

POSTACADEMISCHE OPLEIDING

# FUTUREPROOF DATABEHEER

WAARHEIDSGETROUWHEID IN DATABASES EN AI

10 oktober 2024 – 19 december 2024



UNIVERSITEIT  
GENT

Het gebruik van big data en AI werpt vragen op over **waarheidsgetrouwheid (veracity) van resultaten**. Mede hierdoor behoren het adequaat inschatten en gepast omgaan met **datakwaliteit** tot de grootste IT-uitdagingen van organisaties. **In deze cursus leer je de aandachtspunten, beperkingen en baten bij het gebruik van de nieuwste databanktechnologie met het oog op waarheidsgetrouwe resultaten. Door bewust te leren omgaan met datakwaliteit kan je data-gedreven applicaties of AI-toepassingen beter inzetten en gebruikers beter informeren over de waarheidsgetrouwheid van de bekomen resultaten, wat wordt gezien als een aanzienlijke (concurrentiële) meerwaarde bij beslissingsondersteuning.**

Door nieuwe manieren van verzamelen van data en omgaan met data verkrijgen we **grotere en complexere databronnen**. Deze worden almaar meer aangewend door **nieuwe toepassingen en artificiële intelligentie** en hebben vaak tot doel om te zorgen voor **verbeterde beslissingsondersteuning**. De huidige databanktechnologie anticipeert op deze veranderingen maar vraagt **bijzondere aandacht voor de waarheidsgetrouwheid** van teruggegeven resultaten.

**Deze opleiding is gericht op het verwerven van inzicht in de huidige technieken voor databeheer en datagebruik.** We leren je de praktische **voor- en nadelen van conventionele SQL-databasesystemen** en zogenaamde **NoSQL-databasesystemen** en hoe je deze systemen nuttig samen kan inzetten. Verder besteden we ruim aandacht aan het **efficiënt beheer en gebruik van tekstuele data**. **Nieuwe technologie** zoals vectordatabases en op taalmodellen gebaseerde toepassingen zoals ChatGPT worden toegelicht **met bijzondere aandacht voor waarheidsgetrouwheid en praktische inzetbaarheid**.

De opleiding wordt georganiseerd onder de vorm van hoorcolleges die afgewisseld worden door praktische demoesessies die interactief kunnen worden gevolgd. Deelnemers kunnen vrijblijvend zelf software op hun laptop installeren om bepaalde zaken praktisch uit te testen.

## DOELPUBLEK

Deze opleiding speelt in op de nieuwe trends in databeheer. Daardoor is de opleiding **breed toegankelijk** en volstaat enige basiskennis over databanken en AI als voorkennis.

De opleiding is daarom onder andere geschikt voor:

- **Datawetenschappers en data-analisten** die hun kennis willen uitbreiden naar geavanceerde technieken en toepassingen op het gebied van databeheer, het beheren van tekstuele data en AI.
- **Software-ontwikkelaars** die databasefunctionaliteit willen integreren om hun applicaties en systemen klaar te maken voor AI-toepassingen.
- **Bedrijfsanalisten en managers** die inzicht willen krijgen in hoe geavanceerde databeheer-, tekstbeheer- en AI-technologieën kunnen worden toegepast om waarde te genereren voor hun organisaties.

- **Studenten en academici** die zich specialiseren in datagerelateerde disciplines zoals informatica, datawetenschap, artificiële intelligentie of bedrijfsinformatiesystemen.
- **Professionals** uit verschillende branches die hun kennis willen uitbreiden om **geavanceerde analytische en voorspellende mogelijkheden** te benutten in hun werk zoals bijvoorbeeld in de gezondheidszorg, financiën, marketing, enz.
- **Professionals** die hun kennis willen uitbreiden over het **integreren, beheren, interpreteren en analyseren van tekstuele data** voor informaticatoepassingen.

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.

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
- **Maxim Deforche**, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- **Guy De Tré**, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- **Els Lefever**, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- **Tom Geudens**, Neo4J
- **Bart Mesuere**, Vakgroep Toegepaste Wiskunde, Informatica en Statistiek, Universiteit Gent

**MEER INFO EN INSCHRIJVEN**  
**[WWW.UGAIN.UGENT.BE/DATABASEHEER](http://WWW.UGAIN.UGENT.BE/DATABASEHEER)**

## PROGRAMMA

### Waarom heb je NoSQL-systemen nodig?

In de introductie staan we stil bij 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 staan we stil bij technieken van horizontaal schalen, schemaloze databases, No-ACID systemen en technieken voor het omgaan 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 en gaan we dieper in op key-value stores, document stores en kolom stores die we vergelijken in functie van hun sterktes en zwaktes met het oog waarheidsgetrouwheid.

### Document stores in de praktijk

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 **illustreer we het praktisch gebruik van document stores**, hun voor- en nadelen met betrekking tot datakwaliteit en geven we aan in welke situaties ze best kunnen worden gebruikt. Daarbij worden er door middel van demo's een aantal fundamentele problemen aangegeven 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 bespreken hoe een deze problemen worden aangepakt in een conventionele, relationele databank.

### Column stores in de praktijk

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 demo.

### 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 zwakte punten van een datavoorstelling identificeren en nadien zelf aan de slag gaan om betere alternatieven uit te werken. Bijzondere aandacht gaat naar aspecten van waarheidsgetrouwheid bij datavisualisatie.

### Graph databases in de praktijk

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 niet 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 ruim aandacht voor de toepassingsgebieden. Op dit moment is Neo4j wereldwijd het meest gebruikte graafdatabanksysteem. 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 databank? Hoe modeleer je een graaf voor een specifieke toepassing? Hoe modeleer 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.

### Efficiënt omgaan met tekstuele data

In deze lesavond behandelen twee technologieën die een belangrijke rol spelen bij het omgaan met tekstuele data: **NLP en AI-taalmodellen**. NLP, of Natural Language Processing, is een gebied binnen de computerwetenschappen dat zich bezighoudt met de interactie tussen computers en menselijke taal. Het doel van NLP is om computers te helpen menselijke taal te begrijpen, analyseren en genereren op een manier die natuurlijk aanvoelt voor mensen. Dit omvat taken zoals automatisch vertalen, sentimentanalyse, spraakherkenning en chatbots.

**AI-taalmodellen** zijn computerprogramma's die ontworpen zijn om **menselijke taal te begrijpen en te produceren**. Ze maken gebruik van geavanceerde algoritmen en machine learning-technieken om patronen in tekstgegevens te herkennen en vervolgens natuurlijke taal te genereren die vergelijkbaar is met menselijke taal. Deze modellen kunnen worden ingezet voor taken zoals tekstgeneratie, vertaling, samenvatting en vraag-antwoordsystemen.

Bij de bespreking van beide technologieën besteden we **ruim aandacht voor de waarheidsgetrouwheid van resultaten, voor- en nadelen, praktisch gebruik en toekomstige verwachtingen**.

### Vector databases en semantische indexen

Eén van de grootste **uitdagingen** op vlak van dataverwerking ligt in het op een **betekenisvolle manier integreren en verbinden van data**. Dit geldt temeer wanneer we ook rekening moeten houden met tekstuele documenten. Hoe kunnen **de belangrijke 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, integreren en op te vragen via DQL (een semantisch verrijkt SQL-dialect). We gaan ook dieper in de op de mogelijkheden die Dynizer biedt op vlak van documentanalyse, samenvatting en pseudonimisatie. Tijdens de tweede lesavond gaan we aan de slag met **een aantal demo's en voorbeelden**. We zullen een paar datasets en documenten met Dynizer verwerken 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. De lesavond sluiten we af met een kritische noot en bespreking van waarheidsgetrouwheid van data in Dynizer.

## PRAKTISCH

### Prijs

De deelnameprijs bedraagt **1.600 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](http://www.ugain.ugent.be/annulatievoorwaarden)

## KMO-portefeuille

Universiteit Gent aanvaardt betalingen via de KMO-portefeuille: [www.kmo-portefeuille.be](http://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.
- Er is geen les op 31 oktober 2024.
- Data onder voorbehoud van wijzigingen om onvoorziene omstandigheden.

#### PROGRAMMA

10 en 17 oktober 2024	<b>Waarom heb je NoSQL systemen nodig?</b> Guy De Tré
24 oktober 2024	<b>Document stores in de praktijk</b> Toon Boeckling & Maxime Deforche
7 november 2024	<b>Column stores in de praktijk</b> Bart Mesuere
14 november 2024	<b>Datavisualisatie</b> Bart Mesuere
21 en 28 november 2024	<b>Graph databases in de praktijk</b> Tom Geudens
5 december 2024	<b>Efficiënt omgaan met tekstuele data</b> Els Lefever
12 en 19 december 2024	<b>Vectordatabases en semantische indexen</b> Michael Brands en Guy De Tré

### Organisatie

#### Universiteit Gent

UGain (UGent Academie voor Ingenieurs)  
Technologiepark 60  
9052 Zwijnaarde  
09 264 55 82  
[ugain@ugent.be](mailto:ugain@ugent.be) - [www.ugain.ugent.be](http://www.ugain.ugent.be)

met steun van 

MEER INFO EN INSCHRIJVEN

[WWW.UGAIN.UGENT.BE/DATABASEER](http://WWW.UGAIN.UGENT.BE/DATABASEER)