

Roche's Maxim of Data Transformation - by Example

Alexander Arvidsson | @arcticdba





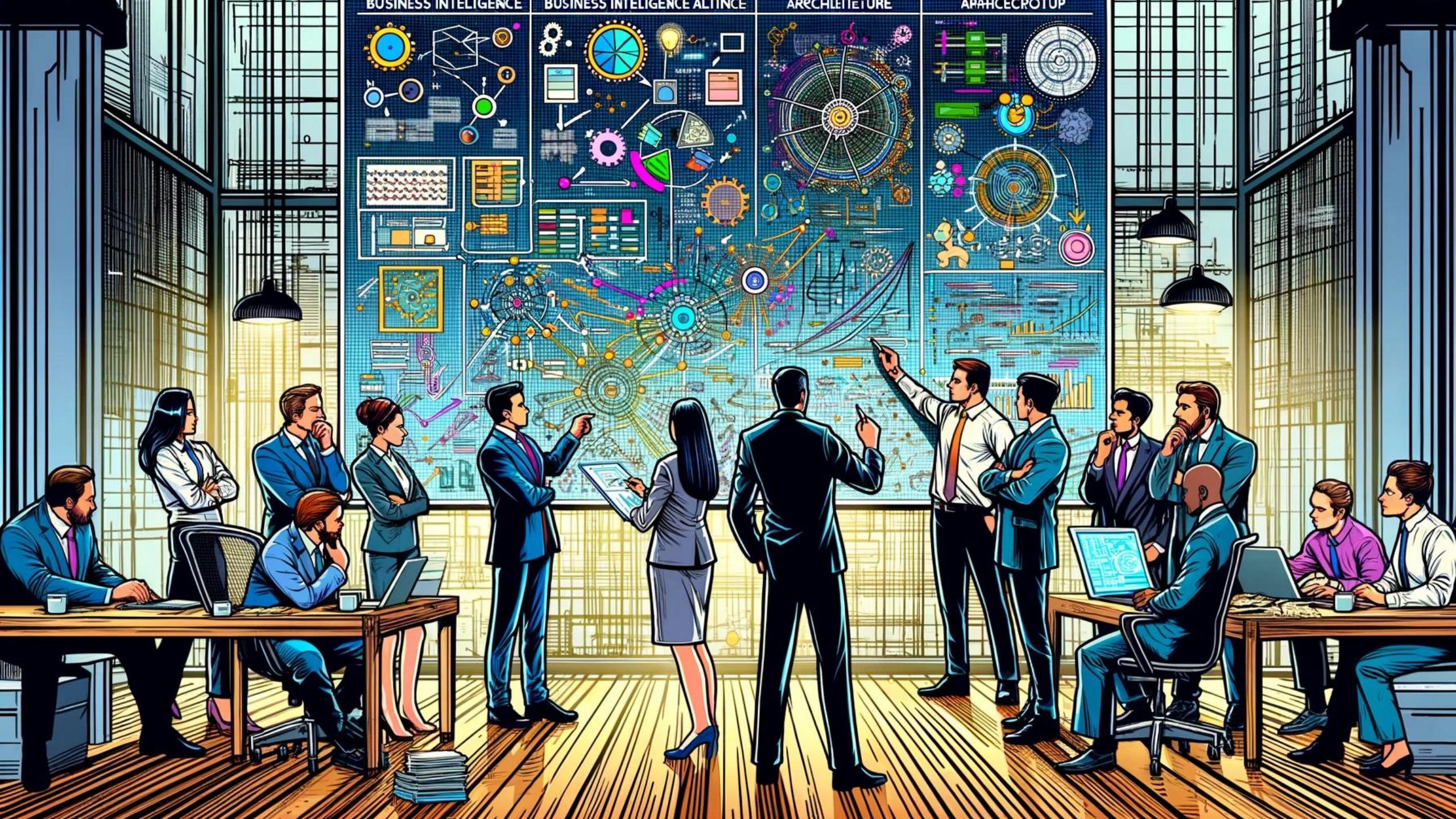




Discuss data engineering
challenges

Explore Roche's Maxim

Apply Roche's Maxim





Alexander Arvidsson

(he/him)

Principal Solutions Architect @
Data Masterminds

| Data Platform MVP | MCT | Speaker |



DATAmasterminds

KNEE-DEEP
IN
TECH

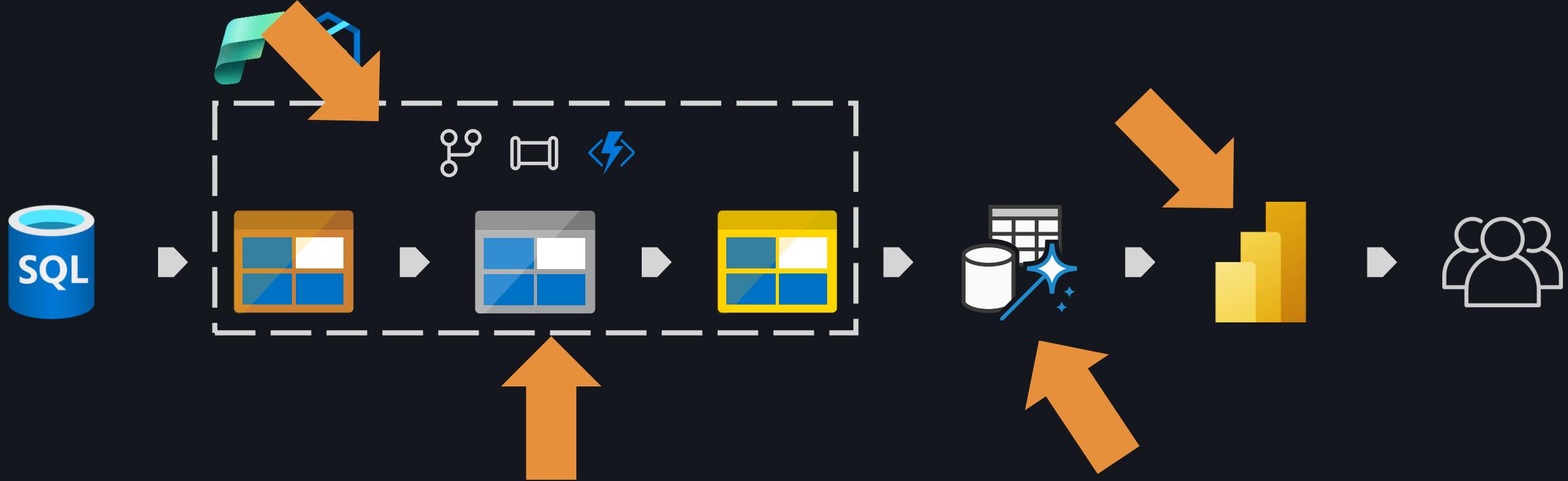


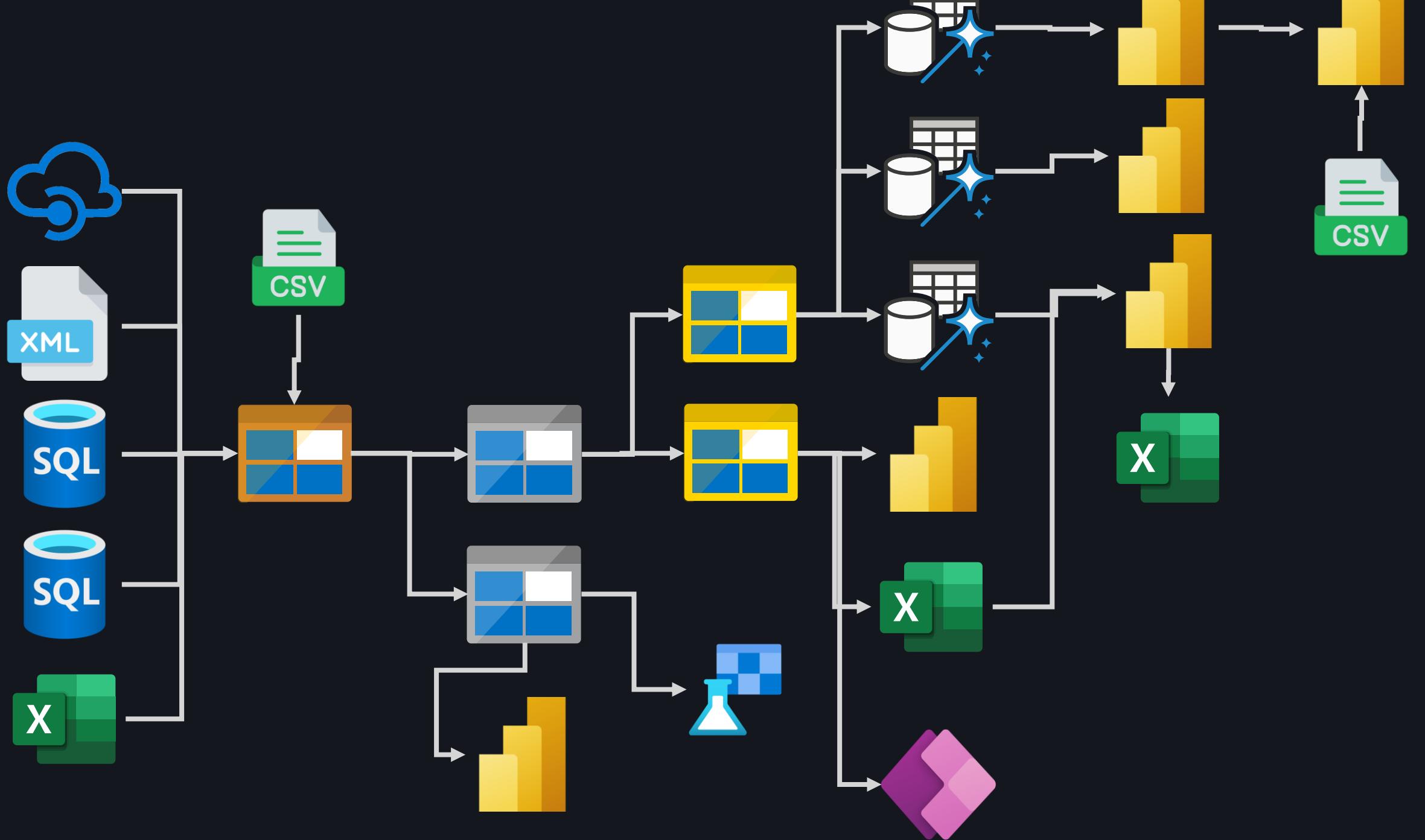


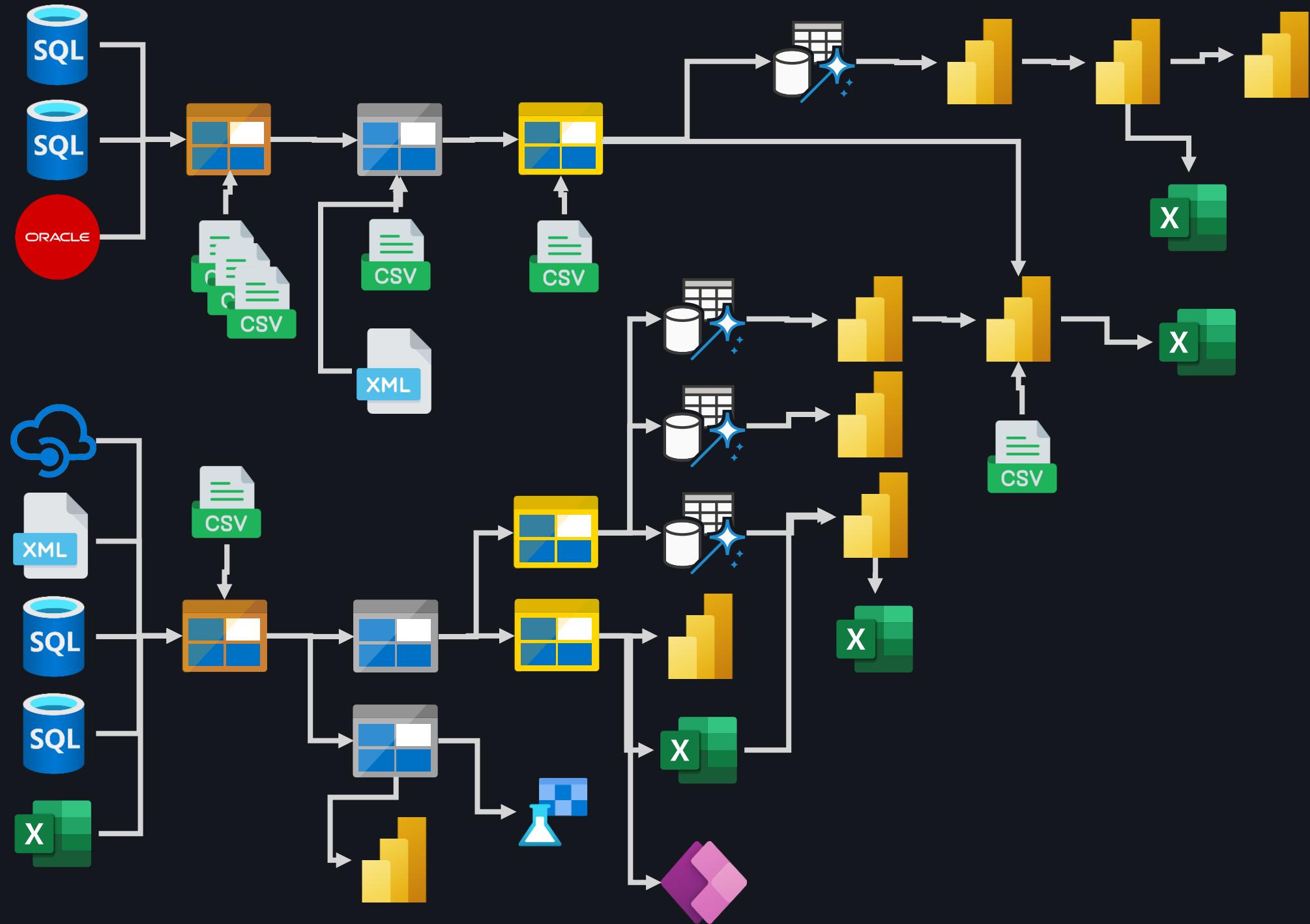
Herd^{ing} **cats** of all
shapes and **sizes**,
sleeping and **awake**, from
where they **do not** want to
leave to where they do not
want to **go** -
and making them **fit** in a **box**.













Modern
Data
Warehouse

Relational data **sources**

Relational data **sources**

SQL Server

Oracle

PostgreSQL

mySQL, etc.

Non-relational data **sources**

Non-relational data **sources**

Flat files

API calls (REST)

JSON / XML



Doing data engineering

SQL

Python

ADF/Synapse/Fabric pipelines

Azure functions

Power Query

DAX

“Data should be **transformed**
as far upstream as **possible**,
and as far downstream as
necessary.”

- Matthew Roche

What is a **maxim**?

A **maxim** is a short, often quirky statement **expressing** a general **truth, principle** or a **rule of conduct**

The man is neither.

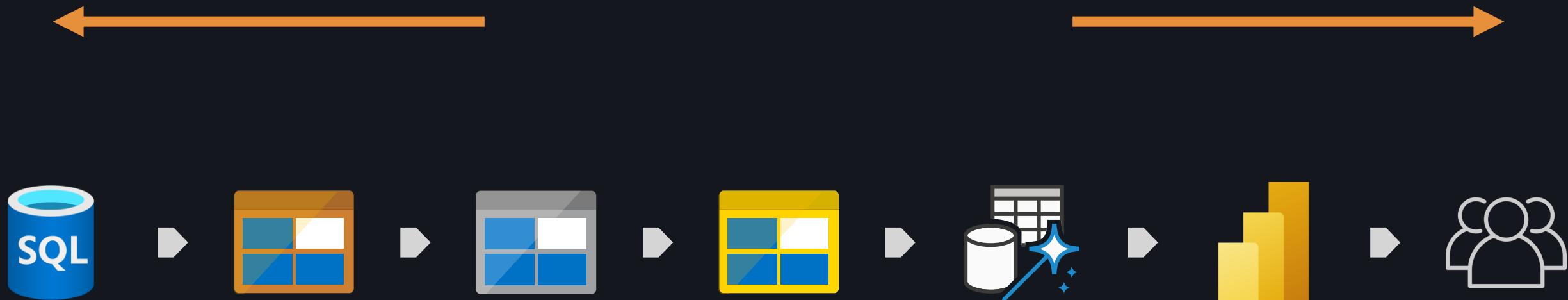
Who the heck is
Matthew Roche?



Define
upstream and **downstream**

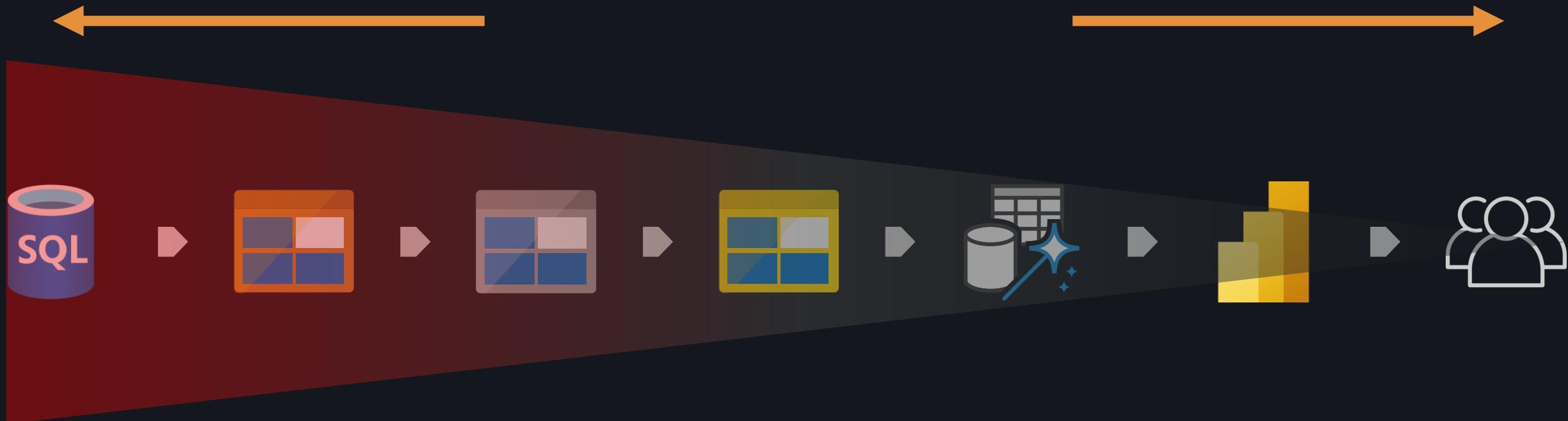
Upstream

Downstream



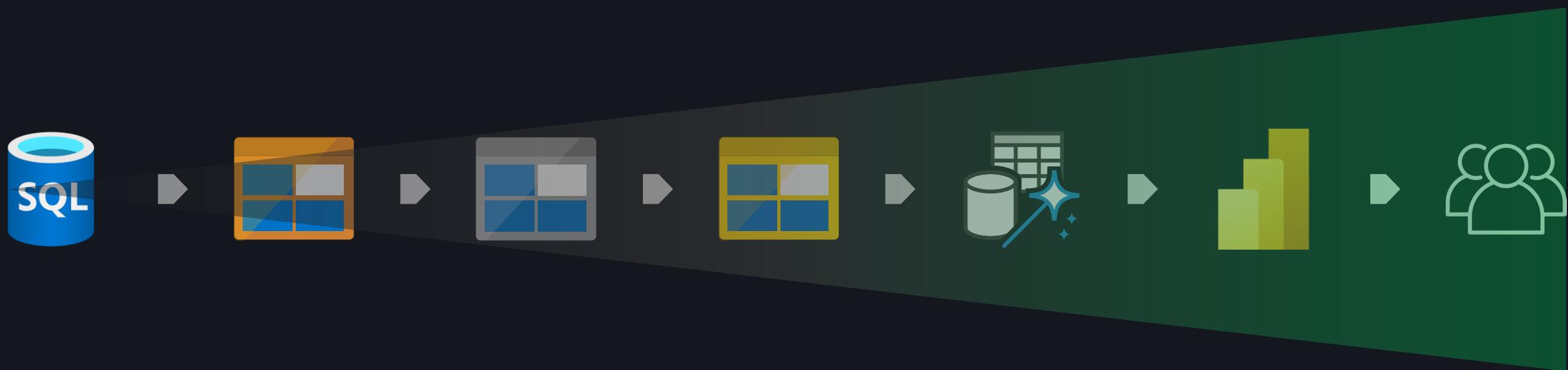
Upstream

Downstream



Upstream

Downstream



As far **upstream** as **possible**

As far **upstream** as **possible**

Skills

Time

Priorities

Available resources

Access

Organizational reasons

As far downstream as necessary

As far **downstream** as **necessary**

Skills

Time

Performance

Access

Data not available upstream

Organizational reasons

“You’ve got to pay the **cost** of
data transformations;
your only **choice** is **when** and
where you pay that **cost.**”

- Stuart Box

“The **cost** of data transformation might not necessarily be **technical**.”

- Alexander Arvidsson



Benefits

Performance

Cost

Maintenance

Consistency

Reusability

Efficiency

Simplicity

Example:
Ancient **Oracle**

Example:
Improving **readability**

Example:
Trend **analysis**

Example:
JSON **shredding**

Example:
The double-sided sword
of **APIs**

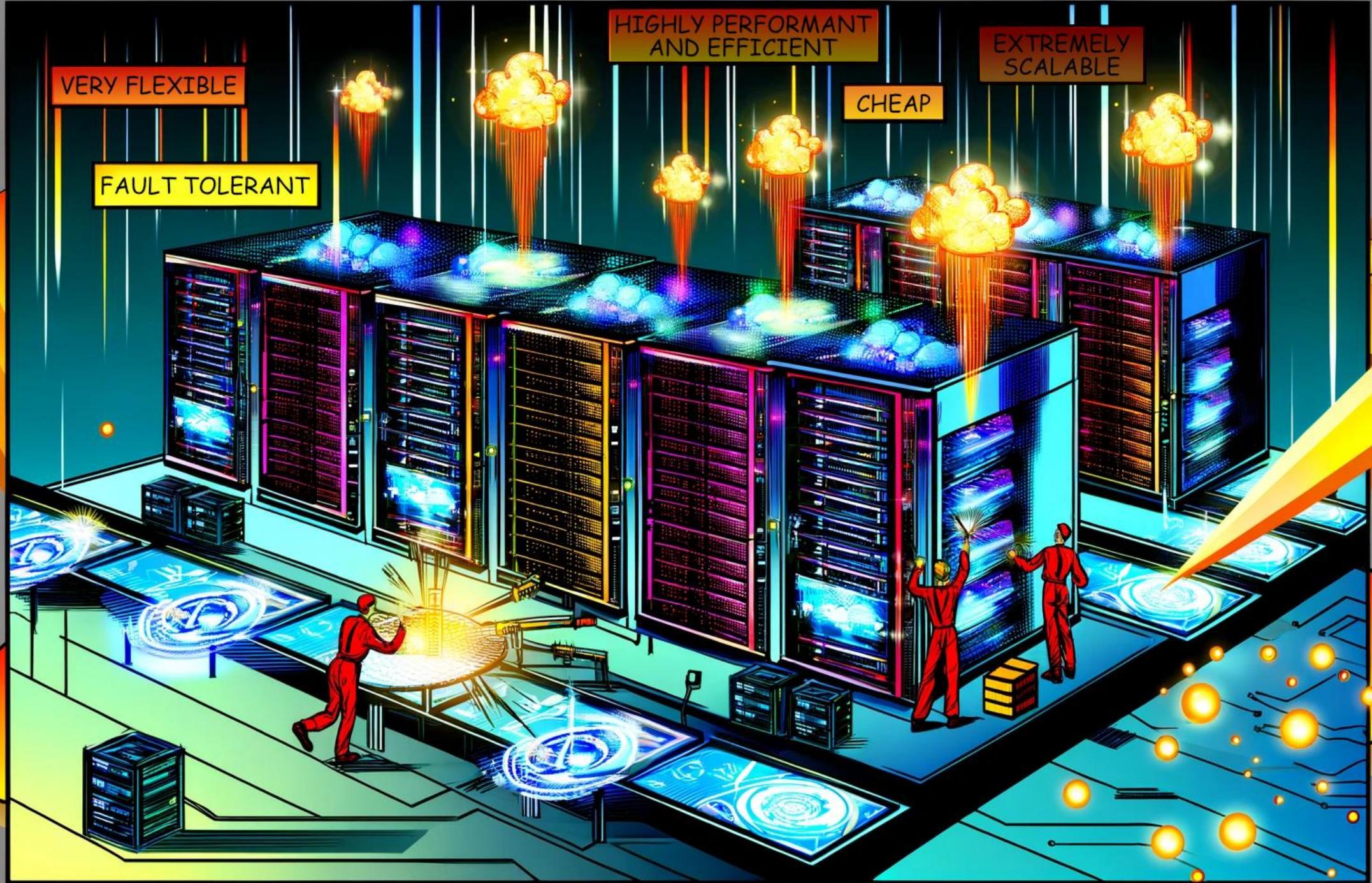
Example:
Flat files for context

The **maxim** applies equally well
to **people** and **processes**

It still depends

Cost comes in many shapes

Roche's maxim is an
excellent compass







Thank you!

@arcticdba
alexander@arcticdba.se