Data warehousing Module

Analyse & terminologie

Minka Firth
Rob Lavrijssen
Colin Pieper

Fontys Hogescholen
Semester 4 – Data Processing – Module 2
Dec 2021

Inhoudsopgave

Data warehouse

Er zijn 3 verschillende soorten data warehouses.

- Enterprise Data Warehouse
- Operational Data Store
- Data mart

EDW Enterprise Data Warehouse

In een EDW wordt data van de hele organisatie opgeslagen. Deze data bestaat vaak uit operationele en transactie systemen zoals een ERP of CRM. Je kunt met behulp van Business Intelligence tools de data presenteren en analyseren. Naast business Intelligence en ETL-tools zijn er ook een andere hulpmiddelen voor data integratie en API's. Deze informatie wordt uiteindelijk gebruikt om Business vragen te beantwoorden.

ODS Operational Data Store

Een data store zorgt voor het rapporteren van operationele data en functioneert als gegevensbron voor een EDW. Een ODS is ontworpen om meerdere bronnen te integreren om de gegevens aan te vullen voor rapportages, controles en operationele besluitvorming.

Datamart

Een datamart is een onderwerp georiënteerde database, dat meestal een onderdeel is van een Data warehouse. De data wordt hier opgeslagen in snapshots, en is vaak wat meer samengevat dan in een data warehouse. Ze zijn vaak kleiner en meer gericht op een specifieke doelgroep. Op deze manier heeft niet iedereen toegang tot dezelfde informatie, alleen de informatie die ze werkelijk nodig hebben. Dit is daarom een hele veilige optie voor grote data warehouses.

Metadata

Meta data is informatie over de data zelf. Bijvoorbeeld bij een film zou de meta data kunnen zijn wie de regisseur is, welke datum het verfilmd is of wie wat heeft aangepast. Meta data houdt bijvoorbeeld bij waar data heen is geweest en wie en wat en wanneer er veranderd is. Het voordeel hiervan is dat bedrijven op deze manier kunnen vinden waar de data vandaan komt, waarom deze is aangepast of juist niet.

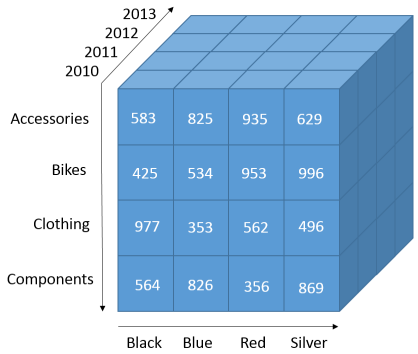
Metadata opslag

De metadata van informatie wordt meestal opgeslagen in een metadata opslagplaats. Dit is een database die gecreëerd is voor het opslaan van meta data. Zo is het mogelijk relaties te bouwen tussen verschillende metadata types, is er betere versie beheer en de data makkelijker is te valideren en meer integer.

Data Cubes

OLAP

OLAP staat voor Online analytical processing. Dit is een techniek voor het analyseren van data. Dit geeft beter en sneller inzicht in de informatie die beschikbaar is. Als er meer dan 3 dimensies zijn, wordt het ook wel een HyperCube genoemd. Een OLAP-kubus bestaat uit feiten en berekeningen en worden gecategoriseerd door dimensies.



Verschillen ROLAP MOLAP HOLAP

	ROLAP	MOLAP	HOLAP
Opslaglocatie voor aggregatie	Relationele database wordt gebruikt als opslaglocatie	Multidimensionale database wordt gebruikt als opslaglocatie	Multidimensionale database wordt gebruikt als opslaglocatie
Rekenkracht	Langzaam	Snel	Snel
Opslagruimte	Veel opslagruimte nodig	Middelmatige opslagruimte nodig	Weinig opslagruimte nodig
Opslaglocatie voor detailgegevens	Relationele database wordt gebruikt	Multidimensionale database wordt gebruikt	Multidimensionale database wordt gebruikt
Vertraging	Laag	Hoog	Middel
Query reactietijd	Langzaam	Snel	Middel

Data warehouse design

Er zijn verschillende factoren die er voorzorgen dat je data warehouse design er anders uit komt te zien. Voordat je begint moet je hier rekening me houden.

De eisen van de business

- Wat is het doel van het project?
- Wat is de scope van de projecten in relatie met je business objecten?
- Wat zullen we in de toekomst nodig hebben en wat hebben we op het moment nodig?
- Bedenk een disaster recovery plan.
- Nadenken over iedere laag van de beveiliging.

Een fysieke omgeving opzetten

Je hebt 3 fysieke omgevingen nodig:

- Develop
- Test
- Productie

Dit doen we omdat we een manier nodig hebben om aanpassingen te kunnen testen zonder dat deze meteen in de productie komt. Bijvoorbeeld kan het beveiliging risico's meenemen. Ook kan het voor problemen veroorzaken als bijvoorbeeld de server crasht door een verkeerd gelopen test of door een bug die voortgekomen is uit een aanpassing.

Een data warehouse model kiezen

Je hebt heel veel verschillende DW-modellen.
De drie meest-voorkomende zijn:

- Snowflake
- Star
- Galaxy

Snowflake

- Niet veel opslagruimte nodig.
- Makkelijker om dimensies te implementeren.
- Omdat er meer tabellen zijn, is de prestatie minder.
- Er is daarom ook meer onderhoud nodig.

Star

- Iedere dimensie wordt gepresenteerd in 1 dimensie tabel.
- De dimensie tabellen moeten eigenschappen hebben.
- De dimensie tabellen moeten met een foreign key verbonden worden met feitentabel.
- De tabellen zijn niet genormaliseerd dus kost het iets meer opslagruimte.
- Dit schema wordt heel goed ondersteunt door BI tools.

Galaxy

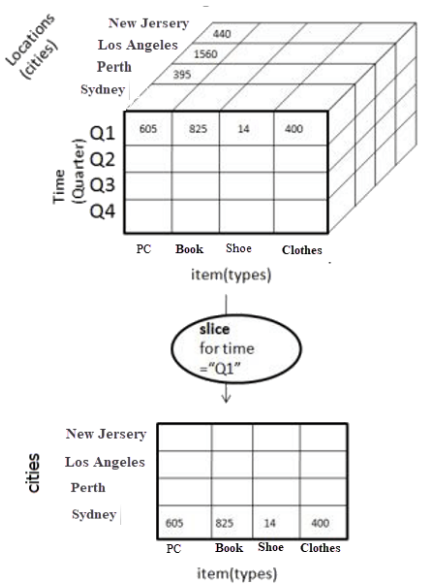
- Kan meer dan 1 feitentabel bevatten.
- Alle dimensie tabellen zijn genormaliseerd tenzij er niet de nodige ruimte voor is.
- De feitentabel heeft feiten en berekeningen.
- De dimensie tabellen hebben foreign keys om met de feitentabel verbinden.
- Het proces om de dimensie tabellen in kleinere dimensie tabellen maken kost meer opslagruimte

OLAP Operations

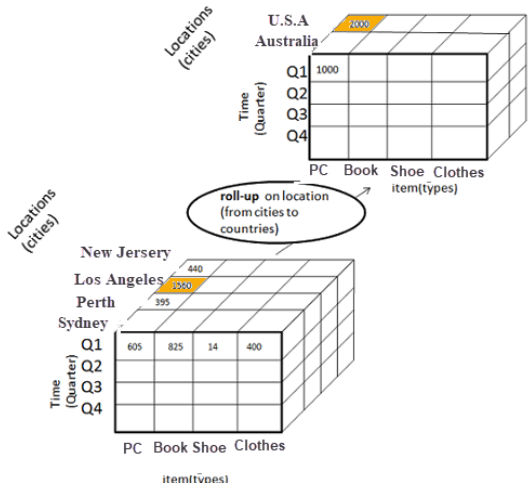
Roll up & drill Down

Roll-up wordt ook wel Consolidatie of Aggregatie genoemd. Deze operatie kan op 2 manieren worden uitgevoerd:
- Door het verminderen van dimensies.
- Door omhoog te gaan in de hiërarchie.

Er wordt dus uitgezoomd op de informatie. Deze wordt meer overkoepelend.

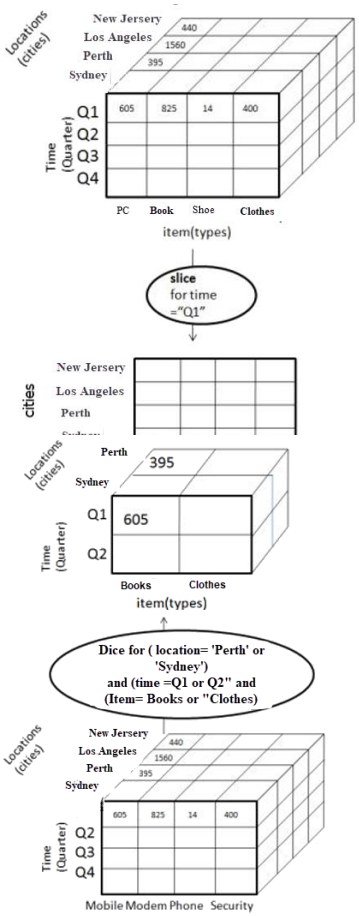


Drill-down is juist het tegenovergestelde. Er wordt ingezoomd op de informatie. Op deze manier kunnen meer details worden gevonden. Dit is mogelijk door naar beneden te gaan in de hiërarchie of door een dimensie toe te voegen.



Slice

Met **Slice** wordt er een klein stuk van de kubus geselecteerd en wordt er een nieuwe gemaakt.
In het voorbeeld laten ze zien dat ze van Q1 een slice hebben gemaakt met alle steden. Zo krijg je een alleen een overzicht van Q1. Dit kan handig zijn in combinatie met andere operations.



Dice

Dice is ongeveer hetzelfde als Slice alleen worden er meerdere dimensie geselecteerd. Hier in het voorbeeld zie je dat ze bijvoorbeeld Q1 en Q2 hebben geselecteerd in combinatie met Perth en Sydney.

ETL vs ELT

Extract: de data wordt uit de originele source gehaald. Bij ETL, wordt deze data in een tijdelijke staging area geplaatst, bij ELT wordt deze data rechtstreeks naar een data lake storage system verplaatst.

Transform: de data wordt getransformeerd zodat het geïntegreerd kan worden met de target data.

Load: de data wordt geladen in een data storage system.

ETL

Bij het ETL-proces wordt de data eerst extract uit de originele data source. Deze data wordt in een tijdelijke 'staging area' geladen. Hier wordt de data daarna schoongemaakt en getransformeerd. Daarna pas wordt de data in een data warehouse geladen.

Voordelen	Nadelen
Makkelijker om dataprivacy te beheren.	Omdat je van tevoren de data moet transformeren, is het misschien niet in het juiste format dat je uiteindelijk nodig hebt.
Een hoop documentatie te vinden.	Tijd kwijt aan het editen van data om het in de juiste vorm te krijgen, als het verkeerd getransformeerd is.
Je transformeert alleen de data die je werkelijk nodig hebt.	Gebruikt geen data lake.
	Werkt beter voor kleinere data sets.
	Loading tijd is vaak langzamer dan ELT.

ELT

Bij het ELT-proces is de eerste stap ook het extracten van de data uit de originele source. In plaats van een staging area, wordt deze data in een "data lake" geladen. Pas daarna wordt de data schoongemaakt en getransformeerd. Data lakes zijn speciale data stores, die in tegenstelling tot OLAP-data warehouses, gestructureerde en ongestructureerde data kan aannemen. Je kan iedere soort of rauwe informatie in een data lake laden, ondanks het format (of het gebrek aan daarvan).

Voordelen	Nadelen
Omdat de data nog niet getransformeerd is, is het laden van de data een stuk sneller.	Omdat dit een nieuw proces is, is er minder documentatie beschikbaar.
Is vaak makkelijker om te automeren.	Het overal proces is vaak duurder dan een ETL-proces.
Gebruikt een data lake.	
Elimineert het 'staging area' proces van ETL.	
Is vaak makkelijker te onderhouden, door automatische oplossingen in plaats van te rekenen op handmatige updates.	

ETL-tools

Een ETL-tool (Extract, Transform en Load) gebruikt een bedrijf voor verschillende redenen.

Tijdbesparend

Een ETL-tool zorgt transformeert en consolideert de data op een geautomatiseerde manier. Dit betekent dat het proces een stuk sneller is afgehandeld.

Handelt complexe data makkelijk af

Omdat de meeste data warehouses bestaan uit data van veel verschillende originele bronnen, gaat er een hoop tijd verloren om van al deze data handmatig af te handelen. Een ETL-tool helpt met het schoonmaken van de data en dit en alleen de bruikbare data eruit te halen.

Minder fouten

Een mens maakt fouten. Zelfs de kleinste fout kan groten gevolgen hebben in een data warehouse. Een ETL-tool zorgt voor een geautomatiseerd proces en verlaagt dus de kans op handmatige fouten die normaal wel voor zouden komen. Denk aan Typo's of verkeerde format.

Verbetert BI en ROI

BI tools kunnen heel goed overweg met ETL-tools. Het gebruik van business intelligence tools is belangrijk om beter inzicht te krijgen in de informatie. Daarom is het belangrijk dat deze tools goed kunnen samenwerken.

Soorten ETL-tools

Batch ETL Tools

Dit type ETL-tool werkt door eerst batches te maken. Deze data wordt gehaald uit een andere bron, dan getransformeerd en geladen in de verzamelplaats voor ETL-opdrachten. Het is een kost efficiënte manier om dat het minimalen middelen gebruikt op een tijdsgebonden manier.

Real-Time ETL Tools

Dit is een beetje het tegenovergestelde van Batch. Hier wordt data juist in real time opgehaald, schoongemaakt en geladen naar de andere systemen. Daardoor is er snellere toegang tot de informatie mogelijk. Deze tools worden laatste tijd steeds populairder omdat er steeds meer vraag is om zo snel mogelijk inzicht te krijgen tot deze informatie.

On-Premise ETL Tools

De meest voornamelijke reden dat meeste bedrijven zowel de Data als de opslag plaats On-Premise hebben is vanwege security. Vandaar dat ze dan ook een ETL-tool willen hebben die On-Prem is zodat er geen verbinding naar buiten staat.

Cloud ETL Tools

Omdat er nu steeds meer van uit huis gewerkt kan worden of bedrijven hebben verschillende locaties vestegingen staan, is de Cloud ETL-tool ook aantrekkelijk. Hierdoor kunnen ze de data makkelijk beheren en versturen naar bijvoorbeeld BI tools of Databases. Hierdoor zijn ze flexibeler en meer agile.

Verschillende ETL-Tools

Om een goede keuze te kunnen maken voor welke ETL-tool we gaan gebruiken moeten we eerst een lijst maken. Hieronder heb ik de top 10 ETL tools genoemd die we hebben gevonden op een onafhankelijke bron. Deze is gefilterd op grootte van het bedrijf. Op basis van deze lijst gaan we daar 3 ETL tools uitkiezen voor de short list en de mogelijkheden daarvan ook bespreken.

- Informatica PowerCenter
- SQL Server Integration Services (SSIS)
- Denodo Platform
- Azure Data Factory
- Nexia Unified Data Operations
- Fivetran
- AWS Glue
- Matillion ETL
- HVR
- Azure Data Factory

ETL-tool kiezen

Als je een ETL-tool wil kiezen moet er eerst gekeken worden naar wat de wens is van de business. Echter zijn er nog een paar extra dingen waar je over na zou kunnen denken

Kosten

De meest voor de hand liggende factor. Hoeveel geld kan er worden gestoken in een ETL-tool?

Connectiviteit

De juiste ETL-tool moet verbinding kunnen maken met alle gegevensbronnen die door het bedrijf gebruikt worden, dit zijn de Built-in connectors.

Makkelijk Interface

Een bug vrij en makkelijk te gebruiken interface geeft een constante en betrouwbare ervaring voor het behandelen van data.

Schaalbaarheid

Aangezien we er van uit gaan dat het bedrijf ook groeit, moet er ook gekeken worden of de ETL-tool hier wel voor gemaakt is. Het zou niet handig zijn als deze bijvoorbeeld een limiet heeft op het aantal gegevensbronnen of data dat hij per maand mag verwerken. De ETL-tool moet ook goed kunnen opschalen als het bedrijf groeit.

Error-Handeling

Een ETL-tool moet ook efficiënt om gaan met fouten en zorgen dat de data nog steeds accuraat is. Ook moet het zorgen dat het feilloos en efficiënt data transformeert en zijn er geen gevallen van verloren data.

Real-Time Data Access

Omdat bedrijven steeds sneller toegang willen tot de data is het handig als de ETL-tool ook bij de data kan doormiddel van bijvoorbeeld een webapplicatie in real-time.

Built-in Monitoring

Een ETL-tool zou ook moeten komen met een built-in monitoringsysteem dat real-time updates geeft over verlopen van het proces en de jobs die hij uitvoert.

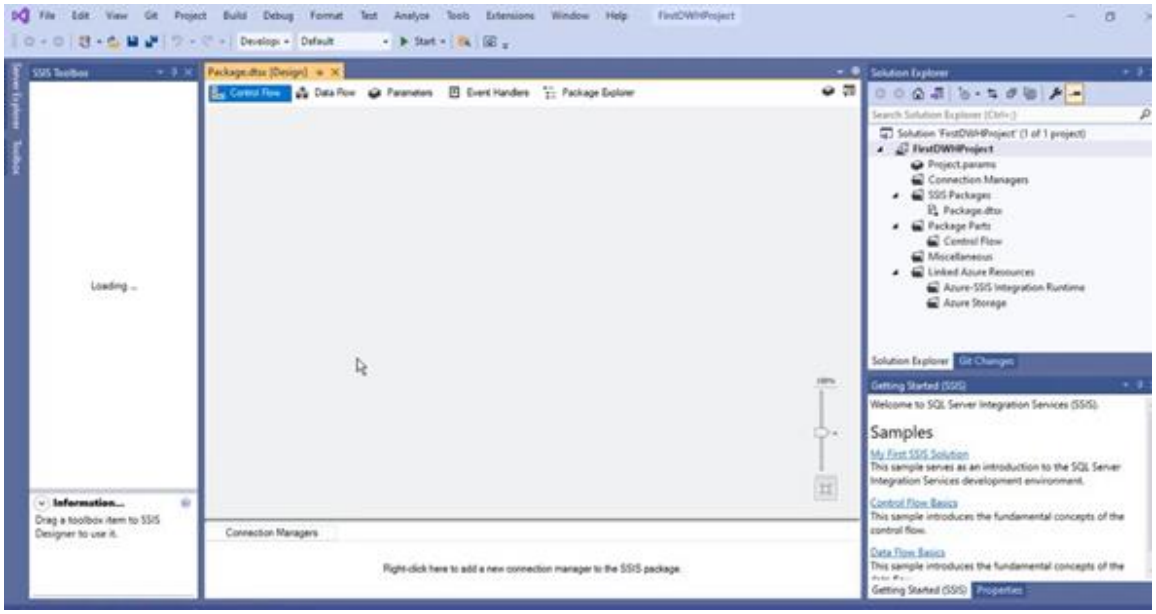
De top 3

De shortlist die je hier onder zal tegen komen bestaat uit de 3 ETL tools die ons persoonlijk het meeste aanspreken. Dit hebben wij gedaan op basis van community (is er veel van op internet te vinden voor veel voorkomende problemen), documentatie (heeft de ontwikkelaar zelf goed gedocumenteerd hoe het programma werkt en eventuele tutorial pagina's of bestanden) en naamsbekendheid (bijvoorbeeld zoiets als Microsoft of wordt het veel benoemt in de community).

SQL Server Integration Services (SSIS)

Deze tool is gratis te gebruiken en van dezelfde makers als Microsoft Visual Code en daarom makkelijk te verbinden met elkaar. Het is ook een goede tool als er gebruik wordt gemaakt van verschillende databronnen en eventuele uitbreidingen zijn ook mogelijk. Het UI is wat zakelijker dan de andere op de lijst, maar dit is een kwestie van smaak en gewenning. Het is ook een tool waar een hoop documentatie en tutorials over te vinden zijn.

Kosten	Inbegrepen bij SQL Server Enterprise (vanaf ~$14.000)
Connectiviteit	Kan met veel gegevensbronnen overweg.
Makkelijk interface	Het heeft een OK interface, niet te moeilijk maar het kan toch iets mooier.
Schaalbaarheid	Het heeft goede schaalbaarheid je kan zo hoog als enterprise niveau schalen.
Error-handeling	Er is wel een ingebouwde error-handeling. Maar deze is niet zo uitgebreid.
Real-time Data Access	Het is wel mogelijk maar daar moet je wel een hoop voor instellen. Het komt niet out of the box.
Built-in Monitoring	Er zijn verschillende monitoring features zoals: logs, reports, views, performance counters en data taps.

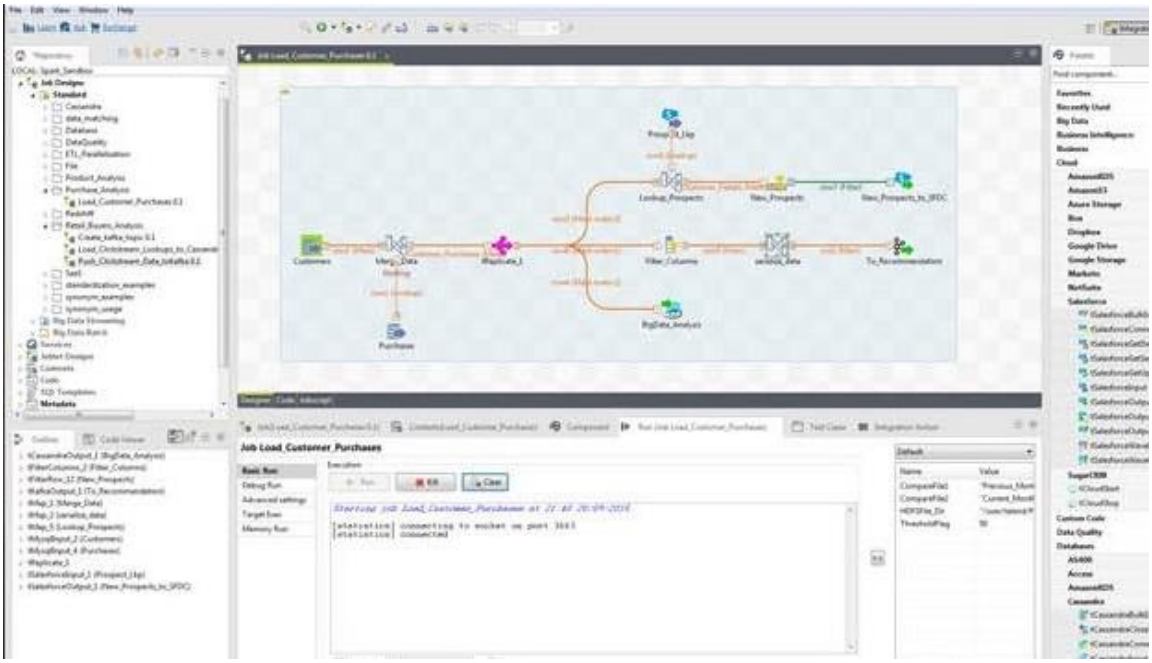


Talend Open Studio

Talend Open Studio heeft ook goede ETL-mogelijkheden, het is makkelijk te gebruiken, en kan ook met veel verschillende gegevensbronnen overweg. Het is Open source, en heeft een UI dat makkelijk te gebruiken is. De tool is gebaseerd op de Eclipse IDE dus als je daar bekent mee bent komt dat wel goed uit. Verder is het installeren van Talend wat lastiger omdat er bijvoorbeeld drivers en instellingen verkeerd kunnen staan. Ook heeft Talend goede support of duidelijke documentatie om te beginnen.

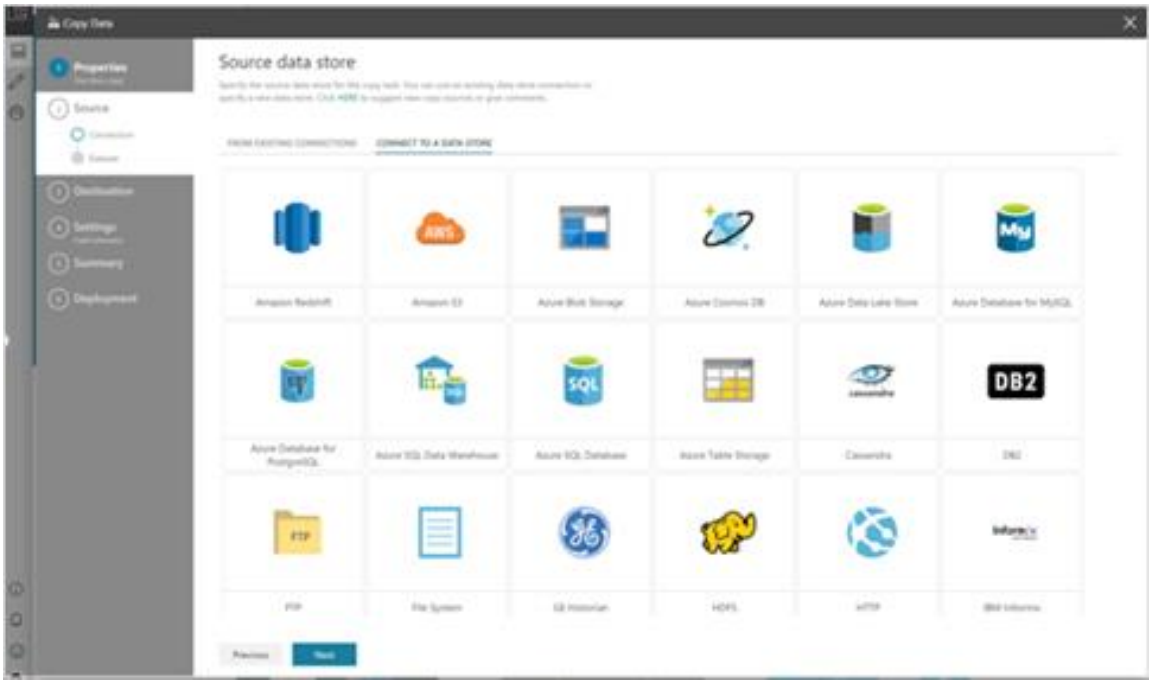
Kosten	Gratis
Connectiviteit	Je kan met veel verschillende gegevensbronnen verbinden en op veel verschillende manier zoals FTP.
Makkelijk interface	Interface is heel gebruiksvriendelijk, je kan alles duidelijk zien en vinden. Het is ook mooi afgewerkt.
Schaalbaarheid	Omdat met talend een grote hoeveelheid data kan worden verwerkt, is de schaalbaarheid aan de goede kant.
Error-handeling	Talend heeft built-in error-handeling met een pagina om dit uit te leggen.
Real-time Data Access	Er is real-time access, maar dit moet apart worden ingesteld en komt niet standaard.
Built-in Monitoring	Er is een standaard monitorings console.

Azure Data Factory



Azure data Factory is een heel uitgebreide tool. Erg gebruiksvriendelijk, en in verhouding met betaalde ETL-tools goedkoop. En kan ook autonoom worden gebruikt. Deze tool kan ook goed samen werken met SSIS. De UI is erg duidelijk en professioneel gemaakt.
De tool is snel en betrouwbaar, heeft veel templates voor pipeline designs en makkelijk te maken ETL workflows.

Kosten	Prijzen verschillen. Er wordt betaald per pijplijnactiviteit en gegevensverplaatsing en hierdoor kunnen prijzen oplopen.
Connectiviteit	Hele goede connectiviteit, het heeft heel veel build-in connectors en werkt goed samen met andere Microsoft-producten
Makkelijk interface	De interface is heel keurig afgewerkt zoals je gewend bent van Microsoft.
Schaalbaarheid	Heel goede schaalbaarheid, het is ook ontwikkeld voor Enterprise niveau. Kan ook in de cloud waardoor schalen van opslag heel eenvoudig gaat.
Error-handeling	De error handling gaat net zoals het gaat bij SSIS en heb je met een heel duidelijk interface opties voor error handling.
Real-time Data Access	Er is realtime access tot de data, daar moeten wel instellingen voor worden verricht.
Built-in Monitoring	Met behulp van extra services (en dus ook vaak extra kosten), zijn er mogelijkheden voor monitoren.



SSIS SQL Server Integration Services

Dit is een onderdeel van de MSSQL Server software. Het is een snelle en flexibele ETL-tool. Het maakt het dus makkelijker om data van de ene bron naar de andere te verplaatsen en, waar nodig, schoon te maken.

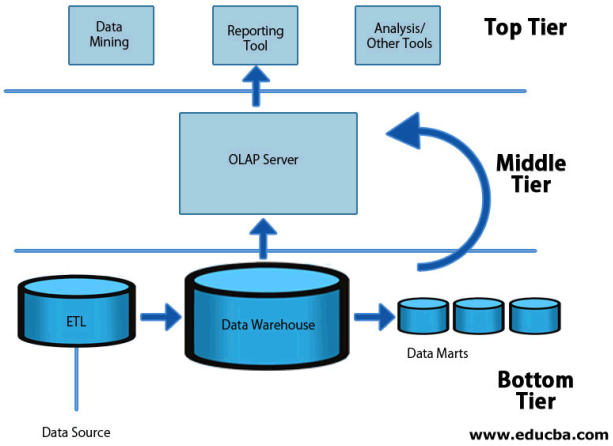
Ons advies

Op basis van het onderzoek en de resultaten daarvan raden we aan om Azure Data Factory. De reden daarvoor is dat het goed voorbereid is voor de toekomst in verband met de schaalbaarheid. De connectiviteit is ook erg goed voor eventuele uitbreidingen/overnames van bedrijven die andere gegevensbronnen gebruiken. De community is groot genoeg en het is van Microsoft, een groot bedrijf met genoeg documentatie.

Echter kunnen wij deze oplossing niet gebruiken voor ons prototype, in verband met de kosten die hieraan zitten. De keuze wordt daarom SSIS. Met SSIS kan je net zo goed met de versie die wij gebruiken goede ETL-opdrachten uitvoeren. Deze zullen dan ook voldoen aan de eisen wat wij eraan stellen.

Lagenverdeling

De meeste data warehouses hebben ongeveer dezelfde lagen, maar er zijn heel veel verschillende manieren om dit te doen. Er is natuurlijk de originele bron waar de data uit wordt gehaald. Deze data wordt in dezelfde laag getransformeerd en in de data warehouse geladen. Dit is de onderste tier. Eventuele datamarts bevinden zich ook in deze laag.



De middelste laag bestaat uit de OLAP-server (pagina 4). Deze laag zorgt ervoor dat de andere lagen kunnen communiceren.

De bovenste laag is de laag die werkelijk door de gebruiker wordt gebruikt. Hier bevinden zich de tools die gebruikt worden om met een data warehouse te communiceren en werken zoals; Query tools, analyse en datamining-tools.

De afbeelding komt enigszins overeen met ons prototype, maar er wordt in ons prototype geen gebruik gemaakt van datamarts. Er is namelijk maar één groep die bij de data kan komen. Als er later nieuwe afdelingen worden toegevoegd, kunnen er altijd datamarts worden toegevoegd.

Security

Omdat er niet alleen financiële gegevens in een data warehouse wordt opgeslagen, maar ook vaak persoonsgegevens, is het beschermen van de data warehouse een prioriteit. Het is vooral belangrijk om persoonsgegevens te bewaren, maar ook intellectuele eigendommen, trade secrets en financiële rapportages. Deze gegevens moeten niet alleen beschermd worden uit belang van de organisatie, maar er zijn ook meerdere wetten die de privacy van mensen beschermen.

Hier zijn verschillende methodes voor. Het is per data warehouse verschillend welke methodes nodig zouden zijn.

Gevoelige gegevens classificeren

Het is belangrijk niet alleen te weten wat de gevoelige data is, maar ook waar deze is. Er zijn meerdere programma's die de data opslagplaats scant en de data sorteert. Dit zijn de classificaties die vaak worden gebruikt:

- Publiek (public)
- Vertrouwelijk (confidential)
- Gevoelig (sensitive)
- Persoonlijk (personal)

Het is ook mogelijk om de classificatie te updaten als de data verplaatst, aangepast of gecreëerd wordt. Het is dan wel belangrijk dat er limieten zijn die ervoor zorgen dat gebruikers de classificaties niet kunnen aanpassen. Alleen gebruikers met de juiste toegang credentials mogen de classificatie aanpassen.

Toegang beperken

Zodra het duidelijk is wat en waar de gevoelige data is, moet er een beleid worden opgezet dat bepaalt wie bij welke data mag komen. Er moet bepaalt worden wat er met deze data gedaan mag worden en consequenties voor het niet houden aan het beleid.

Aan de hand van dit beleid moeten de juiste mensen toegang krijgen tot de juiste informatie. Dit zou bijvoorbeeld kunnen door datamarts te maken of door privileges te geven.

Encryptie

Het is ook mogelijk om een encrypte versie te maken van de data. Op deze manier is de data enigszins gelimiteerd in analyse opties, maar wel beschermd. Dit is vaak een dure optie omdat de meeste encryptie opties specifieke encryptie functies moet aanroepen vanuit de werkelijke applicatie. Dit betekent dat de developer de applicatie door en door moet kennen.

Firewalls

Met een firewall zorg je ervoor dat netwerken van elkaar geïsoleerd zijn. Dit kunnen standalone systemen zijn of bij andere infrastructuur apparaten horen. Ze zorgen er ook voor dat ongewenste gebruikers niet zomaar het netwerk binnen kunnen komen, dit helpt beschermen tegen datalekken van malware en hackers. Aan de hand van de organisaties beleid wordt bepaald of er gebruikers zijn, wie die gebruikers zijn, en of ze moeten verifiëren voor ze het mogen gebruiken.

Proxyserver

Op deze manier wordt er een middelman gemaakt die de verzoeken van gebruikers gunt of afwijst. Een gebruiker maakt een connectie naar de proxyserver, de server evalueert het verzoek en maakt dan een beslissing. Proxyservers worden vaak gebruikt voor traffic-filetering en performance verbetering.

Fysieke bescherming

Er zijn ook manieren om de data warehouse met externe middelen te beschermen. De fysieke en online beveiliging in en om het gebouw van de organisatie is daarom heel belangrijk. Het is bijvoorbeeld een stuk makkelijker een laptop te stelen in plaats van een desktopcomputer. Maar er moet niet alleen beschermd worden tegen diefstal, er moeten ook fysiek en online beschermingen zijn tegen virussen.

Bronvermelding

Aihini, A. (z.d.). *Data Warehouse Security Best Practices*. The Data School. Geraadpleegd op 5 December 2021, van https://dataschool.com/data-governance/data-warehouse-security/

Astera. (z.d.). *What is ETL Tool And Why Do You Need It?* Geraadpleegd op 5 december 2021, van https://www.astera.com/type/blog/what-is-etl-tool/

Corporate Finance Institute. (z.d.). *Data Warehouse*. Geraadpleegd op 5 december 2021, van https://corporatefinanceinstitute.com/resources/knowledge/data-analysis/data-warehouse/

EDUCBA. (z.d.). *Galaxy Schema*. Geraadpleegd op 5 december 2021, van https://www.educba.com/galaxy-schema/

Gartner. (z.d.). *Data Integration Tools Reviews and Ratings*. Geraadpleegd op 13 december 2021, van https://www.gartner.com/reviews/market/data-integration-tools

GeeksforGeeks. (z.d.). *Difference between ROLAP, MOLAP and HOLAP*. Geraadpleegd op 5 december 2021, van https://www.geeksforgeeks.org/difference-between-rolap-molap-and-holap/

Guru99. (z.d.). *Star and Snowflake Schema in Data Warehouse with Model Examples*. Geraadpleegd op 5 december 2021, van https://www.guru99.com/star-snowflake-data-warehousing.html?utm_source=xp&utm_medium=blog&utm_campaign=content

Snowflake. (z.d.). *What is an Enterprise Data Warehouse?* Geraadpleegd op 5 december 2021, van https://www.snowflake.com/guides/what-enterprise-data-warehouse

Wikipedia. (z.d.). *Metadata*. Geraadpleegd op 5 december 2021, van https://nl.wikipedia.org/wiki/Metadata

Wikipedia. (z.d.). *Microsoft Analysis Services*. Geraadpleegd op 5 december 2021, van https://en.wikipedia.org/wiki/Microsoft_Analysis_Services

Wikipedia. (z.d.). *OLAP cube*. Geraadpleegd op 5 december 2021, van https://en.wikipedia.org/wiki/OLAP_cube

Wikipedia. (z.d.). *SQL Server Integration Services*. Geraadpleegd op 5 december 2021, van https://nl.wikipedia.org/wiki/SQL_Server_Integration_Services