PoC Federatieve event distributie

basic data infrastructure

Topsector Logistiek

Colofon

PoC Federatieve event distributie

Auteur
H. Alblas

Januari 2024

basic data infrastructure

Topsector Logistiek

Samenvatting

BDI: Afsprakenstelsel

Het afsprakenstelsel BDI geeft partijen de goede balans tussen beheersing (wie mag wat aan data inzien of gebruiken?) en efficiëntie (geautomatiseerd afhandelen), door:

- Digitaal vertrouwen
  - controle over eigen data door 'data bekijken bij de bron';
  - zekerheid over identiteit;
  - hoge security eisen;
  - contractuele basis voor computers die zonder mensenhanden met elkaar communiceren.

- Afspraken over elkaars 'taal' en begrippen zodat automatisch verwerken mogelijk wordt
  - semantiek.

- Tijdig automatisch gewaarschuwd worden ('events')
  - vanzelf een signaal krijgen als het nodig is maakt het werken zeer efficiënt;
  - als in de werkelijkheid iets gebeurd of veranderd is wat relevant kan zijn.

Het laatste aspect, 'event driven' communiceren is een andere manier om de logistiek van informatie-uitwisseling tussen IT-systemen van bedrijven en overheidsdiensten te structureren.

In plaats van het sturen van berichten ('fire and forget') krijgen alle betrokken partijen een signaal dat er iets relevants gebeurd is ('publish event'). Dat event bevat meta-data, en een link naar de bron van de data.

De vraag was of bestaande middleware voor het invullen van de publish/subscribe functie geschikt is, en in corporate omgevingen over firewalls heen makkelijk toegepast kan worden.
De deelvragen waren:

1 Wat zijn de functionele kerneisen in deze context voor het uitwisselen van events en streams?
2 Hoe scoren bestaande opensource event en stream middleware op deze functionele kerneisen?
3 Hoe wordt een event/stream gemodelleerd?
4 Hoe werkt het distribueren van events/streams met open source technologie in een PoC-opstelling samen met iSHARE en de DIAC (Distributed Information Access Control) request-reply functionaliteit?

De bevindingen zijn:

- Het is mogelijk om met bestaande middleware een publish/subscribe event distributie op te zetten die geschikt is voor de BDI.
- Het is mogelijk om dat over normale openstaande poorten in corporate firewalls te laten lopen, wat de deployment barrière fors vermindert.
- De websocket extensie van Apache pulsar was nog niet goed genoeg uitgewerkt om secure met tokens om te gaan. Hierop is een change request op de websocket extensie van Apache pulsar opgestart, waarmee een websocket verbinding wordt afgesloten wanneer een token verlopen is. Met dit Pulsar Improvement Process wordt de aanpassing onderdeel van het officiële Apache Pulsar project.
- Apache Pulsar bevat een component voor het archiveren van pulsen. Dit kan onderzocht worden om mogelijk de onweerlegbaarheid requirement verder kracht bij te zetten.

Inhoudsopgave

1 Introductie

1.1 BDI

BDI: Afsprakenstelsel voor beheerst data delen voor zakelijke ecosystemen (Controlled Data Sharing for Professional Ecosystems). BDI (www.bdinetwork.org) is gericht op het makkelijk, effectief en beheerst data delen in zakelijke ecosystemen ('federatieve data spaces').

In een hoogontwikkelde economie is er sprake van een hoge mate van specialisatie van bedrijven. Die moeten in hun netwerk van relaties hoogfrequent communiceren met elkaar om het resultaat te halen, zoals het soepel laten lopen van intercontinentale supply chains. En ze moeten de overheidsdiensten die toezicht houden op tijd van informatie voorzien.

Het afsprakenstelsel BDI geeft partijen de goede balans tussen beheersing (wie mag wat aan data inzien of gebruiken?) en efficiëntie (geautomatiseerd afhandelen), door:

- Digitaal vertrouwen
  - controle over eigen data door 'data bekijken bij de bron';
  - zekerheid over identiteit;
  - hoge security eisen;
  - contractuele basis voor computers die zonder mensenhanden met elkaar communiceren.

- Afspraken over elkaars 'taal' en begrippen zodat automatisch verwerken mogelijk wordt
  - semantiek.

- Tijdig automatisch gewaarschuwd worden ('events')
  - vanzelf een signaal krijgen als het nodig is maakt het werken zeer efficiënt;
  - as in de werkelijkheid iets gebeurd of veranderd is wat relevant kan zijn.

Het laatste aspect, 'event driven' communiceren is een andere manier om de logistiek van informatie-uitwisseling tussen IT-systemen van bedrijven en overheidsdiensten te structureren.

In plaats van het sturen van berichten ('fire and forget') krijgen alle betrokken partijen een signaal dat er iets relevants gebeurd is ('publish event'). Dat event bevat meta-data, en een link naar de bron van de data. Iedere partij die het event ontvangt evalueert dat, en kan desgewenst zich melden bij de bron. Op basis van de digitale identiteit van die partij die zich meldt kan de bron:

- Verifieren of deze specitieke partij betrokken is, en in welke rol.
- Bepalen welke data deze partij dan mag ontvangen.

Het iSHARE Trust Framework is hier de basis voor.

Deze opzet is veel effectiever en efficienter, en geeft de data eigenaar veel meer controle.

Deze opzet kan in principe gebruik maken van bestaande systemen om events te distribueren op basis van publish/subscribe mechanismes. De vraag is of de bestaande open source oplossingen geschikt zijn voor grootschalige professionele toepassing. Een van de kernvragen daarbij is of voorkomen kan worden dat er aanpassingen nodig zijn in corporate firewalls.

2 Vraagstelling

2.1 Hoofdvraag

Bestaat er geschikte open source middleware voor het federatief distribueren van events en streams middels publish & subscribe?

2.2 Deelvragen

1 Wat zijn de functionele kerneisen in deze context voor het uitwisselen van events en streams?
2 Hoe scoren bestaande opensource event en stream middleware op deze functionele kerneisen?
3 Hoe wordt een event/stream gemodelleerd?
4 Hoe werkt het distribueren van events/streams met open source technologie in een PoC-opstelling samen met iSHARE en de DIAC (Distributed Information Access Control) request-reply functionaliteit?

3 Uitwerking

3.1 Functionele kerneisen

Wat zijn de functionele kerneisen voor het uitwisselen van events en streams?

Bij gebruik maken van de BDI moeten organisaties de mogelijkheid hebben om te kunnen subscriben op nieuwe relevante data bij data eigenaren.

Hiervoor zijn bestaande en gevestigde Open source softwarepakketten vergeleken. De vergelijkings-criteria zijn opgesteld vanuit het uitganspunt dat 'direct' gestart moet kunnen worden met de software in een productie omgeving.

- **Ontwikkelstatus:** De software release moet stabiel zijn, niet in Beta.
- **OSS licentie:** De software moet opensource en herbruikbaar zijn.
- **Pub/Sub:** Het publish/subscribe patroon moet ondersteund worden.
- **Websocket support:** Websockets ondersteuning voor communicatie over standaard poorten.
- **Guaranteed delivery:** Guaranteed delivery moet ondersteund zijn.
- **Client footprint:** Client moet simpel en light weight zijn.
- **Authentication support:** iSHARE moet toe te voegen zijn als autorisatiemechanisme.
- **Client language:** De client moet in meerdere programmeertalen beschikbaar zijn.

3.2 Score

Hoe scoren bestaande opensource event en stream middleware op deze functionele kerneisen?

De volgende Opensource event en stream middleware zijn vergeleken:

- RabbitMQ
- ActiveMQ Artemis
- SwiftMQ
- Qpid
- AMQ Broker
- Mosquitto
- Amlen Broker
- NanoMQ
- EMQX
- Pulsar
- Cyclone DDS
- CATIA Magic
- CoreDX DDS
- Fast DDS
- HiveMQ
- Storm
- Kafka
- ZeroMQ

Het eenvoudig kunnen uitbreiden van de authenticatie met iSHARE is een harde eis, daarop vielen veel opties af. Apache Pulsar kwam als beste uit de vergelijking. In de bijlage staat het gehele overzicht.

3.3 Modellering

Hoe wordt een event/stream gemodelleerd?

Events worden gemodelleerd als zijnde een puls, waarin enkel kenbaar wordt gemaakt dat data in het bronsysteem bijgewerkt is. De puls bevat voldoende informatie om geautomatiseerd de data op te halen. Hierbij wordt bij het ophalen de autorisatie conform de iSHARE protocollen gecontroleerd. Dit maakt het bijvoorbeeld mogelijk om pulsen naar een breed publiek door te sturen, want enkel de geautoriseerde partijen kunnen de daadwerkelijke data ophalen.

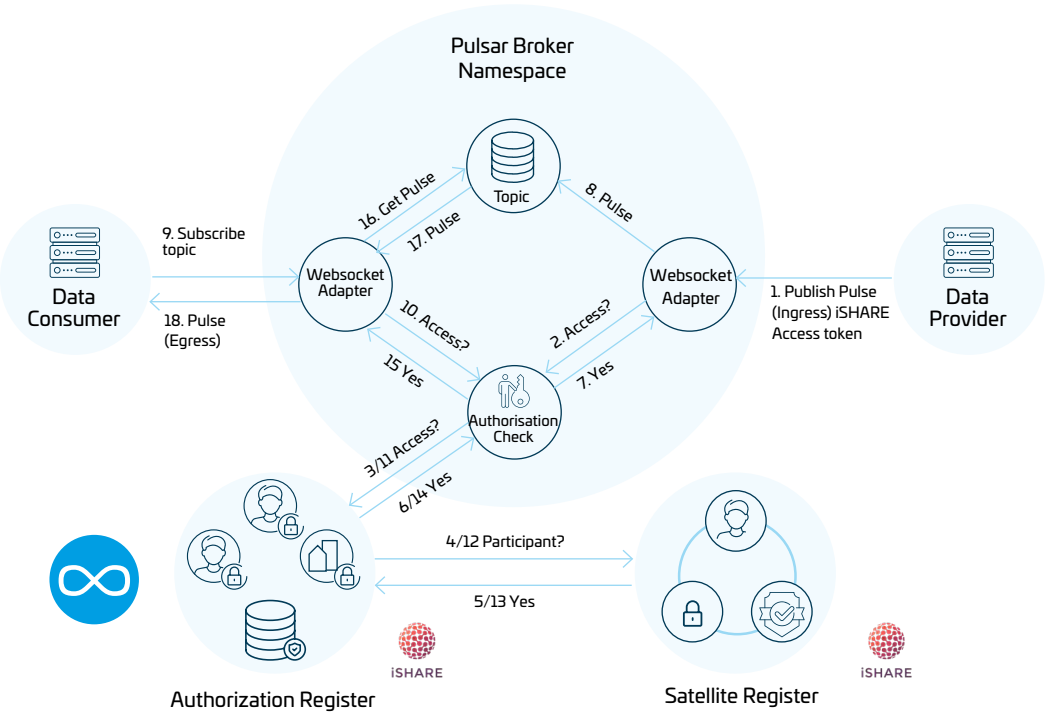
De puls moet zo opgezet worden dat verschillende technologieën gebruikt kunnen worden voor het ophalen van de data. Hierin komt een tegenstelling naar voren tussen enerzijds een volledig uit gespecificeerd model, en anderzijds een formattering die direct door een client gebruikt kan worden. Er is kozen om het model volledig uit te specificeren, waarbij de endpoints en data identifiers los van elkaar gemodelleerd zijn.

Het datamodel, en een voorbeeld hierbij zijn te vinden op de Github van Topsector Logistiek.

3.4 Distributie

Hoe werkt het distribueren van events/streams met open source technologie in een PoC opstelling samen met iSHARE en de DIAC request-reply functionaliteit?

Figuur
Werking distributie



1. De Data provider maakt via Websockets met een iSHARE token verbinding met de Pulsar Broker.
2. De Broker voert een autorisatie check uit om te controleren of het iSHARE token publish rechten heeft op gevraagde topic.
3. De autorisatie plugin maakt verbinding met het externe autorisatie register, en checkt de autorisatie.
4. Het autorisatie register controleert of de organisatie genoemd in het iSHARE token geregistreerd is in de satelliet (register van onboarded deelnemers aan de BDI).
5. De satelliet bevestigt dat de organisatie geregistreerd is.
6. Het autorisatie register bevestigt dat de organisatie uit het iSHARE token gemachtigd is om te publiceren op het topic.
7. De autorisatie plugin bevestigt dat data provider verbinding mag maken, en daarmee mag publiceren.
8. De puls van dataprovider wordt gepersisteerd op de broker (en uitgestuurd naar eventuele subscribers).
9. De dataconsumer maakt via Websockets met een iSHARE token verbinding met de Pulsar Broker.
10. De Broker voert een autorisatie check uit om te controleren of het iSHARE token subscribe rechten heeft op gevraagde topic.
11. De autorisatie plugin maakt verbinding met het externe autorisatie register, en checkt de autorisatie.
12. Het autorisatie register controleert of de organisatie genoemd in het iSHARE token geregistreerd is in de satelliet.
13. De satelliet bevestigt dat de organisatie geregistreerd is.
14. Het autorisatie register bevestigt dat de organisatie uit het iSHARE token gemachtigd is om te subscriben op het topic.
15. De autorisatie plugin bevestigt dat data consumer verbinding mag maken, en daarmee berichten van het gespecificeerde topic mag ontvangen.
16. De broker controleert of er historische berichten op het topic opgeslagen staan.
17. Er is een opgeslagen bericht.
18. Het bericht wordt via de websocket verbinding uitgestuurd naar de data consumer.

3.5 Test en gebruik van standaard poorten in firewalls

Eventuitwisseling kan in corporate IT omgevingen uitdagend zijn door strak afgestelde firewalls. Hierom is een test uitgevoerd om de event uitwisseling via de poorten voor standaard web verkeer te laten verlopen, namelijk port 443. Hiervoor is gebruik gemaakt van de websocket extensie van Apache Pulsar. Het nadeel van de websocket extensie is dat die minder functies bevat dan het native pulsar endpoint, welke gebruik maakt van port 6651. Een van de functies is voor het publiceren van dit rapport als essentieel beschouwd, en aangedragen als toevoeging aan het opensource project.

In de test opzet was de Data Consumer aanwezig binnen het corporate netwerk van CGI, de Data Provider binnen het netwerk van Connekt, en de Broker binnen Azure. Aan de firewalls van CGI en Connekt zijn geen aanpassingen gedaan. Op de Azure instance staan de standaard HTTP en HTTPS porten (80, 443) open voor inkomend verkeer. Met het gebruik van de websocket extensie was het mogelijk om met de standaard firewall instellingen Pulses te versturen en te ontvangen.

4 Bevindingen

- Het is mogelijk om met bestaande middleware een publish/subscribe event distributie op te zetten die geschikt is voor de BDI.
- Het is mogelijk om dat over normale openstaande poorten in corporate firewalls te laten lopen, wat de deployment barrière fors vermindert.
- De websocket extensie van Apache pulsar was nog niet goed genoeg uitgewerkt om secure met tokens om te gaan. Hierop is een change request op de websocket extensie van Apache pulsar opgestart, waarmee een websocket verbinding wordt afgesloten wanneer een token verlopen is. Met dit Pulsar Improvement Process wordt de aanpassing onderdeel van het officiële Apache Pulsar project.
- Apache Pulsar bevat een component voor het archiveren van pulsen. Dit kan onderzocht worden om mogelijk de onweerlegbaarheid requirement verder kracht bij te zetten.
- Het event model is techniek onafhankelijk gemodelleerd. Dit geeft echter extra complexiteit bij het gebruik, omdat er meer geparsed moet worden. In de volgende versie kan het model bijgewerkt worden, zodat de complexiteit kan worden gereduceerd.
- Voor een soepele demo is er een betere integratie nodig aan de bron kant. Wanneer er data aangepast wordt moet er automatisch een juiste puls uitgestuurd worden: die integratie is nu buiten beschouwing gelaten.

Appendix A

Opensource middleware score matrix

PRODUCT	STAN-DAARD	LEVERAN-CIER	STATUS	OSS LICENTIE	BETAALDE ENTERPRISE VERSIE	PUB/ SUB	STREA-MING	WEB SOCKET	GUARAN-TEED SUPPORT	CLIENT FOOTPRINT DELIVERY	SECURITY	AUTHEN-TICATION SUPPORT	CLIENT SERVER LANGUAGE
RABBITMQ	AMQP 0.9	VMWare	GA	MPL 2.0	Ja	Ja	Nee	Nee	Ja	H+	SSL/TLS	SASL	Java, .NET, Python & more
ACTIVEMQ ARTEMIS	AMQP 1.0	Apache Foundation	GA	AL 2.0	Nee	Ja	Nee	Nee	Ja	H+	SSL/TLS	SASL	Java, .NET, Python & more
SWIFTMQ	AMQP 1.0	Apache Foundation	GA	AL 2.0	Ja	Ja	Nee	Nee	Ja	M	SSL/TLS	SASL	Java
QPID	AMQP 1.0	Apache Foundation	GA	AL 2.0	Nee	Ja	Nee	Ja	Ja	M	SSL/TLS	SASL	Java, .NET, Python & more
AMQ BROKER	AMQP 1.0	Redhat	GA	-	Ja	Ja	Nee	Nee	Ja	L	SSL/TLS	SASL	Java, .NET, Python & more
CYCLONE DDS	DDS	Eclipse Foun-dation, Zetta-scale, Adlink	GA	EPL 2.0	Nee	Ja	Ja	Nee	Nee	H	SSL/TLS	PKI Crypto	Java, .NET, Python & more
CATIA MAGIC	DDS	Dassault Systems	GA	-	Ja	Ja	Ja	Nee	Nee	L	SSL/TLS	PKI Crypto	Java
COREDX DDS	DDS	Twin Oaks Computing	GA	-	Ja	Ja	Ja	Nee	Nee	M	SSL/TLS	PKI Crypto	C, C++, C#, Java
FAST DDS	DDS	eProsima	GA	AL 2.0	Ja	Ja	Ja	Nee	Nee	H	SSL/TLS	PKI Crypto	C++, Python
HIVEMQ	MQTT 5.x	HiveMQ	GA	AL 2.0	Ja	Ja	Ja	Ja	Nee	M	SSL/TLS	SASL	Java, .NET, Python & more
MOSQUITTO	MQTT 5.x	Eclipse Foundation	GA	EPL 2.0	Ja	Ja	Ja	Ja	Nee	H	SSL/TLS	SASL	Java, Python
AMLEN BROKER	MQTT 5.x	Eclipse Foundation	Incuba-tion	EPL 2.0	Nee	Ja	Ja	Nee	Nee	M	SSL/TLS	SASL	C
NANOMQ	MQTT 5.x	EMQ	GA	MIT	Ja	Ja	Ja	Nee	Nee	H	SSL/TLS	Basic/JWT	C
EMQX	MQTT 5.x	EMQ	GA	AL 2.0	Ja	Ja	Ja	Nee	Nee	H+	SSL/TLS	HTTP/JWT/ LDAP	Java, .NET, Python & more
ZEROMQ		iMatix	GA	LGPLv3	Ja	Ja	Ja	Nee	Nee	M	SSL/TLS	SASL/ZAP	Java, .NET, Python & more
KAFKA		Apache Foundation	GA	AL 2.0	Nee	Ja	Ja	Nee	Ja	H+	SSL/TLS	SASL/mTLS/ Basic	Java, .NET, Python & more
STORM		Apache Foundation	GA	AL 2.0	Nee	Ja	Ja	Nee	Ja	M	SSL/TLS	SASL	Java, .NET, Python & more
PULSAR		Apache Foundation	GA	AL 2.0	Ja	Ja	Ja	Ja	Ja	H+	SSL/TLS	Basic/JWT/ OAuth 2.0/ Kerberos/ Athenz	Java, .NET, Python & more

Appendix B

Pulse example

```
prefix ex:      <http://example.com/id/>
prefix def:     <http://example.com/def/>
prefix bdi:     <https://ontology.bdinetwork.org/model/def/>
graph <ex:sampleData>{
   ex:message5 a def:Pulse .
   ex:message5 def:timestamp '18-02-1900 16:33:00'^^xsd:dateTime ;
            def:apiEndpoint <https://api.somewhere.nl/v1/> ;
            def:sparqlEndpoint <https://sparql.somewhere.nl/sparql/> ;
            def:targetProperties
               [
                  a def:KeyValuePair ;
                  def:key bdi:type;
                  def:value bdi:Vrachtwagen ;
               ] ,
               [
                  a def:KeyValuePair ;
                  def:key bdi:id;
                  def:value <https://picobellobv.nl/truck1>;
               ] .
}
```

Topsector Logistiek
Ezelsveldlaan 59
2611 RV Delft
+31 15 251 65 65
www.topsectorlogistiek.nl

Topsector Logistiek