

Opleiding : Big Data in realtime verwerken met Spark en Storm

Praktijkcursus - 3d - 21u00 - Ref. DSS

Prijs : 2010 € V.B.

Big Data, bekend om zijn vermogen om enorme hoeveelheden gegevens te verwerken, bevat nu een real-time component met de komst van tools zoals Spark en Storm. Je leert over de voordelen van deze tools, hun real-time gedistribueerde computersystemen en het concept van Streaming Big Data.

Pedagogische doelstellingen

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

- ✓ De grondbeginselen van real-time big data applicatieontwikkeling begrijpen
- ✓ Spark en Storm evalueren
- ✓ Toepassing van Storm en Spark real-time gedistribueerde computersystemen
- ✓ Grote hoeveelheden gegevens in realtime verwerken

Doelgroep

Ontwerpers, ontwikkelaars, architecten.

Voorafgaande vereisten

Goede kennis van softwareontwikkeling. Begrip van big data is een pré.

Opleidingsprogramma

1 Inleiding tot real-time architectuur

- Real-time verwerking.
- Lambda-architecturen.
- Kappa-architecturen.
- SMACK-architecturen.

Praktisch werk

Onderzoek naar de implementatie van een Kappa architectuur voor Spark en Storm.

DEELNEMERS

Ontwerpers, ontwikkelaars, architecten.

VOORAFGAANDE VEREISTEN

Goede kennis van softwareontwikkeling. Begrip van big data is een pré.

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.

2 Kafka-architectuur

- Het overzicht van Kafka Producenten, Makelaars, Consumenten.
- Kafka's logbestanden.
- Avro schema's. Met behulp van ZooKeeper.

Praktisch werk

Studie van de Kafka-configuratie in de Kappa architectuur.

3 Apache Storm-architectuur

- Definitie van de ontwikkelomgeving.
- Creatie van Storm-gebaseerde projecten.
- Definitie van Stormonderdelen (Uitloop en Bout).
- Definitie van Stormstromen.
- Gegevensmodel (sleutel, waarde).
- De rol van Nimbus en ZooKeeper.

Casestudy

Studie van de implementatie van de Kappa architectuur voor Storm.

4 Afhandeling van Storm-berichten

- Programmeerdiensten met Clojure, Java, Python.
- Levenscyclus van berichten.
- De Storm API voor het definiëren van betrouwbaarheid.
- Betrouwbaarheidsimplementatiestrategie voor een toepassing die gebruikmaakt van Big Data.

Praktisch werk

Implementatie van een real-time sociaal netwerkverwerkingsproject in de Kappa architectuur.

5 Apache Spark-architectuur

- De verschillende versies van Spark (Scala, Python, R en Java).
- Vergelijking met de Storm-omgeving.
- De verschillende Spark modules.
- De verschillende soorten architectuur: Standalone, Apache Mesos of Hadoop YARN.

Praktisch werk

Studie van de implementatie van de SMACK architectuur voor Spark.

6 Real-time met Spark Streaming

- Presentatie van veerkrachtige gedistribueerde dataset (RDD)?
- RDD's maken, verwerken en hergebruiken.
- Accumulatoren en uitzendvariabelen.
- Hoe het werkt.
- De verschillende soorten bronnen.
- Vergelijking met Apache Storm.

Praktisch werk

Implementatie van een project om sociale netwerken in realtime te verwerken.

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.

7 Andere marktspelers

- Vergelijking van alle streamingtools in het ecosysteem (Storm, Spark Streaming, Flink, Samza).
- Focus op Samza-architectuur.

Praktisch werk

Onderzoek naar de implementatie van de Kappa architectuur met Samza.