

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 Context	1
1.1 Relevantie monitoren mobiele applicaties	1
1.2 Development Scenarios	4
2 Doelstelling	7
2.1 Library	7
2.2 Data Visualisatie	10
3 Architectuur	13
3.1 Component Diagram	13
3.2 Deployment Diagram	15
3.3 Uitbreidingen	16
3.4 Belangrijkste flows	18
4 Implementatie	21
4.1 Details Implementatie	21
4.2 Klassediagram	25
4.3 Documentatie	28
4.4 Aggregatie	31
4.5 Openstaande uitdagingen	33
5 Evaluatie	37
5.1 Performance	37
5.2 Schaalbaarheid	38
5.3 Developer effort	38
Bibliografie	39

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

3.1	Component diagram Tracklytics library.	14
3.2	Deployment diagram Tracklytics library.	16
3.3	Flow diagram 1.	19
3.4	Flow diagram 2.	19
3.5	Flow diagram 3.	20
4.1	Klassediagram Tracklytics library.	26

Lijst van tabellen

Hoofdstuk 1

Context

1.1 Relevantie monitoren mobiele applicaties

Op het moment van schrijven is het 2016; het jaar waarin ongeveer één op de vijf mensen een smartphone gebruikt in het dagelijkse leven [7]. Een smartphone bevat meerdere mobiele applicaties, ontwikkeld door developers wereldwijd. In de App Store van Apple staan ongeveer 1.5 miljoen verschillende applicaties. Een karakteristiek van de mobiele app wereld is dat er voor een bepaalde app vaak een of meerdere alternatieven bestaan. Dit wil zeggen dat de eigenaars van de app zich moeten proberen onderscheiden van de andere apps. Dat kan op meerdere manieren: door een praktisch design, door een snellere app te hebben, etc. Door tijd en/of geld te investeren in het verbeteren van de applicatie kan de applicatie de meest gebruikte applicatie worden en blijven in zijn categorie.

Met het design wordt de structuur van de gebruikersinterface van de applicatie bedoeld. Dit design bepaalt waar welke UI elementen komen te staan. Er wordt hier niet bedoeld op de esthetiek van de gebruikersinterface. Het design wordt meestal op voorhand vastgelegd alvorens de applicatie ontwikkeld wordt, zodat de developer dit kan gebruiken bij het implementeren. Er kan niet objectief gezegd worden wat een goed of een slecht design is door een ontwikkelaar omdat dit een persoonlijke mening is. Het design kan enkel écht beoordeeld worden door de gebruikers van de applicatie. Dit komt omdat er een verschil kan zijn in hoe de eigenaars van de applicatie denken dat de applicatie gebruikt wordt en in hoe de gebruikers de applicatie gebruiken. Als er een verschil in deze denkwijze zit, dan zou het design best aangepast worden naar de smaak van de gebruikers. Dit bevordert enerzijds de kwaliteit van de applicatie en anderzijds creëert het een band tussen gebruiker en developer. Indien er geen input komt van de gebruikers kan het nog steeds zijn dat de applicatie niet gebruikt wordt hoe de eigenaar het denkt. Er moet dan ontdekt worden welke elementen gebruikt worden en welke niet. Het monitoren van deze elementen kan ervoor zorgen dat er een goed beeld gevormd wordt voor de eigenaar die met deze informatie zijn inzicht in de applicatie kan veranderen. Dit zorgt ervoor dat de eigenaar betere beslissingen kan maken omtrent de toekomst en de verdere

ontwikkeling van de applicatie. Zelfs al zijn er reviews van de gebruikers, dan nog is het voor de eigenaar meestal onmogelijk om te ontdekken welke elementen in de applicatie gebruikt worden en welke niet. Het is dus belangrijk om de applicatie te monitoren, ookal is er veel input van de gebruikers.

De prestaties van een applicatie zijn belangrijk voor een eigenaar omdat deze een impact hebben op de gebruiksvriendelijkheid van de applicatie. De meeste problemen die te maken hebben met prestaties zijn op te lossen door voldoende testen te schrijven en hiermee bugs en trage code sequenties uit de code te halen, maar sommige prestatie problemen doen zich enkel voor bij een significant gebruikersaantal. Drie uit de top tien van meest vermelde klachten van mobiele applicaties hebben te maken met de prestaties van een applicatie [5], namelijk: *Resource-Heavy*, *Slow or lagging* en *Frequent Crashing*. Deze klachten worden vaak in reviews van gebruikers geuit. De eigenaars kunnen hieruit opmaken dat er een probleem is, maar developers weten niet zeker waar het fout loopt uit die reviews. Om uit te vinden waar de prestatie problemen zitten is het voor de developers handig om de applicatie te monitoren waar de bottlenecks kunnen ontstaan. Hierdoor kunnen ze de exacte oorzaak van het probleem vinden. Op deze manier kan een potentiële probleem al ontdekt worden voor deze in de reviews van de gebruikers opduikt. Dit zorgt ervoor dat de tevredenheid niet naar beneden gaat, omdat het probleem op voorhand al ontdekt wordt.

De reviews die door de gebruikers worden gelezen moeten met een korrel zout genomen worden; mensen hebben de intentie om te overdrijven. De eigenaars kunnen hiervoor gebruik maken van een methode, ontwikkeld door de Carnegie Mellon Universiteit [3], die een analyse uitvoert van deze reviews en de inconsistente reviews eruit filtert. Zo kan er een goed standpunt gevormd worden rond de kwaliteit van de applicatie. Maar zelfs met die methode zijn reviews nog steeds subjectief als het gaat over prestaties. De enige manier om objectief te redeneren hierover is door middel van metingen uit te voeren. Er zijn meerdere mechanismen om zo'n data te collecteren, met elk zijn voor en nadelen, namelijk: *Built-in OS support*, *een monitoring library*, *een dedicated test environment*, etc.

Built-in OS support is een mechanisme dat door de ontwikkelaars van het besturingssysteem ingebouwd moet worden. Dit mechanisme kan dan verschillende data meten en collecteren, zoals bv. CPU load, gebruikt RAM geheugen, netwerkgebruik, etc. Het voordeel aan zo'n systeem is dat dit low level kan ingebouwd worden in de code van het besturingssysteem. Dit wil zeggen dat deze data snel gecollecteerd kan worden. Deze data moet dan nog doorgegeven kunnen worden aan de developers zodat zij deze kunnen verwerken en beoordelen. Het nadeel hieraan is dat developers niet kunnen aangeven welke data ze gecollecteerd willen zien. Ze kunnen zich enkel baseren op de data die er voor hen gecollecteerd wordt door het besturingssysteem. De developer is afhankelijk van de ontwikkelaar van het besturingssysteem om deze data te kunnen collecteren, indien die ontwikkelaar enkel lokale data collectie aanbiedt is het niet mogelijk voor de developer van de applicatie

om deze applicatie te monitoren in productie. Met dit mechanisme is het mogelijk om problemen te ontdekken, maar is het moeilijker om deze te identificeren in de applicatie, omdat bv. het mechanisme kan aangeven dat er enorm veel RAM gebruikt wordt, maar men weet dan niet meteen waar de bug zich bevindt die het RAM geheugen vol heeft gestoken.

Een monitoring library is een high level softwarepakket dat ingebouwd moet worden in de applicatie en waarmee developers hun applicatie kunnen monitoren terwijl de applicatie al in productie genomen is. De developer kan meetpunten aangeven om data te kunnen collecteren. Deze data moet dan worden weergegeven aan de developer. Deze weergave verschilt per monitoring library. Het voordeel hieraan is dat developers enkel de data collecteren die ze nodig hebben van de applicatie en zo een compacter overzicht krijgen van deze data. Dit zorgt voor meer vrijheid voor de developers. Het nadeel is dat deze implementatie trager is dan dat dit in het besturingssysteem ingebouwd zou worden. Met een monitoring library kan er ook enkel high level data gecollecteerd wil worden, wat wil zeggen dat vele informatie niet beschikbaar is omdat de API van het besturingssysteem dit niet aanbiedt. Indien een monitoring library een bug ontdekt als de applicatie in productie is, dan kunnen de developers dit probleem sneller identificeren, omdat ze weten waar ze de meetpunten geplaatst hebben.

Een dedicated test environment is een omgeving waar er verschillende tests op de applicatie worden uitgevoerd. Via deze methode kan men, alvorens men de applicatie in productie brengt ervoor zorgen dat bugs en sommige performance issues uit de applicatie kunnen gehaald worden. Een dedicated test environment kan enkel in de development fase gebruikt worden en is eigenlijk een voorbereiding voor het in productie brengen van de applicatie. Nadat de applicatie in productie is kan de test environment enkel gebruikt worden om proberen problemen op te lossen die binnengekomen zijn als bug reports van klanten. Via dit test environment kunnen ze dan proberen om deze bug te identificeren, repliceren en op te lossen.

De developer zal dus een keuze moeten maken welke oplossingen hij gebruikt in zijn applicatie. Het beste zou zijn om deze oplossingen te combineren, omdat er dan voor gezorgd kan worden dat bijna elke bug eruit kan gehaald worden en dat developers bugs kunnen ontdekken (en waar ze zich situeren) indien de applicatie al in productie is. Dit zorgt ervoor dat er veel tijd kan bespaard worden in het zoeken naar bugs. Er moet wel een evenwicht gevonden tussen het ontdekken en oplossen van bugs en de moeite die een developer moet doen om ervoor te zorgen dat dit ontdekt kan worden. Dit staat geheel in het kader van DevOps, omdat men zo snel mogelijk nieuwe versies wil uitrollen en indien de developer dan veel moeite moet steken in het ontdekken van bugs in de applicatie alvorens en tijdens de applicatie in productie is genomen, dan kan de uitrol niet zo snel gebeuren als gehoopt.

Het monitoren van mobiele applicaties is anders dan het monitoren van websites, internet applicaties en pc applicaties omwille van een aantal factoren. Een mobiele

applicatie draait op een besturingssysteem dat speciaal ontworpen is voor de mobiele telefoons. Android en iOS zijn de twee besturingssystemen die het waard zijn om te vermelden [6]. Een mobiele applicatie wordt dus meestal voor beiden ontwikkeld om een zo groot mogelijk aantal gebruikers te bereiken. Ontwikkelen voor de andere besturingssystemen is meestal niet rendabel voor de moeite en het geld dat erin moet gestoken worden. De monitoring library draait dus maar op twee verschillende systemen, terwijl er tientallen web browsers bestaan waar websites en internet applicaties in draaien.

Indien de applicaties native worden ontwikkeld, dan moet de applicatie een keer voor Android en een keer voor iOS ontwikkeld worden. Dit zorgt ervoor dat er verschillende bugs en/of bottlenecks in de verschillende applicaties kunnen zitten. Het zou verkeerd zijn om gecollecteerde data van de twee applicaties samen te verwerken. Een desktop computer of laptop heeft in de meeste gevallen ofwel een WiFi verbinding of een ethernet verbinding met het internet. De internetverbinding van een smartphone varieert tussen WiFi en een mobiel netwerk (4G, 3G, Edge, ...). Deze connecties hebben een verschillende snelheid en moeten dus anders geïnterpreteerd worden.

Bij het monitoren van mobiele applicaties is het noodzakelijk dat een monitoring library zo weinig mogelijk resources (CPU, batterij, netwerk) gebruikt. Zoals eerder al vermeld is dit één van de meest voorkomende klachten bij gebruikers van een mobiele applicatie. Bij pc applicaties wordt hier bijna nooit rekening mee gehouden, omdat de prestaties van de resources hier significant hoger zijn dan bij smartphones. Bij internet applicaties is dit niet van toepassing, omdat dit heel erg af hangt van de implementatie van de browser.

Het monitoren van mobiele applicaties is op veel vlakken een handige feature. Enerzijds om als ondersteuning te dienen voor de eigenaars om de gebruikers van de applicatie tevreden te stellen en de applicatie te kunnen verbeteren/veranderen in functie van de gebruiker en hoe de applicatie gebruikt wordt. Anderzijds dient het monitoren van de applicatie als een hulpmiddel voor de developers om te kunnen ontdekken of er bugs in de applicatie zitten en ook de plaats van voorkomen te identificeren zodat deze opgelost kunnen geraken in een volgende versie. Er moet een keuze gemaakt worden tussen het gebruik van een bestaand monitoringsysteem of het zelf ontwikkelen van zo'n systeem met alle voordelen en nadelen in rekening gebracht.

1.2 Development Scenarios

In deze sectie worden er enkele scenario's uitgelegd die aantonen dat het monitoren van een mobiele applicatie belangrijk is. Deze sectie wordt opgedeeld in twee verschillende onderdelen die gemonitord kunnen worden, namelijk: *prestaties en*

gebruik.

1.2.1 Prestaties

Zoals eerder al vermeldt zijn de prestaties van een mobiele applicatie één van de belangrijkste klachten. Developers moeten er dus voor zorgen om de prestaties van de applicatie zo hoog mogelijk te houden en de performance bugs op te lossen.

Om performance problemen te ontdekken moeten deze ontdekt kunnen worden. Hier kan een monitoring systeem een oplossing bieden. Dit kan een monitoring systeem zijn dat geïmplementeerd is in het besturingssysteem (*Built-in OS support*), een third party monitoring library of via een dedicated test environment. De informatie die de developer kan terugkrijgen hangt af van welk systeem de developer gekozen heeft.

Er zijn een aantal zaken die de applicatie traag kunnen maken, namelijk: CPU-intensieve code, veel RAM geheugen, content die traag van het internet kan gehaald worden, ...

CPU-intensieve code gebruikt veel CPU tijd en kan de applicatie traag doen lijken. Dit kan ontdekt worden met een monitoringssysteem en afhankelijk van welke keuze van systeem kan de developer snel vinden waar deze code zich bevindt. De developer zou dan best deze code, indien mogelijk, herschrijven om de applicatie sneller te doen lopen.

Indien de applicatie **veel RAM geheugen** gebruikt, dan kan dit de applicatie traag doen lopen, omdat het RAM geheugen naar disk moet geswapt worden en deze acties de applicatie vertragen. De developer moet er dan voor zorgen dat de applicatie minder RAM geheugen gebruikt of de bug, waar veel RAM geheugen gebruikt wordt (bv een oneindige while loop die in de achtergrond loopt), oplossen.

Mobiele applicaties halen vaak data van het internet of versturen data naar het internet. Indien deze acties veel tijd kosten hebben deze een impact op de snelheid van de applicatie. Dit soort problemen kunnen bijvoorbeeld voorkomen indien het aantal gebruikers van de applicatie snel omhoog gaat en de servers de bijkomende trafiek niet aankunnen. Het is belangrijk voor de developer om te ontdekken dat deze problemen zich voordoen. Smartphones hebben een WiFi verbinding of een mobiele dataverbinding (4G, 3G, Edge,...). Deze karakteristiek moet mee in rekening genomen worden bij het beoordelen of er een probleem is en waar het probleem zich voordoet.

Buiten deze voorbeelden zijn er nog andere kenmerken die belangrijk zijn voor een mobiele applicatie, maar deze voorbeelden schetsen het beeld waarom het monitoren van mobiele applicaties belangrijk is.

1.2.2 Gebruik

Er bestaat een verschil in de gedachte van de ontwikkelaars van een applicatie over hoe de mobiele applicatie gebruikt wordt. Deze mismatch zorgt ervoor dat ontwikkelaars verkeerde beslissingen kunnen nemen in verband met de mobiele applicatie. Ze kunnen bijvoorbeeld een veelgebruikte feature uit de applicatie halen, omdat ze niet weten hoe de applicatie gebruikt wordt. Dit probleem komt voor in de lijst met meest voorkomende klachten over mobiele applicaties [5].

Om te weten te komen wat er gebruikt wordt in de applicatie kan er bijvoorbeeld geteld worden hoe vaak er op welke *button* gedrukt wordt in de applicatie. Zo kan men zien welke *buttons* er het meeste gebruikt worden. Men kan bijvoorbeeld ook gaan tellen hoe vaak een bepaald scherm bezocht wordt. Met deze informatie kan een ontwikkelaar ervoor kiezen om een bepaald scherm eruit te halen of weten dat ze een bepaald scherm zeker niet uit de applicatie mogen verwijderen.

Dit zijn enkele voorbeelden van hoe een ontwikkelaar een voordeel kan halen door een monitoringsysteem in zijn applicatie te verwerken. Er zijn natuurlijk nog andere scenario's waarin een monitoringsysteem belangrijk kan zijn in het ontwikkelen van mobiele applicaties en zo de klachten van gebruikers tot een minimum te houden.

Hoofdstuk 2

Doelstelling

Het onderzoek van deze thesis baseert zich op twee invalshoeken, enerzijds hoe de developers en eigenaars van een mobiele applicatie geholpen kunnen worden bij het ontwikkelen en verbeteren van die applicatie en anderzijds wat de trade-off is tussen enerzijds hoeveel de applicatie gemonitord wordt en anderzijds de impact op de performance van de applicatie, de developer effort, etc.

De eerste invalshoek gaat over hoe we developers kunnen helpen bij het ontwikkelen van een mobiele applicatie door de developers problemen te laten identificeren (zoals trage code sequenties) en de eigenaars te laten ontdekken welke componenten van de applicatie het meest gebruikt worden en welke het minste. Met deze informatie weten de developers waar ze verbeteringen kunnen aanbrengen en weten eigenaren in welke functies ze wel of niet moeten investeren.

Langs de andere kant is het belangrijk om te weten wat de impact is van het verkrijgen van de informatie op de prestaties van de applicatie en hoeveel moeite en tijd er nodig is van de developer om deze informatie te kunnen vergaren.

Om dit onderzoek te doen slagen wordt er een library gebouwd die developers in hun applicaties kunnen inbouwen om hun applicatie te monitoren. In de volgende sectie wordt uitgelegd wat het doel is van deze library. In het verdere verloop van deze thesis wordt uitgelegd hoe de architectuur achter de gebouwde library in elkaar zit en waarom de keuzes zijn gemaakt voor die structuur. Nadat deze architectuur beschreven is wordt er een uitleg gegeven over de implementatie van de library. Ten slotte wordt de library geëvalueerd om te kijken of de kwaliteit van de library voldoende is om gebruikt te worden in de applicaties.

2.1 Library

Dit onderzoek baseert zich op developers die mobiele applicaties ontwikkelen en eigenaars van die applicaties die meer informatie over die applicaties willen verzamelen.

Het doel van de library is om developers deze library te laten inbouwen in hun applicatie en met behulp van de methodes aangeboden door deze library de applicatie te kunnen monitoren. Deze methodes moeten voldoende zijn om alle aspecten van de applicatie te kunnen monitoren. De library biedt een aantal methodes aan om de developer de mogelijkheid te geven om meetpunten in zijn applicatie in te bouwen. Deze methodes staan beschreven in het hoofdstuk over de implementatie 4.3. De library is open source gemaakt om ervoor te zorgen dat developers een beeld kunnen krijgen van welke informatie er verzameld wordt en hoe deze gebruikt wordt. Met deze informatie kunnen de developers er zeker van zijn dat privacygevoelige informatie niet verzonden wordt naar de servers. Het feit dat de library open source is wil ook zeggen dat andere developers uitbreidingen of verbeteringen kunnen aanbrengen.

Om een kwaliteitsvolle library te bouwen zijn er een aantal aspecten die belangrijk zijn om in het achterhoofd te houden:

- performance impact
- schaalbaarheid
- bruikbaarheid
- beschikbaarheid

Deze aspecten worden besproken in het komende deel van dit hoofdstuk.

2.1.1 Performance impact

Het is van cruciaal belang dat een applicatie niet traag is en responsief is naar de gebruiker toe. De verwerkingskracht (CPU) van een mobiel apparaat is relatief beperkt. Deze twee gegevens gecombineerd zorgt ervoor dat de library die gebouwd wordt zo beperkt mogelijk impact mag hebben op deze verwerkingskracht. Indien deze library relatief veel CPU tijd inneemt, dan blijft er minder CPU tijd over voor de applicatie, wat ervoor zorgt dat de applicatie trager wordt indien deze CPU-intensief is. Het is dus van cruciaal belang dat de library geoptimaliseerd wordt om zo weinig mogelijk CPU tijd te verbruiken. Een bijkomend voordeel dat verschijnt indien de library zo minimaal mogelijk CPU tijd nodig heeft is dat de impact van de applicatie op het batterijverbruik verminderd wordt.

Naast het CPU verbruik is het ook van belang dat het netwerkgebruik zo efficiënt mogelijk gebruikt wordt. In een mobiel apparaat wordt de netwerkinterface in slaapstand gehouden tot deze gebruikt moet worden en op dat moment wordt deze interface geactiveerd. Dit wordt gedaan omdat de netwerkinterface veel energie verbruikt indien deze aanstaat. Indien de netwerkinterface zo efficiënt mogelijk gebruikt wordt kan deze vaker in slaapstand gezet worden en is de impact van de library op het batterijgebruik minder dan indien deze inefficiënt gebruikt wordt. Deze efficiëntie zorgt er ook voor dat de applicatie de netwerkinterface zo vaak mogelijk kan gebruiken en hier zo weinig mogelijk vertraging opgelopen wordt als de applicatie

het netwerk wil gebruiken.

Niet enkel de efficiëntie van het gebruik van de netwerkinterface is van belang, maar ook de snelheid waarmee de back end reageert op de request en deze een resultaat kan terug geven. Indien het relatief lang duurt eer de back end het resultaat terug geeft, is de optimalisatie van het gebruik van de netwerkinterface voor niets geweest. De snelheid van de back end is dus ook een belangrijk aspect.

De impact op de performance van de applicatie is het belangrijkste aspect in het bouwen van de library. Indien deze impact relatief groot is, dan is het voor developers onmogelijk om deze library in de applicatie in te bouwen en nog steeds een degelijke responsiviteit en snelheid van deze applicatie te behalen.

2.1.2 Schaalbaarheid

Zoals hiervoor gezegd is de snelheid van de back end een belangrijk aspect inzake performance van de netwerkinterface. Één aspect van deze snelheid is de implementatie van deze back end. De implementatie moet geoptimaliseerd worden zodat deze zo snel mogelijk een resultaat kan teruggeven aan de library. Het andere aspect in de snelheid van de back end is de load die de servers aankunnen om al de requests te verwerken. De mate van schaalbaarheid is hier een belangrijke factor in.

Een applicatie wordt simultaan gebruikt door gemiddeld N gebruikers waarin de library aanwezig is. Deze applicatie verstuurt per keer gemiddeld X metingen. De library wordt gebruikt door M verschillende applicaties. De applicaties zijn geconfigureerd met een gemiddeld synchronisatie-interval T . Dit geeft dat het gemiddeld aantal requests per tijdseenheid neerkomt op: $RpT = \frac{M * N * X}{T}$. Hoe hoger dit getal, hoe hoger het aantal requests de servers moeten beantwoorden en hoe meer load er per server komt. Elke server S heeft een limiet kwa aantal requests die simultaan kunnen beantwoord worden R . De volgende vergelijking geeft de relatie aan tussen het gemiddeld aantal requests en het aantal requests dat beantwoord kan worden: $S * R > RpT$. Dit wil zeggen dat er (altijd) meer requests moeten kunnen beantwoord worden dan dat er gemiddeld per tijdseenheid aan komen.

2.1.3 Bruikbaarheid

Het primaire doel van een library is om de taken uit te voeren die worden gegeven aan de library. Naast dit primaire doel is het ook de bedoeling dat developers de library kunnen gebruiken zonder veel moeite. De bruikbaarheid van de library moet zo hoog mogelijk zijn zodat developers zonder veel tijd en moeite deze kunnen integreren in een applicatie. De volgende punten zijn hierin belangrijk:

- Installatie
- Configuratie

- Documentatie

Installatie Developers zouden de library out-of-the-box moeten kunnen gebruiken. De installatie van de library in de applicatie zou niet lang mogen duren en niet complex mogen zijn.

Configuratie De configuratie van de library moet straightforward zijn. Enkel triviale parameters, zoals de identificatie van de applicatie en het synchronisatie-interval, zijn belangrijk om te configureren.

Documentatie De library moet zo gebruiksvriendelijk mogelijk zijn. Een goede documentatie is een startpunt om developers het gemakkelijk te maken om de library te gebruiken. Hier moeten de belangrijkste methodes uitgelegd worden en ervoor gezorgd worden dat het duidelijk is hoe de library in de applicatie ingebouwd moet worden.

2.1.4 Beschikbaarheid

Met beschikbaarheid wordt bedoeld dat alle gecollecteerde data opgeslagen wordt op de servers. De data die gecollecteerd wordt is waardevol. Het is de bedoeling dat er zo'n goed mogelijk beeld geschetst wordt voor de developer. Indien een deel van de data niet opgeslagen wordt in de database kan het zijn dat mogelijke uitzonderlijke data niet opgeslagen wordt, terwijl deze uitzonderlijke data juist de belangrijkste is. Er moet dus voor gezorgd worden dat 100% van de gecollecteerde data gesynchroniseerd wordt naar de server om een zo goed mogelijk beeld te vormen voor de developers en de eigenaars van de applicatie.

Deze aspecten moeten in rekening genomen worden bij het ontwerpen en ontwikkelen van een mobiele monitoring library. De belangrijkste aspecten zijn de performance aspecten, omdat in mobiele applicaties de verwerkingskracht en batterij de bottlenecks zijn.

2.2 Data Visualisatie

De library beschreven in de vorige sectie collecteert de data van de applicaties op de mobiele toestellen en verstuurt deze naar de server die deze data dan opslaat. Om deze data overzichtelijk voor de developers en de eigenaars te maken moet deze data gevisualiseerd worden. De data van de library wordt gevisualiseerd in een dashboard dat per applicatie de data weergeeft. De implementatie en de verschillende voorstellingen die de library gebruikt wordt uitgelegd in de implementatie sectie [4.1.3](#). Deze implementatie is gekozen voor de library omdat deze een handig en algemeen over-

zicht geeft. Andere implementaties geven mogelijk een beter resultaat in bepaalde situaties.

Hoofdstuk 3

Architectuur

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.

Zoals getoond op de volgende figuur 3.1 bestaat de Tracklytics library uit vier verschillende componenten:

- **UserApplication**: De applicatie die de developer wil monitoren
- **TracklyticsCore**: Stelt de library voor waarmee de UserApplication communiceert om te kunnen monitoren.
- **TracklyticsBackend**: De backend die de informatie verwerkt die van de TracklyticsCore komt.
- **Dashboard**: Een dashboard om alle informatie weer te geven.

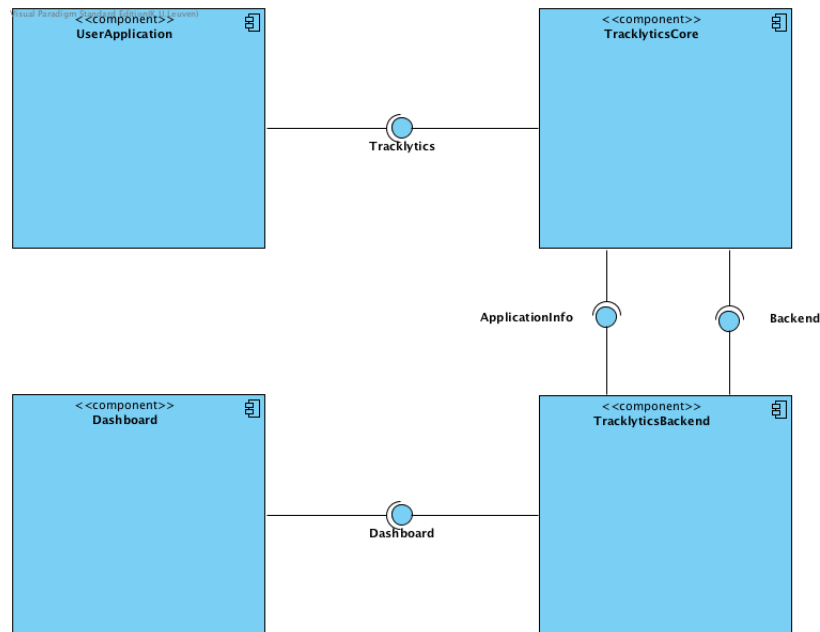
Deze componenten vormen samen de core van de Tracklytics library. Deze communiceren onderling via de aangeboden interfaces. In de rest van dit hoofdstuk worden deze componenten beschreven en worden de architecturale beslissingen uitgelegd met een motivatie waarom die beslissingen gemaakt zijn.

3.1 Component Diagram

Het component diagram 3.1 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.

3.1.1 UserApplication

De UserApplication component representeert een applicatie ontwikkeld door developers. De UserApplication gebruikt de Tracklytics interface om metingen uit te



FIGUUR 3.1: Component diagram Tracklytics library.

voeren en de applicatie te monitoren. Developers moeten zelf aangeven welke elementen er gemonitord moeten worden.

3.1.2 TracklyticsCore

De TracklyticsCore component stelt een monitoring library voor die in deze thesis ontwikkeld wordt. De TracklyticsCore component verzamelt informatie die hij verkrijgt van de UserApplication component, slaat deze tijdelijk op het toestel op om deze dan later, na een bepaald tijdsinterval, door te sturen naar de TracklyticsBackend component via de Backend interface.

Interface De TracklyticsCore component biedt een Tracklytics interface aan. De UserApplication kan deze interface gebruiken om de applicatie te monitoren. De TracklyticsCore stuurt deze data door naar de server.

De interface bestaat uit methodes om de applicatie te monitoren. Tracklytics biedt vijf elementen aan om de applicatie te kunnen monitoren, namelijk:

- Counter: Kan gebruikt worden om evenementen te tellen
- Meter: Kan gebruikt worden om periodiek waarden op te halen.
- Histogram: Kan gebruikt worden om een waarde te verzamelen dat in het dashboard in een histogram gegoten kan worden.

- Timer: Kan gebruikt worden om de lengte van een evenement te meten.
- Gauge: Kan gebruikt worden om een bepaalde waarde te meten.

Deze elementen vormen de kern van de Tracklytics library en worden verder besproken in het volgende hoofdstuk om duplicatie van tekst te vermijden [4.2](#).

3.1.3 TracklyticsBackend

De TracklyticsBackend component representeert de server kant van de library. Hierop draait de code om de gemonitorde data op te slaan in de database. De TracklyticsBackend component bevat dus de database om alle informatie in op te slaan. De TracklyticsBackend component biedt drie interfaces aan, namelijk: ApplicationInfo, Backend en Dashboard. Deze worden in de volgende paragrafen besproken.

Interfaces

ApplicationInfo Deze interface wordt gebruikt door de TracklyticsCore om instellingen van de developer op te halen die gemaakt zijn in het dashboard, zoals bv. bij het On Demand aan- of uitzetten van het monitoren zoals iets verder besproken [3.3](#).

Backend Deze interface wordt gebruikt om data uit te wisselen tussen de TracklyticsCore en de TracklyticsBackend. Dit is de gemonitorde data die de TracklyticsCore verzameld heeft. De TracklyticsBackend verwerkt deze informatie en slaat deze op in de database.

Dashboard Deze interface wordt gebruikt door het dashboard om de data uit de database op te halen om deze weer te kunnen geven aan de developers.

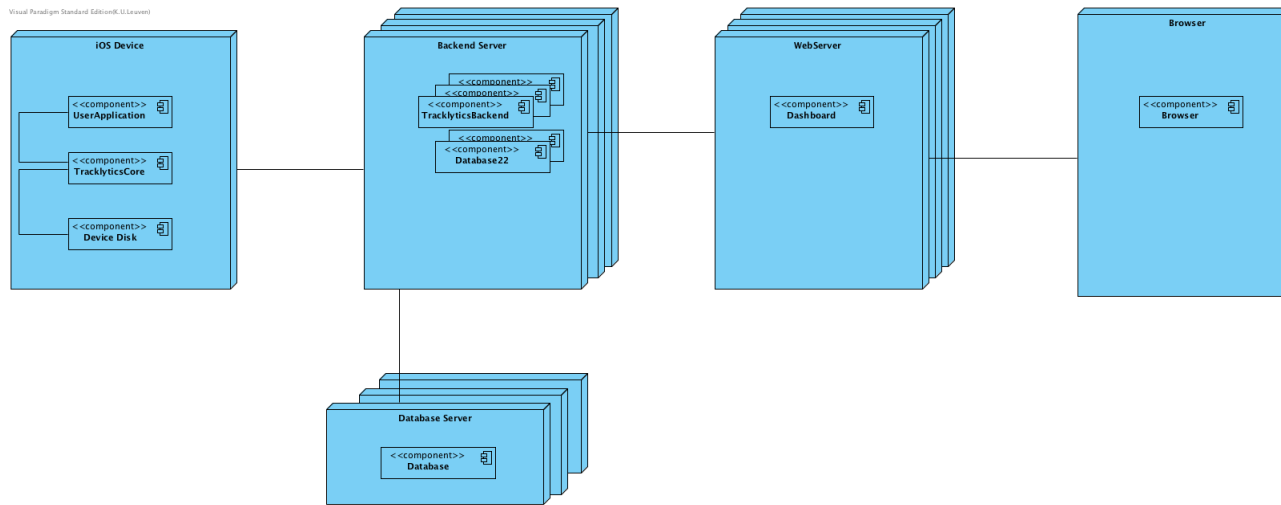
3.1.4 Dashboard

De Dashboard component representeert een applicatie om de gemonitorde data van de Tracklytics library weer te kunnen geven aan de developers en de eigenaars. Met behulp van grafieken en verwerkte data kunnen de developers en de eigenaars beslissingen nemen omtrent de applicatie.

3.2 Deployment Diagram

Het deployment diagram toont op welke nodes de componenten besproken in vorige sectie draaien. Zoals je kan zien draait de UserApplication en de TracklyticsCore op

3. ARCHITECTUUR



FIGUUR 3.2: Deployment diagram Tracklytics library.

het mobiele toestel van de gebruiker zelf. 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.

3.3 Uitbreidingen

In deze sectie worden een aantal uitbreidingen besproken die niet standaard in een monitoring library zitten, maar zeer handig kunnen zijn om developers te helpen bij het monitoren van de applicatie.

Data availability Smartphones vallen weleens zonder internet connectiviteit door het wegvallen van een WiFi verbinding of het wegvallen van de mobiele verbinding. De Tracklytics library zendt de gemonitorde 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 mobiele 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. Zo wordt ervoor gezorgd dat de data in ieder geval in de database terecht komt.

On Demand Het is de bedoeling de developer zoveel mogelijk vrijheid te geven in het monitoren van de applicatie. Om deze vrijheid te verhogen is het On Demand aan- of uitzetten van het monitoren aan Tracklytics toegevoegd. Een developer kan vanuit het dashboard aangeven of de applicatie op dit moment 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 hiermee al een deel van de verwerking gedaan wordt en de backend bij het opvragen van de data vanuit het dashboard minder werk heeft om deze samen te voegen en verwerken. Er is dus een prestatiewinst langs de serverkant indien de aggregatie op de smartphone uitgevoerd wordt. Een nadeel is dat deze aggregatie meer CPU tijd van de smartphone gebruikt dan als we de aggregatie in de backend uitvoeren en dit kan ervoor zorgen dat de applicatie traag gaat overkomen bij de gebruikers. 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 versie B geleidelijk verloopt. De gebruikers van de software worden dus in twee (of meerdere) groepen opgedeeld door een bepaalde 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 tot het probleem in versie B opgelost is.

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 met nieuwe features en code uit te rollen. Dit staat pal tegenover websites die hun webpaginas en code rechtstreeks van een server halen en waar de aanpassingen aan de versies meteen zichtbaar zijn bij de gebruikers. Een workaround van dit probleem is dat de mobiele applicatie de code van de nieuwe versie (B) al

kan bevatten en ook nog de code van de oude versie (A) 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 aspect is de privacy van de gebruiker. Zoals eerder al aangehaald is NewRelic een closed source library, wat wil zeggen dat developers niet weten welke informatie er doorgestuurd wordt naar de backend. Dit gegeven zorgt ervoor dat privacygevoelige applicaties deze library niet zouden mogen gebruiken, omdat er gebruikersinformatie gelekt zou kunnen worden naar de backend van NewRelic zoals bijvoorbeeld email-adressen, creditcard gegevens, etc. 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. Tracklytics verzamelt enkel metadata van de telefoon en de applicatie. De metadata die gecollecteerd wordt door Tracklytics bestaat uit: het type toestel, de versie, de naam van de applicatie, een identifier van de applicatie, een identifier van het toestel om deze van elkaar te kunnen onderscheiden, de datum en het type connectie waarop het toestel zich bevindt. Dit gegeven zorgt ervoor dat Tracklytics de privacy van de gebruiker garandeert..

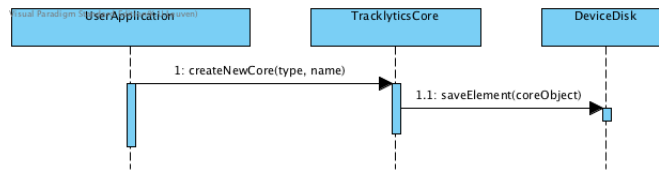
Naast privacy is security ook een belangrijk aspect. Er bestaat namelijk een verbinding tussen de mobiele applicatie en de backend over een onveilig netwerk. Het zou dus niet mogen dat de doorgezonden data door een onrechtmatig iemand gecollecteerd wordt of dat hiermee geknoeid wordt. Om deze situaties 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 de backend. Deze combinatie zorgt ervoor dat de gegevens niet onderschept kunnen worden door onrechtmatige personen.

3.4 Belangrijkste flows

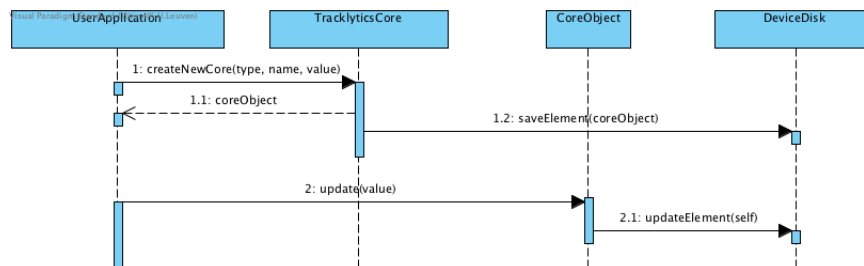
In deze sectie worden de belangrijkste flows in de Tracklytics library beschreven. Deze geven aan hoe de data flow binnen Tracklytics werkt. De belangrijkste flows door Tracklytics zijn:

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

3.4.1 Tracking en monitoring van gegevens



FIGUUR 3.3: Flow diagram 1.



FIGUUR 3.4: Flow diagram 2.

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

De eerste flow [3.3](#) 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 [3.4.2](#). 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 [4.2](#).

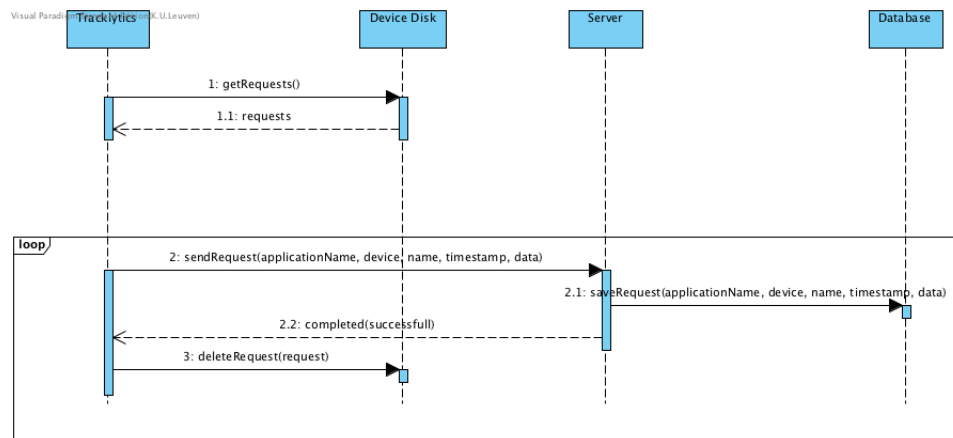
De tweede flow [3.4](#) toont een uitgebreidere flow van de vorige flow. Er wordt weer een bepaalde waarde gecollecteerd en deze wordt opgeslagen op de harde schijf van het toestel. Bij deze flow wordt een Core Object terug gegeven waar verdere acties mee mogelijk zijn. Deze acties worden gebundeld in 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 [4.2](#)

3.4.2 Verzenden en opslaan van de gegevens

In het volgende diagram [3.5](#) wordt getoond hoe het verzenden van de gegevens van het toestel naar de server in zijn werk gaat en hoe de gegevens worden opgeslagen

3. ARCHITECTUUR

in de database.



FIGUUR 3.5: Flow diagram 3.

De Tracklytics library haalt eerst alle meetgegevens van de harde schijf van het mobiele toestel. De Tracklytics library synchroniseert hierna alle meetgegevens naar de Tracklytics backend. De server waar de request terecht komt verwerkt deze gegevens en slaat deze op in de database. Nadat de server antwoord heeft gegeven aan de Tracklytics library dat de request succesvol behandelt is, verwijdert de library dit meetgegeven van de harde schijf. Indien de server antwoordt dat de operatie mislukt is, dan mag dit meetgegeven niet verwijderd worden zoals eerder besproken.

Hoofdstuk 4

Implementatie

In het vorige hoofdstuk is er beschreven hoe de architectuur van een monitoring library eruit kan zien. In dit hoofdstuk wordt er gekeken naar de implementatie van de Tracklytics library [4]. Er wordt gekeken naar welke technologieën er gebruikt zijn bij het ontwikkelen van de library. Daarnaast wordt er uitgelegd hoe developers de library kunnen gebruiken in hun applicaties. Er wordt dieper ingegaan op de communicatie tussen de Tracklytics library en de server.

4.1 Details Implementatie

In deze sectie wordt er besproken welke technologieën er gebruikt zijn bij het ontwikkelen van de Tracklytics library.

4.1.1 iOS Library

De Tracklytics library beschreven in deze thesis is ontwikkeld voor het iOS besturingssysteem. Voor het iOS besturingssysteem bestaan er twee programmeertalen om een applicatie, of in dit geval een library, te ontwikkelen, namelijk: Swift en Objective-C. Swift is een redelijk recente taal, op het moment van schrijven is deze nog geen twee jaar oud. Omdat de ondersteuning van Swift door de meerderheid van de developers nog zeer karig is, is er besloten om de library in Objective-C te schrijven. Objective-C en Swift hebben dezelfde functionaliteiten, maar programmeren in Swift is overzichtelijker door de andere syntax. Swift is meer geïntegreerd in de IDE die Apple aanbiedt om applicaties mee te ontwikkelen. Swift is veiliger en sneller dan Objective-C. Al deze redenen pleiten om Swift als programmeertaal te kiezen. Het grote nadeel is dat Swift nog niet in veel applicaties aanwezig is en dat het moeilijker is om in een Swift applicatie een Objective-C library te integreren dan andersom. Dit is de grote reden waarom er voor Objective-C gekozen is.

De Tracklytics library slaat tijdelijk de data op op de harde schijf van het toestel. Dit heeft twee voordelen, namelijk: er gaat geen data verloren en de library gebruikt minder RAM. Als het toestel geen internet connectie zou hebben, kan de

library de data niet verzenden naar de back end en indien de applicatie dan sluit zijn alle gegevens verloren. Dit scenario kan opgelost worden door de data tijdelijk op de harde schijf op te slaan. De data zal dan pas verwijderd worden als de library er zeker van is dat de data succesvol naar de back end is verstuurd. Er is voor gekozen om de developer hier de keuze aan te laten. Hij kan via het dashboard aangeven of de gegevens tijdelijk op harde schijf opgeslagen moeten worden.

Een mobiel toestel heeft een beperkte grootte RAM geheugen en applicaties kunnen hier niet volledig gebruik van maken omdat het besturingssysteem dit geheugen ook gebruikt. Als de applicatie teveel RAM geheugen gebruikt, dan gaat heel de applicatie deze trager werken, omdat het RAM geheugen dan uitgeswapt wordt naar de harde schijf. Het is dus noodzakelijk om het RAM gebruik zo laag mogelijk te houden. Indien alle data in RAM geheugen zou gehouden gestoken worden, dan kan dit snel vollopen. Het opslaan van de data op de harde schijf verhelpt dit probleem. Het is dan wel noodzakelijk dat het opslaan van de data op schijf in de achtergrond gebeurt, omdat I/O operaties relatief lang duren.

De methodes van de Tracklytics library die gebruikt kunnen worden zijn statische methodes. De keuze hiervoor is gebaseerd op twee redenen.

De Tracklytics library klasse die de methodes aanbiedt is geen objectgerichte klasse. Indien er een object van deze klasse zou gemaakt worden, zou deze elke keer na creatie bijna onmiddellijk niet meer gebruikt worden. Deze creatie van het object zorgt voor een overhead, de welke is weggewerkt door het statisch maken van alle methodes die bruikbaar zijn door de developers.

Een tweede reden is de gebruiksvriendelijkheid. Een library call neemt maar één lijn code in om een methode uit te voeren in plaats van twee. Dit zorgt voor minder code. De impact hiervan hangt af van het aantal monitoring punten die in de applicatie zijn ingevoerd. Dit verbetert ook de leesbaarheid van de code.

4.1.2 Back end

De data die de Tracklytics library doorstuurt vanaf het toestel van de gebruiker moet opgeslagen worden in een database. Zo kan deze data later verwerkt worden en worden weergegeven in het dashboard. Er moet een keuze gemaakt worden over het besturingssysteem, de database en de programmeertaal van de back end.

De server draait in OpenStack, een cloud platform. Deze is gedeployed als infrastructure-as-a-service (IaaS). Dit wil zeggen dat dit meerdere virtuele servers kan aanbieden (zelfs meerdere virtuele servers als fysieke servers). Dit biedt een abstractie en schermt de virtuele server af van andere virtuele servers. Een gebruiker kan zelf nieuwe virtuele servers aanmaken. Elke virtuele server heeft zijn eigen besturingssysteem, te kiezen uit een lijst van images aangeboden door OpenStack.

In de Tracklytics architectuur is gekozen voor een linux distributie (in dit geval Ubuntu). Deze keuze is gemaakt op basis van de gebruiksvriendelijkheid van dit

besturingssysteem. Zo is het gemakkelijk om snel een webserver op te zetten en een database te installeren. Er is voor Ubuntu gekozen, omdat dit de meest gekende en meest gebruikte linux distributie is.

In de database wordt alle data opgeslagen, wat wil zeggen dat dit een van de meest belangrijke onderdelen van de Tracklytics architectuur vormt. De metadata die gecollecteerd wordt per applicatie is in vele gevallen hetzelfde. De parameters die verschillen zijn: het type toestel, het type internet connectie en de versie van de applicatie. Dit gegeven zorgt ervoor dat de metadata vaak hetzelfde is. Indien de metadata uit de data komende van de Tracklytics library wordt uitgehaald kan er relatief veel opslagruimte gespaard worden, omdat deze metadata niet in elke entry in de database aanwezig moet zijn, enkel een verwijzing naar waar die metadata staat. Er is gekozen om een MySQL database te gebruiken, omdat deze een gestructureerde tabellenstructuur heeft en zo de metadata gemakkelijk van de data kan scheiden. De data kunnen door SQL queries gecombineerd worden in views om de data overzichtelijk te maken. Een alternatief is een NoSQL database. Dit alternatief past niet in de manier waarop de data opgeslagen wordt, omdat deze een niet-gestructureerde database aanbiedt.

De back end heeft in de Tracklytics library twee functies, namelijk: het verwerken en opslaan van de data in de database en het opvragen van data uit de database om het dashboard van de data te voorzien. De keuze is gevallen op PHP als programmeertaal. Met PHP is het simpel om een database connectie op te zetten en via SQL queries deze dat in of uit de database te krijgen. Een tweede voordeel van PHP is dat deze met POST data om kan. Via deze manier kan de data in de request van de Tracklytics library verborgen worden in de request en moet deze niet rechtstreeks doorgegeven worden in bijvoorbeeld de URL zelf. Zo is er meer veiligheid en privacy van de data.

4.1.3 Dashboard

Het Tracklytics dashboard is ontworpen om de gegevens van de applicatie in grafieken en details weer te geven om hieruit een conclusie te kunnen trekken over het functioneren van de applicatie. Het dashboard is ontworpen als webapplicatie in plaats van een desktop applicatie. De voordelen hiervan zijn: software updates automatisch worden doorgevoerd zonder dat er tussenkomst van de gebruiker nodig is, het is cross platform, omdat het in de browser draait en er is geen installatie vereist.

Om de webapplicatie te ontwikkelen is er gebruik gemaakt van AngularJS. De reden hiervoor is dat dit een zeer handig framework is voor het ontwikkelen van dynamische websites. Dit framework bindt stukken HTML code aan JavaScript code, wat ervoor zorgt dat het DOM gemakkelijk manipuleerbaar is door JavaScript. Een tweede reden waarom AngularJS in deze situatie voordelig is, is dat met AngularJS het gemakkelijk is om stukken HTML te laten herhalen met andere gegevens in. Zo

kunnen de verschillende grafieken onder elkaar geplaatst worden zonder telkens de HTML code in JavaScript aan te moeten passen. Een bijkomstig voordeel is dat de data in de verschillende tabbladen maar eenmalig ingeladen moet worden, omdat AngularJS ervoor zorgt dat als je op een tab drukt niet de hele webpagina opnieuw wordt ingeladen, maar enkel de view die verandert ingeladen wordt.

Om de grafieken weer te kunnen geven is ervoor gekozen om Chart.js (voor AngularJS) te gebruiken. Dit framework neemt data die in AngularJS variabelen gezet worden en geeft deze weer in een gekozen grafiek (bv. lijn-grafiek of staafdiagram). Chart.js is een simpele manier om snel een grafiek weer te kunnen geven, het werkt out-of-the-box en er zijn een aantal zeer handige opties die kunnen aangepast worden.

Om de histogrammen en de meters bruikbaar te maken zijn hier sliders bij toegevoegd om bij de histogrammen het interval te veranderen en zo een kleinere dataset te gebruiken. Bij de meters kan de dataset verkleind worden door het tijdsinterval te wijzigen. Omdat het dashboard gebouwd is in AngularJS en de histogrammen en meters herhaald worden over de pagina moest er een oplossing gevonden worden die compatibel was met AngularJS. Een andere moeilijkheid was dat de slider langs twee kanten zou moeten kunnen sliden om een minimum en maximum te kunnen definiëren. Door deze twee praktische zaken is er een keuze gemaakt voor de AngularJS-slider. Deze biedt alle functionaliteit aan die nodig is om de histogrammen en meters interactief te maken.

De grafieken worden per soort meetobject weergegeven, er bestaat dus een aparte pagina voor de counters, de gauges, de timers, de meters en de histogrammen. Zo heeft een developer een overzicht per type meting. De grafieken die weergegeven worden zijn basisgrafieken, deze zijn enkel opgedeeld door de namen die aan de metingen gegeven zijn. De developer kan hier enkel basisinformatie uithalen. Om een dieper detailniveau te verkrijgen is ervoor gekozen om per grafiek een gedetailleerdere pagina weer te geven. De grafiek wordt dan opgesplitst door verschillende eigenschappen te gebruiken. De volgende eigenschappen worden gebruikt om de grafieken op te delen: het type toestel, de versie van de applicatie en het type connectie. Deze worden onderling nog gecombineerd om nog extra in detail de gemeten data te kunnen bekijken. Aan de hand hiervan kunnen developers een concreet overzicht krijgen over de prestaties en het gebruik van de applicatie.

In het dashboard is ruimte voorzien om de instellingen van de applicatie te wijzigen. De developer kan kiezen of de applicatie gemonitord moet worden, of de aggregatie op het toestel moet gebeuren of in de backend en hij kan kiezen of de data tijdelijk op de harde schijf opgeslagen moet worden of niet.

De data die nodig is om de titels en de gegevens voor de grafieken in te laden moet uit de backend gehaald worden. Deze is zoals in vorige sectie aangehaald geschreven in PHP.

Er is enorm veel werk gekropen in het ontwikkelen van dit dashboard. Dit omdat

het dashboard from scratch ontwikkeld is. De reden hiervoor is dat de bestaande dashboards die vrij te gebruiken zijn geen oplossing gaven voor de visualisatie die ik wou geven. Het ontwikkelen van het dashboard (inclusief het opzetten van alle achterliggende systemen zoals de database en dergelijke) in combinatie met het ontwikkelen en het tweaken van de library heeft meer dan 200 uur tijd in beslag genomen.

4.1.4 Connectie tussen Front end en Back end

Zoals eerder aangegeven stuurt de front end (de Tracklytics iOS library) data naar de back end over een internetverbinding. Om ervoor te zorgen dat deze data veilig overgedragen wordt is ervoor gekomen om een HTTPS verbinding te gebruiken. Dit zorgt er automatisch voor dat er een veilige verbinding tussen client en server is. Een alternatief is de data zelf encoderen aan de client zijde en decoderen aan de server zijde. Omdat HTTPS ervoor zorgt dat de data veilig wordt overgedragen is ervoor gekozen om encoderen en decoderen niet te gebruiken. Een andere beveiligingskeuze die er is gemaakt is om HTTP POST te gebruiken in plaats van HTTP GET. HTTP POST verbergt de data in de request, terwijl HTTP GET de data in de URL van de request zet. De consequenties hiervan zijn dat de data in HTTP POST requests in geen enkele log of geschiedenis voorkomen, wat wel kan gebeuren met HTTP GET requests. Het is ook moeilijker de data in de HTTP POST requests aan te passen dan die van de HTTP GET requests, omdat het gemakkelijker is om een URL aan te passen dan een request zelf.

4.2 Klassediagram

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

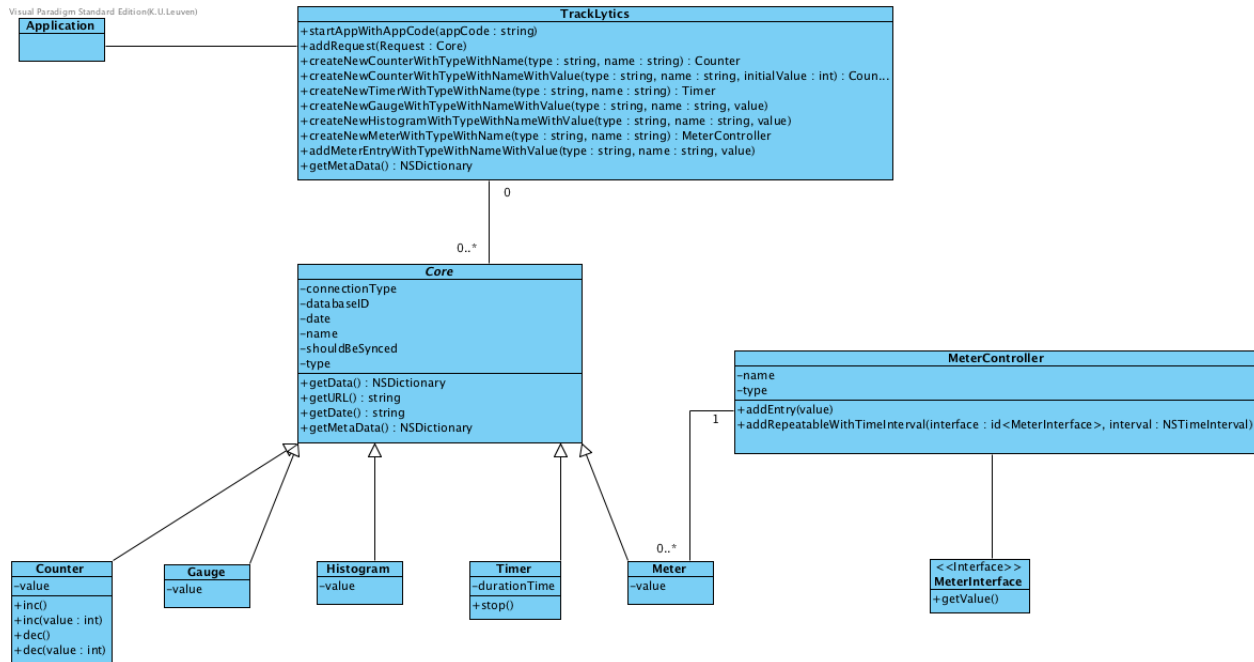
4.2.1 TrackLytics

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 er niet telkens opnieuw een instantie van deze klasse aangemaakt moet worden. Dit bespaart enerzijds code, door niet telkens opnieuw een statement te moeten schrijven die de klasse aanmaakt en anderzijds bespaart dit RAM geheugen, omdat de verschillende Tracklytics objecten in het RAM geheugen op het toestel gaan zitten.

4.2.2 Core

Om de gemeenschappelijke elementen te combineren is ervoor gekozen om die gemeenschappelijke elementen in een abstracte superklasse te steken. De gemeen-

4. IMPLEMENTATIE



FIGUUR 4.1: Klassediagram Tracklytics library.

schappelijke methodes worden ook in de Core klasse gestoken, zodat de subklassen indien nodig dit kunnen overriden.

4.2.3 Counter

De counter klasse stelt een telobject voor. Je kan bij een counter een getal optellen en aftrekken. De `inc()` functie verhoogt de waarde van de counter met 1 terwijl `inc(x)` de waarde van de counter 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 waarde van de counter gaat verlagen.

4.2.4 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.

4.2.5 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 hebben ze dezelfde functionaliteit, enkel de URL naar waar de data verstuurd wordt verschilt. Dit zorgt ervoor dat

in de back end de data op een andere plaats opgeslagen wordt en het dashboard zo deze data kan onderscheiden.

4.2.6 Timer

De timer klasse kan gebruikt worden om de lengte (in tijd) 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 lengte 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.

4.2.7 Meter

Een meter is bedoeld om een reeks van waarden te kunnen meten. Om dit voor de developer makkelijk te maken zijn er 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 3.4. In plaats van dat een Meter object wordt terug gegeven, wordt er een MeterController object terug gegeven. Een MeterController beheert de collectie van de data van Meter waardes.

Meter

Het meter object is het object dat opgeslagen wordt en dat gesynchroniseerd wordt naar de server. Het object houdt de waarde bij en ook het tijdstip van aanmaken. Per meting dat de meter in de applicatie moet er een Meter object aangemaakt worden. Dit is een van de reden dat de MeterController bestaat.

MeterController

De metercontroller zorgt ervoor dat nieuwe meters aangemaakt kunnen worden. De metercontroller bevat de naam en het type van de meter. De keuze om de meters op deze manier aan te maken berust zich op het feit dat zo niet telkens opnieuw het type van de meter moet meegegeven worden, omdat de metercontroller dit al doet en ook om de data periodiek op te kunnen halen. De metercontroller biedt een mogelijkheid aan om via de MeterInterface 4.2.7 periodiek een waarde op te halen. Deze functionaliteit zorgt ervoor dat de developer niet telkens na een bepaald interval zelf de metercontroller aan moet roepen, maar dat dit automatisch gebeurt.

MeterInterface

De meterinterface wordt gebruikt om automatisch een waarde op te halen. De metercontroller roept de enige methode die deze interface aanbiedt (`getValue()`) aan elke keer een gegeven tijdsinterval voorbij is. De developer moet deze methode implementeren in de klasse waar deze data gecollecteerd moet worden.

4.3 Documentatie

In deze sectie wordt er uit de doeken gedaan hoe je als developer de Tracklytics library in je applicatie kunt inbouwen.

4.3.1 Installatie

In deze sectie wordt besproken hoe de installatie van de library in de applicatie in zijn werk gaat.

Om de library snel en gemakkelijk in een applicatie in te bouwen is ervoor gekozen om CocoaPods te gebruiken [2]. Dit is de standaard methode om externe libraries te gebruiken in iOS. De volgende stappen moeten gebeuren om de Tracklytics library in te bouwen in de applicatie.

Allereerst moet er in de root folder van het project (via een terminal) van de applicatie *pod init* uitgevoerd worden. Dit maakt een PodFile bestand aan. Dit bestand moet aangepast worden, het platform moet aangegeven worden als *ios, 7.0* en tussen *target 'AppName' do* en *end* moet de volgende tekst komen: *pod "Tracklytics"*. CocoaPods selecteert automatisch de laatste versie van de library.

Om de library dan te installeren moet in de root folder het commando (via een terminal) uitgevoerd worden: *pod install*. CocoaPods installeert nu de library in de applicatie en zorgt dat alle dependencies geïnstalleerd worden. Dit commando genereert een nieuw project dat voortaan gebruikt moet worden in plaats van het andere project.

4.3.2 Initialisatie

De Tracklytics library moet bij het opstarten van de applicatie gestart worden om het synchronisatieproces te starten. De aanbevolen manier is om dit in de `AppDelegate` klasse te doen in de `(BOOL)application:(UIApplication *)application didFinishLaunchingWithOptions:(NSDictionary *)launchOptions` methode. Deze methode wordt automatisch uitgevoerd nadat de applicatie opgestart is, in welke applicatie dan ook. In deze methode is het dus belangrijk om volgende code uit te voeren: `[TrackLytics startTrackerWithAppCode:appCode withSyncInterval:interval];`. Deze code start het synchronisatieproces van de Tracklytics library. De parameter `appCode` stelt de unieke code voor die Tracklytics gebruikt voor het identificeren van de applicatie. De `interval` parameter geeft de tijd (in seconden) tussen twee synchronisatiecycli weer.

4.3.3 Counter

Een counter kan aangemaakt worden met behulp van volgende call naar de Tracklytics library: `[Tracklytics createNewCounterWithType:type withName:name withValue:value]`, de `withValue:value` is optioneel indien er een initiële waarde in

de counter moet staan. Deze methode geeft een CounterObject terug, wat besproken wordt in volgende sectie.

Parameters

- **type:** Het type evenement (bv. Button als er op een button geklikt wordt, vrij te kiezen door de developer). Deze parameter zorgt ervoor dat de data verdeeld wordt over de verschillende grafieken in het dashboard.
- **name:** De naam van het evenement (bv. de naam die de button heeft). Deze parameter zorgt ervoor dat de data in de grafieken verder opgedeeld worden per naam.
- **value:** De initiële waarde die de counter moet hebben wanneer die gecreëerd wordt. Deze waarde is optioneel en is standaard 0 bij creatie.

CounterObject

Bij de creatie van een Counter wordt er een CounterObject terug gegeven. Dit object kan gebruikt worden om evenementen te tellen. Een counter biedt een methode aan om de counter te verhogen: `inc` met een optionele waarde. Het object biedt ook een methode aan om de counter te verlagen: `dec`, ook met een optionele waarde. De Tracklytics library zorgt ervoor dat het CounterObject in sync blijft met de backend.

4.3.4 Gauge

Aangezien een gauge waarde niet verandert in de tijd, is het niet nodig om hier, zoals bij de counter, een object terug te geven. De waarde moet maar een keer opgeslagen en doorgezonden worden naar de server. Om de waarde op te slaan moet men de volgende methode uitvoeren: `[Tracklytics createNewGaugeWithType:type withName:name withValue:value]`.

Parameters

- **type:** Het type evenement (bv. Search Result bij een zoekopdracht). Deze parameter zorgt ervoor dat de data verdeeld wordt over de verschillende grafieken in het dashboard.
- **name:** De naam van het evenement (bv. het ingevulde woord/zin in de zoekopdracht). Deze parameter zorgt ervoor dat de data in de grafieken verder opgedeeld worden per naam.
- **value:** De waarde die gesynchroniseerd moet worden naar de server (bv. het aantal resultaten van de zoekopdracht).

4.3.5 Histogram

Een histogram heeft dezelfde eigenschappen als een gauge in termen van collectie van data in de Tracklytics library. Een histogram waarde verandert niet in de tijd en het is dus niet nodig om een object hiervan terug te geven. De methode die aangeroepen wordt verschilt enkel in de naam van de methode: `createNewHistogramWithType:type withName:name withValue:value`. De parameters worden nog eens opgesomd om een goed overzicht te hebben.

Parameters

- `type`: Het type evenement. Deze parameter zorgt ervoor dat de data verdeeld wordt over de verschillende grafieken in het dashboard.
- `name`: De naam van het evenement. Deze parameter zorgt ervoor dat de data in de grafieken verder opgedeeld worden per naam.
- `value`: De waarde die gesynchroniseerd moet worden naar de server.

4.3.6 Timer

Een timer meet de tijd dat een evenement nodig heeft. Om een timer aan te maken moet de volgende call gemaakt worden naar de Tracklytics library: `[Tracklytics createNewTimerWithType:type withName:name]`

Parameters

- `type`: Het type evenement. Deze parameter zorgt ervoor dat de data verdeeld wordt over de verschillende grafieken in het dashboard.
- `name`: De naam van het evenement. Deze parameter zorgt ervoor dat de data in de grafieken verder opgedeeld worden per naam.

Deze methode geeft een Timer object terug. Vanaf dat deze methode uitgevoerd is begint de timer te lopen. Deze blijft lopen tot de `stop` methode aangeroepen wordt op het Timer object. De timer heeft dan een exacte waarde hoeveel tijd het evenement in beslag nam. Tracklytics houdt deze waarde bij en synchroniseert deze naar de server.

4.3.7 Meter

Een meter is complexer dan alle andere meetobjecten in de Tracklytics library. Dit is omdat ervoor gekozen is om de mogelijkheid aan te bieden om data automatisch op een bepaald interval te collecteren. Zo moet de developer hier niet meer naar omkijken. Om dit te kunnen verwezenlijken is er een controller gemaakt, de `MeterController`, die de mogelijkheid biedt voor de automatische collectie van data en van de manuele invoer van data. Een `MeterController` can gecreëerd worden door de volgende call naar de Tracklytics library: `[Tracklytics createNewMeter:type]`.

Parameters

- type: Het type evenement. Deze parameter zorgt ervoor dat de data verdeeld wordt over de verschillende grafieken in het dashboard.

MeterController

De metercontroller zorgt voor het managen van de meter values. Het is mogelijk om manueel data toe te voegen of automatisch de data te laten controlleren.

Om data manueel toe te voegen moet de volgende methode uitgevoerd worden: `addEntry: (float) value`. De value parameter is de waarde die moet toegevoegd worden aan de meter.

Om de data automatisch te laten collecteren moet de methode `addRepeatable:(id<MeterInterface>) interface withTimeInterval:(NSTimeInterval) interval` uitgevoerd worden. Deze neemt een object dat de MeterInterface implementeerd als parameter en een interval.

Het MeterInterface object implementeerd de `(float) getValue` methode die een float waarde terug geeft die de MeterController kan collecteren door die methode. Het interval geeft aan (in seconden) hoeveel tijd er tussen twee cyclussen zit waarin de waarde opgehaald wordt.

4.4 Aggregatie

De data die de Tracklytics library opgehaald heeft bij het monitoren van de applicatie moet geaggregeerd worden om weer te kunnen geven in het dashboard. De logische plaats om deze aggregatie uit te voeren is de back end, omdat deze elk meetpunt in de database heeft staan. Het voordeel hieraan is dat de data opgehaald kan worden en enkel verwerkt wordt wanneer deze nodig is. Een nadeel hieraan is dat dit zeer veel CPU kan kosten en zo het dashboard enorm traag maakt. Dit nadeel is gelinkt aan de hoeveelheid gebruikers van de applicatie en het aantal meetpunten in de applicatie. Groeit een van beide of beiden, dan treedt er een extra vertraging op in het aggregeren van de gegevens. Dit probleem kan opgelost worden door op voorhand te aggregeren, dit kan op twee manieren: **op het toestel zelf** en **op periodieke intervallen**. Deze sectie berust zich op het feit dat I/O operaties veel trager zijn dan verwerkingsoperaties.

De eerste oplossing is om de aggregatie uit te voeren op het toestel zelf. De Tracklytics library kan aangepast worden om in plaats van de data van de meetpunten zelf door te sturen, eerst deze te verwerken op het toestel zelf en die verwerkte data door te sturen naar de server. De back end zal ook moeten aangepast worden om deze data op te slaan zodat deze herbruikt kan worden. Het voordeel hieraan is dat als het dashboard de data opvraagt van de back end, de back end dit sneller zal kunnen doen, omdat er minder data uit de database opgehaald en verwerkt moet

worden. Zo is het dashboard bruikbaar bij grotere schaal. Een nadeel hieraan is dat de gegevens niet meer optimaal zijn. Bijvoorbeeld: de mediaan kan enkel nog benaderd worden, omdat niet alle waarden aanwezig zijn in de database. Door afrondingsfouten kunnen de gemiddeldes verschillen van de werkelijke gemiddeldes.

Een tweede oplossing is om periodiek (bv. één keer per dag) per applicatie de nieuwe data te gaan aggregeren en samenvoegen met de al geaggregeerde data van de voorbije tijdstippen. Deze functionaliteit moet op de servers geïmplementeerd worden. Het voordeel hieraan is dat indien het dashboard de data opvraagt, enkel de nieuwe data nog moet samengevoegd worden met de al geaggregeerde data. Dit zorgt voor een enorme performance groei van het dashboard. Het nadeel hieraan is dat ofwel er een overzicht moet bijgehouden worden welke data al geaggregeerd is en welke niet, ofwel de verwerkte data telkens uit de database verwijderd moet worden om delen van de dataset niet opnieuw te verwerken. Dit vraagt dus een extra inspanning om te implementeren.

Een derde oplossing is om de twee te combineren. Op het toestel wordt een deel van de aggregatie al gedaan en deze geaggregeerde data wordt dan verwerkt op de servers zoals in het vorige deel. Dit geeft een extra prestatiewinst, de verwerking van de geaggregeerde data op de servers gaat sneller, omdat deze al geaggregeerd is. Er is dus minder CPU-tijd nodig om de opdracht af te werken. Indien het dashboard de data opvraagt aan de back end gaat dit ook veel sneller. De combinatie van de nieuwe data en geaggregeerde data gaat veel sneller, omdat er minder data aanwezig is.

Een laatste oplossing is om de data te aggregeren wanneer het echt nodig is en deze dan op te slaan. Als de developer of eigenaar van de applicatie het dashboard voor de eerste keer opent, dan wordt deze data geaggregeerd en opgeslagen zoals bij de tweede oplossing. Als de developer of eigenaar het dashboard de volgende keer opent moet enkel de nieuwe data geaggregeerd worden en gecombineerd met de al geaggregeerde data. Deze combinatie wordt dan opnieuw opgeslagen. Dit proces wordt herhaald telkens dat het dashboard van die applicatie geopend wordt. Het voordeel hieraan is dat dit sneller is dan telkens opnieuw alle data te aggregeren. Het nadeel hieraan is dat er moet bijgehouden worden welke data verwerkt is en welke niet.

In de Tracklytics library is ervoor gekozen om de developer de keuze aan te bieden tussen het aggregeren van de data in de back end en het aggregeren van de data op het toestel. De Tracklytics library aggregereert data van de volgende meetobjecten op het toestel: **Gauge, Meter, Timer**.

Het Gauge meetobject wordt geaggregeerd op het toestel door tijdens het aggregeren *het gemiddelde, de mediaan, de laagste en de hoogste waarde* te berekenen

op basis van de waarden die gemeten worden. In plaats van telkens een nieuwe Gauge waarde aan te maken wordt deze toegevoegd aan een helper object die deze waarden bij houdt om de mediaan te berekenen en bij het toevoegen van een waarde ineens het gemiddelde berekent en bekijkt of deze waarde de hoogste of laagste waarde is. Het grootste nadeel hieraan is dat de mediaan in de back end enkel nog benaderd kan worden omdat niet alle waarden nog in de back end aanwezig zijn.

Het Meter meetobject wordt als volgt geaggregeerd: in plaats van elke waarde apart op te slaan wordt het aantal metingen bijgehouden en wordt er een gemiddelde bij gehouden. Deze waardes kunnen in de back end gecombineerd worden om een globaal overzicht te krijgen. Het nadeel hieraan is dat er afrondingsfouten kunnen gebeuren bij het berekenen van het gemiddelde en er zo een verkeerde observatie gemaakt kan worden. Ook wordt er niet per meting bijgehouden op welk tijdstip deze voorvalt.

Het Meter meetobject wordt geaggregeerd door de gemiddelde tijd te berekenen bij het stoppen van een timer en bij te houden hoeveel metingen er voorgekomen zijn. Het nadeel is dat er per timer maar één instantie tegelijkertijd kan lopen omdat Tracklytics dezelfde methodes aanbiedt voor een geaggregeerd object en een niet-geaggregeerd object. Net zoals bij het *Meter meetobject* kunnen er hier aggregatiefouten voorkomen.

4.5 Openstaande uitdagingen

De implementatie gegeven in deze thesis dekt niet de volledige architectuur uit de vorige sectie. Er blijven dus nog steeds uitdagingen om de library uit te breiden zodat deze dichter aanleunt aan de architectuur. Er worden ook andere uitbreidingen beschreven die niet in de architectuur beschreven worden, maar van even groot belang zijn dan degene die wel in de architectuur beschreven wordt.

4.5.1 AB Testing

AB testing wordt gebruikt om geleidelijk aan een nieuwe versie van een applicatie of website uit te rollen, zoals beschreven in de architectuur sectie. Het is een uitdaging om dit concept te combineren met de Tracklytics library. Er moet onderzocht worden wat de efficiëntste manier is om AB testing in mobiele applicaties mogelijk te maken. Een manier om AB testing te gebruiken is om eerst het versienummer voor die gebruiker op te halen in de database. Een niet efficiënte manier zou dan zijn om per versie een if statement te gebruiken om zo de verschillende karakteristieken van de versie te bepalen. Deze manier is op drie manieren inefficiënt, namelijk: de code van de applicatie (en dus ook de grootte op schijf) wordt groter naarmate er meer

versies worden gebruikt, door de if statements wordt de applicatie trager en als er een aanpassing moet gebeuren aan een versie of er wordt een versie toegevoegd moet deze nog steeds eerst via de app store als update uitgevoerd worden.

4.5.2 Operating systems

De Tracklytics library is ontwikkeld voor iOS, het besturingssysteem van Apple. Er bestaat buiten iOS nog een besturingssysteem dat de moeite waard is om te bekijken, namelijk Android. Op het moment van schrijven is Android de marktleider met een marktaandeel van 59.65% in de mobiele markt, gevolgd door iOS (32.28%). De andere mobiele besturingssystemen zijn niet de moeite waard om te bekijken, omdat deze een bijna verwaarloosbare marktaandeel hebben. De toestellen die Android gebruiken zijn verdeeld over vele versies. Op moment van schrijven draait ongeveer een derde van de toestellen op een twee jaar oud besturingssysteem (KitKat 4.4). In contrast, 79% van de toestellen die op iOS draaien hebben de laatste versie geïnstalleerd (op moment van schrijven). Elke nieuwe versie brengt veranderingen en vernieuwingen met zich mee, wat er voor zorgt dat sommige taken sneller/trager uitgevoerd worden in de ene versie dan in de andere versie. Dit heeft mede de keuze voor iOS gemaakt.

De uitdaging die zich hier voordoet is het schrijven van een library voor Android. Dezelfde architectuur kan worden gebruikt voor de library. Er moeten enkele aanpassingen gedaan worden om dit te doen werken. Allereerst moet er een onderscheid gemaakt kunnen worden tussen iOS en Android data. Door telkens het type besturingssysteem door te sturen als metadata is dit mogelijk om het zo in de database te stoppen. Anderzijds is het noodzakelijk in het Android geval om de versie van het besturingssysteem mee door te sturen, zodat er in het dashboard een duidelijke opdeling hiertussen kan gemaakt worden. Dit zorgt er dan weer voor dat het dashboard complexer wordt om te implementeren.

4.5.3 Alarmen

Om het concept en nut van alarmen uit te leggen is het handig om met een voorbeeld te beginnen.

Een eigenaar heeft een applicatie ontwikkeld die communiceert met de back end. De eigenaar heeft ervoor gekozen om servers te huren in plaats van zijn eigen servers te bouwen en onderhouden. De eigenaar wil dat er zo weinig mogelijk servers gehuurd moeten worden en dat de gebruiker van de applicatie geen zichtbare vertraging heeft (er toch genoeg servers zijn om de requests te behandelen). Zo kunnen de kosten geminimaliseerd worden en dus de winst gemaximaliseerd. Developers moeten dus servers toevoegen als de servers de requests niet meer aan kunnen en een server verwijderen indien die server overbodig wordt om het werk gedaan te krijgen.

Om dit mogelijk te maken moet er een trigger of alarm komen die de developers waarschuwt indien de huidige servers het werk niet meer snel genoeg kunnen doen.

Dit alarm kan geïmplementeerd worden in de servers of in de Tracklytics library.

Er bestaan al meerdere oplossingen die deze functionaliteit aanbieden op het vlak van servers. Hier wordt niet verder op ingekeken omdat dit niet in de context van deze thesis past.

In de Tracklytics library kan er een alarm functionaliteit aangeboden worden. Een developer moet dan in het dashboard aangeven welke parameter welke waarde maximum mag krijgen. Het is ook handig dat de developer een maximum aantal schendingen van deze parameter kan aangeven en dat hij enkele types van metadata kan uitsluiten. (bv. de tijd dat het kost om data van de back end op te halen mag maximum bij 5 verschillende gebruikers boven de 10 seconden liggen op WiFi en 4G). De developer moet ook de tijdspanne aanduiden waarin deze parameters niet overschreden mogen worden. Indien deze drempel overschreden wordt, moet er een waarschuwing gestuurd worden naar de developer. Hierdoor moeten de drie componenten samenwerken (Dashboard, Back end en mobiele library). Het dashboard moet de gegeven paramaters doorgeven aan de back end zodat deze opgeslagen kunnen worden in de database. De mobiele library moet, via de back end, de parameters ophalen. Indien er dan een match is wanneer er nieuwe gegevens worden doorgegeven via de applicatie, dan moet de library nagaan of deze de drempel niet overschrijden. Indien de drempel wordt overschreden, dan moet dit gemeld worden aan de back end. De Tracklytics library moet dan een call doen naar de back end. De back end slaat deze melding op in de database en gaat na of er een alarm verstuurd moet worden. De back end kijkt dus of er in de gegeven tijdspanne meer schendingen geweest zijn dan dat de developer als maximum aangaf. In plaats van elke keer dat er een schending is na te gaan of er een alarm verstuurd moet worden, kan de back end periodiek nagaan of de parameters overschreden worden. Het voordeel hieraan is dat de back end sneller is, omdat de controle niet telkens moet uitgevoerd worden. Het nadeel is dat het alarm bijna een volledige periode later uitgezonden wordt indien de schending van de parameters aan het begin van de periode plaatsvindt.

De alarm functionaliteit kan ervoor zorgen dat er zo weinig mogelijk servers gebruikt worden en het geen impact heeft op de kwaliteit van de applicatie en niet zichtbaar is voor de gebruikers. Zo kunnen de kosten geminimaliseerd worden en de winst gemaximaliseerd.

Hoofdstuk 5

Evaluatie

In deze sectie wordt de ontwikkelde library geëvalueerd. Allereerst wordt er gekeken naar wat de impact van de library op de performance van de applicatie is. Verder in dit hoofdstuk wordt de schaalbaarheid van de library bekeken. Ten slotte wordt er nog gekeken hoeveel effort er van de developer nodig is om de library te integreren in een applicatie.

5.1 Performance

In het verloop van deze thesis is al enkele keren aan bod gekomen dat de prestaties van een applicatie één van de belangrijkste eigenschappen van een applicatie is. De impact op de performance van de library op de performance moet aanvaardbaar zijn en zo klein mogelijk om de applicatie niet significant te doen vertragen. In deze sectie wordt deze impact onderzocht en geëvalueerd.

Om de performance te meten zijn er twee verschillende applicaties gekozen, namelijk een stopwatch applicatie en een angry birds kloon [1]. De stopwatch wordt gebruikt omdat dit een niet CPU intensieve applicatie is en zo de resultaten niet vertekend zijn. De angry birds applicatie is CPU intensiever en vereist een bepaalde framerate, aangezien dit een spel is. Deze applicatie wordt gebruikt om de framerate te meten en de impact van de applicatie hierop.

In de volgende secties worden de metingen uitgelegd en uitgevoerd. Tenslotte wordt hier ook een conclusie uit getrokken. Er gaat gemeten hoe lang het gemiddeld duurt per methode die gebruikt kan worden van de library. Dit wordt gemeten langs twee aspecten: de data wordt wel tijdelijk op harde schijf opgeslagen en de data wordt niet tijdelijk op harde schijf opgeslagen. Nadien wordt er gekeken naar de impact van de library op de framerate van de angry birds kloon. Dit laatste combineren we met het wel of niet aggregeren van de data op het toestel zelf.

5.1.1 Call duration

In deze sectie wordt uitgezocht hoelang het gemiddeld duurt voor elke methode om volledig uitgevoerd te worden. Zo kan er een beeld gevormd worden van hoe sterk de library de applicatie zou vertragen.

5.2 Schaalbaarheid

5.3 Developer effort

Bibliografie

- [1] BoloYoung. Angry birds. <https://github.com/BoloYoung/AngryBirds>.
- [2] CocoaPods. <https://www.cocoapods.org>.
- [3] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh. Why people hate your app: Making sense of user feedback in a mobile app store. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1276–1284. ACM, 2013.
- [4] S. V. Gils. <https://github.com/Stefvg/Tracklytics>.
- [5] K. MacLaine. Why your app sucks: The 20 most common complaints about mobile apps. <http://appealingstudio.com/why-your-app-sucks-the-20-most-common-complaints-about-mobile-apps-2/>.
- [6] NetMarketShare. Mobile/tablet operating system market share. <https://www.netmarketshare.com/operating-system-market-share.aspx?qprid=8&qpcustomd=1>.
- [7] Statista. Number of smartphone users worldwide from 2014 to 2019 (in millions). <http://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>.

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