



Open Data Platform voor steden en gemeenten

Technische beschrijving
versie 4, 2021



INTRODUCTIE

Steden en gemeenten in Vlaanderen willen efficiënter en duurzamer worden om een aangenamer leven te creëren voor hun burgers. Lokale besturen staan daarbij in nauw contact met de mensen. Het dagelijkse leven van de inwoners speelt zich dan ook af in steden en gemeenten. Maatschappelijke uitdagingen en problemen van de burgers zijn hier dus het meest zichtbaar. Lokale besturen kunnen oplossingen ontwikkelen die rechtstreeks impact hebben op het leven van de Vlaming.

Wanneer lokale besturen investeren in technologie om allerlei uitdagingen aan te pakken, is dat best toekomstgericht: de oplossing moet duurzaam zijn en een basis leggen voor verdere innovatie. Steden en gemeenten bieden al een digitale dienstverlening aan, maar momenteel ontbreekt nog een zekere mate van coördinatie. Diensten worden nog te veel los van elkaar aangeboden, inwoners moeten te vaak dezelfde gegevens indienen en de verschillende toepassingen vereisen vaak andere handelingen. Dat maakt digitale diensten nog niet volledig toegankelijk. Door de burger centraal te stellen, kunnen we technologie nog beter gebruiken. Zo kan technologie de meest duurzame manier worden om het leven in onze steden en gemeenten te verbeteren. Er zijn al voldoende componenten en modellen om smart cities uit te bouwen, de uitdaging is om deze kennis werkbaar te implementeren.

INHOUDSTAFEL

1. Gegevens zijn de toekomst	3
2. Hoe maken we gegevens nuttig	4
3. Soorten data	5
4. Data-ontsluiting	6
4.1 Slimme, eigen toepassingen	6
4.2 Visualisatie	6
4.3 Smart City Dashboard	6
4.4 Het MyCSN Dashboard met standaardtoepassingen	7
4.5 Data als commerciële informatie	7
4.6 Vlaams Open Data Portaal	7
4.7 Eigen Open Data Portaal	8
4.8 Digitale borden	8
4.9 Onze Stad App	8
5. De Levensloop van data	9
6. Open data platform voor steden en gemeenten	10
7. Lambda-architectuur	11
8. Componenten	12
8.1 DC/OS	12
8.2 VerneMQ	12
8.3 Apache Kafka	12
8.4 Apache Spark	12
9. Logische architectuur	13
10. Kernfuncties	14
10.1 Connectiviteit beheren	15
10.2 Inzichten genereren	15
10.3 Geavanceerde analyses	15
10.4 IoT-oplossingen op maat	15
10.5 Cloud-onafhankelijk	15
10.6 Toestellen verbinden	16
10.7 Gegevensbescherming	16
10.8 Niet-functionele vereisten	16
10.9 Cluster resource manager	16
10.10 Automatisatie	16
10.11 Container orchestration	16
10.12 Multi-tenant	17
10.13 Common services	17
10.14 Aanpassingen aan elke tenant	17
10.15 Open Platform	17
10.16 Security	18
10.17 Open Standaarden	19
10.18 Toegepaste Standaarden	19

1. GEGEVENS ZIJN DE TOEKOMST

Al decennialang maken lokale besturen beleidsbeslissingen op basis van beschikbare gegevens. Het verschil tussen vroeger en nu is de vorm en het volume ervan. De evolutie van traditionele data naar de zogenaamde Big Data zorgt ervoor dat conventionele middelen om gegevens te verzamelen, op te slaan, te analyseren en verwerken, niet meer volstaan. Big Data betekent enorme hoeveelheden data die aan hoge snelheid uit verschillende bronnen komen.

Big Data brengt uitdagingen met zich mee omdat de gegevens niet noodzakelijk gestructureerd binnenstromen en de bronnen waar men uit put, nemen alleen maar toe. Bovendien zijn niet alle binnenkomende gegevens waardevol, een deel ervan is mogelijk niet van goede kwaliteit en heeft een mate van onzekerheid. Er zijn bovendien steeds meer data beschikbaar, onder andere door het gebruik van sociale media (Twitter, Facebook, Instagram, YouTube, enzovoort), steeds meer Internet of Things-sensoren, analyses van web-sitebezoeken, een groeiende online hoeveelheid video- en beeldmateriaal en de ononderbroken communicatie via smartphones.

Daardoor is er een kantelmoment aangebroken in hoe lokale besturen omgaan met gegevens die hun beleid ondersteunen. Om met Big Data te werken, is er kennis en geschikte apparatuur nodig. Het MyCSN dataplatform is het centrale, beveiligde punt in de cloud om die data te verzamelen, te beheren, gecontroleerd te delen en te analyseren.

Ontwikkelingen op technologisch gebied en kennisdeling laten zich niet beperken tot de grenzen van een stad of gemeente. Daarom passen we de huidige open standaarden voor gegevensverwerking toe en zullen we die blijven toepassen naarmate er meer afspraken over standaarden komen. Door aan standaarden te voldoen, uniformiseren we data en maken we samenwerking en innovatie mogelijk. Het MyCSN dataplatform is ontwikkeld om de sprong te maken naar de nieuwe behoeften van steden en gemeenten in een wereld van data.

2. HOE MAKEN WE GEGEVENS NUTTIG

Het MyCSN dataplatform, meetapparatuur zoals sensoren en applicaties zijn enkele voorbeelden van de technologie die u kan inzetten als oplossing voor de vraag: Hoe maak ik het leven van de burgers aangenamer? Maar de technologie is slechts stap één. Wat u met de gegevens doet die beschikbaar zijn, is de volgende stap.

Door digitale data uit een groeiend aantal bronnen te verzamelen en te analyseren, kunnen steden en gemeenten hun beleid op een meer efficiënte manier voorbereiden, monitoren en bijsturen. Dat kan door die beschikbare data te analyseren en te combineren met gegevens uit andere domeinen en andere bronnen zoals bijvoorbeeld mobiliteit, milieu, nutsvoorzieningen, afvalverwerking, economie en meer. Hoe u die inzichten deelt met inwoners, uw beleid ermee ondersteunt of ze inzet om slimme toepassingen te maken, dat noemen we data-ontsluiting.



3. SOORTEN DATA

Voor we het hebben over hoe we data ontsluiten, is het interessant om te weten waar de data vandaan komen en welke soorten er zijn. Op het MyCSN dataplatform komen heel wat soorten data samen.

Ten eerste zijn er de meetgegevens uit het zogenaamde Internet of Things. Dat zijn data die worden verzameld in het publieke domein, zoals bijvoorbeeld sensoren in lichtmasten of meetapparatuur in andere infrastructuur.

Een ander soort data zijn gegevens uit interne toepassingen van besturen, organisaties en bedrijven die zij beschikbaar kunnen stellen op het MyCSN dataplatform. Dat gebeurt steeds op basis van goede afspraken en met respect voor de richtlijnen rond gegevensbescherming. Denk bijvoorbeeld aan de realtime bezetting in commerciële parkeergarages.

Eenmaal de gegevens op het MyCSN dataplatform staan, kan het bestuur of de organisatie die ze aanbiedt ervoor kiezen om ze constant bij te werken in een zogenaamde datastroom. Andere gebruikers van het platform kunnen zich hier eventueel op abonneren.

Daarnaast stelt MyCSN andere data ter beschikking. Dat zijn externe, open of aangekochte data die nuttig zijn om toepassingen mee te ontwikkelen. Een voorbeeld van open data is weersinformatie en een voorbeeld van aangekochte data is geanonimiseerde data over het verbruik van wind- en zonne-energie of mobiliteitsdata van Be-mobile.



4. DATA-ONTSLUITING

Deze soorten data zijn allemaal op één centrale plaats beschikbaar: het MyCSN dataplatform. Hoe gebruiken we die data in de praktijk?

4.1 SLIMME, EIGEN TOEPASSINGEN

U kan gegevens gebruiken om een toepassing aan te sturen. Als startpunt kan u bijvoorbeeld meten hoeveel verkeer er voorbijrijdt op bepaalde punten in de stad en dit combineren met hoe het met de luchtkwaliteit gesteld is. Wilt u als bestuur dan inzetten op verminderde uitstoot of meer verkeersveiligheid op het moment dat de kinderen naar school gaan? Op die manier ontstaan ideeën voor slimme toepassingen om het beleid te ondersteunen.

MyCSN ondersteunt zulke toepassingen, zoals buck-e. Dat is een app die kinderen belooft om met de fiets naar school te gaan en ouders kunnen zien wat de impact hiervan is op ecologie en verkeersveiligheid.

4.2 VISUALISATIE

Er bestaan verschillende tools om op een snelle manier de ruwe data te bekijken in begrijpbare grafieken en beelden. Grafana is zo'n visualisatietool waarmee u snel duidelijke informatie uit ruwe gegevens kan halen. Deze applicaties zijn beschikbaar voor de gebruikers van het MyCSN data-platform en bieden visuele inzichten zoals heatmaps en taartgrafieken. Ook dit is een vorm van data-ontsluiting.

4.3 SMART CITY DASHBOARD

Het MyCSN dataplatform biedt een eigen portaal voor iedere stad of gemeente: het Smart City Dashboard. Daar kan elke stad of gemeente een eigen overzicht creëren op basis van interne data, maar ook met informatie van anderen om gecombineerde inzichten te maken. Door in te loggen op MyCSN kan u uw data in real-time bekijken. Er zijn verschillende inzichten over gegevenssets beschikbaar op het Smart City Dashboard: parkinginformatie, mobiliteit, data over de productie van groene energie en binnenkort ook over sluikstorten.

Naast real-time informatie zijn er ook historische grafieken van de gehele gemeten periode zodat u vergelijkingen kan maken om evoluties te zien. Produceert men meer of minder hernieuwbare energie? Hoe druk is het nu in de stad en zijn er piekmomenten? Met vergelijkingen als deze kan u beleidsinformatie vergaren die u kan inzetten om een slim beleid te ondersteunen. MyCSN maakt dit portaal met verschillende standaarddashboards beschikbaar voor alle steden en gemeenten die klant zijn. Daarnaast bieden we op aanvraag ook inzichten op maat aan.

4.4 HET MYCSN DASHBOARD MET STANDAARDTOEPASSINGEN

Op het MyCSN Dashboard kunnen steden en gemeenten de standaardtoepassingen van MyCSN in actie zien. Bestaande toepassingen zoals Hydrosan overstromingsvoorspelling, Cambio deelauto's, buck-e voor veiliger en ecologisch verkeer kan u real-time in de gaten houden voor uw stad of gemeente.

Elke gegevensstroom die u beheert, kan u met het MyCSN Dashboard visualiseren op een overzichtelijke en toegankelijke manier. Ook hier kan u vergelijkingen maken met verschillende periodes of de data opsplitsen per zone.

4.5 DATA ALS COMMERCIËLE INFORMATIE

Vanuit het MyCSN dataplatform kunnen steden en gemeenten hun data commercieel ter beschikking stellen voor derde partijen. Dat gebeurt via de toepassing databroker.global. Zo kunnen data-afnemers zoals ontwikkelaars van toepassingen een contract aanvragen om aan een afgesproken bedrag anonieme data te verkrijgen van de stad of gemeente. Een stad die bijvoorbeeld zelf parkinginformatie opmeet en analyseert, zou die kunnen verkopen aan applicaties voor mobiliteit of parkeerfaciliteiten.

Ook het commerciële proces is volledig geautomatiseerd om het delen van data toegankelijk te maken voor alle steden en gemeenten op het MyCSN dataplatform. Zo blijven alle handelingen met uw data gecentraliseerd, veilig en overzichtelijk.

4.6 VLAAMS OPEN DATA PORTAAL

Een ander soort data-ontsluiting is via het Vlaams Open Data Portaal, de centrale vindplaats voor open data van overheden in Vlaanderen. Iedere gebruiker van het MyCSN dataplatform heeft rechtstreeks toegang tot het Open Data Portaal. U kan er ook zelf uw open data beschikbaar stellen en up-to-date houden.

Open data delen is een waardevolle bron om gecombineerd inzichten te maken. Bepaalde data die gemeten worden in de openbare ruimte zijn per definitie open en moeten dus beschikbaar zijn voor het publiek. Door dit via het MyCSN dataplatform te doen, vermijdt u dat u voor deze data aparte opslag en beheer moet verzorgen.

4.7 EIGEN OPEN DATA-PORTAAL

Als bestuur kan u ook zelf een portaal opzetten om open data ter beschikking te stellen aan anderen. MyCSN stelt open data-portalen ter beschikking aan de steden en gemeenten.

Dat verloopt meestal via CKAN, een online tool waarmee een bestuur een portaal kan maken en aanpassen zodat ze zelf hun open data kunnen beheren. Andere tools om open data te delen, zijn Metabase of Open Data Soft. Ook deze tools laten toe dat steden en gemeenten zelf instaan voor het beheer van hun open data.

4.8 DIGITALE BORDEN

Digitale borden zijn een handige manier om de inzichten uit data te gebruiken als informatie voor de inwoners. Deze schermen kan u op publieke plaatsen zetten, zoals bijvoorbeeld het treinstation, de markt, de bibliotheek of het stedelijke zwembad. De inhoud van de schermen kan de stad of gemeente zelf bepalen, alle toepassingen die ze gebruiken via MyCSN kunnen we visualiseren. Zo delen steden en gemeenten op eenvoudige manier de data die ze op het Smart City Dashboard beheren.

Aan het station Antwerpen-Centraal op de Rooseveltplaats staat bijvoorbeeld een digitaal bord waar reizigers in real-time kunnen volgen waar en wanneer hun bus of tram aankomt, voor een vlotte verbinding. In Lommel kunnen voorbijgangers op digitale borden zien

hoeveel groene stroom er geproduceerd wordt in de regio en of ze meer of minder stroom verbruiken dan omliggende gemeenten.

4.9 ONZE STAD APP

Een tweede manier om de gegevens van op de Smart City Dashboard te delen met de burgers is de Onze Stad App. De mobiele toepassing draagt steeds de naam van de stad of gemeente zelf en kan volledig gepersonaliseerd worden om bij de stadsmarketing te passen. Burgers hebben met een app alle informatie en toepassingen van de stad of gemeente in hun broekzak.



5. DE LEVENSLLOOP VAN DATA

Van ruwe data tot bruikbare informatie; gegevens kennen een levensloop. Dat vatten we even samen: van zodra een stad of gemeente iets gaat meten, ontstaat er data. Die ruwe gegevens vertellen nog niet veel, maar na analyse krijgen we meer inzicht in wat er precies gemeten werd.

Daarna kunnen we data fuseren of combineren met andere data, wat zorgt voor meer verhelderende overzichten en informatie over concrete zaken. Om die reden kan data delen en commercieel ter beschikking stellen interessant zijn.

We bespraken ook dat er verschillende soorten data bestaan. Om ervoor te zorgen dat we data kunnen combineren, delen en kunnen blijven gebruiken, uniformiseren we alle data. MyCSN zorgt ervoor dat alle data op het data-platform op z'n minst voldoen aan de standaarden die Vlaanderen aanbeveelt en zelf gebruikt. We houden ook rekening met nieuwe standaarden die in allerlei werkgroepen in ontwikkeling zijn, ook op internationaal vlak.

Om vervolgens de data goed te begrijpen, is visualisatie in applicaties of dashboards essentieel. Hoe meer data er is en hoe langer de gemeten periode, hoe meer we in staat zijn om vergelijkingen en voorspellingen te maken.

Data wordt dus steeds interessanter en waardevoller. Zo stijgt de waarde op commercieel gebied en uiteraard als beleidshefboom.

Omdat al deze stappen op het MyCSN data-platform gecentraliseerd verlopen, behoudt u de controle en het overzicht over uw data. We kunnen samen bouwen aan uw smart city-project.

6. OPEN DATA PLATFORM VOOR STEDEN EN GEMEENTEN

MyCSN heeft een dataplatform ontwikkeld om aan steden en gemeenten tegemoet te komen. Het MyCSN dataplatform is een centrale, toegankelijke plek waar steden en gemeenten eigenaar blijven van hun gegevens en hun data gecontroleerd kunnen uitwisselen. Omdat alle gegevens op het MyCSN dataplatform voldoen aan de open standaarden, creëren we kennis over domeinen heen. Zo bouwen we samen aan de toekomst en aan een slimme regio door technologie in te zetten voor beter beleid. In dit document bespreken we de technische specificaties van het MyCSN dataplatform. Zo schep- pen we een beeld van de functies en de mogelijkheden die het MyCSN dataplatform biedt.

Het MyCSN dataplatform is ontwikkeld om grote hoeveelheden gegevens te verwerken en geavanceerde analyses uit te voeren. Achterliggend is het platform gebaseerd op beproefde openbrontechnologie die toelaat om de verwerkingscapaciteit ogenblikkelijk aan te passen aan de behoefte. Commerciële bedrijven zoals LinkedIn, Facebook of Airbnb toonden aan dat niet alleen traditionele bedrijfssoftware, maar ook hun openbrontechnologie een antwoord kan bieden op de vraag naar zeer flexibele capaciteit en het verdelen ervan over clusters in de cloud.

Bij het verwerken van grote hoeveelheden semigestructureerde gegevens is er altijd een vertraging tussen het moment waarop ze worden verzameld en de beschikbaarheid ervan in dashboards. Vaak ontstaat de vertraging uit de noodzaak om ruwe gegevens te valideren of op zijn minst correct te identificeren. In sommige gevallen is het echter belangrijker om snel op nieuwe gegevens te kunnen reageren in plaats van volledige zekerheid te hebben over de geldigheid van de gegevens.



7. LAMBDA-ARCHITECTUUR

Door een lambda-architectuur te gebruiken, kan het platform elke vorm van gestructureerde en ongestructureerde data opnemen en verwerken zonder vooraf strikte datamodellen en -schema's te definiëren. Dit zorgt voor meer flexibiliteit om nieuwe gegevensbronnen toe te voegen.

De opbouw van het MyCSN dataplatform is gebaseerd op bouwblokken in een zogenaamde lambda-architectuur. Dat is een gedocumenteerde opstelling voor de verwerking van enorme hoeveelheden gegevens en het snel visualiseren ervan. Deze architectuur zoekt een evenwicht in vertraging, doorvoer en fout-tolerantie; drie kenmerken van gegevensverwerking die men voordien moeilijk in balans kreeg. De batch processing in de lambda-architectuur biedt uitgebreide en nauwkeurige weergave van de gegevens. Visualisatie van onlinegegevens gebeurt door real-time stream processing. De resultaten van beide vormen van gegevensverwerking (batch en stream), voegen we samen voor verdere visualisatie of verwerking.

We overlopen de technologie die het MyCSN dataplatform zijn unieke eigenschappen geeft. In de volgende hoofdstukken bespreken we in detail de bouwstenen.

De basislaag van het MyCSN dataplatform wordt geleverd door DC/OS (Distributed Cloud Operating System). Dat is een openbron, gedistribueerd besturingssysteem dat de cloud-infrastructuur beheert en live houdt.

De message broker van het dataplatform zorgt ervoor dat de communicatie tussen alle machines vlot verloopt en is vooral gericht op communicatie met de buitenwereld. We gebruiken hiervoor VerneMQ omwille van de flexibiliteit en de kracht ervan.

Met Apache Kafka beheren we de interne distributie van de datastromen, wat een architectuur van microservices mogelijk maakt. Het betekent dat elke applicatie van de cloudomgeving op een eigen infrastructuur steunt, met betrouwbaarheid en flexibiliteit als belangrijkste voordelen.

Dan zijn er nog de tools die het MyCSN dataplatform de vorm van een Software as a Service (SaaS) geven. De data-basesoftware Apache Druid in combinatie met een PostgreSQL datawarehouse en de Cloud Object Storage van Amazon zorgen voor data lake-functies die zaken zoals historische analyses mogelijk maken. Een andere belangrijke bouwsteen van het MyCSN dataplatform is het Apache Spark ecosysteem voor de analyse van Big data en machine learning. Dat maakt geavanceerde analyses mogelijk op de gegevens van het MyCSN dataplatform. Dat kan zowel op real-time gegevens als op historische gegevens met behulp van Apache Spark Streaming of Apache Flink.

8. COMPONENTEN

We overlopen een aantal bouwstenen van het MyCSN data-platform met hun functies en eigenschappen.

8.1 DC/OS

DC/OS (Distributed Cloud Operating System) is een open-bron, gedistribueerd besturingssysteem gebaseerd op Apache Mesos en ontstaan als een onderzoeksproject aan de universiteit van Berkeley. DC/OS beheert meerdere nodes in de cloud of in een on-premises datacenter vanuit één grafische interface waarop MyCSN en alle tenants hun containerized applicaties hosten. DC/OS host als hypervisor ook gedistribueerde containerized services of third party applicaties, biedt het beheer van netwerk en service discovery en verzorgt resource management en scheduling om de services draaiende te houden en met elkaar te laten communiceren. DC/OS kent steeds meer succes gezien de mogelijkheid om zowel streaming data als processing workloads (zoals Apache Spark en GPU-geactiveerde TensorFlow) te draaien als containerized microservices in één platform.

8.2 VERNEMQ

VerneMQ is de MQTT-broker van het MyCSN dataplatform. Deze broker dient als protocol adapter die de communicatie met de buitenwereld en het internet mogelijk maakt. Met het MQTT-protocol is bidirectionele communicatie mogelijk tussen de verbonden apparaten en mobiele of webapplicaties. Het protocol biedt een eenvoudige methodiek om berichten uit te wisselen via een publish-subscribe model waarbij geen directe connectie bestaat tussen de verzender van de berichten en de ontvangers ervan. Dit maakt het geschikt voor meerdere Internet of Things datastromen te verbinden, zoals sensoren op laag vermogen, mobiele apparaten, embedded computers of microcontrollers. MQTT is een open en gestandaardiseerd protocol ontwikkeld door de technische commissie van het non-profit consortium OASIS. Dit maakt het gemakkelijk te gebruiken voor een grote verscheidenheid aan IoT-apparaten, platforms en besturingssystemen.

8.3 APACHE KAFKA

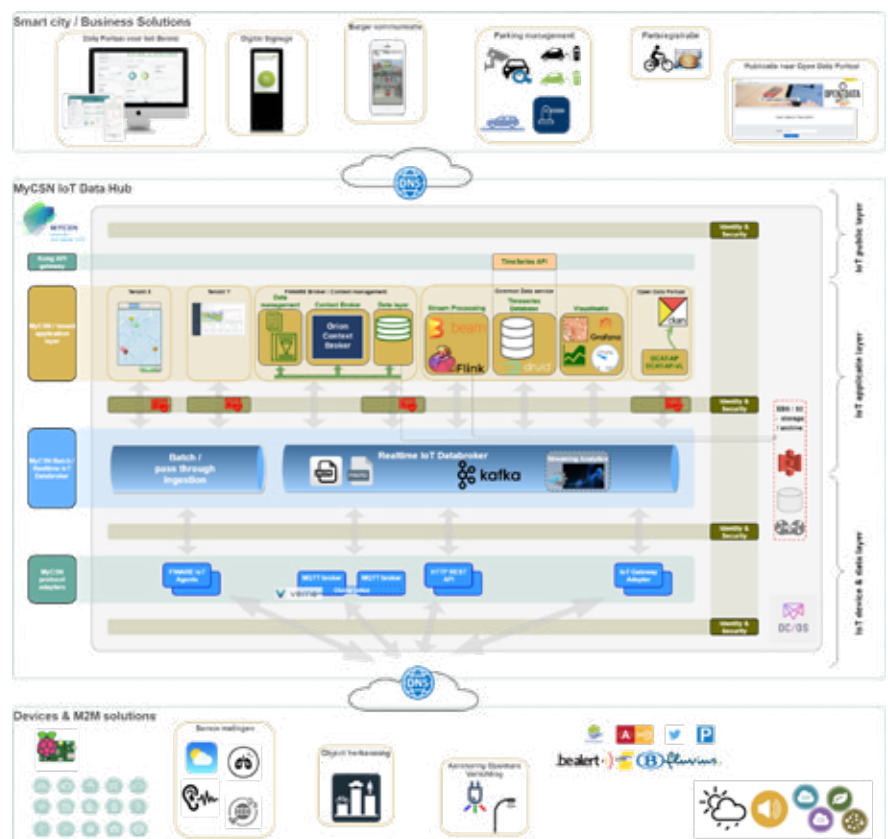
Om de kracht en de flexibele capaciteit van het MyCSN dataplatform te verhogen, hebben we het aangevuld met Apache Kafka als databroker voor de interne communicatie. Een tiental jaar geleden is Apache Kafka door LinkedIn beschikbaar gesteld als openbronsoftware en intussen geëvolueerd naar een gedistribueerd streamingplatform dat miljarden boodschappen per dag kan verwerken. Het biedt data pipelines met zeer lage vertraging, hoge verwerkingscapaciteit en fouttolerantie. Dat stelt het in staat om real-time datastromen te verwerken.

Alle datastromen worden bijgehouden in de vorm van topics, zoals het geval is bij wel meer publish-subscribe systemen. Alle partijen die geïnteresseerd zijn in een datastroom of een topic, kunnen zich erop abonneren om de data te lezen of te versturen. Apache Kafka is een gedistribueerd systeem, wat inhoudt dat de topics verdeeld en gedupliceerd over verschillende nodes. Op die manier geeft Apache Kafka de garantie dat de data beschikbaar blijft bij uitval van een node. Een ander voor-

deel van Apache Kafka is dat elke partitie binnen een topic wordt behandeld als een logboek, een geordende reeks boodschappen. Zo kan men bijhouden wie welke boodschap heeft gestuurd naar de Apache Kafka broker en bezorgt het systeem een methodiek voor back-up en restore.

8.4 APACHE SPARK

Apache Spark is een enorm krachtige engine voor de verwerking van grote gegevenshoeveelheden, gebouwd voor snelheid, gebruiksgemak en geavanceerde analyses. Oorspronkelijk ontwikkeld op de universiteit van Berkeley, bracht Apache Spark een revolutie teweeg in de wereld van Big Data. De openbronsoftware is sneller en krachtiger dan vele andere commerciële oplossingen. Apache Spark bevat een stapel libraries zoals SQL, Data Frames, MLib (Machine Learning Libraries), GraphX, Spark-streaming en heeft veel plug-ins om data rechtstreeks te lezen en te schrijven naar data lakes, zoals dat van Amazon S3. Het leest en schrijft zo ook rechtstreeks naar Apache Kafka, de streaming service van het MyCSN dataplatform.



9. LOGISCHE ARCHITECTUUR

De bouwstenen van het MyCSN dataplatform vormen samen een logische architectuur. De basis wordt gevormd door een backend internetlaag, vervolgens een platform- of applicatielaag en tot slot een front-end internetlaag. De lagen hebben elk hun afgeijnde karakteristieken en functies:

- de backend internetlaag dient om de datastromen uit alle sensoren en meetapparatuur te laten samenkomen alvorens ze op een uniforme manier bij de gebruiker van het MyCSN dataplatform arriveert;
- de platform- of applicatielaag zorgt voor de verwerking van de data. De datavelden die van toepassing zijn worden geselecteerd, de gegevens indien nodig anoniem gemaakt of gevalideerd. Verdere verwerking maakt analyses en correlaties om meer inzicht te creëren. Door de data op deze manier op te waarderen, wordt het interessanter om deze te delen met de andere gebruikers van het MyCSN dataplatform en daarbuiten. Het is ook mogelijk om deze gegevens te commercialiseren;
- de front-end internetlaag is waar de data, inzichten en visualisaties bruikbaar worden voor de eindgebruiker. Denk bijvoorbeeld aan grafieken over parkeerbezetting, opvolging van sluikstortmeldingen, allerlei visuele analyses of het uitbouwen van smart city-applicaties.



10. KERNFUNCTIES

Het MyCSN dataplatform is ontworpen en gebouwd volgens een aantal basisprincipes. Dat weerspiegelt zich in de functies die we aan steden en gemeenten aanbieden om op een duurzame, toegankelijke en slimme manier met data om te gaan:

- een open platform, gebaseerd op open-bronsoftware waarbij elke gebruiker een specifiek gedefinieerde toegang heeft op een deel van alle gegevens en toepassingen (multi-tenant) om co-creatie en samenwerking mogelijk te maken;
- een modulair platform om nieuwe componenten en diensten eenvoudig te kunnen toevoegen;
- een toolbox met een verscheidenheid aan onderliggende componenten op maat van de klant of het project;
- het gebruik van best practices en bewezen technologie;
- de mogelijkheid bieden om geleidelijk wijzigingen aan te brengen in het ontwerp het van platform en functies toe te voegen op basis van gebruikersfeedback;
- toelaten om te valideren en uit te breiden, zodat we blijven leren uit proefprojecten.

Het MyCSN dataplatform biedt verschillende bouwstenen (een IoT-toolbox) om allerlei smart city-projecten op te zetten. Ook real-time dashboarding, rapportering van tijdreeksen en geavanceerde analyses zijn voorzien. Tot slot zijn er gebruiksvriendelijke workflow-modules voor prototyping en snel testen van IoT-projecten.



10.1 CONNECTIVITEIT BEHEREN

Het MyCSN platform maakt het mogelijk om datastromen te verbinden en te verzamelen vanuit verschillende sensoren, applicaties en andere contextuele data (denk bijvoorbeeld aan weersverwachtingen). Om deze data binnen te halen, gebruiken we vooral het MQTT-protocol. MQTT is een zeer compact protocol en dat ook ondersteund wordt door de kleinste meet- en bewakingsapparatuur. Wanneer nieuwe protocollen op de markt komen, kunnen we extra protocoladapters voorzien en mee inschakelen.

10.2 INZICHTEN GENEREREN

Er zijn op het MyCSN dataplatform verschillende mogelijkheden om data op te slaan en om inzichten te genereren zoals bijvoorbeeld door datastromen te aggregeren en correlaties te leggen. Het MyCSN dataplatform heeft daartoe het volgende:

- een relationeel datawarehouse (PostgreSQL) voor het opslaan van gestructureerde data en de analyseresultaten voor rapportering in BI-tools;
- tijdreeksen (Apache Druid database) om historische sensorwaarden en aggregaten op te slaan aangevuld met gemiddelden, minimum- en maximumberekeningen;
- een data lake voor de opslag van onbewerkte data als historisch archief voor geavanceerde analyses;
- de mogelijkheid tot real-time rapportage met dashboards van bijvoorbeeld tijdreeksen (met gebruik van Apache Superset, Grafana, Freeboard).

10.3 GEAVANCEERDE ANALYSES

Op basis van data uit verschillende bronnen kan men via het MyCSN dataplatform dashboards aanmaken. Dat kan ook in real-time en de data kunnen vooraf nog bijkomend ge-

analyseerd worden. Dit gebeurt met Apache Spark in combinatie met Apache Zeppelin of Apache Jupyter Notebook. Deze notebooksoftware maakt data-analyse voor data scientists en anderen toegankelijk via een online gebruikersportaal waar documentatie, analytische logica en de resultaten beschikbaar zijn in één omgeving.

10.4 IOT-OPLOSSINGEN OP MAAT

MyCSN heeft een IoT-toolbox om snel projecten of proefopstellingen te bouwen:

- het visueel programmeren van workflows gebeurt met rules engines zoals Node-RED of Apache Nifi.
- API-integratie voor mobiele apps en webapplicaties gebeurt op basis van Kong Gateway.
- in een selfservice portaal is het mogelijk om toepassingen van derden onder te brengen.

Sommige bouwblokken zijn al op maat gemaakt of kunnen verder ontwikkeld worden om te voldoen aan de vereisten van het IoT-project.

10.5 CLOUD-ONAFHANKELIJK

Het MyCSN dataplatform is momenteel opgezet in de cloud van Amazon waarvan we alleen enkele basiscomponenten gebruiken, zoals bijvoorbeeld EC2 en S3. Het ontwerp van het MyCSN platform en de gebruikte technologie (zoals DC/OS) zijn echter leveranciersonafhankelijk. Het betekent dat het mogelijk is om het platform over te plaatsen naar een andere leverancier (Microsoft, Google...). Dit vereist een beperkt aantal aanpassingen aan code van de automatisatielaag om andere API's aan te spreken. MyCSN is dus niet afhankelijk van een bepaalde leverancier en data blijven altijd in het bezit van de eigenaar ervan.

10.6 TOESTELLEN VERBINDEN

Er zijn meerdere protocollen mogelijk om gegevens uit databronnen te verzamelen op het MyCSN dataplatform. De manier waarop gegevens op het platform gebracht worden, hangt af van de behoeften binnen het project. Het MyCSN dataplatform is uitgebreid getest en kan sensoren en andere toestellen verbinden met behulp van LPWAN (LoRa en Sigfox), IP (bijvoorbeeld Raspberry Pi), mobiele IoT-netwerken (via 3G, 4G en LTE op iOS en Android) en andere databronnen via API's.

10.7 GEGEVENS- BESCHERMING

Het MyCSN dataplatform kan alle vereisten van de Algemene verordening gegevensbescherming (GDPR) correct verwerken. We kunnen beroep doen op de diensten van Cranium om de implementatie van de GDPR te begeleiden of te valideren.

10.8 NIET-FUNCTIO- NELE VEREISTEN

Naast de functionele vereisten van een databroker zijn er ook enkele niet-functionele eisen waaraan het platform voldoet, namelijk:

10.9 CLUSTER RESOURCE MANAGER

De basislaag van het MyCSN dataplatform is een op maat gemaakte DC/OS omgeving met verschillende modules die de belangrijkste functies bieden voor het databroker platform. DC/OS is een cluster- en resource management systeem dat toelaat om meerdere container workloads te verwerken. DC/OS beheert alle bronnen van een node binnen de cluster. Dit zijn het

aantal CPU's, de hoeveelheid RAM-geheugen of de opslag. Wanneer een node van de cluster bijvoorbeeld tegen 80% van zijn capaciteit draait, zorgt DC/OS en zijn orchestrator ervoor dat nieuwe containers of applicaties naast de node worden verwerkt. Dit dient om de belasting over de cluster te spreiden en ervoor te zorgen dat er genoeg capaciteit vrij is om bijvoorbeeld het falen van een node op te vangen.

10.10 AUTOMATISATIE

Om een mogelijke onbeschikbaarheid te herstellen, extra nodes of capaciteit toe te voegen, zijn alle implementatie- en configuratietaken zoveel mogelijk geautomatiseerd. Het betekent dat we met een druk op de knop een volledig functionele versie van het MyCSN dataplatform kunnen opstarten. In het geval van een grote onderbreking, bijvoorbeeld een storing in het datacenter, is het mogelijk het volledige dataplatform op maximaal enkele uren te herstellen, in plaats van dagen of weken. Daarnaast kunnen we geautomatiseerd extra tools toevoegen, zoals voor extra gebruikers of een toename van opslag- of verwerkingscapaciteit.

10.11 CONTAINER ORCHESTRATION

Marathon is de container orchestrator van het DC/OS besturingssysteem. De orchestrator bepaalt op welke node een applicatie of container zal draaien en maakt deze ook bereikbaar voor de rest van de cluster en eventueel voor de buitenwereld of het internet. Marathon zorgt virtuele IP-routing en kent deze adressen plus poorten toe aan de containers of applicaties.

Docker is de gebruikte containertemplate. Wanneer een applicatie wordt ontwikkeld en beschreven volgens de standaarden van Docker zal deze met al zijn afhankelijkheden op meerdere hosting-omgevingen kunnen draaien. Dit maakt de Docker-technologie zo goed als de standaard in de containerized IT-omgevingen.

10.12 MULTI-TENANT

Het MyCSN dataplatform is opgezet als een multi-tenant omgeving. Het betekent dat meerdere tenants of gebruikers dezelfde bronnen, tools en functies delen. Dat gaat over CPU, RAM, opslagruimte en netwerk, maar ook enkele common service tools van het dataplatform. Uiteraard zijn op alle lagen van het platform alle regels in acht genomen om een strikte scheiding tussen de gebruikers te garanderen, meestal op basis Role-based access control (RBAC). DC/OS Calico zorgt er ook voor dat elke gebruiker van het MyCSN dataplatform een eigen netwerk of IP-range krijgt op basis van software defined network-technieken zodat elke gebruiker van elkaar is geïsoleerd.

10.13 COMMON SERVICES

De softwarediensten die MyCSN inzet voor alle gebruikers noemen we de common services. Dankzij de multi-tenant opzet is het niet nodig om elke service per gebruiker te doen werken. De common service laag is dus kostenefficiënter, maar vereist wel dat er ook multi-tenancy is op applicatieniveau. De voornaamste common services zijn DC/OS, MQTT message broker, Kafka data streaming service, Apache Druid timeseries database en Kong Gateway. We vatten samen welke functies deze tools bieden:

- DC/OS laat toe om elke tenant apart te configureren zodat er vrije keuze is welke services gemeenschappelijk of per tenant aangeboden worden. In de platformlaag van het MyCSN dataplatform zorgt DC/OS ervoor dat er meerdere platforminstanties kunnen draaien. Het biedt extra mogelijkheden om het MyCSN dataplatform uit te breiden op vraag van klanten of naar andere regio's. Deze flexibiliteit is uniek en bestaat nauwelijks in de andere IoT-datahubs.
- VerneMQ is de messagebroker die volgens het MQTT-protocol werkt. Zoals eerder beschreven, is het MQTT-protocol uiterst geschikt om datastromen binnen te halen vanuit allerlei sensoren en IoT-toestellen.

- Apache Druid is de al beschreven database voor tijdreeksen.
- Kong Gateway zorgt ervoor dat interne datastromen van klanten aan de buitenwereld ter beschikking kunnen gesteld worden. Op deze manier kunnen de gebruikers of derden deze gegevens gebruiken in hun applicaties. Kong Gateway beheert en controleert de rechten op alle gegevensstromen. Andere functies van Kong Gateway zijn band-breedtebeperking, traffic control, plus monitoring, logging en statistieken rond dataverbruik.

10.14 AANPASSINGEN AAN ELKE TENANT

Elke tenant heeft een eigen omgeving in het MyCSN data-platform en kan die uitbreiden met applicaties en tools die geen multi-tenancy ondersteunen. Die tools zijn dan volledig geïsoleerd en niet bereikbaar voor de andere gebruikers van het MyCSN dataplatform. Gebruikers kunnen kiezen uit verschillende beschikbare tools, afhankelijk van het project. Het MyCSN dataplatform gebruiken kan variëren van een eenvoudige setup voor gegevensvisualisatie tot geavanceerde data science op basis van Apache Spark. Zo is er voor elke gebruiker personalisatie met uitgebreide mogelijkheden om projecten te ontwikkelen en te ondersteunen.

10.15 OPEN PLATFORM

Op het MyCSN dataplatform kunnen gebruikers ook eigen applicaties hosten. Daarnaast kunnen gebruikers applicaties en software van anderen hosten als de bronsoftware in een Docker container-template verpakt is. Het MyCSN dataplatform is dus open en gebruikers kunnen software uploaden naar de MyCSN container registry (daarvoor gebruiken we Harbor) of hosten in de omgeving van de gebruiker. Op deze manier kan men nieuwe services en applicaties gebruiken via het MyCSN dataplatform in plaats van de gegevens te moeten exporteren naar externe systemen. Zo blijft de data veilig en gecentraliseerd.

10.16 SECURITY

Beveiliging is voorzien op verschillende niveaus van het MyCSN dataplatform.

Sensoren en andere toestellen

De sensoren en andere toestellen die gegevens verzamelen, communiceren met de VerneMQ MQTT-broker via beveiligde en versleutelde Transport Layer Security (TLS)-verbindingen. Dit gebeurt met een geldig mycsn.be certificaat om de authenticiteit van het platform te valideren. Voorts moeten alle toestellen zich authentifieren met behulp van bepaalde tokens per toestel of sensor die na een bepaalde tijdsspanne of bij verdachte omstandigheden ingetrokken kunnen worden. Met een token kunnen de sensoren en toestellen alleen maar participeren op de topics die aan hen zijn toegekend. Dit wordt gecontroleerd met Access control lists (ACL's). Aanvullende beveiligingsmaatregelen zoals bandbreedtebeperking, DDOS-bescherming of anomaliedetectie kunnen worden toegevoegd op basis van de vereisten van het project.

API's en gebruikers

Wanneer gebruikers of API's een verbinding maken, gebruiken ze een HTTPS Transport Layer Security (TLS)-verbindingen, ook met een geldig mycsn.be certificaat om de authenticiteit van het platform te valideren. De authenticatie van de gebruikers gebeurt op basis van een gebruikersnaam en wachtwoord of met API-access keys. Ook hier is alle communicatie versleuteld, inclusief de authenticatie. De loggegevens van gebruikers of de API keys kunnen op elk moment worden ingetrokken.

Security rond gegevensopslag

Er zijn verschillende manieren waarop gegevens op het MyCSN dataplatform beveiligd zijn, afhankelijk van het type opslag:

- raw object storage: elke gebruiker heeft unieke toegangsleutels om data naar de Amazon S3 object storage weg te schrijven. Op verzoek kunnen aanvullende Identity and Access Management (IAM) gebruikers en rollen worden gecreëerd

om meer verfijnde toegangsrechten tot subsets van de gegevens te bieden. Bovendien kan een volledige encryptie van de data worden ingeschakeld;

- timeseries: de gemeenschappelijke Apache Druid timeseries database is ontworpen als een multi-tenant tool en vereist authenticatie voor elke applicatie die data wenst weg te schrijven. Ook de data van de verschillende tenants is volledig van elkaar gescheiden;
- PostgreSQL: het gemeenschappelijke datawarehouse is multi-tenant qua design en vereist authenticatie van elke applicatie met gebruik van inloggegevens per tenant. Het isoleert tenantgegevens in de database;
- lokale opslag: sommige containerservices van tenants vereisen een apart volume om data weg te schrijven. Ook hier zorgt MyCSN ervoor dat deze volumes enkel toegankelijk zijn binnen de tenant en kan er, indien gewenst, volledige diskencryptie opgezet worden.



10.17 OPEN STANDAARDEN

Gegevens zijn een snelgroeiende bron aan mogelijkheden. Bijna 90% van de gegevens in de wereld werden de afgelopen twee jaar gegenereerd. Een recent onderzoek van IDC en DellEMC voorspelt dat digitale gegevens tegen 2025 tien keer exponentieel zullen groeien tot 163 zettabytes (miljard gigabytes). We hebben echter nog een lange weg te gaan om de kracht van gegevens te benutten. Uit hetzelfde rapport blijkt dat slechts 1% van de gegenereerde gegevens effectief wordt gebruikt. Een van de belangrijkste belemmeringen daarvoor is de inconsistentie in de gegevens. Dat blokkeert het gemak waarmee we data benutten.

Ook lokale besturen beschikken over een schat aan gegevensbronnen. Kaartverkoop op het openbaar vervoer, lokale belastinginformatie, politierapporten, weerstations, afvalverwerkingsinstallaties en verkeersinformatie zijn slechts een paar voorbeelden. Ook uit video- en fotobestanden kan men nu gegevens verzamelen dankzij artificiële intelligentie. Met patroonherkenning kan men bijvoorbeeld verkeersopstoppingen verbeteren of criminaliteit terugdringen. Stel u voor wat we kunnen waarmaken als we in staat zijn om meer gegevens effectief in te zetten. We moeten verder gaan dan de 1% door schonere gegevens te creëren en deze te gebruiken om beslissingen te sturen. Het obstakel waar ook steden en gemeenten tegen botsen, zijn inconsistente gegevens die naast elkaar bestaan. Om slimme gemeenten, steden en regio's te ontwikkelen, moeten we verbondenheid creëren via geaggregeerde datasets.

Gestandaardiseerde datamodellen zijn essentieel om een eengemaakt aanbod van oplossingen te creëren die met elkaar samenwerken. Zo kunnen we domeinen overschrijden en kennis delen. Niet alleen voor smart cities, maar ook in de voedselvoorziening, voor nutsbedrijven en in andere sectoren is dit een waardevol doel. Standaardmodellen vormen de technische

grondslag voor open en duurzame innovatie. Om die reden lanceerden de non-profitorganisatie TM Forum en de FIWARE Foundation het Front-runner Smart Cities Program, dat streeft naar het harmoniseren van datamodellen in smart cities.

Op het MyCSN dataplatform worden alle data omgezet naar standaarden en kan iedere gebruiker op een veilige manier data delen en combineren.

10.18 TOEGEPASTE STANDAARDEN

MyCSN streeft ernaar om de datamodellen en referentie-infrastructuren te volgen die vooropgesteld worden door Vlaanderen, federaal België en de EU. Om alvast up-to-date te blijven, volgt MyCSN een hele reeks organisaties, zoals onder meer:

- Open & Agile Smart Cities (OASC), met OASC Context Information Management MIM, OASC Data Models MIM, OASC Ecosystem Transaction Management MIM, Security MIM en Storage MIM
- SynchroniCity, met NGSI-LD
- FIWARE
- Open Standaarden voor Linkende Organisaties (OSLO)
- Open City of Things
- Vlaamse Open City Architectuur (VLOCA), MyCSN is lid van het co-creatieproject
- COOCK-project Open Stad, MyCSN is een van de zeven pionierbedrijven in dit project
- Open Data Charter

MyCSN is actief betrokken bij een aantal van deze organisaties om de gebruikers van het MyCSN dataplatform te begeleiden in de keuze voor de meest geschikte standaard voor hun project. Het MyCSN dataplatform is volledig op standaarden gebaseerd en we voorzien diensten om per project extra standaarden te implementeren.



Contacteer ons

<https://mycsn.be/contact/>

MyCSN

Franklin Rooseveltplaats 12 box 14

2060 Antwerpen

E-mail: info@mycsn.be

Algemene vragen: +32 14 39 70 70

Technische vragen en support: +32 14 39 70 71

www.mycsn.be