



Cloud Datawarehousing

Ontwerp, implementatierichtlijnen en best practices

Praktisch halfdags seminar met Rogier Werschull

AdeptEvents

- Architectuurprincipes en verhouding tot traditionele datawarehousing, data lake, data lakehouse
- De rol van de data scientist en het belang daarvan
- Opzetten van een modern datawarehouse met een cloud native analytische database
- Vergelijking van de meest gebruikte cloud analytische databases: Snowflake, Google BigQuery, Amazon Redshift en Azure Synapse
- Additionele tooling om een cloud-gebaseerd datawarehouse mee te bouwen, bijv. Stitch, Fivetran, DBT en Matillion
- Het belang van datawarehouse automation bij de bouw van het moderne datawarehouse
- Lessons learned op het vlak van implementatie en techniek
- De 10 Do's en Don'ts van cloud datawarehousing

TAAL
Nederlands of Engels

LOCATIE
Online / omgeving Utrecht / Hilversum

TIJD
Van 9:00 tot 13:00 uur

REGISTRATIE
www.adeptevents.nl



Cloud Datawarehousing

Ontwerp, implementatierichtlijnen en best practices

Rond 2015 begonnen Nederlandse bedrijven met het migreren van *on-premise* datawarehouses naar cloud-omgevingen. Maar hoe ontwerpen we eigenlijk een goed op cloud technologie gebaseerd datawarehouse dat daadwerkelijk gebruik maakt van de nieuw beschikbare technologische mogelijkheden? Hoe zorgen we ervoor dat zo'n datawarehouse voldoet aan de veranderende informatiebehoefte en hedendaagse klantwensen? En tenslotte: theorie is belangrijk, maar wat zijn nu de praktische lessons learned bij het bouwen van een modern datawarehouse?

Steeds meer organisaties '*wagen de sprong*' om hun datawarehouse te verplaatsen naar de cloud. Een belangrijke reden hiervoor is dat het duidelijk is geworden dat cloud-native analytische databases, de drijvende technologie achter het cloud gebaseerde datawarehouse, op het gebied van schaalbaarheid en TCO een on-premise systeem in veel gevallen verslaan.

Eén van de belangrijkste voordelen van cloud analytische databases is hun vrijwel '*eindeloze*', onafhankelijk schaalbare opslag en rekenkracht. Veel organisaties komen er echter achter dat het nog niet zo eenvoudig is om optimaal gebruik te maken van deze nieuwe mogelijkheden als ze de migratie uitvoeren volgens een *lift-and-shift* aanpak. Dit komt doordat veel gebruikte data architectuur ontwerp principes van het '*klassieke datawarehouse*' vaak niet helemaal logisch meer zijn binnen een cloud context.

Een tweede probleem van de '*lift en shift*' migratie is een mogelijke mis-match op het gebied van hedendaagse consumptiepatronen. Een datawarehouse dat alleen nog maar focust op tactische / strategische informatievoorziening voldoet niet meer. In veel organisaties bestaat een concrete vraag naar operationele (meer realtime) analytische informatie, met daarnaast ook nog eens wensen op het gebied van goed vindbare, ruwe data ten behoeve van de

'*explorerende*' data scientist. Veel data architecten zijn zich gaan realiseren dat '*de data scientist*' momenteel veel werk verricht dat prima past onder de taken van 'datawarehousing'. Een goed ontworpen modern datawarehouse ziet deze data scientist dan ook als een primaire klant.

Seminar

In dit half-daagse seminar zal Rogier Werschull u aanknopingspunten geven waarmee u een begin kunt maken met het naar de cloud brengen van uw huidige datawarehouse. Op basis van een data architectuur die past bij hedendaagse use cases en gebruikers consumptiepatronen.

Hiernaast krijgt u inzicht hoe de verschillende (native) cloud analytische databases zich tot elkaar verhouden. Welke oplossing en bijbehorende toolset moeten we kiezen en waarom? Bij de vergelijking van deze oplossingen beperken wij ons tot de belangrijkste spelers van dit moment:

Amazon Redshift, Google BigQuery, Snowflake en Azure SQL datawarehouse / Synapse. De ondersteunende tools die voorbij zullen komen vallen in de categorie *Extract-Load* tools zoals Fivetran en Stitch en nieuwe data transformatie tools, zoals Data Build Tools (DBT) en Matillion.

Het belangrijkste doel van dit seminar is om u een 'kickstart' te geven waarmee u concreet mee aan de slag kunt. Dat u eerst gaat beseffen wat het implementeren van een op (native) cloud technologie gebaseerde ELT (niet ETL!) data architectuur voor analytische consumptie patronen in de praktijk betekent. En hoe u vervolgens op basis van nieuwe architectuur principes kunt beginnen met de technische implementatie. Daarnaast zal de trainer zijn persoonlijke 'lessons learned' en best practices met u delen.

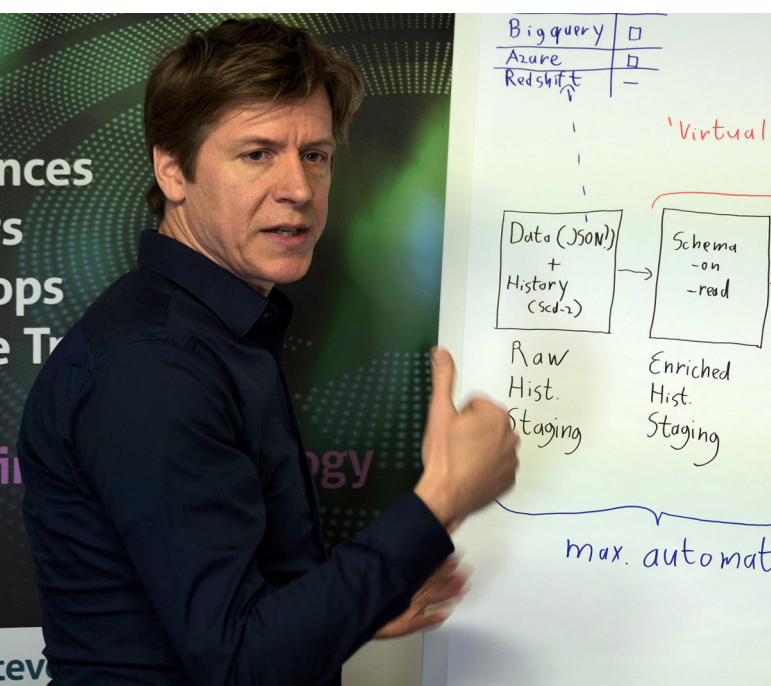
Leerdoelen

In deze training krijgt u antwoord op de volgende vragen.

- Hoe zit het ook alweer met datawarehousing als architecturale discipline, hoe overlapt dit met recentere concepten als het 'data lake' of het 'data lakehouse'?



- Wat is een 'moderne manier van datawarehousing' en hoe verschilt dit van de klassieke aanpak? We kijken hier met name naar kwaliteitseisen en gebruiker toegangspatronen.
- Wat is de rol van de 'Data Scientist' in deze context en waarom is deze essentieel?
- Waarom kan het opzetten van een modern datawarehouse beter in een cloud native analytische database dan op basis van 'on-premise' technologie?
- Wat bedoelen we met 'separation of concerns', het primaire architectuur principe achter het ontwerp van een modern datawarehouse?
- Wat zijn de belangrijkste data architectuur principes, voortvloeiend uit 'separation of concerns' en het effect hiervan op de praktische implementatie van het moderne datawarehouse?
- Wat zijn de belangrijkste voor- en nadelen van de meest gebruikte cloud analytische databases: Snowflake, Google BigQuery, Amazon Redshift en Azure Synapse.
- Welke additionele tooling komen we tegen om een cloud-gebaseerd datawarehouse mee te bouwen (bijv. Stitch, Fivetran, DBT en Matillion)?
- Hoe ziet de cloud agnostische datawarehouse architectuur er uit die de trainer sinds 2016 inzet?
- Waarom datawarehouse automation een essentiële rol zou moeten spelen bij de bouw van het moderne datawarehouse, en hoe dit globaal gezien zou moeten gebeuren.
- Wat is de rol van elk van de vier lagen van het moderne datawarehouse?
- Wat zijn de gebruikte fysieke data modellerings technieken per datawarehouse laag?
- Wat zijn de 'lessons learned' op het vlak van implementatie en techniek?
- Wat zijn de 10 Do's en Don'ts van cloud datawarehousing?



Bestemd voor ú

Programma- en projectmanagers voor datamanagement en analytics, datawarehouse-professionals, architecten, technologieplanners. Tevens dataontwerpers, data-integratiespecialisten, data-engineers, datawarehouse / ETL developers. Eenieder die een rol heeft bij de migratie van legacy datawarehouses naar de cloud.

Programma

U ontvangt van ons ruim vooraf de documentatie (fysieke map of PDF) en uiteraard ook de uitnodiging met hyperlink voor toegang tot de virtuele sessie. Deze training in één ochtend bestaat uit een mix van theorie en praktijk.

Het programma start in de ochtend om 09:00 uur en duurt tot 13:00 uur. Log tijdig in en controleer vooraf uw geluid- en video instellingen.

Data Warehousing in the Cloud - revisited

- Data warehousing through the years
 - Reasons for failure
 - On Architecture VS Technology
- The Data Lake versus Data warehouse versus Data Lakehouse
 - Dissecting the hype words
 - On 'separation of concerns'
 - The Data Warehouse in Ronald Damhof's DQM model
- Issues with *classical* data warehousing
 - Costs
 - Scaling
 - Ease of use
- Why data warehousing in the cloud?
 - Costs
 - Scaling
 - Ease of use
- What is it that makes a modern data warehouse?

- Separate concerns: from an ETL to ELT patterns
- Changing consumption patterns: more '*near realtime*'
- Different users: enter the Data Scientist
- Data security / privacy (by design?)
- Changing tool sets
- What does not / should not change?
- What could be coming next?

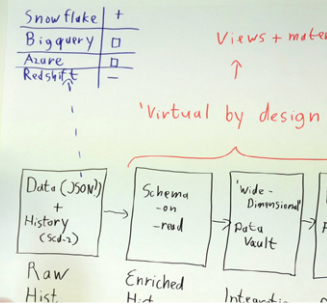
Comparing cloud analytical databases and supporting SAAS tools

- Cloud analytical Databases
 - BigQuery
 - Snowflake
 - Redshift
 - Azure SQL DWH / Synapse
- Comparing the solutions
 - Costs
 - Scaling
 - Development / deployment and maintenance
 - Privacy
- Supporting tools
 - Extract-load solutions
 - Transform-load solutions
 - CI/CD
 - Data catalogs
 - BI tools



Programma

Conferences
Seminars
Workshops
In-house T



The modern data warehouse layers: architecture principles and concerns

- The modern data warehouse architecture
 - Business architecture principles / rationales
 - Data architecture principles / rationales
 - Business / Logical data modeling
 - Separate concerns: the *what* and *why* of the modern data warehouse layers
 - Raw historized staging layer (RHS)
 - Enriched staging layer (EHS)
 - Integration layer (INT)
 - Presentation layer (PRE)
- The need for data warehouse automation in the
 - RHS layer
 - EHS layer
 - INT layer
 - The potential role of '*data virtualization*' in the
 - EHS layer
 - INT layer
 - PRE layer

The modern data warehouse layers: physical data model, technical implementation aspects and lessons learned

- Separate concerns: the what does each layer *look* like?
- Physical data model
- Metadata
- Separate concerns: the *how* to technically implement each layer guidelines
- Implementation principles
- Code patterns / suggestions
- Some lessons learned per layer
 - The need for RHS stability
 - Automation in the EHS
 - The always present complexity of INT
 - The need for agility of PRE
- Wrap-up: the 10 do's / don'ts of data warehousing in the cloud



ROGIER WERSCHULL

ROGIER WERSCHULL is data-solution architect met ruim 20 jaar ervaring met consultancy en het geven van training in het vakgebied van business-intelligence, datawarehousing en analytics. Functionaliteit vóór technologie. Open samenwerking en het actief delen van kennis. Dit is waar hij voor staat en gaat. Vanuit deze principes is hij de laatste 5 jaar voornamelijk bezig om als *hands-on* data en solution architect moderne op de cloud gebaseerde datawarehouse / analytics oplossingen neer te zetten of te migreren in diverse organisaties.

Als '*Head of Data*' heeft Rogier dit werk gedaan bij Spilgames (web en mobile gaming), dat in 2016/2017 als een van de eerste organisaties in Nederland volledig naar de Google Cloud migreerde (o.a. PubSub / Dataflow / BigQuery). Bij Property Finder, '*de Funda van Dubai*', werkte hij als lead aan de migratie van Amazon Redshift naar Snowflake (AWS). Vanaf 2020 is hij lead data architect bij 6 Gorilla's (6G), een healthcare startup die onderdeel is van de Tenzinger groep. 6G bouwt aan een dataplatform voor de zorg onder '*privacy by design*', op basis van o.a. Snowflake (Azure) en is hiermee de eerste grote '*health data*' klant van Snowflake in de EMEA regio. Rogier heeft gedurende zijn consultancy tijd bij Centennium in Den Haag jaren lang training gegeven op het gebied van Data Architectuur, Data Vault en Dimensioneel Modelleren en verzorgt nog steeds trainingen op gebied van datamodelleren.

Rogier spreekt regelmatig over datawarehousing in de cloud op internationale conferenties zoals Data Modeling Zone Europe, DataCon Dubai, The Knowledge Gap en de Datawarehousing & BI Summit. Ook blogt hij met enige regelmaat op LinkedIn.

DATUM EN TIJD

Het seminar vindt periodiek plaats in het voorjaar en/of najaar. De exacte data en aanvangstijden vindt u op onze website. Het programma start om 9:00 uur en duurt tot 13:00 uur. Inloggen is mogelijk vanaf 8:30 uur.

Bij voldoende belangstelling kan dit seminar in het Engels worden verzorgd.

PLAATS

Adept Events werkt met verscheidene accommodaties in de omgeving van Amersfoort en Utrecht. Op de website en in de herinneringse-mail die u een week voor het event ontvangt wordt altijd de definitieve accommodatie vermeld. Controleer dit voor uw vertrek.

De virtuele seminars en workshops worden aangeboden via een live video stream vanuit onze lokale cursusruimte annex videostudio. Omdat wij werken met een tweepersoons crew en meerdere professionele camera's en microfoons, is het mogelijk om de traditionele cursusruimte waar nodig volledig na te bootsen en daarbij vooral ook de interactie met deelnemers op peil te houden.

AANMELDEN

Aanmelden kan via ons online inschrijfformulier op www.adeptevents.nl. Geeft u de voorkeur aan schriftelijk aanmelden? Stuur de PDF van uw aanmelding of inkoopopdracht naar seminars@adeptevents.nl. Vermeld altijd duidelijk het e-mailadres van de deelnemer(s) alsmede dat van de crediteurenafdeling. Na ontvangst van uw aanmelding krijgt u de bevestiging en factuur per e-mail toegestuurd.

KOSTEN

Deelname aan dit seminar kost € 297 per persoon indien u zich uiterlijk 30 dagen voor aanvang registreert en daarna € 330,- (exclusief BTW)* inclusief documentatie.

Werkt u bij een gemeente of provincie? Dan kunt u BTW terugvorderen via het BTW compensatiefonds.

AANTREKKELIJKE KORTINGEN

Meldt u tegelijkertijd meerdere personen van één bedrijf aan voor hetzelfde evenement, dan geldt al vanaf de tweede deelnemer een korting van 10% per deelnemer. Vanaf vier deelnemers ontvangen alle deelnemers 15% korting (de deelnemers dienen op dezelfde factuur te staan)*.

*) Prijzen of kortingen in deze pdf brochure kunnen (tijdelijk) afwijken van de website. In dat geval prevaleert altijd de informatie op de website.

ANNULEREN

Annuleren dient schriftelijk te geschieden. U kunt annuleren tot drie weken voor het evenement plaatsvindt. Er wordt echter wel € 75,- (excl. BTW) administratiekosten in rekening gebracht. Annuleren is niet meer mogelijk vanaf drie weken voordat het evenement plaatsvindt. Vervanging door een ander dan de aangemelde persoon is te allen tijde mogelijk.

MEER INFORMATIE



+31(0)172 742680



<http://www.adeptevents.nl/cdw>



seminars@adeptevents.nl



@AdeptEventsNL / <https://twitter.com/AdeptEventsNL>



<http://www.linkedin.com/company/adept-events>



<https://www.facebook.com/AdeptEventsNL>



Bezoek onze Business Intelligence en Datawarehousing website www.biplatform.nl en download de App



Bezoek ook onze website over Software Engineering in al zijn facetten: www.release.nl en download de App

IN-HOUSE SESSIES VOOR UW MEDEWERKERS

Wilt u dit seminar binnen uw organisatie aanbieden als in-house sessie voor een groep medewerkers? Een zeer aantrekkelijke optie voor uw organisatie! Bel of stuur een e-mail via ons contactformulier. Op onze website vindt u meer informatie over de mogelijkheden van In-house seminars en workshops.