

twoday capacity

Creating a full Microsoft Fabric Lakehouse in 10 minutes

Data Community Austria Day
January 2024

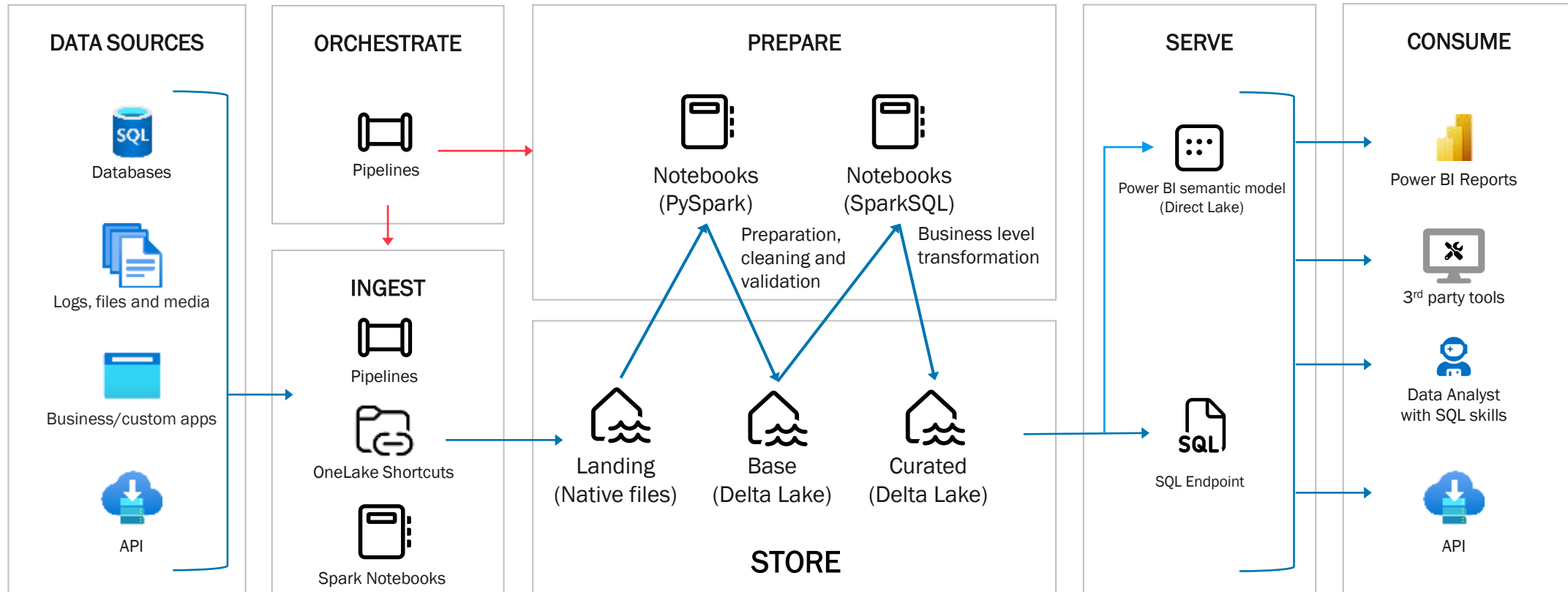




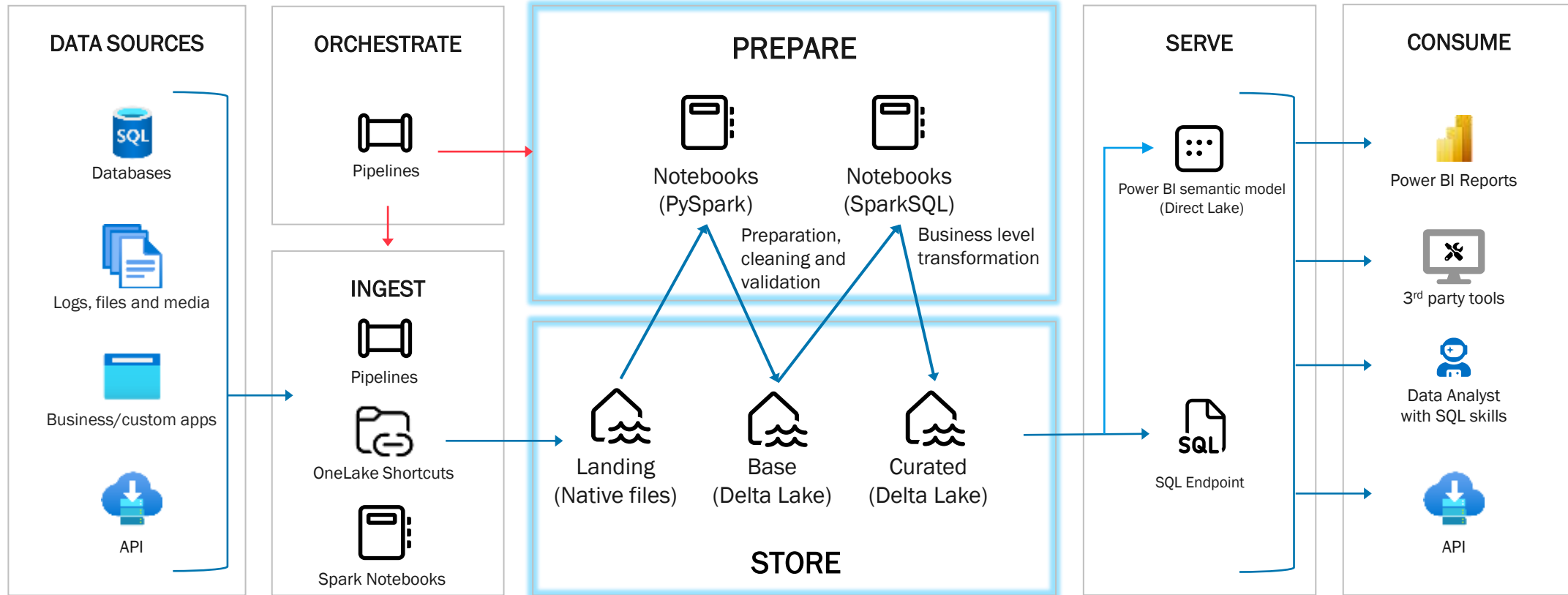
Agenda

- Architecture: Fabric Data Lakehouse
- AquaShack: Data Lakehouse accelerator
- Layers in the Data Lakehouse
- Metadata driven from Landing to Base

Architecture: Fabric Data Lakehouse



Architecture: Fabric Data Lakehouse



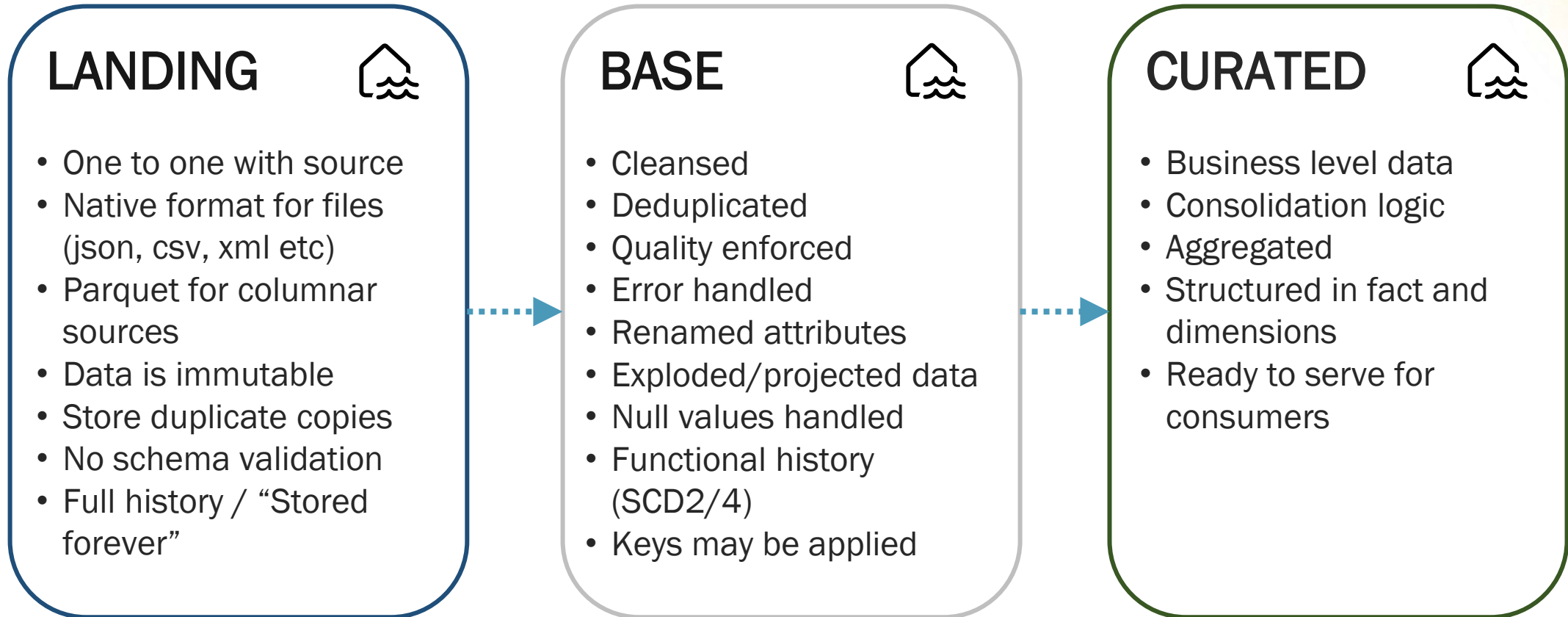
AquaShack: Data Lakehouse accelerator

- Starting point for a more advanced framework. Little brother to AquaVilla 😊
- The textbook example of a Data Lakehouse has three layers, where data is moved between.
 - Landing: Holds the data in original format if possible. Relation sources are stored in parquet. Delta is not needed here, as the layer is not intended for re-writing or querying.
 - Base: Aligned clean data, all stored in Delta tables.
 - Curated: Data for serving, business logic is applied, star schemas are defined etc.
- The notebooks builds upon functions defined in AquaShack_functions.
- We are reading metadata from JSON, it could have been from a database or YAML file as well. The important part is it ends in a common structure.
- Credit to Christian Henrik Reich



Learn more about AquaVilla: <https://www.youtube.com/watch?v=0fvLLYW5yW4> and <https://www.kapacity.com/knowledge/from-ingest-to-insights-building-robust-data-lakehouses-with-microsoft-fabric>

Layers in the Data Lakehouse



Metadata driven from Landing to Base

1. Source location

- File path
- Can be a specific file or folder

2. Source format

- Supported types are CSV and Parquet

3. Destination name

- Table name

4. Projected columns

- Default is all columns
- You can also specify specific columns

```
[
  {
    "source": "Files/External/Sample data/Sales/Customers/Customers.csv",
    "format": "csv",
    "destination": "sales_Customers",
    "projected_columns": ["CustomerID", "FirstName", "LastName"]
  },
  {
    "source": "Files/External/Sample data/Sales/Products/Products.parquet",
    "format": "parquet",
    "destination": "sales_Products",
    "projected_columns": []
  },
  {
    "source": "Files/External/Sample data/Sales/Transactions/",
    "format": "csv",
    "destination": "sales_Transactions",
    "projected_columns": []
  }
]
```


A portrait of Just Blindbæk, a man with a beard and short hair, wearing a dark blue V-neck t-shirt and dark trousers. He is standing with his hands in his pockets against a light blue background.

Just Blindbæk

- Principal Architect at twoday capacity
- Microsoft Data Platform MVP
- Found and organizer of
 - Danish Microsoft BI Community (MsBIP.dk)
 - Power BI UG Denmark (PowerBI.dk)
 - Power BI Cruise, Power BI Next Step and Data Platform Next Step



Any Questions?



twoday capacity