

Opleiding : Hadoop, installatie en beheer

Praktijkcursus - 4d - 28u00 - Ref. HOD

Prijs : 2520 € V.B.



Het Apache Hadoop platform was de eerste oplossing die (gedistribueerde) verwerking van enorme hoeveelheden gegevens echt mogelijk maakte. Deze cursus laat zien hoe je een Hadoop-cluster en andere componenten van het ecosysteem (Hive, Pig, HBase, Flume, enz.) installeert, configureert en beheert.

Pedagogische doelstellingen

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

- ✓ Ontdek de concepten en uitdagingen van Hadoop
- ✓ Begrijpen hoe het platform en zijn onderdelen werken
- ✓ Het platform installeren en beheren
- ✓ Het platform optimaliseren

Doelgroep

Hadoop clusterbeheerders, ontwikkelaars.

Voorafgaande vereisten

Goede kennis van Linux administratie. Vereiste ervaring.

Praktische modaliteiten

Praktisch werk

Installatie en configuratie van het Hadoop-cluster.

Leer methodes

Lesmethode " met praktische oefeningen na elk uitgelegd concept of groep concepten.

Opleidingsprogramma

DEELNEMERS

Hadoop clusterbeheerders, ontwikkelaars.

VOORAFGAANDE VEREISTEN

Goede kennis van Linux administratie. Vereiste ervaring.

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 Presentatie van het Apache Hadoop-raamwerk

- De uitdagingen van big data en de voordelen van het Hadoop framework.
- Inleiding tot de Hadoop-architectuur.
- Beschrijving van de belangrijkste onderdelen van het Hadoop-platform.
- Presentatie van de belangrijkste distributies op de on-premise en on-cloud markt, en de hybride aanpak.
- Voordelen/nadelen van het platform ten opzichte van alternatieve oplossingen.
- Overzicht van de verschillende native en aanvullende componenten en vergelijkingen (Storm, Flink, Spark, enz.).

2 Het Hadoop-cluster voorbereiden en configureren

- Bedieningsprincipes van het Hadoop Distributed File System (HDFS).
- Werkingsprincipes van MapReduce.
- Ontwerp "type" van het cluster.
- Criteria voor apparatuurselectie.

Praktisch werk

Het Hadoop-cluster configureren.

3 Een Hadoop-platform installeren

- Type inzet.
- Hadoop installeren.
- Installatie van andere componenten (Hive, Pig, HBase, Nifi, enz.).
- Presentatie en vergelijking van historische (HDP, HDF, CDH) en huidige (CDP/CDSW...) softwarestacks.
- Kappa, Lambda, SMACK-architecturen (Spark, Mesos, Akka, Cassandra, Kafka).

Praktisch werk

Een Hadoop-platform en de belangrijkste onderdelen ervan installeren.

4 Een Hadoop-cluster beheren

- Hadoop-clusternodes beheren.
- MapReduce V2 (Yarn, Resource Manager, Application Master, Node Manager, enz.).
- Resource managers (Yarn versus Mesos).
- Taakbeheer via schedulers.
- Logboekbeheer.
- Procesplanning (Oozie).
- Gebruik een manager.

Praktisch werk

Taken weergeven, wachtrijstatus, taakstatus, taakbeheer, WebUI-toegang.

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.

5 Gegevensbeheer in HDFS

- Externe gegevens (bestanden, relationele databases) importeren in HDFS.
- Omgaan met HDFS bestanden.
- Bestandsformaten (SequenceFile, ORC, Parquet, etc.), hun gebruik en hun relatie tot prestaties.
- Databaseopslag (gestructureerd of ongestructureerd): NoSQL Hbase, SQL met Impala, Hive, Hive LLAP, enz.

Praktisch werk

Importeer externe gegevens met Flume of Nifi, importeer gegevens uit relationele databases met Sqoop.

6 Geavanceerde configuratie

- Autorisaties en beveiliging: administratie, authenticatie, autorisaties, audit, gegevensbescherming.
- Componenten betrokken bij beveiliging: Ranger, Knox, Kerberos, KMS, etc.
- NameNode hoge beschikbaarheid (MRV2/YARN).

Praktisch werk

Configuratie van een authenticatie op serviceniveau (SLA) en een toegangscontrolelijst (ACL).

7 Bewaking en optimalisatie/tuning

- Monitoring (Ambari, Cloudera Manager, Ganglia, enz.).
- Soorten benchmark (DFSIO, Teragen/TeraSort/TeraValidate) en online beschikbare resultaten (TPCx-HS, etc.)
- Vergelijking tussen MapReduce en TEZ.
- Voorbeelden van optimalisatie en optimalisatietools (CDP-adviseur, enz.).
- Keuze van blok grootte.
- Andere afstelopties (gebruik van compressie, geheugenconfiguratie, enz.).

Praktisch werk

Het opzetten, starten en analyseren van benchmarks en het begrijpen van nieuwe opdrachten voor clustermonitoring en optimalisatie.

8 De voordelen van Hadoop v3

- Objectopslagbenaderingen (Ozone).
- Erasure-codering.
- Federatie van garen.
- Migratiescenario's, aspecten om rekening mee te houden en een paar voorbeelden (Hortonworks naar Cloudera...).

Data en plaats

KLAS OP AFSTAND
2026 : 23 juni, 22 sep.

PARIS LA DÉFENSE
2026 : 23 juni, 22 sep.