ONDERZOEKSVOORSTEL

Analyse, architectuur en PoC van relevante labo's met betrekking tot data storage en efficiënte queries.

Bachelorproef, 2022-2023

Thomas Vanderveken

E-mail: thomas.vanderveken@student.hogent.be

Samenvatting

Dit onderzoek richt zich op de impact van verschillende databanktechnologieën op dataopslag en queryverwerking, met specifieke aandacht voor de educatieve context. De voortdurende evolutie van software en hardware vereist begrip van optimalisatietechnieken, vooral in kleinere use cases zoals labo-opdrachten. Het doel is om deze complexe concepten overtuigend over te brengen in het onderwijs door diepgaande analyse, architecturale implementatie en Proof of Concept (PoC) van relevante labo-opdrachten.

De onderzoeksopzet omvat de analyse van diverse databanktechnologieën binnen de context van een specifieke use case, zoals een webshop. De focus ligt op het meten en vergelijken van prestaties om voor- en nadelen grondig in kaart te brengen. Het uiteindelijke doel is waardevolle inzichten te genereren voor zowel onderwijs als bedrijfsleven, met een praktische toepassing in de vorm van een PoC voor lesdemonstraties.

Dit onderzoek is van belang voor zowel onderwijs als bedrijfsleven, en biedt niet alleen diepgaand inzicht in databanktechnologieën, maar ook praktische implicaties voor kosteneffectieve en performante implementaties. De resultaten zullen bijdragen aan zowel academisch begrip als praktische toepassing, met als focus een verbeterde educatieve ervaring.

Keuzerichting: Al & Data Engineering

Sleutelwoorden: data storage, sql, nosql, educationals

Inhoudsopgave

	iding	
		•
1.3	Onderzoeksopzet	2
1.4	Doelstelling van het Onderzoek	2
1.5	Relevantie van het Onderzoek voor	
	Onderwijs en Bedrijfsleven	2
1.6	Onderzoeksvragen	2
1.7	Toepassingsgerichte Use Cases voor	
	Lesillustratie	2
Lite	ratuurstudie	2
2.1	Evoluerend Landschap van Databank	
	technologieën	2
2.2	Impact van Prestatie-optimalisatie	3
2.3	Relevante Databanktechnologieën er	1
2	.3.1 NoSQL-databases	2
2	.3.2 Relationele Databases (SQL)	4
2	.3.3 Voor- en Nadelen	4
Met	thodologie	4
	Identificatie	4
3.2	Fase 2: Selectie van Databanken en	
		4
	•	
	·	_
	Fase 4: Uitvoeren van de POC en Edu-	
J. 1	rase i. Oltvocicii vairae i occii Laa	
	1.2 1.3 1.4 1.5 1.6 1.7 Lite 2.1 2.2 2.3 Med 3.1 3.2	Literatuurstudie

	3.5 Fase 5: Analyse van de Gegevens en	
	Educatieve Effectiviteit	5
	3.6 Fase 6: Conclusies en Aanbevelingen	
	voor Lesgebruik	5
4	Verwachte resultaten	5
5	Discussie, conclusie	5
	Referenties	6

1. Inleiding

1.1. Achtergrond en Context

Met de voortdurende evolutie van software en hardware is het essentieel om de impact van optimalisatietechnieken op databanktechnologieën te begrijpen, vooral met betrekking tot dataopslag en efficiënte queries. Het onderwijs staat voor de uitdaging om deze complexe concepten overtuigend over te brengen, met name in kleinere use cases en labo-opdrachten. Dit onderzoek beoogt deze uitdaging aan te pakken door diepgaande analyse, architecturale implementatie en Proof of Concept (PoC) van relevante labo-opdrachten met betrekking tot dataopslag en efficiënte queries.

1.2. Probleemstelling

In het onderwijs blijft het moeilijk om de impact van optimalisatietechnieken en verschillende databanktechnologieën overtuigend te presen-



Dit onderzoek streeft naar een diepgaand inzicht in de impact van diverse databanktechnologieën op de efficiëntie van dataopslag en queryverwerking, met specifieke aandacht voor praktische toepassingen zoals webshops. De vertaling van deze inzichten naar praktijkgerichte demonstraties is van cruciaal belang.

1.3. Onderzoeksopzet

Het hoofddoel van dit onderzoek is een kritische analyse uit te voeren van verschillende databanktechnologieën en hun impact binnen de context van een specifieke use case, zoals een webshop. Door de prestaties van verschillende databanktechnologieën te meten en te vergelijken, beoogt dit onderzoek de voor- en nadelen grondig in kaart te brengen. Het uiteindelijke doel is waardevolle inzichten te genereren die zowel relevant zijn voor het onderwijs als voor het bedrijfsleven. Bovendien zal er een Proof of Concept (PoC) worden ontwikkeld om de verschillen in de praktijk te demonstreren.

1.4. Doelstelling van het Onderzoek

De voornaamste doelstelling van dit onderzoek is een kritische analyse uit te voeren van verschillende databanktechnologieën binnen de context van een specifieke use case, zoals een webshop. Door prestaties van diverse databanktechnologieën te meten en te vergelijken, beoogt dit onderzoek de voor- en nadelen grondig in kaart te brengen. Het uiteindelijke doel is waardevolle inzichten te genereren die zowel relevant zijn voor het onderwijs als voor het bedrijfsleven, met als concrete toepassing een Proof of Concept (PoC) die in de les kan worden gedemonstreerd.

1.5. Relevantie van het Onderzoek voor Onderwijs en Bedrijfsleven

Dit onderzoek is van groot belang voor zowel het onderwijs als het bedrijfsleven. Het biedt niet alleen diepgaand inzicht in de invloed van databanktechnologieën op de efficiëntie van dataopslag en queryverwerking, maar draagt ook praktische implicaties aan voor bedrijven die streven naar kosteneffectieve en performante databankimplementaties. De bevindingen zullen niet alleen bijdragen aan het academisch begrip van deze concepten, maar ook aan de demonstratie van praktische scenario's met een Proof of Con-

1.6. Onderzoeksvragen

Dit onderzoek zal zich concentreren op het beantwoorden van de volgende vragen:

· Welke impact hebben verschillende data-

- teren, vooral in kleinere use cases zoals labo-opdrachten. operaties binnen de context van een specifieke use case, zoals een webshop?
 - · Hoe variëren de prestaties van databanktechnologieën bij het verwerken van aanzienlijke hoeveelheden data?
 - · Hoe beïnvloeden architecturale keuzes, zoals de keuze tussen cloud en lokale opslag, de snelheid en kosteneffectiviteit van databankimplementaties?
 - · Welke inzichten kunnen worden afgeleid uit de vergelijking van verschillende databanktechnologieën, en hoe kunnen deze inzichten zowel het onderwijs als het bedrijfsleven ten goede komen?

Door deze vragen grondig te beantwoorden, beoogt dit onderzoek een diepgaand begrip te verschaffen van de impact van databanktechnologieën op de efficiëntie van dataopslag en queryverwerking, met als resultaat praktisch toepasbare waarde voor zowel het onderwijs als het bedrijfsleven.

1.7. Toepassingsgerichte Use Cases voor Lesillustratie

Een belangrijk aspect van dit onderzoek is het identificeren van use cases die geschikt zijn voor lesillustratie. Hierbij wordt onderzocht welke specifieke situaties en oefeningen geschikt zijn om de impact van diverse databanktechnologieën op een boeiende manier te demonstreren. Het doel is om praktijkgerichte voorbeelden te presenteren die verder gaan dan de traditionele "hellowereld"demo's en quickstarts, waardoor de educatieve waarde van het onderzoek wordt versterkt. Bovendien zal de ontwikkelde Proof of Concept (PoC) dienen als een tastbaar voorbeeld, ook relevant voor bedrijven die streven naar geoptimaliseerde databankimplementaties.

2. Literatuurstudie

2.1. Evoluerend Landschap van Databanktechnologieën

In het informatietijdperk is het cruciaal geworden om grote hoeveelheden gegevens op te slaan en te beheren voor bedrijfsactiviteiten. Traditionele databankbeheersystemen zijn echter niet meer toereikend voor de complexe datasets gegenereerd door moderne toepassingen (Chaturvedi (2023); Gupta e.a. (q.d.)). De opkomende trends in databanken vereisen een nieuwe benadering om de interactie van deze componenten met de rest van de bedrijfsarchitectuur te begrij-

Een cruciale ontwikkeling is de opkomst van banktechnologieën op de snelheid van CRUD- multi-model databases, zoals benadrukt door Chaturvedi (2023). Deze databases zijn ontworpen als



ondersteunen meerdere datamodellen. Door verschillende modellen te integreren, kunnen IT-teamsen bedrijfsstrategie. en bedrijfsgebruikers meerdere toepassingen ondersteunen zonder voor elke toepassing een afzonderlijk databasesysteem te implementeren.

Een andere opkomende trend is het gebruik van cloud-native databases, specifiek ontworpen om de mogelijkheden van cloudtechnologie en gedistribueerde systemen ten volle te benutten. Deze databases bieden voordelen zoals schaalbaarheid, elasticiteit, veerkracht, ondersteuning voor automatisering en verbeterde toegankelijkheid (Chaturvedi (2023); Gupta e.a. (g.d.)).

2.2. Impact van Prestatie-optimalisatie

Prestatie-optimalisatie is cruciaal, waarbij gebruikers snel en naadloos door grote datasets moeten kunnen navigeren. Zoals FasterCapital (2023) benadrukt, is het begrijpen van het belang van prestatie-optimalisatie essentieel in de hedendaagse digitale wereld. Gebruikers hebben weinig tolerantie voor trage interfaces en langzame laadtijden, wat de directe invloed van prestatie-optimalisatie op de gebruikerservaring benadrukt.

Prestatie-optimalisatie heeft ook aanzienlijke zakelijke implicaties. Het voorbeeld van Google, aangehaald door FasterCapital (2023), illustreert dat zelfs kleine vertragingen in laadtijd kunnen leiden tot aanzienlijke vermindering van betrokkenheid en conversies. Deze zakelijke impact onderstreept de noodzaak voor ontwikkelaars en bedrijven om prestatie-optimalisatie te omarmen als een integraal onderdeel van hun ontwikkelingsen bedrijfsstrategie.

FasterCapital (2023) suggereert enkele cruciale technieken voor prestatie-optimalisatie, waaronder het minimaliseren van de belasting van bronnen, caching, databasequery-optimalisatie, lazy loading en het gebruik van Content Delivery Networks (CDN's). Deze technieken bieden praktische benaderingen om laadtijden te verkorten, serverbelasting te verminderen en een snelle, responsieve gebruikerservaring te waarborgen.

Prestatie-optimalisatie is cruciaal, waarbij gebruikers snel en naadloos door grote datasets moeten kunnen navigeren. Zoals FasterCapital (2023) benadrukt, is het begrijpen van het belang van prestatie-optimalisatie essentieel in de hedendaags **Redis (Intellectsoft (2020)):** digitale wereld. Gebruikers hebben weinig tolerantie voor trage interfaces en langzame laadtijden, wat de directe invloed van prestatie-optimalisatie op de gebruikerservaring benadrukt.

Prestatie-optimalisatie heeft ook aanzienlijke zakelijke implicaties. Het voorbeeld van Google, aangehaald door FasterCapital (2023), illustreert dat zelfs kleine vertragingen in laadtijd kunnen leiden tot aanzienlijke vermindering van betrokkenheid en conversies. Deze zakelijke impact onderstreept de noodzaak voor ontwikkelaars en

veelzijdige platforms voor gegevensverwerking en bedrijven om prestatie-optimalisatie te omarmen als een integraal onderdeel van hun ontwikkelings-

> FasterCapital (2023) suggereert enkele cruciale technieken voor prestatie-optimalisatie, waaronder het minimaliseren van de belasting van bronnen, caching, databasequery-optimalisatie, lazy loading en het gebruik van Content Delivery Networks (CDN's). Deze technieken bieden praktische benaderingen om laadtijden te verkorten, serverbelasting te verminderen en een snelle, responsieve gebruikerservaring te waarborgen.

2.3. Relevante Databanktechnologieën en Toepassingsgebieden

Dit gedeelte biedt een diepgaande analyse van belangrijke NoSQL-databases zoals MongoDB, Redis, Cassandra, en Neo4j, evenals prominente SQLdatabases zoals Microsoft SQL Server (MSSQL), MySQL, en PostgreSQL.

2.3.1. NoSQL-databases MongoDB (K (2019)):

- · Sterktes:
 - Schema-less NoSQL database.
 - Geen complexe joins nodig; geen relatie tussen gegevens.
 - Eenvoudig schaalbaar.
 - Eenvoudig te installeren en op te zetten.
 - Ondersteunt opslag van arrays en objecten via JSON-formaat.
 - Hogere prestaties vergeleken met veel relationele databases.
 - Gebruikt interne geheugenopslag voor snellere toegang.
- · Zwaktes:
 - Vereist aanzienlijk geheugen voor gegevensopslag.
 - Documentgrootte is beperkt tot 16 MB.
 - Gebrek aan transactieondersteuning in MongoDB.

- Sterktes:
 - Uitstekende keuze voor caching met geavanceerde datastructuren.
 - Persistente systeemstructuur voor "hot data"verwerking.
 - Effectief berichtentransportsysteem.
- · Zwaktes:



- flexibiliteit in het kiezen van welke gegevens op te slaan..
- Clustering kan moeilijk zijn vanwege de afzonderlijke structurele eenheden die in paren zijn verbonden.

Cassandra (Hegde e.a. (2023)):

- · Sterktes:
 - Flexibel en tolerant ten opzichte van fouten.
 - Schaalbaarheid waarbij lees- en schrijfdoorvoer lineair toenemen.
 - Beveiliging en observatie met audit log-
 - Zero Copy Streaming versnelt operaties tot vijf keer sneller.

· Zwaktes:

- Gebrek aan SQL-queryfunctionaliteiten.
- Onvermogen om joins uit te voeren, waardoor het uitdagend is voor complexere taken zoals online transacties.
- Minder geschikt voor OLTP-type transactiegerichte en hoog-concurrente systemen.

Neo4j (GeeksForGeeks (2022)):

- · Sterktes:
 - Eenvoudige representatie van verbonden gegevens in een graf.
 - Snelle ophaling, navigatie en traversering van verbonden gegevens.
 - Gebruikt een eenvoudig en krachtig gegevensmodel.
 - Kan semi-gestructureerde data gemakkelijk vertegenwoordigen.

7waktes:

- Ondersteuning voor OLAP in deze databases is niet goed geïmplementeerd.

MongoDB biedt een schaalbare, schema-loze omgeving met snelle toegang, maar vereist aanzienlijk geheugen en mist transactieondersteuning. Redis is ideaal voor caching, maar beperkt opslagflexibiliteit en kan clustering compliceren. Cassandra excelleert in flexibiliteit en schaalbaarheid maar mist SQL-query's en is minder geschikt voor complexe transacties. Neo4j vertegenwoordigt verbonden gegevens efficiënt in een graf, maar heeft beperkte ondersteuning voor OLAP. Deze analyse benadrukt de diversiteit in sterke en zwakke meest relevante databanktechnologieën geselecpunten van NoSQL-databases, cruciaal bij het kiezen van een geschikte databaseoplossing.

- Verplichte opslag voor alle gegevens, geen 2.3.2. Relationele Databases (SQL)

Relationele databases zoals Microsoft SQL Server (MSSQL), MySQL en PostgreSQL zijn veelgebruikte relationele databases. Deze databases volgen een schema en gebruiken tabellen om gestructureerde gegevens op te slaan. Voorbeelden van SQL-databases zijn onder andere Oracle, PostgreSQL, MySQL en SQL Server.

Kenmerken van Relationele Databases:

- · Strikte schema's zorgen voor voorspelbaarheid en gemakkelijke beoordeling van gegevens.
- · ACID-naleving garandeert atomiciteit, consistentie, isolatie en duurzaamheid.
- · Goed gestructureerd en verminderen de kans op fouten.

2.3.3. Voor- en Nadelen

Voor- en Nadelen van Relationele Databases:

- · Verdiensten zijn onder andere voorspelbaarheid, strikte schema's en ACID-naleving.
- · Nadelen omvatten beperkingen bij het opslaan van grote hoeveelheden internetgegevens en moeilijkheden bij horizontaal schalen

3. Methodologie

3.1. Fase 1: Literatuurstudie en Use Case Identificatie

In deze initiële fase wordt niet alleen een grondige literatuurstudie uitgevoerd om een stevig theoretisch kader te creëren, maar wordt ook specifiek onderzoek gedaan naar use cases die bijzonder geschikt zijn voor lesillustratie. Hierbij wordt onderzocht welke specifieke situaties en oefeningen zich goed lenen om de impact van diverse databanktechnologieën te demonstreren in een lesformaat. Criteria voor de selectie van deze use cases omvatten geschiktheid voor educatieve doeleinden, relevantie voor de beoogde doelgroep, en de mogelijkheid om boeiende en instructieve scenario's te creëren.

Resultaat: Een uitgebreid literatuuroverzicht, inclusief geïdentificeerde use cases voor lesillustratie, met duidelijke specificaties van de bijbehorende data, hoeveelheden en oefeningen.

3.2. Fase 2: Selectie van Databanken en Use Case Implementatie

Gebaseerd op de literatuurstudie en de geïdentificeerde use cases, worden zorgvuldig de teerd. De selectiecriteria omvatten niet alleen



technologische geschiktheid, maar ook de bruikbaarheid voor de gekozen educatieve use cases. Vervolgens wordt in deze fase de implementatie van de geselecteerde use cases uitgevoerd. Dit omvat het modelleren van data, het opzetten van scenario's en het voorbereiden van oefeningen die de lesillustratie zullen ondersteunen.

Resultaat: Een weloverwogen lijst van data-Resultaat: Een samenhangende set conclubanktechnologieën, onderbouwd door literaire bronsies en praktische aanbevelingen voor de acadenen, en de implementatie van relevante use cases mische en educatieve gemeenschap. voor educatieve doeleinden.

3.3. Fase 3: Opzetten van de POC met **Educatieve Focus**

In deze fase wordt de Proof of Concept (PoC) ontwikkeld met een specifieke focus op de educatieve aspecten van de geselecteerde use cases. Het experimentele ontwerp wordt aangepast om de impact van databanktechnologieën op lesillustratie te meten. Hierbij worden parameters zoals begrijpelijkheid, betrokkenheid van studenten en effectieve demonstratie van concepten nauwlettend in overweging genomen.

Resultaat: Een methodologisch verantwoorde PoC met duidelijk gedefinieerde parameters voor educatieve evaluatie en meetpunten.

3.4. Fase 4: Uitvoeren van de POC en **Educatieve Evaluatie**

De aangepaste PoC wordt uitgevoerd volgens het ontwikkelde experimentele ontwerp, met specifieke aandacht voor educatieve doelen. Gedetailleerde gegevens worden verzameld, niet alleen over de prestaties van de databanktechnologieën, maar ook over de effectiviteit van de lesillustratie. Studentenfeedback en educatieve resultaten worden zorgvuldig geanalyseerd.

Resultaat: Verzamelde experimentele gegevens die zowel de prestaties van de geselecteerde databanktechnologieën als de effectiviteit van de lesillustratie weerspiegelen.

3.5. Fase 5: Analyse van de Gegevens en Educatieve Effectiviteit

De verzamelde gegevens worden systematisch geanalyseerd, met specifieke aandacht voor de educatieve effectiviteit van de use cases. De nadruk ligt niet alleen op technische prestaties, maar ook op de mate waarin de lesillustratie de beoogde educatieve doelen heeft bereikt.

Resultaat: Gedegen analyse van experimentele gegevens met onderbouwde conclusies, inclusief educatieve effectiviteit.

3.6. Fase 6: Conclusies en Aanbevelingen voor Lesgebruik

Conclusies worden getrokken op basis van zowel technische prestaties als educatieve effectiviteit. Deze conclusies worden gecontextualiseerd binnen het theoretisch kader, de bestaande literatuur en de specifieke context van lesgebruik. Aanbevelingen voor verdere verbeteringen en toepassingen in lesomgevingen worden geformuleerd, rekening houdend met de educatieve impact van verschillende databanktechnologieën.

4. Verwachte resultaten

- Databankprestatievergelijking en Educa**tieve Impact:** Onderzoek naar superieure databanktechnologieën voor CRUD-operaties, met meetbare criteria zoals snelheid en systeembronnen, gericht op hun effectiviteit bij lesillustratie. Evaluatie van de educatieve impact van deze demonstraties op studentenbegrip en betrokkenheid.
- · Effect van Databanktechnologie op Querysnelheid in Educatieve Context: Onderzoek naar de prestaties van verschillende databanktechnologieën (SQL Server, Postgres, MongoDB, Redis, Neo4j, en anderen) bij het verwerken van grote hoeveelheden data en hun impact op guery-efficiëntie, met bijzondere aandacht voor de overdraagbaarheid van deze concepten naar educatieve scenario's.
- · Invloed van Architecturale Keuzes in Les**gebruik:** Analyse van hoe architecturale keuzes, zoals cloud versus lokale opslag, de snelheid en kosteneffectiviteit van databankimplementaties beïnvloeden, met specifieke overwegingen voor de integratie van deze concepten in lesmateriaal.
- · Praktisch Toepasbare Inzichten voor Educatie: Conclusies en aanbevelingen bieden waardevolle inzichten voor geïnformeerde besluitvorming over databanktechnologiekeuzes, zowel voor het onderwijs als het bedrijfsleven. De nadruk ligt op hoe deze inzichten kunnen worden geïntegreerd in lesprogramma's en hoe ze de educatieve ervaring van studenten kunnen verbeteren.

5. Discussie, conclusie

In deze bachelorproef wordt een diepgaand onderzoek uitgevoerd naar de impact van diverse databanktechnologieën op dataopslag en gueryverwerking, met specifieke aandacht voor de educatieve context. De literatuurstudie illustreert het evoluerende landschap van databanktechnologieën, waarbij zowel NoSQL-databases als relationele databases worden onderzocht. De voor- en nadelen van verschillende databases, zoals MongoDB,



Redis, Cassandra, en Neo4j, worden kritisch geanalyseerd, evenals de kenmerken van relationele databases.

De methodologie omvat de selectie van relevante use cases voor het mogelijk gebruik in de les om verschillende aspecten en eigenaardigheden te demonstreren. Dit omvat de implementatie van deze cases met zorgvuldig geselecteerde databanktechnologieën en de ontwikkeling van een Proof of Concept (PoC) met een educatieve focus.

De verwachte resultaten omvatten inzichten in databankprestaties, de impact van architecturale keuzes en praktisch toepasbare aanbevelingen voor zowel het onderwijs als het bedrijfsleven. Deze bachelorproef draagt bij aan het begrip van databanktechnologieën en biedt concrete toepassingen voor lesgebruik, waarbij zowel technische als educatieve aspecten worden belicht

Referenties

- Chaturvedi, V. (2023, augustus 6). 6 Emerging Trends in Databases For Every Industry: Multi Model Database and more. https://www.redswitches.com/blog/database-trends/#:~:text=2.-,Hybrid % 20and % 20Multi % 2Dcloud % 20Database % 20Architectures, these%20ideas%20in%20some%20detail.
- FasterCapital. (2023, november 22). Prestaties maximaliseren optimalisatietechnieken voor vermeldingsapplicaties. https://fastercapital.com/nl/inhoud/Prestaties-maximaliseren-optimalisatietechnieken-voor-vermeldingsapplicaties. html#Inleiding-tot-de-optimalisatie-van-vermeldingsapplicaties
- GeeksForGeeks. (2022, oktober 17). Neo4j Introduction. https://www.geeksforgeeks.org/ neo4j-introduction/
- Gupta, A., Shrivastava, R., & Jain, J. K. (g.d.). Emerging Trends in Database and its Applications in MNC's. https://doi.org/10.5815/ijeme. 2023.04.05
- Hegde, A., Bhat, M., Bavalatti, A., Hegde, P., Sahadevan, D., & Hegde, A. (2023). Apache Cassandra Overview (Features, Pros, and Cons) [Website: https://www.erp-information.com/].
- Intellectsoft. (2020, februari 3). Redis vs MongoDB: How to differ and what to choose? https: //www.intellectsoft.net/blog/redis-vsmongodb/
- K, J. (2019, juli 30). ALL ABOUT MONGODB: THE NOSQL DATABASE. https://acodez.in/mongodb-nosql-database/#Advantages_and_Disadvantages_of_MongoDB

