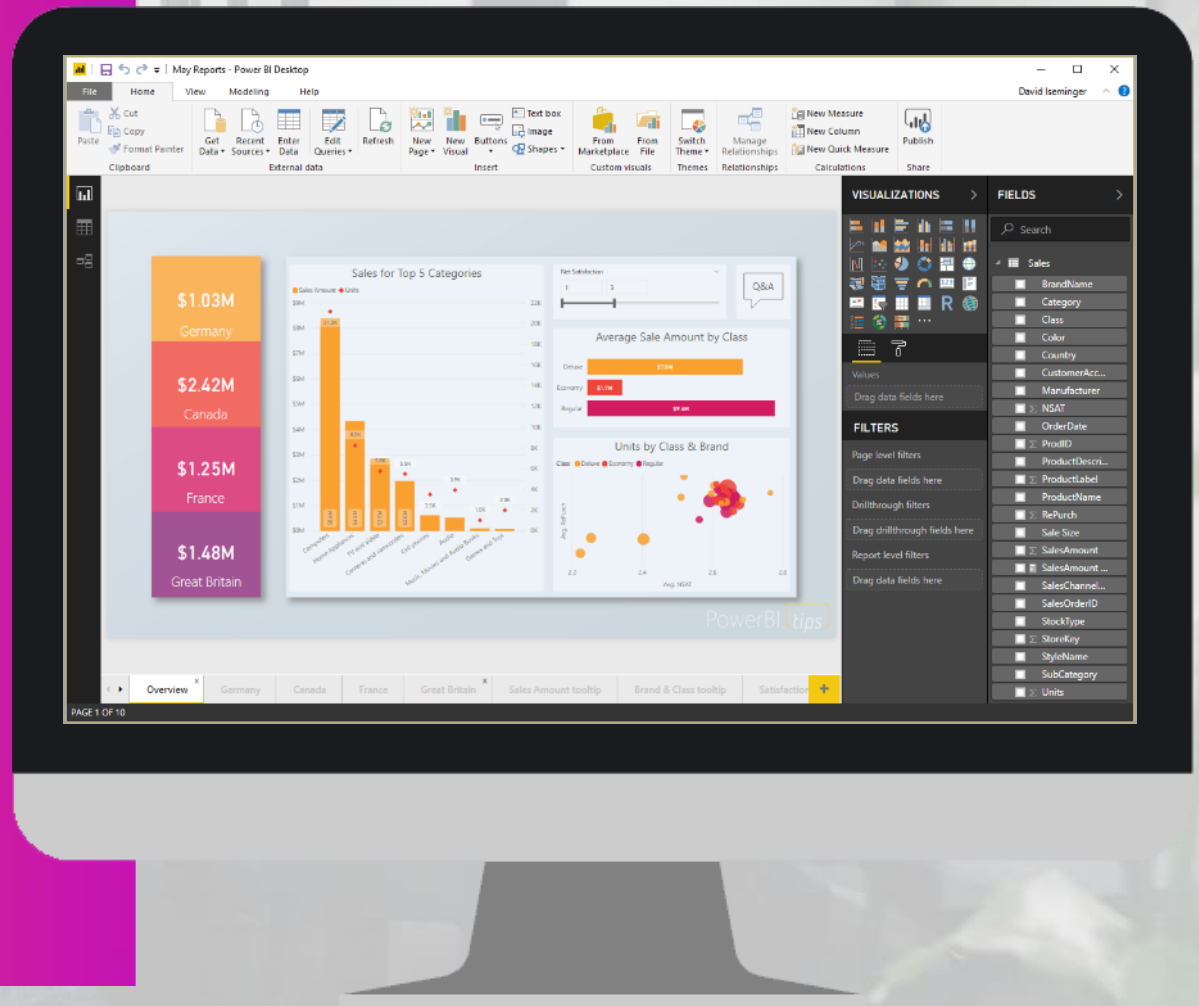
Advanced Refresh in Power BI with ADF/Synapse

Mathias Halkjær



De Power BI Gebruikersdagen worden mede mogelijk gemaakt door:

PBIG
POWER BI GEBRUIKERSGROEP

PLATINUM

delaware

Smart. wortell

ilionx

easydash

GOUD

Nova Silva

creates.

Tabular Editor

EIFFEL

Voorblijven. Niet bijblijven.

KASPAROV FINANCE & BI

macaw

rockfeather

GET RESPONSIVE

ZILVER

VICTA
BUSINESS INTELLIGENCE

Dynamic People

Motion10
AN hso COMPANY

VisionBI
Smart Data Experts

VALID
STAY AHEAD

inergy

valcon

iqbs

Quanto
collective analytics

COMMUNITY

Kimura
data intelligence

DashData.

volda;
INFORMATIESPECIALISTEN

Power BI Connector
by DAVISTA

AZURROFINANCE

Microsoft® Most Valuable Professional

sessionize MOST ACTIVE SPEAKER 2022

Fellowmind

MATHIAS HALKJÆR PETERSEN

Principal architect, Data & Analytics

Data enthusiast with a great passion for the synergy between data analytics and cloud data platforms. Principal Architect for Fellowmind working with building up our clients' data capabilities, implementing data platform technologies and all-in-all maximize the value they get out of their data.

- **Microsoft MVP**, Data Platform
- **MSc in Product Development and Innovation**
- **Certified Associate in Project Management (CAPM)®**
- **Sergeant**, Royal Danish Air Force
- Parent, Foodie, Casual Gamer, Tech-aficionado

FAVORITE TOOLS:

www.linkedin.com/in/mhalkjaer

Fellowmind

Agenda

1. Why go beyond Power BI?
2. Exploring the toolbox – what are our options?
3. **Example implementation –** Asynchronous Power BI refresh schedules orchestrated by ADF
4. **Demonstration –** Practical implementation walkthrough
5. Thoughts?

01 Why go beyond Power BI?

Capabilities in the Power BI service

Half hour intervals (with delays)

Refreshes per day: 8 for Pro, Unlimited (48) for Premium

Challenges

Schedule may be out of sync with DWH

Data may be old when arriving due to poorly timed orchestration.

Schedule time is imprecise

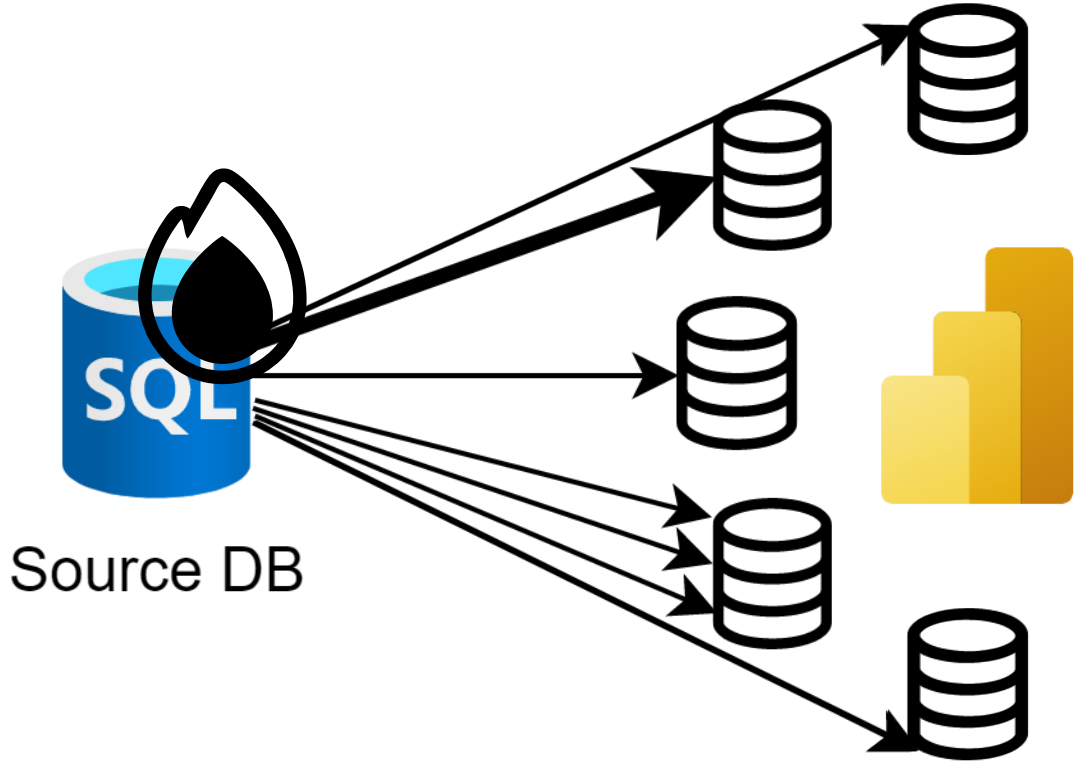
Not very fine-grained control over when the schedule happens, several datasets may have different versions of the data.

Intermittent problems may deactivate scheduled refresh

In only 2 hours a dataset may go from problem free to deactivated.

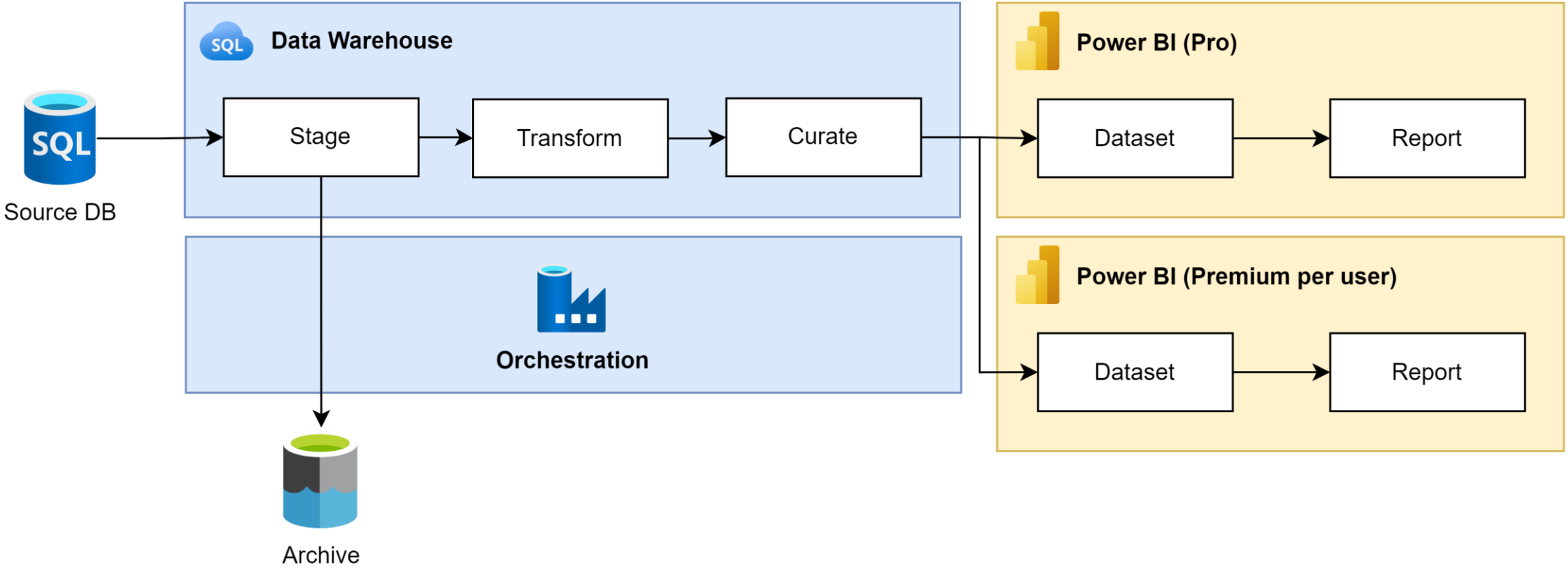
4 consecutive attempts, cannot change threshold.

The Case



PBIG
POWER BI GEBRUIKERSGROEP

The Case



02 Exploring the toolbox

– what are our options?

PBIG
POWER BI GEBRUIKERSGROEP

Tools

Rest API
- Trigger Dataset refresh
- Monitor Dataset refresh

Power BI Service
- Dataset refresh schedule

Power Automate
- Power BI: Refresh a dataset

Integration Pipelines
- Orchestration
- Pipelines
- Web activity task

Azure SQL
- Tables
- Stored Procedures

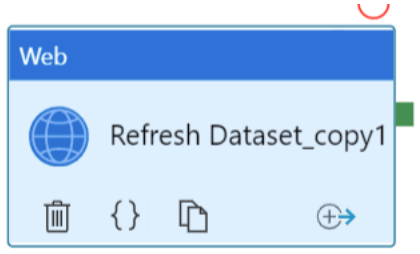
Data Lake
- Azure Table Storage

Demo: Refresh from ADF

Refresh Power BI Dataset from Azure Data Factory

Steps:

1) Create AD Security Group (E.g. SG-PBI-API-Access)
2) Fetch Managed Identity Object ID from your ADF or Synapse Pipeline
3) Add Managed Identity to the security group
4) Go to Power BI service, modify the admin setting “Allow service principals to use Power BI APIs”
    a. Enable it for specific security groups
    b. Add the security group from step 1
5) Add the Managed Identity as “Member” or “Administrator” to the workspaces you want it to work with
6) Go to ADF, create a pipeline with a web task
    a. Use the API endpoint (POST), replace groupId and datasetId with their GUID Id:
    https://api.powerbi.com/v1.0/myorg/groups/{groupId}/datasets/{datasetId}/refreshes
    b. Add body:
    {
    notifyOption: "“
    }
    c. Choose System assigned managed identity as auth and add resource:
    https://analysis.windows.net/powerbi/api



General | **Settings** | User properties

Fields for Http Request Timeout, Datasets, Linked Services, Disable Certificate Validation are moved to Advanced section.

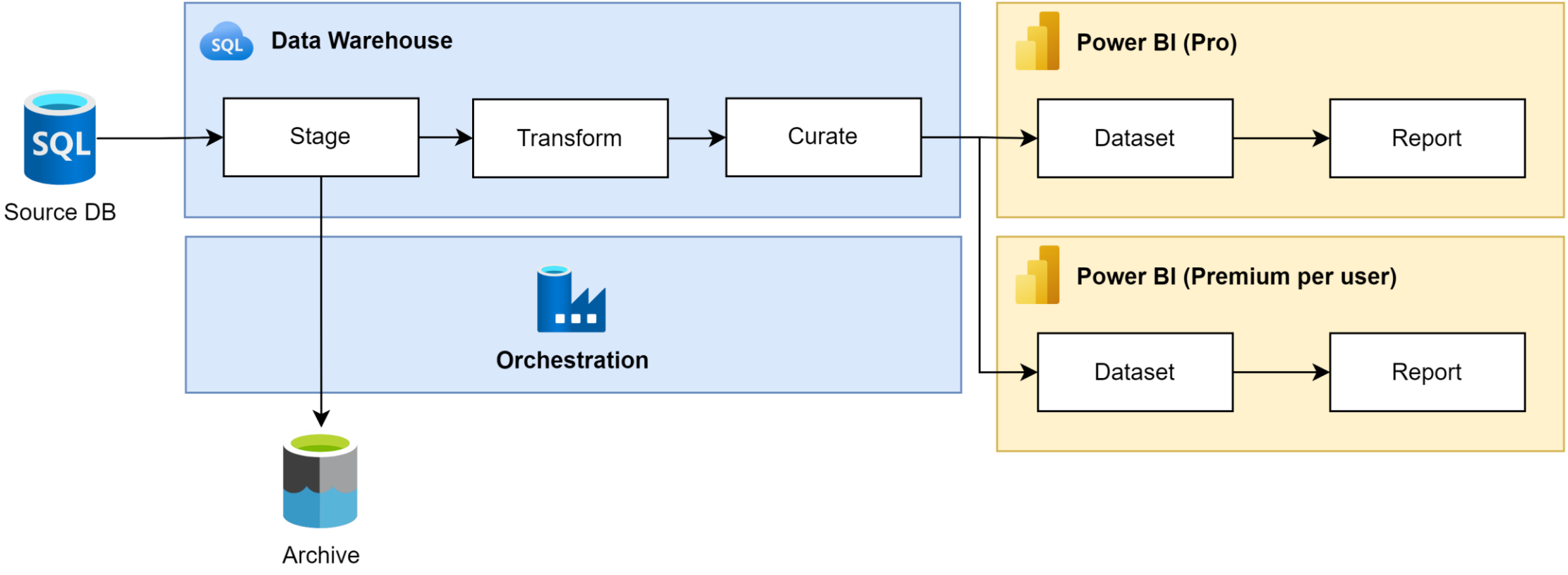
URL *	https://api.powerbi.com/v1.0/myorg/group
Method *	POST
Body	{ notifyOption: "" }
Authentication	System Assigned Managed Identity
Resource *	https://analysis.windows.net/powerbi/api
Headers	+ New

03 Example implementation

– Asynchronous Power BI refresh schedules orchestrated by ADF

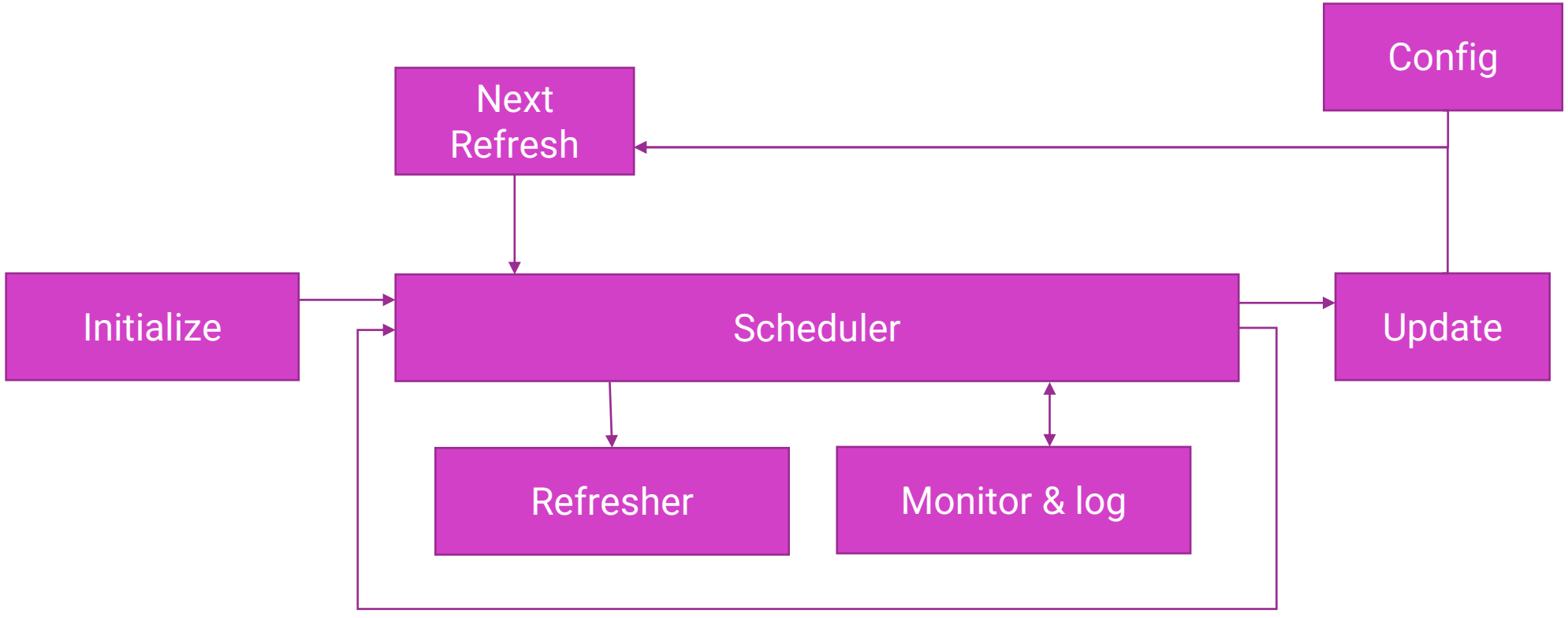
PBIG
POWER BI GEBRUIKERSGROEP

The Case



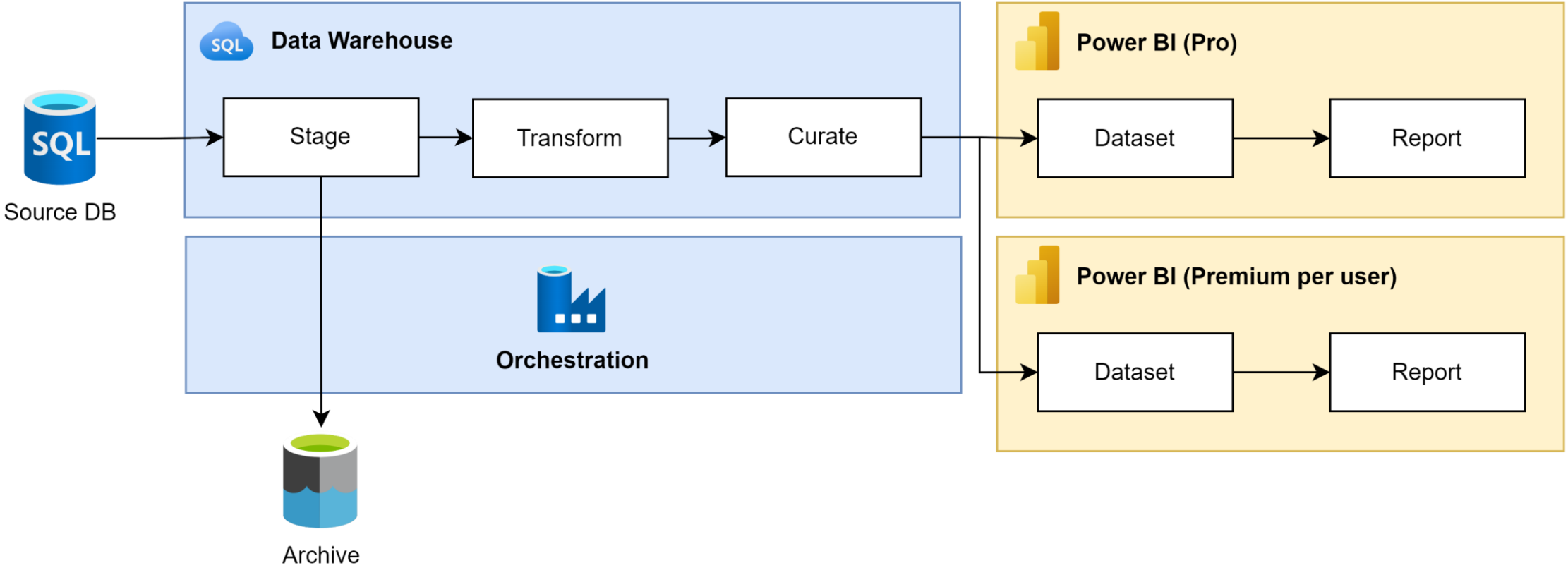
PBIG
POWER BI GEBRUIKERSGROEP

Concept

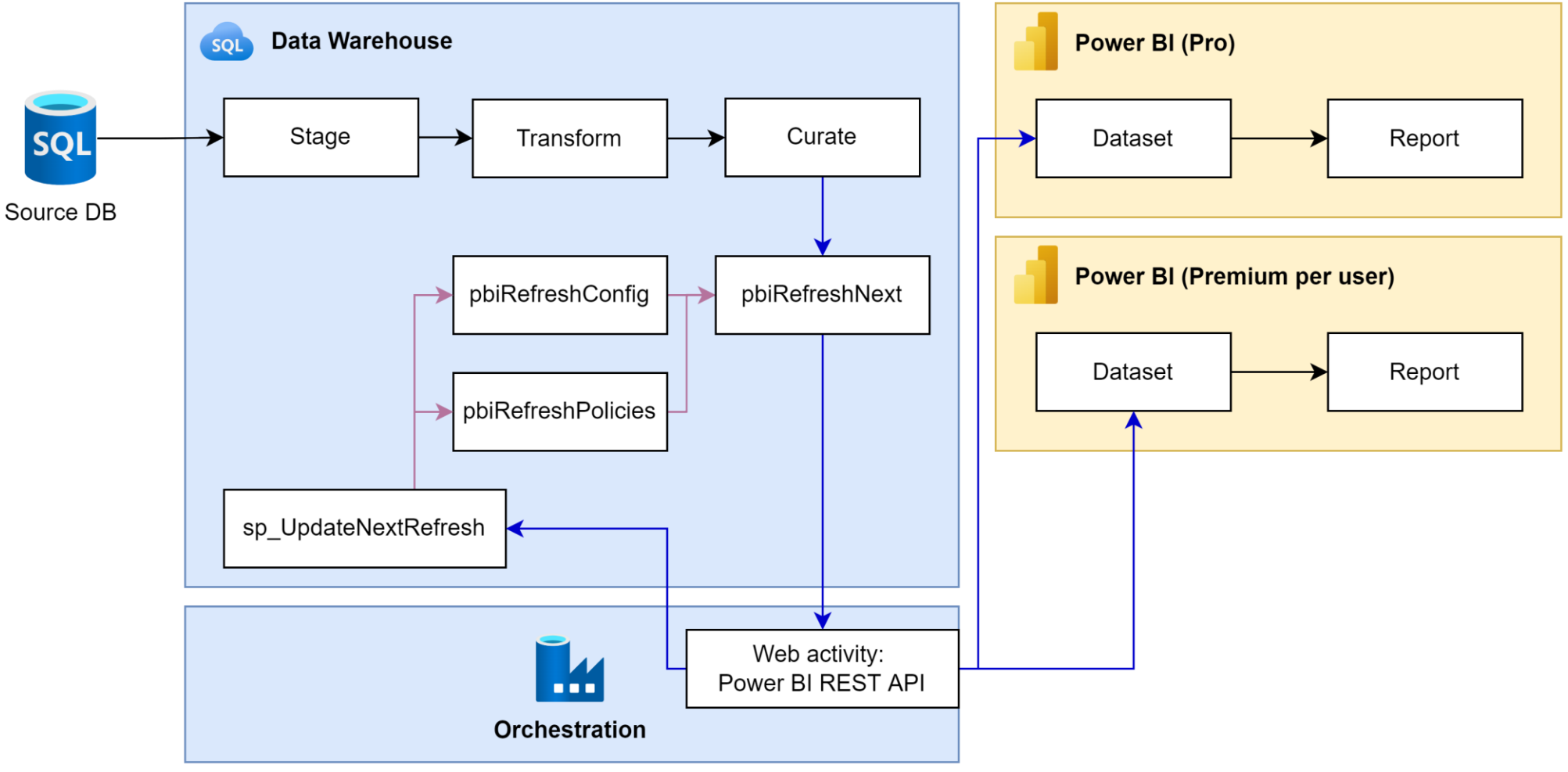


PBIG
POWER BI GEBRUIKERSGROEP

The Case



The Case


Data Warehouse
Source DB
Stage
Transform
Curate
pbiRefreshConfig
pbiRefreshNext
pbiRefreshPolicies
sp_UpdateNextRefresh
Orchestration
Web activity:
Power BI REST API
Power BI (Pro)
Dataset
Report
Power BI (Premium per user)
Dataset
Report

Stored procedure – Update PBINextRefresh

```
CREATE PROCEDURE [dbo].[SP_Reload_PBINextRefresh] @dataset uniqueidentifier, @workspace uniqueidentifier
AS
BEGIN
UPDATE rc
SET nextRefresh = rp.refreshDateTime
FROM pbiRefreshConfig rc
LEFT JOIN (
select refreshPolicy, min(date+cast(refreshTime as datetime)) as 'refreshDateTime' from pbiRefreshPolicies
CROSS JOIN (VALUES(cast(cast(DATEADD(hour,2,getdate()) as date)as
datetime)),((cast(cast(DATEADD(hour,26,getdate()) as date)as datetime)))) t(date)
where date+cast(refreshTime as datetime) > DATEADD(hour,2,getdate())
group by refreshPolicy
) rp on rc.refreshPolicy = rp.refreshPolicy
where datasetID = @dataset and workspaceID = @workspace
END
GO
```

04 Demonstration

– Practical implementation walkthrough

05 Thoughts?

Additional Reading

Data refresh in Power BI - Power BI | Microsoft Docs

Chris Webb's BI Blog: Delaying Power BI Dataset Refresh Until The Source Data Is Ready Chris Webb's BI Blog (crossjoin.co.uk)

Refresh your Power BI dataset using Microsoft Flow | Microsoft Power BI Blog | Microsoft Power BI

Azure Table Storage Tutorial | Easy and scalable NoSQL database - YouTube

How to refresh power bi datasets from data factory with managed identity. (tackytech.blog)

Building blocks

- Rest API
  - Refresh dataset
  - Monitor refresh status
  - Refresh individual tables
- Integration Pipelines
  - Orchestration
  - Web activity
- Config settings
  - SQL DB
  - Table storage (Data lake)
  - Power BI refresh schedule
- SQL DB
  - Schedule updater Stored Procedures

Microservices (components):

- Monitor & Log refresh
- Trigger refresh
- Schedule checker
- Schedule updater
- Authorization/Access (Managed principal?)

Agenda:

- Welcome
- Sponsor
- Me (+ Credit to Casper)
- Problem – why not just do it in Power BI
- Blueprint (Overview - Architecture drawing)
- Prerequisites
  - Authorization
    - Managed Identity
    - Workspace membership (Member+?)
  - ADF/Synapse, Data Lake/SQL, Power BI dataset (Obviously)
- Components
  - Trigger (Listener, Orchestrator, Input: Schedule, Output: Pipeline start)
  - Scheduler (Input:Trigger, Action: Check, Output: Refresh request)
  - Refresher (Input: Request, DatasetID, Auth?, Action: Trigger Dataset refresh)
  - Monitor & Log
- Blueprint (Deep-dive) / Demo
- Discussion/Questions
- Wrap up