# Fabric for the Data Engineer

## The Data ~~Mesh~~ Mess

# Brian Bønk Rueløkke

Principal & Enterprise Architet
Fellowmind

https://linkedin.com/in/brianbonk
https://brianbonk.dk

Microsoft MVP
Microsoft Most Valuable Professional

Microsoft
FastTrack Recognized
Solution Architect
Power BI
2022 >>

Microsoft
Certified Trainer
Data Platform

2018 >>

# Table of contents

# Fabric in a flash

Data Factory

Synapse Data Engineering

Synapse Data Warehouse

Synapse Data Science

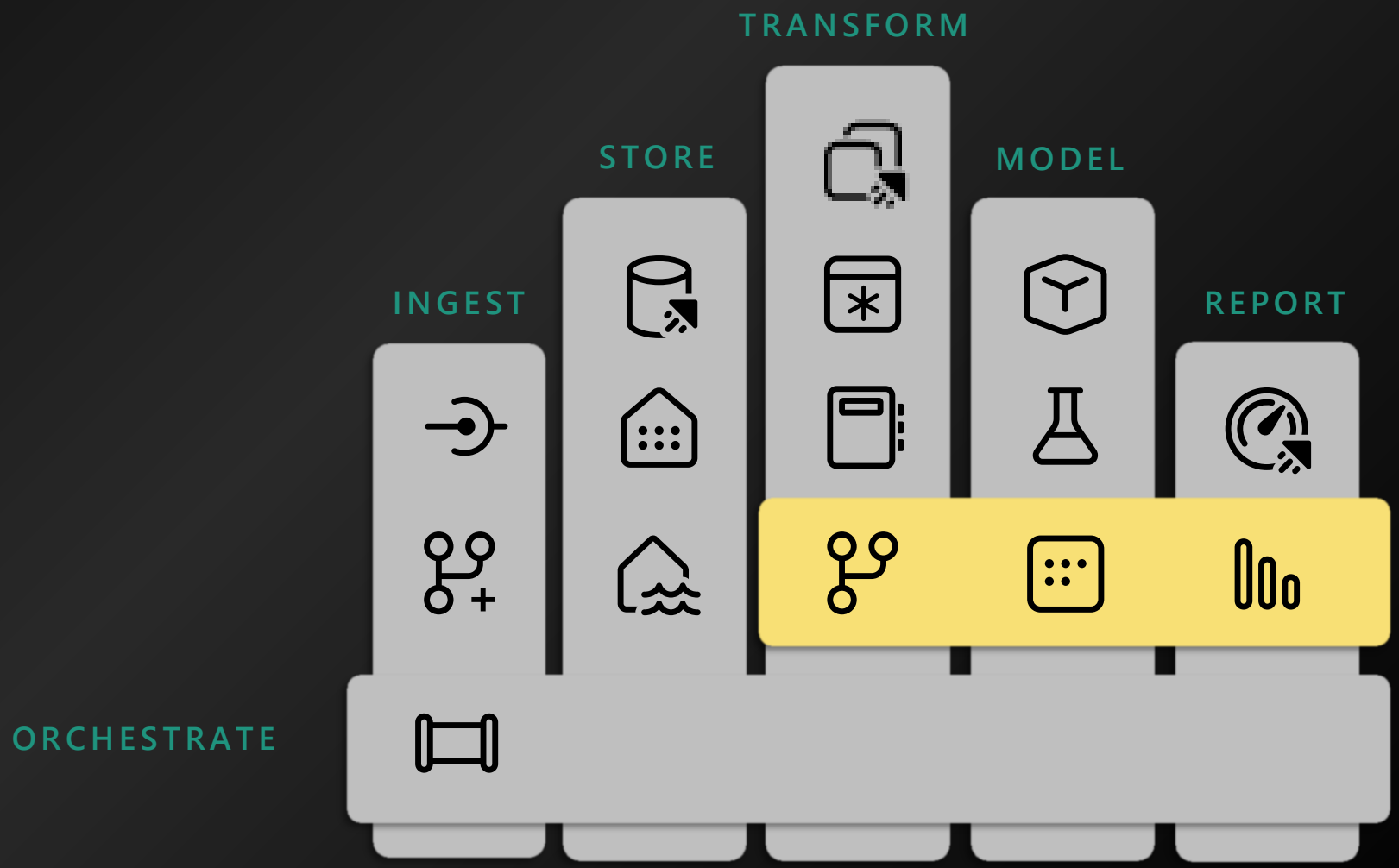
Synapse Real-Time Analytics

Power BI

Data Activator (coming soon)

OneLake

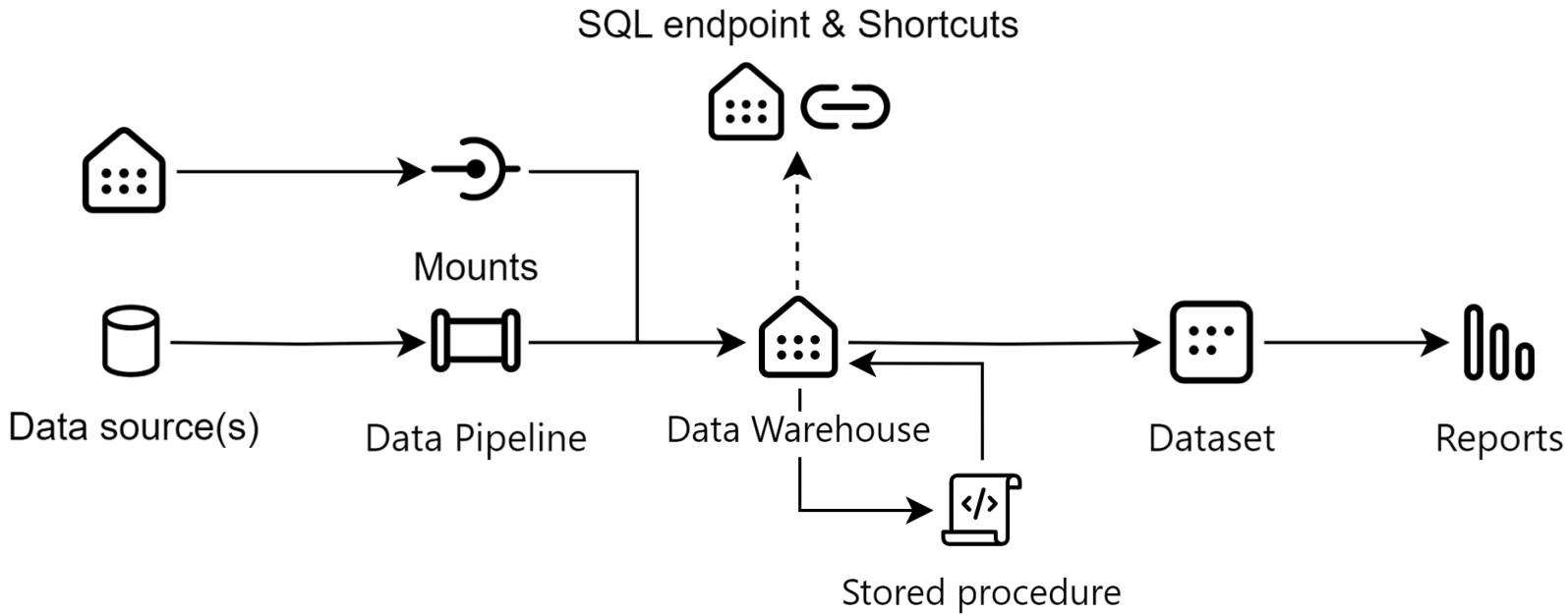
# Fabric in a flash



# The (new) services

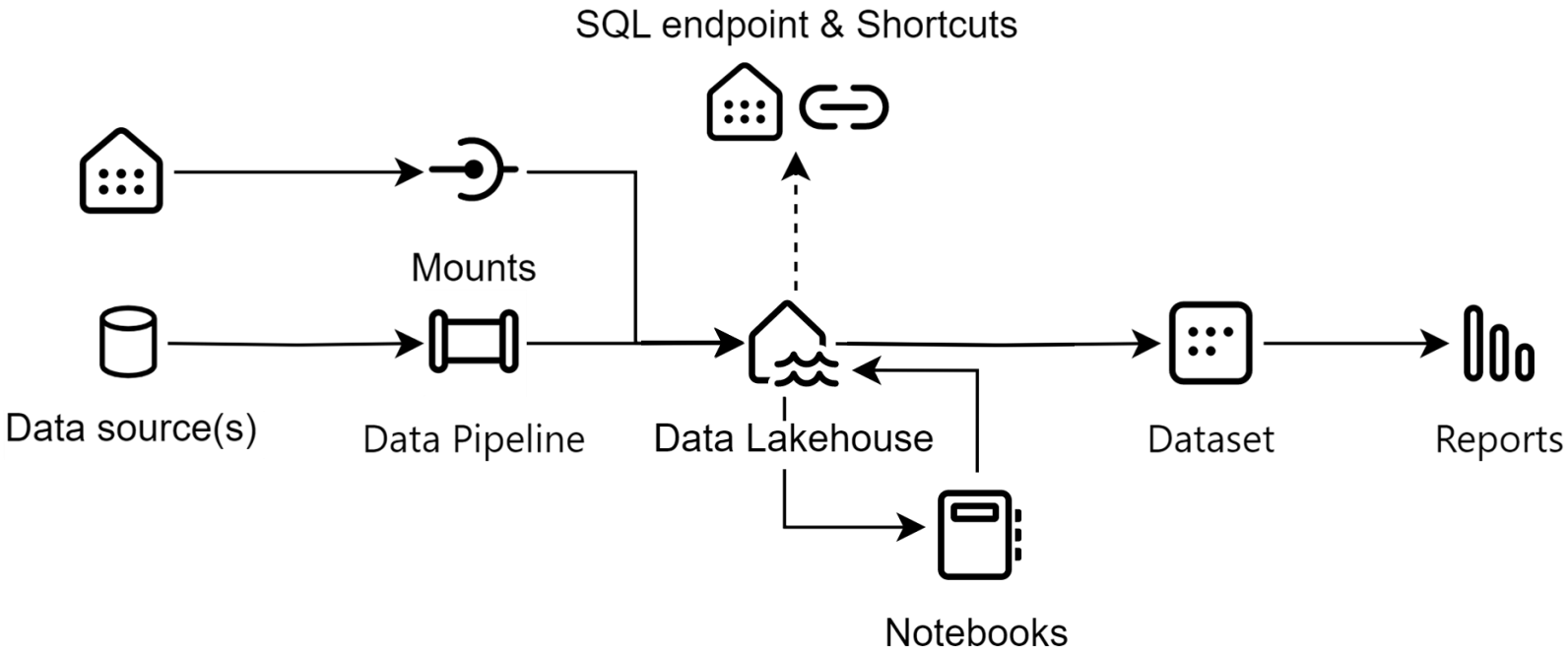
- Warehouse
- Notebooks
- Dataflows gen 2
- Lakehouse
- Pipelines
- KQL querysets

# Architecture blueprints



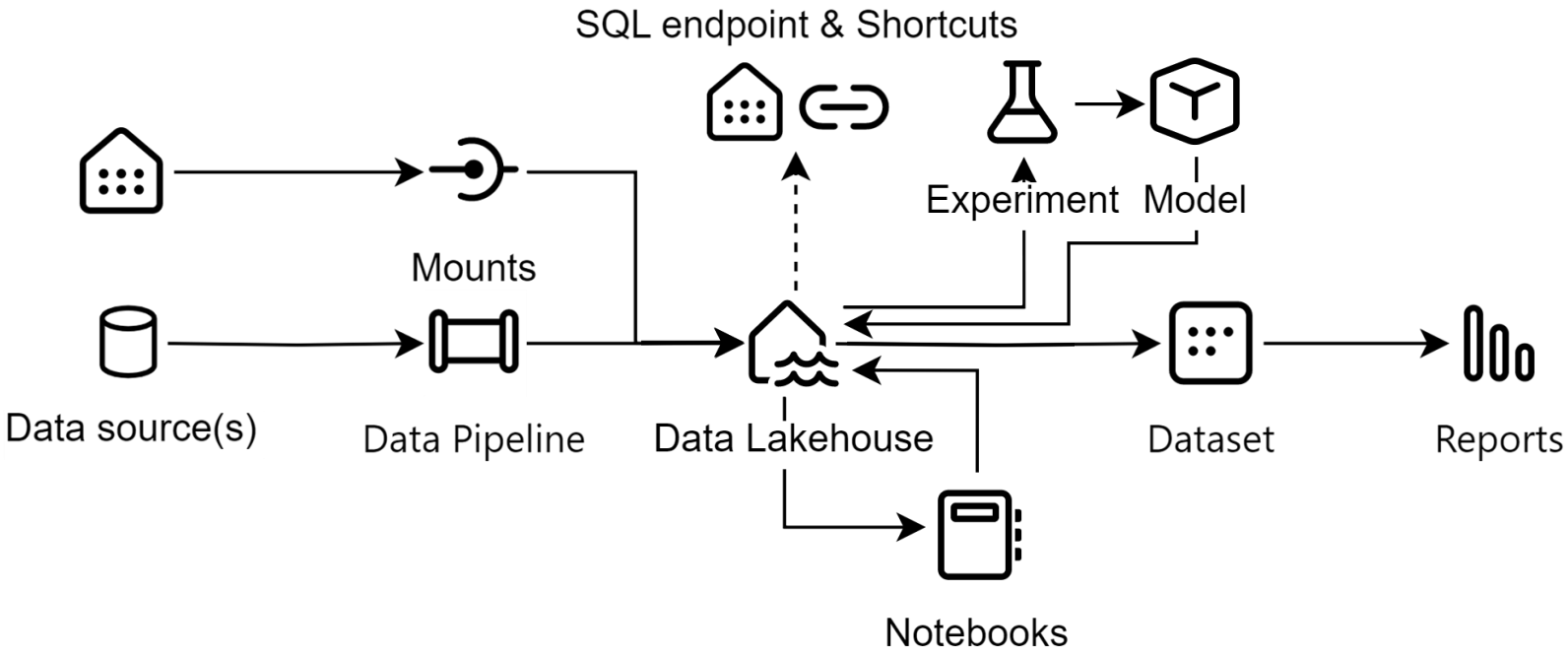

Modern SQL Data Warehouse

# Architecture blueprints




Data Lakehouse

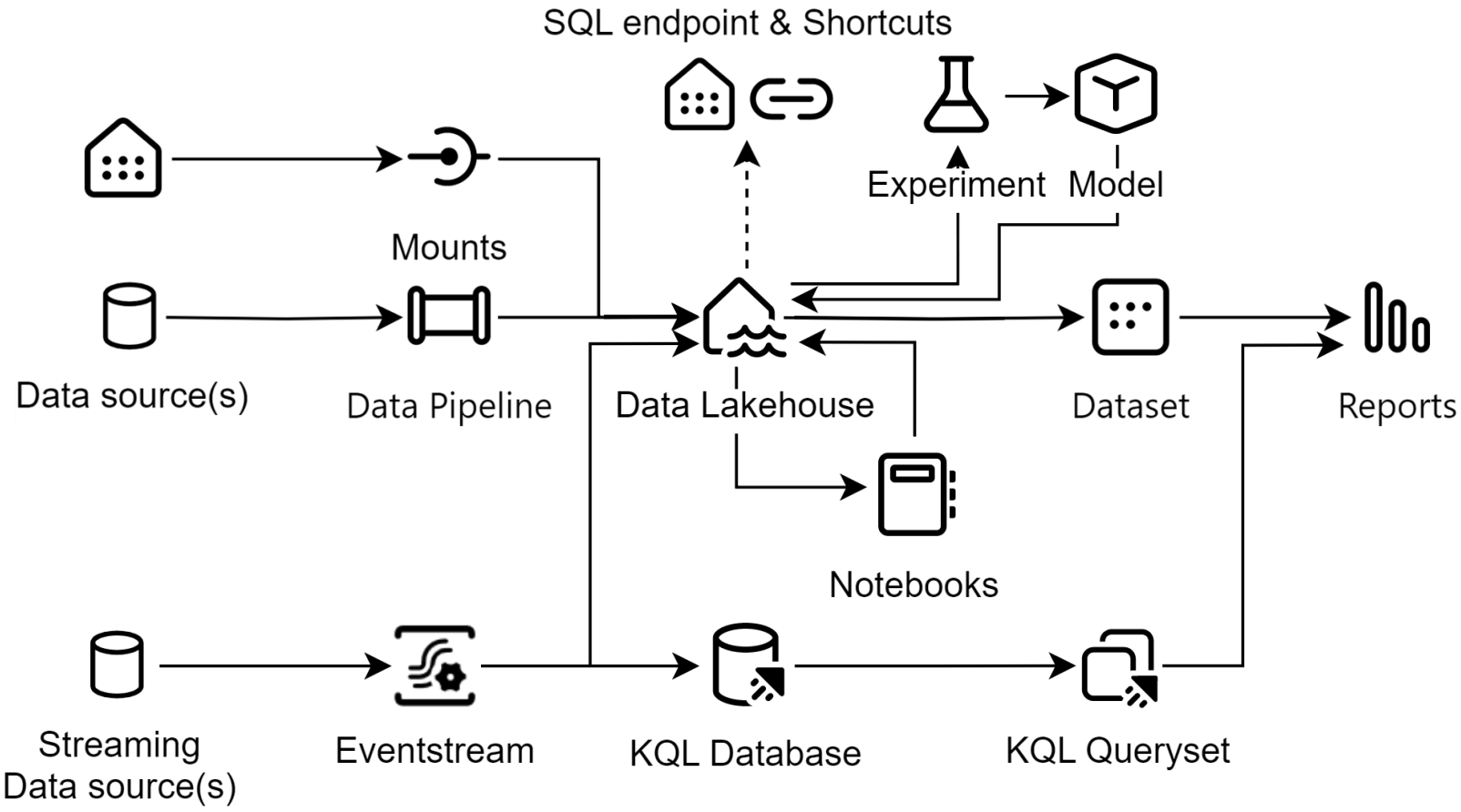
# Architecture blueprints



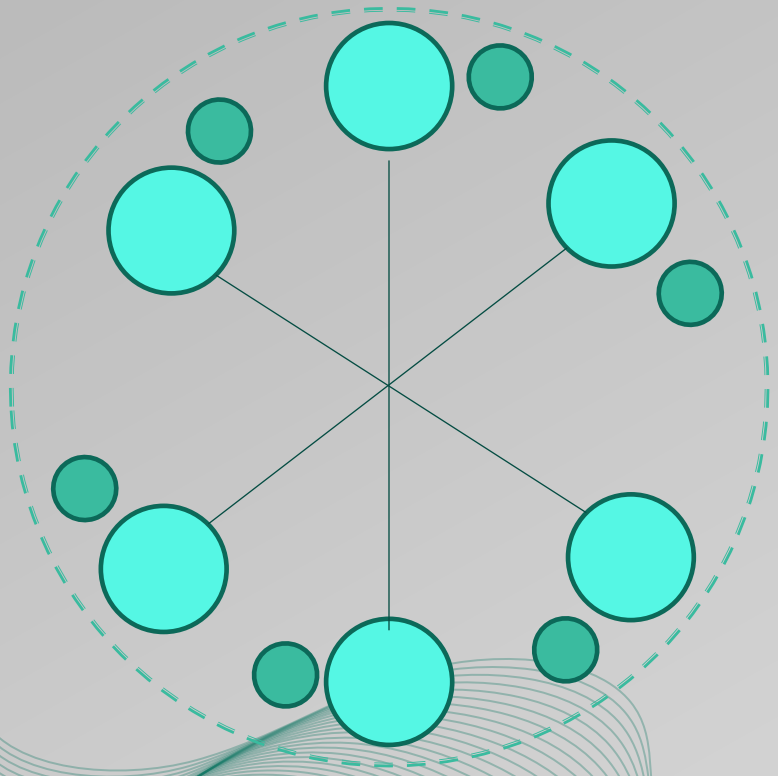

Data Lakehouse
+ Machine Learning Workflow

# Architecture blueprints

Data Lakehouse
+ Machine learning workflow
+ Real-Time Analytics



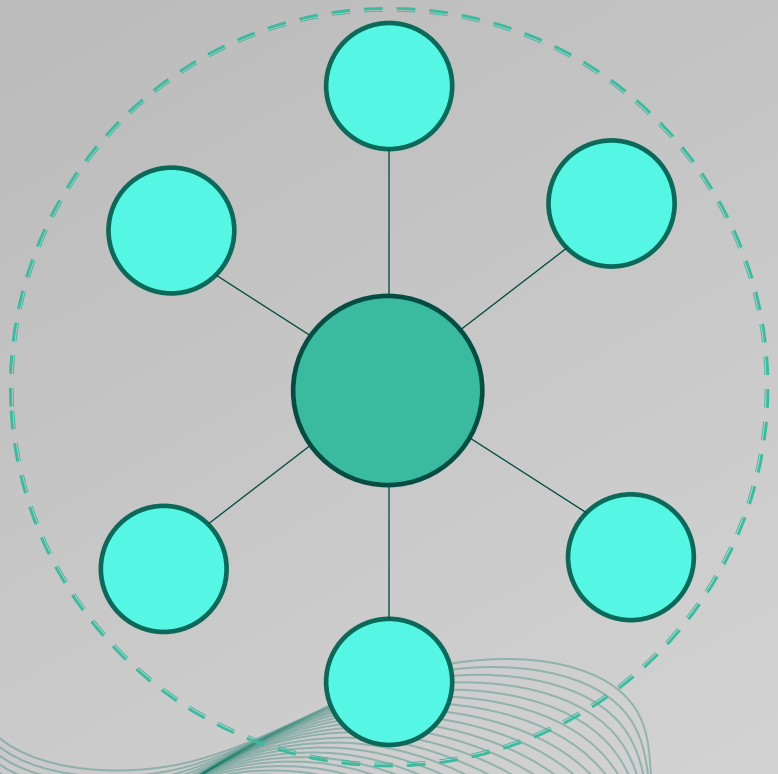
# Organization



**Self-service**

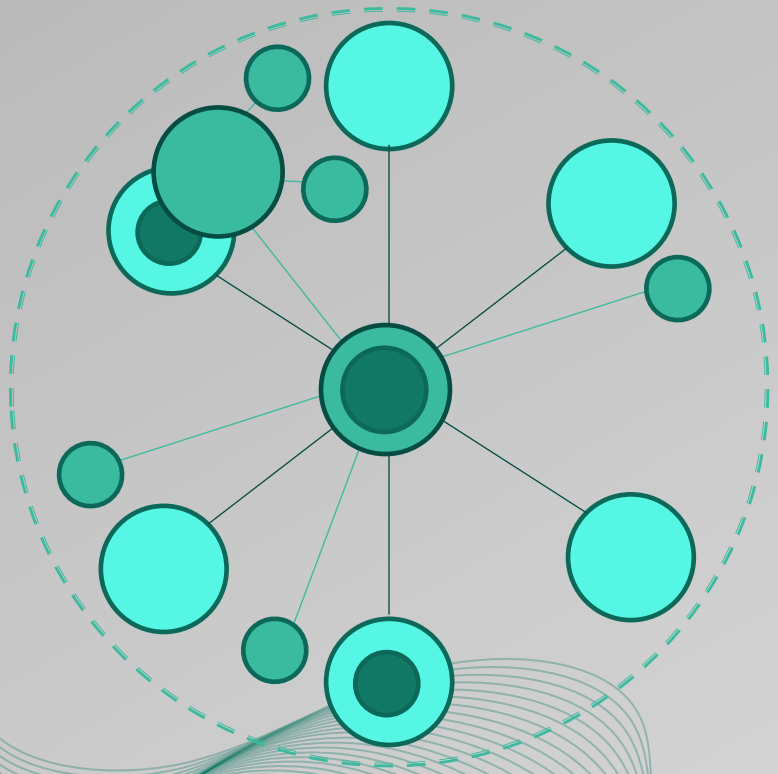
(Decentralized)

# Organization



**Enterprise**
(Centralized)

# Organization

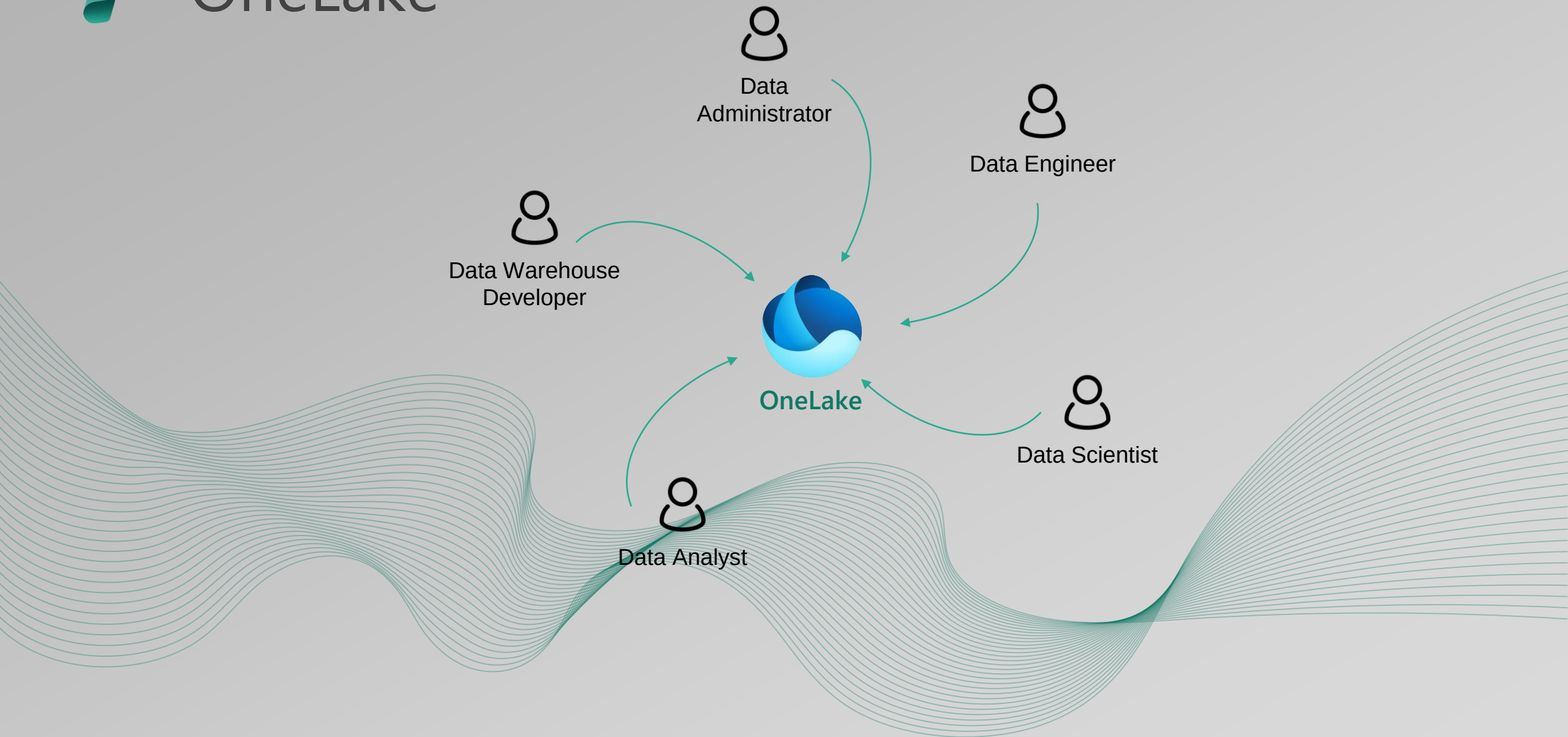


**Hybrid**

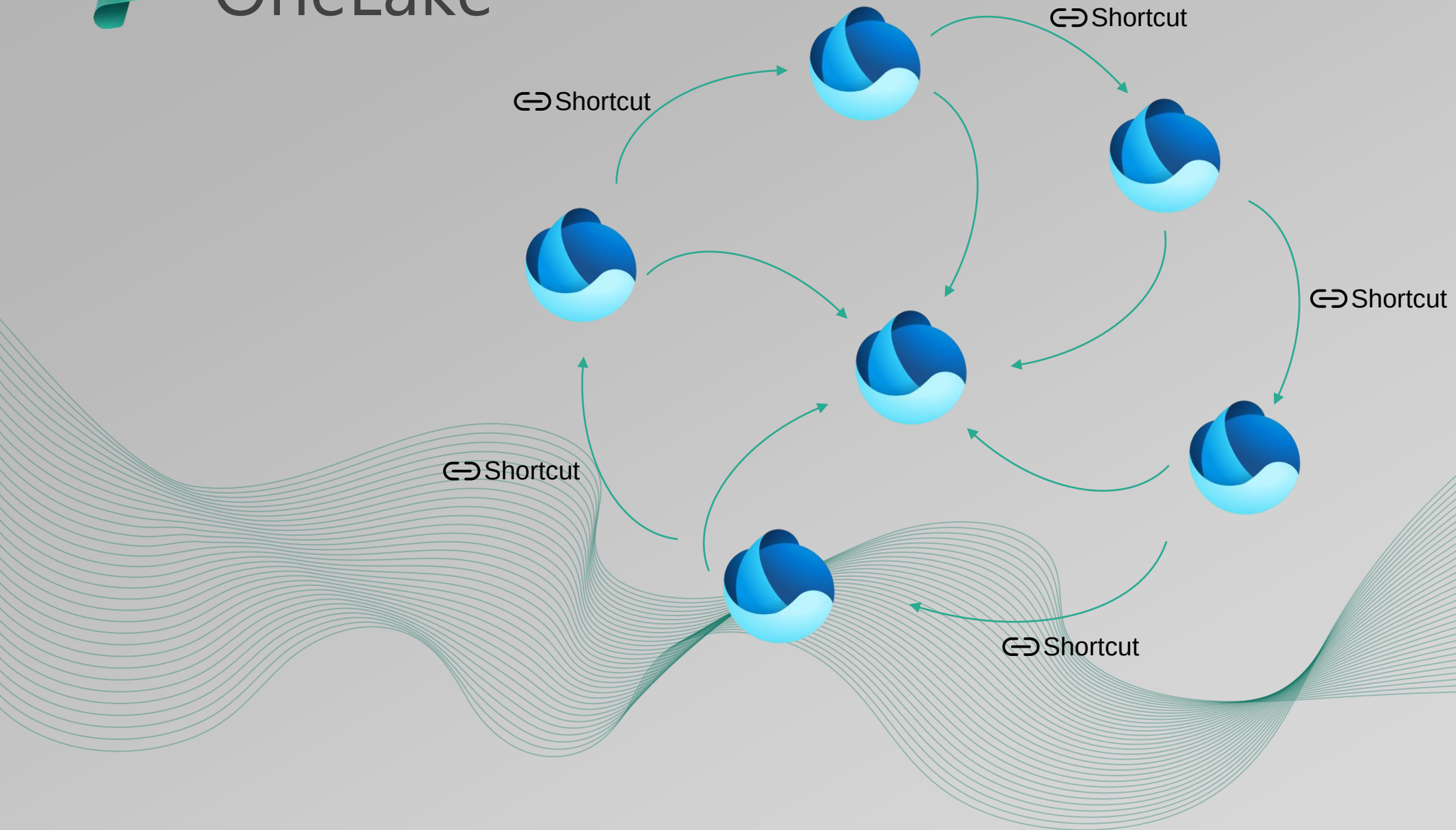
(Mixed)

# OneLake

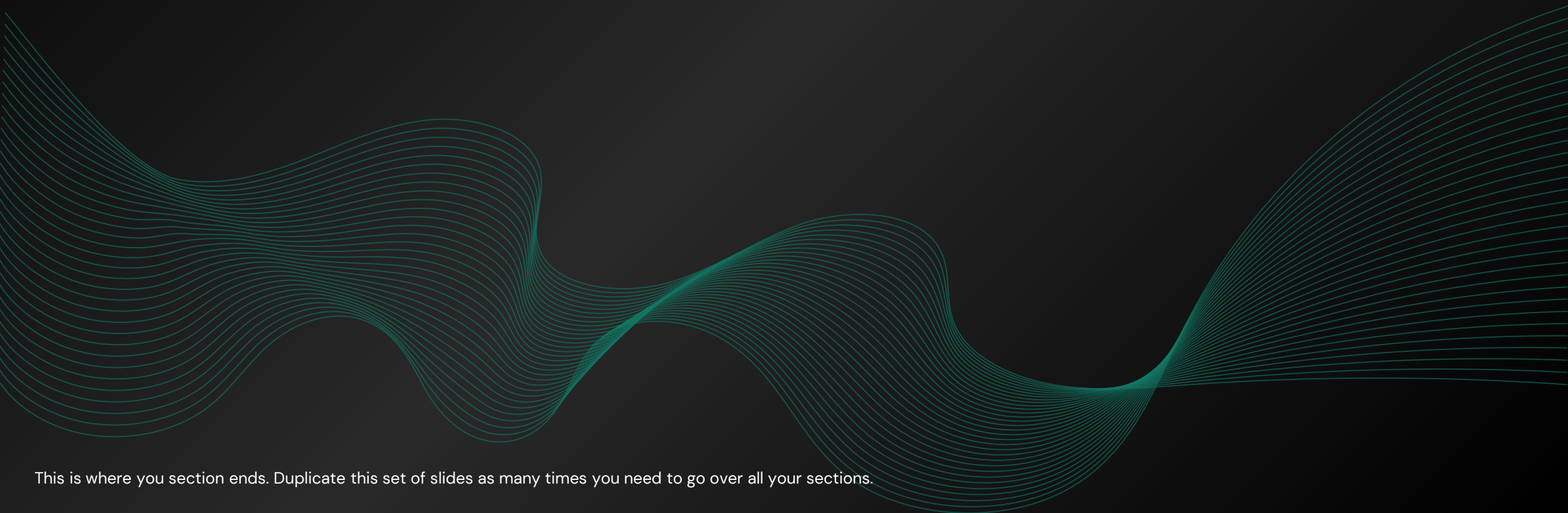
# OneLake


Data Administrator
Data Engineer
Data Warehouse Developer
OneLake
Data Scientist
Data Analyst

# OneLake



# Demo time



This is where you section ends. Duplicate this set of slides as many times you need to go over all your sections.

# LakeHouse SQL Endpoint vs Warehouse

| SQL Engine | |
|---|---|
| LakeHouse | Warehouse |
| Tables in Delta Parquet | Tables in Delta Parquet |
| Defaults to CS collation | Defaults to CS collation |
| Supports views | Supports views |
| | DDL / DML |

Best practices for serverless SQL pool - Azure Synapse Analytics | Microsoft Learn

# Database collation

Default (and for now, only) database collation is Latin1_General_100_BIN2_UTF8

## Use proper collation to utilize predicate pushdown for character columns

Data in a Parquet file is organized in row groups. Serverless SQL pool skips row groups based on the specified predicate in the WHERE clause, which reduces IO. The result is increased query performance.

Predicate pushdown for character columns in Parquet files is supported for Latin1_General_100_BIN2_UTF8 collation only. You can specify collation for a particular column by using a WITH clause. If you don't specify this collation by using a WITH clause, the database collation is used.

# Brian Bønk Rueløkke

Principal & Enterprise Architet
Fellowmind

https://linkedin.com/in/brianbonk
https://brianbonk.dk

Microsoft MVP
Microsoft Most Valuable Professional

Microsoft
FastTrack Recognized
Solution Architect
Power BI
2022 >>

Microsoft
Certified Trainer
Data Platform

2018 >>