

## Werkplan Masterproef

Titel	Streamlining the deployment of Cloud-Native Data-Driven Applications
Naam Student	
Email	
Bedrijf/ Onderzoeksgroep	IDLab, Ghent University – imec
Promotoren	
Begeleiders	

### Opsplitsing per semester:

	Semester 1	Semester 2
# studiepunten vakken	27	15
# studiepunten masterproef	6	12

### Bestaande situatie en probleemstelling:

Het efficiënt verwerken van grote hoeveelheden data is essentieel voor data-gedreven applicaties. Dit vereist platforms zoals Obelisk, die een basis vormen voor schaalbare en robuuste data-oplossingen. Hoewel de vraag naar dergelijke platforms groot is, vormt de complexiteit van de installatie een belangrijke belemmering.

Obelisk is een cloud-native platform voor het bouwen van schaalbare, data-gedreven applicaties en diensten dat draait in een Kubernetes-omgeving. Het verbindt dataproducenten en -consumenten via API's, waarmee grote hoeveelheden data op een georganiseerde en veilige manier kunnen worden verzameld, opgevraagd en gestreamd.

Momenteel worden de diensten van Obelisk uitgerold met behulp van één enkele Helm chart. De afhankelijkheden, zoals de databases, de berichtenservices en andere componenten, worden uitgerold en beheerd met een combinatie van Operators en Helm charts. Naast de geautomatiseerde uitrol via Operators en Helm, zijn er nog verschillende manuele stappen nodig om Obelisk volledig functioneel te maken. Zo moeten er verschillende onderdelen handmatig worden geïmporteerd en ingesteld, zoals gebruikersbeheer en databaseschema's.

Hoewel Helm veel flexibiliteit biedt in Kubernetes, worden ook problemen ondervonden in de huidige aanpak. Het werken met Helm kan omslachtig zijn, vooral wanneer het gaat om het beheren van afhankelijkheden en het correct doorgeven van bijvoorbeeld database-URL's en credentials naar Obelisk. Hiervoor is vaak aangepaste configuratie nodig, zoals het gebruik van aanvullende scripts. Operators bieden in sommige gevallen een oplossing, maar dit gaat

ten koste van controle over de uitrol. Bovendien zorgt het gebruik van operators voor extra complexiteit bij het debuggen van problemen.

Er is behoefte aan een gestroomlijnde en geautomatiseerde aanpak die handmatige stappen minimaliseert en de complexiteit reduceert, zodat gebruikers het opzetten en beheren van Obelisk efficiënter kunnen realiseren.

## Doelstelling van de masterproef:

De masterproef richt zich op het vereenvoudigen van de implementatie en configuratie van Obelisk-instanties binnen een Kubernetes-omgeving. In de eerste fase zal er een verkennend onderzoek worden uitgevoerd naar de meest geschikte hulpmiddelen en technologieën om dit te realiseren. Hierbij wordt gekeken naar verschillende methoden en technologieën, zoals Helm, Operators, Werf, Flux en andere. Het doel van deze verkenning is geen uitgebreide vergelijking, maar het identificeren van de technologieën die het beste aansluiten bij de specifieke eisen van Obelisk en die kunnen bijdragen aan het vereenvoudigen of gedeeltelijk oplossen van de huidige problemen, zoals de complexiteit in beheer en automatisering.

Daarnaast zal er een deelimplementatie van Obelisk uitgevoerd worden met de geselecteerde technologieën. Dit houdt in dat er een subset van Obelisk wordt opgezet, deze implementatie gebeurt met een selectie van de technologieën die voortkomen uit het verkennende onderzoek. Het doel is om de functionaliteit en efficiëntie van de geselecteerde technologieën in de praktijk te evalueren, zodat op basis hiervan een definitieve keuze kan worden gemaakt.

Op basis van de uitkomsten van dit verkennend onderzoek wordt een weloverwogen keuze gemaakt voor de meest geschikte technologie of een combinatie van meerdere. Vervolgens zal de masterproef zich richten op het grondig verder ontwikkelen en configureren van deze geselecteerde technologieën, zodat deze worden afgestemd op de vereisten en ingezet kunnen worden voor de implementatie van Obelisk.

Het uiteindelijke doel is om een gebruiksvriendelijke en grotendeels geautomatiseerde omgeving te creëren, waarin nieuwe gebruikers binnen een bepaalde tijdspanne een Obelisk-instantie op Kubernetes kunnen opzetten. Dit omvat het ontwikkelen van een duidelijke "getting started", waarbij handmatige stappen tot een minimum worden beperkt.

De configuratie zal generiek genoeg zijn om ook inzetbaar te worden op andere huidige en toekomstige platforms, mogelijks met een set van best practices en richtlijnen om de implementatie van Obelisk en vergelijkbare projecten te ondersteunen.

Planning en mijlpalen:

Taak 1	3 weken	<i>Deadline:</i> 21 oktober 2024	<b>Werkplan en technologieverkenning</b>
<b>Inhoud</b>			
<p>Het verkennen van de technologieën en tools die momenteel worden gebruikt voor de implementatie en het beheer van Obelisk, zoals Helm en Operators. Daarnaast wordt gestart met het verkennen van alternatieve tools zoals Werf, Flux, Keel en andere die mogelijk kunnen zorgen voor een verbeterde efficiëntie of automatisering.</p> <p>Ook wordt in deze fase het werkplan opgesteld,.</p>			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Het werkplan is opgesteld, nagekeken door de begeleider(s) en promotor en ingediend op Plato.</li> <li>• Inzicht verworven in de huidige implementatie van Obelisk en zijn afhankelijkheden en de huidige technologieën die gebruikt worden.</li> <li>• Een eerste verkenning van mogelijke alternatieve tools en technologieën die in latere fasen verder onderzocht zullen worden.</li> </ul>			

Taak 2	2 weken	<i>Deadline:</i> 4 november 2024	<b>Analyse van de huidige implementatie en werkwijze</b>
<b>Inhoud</b>			
<p>Een analyse van de huidige implementatie en het beheer van Obelisk binnen de Kubernetes-omgeving. Dit omvat het volgen van het bestaande stappenplan om Obelisk volledig op te zetten, inclusief de handmatige stappen zoals het instellen van gebruikersbeheer, databaseschema's,.... Ook zal worden gestart met het schrijven van de literatuurstudie voor de scriptie.</p>			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Een begrip van de huidige deployment cyclus en beheer van Obelisk.</li> <li>• Identificatie van de problemen en inefficiënties in de bestaande werkwijze.</li> </ul>			

Taak 3	3 weken	<i>Deadline:</i> 18 november 2024	<b>Verkenkend onderzoek naar technologieën</b>
<b>Inhoud</b>			
<p>Deze taak omvat een analyse van verschillende technologieën die kunnen bijdragen aan de verbetering van de implementatie en het beheer van Obelisk. De focus ligt op het onderzoeken van hoe deze tools omgaan met afhankelijkheden en hoe manuele stappen kunnen worden geminimaliseerd of geautomatiseerd. Ook zal er worden verder gewerkt aan de literatuurstudie voor de scriptie.</p>			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Een overzicht van technologieën en methoden, met specificaties.</li> <li>• Een deel van de literatuurstudie opgenomen in de scriptie.</li> </ul>			

Taak 4	4 weken	<i>Deadline:</i> 9 december 2024	<b>Deelimplementatie van Obelisk met verschillende geselecteerde technologieën</b>
<b>Inhoud</b>			
Er wordt een subset van Obelisk gedeployed, o.a. door bijvoorbeeld PostgreSQL op te zetten, schema's te initialiseren,.... Dit gebeurt met een selectie van de tools en technologieën die voortkomen uit het verkennende onderzoek in de vorige stap. Het doel is om de functionaliteit en efficiëntie van de geselecteerde tools in de praktijk te evalueren, zodat op basis hiervan een definitieve keuze kan worden gemaakt.			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Een tweetal deelimplementatie van Obelisk met verschillende geselecteerde technologieën.</li> <li>• Inzicht in de verschillen, sterke en zwakke punten .</li> <li>• Een definitieve keuze voor de technologie(ën) die verder gebruikt zal/zullen worden in het vervolg van de masterproef.</li> </ul>			

Taak 5	2 weken	<i>Deadline:</i> 16-21 december 2025	<b>Tussentijdse presentatie</b>
<b>Inhoud</b>			
Opstellen en voorbereiden van de tussentijdse presentatie.			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Tussentijdse presentatie is opgesteld en voorbereid.</li> <li>• Een duidelijk beeld geven van de status van het project aan de promotor en begeleiders.</li> </ul>			

Taak 6	3 weken	<i>Deadline:</i> 3 maart 2025	<b>Verdiepen in de gekozen technologie</b>
<b>Inhoud</b>			
Het verder verdiepen in de technologie die in de vorige fase is gekozen om verder mee te werken. Hierbij wordt meer gedetailleerd onderzocht hoe deze de huidige problemen minimaliseert, welke configuratiemogelijkheden ze bieden, etc. Er wordt een aanpak ontwikkeld voor de implementatie ervan en hoe die in het verdere verloop van de masterproef zal worden gebruikt.			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Een plan voor hoe de technologie gebruikt zal worden in het verdere verloop van deze masterproef.</li> </ul>			

Taak 7	3 weken	<i>Deadline:</i> 31 maart 2025	<b>Schrijven van scriptie – eerste 25 bladzijden</b>
<b>Inhoud</b>			
Het verder schrijven van ten minste de eerste 25 pagina's van de scriptie.			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
De scriptie, die uit minstens 25 pagina's bestaat, zal ter review worden verstuurd naar de promotor en begeleiders om er feedback over te ontvangen.			

Taak 8	6 weken	<i>Deadline:</i> 21 april 2025	<b>Ontwikkeling van een geautomatiseerde Setup-Flow</b>
<b>Inhoud</b>			
<p>Ontwikkelen van een gestroomlijnde en geautomatiseerde setup-flow voor nieuwe Obelisk-instanties, gebruikmakend van inzichten uit eerdere fases.</p> <p>De setup-flow zal onder andere volgende elementen bevatten: automatisering van credentials, bootstrap van databases en schema's, namelijk het ontwikkelen/gebruiken van mechanismen om automatisch de benodigde configuraties in te stellen bij het opzetten van een nieuwe Obelisk-instantie.</p>			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Een functionerende setup-flow voor nieuwe Obelisk-instanties.</li> <li>• Zo geautomatiseerd mogelijk beheer van credentials en connection strings.</li> <li>• Dit geldt ook voor het initialiseren van de databaseconfiguraties en configuraties van andere afhankelijkheden.</li> </ul>			

Taak 9	3 weken	<i>Deadline:</i> 28 april 2025	<b>Ontwikkeling van een "Getting Started" gids en best practices</b>
<b>Inhoud</b>			
<p>Het creëren van een duidelijke "getting started" gids zodat nieuwe gebruikers snel en efficiënt aan de slag kunnen gaan met het opzetten van hun omgeving. De gids zal stapsgewijze instructies bevatten, die het installatieproces gebruiksvriendelijker maken.</p> <p>Mogelijks wordt er ook een set van best practices en richtlijnen geschreven die niet alleen de implementatie van Obelisk ondersteunen, maar ook toepasbaar zijn op vergelijkbare projecten. Deze richtlijnen kunnen aanbevelingen bevatten over configuratie en efficiënt gebruik van de gebruikte technologie in deze masterproef.</p>			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Er is een eerste versie van een "getting started" gids beschikbaar die zal worden doorgestuurd naar de begeleiders en promotor en zal worden getest door enkele personen.</li> <li>• Mogelijks zal er ook een document met best practices en richtlijnen zijn.</li> </ul>			

Taak 10	2 weken	<i>Deadline:</i> 5 mei 2024	<b>Evaluatie van prestaties en gebruiksgemak</b>
<b>Inhoud</b>			
<p>Na de implementatie zal feedback worden verzameld van gebruikers die de gids hebben gevolgd om hun omgeving op te zetten. Deze feedback zal worden gebruikt om verbeterpunten te identificeren. Ook zullen de prestaties van het installatieproces, zoals de snelheid en efficiëntie van de setup, worden beoordeeld. Op basis van deze evaluatie zullen verbeteringen worden gedaan in de gids en installatieproces.</p>			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• Na de evaluatie zullen verbeterpunten geïdentificeerd worden.</li> <li>• De toegepaste feedback zal de gebruiksvriendelijkheid verhogen.</li> </ul>			

Taak 11	3 weken	<i>Deadline:</i> 22 mei 2025	<b>Indienen van de eerste versie van de scriptie (95% afgewerkt)</b>
<b>Inhoud</b>			
Het verder werken aan de eerste versie van de scriptie, waarin alle belangrijke onderdelen zijn opgenomen, de scriptie zal voor ongeveer 95% afgewerkt zijn.			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• De scriptie (95% afgewerkt) wordt doorgestuurd naar de promotor en begeleiders voor feedback.</li> <li>• Ontvangen feedback van de promotor en begeleiders voor verdere verbetering.</li> </ul>			
Taak 12	2 weken	<i>Deadline:</i> 5 juni 2025	<b>Finaliseren van de scriptie</b>
<b>Inhoud</b>			
Het afronden van de scriptie en het toepassen van de ontvangen feedback van de promotor en begeleiders.			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• De definitieve versie van de scriptie.</li> <li>• De scriptie wordt ingediend op Plato.</li> </ul>			
Taak 13	3 weken	<i>Deadline:</i> 25-27 juni 2025	<b>Masterproefexamen</b>
<b>Inhoud</b>			
Het opstellen van de presentatie voor het masterproefexamen en het voorbereiden van de mondelinge verdediging.			
<b>Belangrijkste resultaten, deliverables of inzichten na deze fase:</b>			
<ul style="list-style-type: none"> <li>• De presentatie van de masterproef is opgesteld, inclusief voorbereiding op vragen van de jury.</li> <li>• Presentatie en verdediging zelf.</li> </ul>			

### Contactmomenten:

De communicatie met de begeleiders (en promotor) vindt om de twee weken plaats in iGent. Afhankelijk van de voortgang en de duidelijkheid van de situatie kan de tijd tussen de meetings wat langer zijn. Ook is het mogelijk om deze contactmomenten via Teams te doen. Indien er geen fysieke afspraak plaatsvindt, zal er minstens een tweewekelijkse rapportering via e-mail plaatsvinden. Alle e-mailcommunicatie zal gericht zijn aan zowel de begeleiders als de promotor, zodat iedereen steeds op de hoogte is van de voortgang.

Gantt chart:

	Taak1	Taak2	Taak3	Taak4	Taak5	Taak6	Taak7	Taak8	Taak9	Taak10	Taak11	Taak12	Taak13
Week 1: 23/09/24													
Week 2: 30/09/24	X												
Week 3: 7/10/24	X												
Week 4: 14/10/24	X												
Week 5: 21/10/24		X											
Week 6: 28/10/24		X	X										
Week 7: 4/11/24			X										
Week 8: 11/11/24			X	X									
Week 9: 18/11/24				X									
Week 10: 25/11/24				X									
Week 11: 2/12/24				X	X								
Week 12: 9/12/24					X								
Week 13: 16/12/24					X								
Kerstvakantie													
Examenperiode januari													
Week 1: 10/02/25						X							
Week 2: 17/02/25						X							
Week 3: 24/02/25						X		X					
Week 4: 3/3/25								X					
Week 5: 10/3/25							X	X					
Week 6: 17/3/25							X						
Week 7: 24/3/25							X						
Week 8: 31/3/25								X					
Paasreces w1: 7/4/25								X	X				
Paasreces w2: 14/4/25								X	X				
Week 9: 21/4/25									X	X			
Week 10: 28/4/25										X			
Week 11: 5/5/25											X		
Week 12: 12/5/25											X		
Week 13: 19/5/25											X		
Week 14: 26/5/25												X	
Week 15: 2/6/25												X	
Week 16: 9/6/25													X
Week 17: 16/6/25													X
Week 18: 23/6/25													X