Microsoft Fabric

Data Engineering Game Changer or Just a Fad?

Vienna Data Engineering Meetup, 15-10-2024

Tim Bossenmaier

This is me

Tim Bossenmaier

💼 Data Engineer @ cloudflight

🏔️ located in Innsbruck

👨🏻‍💻 Open-Source Contributor
Apache StreamPipes
Datahub

bossenti

Tim Bossenmaier

cloudflight

Today's Agenda

01

Motivation:
A Data Platform for Cloudflight

02

Overview of MS Fabric

03

Painting an Ideal World
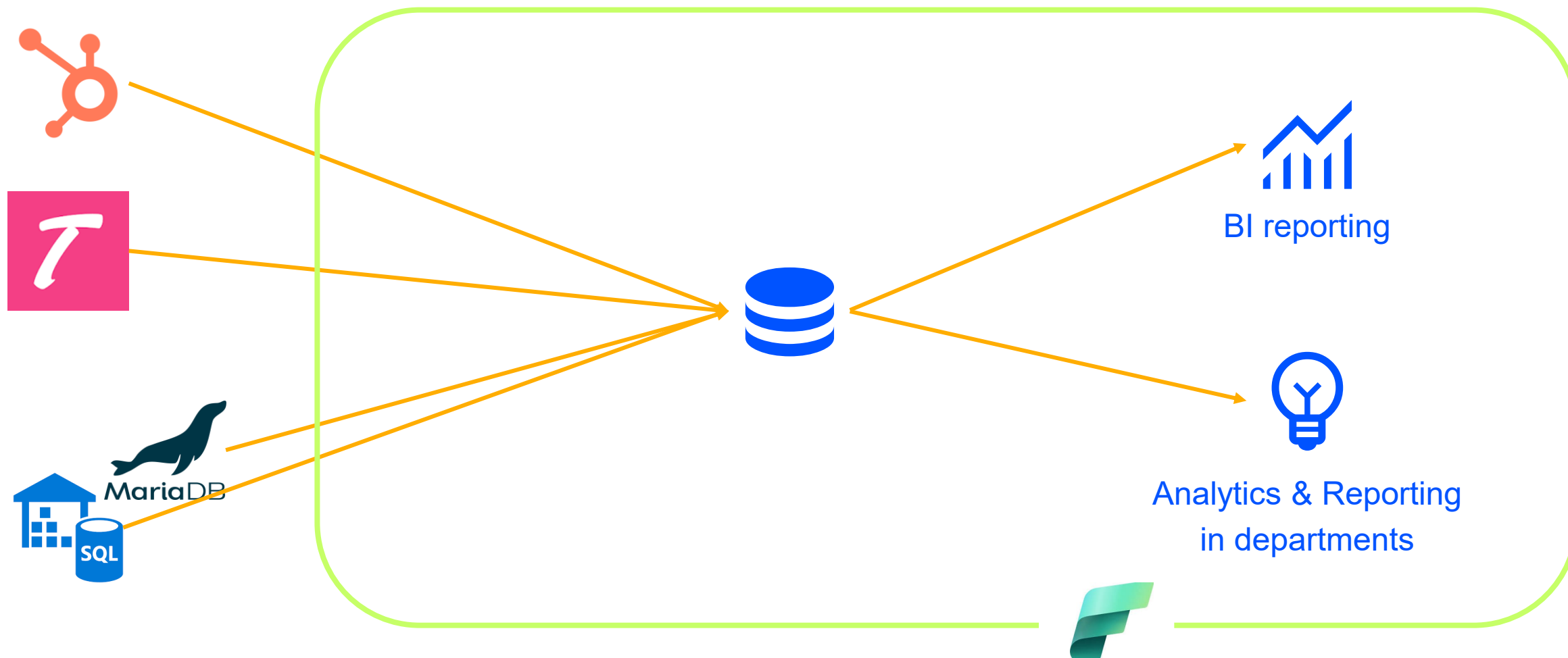
04

Problems Arise

05

How We Arranged

06

Conclusion & Discussion

01

Motivation

Goal: A central data platform for Cloudflight

BI reporting

Analytics & Reporting
in departments

02

Overview MS Fabric

New Kid on the Block: MS Fabric

Microsoft Fabric is an end-to-end analytics and data platform designed for enterprises that require a unified solution. It encompasses data movement, processing, ingestion, transformation, real-time event routing, and report building.
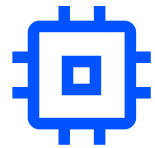
➔ A one-stop-shop for data use cases

➔ SaaS solution with a lot of No/Low Code functionalities

➔ Our internal platform is the perfect opportunity for us to experiment with Fabric and assess whether it is something we can recommend for use in customer projects

A collection of Azure products
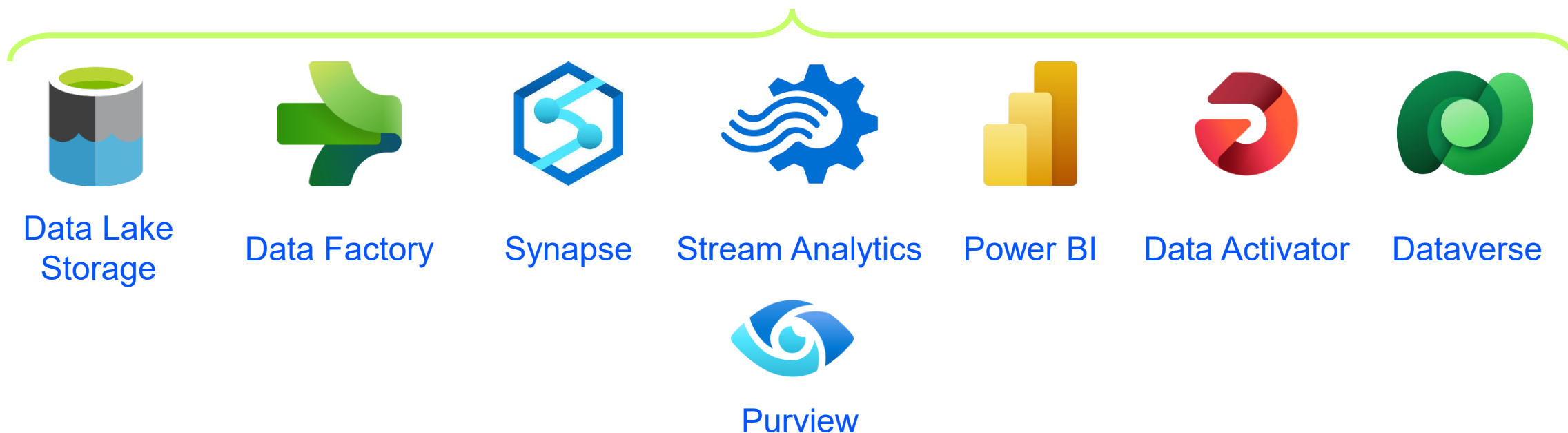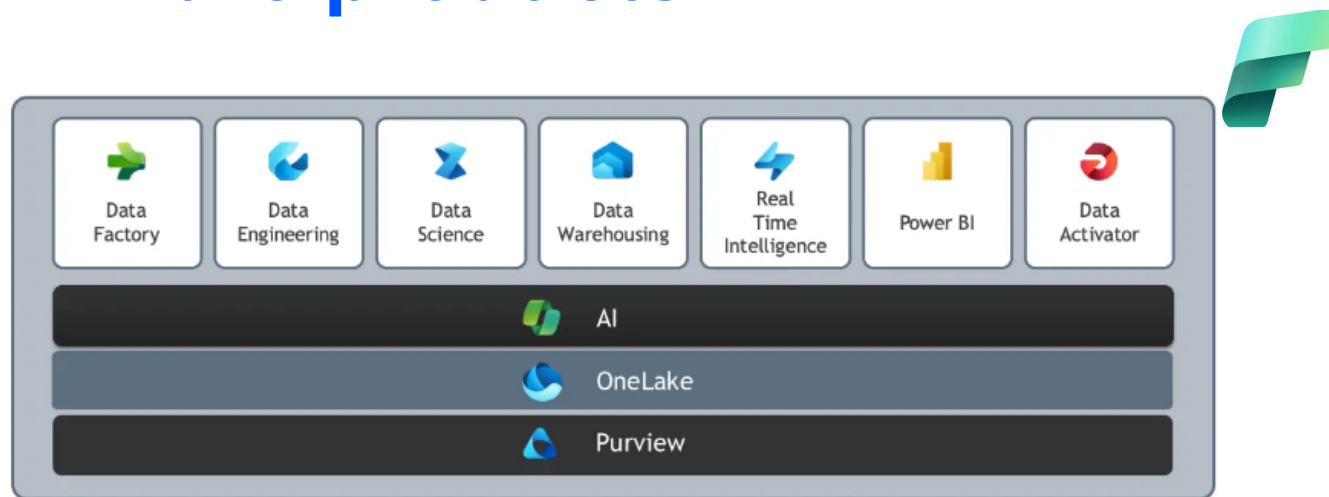
Quanta costa?

Compute

shared pool of capacity that powers all capabilities in Microsoft Fabric

Consumption is modeled with capacity units (CUs)

Different capacity sizes are available, easily scale down & up on demand

Storage

Data storage used consumed in One Lake

Billed based on GB/month

Coming Soon

Networking

03

Painting an Ideal World

How do we do transformations?

Spark

Using Fabric Notebooks

- Requires Spark knowledge
- Warm-Up time for pools
- Overkill for our amount of data

SQL in Warehouse

Define SQL Objects in Warehouse

- Working in Warehouse UI
- Hard to establish dev workflow

No code

Use Pipelines in Data Factory

- Few building blocks in Fabric available
- Hard to establish dev workflow

dbt

Define dbt models

- Only SQL skills required
- Easy to establish dev workflow
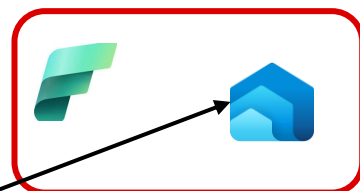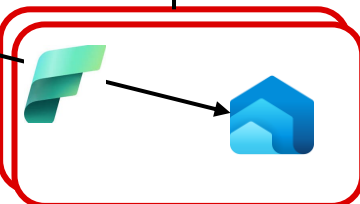- Less platform lock-in

Our Fabric Masterplan
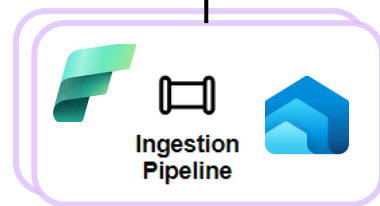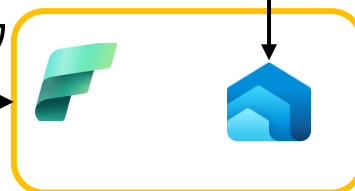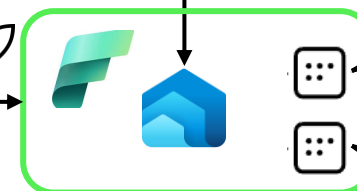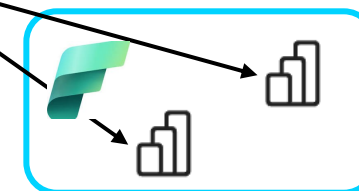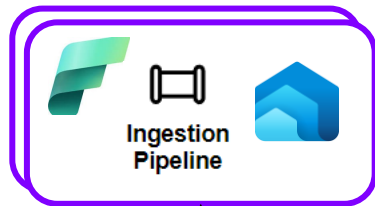

WS = Workspace

04

Problems arise

Problem #1

Landing WS PRD
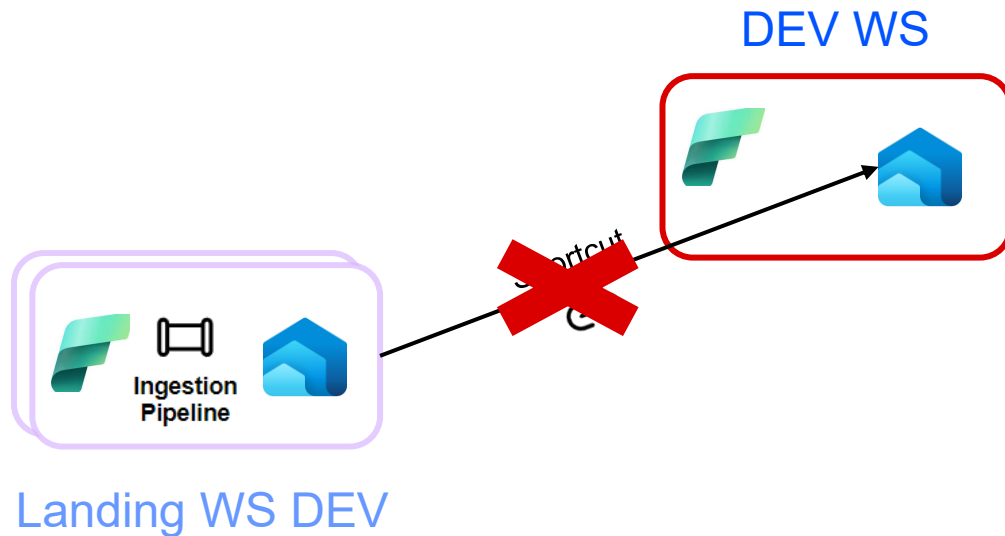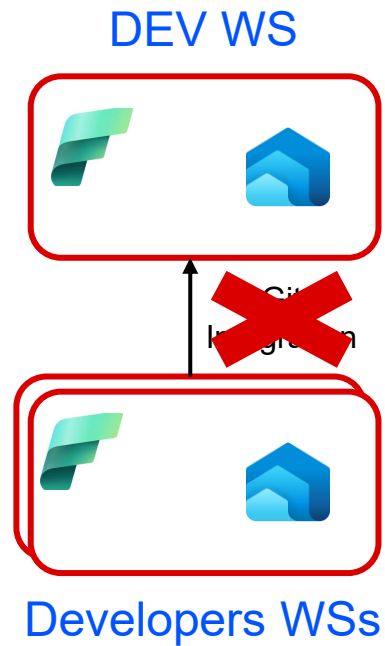

Landing WS DEV

- Deployment Pipelines only support a fraction of Fabric items

- Most element types, data pipelines, and warehouses, a.o.,
  are only partially/not supported
  (often not covering deployment rules for configuration)

Problem #2

DEV WS

Landing WS DEV

- Shortcuts are only support for Lakehouses

- Lakehouses only support READ operations in their SQL endpoint

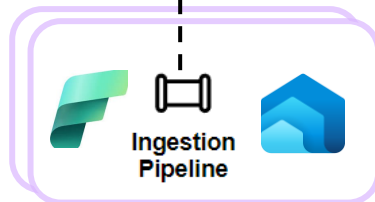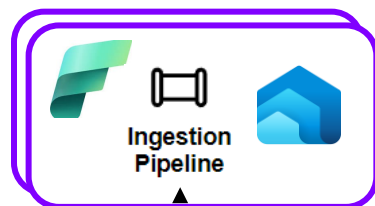   - > not compatible with the usage of dbt

Problem #3

DEV WS

Developers WSs

- Syncing a Warehouse from one workspace to another with the git integration does not work

- Multiple support cases and alleged fixes later, the problem still exists

Problem #4

Landing WS PRD

Landing WS DEV

- Pipelines can only trigger pipelines within the same workspace directly

- Pipelines can trigger other pipelines by using the REST API

- API does not support authentication via service principles

What gives us a pit in our stomach additionally …

- No real support for infrastructure as code, etc., there is no terraform provider (or similar)

- It's hard to establish review-processes

- Semantic models require regular refreshing even without changes
  to prevent breaking visuals in Power BI

- No integration of Azure Key Vault

05

How we arranged

Our simplified Fabric Setup
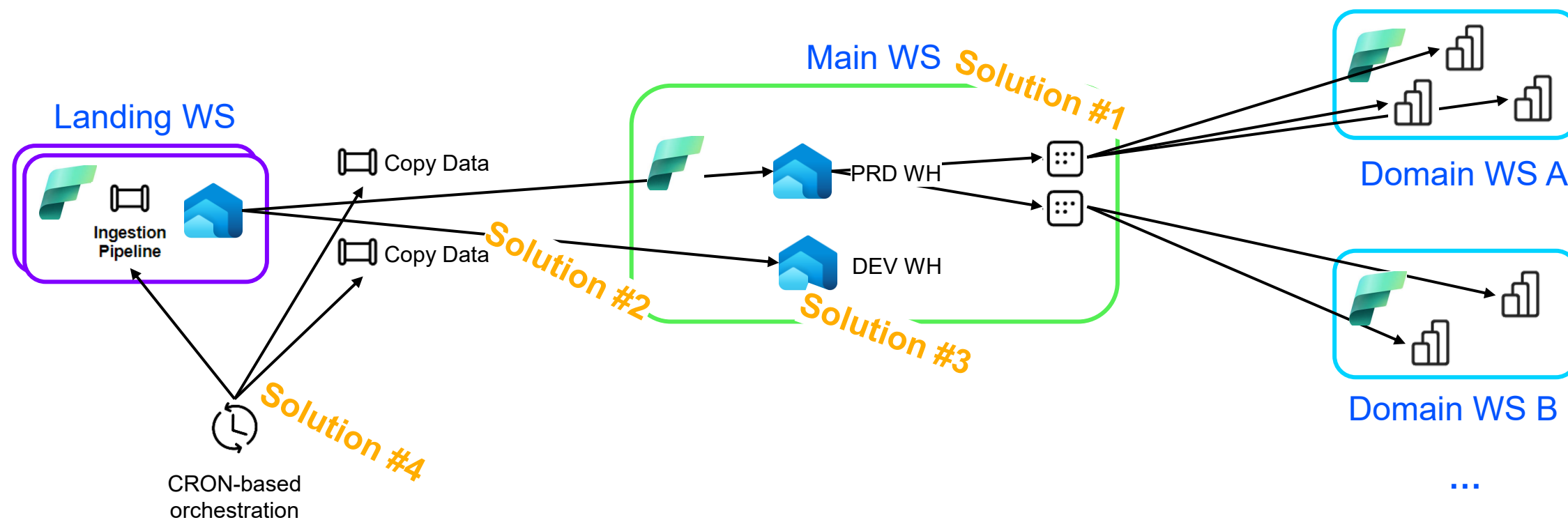
There is light on the horizon

- There is an experimental Terraform provider available

- Authentication via service principle is announced for Q3/Q4 2024

- Fix for DWH sync issue is announced for first week of November

… find more on their Roadmap

06

Conclusion

Quo vadis, Fabric?

cloudflight

Get the Slides

Thank you!

Questions & Discussion