



HoGent

Faculteit Bedrijf en Organisatie

Wanneer is EventSourcing een meerwaarde voor een bedrijf zoals Skedify, waar ze gespecialiseerd zijn in online scheduling?

Robin Malfait

Scriptie voorgedragen tot het bekomen van de graad van
professionele bachelor in de toegepaste informatica

Promotor:
Lieven Smits
Co-promotor:
Mathias Verraes

Instelling: —

Academiejaar: 2016-2017

Tweede examenperiode

Faculteit Bedrijf en Organisatie

Wanneer is EventSourcing een meerwaarde voor een bedrijf zoals Skedify, waar ze gespecialiseerd zijn in online scheduling?

Robin Malfait

Scriptie voorgedragen tot het bekomen van de graad van
professionele bachelor in de toegepaste informatica

Promotor:
Lieven Smits
Co-promotor:
Mathias Verraes

Instelling: —

Academiejaar: 2016-2017

Tweede examenperiode

Samenvatting

De maatschappij waarin we leven draait om data. Data en informatie zijn zeer veel geld waard, daarom is het belangrijk dat er goed mee om wordt gegaan. In veel systemen waar relationele modellen gebruikt worden gaat er data verloren. Telkens wanneer er een UPDATE of DELETE sql statement wordt uitgevoerd is deze informatie er niet meer. Door middel van EventSourcing gaat er geen data verloren. Het is belangrijk om informatie bij te houden, dit kan altijd interessant zijn in de toekomst wanneer er specifieke business vragen komen waar er momenteel nog geen antwoord op is. Eerst wordt er gekeken naar al de delen die deel uitmaken van een EventSourced applicatie. Dit gaat van CQS, CQRS tot de effectieve onderdelen van EventSourcing. Er wordt ook uitleg gegeven over Skedify zodat de concrete businesscase duidelijk gescoped wordt.

Voorwoord

Ik ben al een lange tijd gepassioneerd door data en informatie. Door gebruik te maken van EventSourcing gaat er geen informatie verloren, en is er een mogelijkheid om te tijdreizen naar het verleden tot op het heden. Het idee van deze bachelorproef kwam uit gesprekken met een aantal mensen via sociale media Twitter.

Dankzij Twitter heb ik ook mijn co-promotor, Mathias Verraes leren kennen en ik zou hem heel graag willen bedanken voor het realiseren van deze bachelorproef. Ik kon altijd met al mijn vragen raad bij hem, en hij zorgde er ook voor dat ik naar Meetups kon gaan om meer informatie te verschaffen voor mijn bachelorproef.

Daarnaast zou ik ook graag mijn promotor, Lieven Smits willen bedanken voor het benadrukken van mijn spellingsfouten en deze bachelorproef tot een goed einde te brengen.

Inhoudsopgave

1	Inleiding	9
1.1	Stand van zaken	9
1.2	Probleemstelling en Onderzoeksvragen	9
1.3	Opzet van deze bachelorproef	10
2	Methodologie	11
2.1	Literatuurstudie	11
2.2	Proof of concept	12
3	Skedify	13
4	CQRS	15
5	EventSourcing	17
5.1	Aggregates	17

5.2	Value Objects	18
5.3	Domain Events	18
5.4	Invariants	18
5.5	Eventual Consistency	19
5.6	Event Store	19
5.7	Audit Log	19
5.8	Projections	20
5.9	Schaalbaarheid	20
5.10	Voordelen	20
6	Conclusie	23
	Bibliografie	25

1. Inleiding

De inleiding moet de lezer alle nodige informatie verschaffen om het onderwerp te begrijpen zonder nog externe werken te moeten raadplegen (Pollefliet, 2011). Dit is een doorlopende tekst die gebaseerd is op al wat je over het onderwerp gelezen hebt (literatuuronderzoek).

Je verwijst bij elke bewering die je doet, vakterm die je introduceert, enz. naar je bronnen. In \LaTeX kan dat met het commando `\textcite{}` of `\autocite{}`. Als argument van het commando geef je de “sleutel” van een “record” in een bibliografische databank in het Bib \TeX -formaat (een tekstbestand). Als je expliciet naar de auteur verwijst in de zin, gebruik je `\textcite{}`. Soms wil je de auteur niet expliciet vernoemen, dan gebruik je `\autocite{}`. Hieronder een voorbeeld van elk.

Knuth (1998) schreef een van de standaardwerken over sorteer- en zoekalgoritmen. Experts zijn het erover eens dat cloud computing een interessante opportuniteit vormen, zowel voor gebruikers als voor dienstverleners op vlak van informatietechnologie (Creager, 2009).

1.1 Stand van zaken

1.2 Probleemstelling en Onderzoeksvragen

Data is belangrijk, rapportering is belangrijk. Voorspellen van vragen die de business kan vragen binnen dit en 5 jaar is onmogelijk. Daarom is het belangrijk om alle informatie bij te houden om hier interessante rapportering op te kunnen uitvoeren. In een systeem

met een standaard relationele databank wordt niet alle informatie bijgehouden. Bij het verwijderen of wijzigen van informatie is de voorgaande informatie verloren gegaan. Alle informatie bijhouden is dus interessant, maar wat zijn hier de voor en nadelen van?

De onderzoeksvraag luidt als volgt: Wanneer is EventSourcing een meerwaarde voor een bedrijf zoals Skedify, waar ze gespecialiseerd zijn in online scheduling?

Hierbij worden ook volgende subvragen beantwoord:

- Wat is het verschil tussen een systeem met relationele databank modellen en een systeem met EventSourcing?
- Hoe kan een applicatie die gebruik maakt van EventSourcing getest worden?
- Welke delen moeten er EventSourced worden, en hoe kan je deze delen bepalen?

1.3 Opzet van deze bachelorproef

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt de methodologie toegelicht en worden de gebruikte onderzoekstechnieken besproken om een antwoord te kunnen formuleren op de onderzoeksvragen.

In Hoofdstuk 6, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeksvragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

2. Methodologie

Omdat er met EventSourcing gewerkt zal worden om een proof of concept te maken, moet er eerst gekeken worden naar wat EventSourcing is en wat het allemaal inhoudt. Naast EventSourcing zijn er ook concepten zoals CQRS die aan bod moeten komen. Daarnaast draait het in dit onderzoek rond Skedify, het bedrijf die als concrete case gebruikt zal worden. In het deel rond Skedify, kan er meer te weten gekomen worden in verband met wat ze doen en hoe ze te werk gaan. Daarna zal er gekeken worden naar de huidige manier van werken. Tot slot gaan we alles samen voegen: Skedify en de huidige manier van werken, en Skedify en EventSourcing. Hierna kunnen we dan de voor- en nadelen van beide mogelijkheden afdelen om te zien of EventSourcing nu effectief een meerwaarde biedt voor Skedify. Het onderzoek is dus in grote delen opgesplitst: Skedify, EventSourcing, huidige manier van werken en de conclusies.

In het deel over Skedify zal er onderzocht worden wat ze allemaal doen, en hoe ze dit doen. Er zal ook een gesprek komen met de business kant van Skedify om vragen te kunnen beantwoorden en om duidelijke conclusies te kunnen trekken.

In het deel over EventSourcing gaan we iets dieper in hoe EventSourcing juist werkt en hoe het gebruikt kan worden.

2.1 Literatuurstudie

...TODO...

2.2 Proof of concept

Voor de proof-of-concept zullen bepaalde functies van Skedify geschreven worden in het huidige systeem en in een EventSourcing systeem. Daarbij zal er gekeken worden naar voordelen en nadelen van beide systemen.

3. Skedify

Skedify is een Gentse start up gespecialiseerd in het maken van afspraken. Deze bachelorproef gaat op zoek naar een antwoord of EventSourcing een meerwaarde kan bieden voor Skedify. Skedify zorgt er voor dat een persoon in eender welke sector een afspraak kan maken op zeer korte tijd voor eender welk probleem, en dit met de juiste contact persoon van het bedrijf. Skedify is een business-to-business tool om dynamisch formulieren te gaan opbouwen, in meerdere talen, voor verschillende onderwerpen. Op deze manier is de tool heel dynamisch en kan ze dus voor heel veel doeleinden worden gebruikt. Alles wordt in een relationele databank opgeslagen.

Een van de mogelijkheden die Skedify biedt is het wijzigen van een afspraak, het kan dus interessant zijn om te weten hoeveel keer dit gebeurt en of er specifieke maatregelen kunnen genomen worden.

4. CQRS

CQRS is een term uitgevonden door Greg Young en is de opvolger van CQS. CQS betekend Command Query Seperation en is een principe die uitgevonden is door Bertrand Meyer (1988). Als er gekeken wordt naar CQS is dit al een eerste vorm van goede code schrijven. CQS zorgt er voor dat getters en setters gescheiden zijn. Getters zijn strict bedoeld om een waarde uit de huidige state te halen en deze terug te geven. Setters zijn bedoeld om een wijziging te doen (of een algemene actie uit te voeren), setters geven geen waarde terug maar void.

Het grootste probleem met CQS is dat developers acties uitvoeren in getters wat voor bugs kan zorgen, bijvoorbeeld het huidige project is een spel en de totale score wordt opgevraagd. Dan ziet de getter naam er als volgt uit `getTotalScore`, maar als deze getter een andere methode oproept zoals `calculateTotalScore`, en deze methode wijzigt de state, dan kan elke keer dat `getTotalScore` opgeroepen wordt het resultaat anders zijn.

Een tweede probleem is dat getters en setters in een en dezelfde klasse zijn gedefinieerd. Er is geen stricte scheiding tussen de lees kant en de schrijf kant.

Dit is waar CQRS komt kijken, Command Query Responsibility Segregation. CQRS zorgt er voor dat de lees kant en de schrijf kant strict gescheiden zijn. Het zijn bijna 2 applicaties die naast elkaar staan. Een command ligt aan de schrijfkant en een query aan de leeskant. De leeskant gaat zijn informatie halen bij de databank (of een andere vorm van opslagmechanisme), dit kan via sql queries, ORM tools, enzovoort. De manier waarop dit gebeurt staat volledig los van hoe de schrijfkant communiceert met het opslagmechanisme.

Heel veel applicaties, onder andere ook die van Skedify zijn veel intensiever aan de leeskant dan aan de schrijfkant. Vermits de lees- en schrijfkant nu strict gescheiden zijn kan er gebruik gemaakt worden van schalingsmechanismen. Beter nog, de leeskant en schrijfkant

kunnen individueel geschaald worden.

CQRS speelt een grote rol bij EventSourcing, vandaar dit korte hoofdstuk.

5. EventSourcing

EventSourcing is geen nieuwe uitvinding, maar wordt al jaren gebruikt in andere sectoren. Bekende voorbeelden zijn de bankindustrie, wetgeving, en patiënten fisches die opgeslagen zijn bij dokters. EventSourcing, is zoals de naam al verklapt, dat events de bron zijn van heel de applicatie. Als er naar de bankindustrie wordt gekeken dan is het bedrag op iemand zijn rekening niet een getal dat opgeslagen is in een databank, of toch niet de enige vorm hier van. Er is een lijst van transacties die tot dit getal komen. Het zal ook wel in een databank opgeslagen zijn, maar dit is den puur een caching mechanisme.

Het is zo dat nu pas, de laatste jaren, EventSourcing opkomt in de informatica sector. In de volgende hoofdstukken zal er dieper ingegaan worden op onderdelen van EventSourcing.

5.1 Aggregates

Aggregaten zijn de entiteiten die behoren tot het domein. Een aggregaat is een entiteit die zal zorgen dat invariants (besproken in 5.4) gegarandeerd zullen worden. Eens dit het geval is zal die er voor zorgen dat de juiste domain events opgeslagen worden en dus de state van de applicatie wijzigt.

Deze aggregaten moeten goed gekozen worden, zodat er geen gevaar is dat er 'god' aggregaten ontstaan. Een 'god' aggregaat is een aggregaat die alles in een applicatie doet, in veel gevallen is dit een 'User' aggregaat.

In het geval van Skedify zijn goede aggregaten: Appointment, Subject, Question, Answer, Office, ...

5.2 Value Objects

Value objects zijn objecten die zorgen dat een bepaalde waarde altijd in een geldige staat is. Bijvoorbeeld, er kan een value object zijn voor een Email. Dit email adres moet ten alle tijden geldig zijn. Dit kan afgedwongen worden door in de constructor van een klasse een waarde te ontvangen en deze te controleren. Indien deze waarde fout is wordt er een exceptie gegoooid. Deze exceptie zal dan opgevangen worden in de lagen daarboven.

Een belangrijke regel bij Value objects is dat deze geen identiteit bevatten, ze hebben geen unique identifier en zijn dus geen entiteit. Value objects kunnen wel tot een entiteit behoren zoals Bijvoorbeeld een User entiteit.

5.3 Domain Events

Domain events zijn een heel belangrijk onderdeel van EventSourcing. Domain events zijn de effectieve events die zullen opgeslagen worden in een append-only database. Een event is iets dat gebeurt is en nooit meer kan veranderen. Er zijn een paar heel belangrijke eigenschappen aan deze events.

- Ze bevatten een unieke id, die op voorhand vastgelegd is. Dit kan een GUID (Globally Unique Identifier) zijn.
- Ze bevatten enkel de data die gewijzigd is ten opzichte van de vorige versie.
- Alle data die ze bevatten, is correct en kan niet meer aangepast worden.

Domain events worden opgeslagen in een append-only database, maar wat als er een fout gemaakt is gemaakt? Indien een fout is opgetreden, moet er een nieuw domain event gemaakt worden, die het vorige event corrigeerd. Op deze manier blijft al je data correct, en verlies je geen belangrijke informatie. Er is ook niet geprutst met de historie van deze events, wat heel belangrijk is.

Domain events hebben ook een naam, deze naam is heel specifiek in wat er gebeurt is. Het is ook belangrijk dat deze naam in de verleden tijd is opgesteld. Een event is tenslotte gebeurd. Als er in context van Skedify gesproken wordt, dan is AppointmentWasRescheduled een goede naam voor een domain event. Het is ook belangrijk dat er geen CRUD (Create Read Update Delete) events gemaakt worden zoals OfficeWasCreated, want we hebben niet effectief een office gemaakt, we hebben er een toegevoegd. Een betere naam zou zijn OfficeWasAdded.

5.4 Invariants

Invariants zijn regels die opgelegd zijn door de business. Invariants worden gecontroleerd alvorens een domain event opgeslagen wordt. Elke invariant moet goedgekeurd, want eens een domain event opgeslagen is, kan dit niet meer ongedaan gemaakt worden. Er kan wel een nieuw domain event opgeslagen worden om deze wijziging te niet te doen. Invariant

controle wordt uitgevoerd, en telkens wanneer een input niet aan deze invariant voldoet moet er een exception gegoooid worden.

5.5 Eventual Consistency

...TODO...

5.6 Event Store

De EventStore is een belangrijk concept bij EventSourcing. De EventStore is de plaats waar alle DomainEvents opgeslagen zullen worden. De EventStore kan elk soort databank zijn, er is ook een speciaal gemaakt en is te vinden op <https://geteventstore.com>. Voor deze bachelorproef, en de proof-of-concept zal er gebruik gemaakt worden van een simpele EventStore in mysql. Een EventStore is helemaal niet zo moeilijk qua structuur, het bevat volgende minimale velden:

- id, een simpele id, die autoincrementeel kan zijn
- stream_id, een id die bij een Aggregate hoort, dit is een GUID. Alle events die horen tot een bepaalde aggregate zullen dezelfde stream_id hebben.
- stream_version, een versie die incrementeel is per aggregaat. Dit zorgt er voor dat event in de juiste volgorde kunnen worden afgespeeld indien nodig.
- event_name, de naam van het DomainEvent dat opgeslagen moet worden. Via deze naam kan het juiste object weer opgebouwd worden in onze code.
- payload, dit bevat de effectieve data van het DomainEvent. Dit kan een json object zijn. Dit wordt mee gebruikt, naast de naam om het object weer op te bouwen.
- recorded_at, een timestamp waarop het event opgeslagen is geweest. Dit kan handig zijn voor de audit log.

5.7 Audit Log

Wanneer er gebruik gemaakt wordt van EventSourcing, krijg je door de append-only EventStore gratis een audit log. Deze audit log is een bewijs van alle events die tot een bepaalde waarde (projection) lijden. Wanneer de EventStore op een WORM drive (Write Once Read Many) gezet wordt, dan kan er niet geknoeid worden met deze lijst van events. Het voordeel hieraan is dat er gemakkelijk bewijs kan geleverd worden naar andere partijen indien dat nodig is. Stel er komt een rechtzaak tussen 2 bedrijven omdat er geknoeid is met de data. Dan kan dit zwart op wit bewezen worden dat dit niet het geval is.

Deze audit-log heeft ook nog andere voordelen, als er een fout zit in de applicatie en de applicatie moet gedebugged worden, dan kan men de audit log nemen en deze opnieuw afspelen tot de huidige staat. Zo kan er per event gekeken worden of de uitkomst van de huidige staat al dan niet correct is.

5.8 Projections

Projections zijn een projectie van de huidige staat die berekend is van alle voorbijgaande events. Bij een systeem waar er met een relationele databank wordt gewerkt, is de databank zelf de projectie. Met als grote verschil dat daar ook de databank de enige echte bron van waarheid is. Dat is niet het geval bij een systeem met EventSourcing, daar is de bron van waarheid de events die opgeslagen zijn.

Projections hebben een heel groot voordeel dat er meerdere kunnen zijn. Er kan een databank gegenereerd worden voor de business kant die dan allerlei statistieken kan bekijken. Er kan ook een databank gegenereerd worden voor werknemers die met een bepaalde applicatie werken. Er kan zelf een databank of ander opslagsysteem gevuld worden los van de huidige databanken voor de andere partijen. Op deze manier staan we los van de databank.

Deze projecties zijn een berekening van de events, dit wilt ook zeggen dat we niet moeten optimaliseren voor de schrijfkant (zie cqrs). Er kan geoptimaliseerd worden voor de leeskant, er is geen nood om een aantal tabellen elke keer te gaan joinen, als er rechtstreeks naar 1 tabel geschrijven wordt.

5.9 Schaalbaarheid

Een systeem dat gebruik maakt van EventSourcing, kan geschaald worden, het schalen is relatief gemakkelijk door volgende redenen. De EventStore is een append-only database, er kan altijd een kopie genomen worden van deze database. Er moet niet gekeken worden of er al dan niet wijzigingen zijn in de data en welke de correcte data is. Synchroniseren van meerdere EventStores is een kwestie van ontbrekende rijen aan te vullen. Hiervoor kan er een mechanisme gebruikt worden zoals "geef alle events na id X". Schalen van de applicatie zelf kan als volgt gebeuren. Er kunnen projecties bijgemaakt worden die zichzelf gaan 'voeden' met alle events uit de EventStore. Van zodra ze up-to-date zijn met de EventStore kunnen ze als projection listener geregistreerd worden, om live naar de events te luisteren.

5.10 Voordelen

Er zijn heel wat voordelen verbonden aan EventSourcing. Een groot voordeel is dat je van technologie kan wisselen voor de projecties. Stel er werd gebruik gemaakt van MySQL om de huidige staat in op te slaan. Wanneer er gemerkt wordt dat er nood is aan een graph database in plaats van een relationele database dan kan er een projectie bijgemaakt worden die simultaan met de mysql database loopt. Zodra de graph database helemaal gevoed is met de events uit de EventStore, dan kan de MySQL databank weggegooid worden. Vermits een projectie opgebouwd is in code, en in een versiebeheersysteem zit, moet er niet nagedacht worden om een eventuele back-up te nemen van deze MySQL databank. Indien

er later opnieuw van deze databank gebruik gemaakt moet worden, dan kan deze projectie terug ingeladen worden en helemaal opnieuw opgebouwd worden. Er werd net gesproken over de graph database, stel dat de applicatie al 3 jaar draait en nu komt deze graph database er bij. Zodra de events allemaal gevoed zijn in de graph database, dan staat deze op het punt alsof ze al van in begin geprogrammeerd was. Dit is het grote voordeel van al de events bij te houden. Een ander voordeel van EventSourcing is dat je enkel de EventStore moet backuppen. Indien elke server corrupt of kapot gaat en alle MySQL instanties zijn kapot, dan kunnen deze opnieuw opgebouwd worden. Dit is allemaal mogelijk omdat de EventStore, de single source of truth is.

6. Conclusie

...TODO...

Bibliografie

- Creeger, M. (2009). CTO Roundtable: Cloud Computing. *Communications of the ACM*, 52(8), 50–56.
- Knuth, D. E. (1998). *The art of computer programming, volume 3: (2nd ed.) sorting and searching*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc.
- Polleffiet, L. (2011). *Schrijven van verslag tot eindwerk: do's en don'ts*. Gent: Academia Press.

Lijst van figuren

Lijst van tabellen