimensions.
- Vectorisation.

Praktisch werk

Bereid gegevens voor op analyse.

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 ML-levenscyclus met MLflow

- Levenscyclus van een machine-leerproject.
- Presentatie van het MLflow open source platform.
- De belangrijkste onderdelen van MLflow: Volgen, Modellen en Projecten.
- Parameters, metriek, tags en artefacten.

Praktisch werk

Een machine-learningproject maken en gebruiken.

7 Machinaal leren

- MLlib, Spark's machine learning bibliotheek en de beschikbare algoritmes.
- Een dataset opsplitsen.
- Configureer een model en voer het uit.
- Interpretatie en validatie van leerresultaten.
- Inleiding tot Spark Streaming.

Praktisch werk

Machinaal leren implementeren.

8 Praktijkvoorbeelden

- Doe aanbevelingen.
- Verkoopvoorspellingen maken.
- Semantische analyse.
- Computer vision met Spark en PyTorch.
- Analyse temps réel avec Spark et Kafka.

Casestudy

Voer de verschillende voorgestelde casestudies uit.

Data en plaats

KLAS OP AFSTAND

2026 : 18 mei, 28 sep.

PARIS LA DÉFENSE

2026 : 18 mei, 28 sep.