

Big data, een stand van zaken

Seminar van 2 dags - 14h

Ref : BGA - Prijs 2025 : 2 140 excl. BTW

De voortdurende toename van digitale gegevens in organisaties heeft tot de opkomst van big data geleid. Dit concept doet vragen rijzen over de opslag en bewaring van grote hoeveelheden gegevens, maar ook over de potentiële waarde van deze massa's gegevens. In het kader van dit seminarie worden de specifieke problemen van big data en de mogelijke technische oplossingen behandeld, van gegevensbeheer tot de verschillende soorten verwerking.

PEDAGOGISCHE DOELSTELLINGEN

Na afloop van de opleiding kan de cursist:

- Kennismaken met de belangrijkste concepten van big data
- Economische uitdagingen identificeren
- Voor- en nadelen van big data beoordelen
- De belangrijkste problemen en mogelijke oplossingen begrijpen
- De belangrijkste methoden en toepassingsgebieden van big data te identificeren

HET PROGRAMMA

laatste update: 08/2024

1) Inleiding

- De oorsprong van big data: een wereld van digitale gegevens, e-health, chronologie.
- Een definitie aan de hand van de vier V's: de herkomst van gegevens.
- Een breuk: veranderingen in kwantiteit, kwaliteit, gewoonten.
- De waarde van gegevens: een verandering in belang.
- Gegevens als grondstof.
- Het vierde paradigma van de wetenschappelijke ontdekking.

2) Big data: verwerking van de verzameling tot het resultaat

- Stroom van de acties. Verzameling.
- Verzameling van gegevens: crawling, scraping.
- Beheer van gebeurtenisstromen (Complex Event Processing (CEP)).
- Indexering van de inkomende stroom.
- Integratie in oude gegevens.
- Gegevenskwaliteit: een vijfde V?
- De verschillende soorten verwerking: zoeken, leren (Machine Learning, transactieverwerking, datamining).
- Andere stroommodellen: Amazon, e-Health.
- Een of meer gegevensbibliotheken? Van Hadoop tot in-memory.
- Van toonanalyse tot kennisontdekking.

3) Relaties tussen cloud en big data

- Architectuurmodel van openbare en privéclouds.
- XaaS-services.
- Doelstellingen en voordelen van cloudarchitecturen.
- Infrastructuren.
- Gelikheden en verschillen tussen cloud en big data.

DEELNEMERS

IS-afdelingen, technisch directeurs, projectleiders, architecten, IS-managers.

VOORAFGAANDE VEREISTEN

Basiskennis van technische architecturen.

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.

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.

- Opslagclouds.
- Classificatie, beveiliging en vertrouwelijkheid van gegevens.
- Structuur als classificatiecriterium: ongestructureerd, gestructureerd, halfgestructureerd.
- Classificatie volgens levenscyclus: tijdelijke of permanente gegevens, actieve archieven.
- Moeilijkheden op het gebied van beveiliging: toenemende volumes, distributie.
- Mogelijke oplossingen.

4) Inleiding tot open data

- Filosofie van open gegevens en doelstellingen.
- Vrijgave van openbare gegevens.
- Implementatiemoeilijkheden.
- Essentiële kenmerken van open gegevens.
- Toepassingsgebieden. Verwachte voordelen.

5) Hardware voor opslagarchitecturen

- Servers, schijven, netwerken en gebruik van SSD-schijven, belang van netwerkinfrastructuur.
- Cloudarchitecturen en traditionelere architecturen.
- Voordelen en moeilijkheden.
- TCO. Stroomverbruik: servers (IPNM), schijven (MAID)
- Objectopslag: principe en voordelen.
- Objectopslag vergeleken met traditionele NAS- en SAN-opslag.
- Softwarearchitectuur.
- Indelingsniveaus van opslagbeheer.
- "Software Defined Storage".
- Gecentraliseerde architectuur (Hadoop File System).
- Peer-to-peerarchitectuur en gemengde architectuur.
- Interfaces en connectors: S3, CDMI, FUSE, enz.
- Toekomst van andere opslagmethoden (NAS, SAN) vergeleken met objectopslag.

6) Gegevensbescherming

- Bewaring in de tijd tegenover toenemende volumes.
- Back-ups maken, online of lokaal?
- Traditioneel archief en actief archief.
- Verbanden met opslaghiërarchiebeheer: toekomst van magnetische banden.
- Replicatie over meerdere locaties.
- Slechtere prestaties van opslagmedia.

7) Verwerkingsmethoden en toepassingsgebieden

- Classificatie van analysemethoden volgens gegevensvolume en verwerkingskracht.
- Hadoop: het Map Reduce-verwerkingsmodel.
- Hadoop-ecosysteem: Hive, Pig. Moeilijkheden van Hadoop.
- OpenStack en Ceph-gegevensbeheer.
- Complex Event Processing: een voorbeeld? Storm.
- Van BI tot big data.
- Vernieuwde Business Intelligence en transacties: NoSQL-databases. Typologie en voorbeelden.
- Gegevensopname en indexering. Twee voorbeelden: Splunk en Logstash.
- Open Source-crawlers.
- Zoeken en analyseren: Elasticsearch.
- Leren: Mahout. In-memory.
- Visualisatie: al dan niet in real time, in de cloud (Bime), vergelijking QlikView, Tibco Spotfire, Tableau.
- Een algemene dataminingarchitectuur via big data.

8) Gebruikscases aan de hand van voorbeelden en conclusie

- Verwachting: behoeften van gebruikers in bedrijven, onderhoud van uitrustingen.

- Beveiliging van personen, opsporing van fraude (post, belastingen), netwerk.
- Aanbeveling. Marketing- en impactanalyses.
- Routeanalyses. Distributie van video-inhoud.
- Big data voor de auto-industrie? Voor de aardolie-industrie?
- Moet u aan een big-dataproject beginnen?
- Welke toekomst voor gegevens?
- Governance van gegevensopslag: rol en aanbevelingen, Data Scientist, vaardigheden voor een big-dataproject.

DATA

KLAS OP AFSTAND
2025 : 30 sept., 16 déc.

BRUSSEL
2025 : 30 sept., 16 déc.