

Big Data, methoden en praktische oplossingen voor gegevensanalyse

Praktijkcursus van 5 dagen - 35u
Ref : BID - Prijs 2024 : € 3 530 excl. BTW

Deze cursus stelt u in staat om de uitdagingen en voordelen van Big Data te begrijpen, evenals de technologieën om deze te implementeren. U leert hoe u enorme hoeveelheden gestructureerde en ongestructureerde gegevens kunt integreren met behulp van een ETL en vervolgens kunt analyseren met behulp van statistische modellen en dynamische dashboards.

PEDAGOGISCHE DOELSTELLINGEN

Na afloop van de opleiding kan de cursist:

De concepten en de bijdrage van Big Data aan zakelijke uitdagingen begrijpen

Het technologische ecosysteem begrijpen dat nodig is voor een Big Data-project

De technische vaardigheden verwerven die nodig zijn om complexe, ongestructureerde en enorme gegevensstromen te beheren

Statistische analysemodellen implementeren om te voldoen aan bedrijfsbehoeften

Leren hoe u een datavisualisatietool gebruikt om dynamische analyses te maken

EXERCISE

Een Hadoop-platform en de basiscomponenten ervan opzetten, een ETL gebruiken om gegevens te beheren, analysemodellen en dashboards maken.

HET PROGRAMMA

laatste update: 10/2021

1) De concepten en uitdagingen van Big Data begrijpen

- Oorsprong en definitie van Big Data.
- Belangrijke marktcijfers wereldwijd en in Frankrijk.
- De uitdagingen van Big Data: ROI, organisatie, betrouwbaarheid van gegevens.
- Een voorbeeld van Big Data architectuur.

2) Big Data-technologieën

- Beschrijving van de architectuur en componenten van het Hadoop-platform.
- Opslagmodi (NoSQL, HDFS).
- Werkingsprincipes van MapReduce, Spark, Storm...
- Belangrijkste marktdistributies (Hortonworks, Cloudera, MapR, Elastic Map Reduce, Biginsights).
- Een Hadoop-platform installeren.
- Technologieën voor datawetenschappers.
- Presentatie van specifieke technologieën voor Big Data (Tableau, Talend, Qlikview, enz.).

Installatie van een Big Data Hadoop-platform (via Cloudera QuickStart of een ander platform).

3) Gestructureerde en ongestructureerde gegevens beheren

- Bedieningsprincipes van het Hadoop Distributed File System (HDFS).

DEELNEMERS

Gegevensanalisten, statistische onderzoekers, ontwikkelaars, projectmanagers, business intelligence consultants.

VOORAFGAANDE VEREISTEN

Basiskennis van relationele modellen, statistiek en programmeertalen. Basiskennis van Business Intelligence concepten.

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.

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.

- Externe gegevens importeren in HDFS.
- SQL-query's uitvoeren met HIVE.
- Gebruik PIG om de gegevens te verwerken.
- Het ETL-principe (Talend, enz.).
- Beheer van massale datastreaming (NIFI, Kafka, Spark, Storm...)

Implementatie van massale gegevensstromen.

4) Technieken en methoden voor big data-analyse

- Machine Learning, een onderdeel van kunstmatige intelligentie.
- Ontdek de drie families: Regressie, Classificatie en Clustering.
- Datavoorbereiding en feature engineering.
- Genereer modellen in R of Python.
- Ensemble leren.
- Ontdek de tools op de markt: Jupyter Notebook, Dataiku, Amazon Machine Learning, enz.

Zet analyses op met een van de bestudeerde tools.

5) Datavisualisatie en praktijkvoorbeelden

- Definieer de behoefte aan datavisualisatie.
- Gegevensanalyse en visualisatie.
- Kunnen alle soorten gegevens worden gebruikt in DataViz?
- DataViz-tools op de markt.

Installatie en gebruik van een tool voor gegevensvisualisatie om dynamische analyses te maken.

6) Conclusie

- Dingen om te onthouden.
- Samenvatting van best practices.
- Bibliografie.

DATA

KLAS OP AFSTAND
2024 : 01 jul, 18 nov

BRUSSEL
2024 : 01 jul, 18 nov