

POSTACADEMISCHE OPLEIDING

TRENDS IN DATABANKBEHEER

DE PARADIGMA-SHIFT NAAR AI EN DATA SCIENCE APPLICATIES

2 oktober 2023 – 18 december 2023



UNIVERSITEIT
GENT

In deze opleiding leer je hoe je de nieuwste databanktechnologie voor het beheer van 'big' data kan integreren in je huidige databanksystemen om zo je data beter te kunnen inzetten in nieuwe data-gedreven applicaties of AI-toepassingen. Je leert niet alleen de theoretische concepten, maar ook hoe je praktisch aan de slag kan.

Grotere datavolumes, nieuwe manieren van omgaan met data en almaar veranderende datanoden van nieuwe toepassingen en artificiële intelligentie maken dat het werken met databanksystemen de voorbije jaren een aantal ingrijpende veranderingen onderging.

De opleiding is gericht op de essentie van nieuwe NoSQL databanktechnologie en legt een grote nadruk op het gebruik en de inzetbaarheid van deze technologie in de praktijk.

We anticiperen ook op het toenemend belang om tekstuele data automatisch te kunnen gebruiken en te integreren in data-analyses en AI oplossingen.

DOELPUBLIEK

Iedereen met enige informatica-achtergrond en uitgesproken interesse voor databeheer.

Er wordt gewerkt met eigen laptop. Deze moet krachtig genoeg zijn (minimum 8GB RAM) en deelnemers moeten administratierechten hebben voor het installeren van de nodige programma's.

Het aantal deelnemers is beperkt tot 30.

Na deelname ontvangt u een aanwezigheidsattest.

WETENSCHAPPELIJKE COÖRDINATIE

Prof. dr. Guy De Tré, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent

LESGEVERS

- **Toon Boeckling**, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- **Michael Brands**, Consono.ai
- **Maxime Deforche**, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- **Guy De Tré**, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- **Tom Geudens**, Neo4J
- **Bart Mesuere**, Vakgroep Toegepaste Wiskunde, Informatica en Statistiek, Universiteit Gent
- **Yoram Timmerman**, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent

MEER INFO EN INSCHRIJVEN

WWW.UGAIN.UGENT.BE/DATABANKBEHEER

PROGRAMMA

Waarom heb je NoSQL systemen nodig?

In deze introductie bespreken we de recente evolutie in datagebruik en datanoden. We leggen uit wat de karakteristieken zijn van 'big' data en welke uitdagingen dit meebrengt voor databeheer. Vervolgens staan we stil bij de tekortkomingen van conventionele databasesystemen en geven we de beschikbare oplossing voor elke uitdaging. Meer specifiek behandelen we technieken van horizontaal schalen, schemaloze databases, No-ACID systemen en technieken om om te gaan met de waarheidsgetrouwheid van data. Deze oplossingen brengen ook nadelen met zich mee. Deze lichten we toe aan de hand van het CAP theorema en het principe van BASE systemen. Tot slot geven we een bondig overzicht van de bestaande NoSQL oplossingen, die verderop in deze opleiding aan bod komen.

Key-Value stores

Key-value stores vormen de eenvoudigste NoSQL-databaseoplossing. Ze laten het toe om data heel snel weg te schrijven, maar bieden amper ondersteuning voor bevraging en integriteitscontrole. Wanneer je kiest voor deze oplossing moet je dus heel goed de voor- en nadelen tegen elkaar afwegen en anticiperen hoe je hiermee kan omgaan. Key-value store technologie wordt tegenwoordig ook gebruikt in de zogenaamde vectordatabases die gebruikt worden in tal van toepassingen voor machinaal leren.

Column stores

In deze les verkennen we column stores en hun toepassingen. We gaan dieper in op de verschillende noden bij Online Transaction Processing (OLTP) en Online Analytical Processing (OLAP) workloads. Ook het bijhouden van tijdsreeksen, een veel voorkomende toepassing van column stores, komt aan bod. We gaan dieper in op enkele concrete databanksystemen zoals DuckDB en Cassandra in een praktische oefening.

Datavisualisatie

In deze les gaan we dieper in op het belang van datavisualisatie en hoe we data op een effectieve manier kunnen voorstellen. We leren je de taal om te spreken over data en de componenten waaruit een visualisatie bestaat. Doorheen de les werken we met een voorbeeld waarbij we eerst de sterke en zwakke punten van een datavoorstelling identificeren en nadien zelf aan de slag gaan om betere alternatieven uit te werken.

Document stores

Binnen de NoSQL databanken vormen de document stores een belangrijke categorie. Zoals de naam aangeeft, worden dit soort databanken gekenmerkt door het feit dat ze data opslaan in documenten (bv. JSON). Daardoor zijn ze zeer intuïtief en flexibel in gebruik, en schalen ze vrij goed. In deze lessen leren we aan wat document stores zijn, welke voor- en nadelen ze hebben en in welke situaties ze best gebruikt kunnen worden. Daarnaast worden er, door middel van hands-on oefeningen en demo's, een aantal fundamentele problemen aangepakt met betrekking tot het ontwerp van document stores, en het manipuleren en analyseren van data in document stores. We behandelen deze problemen voor verschillende document stores (MongoDB, CouchDB,...), en denken na over hoe eenvoudig deze problemen aangepakt zouden kunnen worden in een meer conventionele, relationele databank.

Graph databases

Met de behoefte aan nieuwe vormen om data te verwerken kwam ook de nood om data efficiënter te connecteren en via navigatie te kunnen doorzoeken en analyseren. Dit blies netwerkdatabase-technologie nieuw leven in en resulteerde in graafdatabanken waarbij de "graaf-met-eigenschappen" (property graph) het dominant databankmodel is met GQL als nieuwe ISO standaard querytaal in wording. De lessen rond graafdatabanken zijn opgedeeld in twee delen. In het eerste deel wordt de kracht van connecties geïllustreerd, bekijken we het databankmodel achter een graafdatabase en hebben we ruime aandacht voor de toepassingsgebieden. Op dit moment is Neo4j wereldwijd het meest gebruikte graafdatabase-systeem. Hoe werkt dit systeem? Hoe integreer je een Neo4j graafdatabase met je bestaande databanken? Wat is Graph data science? Als afsluiter zetten we de stap naar de virtuele Neo4j omgeving waar je zelf kan kennismaken met de Cypher (een voorloper van de GQL standaard) querytaal.

In het tweede deel graven we wat dieper in een graaf en geven we antwoorden op de volgende vragen: In welke opzichten verschilt graafmodellering van modellering in een relationele database? Hoe modelleer je een graaf voor een specifieke toepassing? Hoe modelleer je een graaf voor Graph Data Science? Daarna gaan we terug aan de slag met de virtuele Neo4j omgeving. We laden (bulk) data, ontwikkelen de queries voor een "aanbevelings" applicatie, doen aan path finding en om af te sluiten doorlopen we de stappen om een data science pijplijn voor een graafdatabase op te zetten.

Semantische indexen

De wereld draait om data en data bestaan in alle mogelijke vormen en voorstellingen, databanken, documenten, presentaties, filmpjes enz. Een van de grootste uitdagingen op vlak van data ligt dan ook in het op een betekenisvolle manier integreren en verbinden van data. Hoe kunnen de belangrijkste elementen uit een document gehaald worden? Hoe kan de link gelegd worden tussen velden uit verschillende databanken en datamodellen? Kan de informatie uit documenten automatisch gelinkt worden aan de inhoud van databanken? Hoe kan ik mijn data bruikbaar maken voor AI? Het zijn allemaal vragen die terugkomen bij iedereen die 'iets meer' wil doen met data of die wil starten met AI in zijn organisatie.

Tijdens de eerste lesavond bekijken we hoe Dynizer, een unieke semantisch intelligente dataoplossing van het Belgische bedrijf Consono.ai, semantische abstractie en AI gebruikt om data makkelijk te modelleren, te integreren en op te vragen via DQL (een semantisch verkrijgt SQL-dialect). We gaan ook dieper in op de mogelijkheden die Dynizer biedt op vlak van document analyse, samenvatting en pseudonimisatie.

Tijdens de tweede lesavond gaan we aan de slag met hands-on voorbeelden. We verwerken een paar datasets en documenten met Dynizer om dan via dashboards en DQL samen te ontdekken hoe het systeem automatisch de links tussen de inhoud van documenten en gestructureerde data blootlegt en je komt vertellen wat je nog niet wist over de data.

Hoe gebruik je NoSQL systemen in de praktijk?

In deze les geven we een totaaloverzicht en staan we stil bij een aantal strategische overwegingen bij het opzetten en gebruik van NoSQL systemen. Wat zijn de voor- en nadelen? Met welke valkuilen moet je rekening houden? Hoe zit het met wettelijke verplichtingen? We werpen ook een blik op de toekomst en gaan dieper in op de toenemende nood om adequaat te kunnen omgaan met waarheidsgetrouwheid van data en welke technieken daartoe beschikbaar zijn.

PRAKTISCH

Prijs

De deelnameprijs bedraagt **1.890 euro**.

Deelnameprijs omvat lesgeld, hand-outs, frisdranken, koffie en broodjes.

Betaling geschiedt na ontvangst van de factuur. Alle facturen zijn betaalbaar dertig dagen na dagtekening. Alle vermelde bedragen zijn vrij van BTW.

Laptop

Er wordt gewerkt met eigen laptop. Deze moet krachtig genoeg zijn (minimum 8GB RAM) en deelnemers moeten administratierechten hebben voor het installeren van de nodige programma's.

Korting

- Indien minstens één deelnemer van een bedrijf inschrijft voor de volledige opleiding wordt voor alle bijkomende gelijktijdige inschrijvingen van hetzelfde bedrijf een korting van 20% verleend. Facturatie geschiedt dan d.m.v. een gezamenlijke factuur.
- Aangepaste prijzen voor personeel van UGent
- Kortingen zijn niet cumuleerbaar.

Annulering

Raadpleeg onze annulatievoorwaarden op www.ugain.ugent.be/annulatievoorwaarden

KMO-portefeuille

Universiteit Gent aanvaardt betalingen via de KMO-portefeuille: www.kmo-portefeuille.be; gebruik autorisatiecode DV.0103194.



Tijdstip en locatie

- De lessen worden on campus gegeven **van 17u30 tot 21u**, in 2 delen, gescheiden door een broodjesmaaltijd. Deze vinden plaats aan de Universiteit Gent, UGent Academie voor Ingenieurs, Technologiepark 60, 9052 Zwijnaarde.
- Data onder voorbehoud van wijzigingen om onvoorziene omstandigheden.

PROGRAMMA

2 en 9 oktober 2023	Waarom heb je NoSQL systemen nodig? Guy De Tré
9 oktober 2023	Key-Value stores Toon Boeckling & Guy De Tré
16 oktober 2023	Column stores Bart Mesuere
23 oktober 2023	Datavisualisatie Bart Mesuere
6 en 13 november 2023	Document stores Toon Boeckling, Maxime Deforche & Yoram Timmerman
20 en 27 november 2023	Graph databases Tom Geudens
4 en 12 december 2023	Semantische indexen Michael Brands
18 december 2023	Hoe gebruik je NoSQL systemen in de praktijk? Guy De Tré

Organisatie

Universiteit Gent

UGain (UGent Academie voor Ingenieurs)
Technologiepark 60
9052 Zwijnaarde
09 264 55 82
ugain@ugent.be - www.ugain.ugent.be

met steun van 

MEER INFO EN INSCHRIJVEN

WWW.UGAIN.UGENT.BE/DATABANKBEHEER