

DevOps: Operational data for developers

Stef Van Gils

Thesis voorgedragen tot het behalen
van de graad van Master of Science
in de ingenieurswetenschappen:
computerwetenschappen,
hoofdspecialisatie Software
engineering

Promotor:
Prof. W. Joosen

Assessor:
Dimitri Van Landuyt

Begeleider:
Dimitri Van Landuyt

© Copyright KU Leuven

Zonder voorafgaande schriftelijke toestemming van zowel de promotor als de auteur is overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie i.v.m. het overnemen en/of gebruik en/of realisatie van gedeelten uit deze publicatie, wend u tot het Departement Computerwetenschappen, Celestijnenlaan 200A bus 2402, B-3001 Heverlee, +32-16-327700 of via e-mail info@cs.kuleuven.be.

Voorafgaande schriftelijke toestemming van de promotor is eveneens vereist voor het aanwenden van de in deze masterproef beschreven (originele) methoden, producten, schakelingen en programma's voor industrieel of commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden.

Voorwoord

Dit is mijn dankwoord om iedereen te danken die mij bezig gehouden heeft. Hierbij dank ik mijn promotor, mijn begeleider en de voltallige jury. Ook mijn familie heeft mij erg gesteund natuurlijk.

Stef Van Gils

Inhoudsopgave

Voorwoord	i
Samenvatting	iii
Lijst van figuren en tabellen	iv
1 Architectuur	1
1.1 Architecturale beslissingen	1
1.2 Klassediagram	6

Samenvatting

In dit **abstract** environment wordt een al dan niet uitgebreide samenvatting van het werk gegeven. De bedoeling is wel dat dit tot 1 bladzijde beperkt blijft.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Lijst van figuren en tabellen

Lijst van figuren

1.1	Component diagram Tracklytics library.	2
1.2	Deployment diagram Tracklytics library.	2
1.3	Flow diagram 1.	5
1.4	Flow diagram 2.	5
1.5	Flow diagram 3.	6
1.6	Klassediagram Tracklytics library.	7

Lijst van tabellen

Hoofdstuk 1

Architectuur

1.1 Architecturale beslissingen

Zoals beschreven in vorige secties gaan we een library ontwikkelen om applicaties te kunnen monitoren. Tracklytics is de naam van de library en wordt in het vervolg van dit document gebruikt om de library aan te duiden.

In deze sectie worden de architecturale beslissingen die gemaakt zijn uit de doeken gedaan. Aan de hand van diagrammen wordt besproken hoe Tracklytics conceptueel werkt.

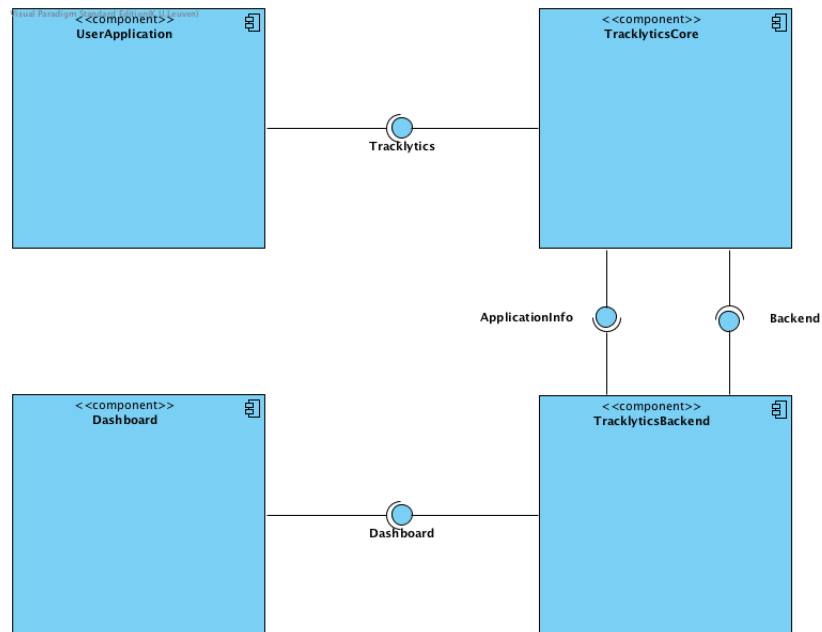
Tracklytics bestaat uit volgende hoofdcomponenten:

- UserApplication: De applicatie die de developer wil monitoren
- TracklyticsCore: De Tracklytics library die in de applicatie verwerkt is.
- TracklyticsBackend: De backend die de informatie verwerkt die van de TracklyticsCore komt.
- Dashboard: Een dashboard om alle informatie weer te geven.

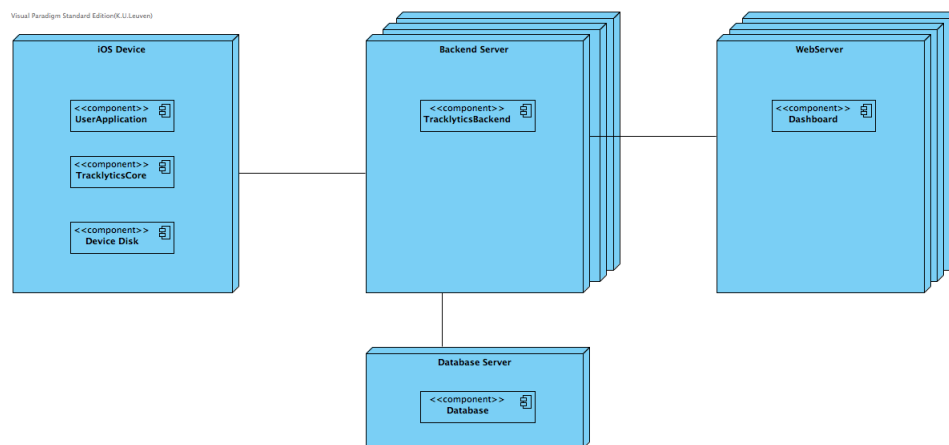
1.1.1 Deployment Diagram

Het deployment diagram toont op welke nodes de componenten draaien. Zoals je kan zien draait de UserApplication en de TracklyticsCore op het toestel van de gebruiker zelf (een iOS device). De TracklyticsBackend draait op een backend server die gerepliceerd kan worden. Deze staat in verbinding met een Database server waar alle data wordt opgeslagen. Het Dashboard draait op een webserver. Deze staat in verbinding met de backend server om de data te kunnen ophalen.

1. ARCHITECTUUR



FIGUUR 1.1: Component diagram Tracklytics library.



FIGUUR 1.2: Deployment diagram Tracklytics library.

1.1.2 Component Diagram

Het component diagram 1.1 duidt aan welke componenten een rol spelen in het bouwen van de Tracklytics library. In de volgende secties worden de verschillende componenten beschreven en hun rol wordt uitgelegd.

UserApplication

De UserApplication component bestaat uit de applicatie ontwikkeld door de developers. Deze verschilt per applicatie. De UserApplication stuurt data door naar de TracklyticsCore component via de Tracklytics interface. De interface wordt in de volgende sectie besproken. De developers kiezen zelf welke elementen getrackt en gemonitord worden. Deze component draait op het toestel van de gebruiker.

TracklyticsCore

Monitoring De TracklyticsCore component stelt de library voor die in deze thesis ontwikkeld wordt. De component draait op het toestel van de gebruiker. De TracklyticsCore component verzamelt informatie die hij doorkrijgt van de UserApplication component, slaat deze tijdelijk op het toestel op om deze dan later door te sturen naar de TracklyticsBackend component via de Backend interface. Deze interface wordt in de volgende sectie besproken.

De TracklyticsCore component biedt een Tracklytics interface aan. De UserApplication kan deze interface gebruiken om data te verzamelen. De TracklyticsCore stuurt deze later naar de server zoals besproken wordt in volgende sectie.

De interface bestaat uit het gebruik van 5 elementen, namelijk:

- Counter
- Meter
- Histogram
- Timer
- Gauge

Deze elementen worden verder besproken in het hoofdstuk over het klassediagram [1.2](#).

Data availability Smartphones vallen weleens zonder internet connectiviteit. Tracklytics zendt de data over het internet naar de backend om deze op te slaan in een database. Als de internetverbinding zou wegvallen, dan zou er mogelijks belangrijke data verloren gaan. Om dit scenario te voorkomen slaan we de data tijdelijk op op de harde schijf van de telefoon. Indien de smartphone dan terug verbinding met het internet heeft gemaakt kan de data alsnog gesynchroniseerd worden met de backend. De data wordt tijdelijk opgeslagen, dus nadat de data gesynchroniseerd is wordt deze van de smartphone verwijderd.

On Demand Het is de bedoeling de developer zoveel mogelijk vrijheid te geven in het monitoren van de applicatie. Om deze vrijheid te verhogen voegen we het On Demand aan- of uitzetten van het monitoren toe aan Tracklytics. Een developer kan vanuit het dashboard aangeven of de applicatie gemonitord moet worden of niet. Dit kan gebruikt worden om op de piekdagen of piekmomenten het monitoren aan te zetten. Zo kan uitgezocht worden of er een bottleneck bestaat als het gebruik van de applicatie piekt.

Indien een developer ontdekt dat er geen bottleneck of problemen zijn met de app en deze niet verder gemonitord moet worden, dan is het handig om de monitoring uit te kunnen zetten. Dit zorgt ook voor een prestatieboost, omdat er niet meer gesynchroniseerd moet worden naar de server en de data moet niet meer op het toestel verwerkt en opgeslagen worden door Tracklytics.

Aggregatie De data die doorgestuurd wordt naar de server zegt op zichzelf niets. Om een goede analyse weer te geven in het dashboard moet deze data geaggregeerd worden. Er zijn twee plaatsen waar deze aggregatie uitgevoerd kan worden, namelijk op de smartphone zelf in de Tracklytics library of in de backend.

De reden om de aggregatie op de smartphone uit te voeren is dat er zo minder data gecollecteerd moet worden in de backend om de data visueel voor te stellen. Er is dus een prestatiewinst om de aggregatie op de smartphone uit te voeren. Een nadeel is dat deze aggregatie meer CPU tijd van de smartphone gebruikt dan als we de aggregatie in de backend uitvoeren. Er is dus een goede afweging nodig waar deze aggregatie gebeurt.

Een functie in het dashboard zou ervoor kunnen zorgen dat de developer de optie krijgt om te kiezen of deze aggregatie op de telefoon of in de backend uit te voeren. Dit geeft de developer ook weer meer vrijheid om deze keuzes te maken.

AB Testing AB testing is een mechanisme dat grote bedrijven, zoals facebook, gebruiken om nieuwe features uit te rollen naar de gebruikers. Het mechanisme werkt als volgt: er bestaat een versie A en een versie B van de software met B de nieuwere versie. AB testing zorgt ervoor dat de uitrol van de versie B geleidelijk verloopt. De gebruikers van de software worden dus in twee (of meerdere) groepen opgedeeld door een eigenschap. Deze eigenschap kan vanalles zijn, namelijk: geografische locatie, ingestelde taal, de gebruikte browser, het type toestel, enz. Indien er dan een probleem met versie B is, dan bestaat dit enkel bij de groep die versie B al verkregen is en dus niet bij alle gebruikers. Hierdoor merkt enkel die groep dat er een probleem is en kan versie B aangepast worden om dit probleem op te lossen of eventueel kan ervoor gezorgd worden dat iedereen terug versie A gebruikt.

In mobiele applicaties is het moeilijker om echt aan AB testing te doen, omdat deze applicaties meestal statisch zijn, er moet een update in de app store komen om een nieuwe versie uit te rollen. Dit staat pal tegenover websites die hun webpaginas van een server halen en dus bij elk vernieuwen van een webpagina helemaal anders

zijn. Een workaround van dit probleem is dat de mobiele applicatie de code van de nieuwe versie al bevat en ook nog de oude code erin heeft staan. Tracklytics kan dan een functie in het dashboard aanbieden om de groepen op te delen in twee (of meerdere) groepen op basis van eigenschappen die de developer kan kiezen. In de mobiele applicatie moeten we dus dat onderscheid kunnen maken welke code aangeroepen moet worden. Het gemakkelijkste is om een codenaam aan de versie toe te voegen, zodat er meerdere versies van de code aanwezig kan zijn. Het nadeel van dit mechanisme is dat de applicatie groter is dan hij zou zijn moest de applicatie enkel de code bevatten die voor die bepaalde groep nodig is.

Security & Privacy Een belangrijk punt is de privacy van de gebruiker. Zoals eerder al aangehaald is NewRelic TODO een closed source library, wat wil zeggen dat developers niet weten welke informatie er doorgestuurd wordt naar de backend. Dit gegeven zorgt ervoor dat er privacygevoelige applicaties deze library niet zou mogen gebruiken, omdat er gebruikersinformatie gelekt zou kunnen worden naar de backend van NewRelic.

Om deze reden is ervoor gekozen dat Tracklytics open source is. Zo kunnen developers zeker zijn welke informatie er wordt doorgestuurd naar de backend en is de privacy van de gebruiker wel gegarandeerd langs de monitoring kant. De metadata die gecollecteerd wordt door Tracklytics bestaat uit: het type toestel, de versie, de naam van de applicatie en de bundle naam, de UDID TODO van het toestel, de datum en het type connectie waarop het toestel zich bevindt. Zo wordt ervoor gezorgd dat de privacy van de gebruiker gegarandeerd wordt en dat er toch genoeg gegevens zijn om deze te representeren op een dashboard.

Naast privacy is security ook belangrijk. Er staat namelijk een verbinding tussen de mobiele applicatie en de backend. Het zou dus niet mogen dat de doorgezonden data door een onrechtmatig iemand gecollecteerd wordt of dat hiermee geknoeid wordt. Om deze situaties te voorkomen is ervoor gekozen om via een HTTPS verbinding te werken. Deze verbinding zorgt uit zichzelf voor een veilige verbinding tussen begin- en eindpunt. Als extra veiligheid werkt Tracklytics met HTTP Post in plaats van HTTP Get, zodat de doorgegeven data niet zichtbaar is in de URL naar een backend bestand.

TracklyticsBackend

De TracklyticsBackend component representeert de server kant van de library. Hierop draait de code om de getrackte en gemonitorde informatie op te slaan in de database. De TracklyticsBackend component bevat dus ook de database om alle informatie in op te slaan.

De TracklyticsBackend component biedt twee interfaces aan, namelijk: Backend en Dashboard.

De Backend interface wordt gebruikt om data uit te wisselen tussen de TracklyticsCore en de TracklyticsBackend. Dit is de data die de TracklyticsCore verzameld heeft van de applicatie. De TracklyticsBackend verwerkt deze informatie en slaat

deze op in de database.

De Dashboard interface wordt door het Dashboard component gebruikt om data op te halen om te kunnen weergeven in een webpagina.

Dashboard

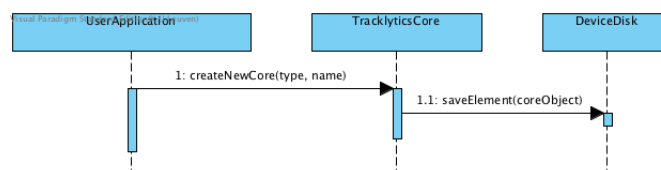
De Dashboard component dient om de verzamelde informatie door de Trashlytics library voor te kunnen stellen in een overzicht. Deze haalt de data op via de, in vorige sectie beschreven, Dashboard interface.

1.1.3 Belangrijkste flows

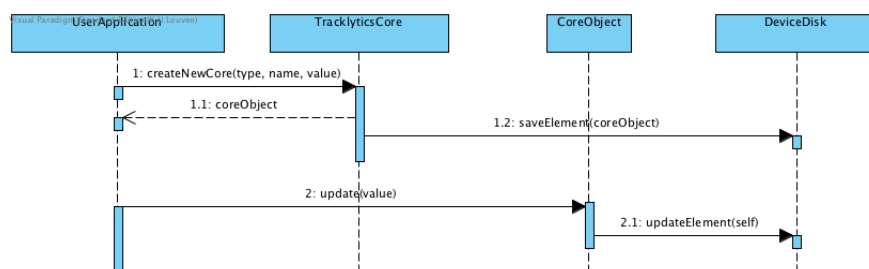
De belangrijkste flows door Tracklytics zijn:

- Tracking en monitoring van gegevens
- Verzenden en opslaan van de gegevens

Tracking en monitoring van gegevens



FIGUUR 1.3: Flow diagram 1.



FIGUUR 1.4: Flow diagram 2.

Het monitoren van de gegevens van de applicatie kan opgesplitst worden in twee verschillende flows.

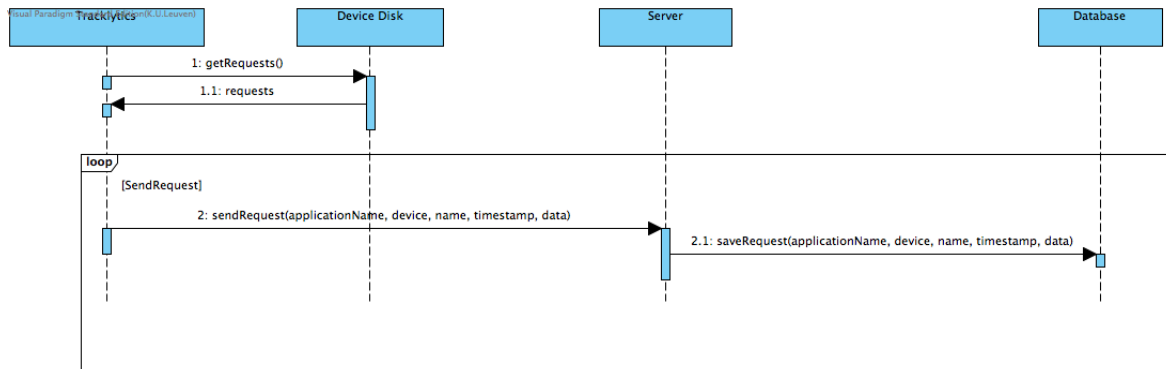
De eerste flow 1.3 toont de meest simpele flow. De developer geeft aan dat een bepaalde waarde gecollecteerd moet worden. De TracklyticsCore slaat deze informatie op op de schijf van het toestel van de gebruiker, zodat deze later verwerkt

kan worden. De verwerking van de data wordt uitgelegd in de volgende sectie 1.1.3. Deze situatie is geldig in de gevallen dat de developer data voor een histogram of een gauge wil collecteren. De uitleg over deze componenten staat in het volgende hoofdstuk 1.2.

De tweede flow 1.4 toont een uitgebreidere flow van de vorige flow. Er wordt weer een bepaalde waarde gecollecteerd en deze wordt opgeslagen op de harde schijf van het toestel. Bij deze flow wordt een Core Object terug gegeven waar verdere acties mee mogelijk zijn. Deze acties worden gebundeld in de **update** call. Deze flow is geldig als de developer data voor een Counter, een Timer of een Meter wil collecteren. De uitleg over deze componenten en de acties die op deze componenten mogelijk zijn staan in het volgende hoofdstuk 1.2

Verzenden en opslaan van de gegevens

In het volgende diagram 1.5 wordt getoond hoe het verzenden van de gegevens van het toestel naar de server in zijn werk gaat en hoe de gegevens worden opgeslagen in de database.



FIGUUR 1.5: Flow diagram 3.

1.2 Klassediagram

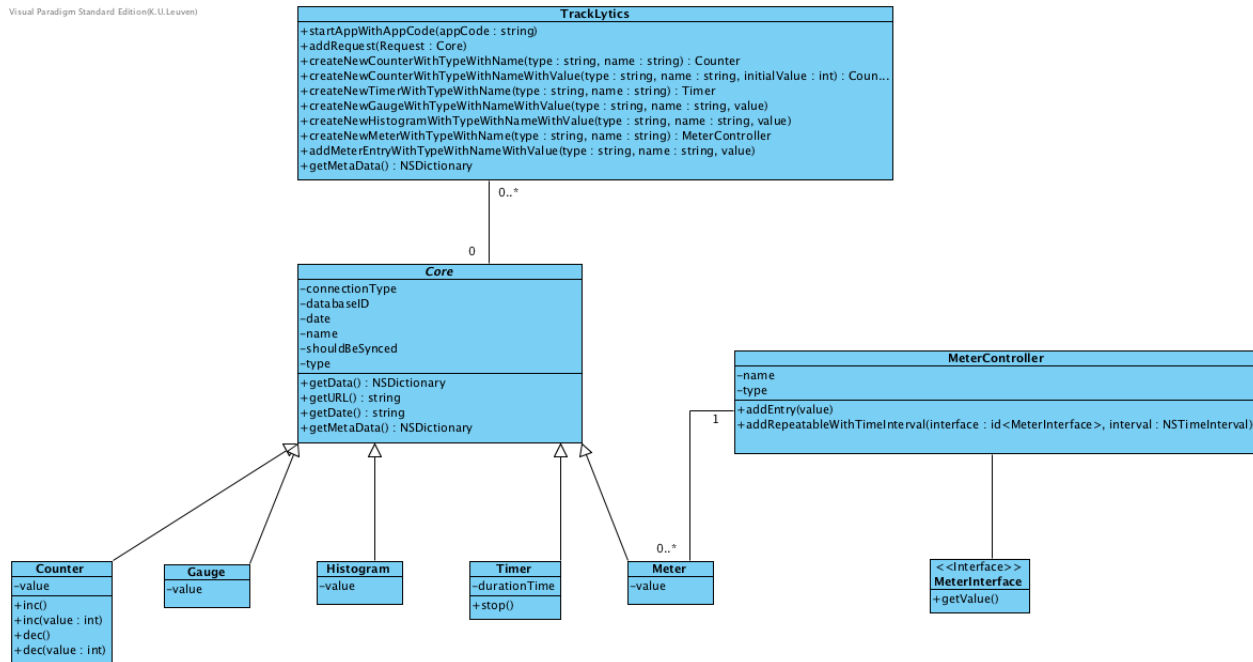
Het klassediagram duidt de klassen aan waaruit de Tracklytics libray bestaat. De verantwoordelijkheden van de verschillende klassen worden uitgelegd in de volgende secties.

1.2.1 TrackLytics

De Tracklytics klasse is de klasse die de synchronisatie met de server verzorgt. De klasse is ook verantwoordelijk voor het creëren en opslaan van de meetobjecten. De klasse heeft enkel statische methodes, dit zorgt ervoor dat de library gemakkelijker te

1. ARCHITECTUUR

Visual Paradigm Standard Edition (K.U. Leuven)



FIGUUR 1.6: Klassediagram Tracklytics library.

gebruiken is omdat er geen object van deze klasse aangemaakt moet worden telkens de library gebruikt zou worden.

1.2.2 Core

Om de gemeenschappelijke elementen te combineren is ervoor gekozen om die gemeenschappelijke elementen in een abstracte superklasse te steken. De gemeenschappelijke methodes worden ook in de Core klasse gestoken, zodat de subclasses indien nodig dit kunnen overriden.

1.2.3 Counter

De counter klasse stelt een telobject voor. Je kan bij een counter een getal optellen en aftrekken. De `inc()` functie verhoogt de value met 1 terwijl `inc(x)` de value verhoogt met `x`. Dit is opgesplitst omdat `inc()` vaker gebruikt gaat worden en zo moet er niet telkens `inc(1)` uitgevoerd worden. Hetzelfde geldt voor `dec` dat de value gaat verlagen. Deze functies zijn een onderdeel van de update call zoals besproken in vorige sectie [1.1.3](#).

1.2.4 Gauge

De gauge klasse stelt een object met een waarde voor. Dit object wordt aangemaakt en opgeslagen met die waarde en gesynchroniseerd naar de server. Er zijn geen

methodes beschikbaar speciaal voor dit object, omdat deze waarde niet aangepast zal worden.

1.2.5 Histogram

De gauge klasse en het histogram zijn maar op 1 punt verschillend en dat is in de back-end. De waarde van het histogram wordt ergens anders opgeslagen dan de waarde van de gauge. In de tracklytics library zijn ze dus hetzelfde, enkel de `getUrl()` methode geeft een andere url terug.

1.2.6 Timer

De timer klasse kan gebruikt worden om de duur van een gebeurtenis te meten. Bij het aanmaken van de timer wordt de huidige timestamp bijgehouden. De `stop()` methode neemt de huidige timestamp en vergelijkt die met degene die werd bijgehouden. Zo weten we de duur van de gebeurtenis. Indien de programmeur de methode `stop()` nooit aanroept wordt de timer niet gesynchroniseerd naar de server zodat er geen foute data tussen de data in de database staat. De `stop()` call is een onderdeel van de update call zoals besproken in vorige sectie [1.1.3](#).

1.2.7 Meter

Een meter is bedoeld om een reeks van waarden te kunnen meten. Om dit voor de developer makkelijk te maken hebben we hiervoor 3 componenten uitgedacht: de Meter, de MeterController en de MeterInterface. De meter component is een uitzondering als er gekeken wordt naar het flow diagram uit de vorige sectie [1.4](#). In plaats van dat een Meter object wordt terug gegeven, wordt er een MeterController object terug gegeven. Deze keuze is gemaakt, omdat het de structuur van de library versimpeld en het de developer gemakkelijker maakt.

Meter

Het meter object is het object dat opgeslagen wordt en dat gesynchroniseerd wordt naar de server. Het object houdt de value bij en ook het tijdstip van aanmaken.

MeterController

De metercontroller zorgt ervoor dat nieuwe meters aangemaakt kunnen worden. De metercontroller bevat de naam en het type van de meter. We kiezen ervoor om meters via deze weg aan te maken, omdat dit het werk van de programmeur vermindert. De programmeur moet enkel de waarde meegeven aan de metercontroller in plaats van telkens opnieuw de type en de naam mee te geven. We hebben het de programmeur nog op een manier gemakkelijk gemaakt. De metercontroller heeft de mogelijkheid om periodiek data op te halen van een functie die de programmeur specificeert. Zo moet de programmeur dit niet zelf gaan doen in een bepaalde methode.

MeterInterface

De meterinterface is de verbinding tussen de periodieke functie van de MeterController en het object waar de data op moet gehaald worden. Er is maar 1 functie, namelijk de `getValue()` die een waarde gaat ophalen. De programmeur vult deze methode dan in om de correcte waarde terug te geven.

Fiche masterproef

Student: Stef Van Gils

Titel: DevOps: Operational data for developers

Engelse titel: DevOps: Operational data for developers

UDC: 621.3

Korte inhoud:

Thesis voorgedragen tot het behalen van de graad van Master of Science in de ingenieurswetenschappen: computerwetenschappen, hoofdspecialisatie Software engineering

Promotor: Prof. W. Joosen

Assessor: Dimitri Van Landuyt

Begeleider: Dimitri Van Landuyt