



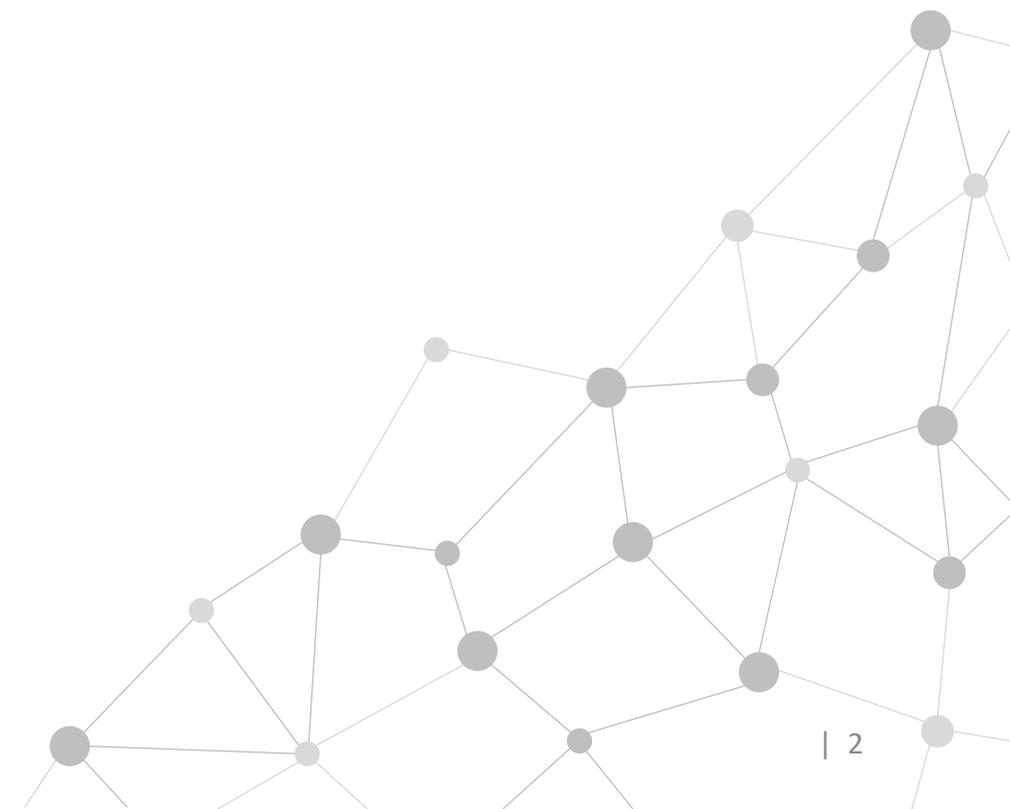
Azure Synapse data lakehouse

Customer presentation





Intro



An aerial photograph of a city street intersection, viewed from a high angle. The image is darkened with a semi-transparent overlay. On the left side, there is a vertical bar with a blue-to-orange gradient. The text is centered in the middle of the image.

OQuila helps organisations to transform to
a data-driven organisation.



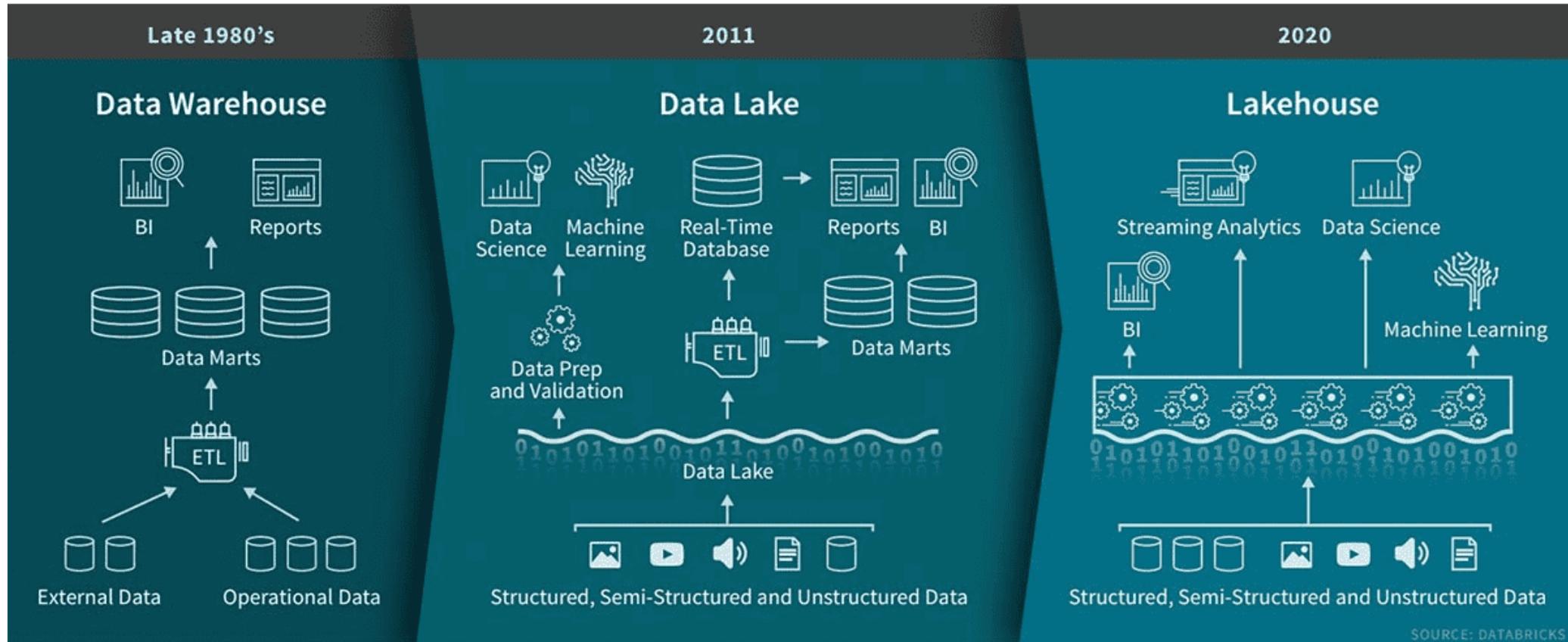
About OQuila

- *Data & Analytics, Internet of Things and Application Innovation solutions*
- Joining forces with established IT company 
- Innovation & transformation with trusted technologies

Gold
Microsoft Partner



Evolution of data platforms





Data Lake vs Data Warehouse

Data Lake

Schema on read; answers also the questions of tomorrow

Scales without limits

Can hold any type of data

≠

Data Warehouse

Schema on write; answers the questions of today

Mainly for relational data (tables and rows)

Can be part of an Enterprise data lake or lakehouse



Overview





General principles OQuila Architecture

1

Use of standard components

4

Use of components within the same ecosystem: e.g. Microsoft Azure Synapse

2

100% Cloud Services: PaaS or SaaS. No installations or Virtual Machines

5

Minimize maintenance by using Services (maintained by Microsoft)

3

No custom development

6

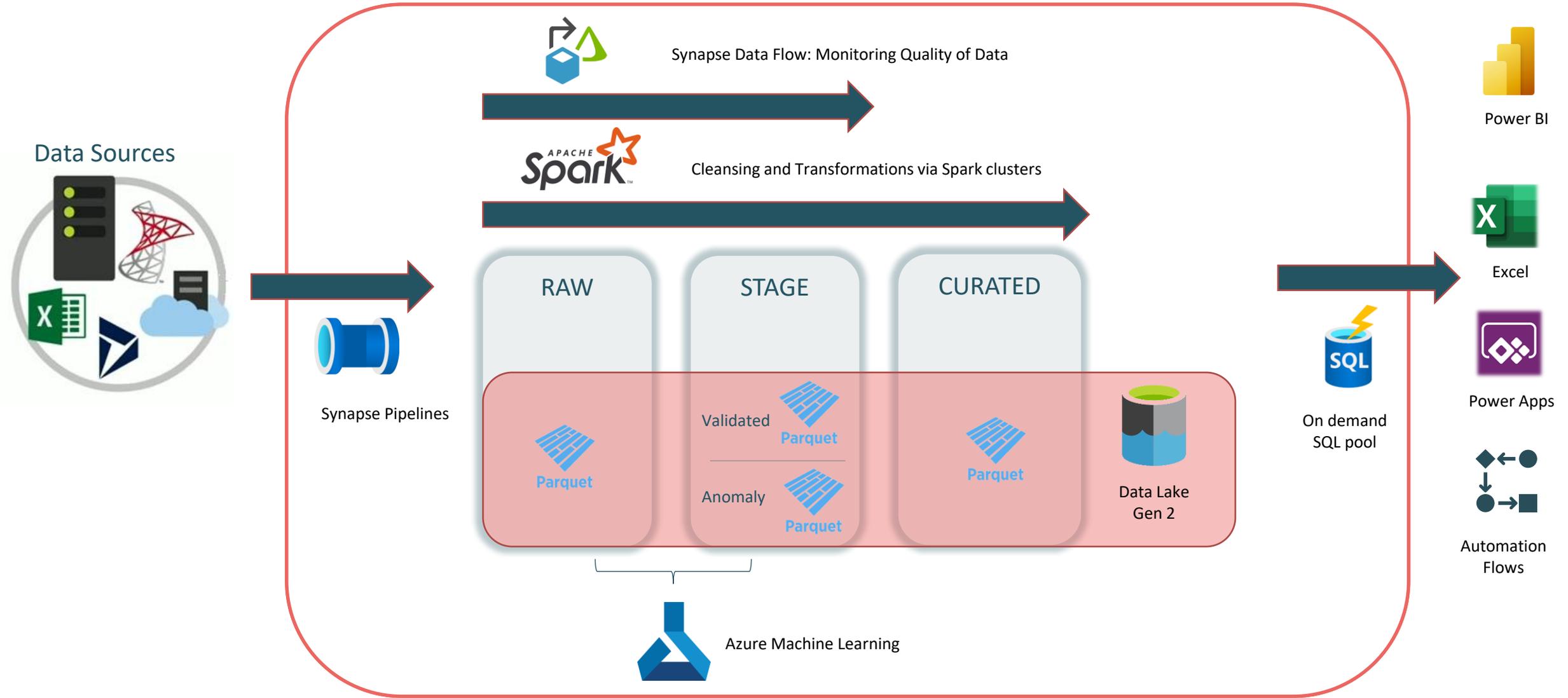
Dynamic and scalable



Agile Data Model

- No traditional schema or fixed model
- RAW, STAGED, CURATED:
 - No rework when adding additional sources
 - RAW and CURATED stores data separately
 - Preparations/calculations are done in STAGED environment and are reusable
 - Supports changes to business rules with ease
- *Schema on read; **answers also the questions of tomorrow***

Azure Synapse Analytics





Synapse components

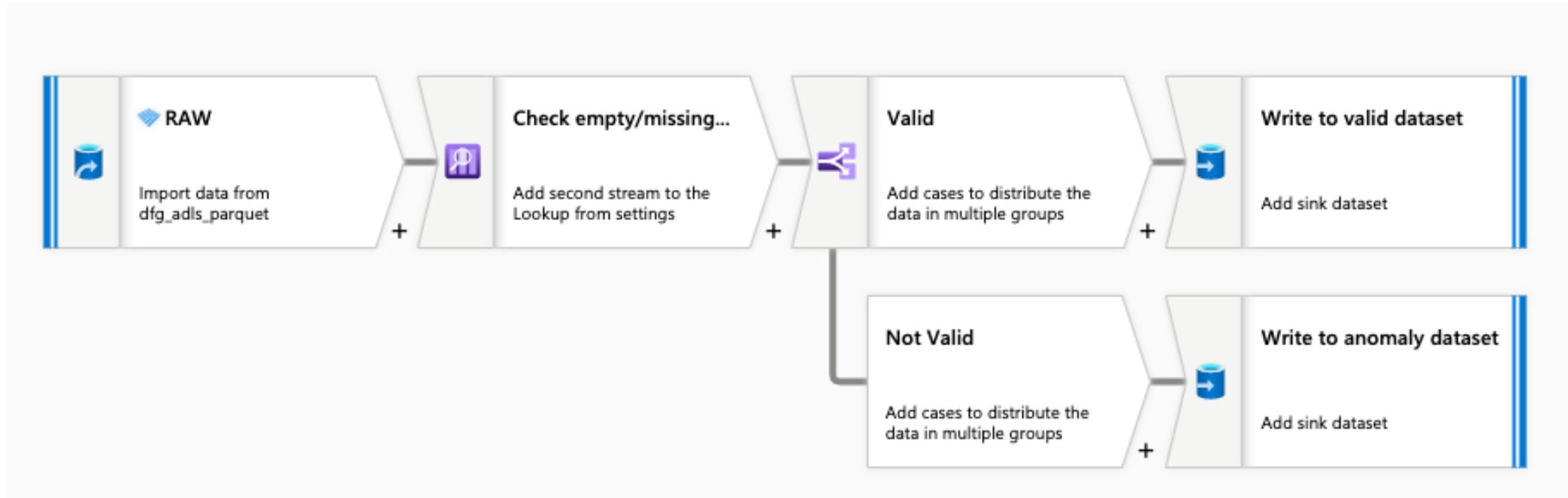
- Data pipelines:
 - A lot of standard connectors (SQL, Oracle, CSV, API, ...)
 - Data extraction from online and on-prem systems
 - Add new systems easily
- Data Lake:
 - RAW, STAGE and CURATED folders (level maturity en correctness data)
 - Parquet files to be able to work efficiently with large amounts of data
- Spark Cluster:
 - Performant transformation and cleansing actions via notebooks
 - Transfers “edited” data to the next stage (RAW, STAGE, CURATED)
- Synapse Data flows:
 - Definition business rules via graphical designer (missing values, inconsistencies, ...)
 - Puts anomalies in a separate STAGE environment



Synapse components

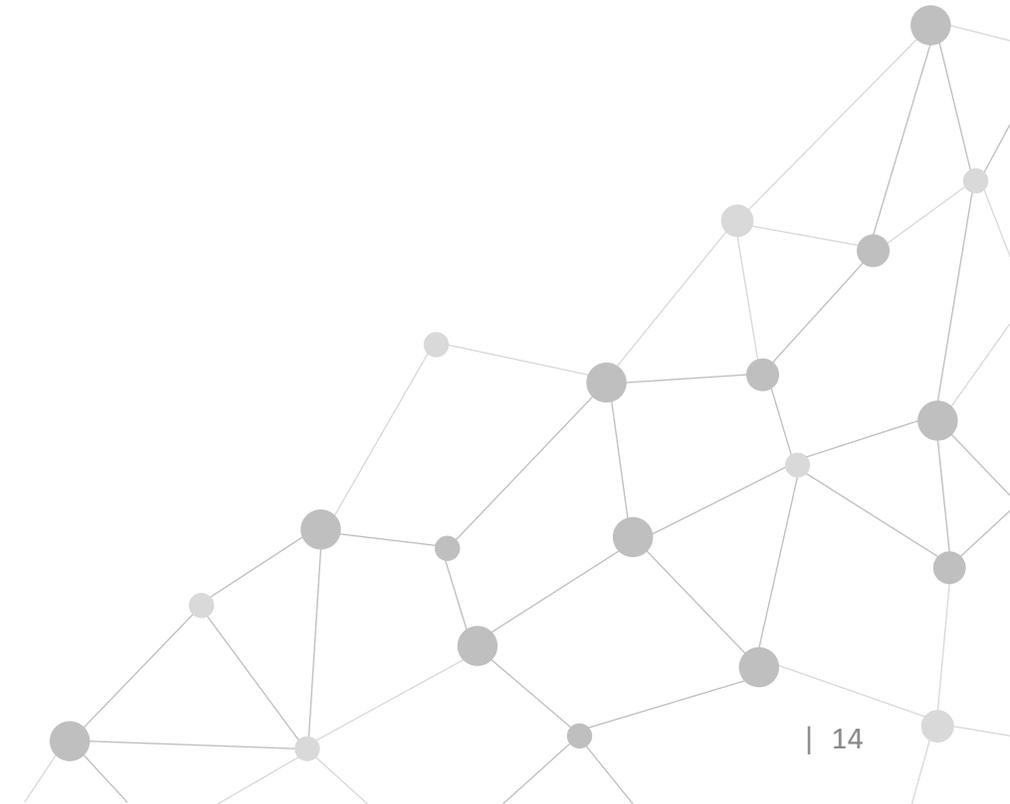
- On demand SQL Pool:
 - Build in in Azure Synapse
 - Links directly to Parquet files in CURATED zone (without having to copy data to tables).
 - Row level security
 - Allows to access data via:
 - Queries
 - Power BI
 - Excel
 - Automation tools
 - ...

Synapse Data Flow



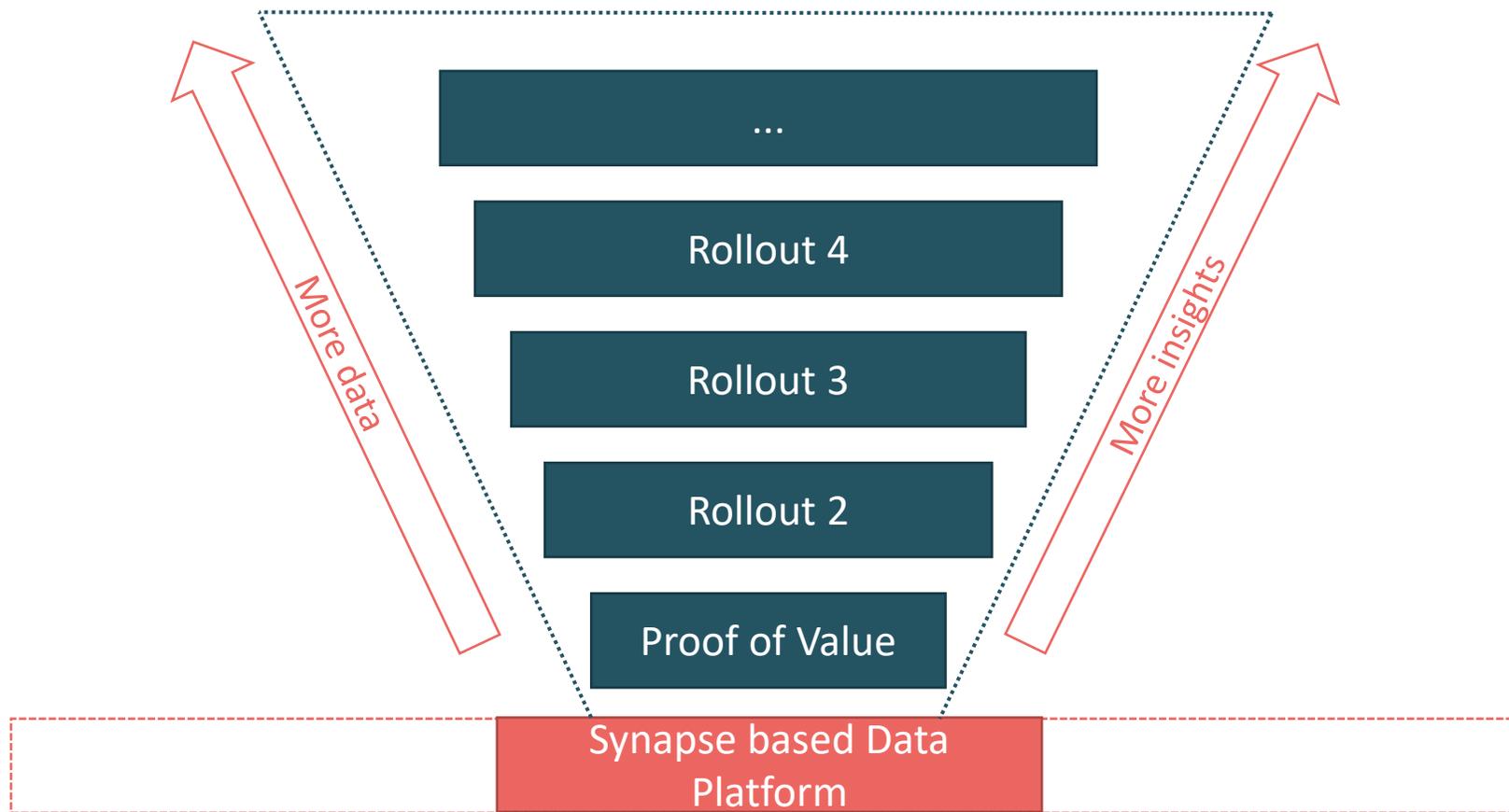


Our PoV/PoC approach





Dream Big, Start Small, Grow Fast





Proof of Concept Project approach

- Make smart choices about the scope
 - Define the 'low hanging fruit' data sources eligible for the PoC
 - Define a quick-win report
- Define a lean & mean project team
- After kick-off – OQuila will
 - Set-up the Azure environment
 - Set-up the OQuila's Synapse Data lakehouse framework
 - Set-up and deploy the selected data pipeline(s)
 - Build the report
 - Document the solution
 - Present the solution
- Ready for use and grow!

An aerial, top-down view of a city intersection. The image is dark and semi-transparent, with a grid of streets and several buildings. A prominent building with a reddish-brown facade is on the left. A tall, white, modern building is at the top center. The streets are filled with cars and trucks. The text "Thank you !" is centered in the middle of the image in a light red color.

Thank you !