

ONDERZOEKSVOORSTEL

Toepassen van een tijdelijke grafiek transformatie op object traceringsdata voor het trainen van een LLM.

Bachelorproef, 2024-2025

Maarten Van der Schueren

E-mail: maarten.vanderschueren@student.hogent.be

Co-promotor: B. Peirens (Tracked, bart.peirens@tracked.be)

Samenvatting

In deze bachelorproef onderzoeken we hoe we een grafiekmodellering kunnen gebruiken een LLM te ontwikkelen, waarbij we bij bijvoorbeeld klachtenafhandeling snel en efficiënt de oorzaak kunnen gaan vragen. Het onderzoek is verdeeld in twee fases: de eerste fase gaat over het omzetten van events volgens EPCIS naar een grafiekmodel in cosmos DB met behulp van Gremlin API. In de volgende fase trainen we een LLM die ons kan vertellen waar, wanneer, wat en hoe een event heeft plaatsgevonden en anomalieën in het proces kan terugvinden. Het gebruik van deze LLM kan tijds- en geldbesparend zijn door handmatig zoekwerk te onderdrukken. Het is belangrijk in dit onderzoek dat alles schaalbaar en performant is voor zowel bedrijven, nationale of landelijke business use cases. Door deze implementatie verwachten we de klantenervaring en efficiëntie te optimaliseren.

Keuzerichting: AI & Data Engineering

Sleutelwoorden: LLM, OTG, EPCIS, CosmosDB, Assistant Agents

Inhoudsopgave

1	Inleiding	1
2	Literatuurstudie	1
2.1	Cosmos DB	1
2.1.1	Waarom Cosmos DB gebruiken	1
2.2	Grafiek modellering	2
2.2.1	Verskil transformatie en aggregatie	2
2.3	EPCIS	2
2.3.1	Waarom EPCIS?	2
3	Methodologie	2
4	Verwacht resultaat, conclusie	2
	Referenties	3

De scope van deze bachelorproef is tweedelig, waarbij deel één eruit bestaat om samen met Tracked vanuit een gestandaardiseerd datamodel die zichtbaarheid events bijhoudt om te zetten naar een grafiekmodel. Deel twee bouwt verder voort op dit grafiekmodel om dan een LLM (large language model) te trainen die eenvoudig informatie eruit kan halen en teruggeven.

De centrale vraag die hier gesteld wordt is: "Hoe we efficiënt en snel grafiekmodellering kunnen toepassen om een LLM te ontwikkelen die in staat is om het waar, wanneer, wat en hoe van gebeurtenissen binnen een proces vast te stellen, ter ondersteuning van klachtenafhandeling bij productfouten."

1. Inleiding

Binnen de Arcelor Mittal Gent bestaan er veel verschillende processen die verspreid zijn over meerdere afdelingen. Hierbij bevat elke afdeling een deel van het proces om van grondstof tot een afgewerkt product te komen. Tussen elke afdeling bevindt zich een deel van de toeleveringsketen, waarbij het resultaat van de ene afdeling het beginpunt is voor de volgende afdeling. Dit zorgt ervoor dat iedere stap zijn invloed heeft op de kwaliteit van het eindresultaat.

Elke afdeling is opgebouwd rond een specifiek proces, zoals de hoogoven die ruwe grondstoffen omzet in ruw ijzer en de staalfabriek die ruw ijzer in staal verandert. Omdat deze afdelingen zeer specifieke processen hebben, hebben ze ook zeer specifieke behoeften en datamodellen. Dit maakt analyse of onderzoek over afdelingen heen zeer complex en tijdsrovend.

2. Literatuurstudie

2.1. Cosmos DB

2.1.1. Waarom Cosmos DB gebruiken

Cosmos DB is een NoSQL-database van microsoft, deze heeft een lage latentie, multi-query-api die makkelijk grote hoeveelheden data kan verwerken en heeft een grote beschikbaarheid zegt van der Put (2020) wat zeer belangrijk is in ons project. Naast deze feiten is CosmosDB horizontaal schaalbaar wat betekent dat we op hoogtepunten een miljoen lees- en schrijfaanvragen kunnen verwerken door het nodige aantal servers toe te voegen. De hoge beschikbaarheid wordt gegarandeerd door replicatie waardoor we snel kunnen overschakelen als er iets fout gaat in de database.

2.2. Grafiek modellering

Cosmos DB biedt ook grafiek modellering aan met behulp van Gremlin. Gremlin is een grafiek database service waarbij we grote grafieken kunnen opslaan die bestaan uit vertices en edges (Microsoft, 2024). Dit gebeurt met een snelheid van miliseconden waardoor we een snelle verwerkingstijd zullen kunnen neerzetten. Ook is Gremlin schaalbaar en kan de consistentie level gekozen worden om een balans te vinden tussen consistentie, beschikbaarheid en latentie. Binnen zo een grafiek model hebben we Vertices (of nodes) die een persoon, plaats of event beschrijven zoals in ons geval bijvoorbeeld een slab die verplaatst wordt van A naar B.

2.2.1. Verschil transformatie en aggregatie

Binnen ons project kunnen bestaansvormen van de producten veranderen. Zo hebben we transformatie waarbij de cokes bijvoorbeeld omgezet worden in vloeibaar staal, dit is niet omkeerbaar. Bij aggregatie kunnen we terug naar de oorspronkelijke bestaansvorm zoals bijvoorbeeld van warmwalserij naar koudwalserij waarbij de platen dunner worden en eventueel een roestvrije laag krijgen (Arcelor Mittal, 2024). In onze grafiek modellering moeten we gebruik maken van tijdelijke edges om ook terug te kunnen kijken naar wat voor het transformatie- of aggregatieproces gebeurd is om anomalieën terug te vinden (Jaewook Byun, 2020).

2.3. EPCIS

Voor dit onderzoek moet alles voldoen aan de EPCIS (Electronic Product Code Information Services) waarden, dit zijn de wat, wanneer, waar, waarom en hoe. Deze waarden zijn ontwikkeld door GSI om gegevens over beweging, status en verandering van een item in de toeleveringsketen (supply chain) vast te leggen en te delen (Devins e.a., 2022). "Met behulp van deze waarden kunnen we real-life objecten omzetten in elektronisch opgeslagen informatie, waarna we dit kunnen communiceren met eindgebruikers." zegt Devins e.a. (2022). Door deze normen toe te passen kunnen we de traceerbaarheid van het product per proces garanderen inclusief de gewenste parameters die opgeslagen worden in ons grafiekmodel zoals tijd (wanneer) en temperatuur (hoe) waar nodig.

2.3.1. Waarom EPCIS?

Huidige Legacy Systemen maken gebruik van ERP, POS, WMS,... Vaak lopen deze niet real-time wat nadelig is als we direct iets willen weten over een product en hoge beschikbaarheid verwachten (Vieweger, g.d.). Doordat iedereen eenzelfde norm gebruikt kunnen we dit schaalbaar houden en makkelijk uitbereiden.

3. Methodologie

Eerst en vooral gaan we aan de slag met een json bestand dat data bevat van Arcelor Mittal waar events in opgeslagen staan. Daarna zetten we dit om naar Cosmos DB en kunnen we via Grammlin API werken om een grafiek op te zetten met tijdsgebonden edges. Dit dient ervoor om te zorgen dat we in tijd terug kunnen om de historie op te vragen en te kijken waar, wanneer en hoe een proces gebeurd is. Daardoor kunnen we achteraf bij onder andere klachtenafhandeling terug kijken in het grafiekmodel wat er mogelijks anders was of fout is gelopen.

Dit grafiekmodel moet ook voldoen aan de normen volgens EPCIS en schaalbaar zijn, daarom gaan we zorgen dat de gebruikte data binnen het grafiekmodel zoals een node een id krijgt in de plaats van een statische naam. Daardoor kan Tracked in latere fases of andere business cases dezelfde technieken gebruiken zonder veel aanpassingen te moeten doen op grote schaal.

Zodra we een duidelijk grafiekmodel hebben waarmee we makkelijk in het verleden kunnen kijken, kunnen we dit trainen met copilot van microsoft. Daardoor kunnen we dan snel zoeken naar waar mogelijks anomalieën zijn in een proces door te vergelijken met andere batches die hetzelfde proces hebben doorlopen. Als laatste kunnen we dan via copilot makkelijk een implementatie starten voor onder andere microsoft teams waarbij iemand van de klantenservice kan vragen wat er fout is gelopen en dat de chatbot zegt waar, wanneer en hoe er mogelijks een fout is geweest in het proces.

4. Verwacht resultaat, conclusie

In dit onderzoek maken we startlaag aan een groter project waarbij het mogelijk is om uit te bereiden naar bijvoorbeeld een planningstool om de meest efficiënte route te zoeken. Door deze chatbot verwachten we dat medewerkers van Arcelor Mittal eenvoudig een vraag kan stellen en een antwoord kan krijgen op een korte tijd. Zonder deze optie moeten ze elk deel van het bedrijf handmatig opbellen en vragen wat er gebeurd is, dat zijn zaken die zeer tijdsslopend zijn en veel geld kosten. Ik hoop ook dat we dit schaalbaar kunnen opstellen en performant kunnen houden waardoor het ook kan gebruikt worden op landelijk of internationaal niveau. Door de implementatie van dit systeem hoop ik dat we klanttevredenheid en besparingen binnen een bedrijf (in ons geval Arcelor Mittal Gent) optimaliseren

Referenties

- Arcelor Mittal. (2024). *Productieafdelingen*. Verkregen november 14, 2024, van <https://belgium.arcelormittal.com/werkomgeving/productieafdelingen/>
- Devins, C., e.a. (2022, juni). *EPCIS standard*. Verkregen november 13, 2024, van <https://ref.gs1.org/standards/epcis/>
- Jaewook Byun, D. K. (2020). *Expert Systems With Applications*.
- Microsoft. (2024, augustus 22). *What is Azure Cosmos DB for Apache Gremlin*. Verkregen november 14, 2024, van <https://learn.microsoft.com/en-us/azure/cosmos-db/gremlin/introduction>
- van der Put, M. (2020, augustus 28). *Azure Cosmos DB*. Verkregen november 14, 2024, van <https://humandigital.nl/nieuws-en-artikelen/azure-cosmos-db/>
- Vieweger, T. (g.d.). *EPCIS: Hoe zichtbaarheid van de voorraad verder gaat dan de winkelmuren*. Verkregen november 14, 2024, van <https://www.nedap-retail.com/nl/epcis-hoe-zichtbaarheid-van-de-voorraad-verder-gaat-dan-de-winkelmuren/>