Cloud Datawarehousing
Practical Approach and Best Practices

• Architectural aspects relating to traditional data warehousing, data lake, data lakehouse
• The role and relevance of the ‘Data Scientist’
• Why can a modern data warehouse be built better using a cloud (native) solution?
• Comparison of the most commonly used cloud analytical databases: Snowflake, Google BigQuery, Amazon Redshift and Azure Synapse
• Additional tooling that can be used to build a cloud based data warehouse, such as Stitch, Fivetran, DBT and Matillion
• Why data warehouse automation should play a vital role
• Lessons learned on technical / implementation aspects per layer

Workshop by
Rogier Werschkull

AdeptEvents

LANGUAGE
Dutch or English
VENUE
ONLINE / Utrecht, The Netherlands
TIME
9:00 – 13:00 hours or 16:30 hours
REGISTRATION
www.adeptevents.nl
Cloud Datawarehousing
Practical Approach and Best Practices

As we stand, engineers have been building cloud tech based data warehouses in The Netherlands since 2015. But how do we actually design a cloud based data warehouse that really leverages the new technological capabilities the cloud brings? How does it differ from the classical, on premise data warehousing approach? And how can we go about implementing such a cloud based data warehouse technically?

Organizations are increasingly looking into the pros and cons of cloud data warehousing as the technology is evolving rapidly and more and more organisations are ‘taking the leap’. One of the main advantages of cloud analytical databases is their virtually ‘endless’, independently scalable (cheap) storage and compute. For most companies, however, migrating an existing data warehouse to the cloud is not that trivial. An important reason for this is that to take advantage of the independently scalable storage and compute, a lift-and-shift approach often does not work. The common cause for this is that the data architectural design principles that were used to design the classical data warehouse are outdated too!

In addition, the ‘modern data warehouse’ often demands a different consumption pattern than that we are used to in the classical data warehouse setup. As they are, these architectures were often focused on tactical / strategical information provisioning. However, in this day and age of data science and advanced analytics, it is only logical for the data scientist to be one of the primary users of the data warehouse. After all, people have started to realize that data scientists currently perform a lot of work that could and should be covered under the label of ‘data warehousing’. A well designed modern data warehouse can therefore offer a competitive advantage for any modern data driven / information supported company.

Seminar
In this seminar, Rogier Werschkull will explain how one should deal with the solution and data architectural changes, offering lessons learned and best practices. Should you indeed move your data warehouse to the cloud? If so, what could be the new design principles forging your ‘modern data warehouse’? How do the different cloud data warehouse vendors compare? What cloud data warehouse solution and surrounding toolset should we choose and why? And how can we start building such a new data architecture, learning from how others have done it so far?

The cloud analytical databases that will be looked at will be restricted to the major players: Amazon Redshift, Google BigQuery, Snowflake and Azure SQL Data Warehouse / Synapse. The supporting tools we will look at, fall in the category of EL tools (for onboarding data) like Fivetran and Stitch and modern ELT tools like Data Build Tools (DBT) and Matillion.

It is important to realize that while we will look at the technical aspects of these new cloud based solutions, the focus of this seminar will be to get a head start with implementing an analytics data architecture that you can really call a ‘modern data warehouse’. The essence of this is to grasp what implementing a cloud based ELT (not ETL!) analytical data architecture means in real life.

Learning objectives
Attendees will learn:
• What a data warehouse really is and what it is not. On architectural versus technology concerns
• What is ‘modern data warehousing’ about? How does it differ from ‘classical’ data warehousing, looking at consumer requirements / access patterns and the required data architecture?
• What is the role of the ‘Data Scientist’ in the context of the modern data warehouse?
• What does the cloud agnostic data warehouse data architecture look like that Rogier has been using since 2016
• What are the core data architectural principles that forge the modern data warehouse
• What is ‘separation of concerns’, the primary architectural principle that drives modern data warehouse design?
• Why data warehouse automation should play a vital role in building the modern data warehouse
• Why can a modern data warehouse be built better using a cloud (native) solution?
• What are the major differences / pro’s / cons of the most commonly used cloud analytical databases: Snowflake, Google BigQuery, Amazon Redshift and Azure Synapse

• What additional tooling can be used to build a cloud based data warehouse with and why?
• What is the physical data modelling approached used per data warehouse layer and why?
• Most important lessons learned from technical and implementation perspective.

Who is It For?
Data management and analytics program and project managers, Data management and analytics architects. Data warehouse architects, Technology strategists and planners, Data engineers, Data warehouse / ETL developers and data integration specialists. Anyone with a role in migrating legacy data warehouses to the cloud.
Data Warehousing in the Cloud - revisited
- Data warehousing revisited
  - Data warehousing throughout the years
  - Why data warehousing?
  - Reasons for failure
  - On Architecture versus Technology
  - Modern Data Warehousing
  - What is it that makes a data warehouse ‘modern’?
  - What has changed?
  - Why do it different
  - The modern Data Warehouse ‘business’ architecture
  - Data-analytics architectures compared
  - On ‘separation of concerns’
  - Dissecting some data-analytics hype words

Implementation principles and logical data architecture / information architecture
- Architecture of the ‘the modern cloud-based data warehouse’
  - ‘Business’ architecture
  - Data / implementation architectural principles

- Principle 1: Separate Data Warehouse Concerns
  - On separating data warehouse data / informational concerns
  - The modern data warehouse layers / sub-products
    - Raw historized staging layer (RHS)
    - Enriched staging layer (EHS)
    - Integration layer (INT)
    - Presentation layer (PRE)

- Principle 2: Build transformation code ‘Virtual by design’
  - Where to apply?
    - From ETL+x(TL) to EL+x(TC)

- Principle 3: Automate what that feasibly can be automated
  - Automation per data warehouse layer

- Principle 4: Data / analytics engineering is a complex, data centric form of software engineering: DataOps
  - The DataOps Manifesto
  - Principles derived from the manifesto

- Principle 5: Any physical data modeling style will do if it services the consumer
  - Suggested data modeling style per layer
• Principle 6: Implementing Data search and lineage is non-optional
• Principle 7: Build for data security and privacy by design

Comparing cloud analytical databases and supporting SAAS tools
• Why data warehousing in the cloud?
  - Outsourcing ‘work’
  - Economies of scale, differentiators
• Cloud analytical Databases overview
  - BigQuery
  - Snowflake
  - Redshift
  - Azure SQL data warehouse / Synapse
• Comparing the databases
  - Costs
  - Scaling
  - Development / deployment and maintenance
  - Privacy

• Supporting tools
  - Extract-load solutions
  - Transform-load solutions
  - CI/CD
  - Data catalogs
  - ‘Reverse-ETL’

Implementation guidelines and best practices
• Guidelines for developing all layers
  - HOW to build
  - WHEN to build / order
  - WHO is building
  - Data Warehouse Database schema setup
  - Building / coding guidelines
• Specific guidelines per layer
  - Making raw historized staging robust
  - Virtualizing and automating enriched staging
  - The ever-present complexity of the integration layer
  - The agile / demand driven presentation layer

Course description

ROGIER WERSCHKULL

ROGIER WERSCHKULL is a data-solution architect with over 20 year of consulting and teaching experience in BI, data warehousing and analytics. Functionality before technology. Constructive and active knowledge sharing. This is what he stands for. Based on these principles, he has spent the last five years as a hands-on data and solution architect developing modern cloud-based data warehouse/analytics solutions in various organisations or migrating accordingly. As ‘Head of Data’ Rogier did this same kind of work at Spilgames (web and mobile gaming), which in 2016/2017 was one of the first organisations in the Netherlands to migrate completely to the Google Cloud (a.o. PubSub / Dataflow / BigQuery). At Property Finder, he worked as a lead on the migration from Amazon Redshift to Snowflake (AWS). As of 2020, he is lead data architect at 6 Gorillas (6G), a healthcare startup that is part of Tenzinger. 6G is building a data platform for healthcare based on Snowflake (Azure) and is the first major ‘health data’ customer of Snowflake in the EMEA region. During his consultancy period at Centennium in The Hague, Rogier has delivered training on Data Architecture, Data Vault and Dimensional Modelling for years and still delivers training on data modelling. Rogier regularly speaks on data warehousing in the cloud at international conferences such as Data Modeling Zone Europe, DataCon Dubai, The Knowledge Gap and the Data Warehousing & BI Summit. He also blogs on a fairly regular basis on LinkedIn.
DATE AND TIME
The workshop will take place once or twice a year with the exact date and time available on our website.
We will send the course materials and meeting instructions well in advance as well as the invitation with hyperlink to join us online.
The seminar will start at 09:00 and lasts until 13:00 if we run two half days. The online meeting will be available at least one half hour earlier so please log in timely in order to check your sound and video settings beforehand.

VENUE
Adept Events works with several venues in and near Amersfoort and Utrecht. We strive to provide you with the location details as soon as possible. The exact venue will be on our website and in the confirmation e-mail that you will receive one week prior to the event. Always check our website prior to your departure to ensure you have the exact location and directions.
The virtual seminars and workshops will be offered through a live video stream from our video studio. The virtual classroom is equipped with multiple professional cameras and microphones and we have a two person crew available so that we can offer a similar experience to that of the traditional classroom and, more important, still offer adequate interaction with the speaker.

HOW TO REGISTER
Please register online at www.adeptevevents.nl. For registering by print, please scan the completed registration form and send this or your Purchase Order to seminars@adeptevevents.nl. We will confirm your registration and invoice your company by e-mail therefore please do not omit your e-mail address when registering.

REGISTRATION FEE
Taking part in this hands-on workshop will only cost € 625,50 when registering 30 days beforehand and € 695 per person after the Early Bird period expires (excl. 21% Dutch VAT). This also covers documentation.
Please note: this seminar can also be offered by virtual delivery or face-to-face combined with live streaming. In that case the registration fee will differ from the above. On our webpage Registration Fee you will find the fees for all available event formats.
In completing your registration form you declare that you agree with our Terms and Conditions.

TEAM DISCOUNTS
Discounts are available for group bookings of two or more delegates representing the same organization made at the same time. Ten percent off for the second and third delegate and fifteen percent off for all delegates when registering four or more delegates (all delegates must be listed on the same invoice).
This cannot be used in conjunction with other discounts.

PAYMENT
Full payment is due prior to the workshop. An invoice will be sent to you containing our full bank details including BIC and IBAN. Your payment should always include the invoice number as well as the name of your company and the delegate name.
Payment by credit card is available for attendees from countries outside the IBAN region. This is not an automated process via our website but requires a manual transaction by phone or Skype. For Credit Card payment please contact our office by e-mail or through our contact form mentioning your phone number to obtain your credit card information. Never mention your credit card details in our registration form, contact form or in e-mail messages.

Cancellation Policy
Cancellations must be received in writing at least three weeks before the commencement of the workshop and will be subject to a € 75,- administration fee. It is regretted that cancellations received within three weeks of the workshop date will be liable for the full workshop fee. Substitutions can be made at any time and at no extra charge.
Cancellation Liability
In the unlikely event of cancellation of the workshop for any reason, Adept Events’ liability is limited to the return of the registration fee only. Adept Events will not reimburse delegates for any travel or hotel cancellation fees or penalties. It may be necessary, for reasons beyond the control of Adept Events, to change the content, timings, speakers, date and venue of the workshop.

INFORMATION
+31(0)172 742680
http://www.adeptevevents.nl/cdw-en
seminars@adeptevevents.nl
@AdeptEventsNL / https://twitter.com/AdeptEventsNL
http://www.linkedin.com/company/adept-events
https://www.facebook.com/AdeptEventsNL
Visit our Business Intelligence and Data Warehousing website www.biplatform.nl and download the App
Visit our website on Software Engineering, www.release.nl and download the App

IN-HOUSE TRAINING
Would you like to run this course in-company for a group of persons? We can provide a quote for running an in-house course, if you offer the following details. Estimated number of delegates, location (town, country), number of days required (if different from the public course) and the preferred date/period (month). Please find more info on the In-house page on our website.

All prices are VAT excluded.