Keeping the *Direct* in Power BI DirectQuery

Benni De Jagere

CAT

Gold Sponsors

Power BI Summit presented to you by

Microsoft

Benni De Jagere

Senior Program Manager | Power BI Customer Advisory Team (PBICAT)

Power BI CAT

dataMinds .be Member

@BenniDeJagere

/bennidejagere

sessionize /bennidejagere

/bennidejagere

#SayNoToPieCharts

CAT

OUR TEAM activities

We engage with strategic Power BI customers and the sales professionals supporting them to showcase and provide technical guidance, help drive product adoption and consumption and influence our engineering roadmap.

BDM SHOWCASE

We own demo resources for CxO-level conversations

PATTERNS & PRACTICES

We curate content on cross-cutting technology topics

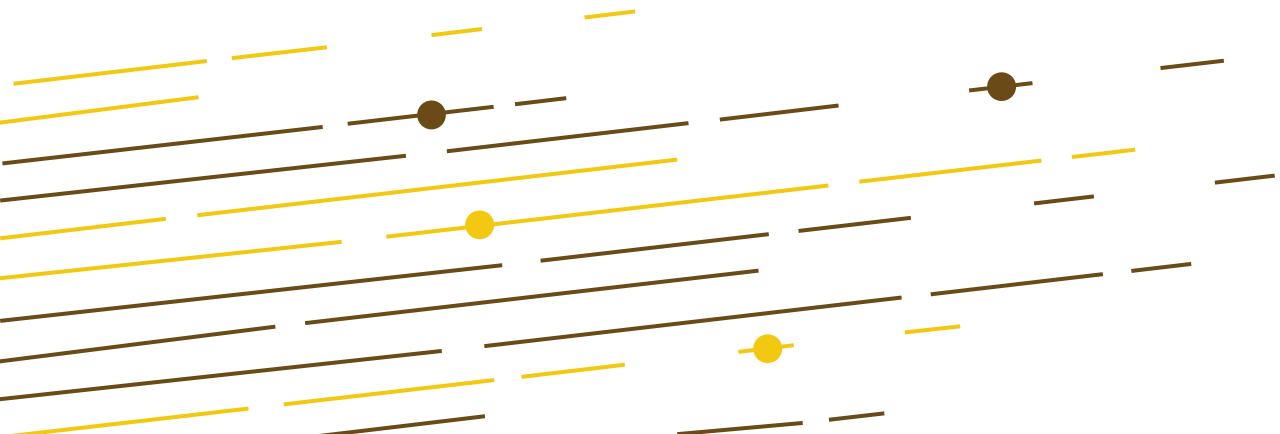
FIELD SUPPORT

We support technical escalations & training opportunities

ENGINEERING FEEDBACK

We influence the products roadmap via feedback rhythms

Session Objectives

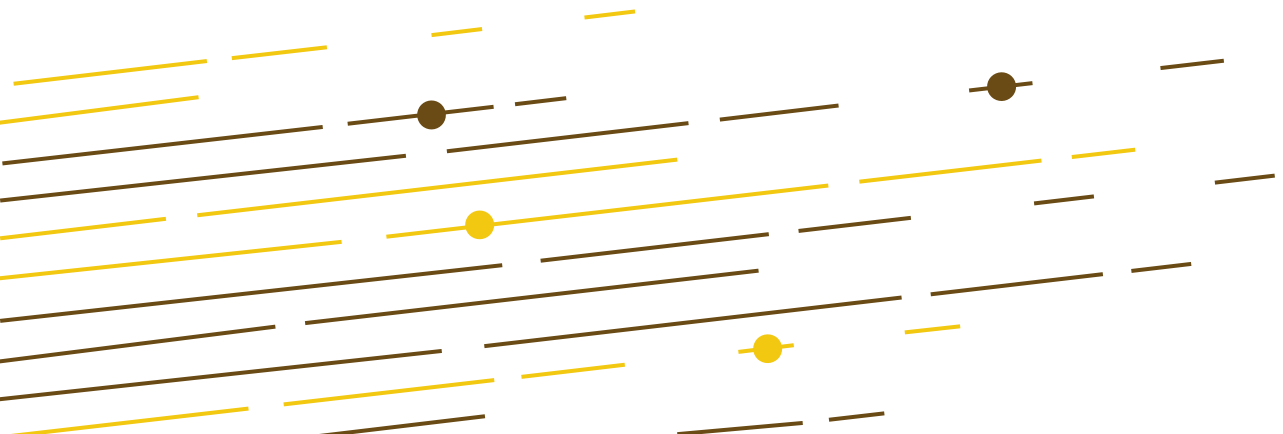


Session Objectives

- Assess problem areas
- Discuss different cogs in the chain
- Help you make better design decisions

Disclaimer: Benchmark your own scenario and results

Where does DirectQuery shine?



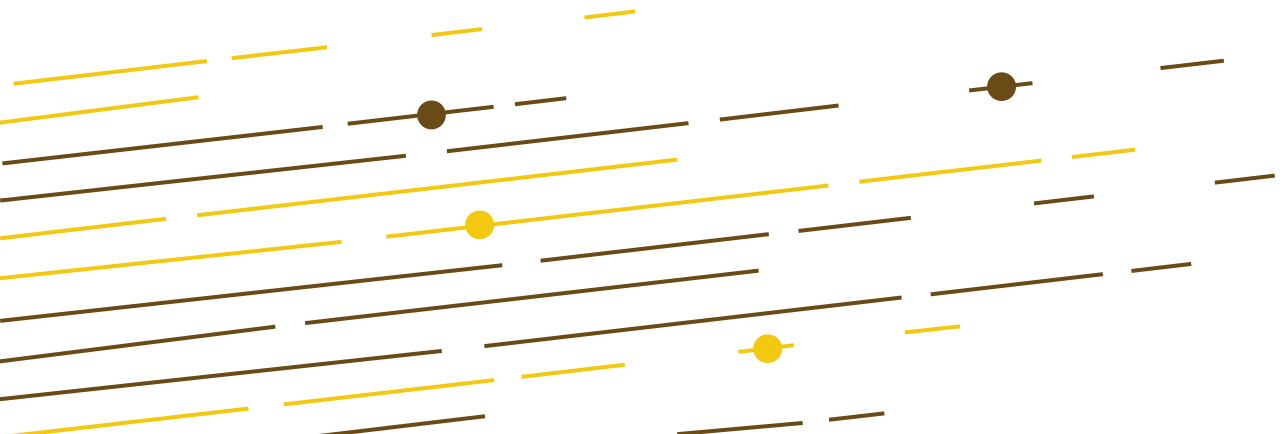
Reasons for opting for DirectQuery

- Data is too large to import
- Data changes too frequently to refresh
- Data Sovereignty restrictions apply
- Data Source has complex security roles
- Data Exploration scenarios
- Philosophical reasons..

But what's the catch?

It's not easy	• DirectQuery scenarios are typically harder to investigate • Schema changes have an immediate impact
Power BI is 'chatty'	• Data Source handles incoming requests • Response time of 3 seconds or less
Query Folding applies	• All transformations need to be transformed into source appropriate SQL • Beware of source specific functionalities
Scoped set of DAX	• Not all transformations are available • Especially 'Time Intelligence' is more complicated
User Education matters	• End users need to briefed on the implications for their actions

Where does it hurt?



User inspired quotes

"My report is slow. Fix it."

"Why can't I see anything on my screen?"

"Seriously, is this the best we can do?"

"I could build a faster report in Crystal Reports"

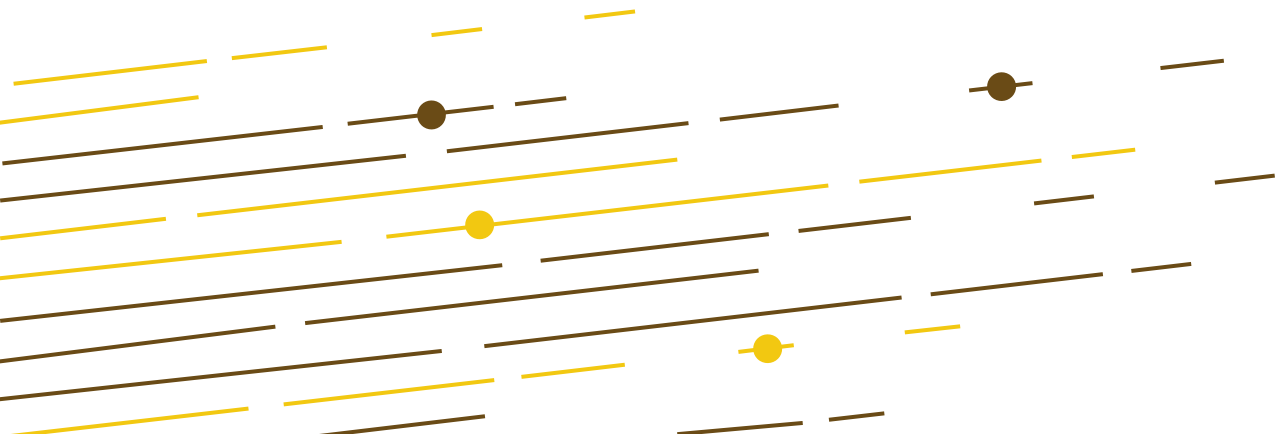
HELLO, IT, HAVE YOU TRIED
TURNING IT OFF AND ON AGAIN?

What's your tool of choice?

- Performance Analyzer Pane
- DAX Studio
- VertiPaq Analyzer (SQLBI.com)
- Tabular Editor
- Report Analyzer
- Power BI Helper
- Bravo
- ALM Toolkit
- Power BI Field Finder
- Power BI Sentinel (€)
- Power Query Diagnostics Tool
- Power BI Sidetools
- Power BI Cleaner
- PBI Tools
- And more ..

Power BI Performance Analyzer Pane
DAX Studio
SQL Profiler

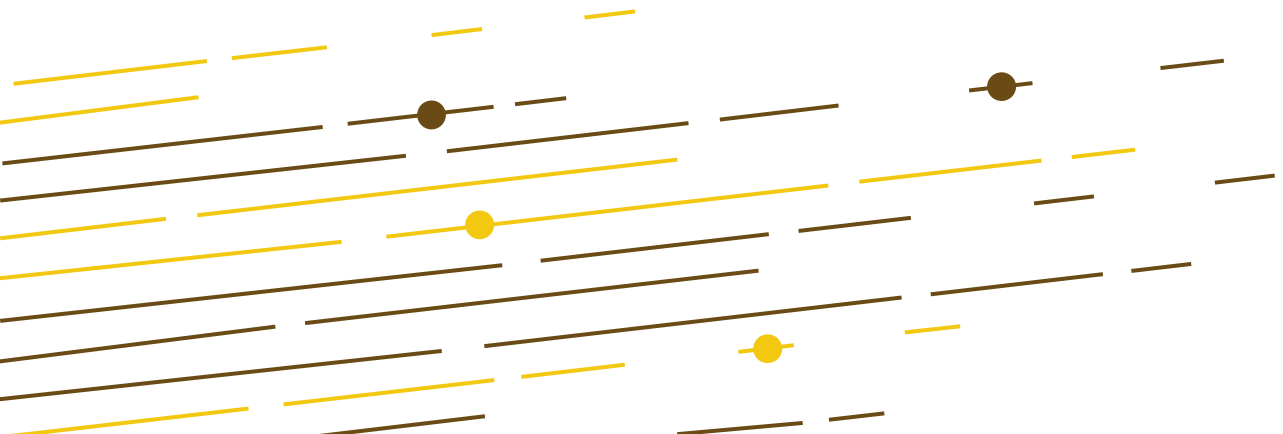
Data Source

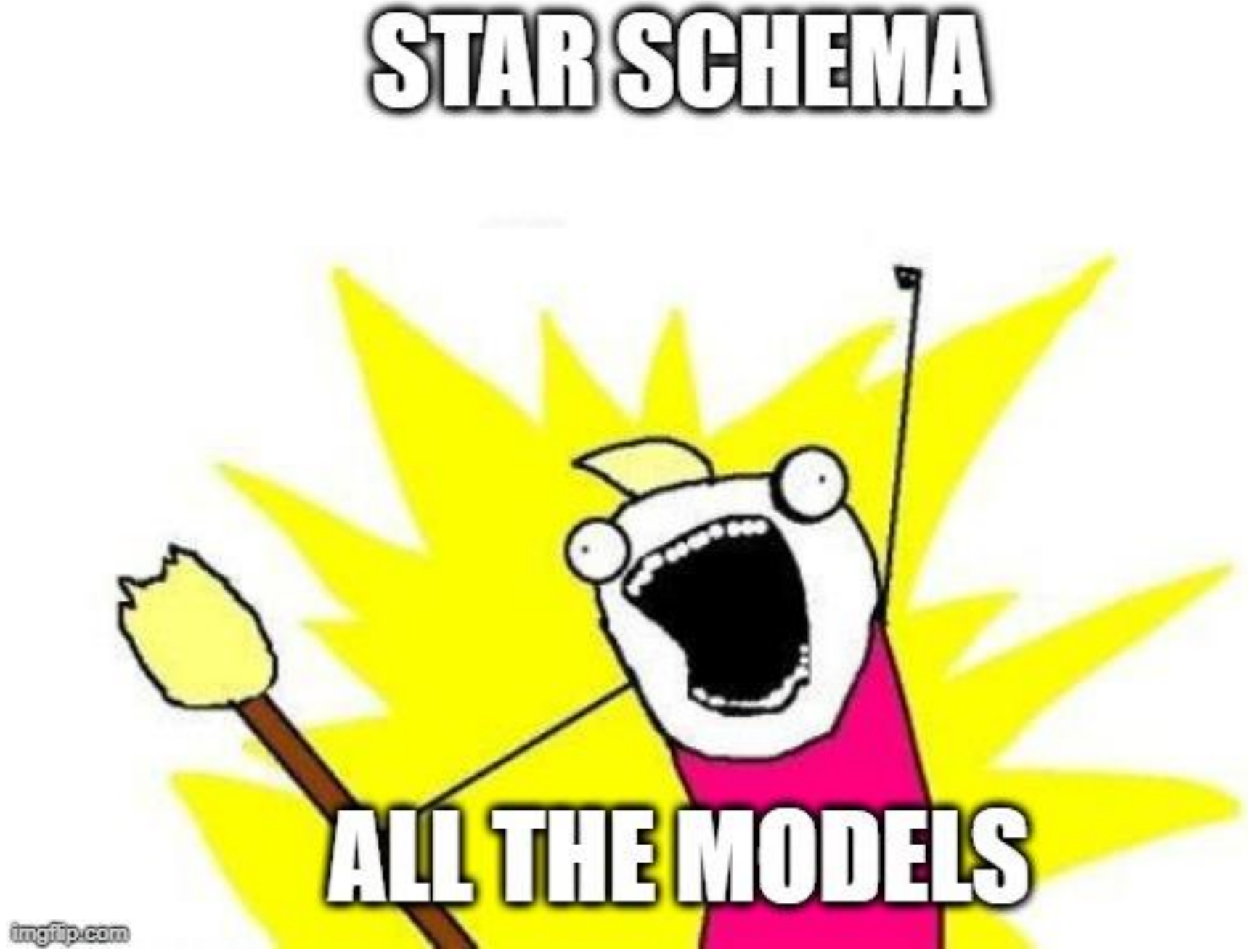


Is your data source ..

Scaled/Sized?	Data Typed?
Modelled?	Caching Results?
Materialized?	Data Integrity?
Indexed?	Limiting access to objects?
Partitioned?	Serverless?

Power BI Dataset (The Model)

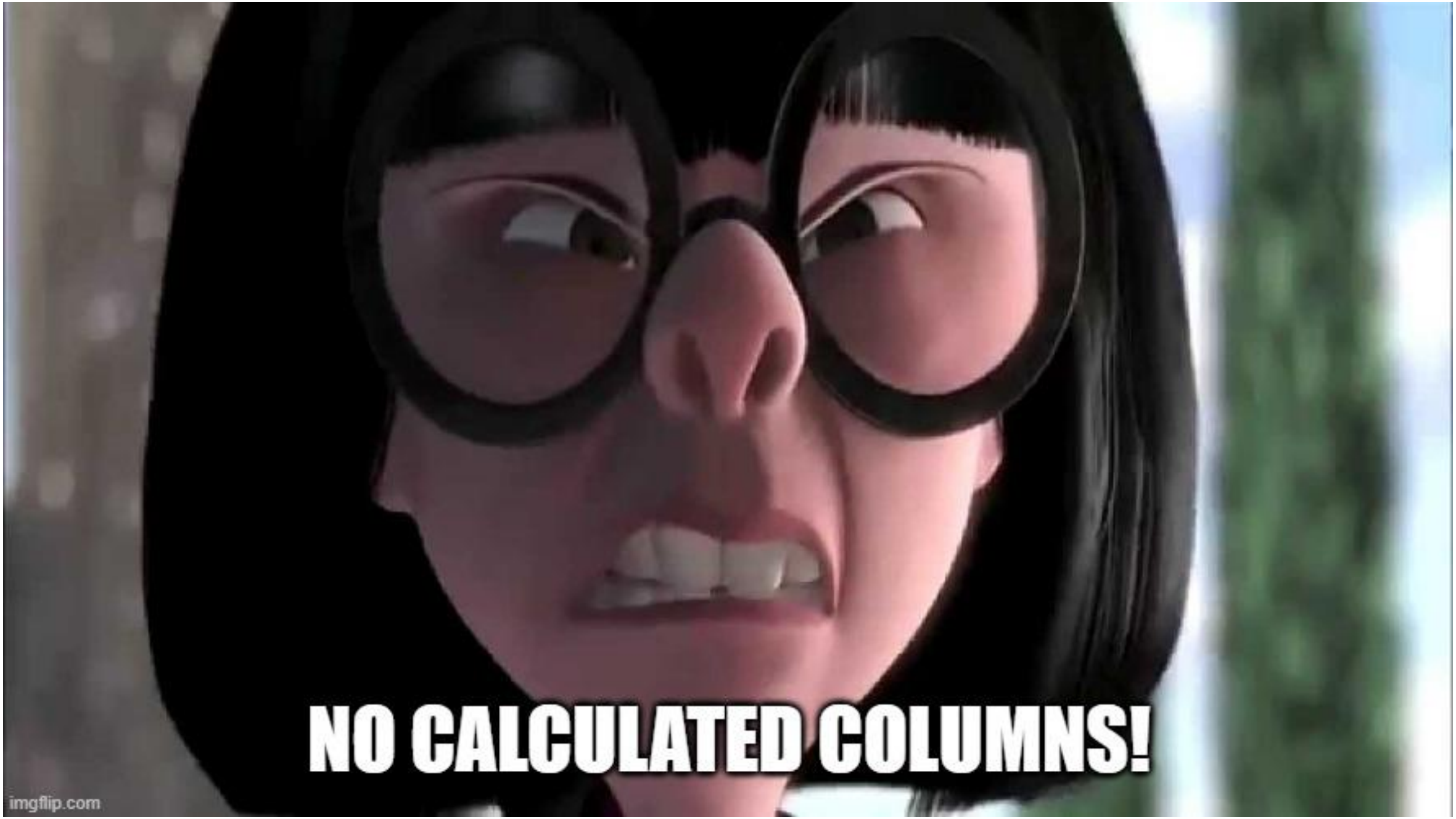




Thanks, @KoVer!

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

Matthew Roche, 2021
(The purple haired sword afficionado)
https://ssbipolar.com/2021/05/31/roches-maxim


NO CALCULATED COLUMNS!
imgflip.com

PowerQuery & Relationships

Push Data Transformations to the source

'Assume Referential Integrity'

Opt for Dual (or Import) mode for dimension tables

Avoid

- bi-directional filtering
- many-to-many relationships
- snowflake dimensions

Model

Keep things simple

Complex DAX is often due to modeling deficiencies

Make your Date dimension ☆ **shine** ☆

Hide the ‘one-side’ of the relationship

- Filtering on the fact generates poor queries

Properties

Consider the Maximum Data Connections	• 10 connections by default • 30 connections when using (💎)
Advanced:	• Premium Capacity Workload Settings (💎) • Analysis Services Server Properties (💎)

!! Use with caution !!

Capacity – Workload Settings

DATASETS

Observe XMLA-based workspace settings (which may override capacity settings)

On

Query Memory Limit (%)

0

Query Timeout (seconds)

240

Max Intermediate Row Count

1000000

Max Result Row Count

2147483(

Max Offline Dataset Size (GB)

0

Automatic page refresh

On

Minimum refresh interval

1 Seconds

Change detection measure

On

Minimum execution interval

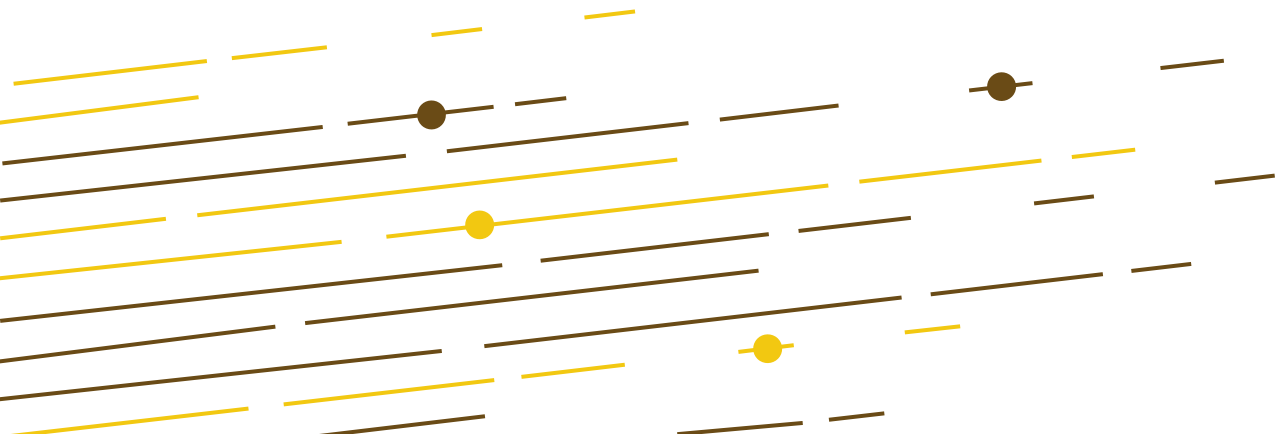
30 Seconds

XMLA Endpoint

Read Write

PQ & Model

Power BI Reports



Be mindful of your Filters and Visuals

Measure filters	Generates multiple queries
Fact attribute filters	Filters over fact table, when elimination through dimension could be enough
TopN	Generates multiple queries
Median	Not supported as aggregate by sources
Multi-Select	Each new selection launches queries to datasource
Visual Totals	Generates multiple queries

Query Reduction Techniques

- Visual Interactions
- Apply Filters
- Apply Slicers
- Where do you filter?

Report Tips

Limit number of visuals (and objects!)

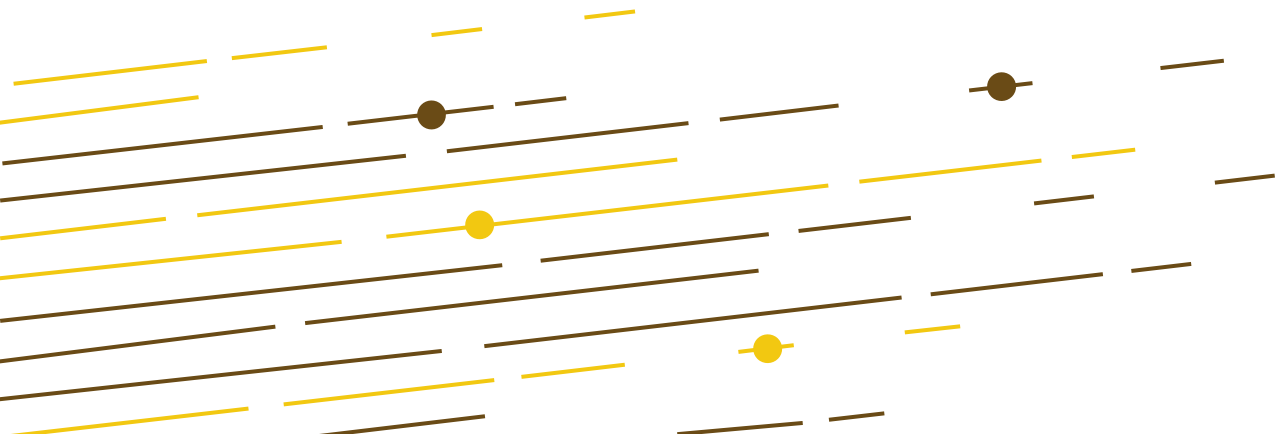
Be mindful of level of detail & granularity

Visible objects get rendered

- Bookmarks
- Drill Down
- Drill Through
- Page Navigation

Power BI Report

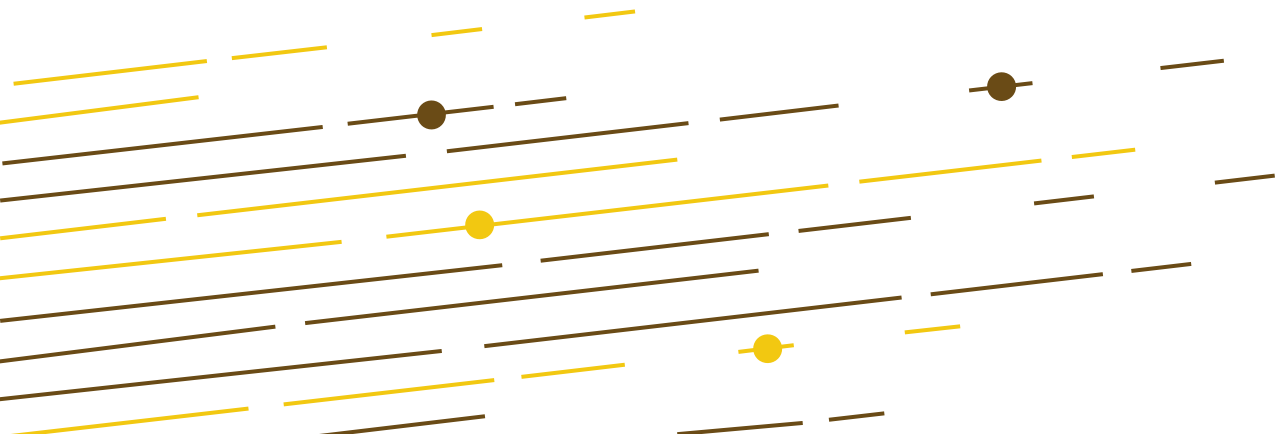
User Education



User Education

- Benchmark, Set expectations
- Filter first, details later
- Benchmark, Set expectations
- “Chatty” users can overload a data source
- Analyze in Excel over DirectQuery is killer

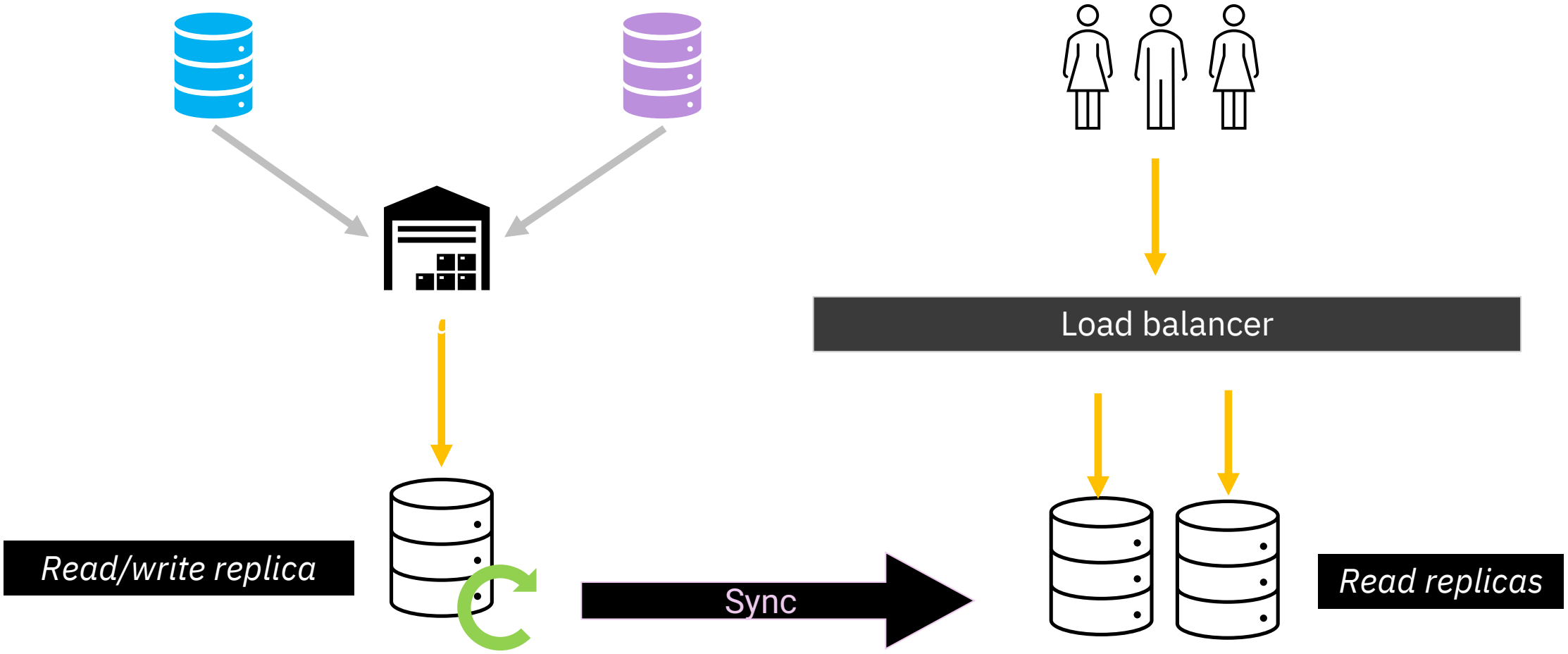
Kicking it up a notch



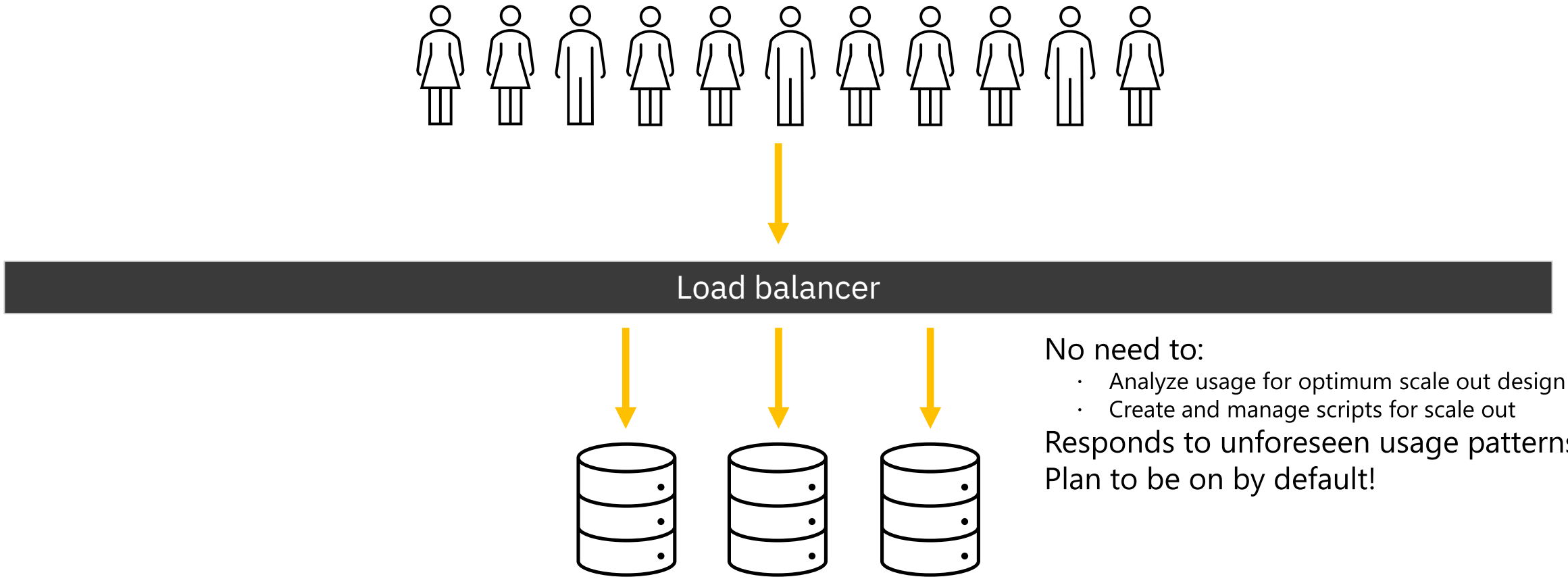
Look into these functionalities

- Composite Models
- User-Defined Aggregations
- Automatic Aggregations ( )
- Hybrid Tables

QSO - Refresh isolation



QSO – High Concurrency - *planned by GA*



No need to:

- Analyze usage for optimum scale out design
- Create and manage scripts for scale out

Responds to unforeseen usage patterns

Plan to be on by default!

Resources

- https://blog.crossjoin.co.uk/2023/01/02/why-dax-window-functions-are-important-for-performance-in-power-bi-directquery-mode/
- https://medium.com/snowflake/best-practices-for-using-power-bi-in-directquery-mode-with-snowflake-bfd1312ca7ab
- https://learn.microsoft.com/en-us/power-bi/guidance/directquery-model-guidance
- https://learn.microsoft.com/en-us/power-bi/guidance/star-schema
- https://blog.crossjoin.co.uk/2021/05/02/measuring-directquery-performance-in-power-bi/

Slides

https://github.com/BenniDeJagere/Presentations/{Year}/{YYYYMMDD}_{Event}

Thank you