DCAC

Denny Cherry
& Associates Consulting

Your Data,

Our Expertise

Power BI: Where Should My Data Live?

Meagan Longoria
Consultant, Denny Cherry & Associates
@mmarie

Microsoft Partner
Microsoft
Gold Cloud Platform
Gold Data Platform
Gold Data Analytics
Silver Application Development
Silver ISV

July 11, 2019

You Have Options for Your Power BI Storage Mode

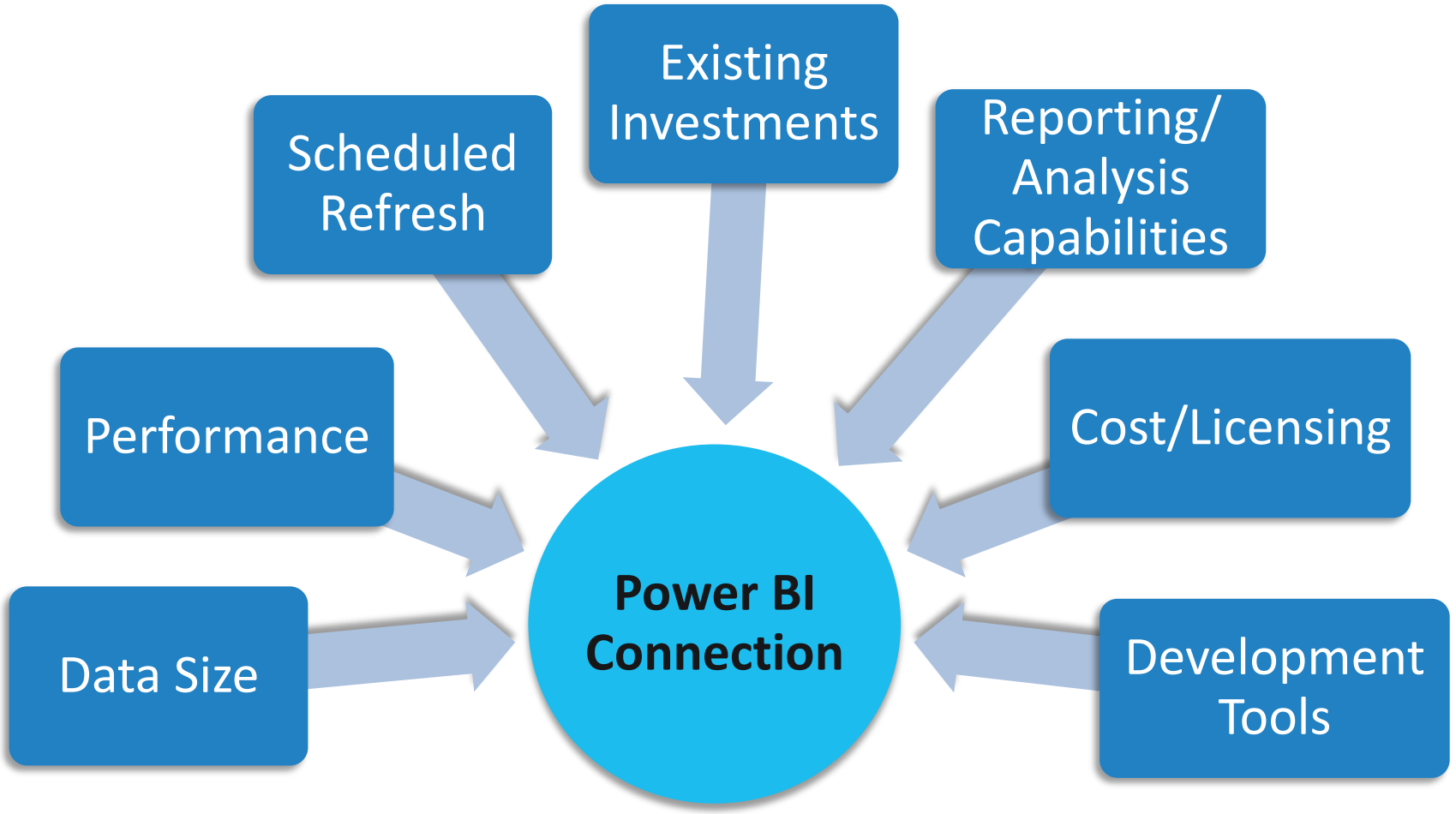
"Traditional" Connection Types

- Imported Model
- DirectQuery
- Live Connection

Game-changing Features

- Power BI Premium
- Data Flows
- Composite Models

Factors To Consider When Choosing a Storage Mode in Power BI



Our Plan For Today

1. Explain the 3 main connection types
2. Identify related decision factors
3. Discuss how game-changing features affect your decision
4. Walk through a few use cases
5. Questions

Power BI Storage Modes

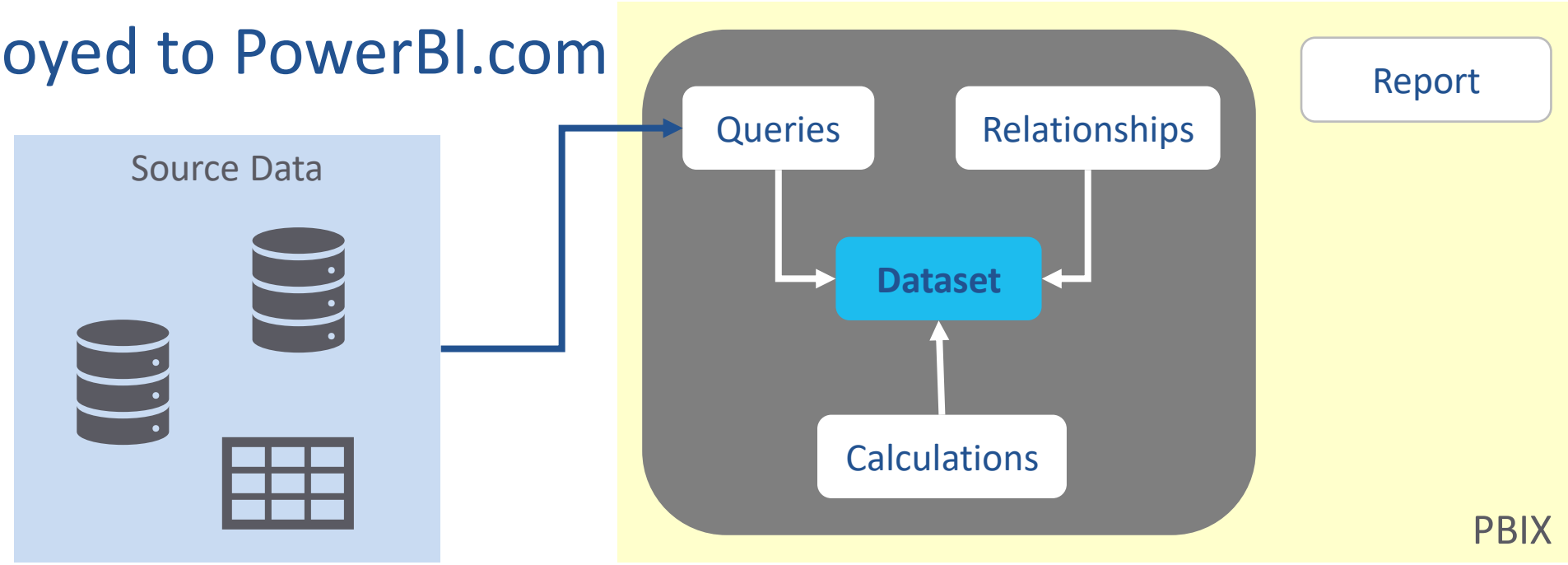
Imported Dataset

Build dataset in Power BI Desktop

A copy of source data is compressed and stored in RAM

All Power Query and DAX features available

Data is deployed to PowerBI.com



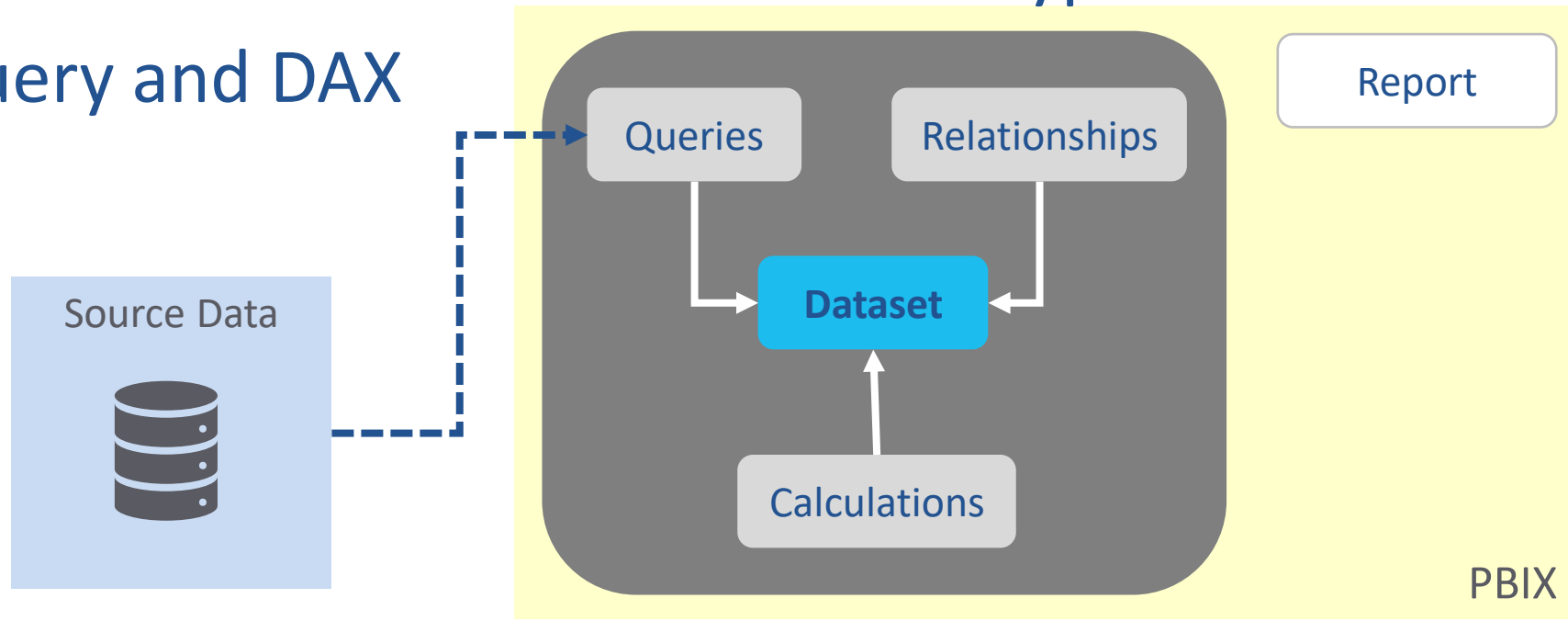
DirectQuery

Build dataset in Power BI Desktop

No data is stored. DAX queries are converted to a query against the external data source.

Single data source of a limited number of source types

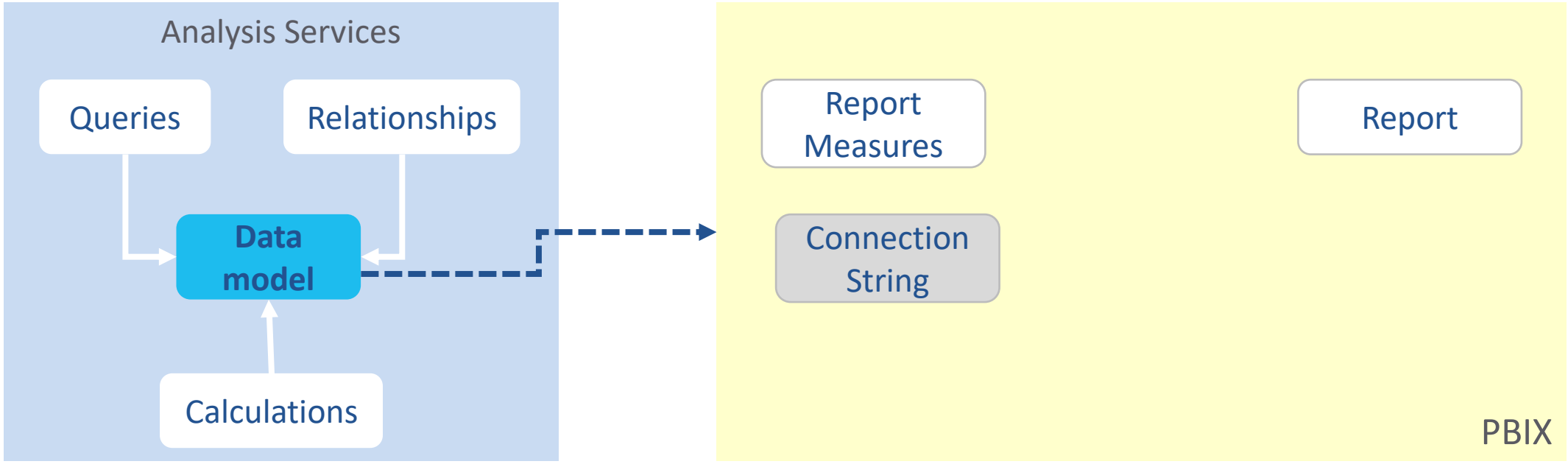
Limited Power Query and DAX



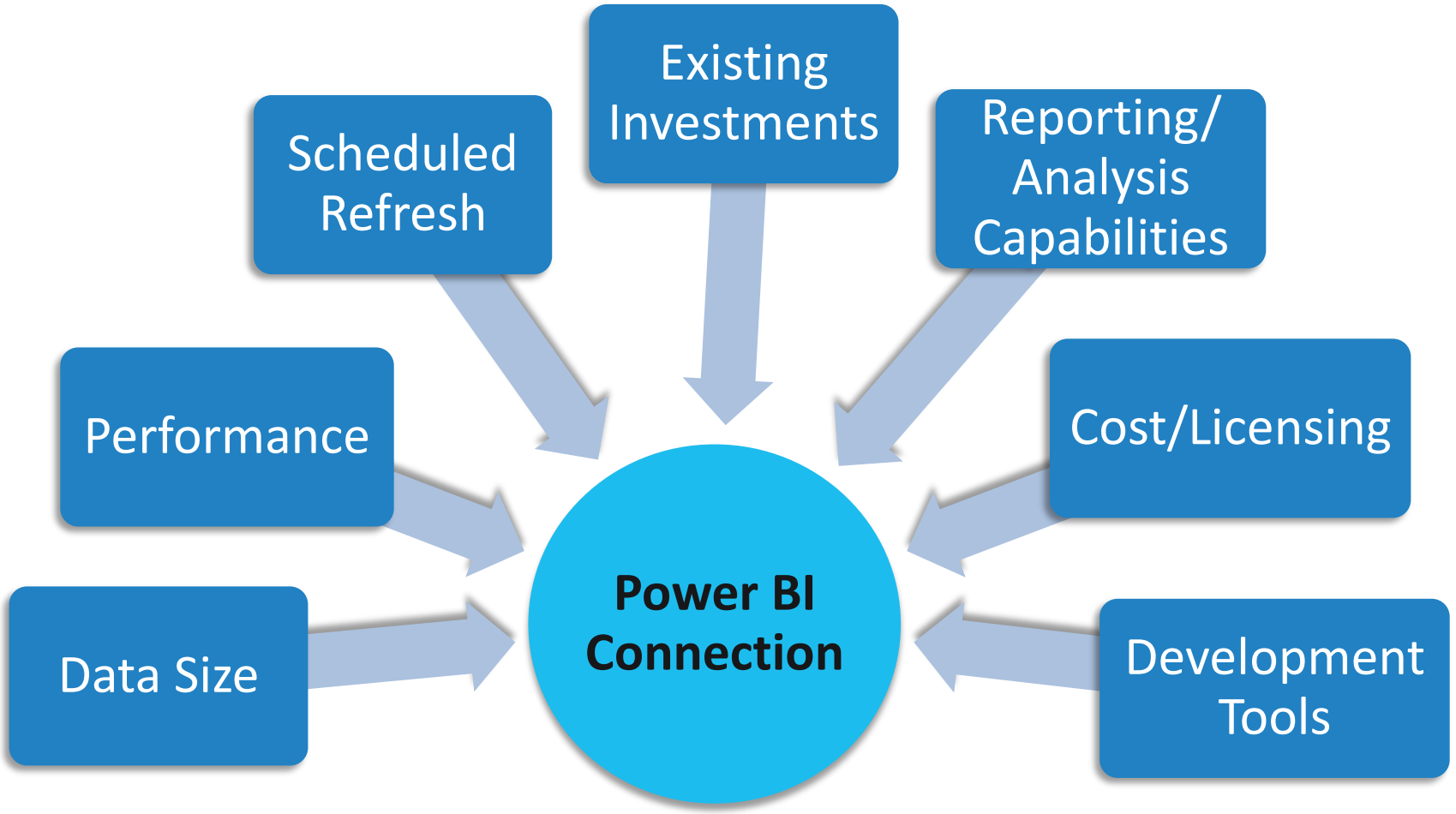
Live Query

Connect directly to a single Analysis Services model

Only report measures may be added



Factors To Consider When Choosing a Storage Mode in Power BI



Dataset Size Limitations

	Data Storage Location	Size Limit
Imported	Power BI	1 GB in shared capacity
DirectQuery	Source System	Server configuration
Live Connection	Source System	Server configuration

Query/Report Performance

Imported	Fast, usually quickest
DirectQuery	Slowest, requires perf tuning and careful data modeling
Live Connection	Fast, depending on server resources

Demo: DirectQuery vs Imported

Scheduled Refresh

Storage Mode	Where	Frequency
Imported	PowerBI.com	8x per day in shared capacity
DirectQuery	None	None
Live Connection	SQL Agent, PowerShell	Limited by speed of refresh

Note: PowerBI.com has a 15-minute SLA from the time of your scheduled refresh

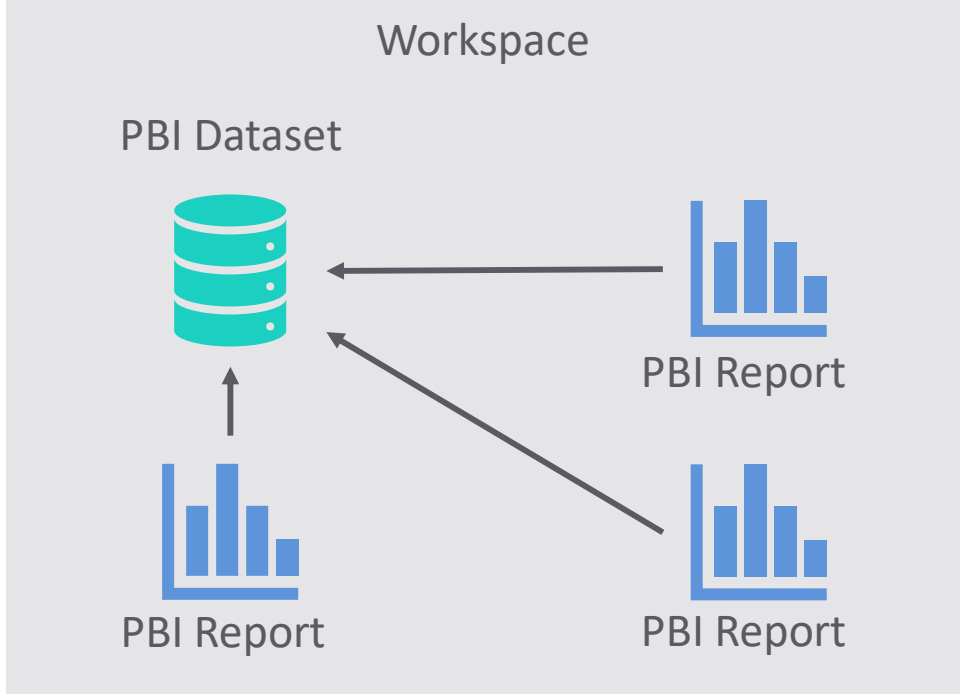
Existing Investments

Don’t throw out existing SSAS or Azure AS models!

Multidimensional works, but may limit features

Be aware of self-service needs

Reuse AS and PBI models to avoid data model sprawl



Shared Datasets!

Power BI Live Service Connection

Data Model Features/Capabilities

PBI Data Mashup/Model

Storage Mode	DAX	Power Query	Relationships
Imported	Full	Full	Full
DirectQuery	Partial	Partial	Partial
Live Connection	Report Measures	None	None

Security

Demo: AS Object Security

Storage Mode	Row-level	Object-level	Location
Imported	✓	X	PBI
DirectQuery	✓	X	PBI or source
Live Connection	✓	✓ (AS 2017+ or AAS)	Source (AS)

Reporting Features/Capabilities

Storage Mode	Imported	DirectQuery	Live Connection
Grouping/ Binning/ Clustering	✓	X	X
Hierarchies	✓	X	X
Q&A	✓	X	AS Tabular 2016+
Quick Insights	✓	X	X
Explore in Excel	✓	May be slow	✓

Cost/Licensing

Power BI Pro licenses (unless you have Premium) for all users

Using Analysis Services?

- Cost of SQL Server or Azure Analysis Services
- Cost of service used to refresh the model

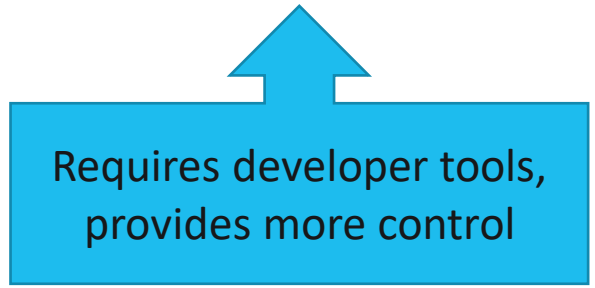
Using data sources in a private network?

- Cost of server to host Power BI Gateway

Includes Azure SQL DB/DW on a VNET

Development Tools and ALM

Storage Mode	Imported	DirectQuery	Live Connection
Dev Tool	PBI Desktop	PBI Desktop/Source	SSDT
Source Control	OneDrive/SharePoint	Azure Repo/Github/Any	Azure Repo/Github/Any
Deployment	Manual or PowerShell	Manual or PowerShell	Manual, PowerShell, Deployment Wizard, Command Line
Environment Variables	Parameters (change after deployment)	Parameters (change after deployment)	Set data source connections and database name during deployment



Game Changers

Power BI Premium

Dedicated capacity nodes

Larger dataset sizes, up to 10GB

Refresh up to 48 times each day

Incremental refresh

Geographic distribution

App consumers do not need a Pro license

Dataflows, AI (preview), Paginated Reports (preview)

XMLA endpoints (preview)

You must manage your capacity. You may need more than a P1.

Lots of potential for 3rd party reporting tools, model documentation, and development tools

Dataset Size Limitations - Revisited

	Data Storage Location	Size Limit	Size Limit with Premium
Imported	Power BI	1 GB in shared capacity	Capacity limit
DirectQuery	Source System	Source system limit	
Live Connection	Source System	Source system limit	

Premium Support for .PBIX Size:
P1: < 3 GB
P2: < 6 GB
P3: < 10 GB

Scheduled Refresh - Revisited

Storage Mode	Where	Frequency	With Premium
Imported	PowerBI.com	8x per day in shared capacity	48x per day, allows incremental
DirectQuery	None	None	None
Live Connection	SQL Agent, PowerShell, Logic App	Limited by speed of refresh	

Note: PowerBI.com has a 15-minute SLA from the time of your scheduled refresh

Cost/Licensing

Power BI Pro licenses (unless you have Premium) for ~~all users~~ producers

~~Using Analysis Services?~~

- ~~Cost of SQL Server or Azure Analysis Services~~
- ~~Cost of service used to refresh the model~~

Using data sources in a private network?

- Cost of server to host Power BI Gateway

Composite Models

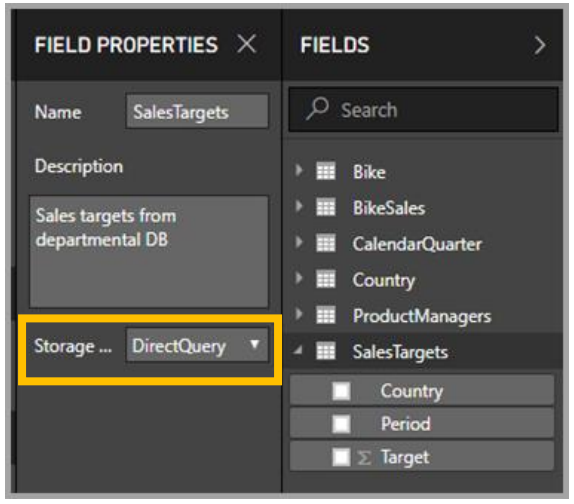
Combine multiple DirectQuery and imported data sources

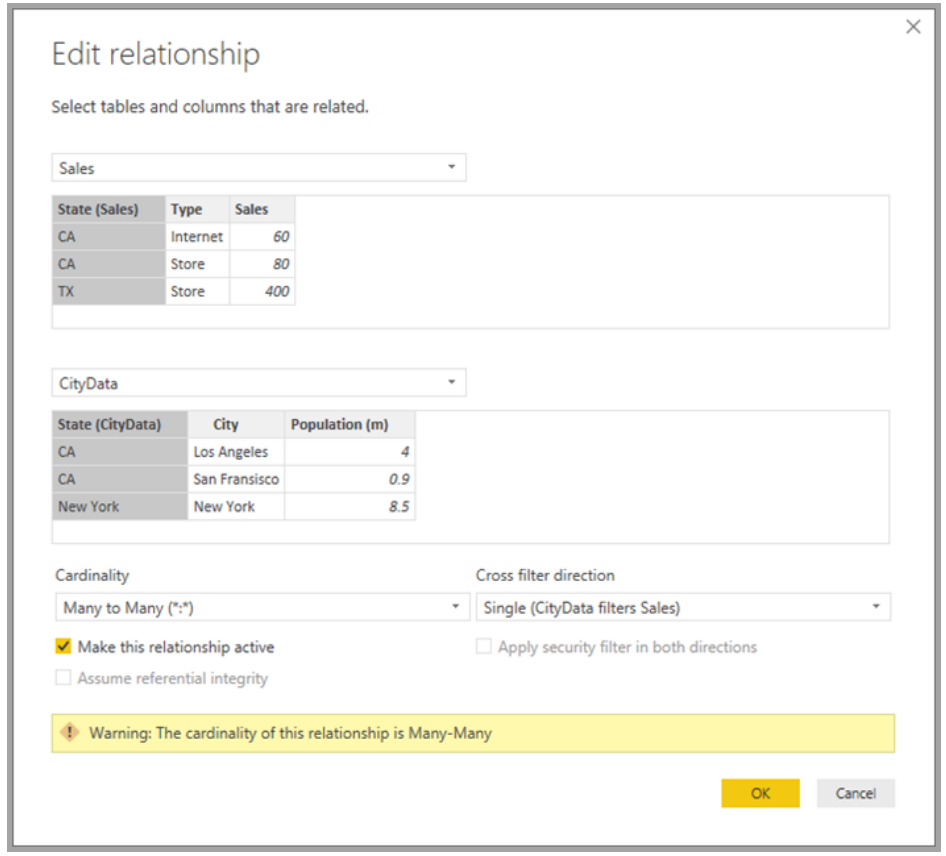
Use Many-to-Many relationships

Use calculated tables on DirectQuery sources

Security implication: cross pollination of data

Performance still a concern





Dataflows

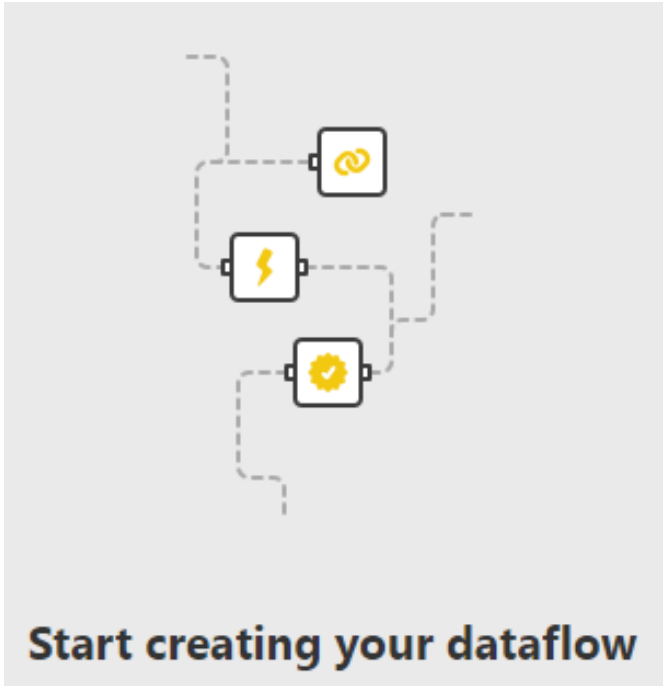
Self-service data prep that creates reusable entities

Use Power Query to define entities that don’t fit on your laptop

Store data in Azure Data Lake Storage Gen2

Schedule refresh through Power BI

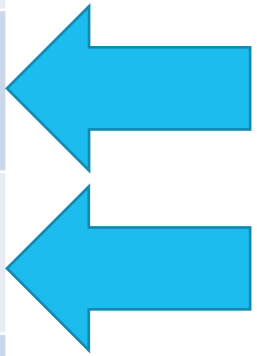
Use with imported/composite models



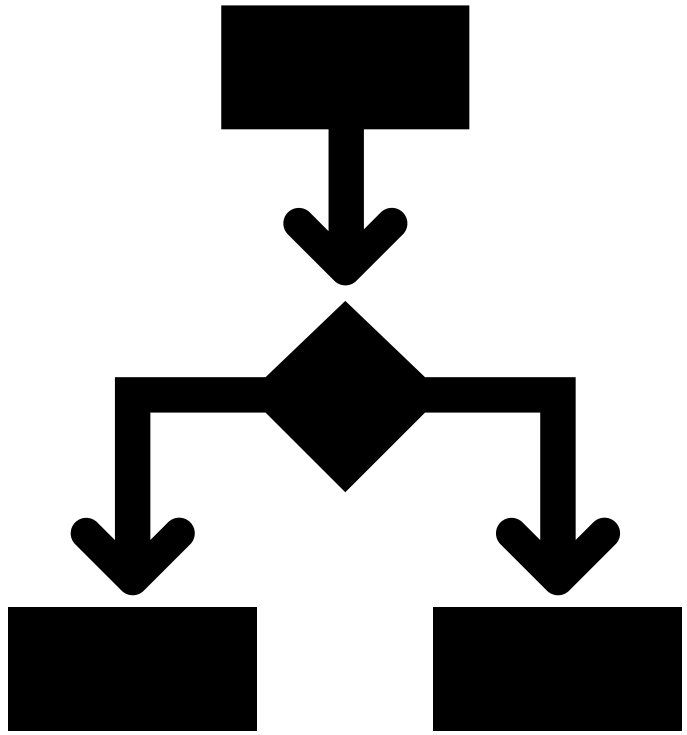
Dataflows with Power BI Premium

Dataflow Capability	Pro	Premium
Connectivity	All connectors to all sources	All connectors to all sources
Storage	10GB per user	100TB for P1 or greater nodes
Data ingestion	Serial ingestion of entities, making data refresh longer	Parallel ingestion of entities
Incremental updates	Not available	Available
References to entities in the same workspace	Not available	Available, allowing the creation of complex data prep processes using multiple dataflows
References to entities across workspaces	Not available	Available, allowing full data consistency across the whole data estate
Calculation engine	Not available, since entities cannot refer to other entities, computed entities cannot be created	Available, allowing computed entities for complex data prep projects with multiple cleansing and enrichment steps
Refresh rates	Up to 8 times a day	Up to 48 times a day

This is cool



From Data Flows in Power BI Whitepaper, Nov 2018



Use Cases

Use Case #1

Departmental Expenses reports

2 GB OLTP Database + 2 Excel workbooks

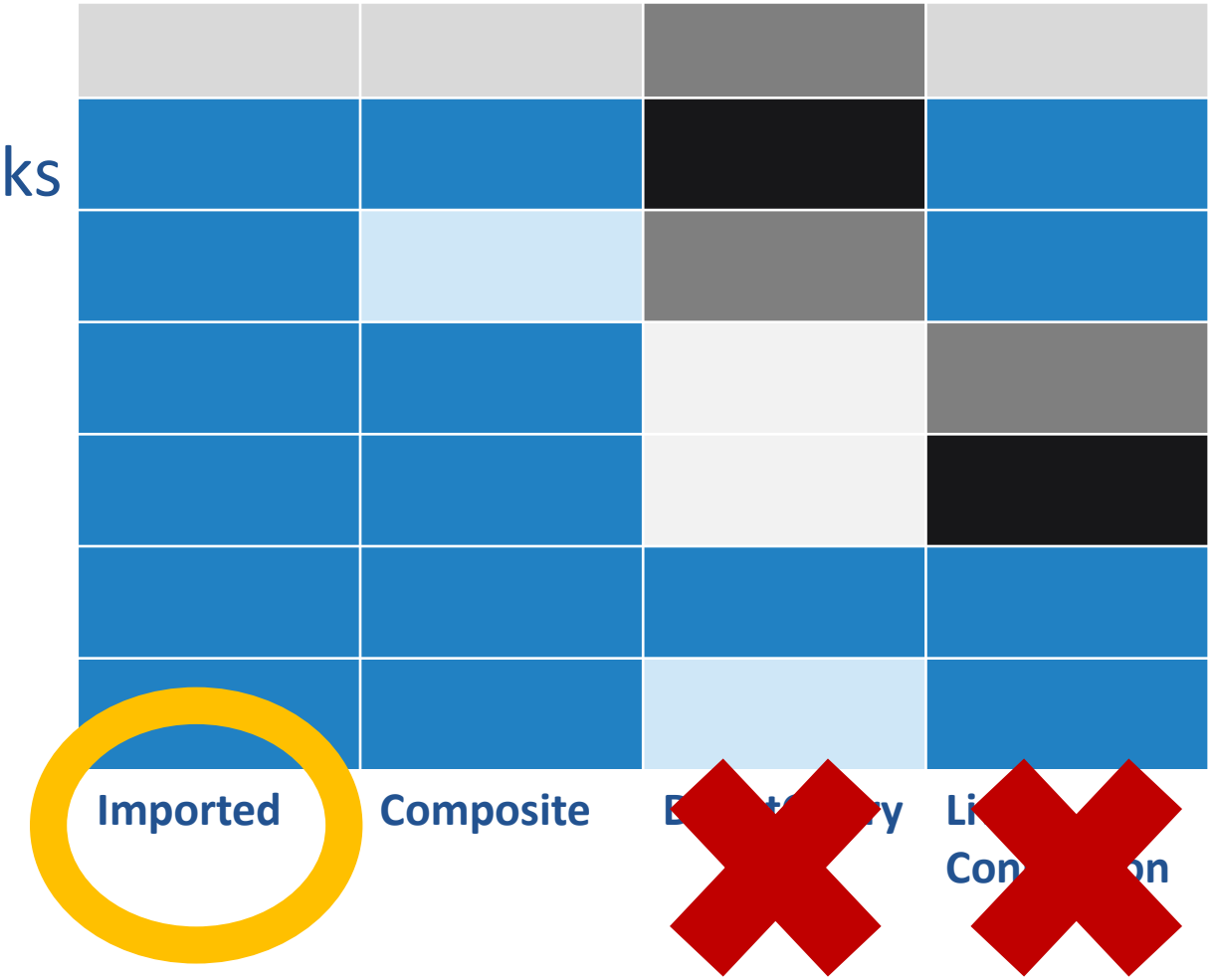
Need to see new data weekly

No existing SSAS instances

Maintained by financial analyst

Multiple reports use the same data

Audience is 15 department managers



Use Case #2

Corporate Profit Dashboard

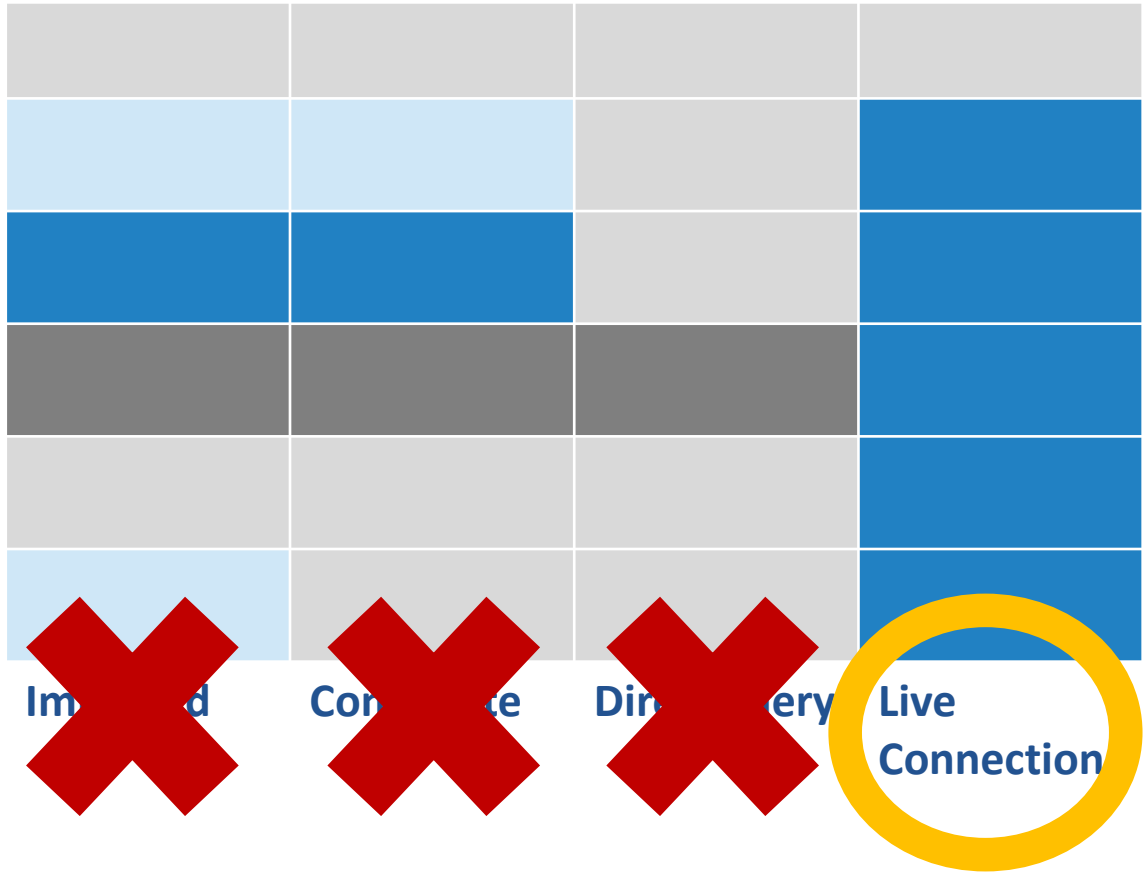
5 GB source data warehouse

Refresh daily

Existing SSAS Tabular

Maintained by IT

Audience of 300 employees



Use Case #3

Trucking company current on-time status

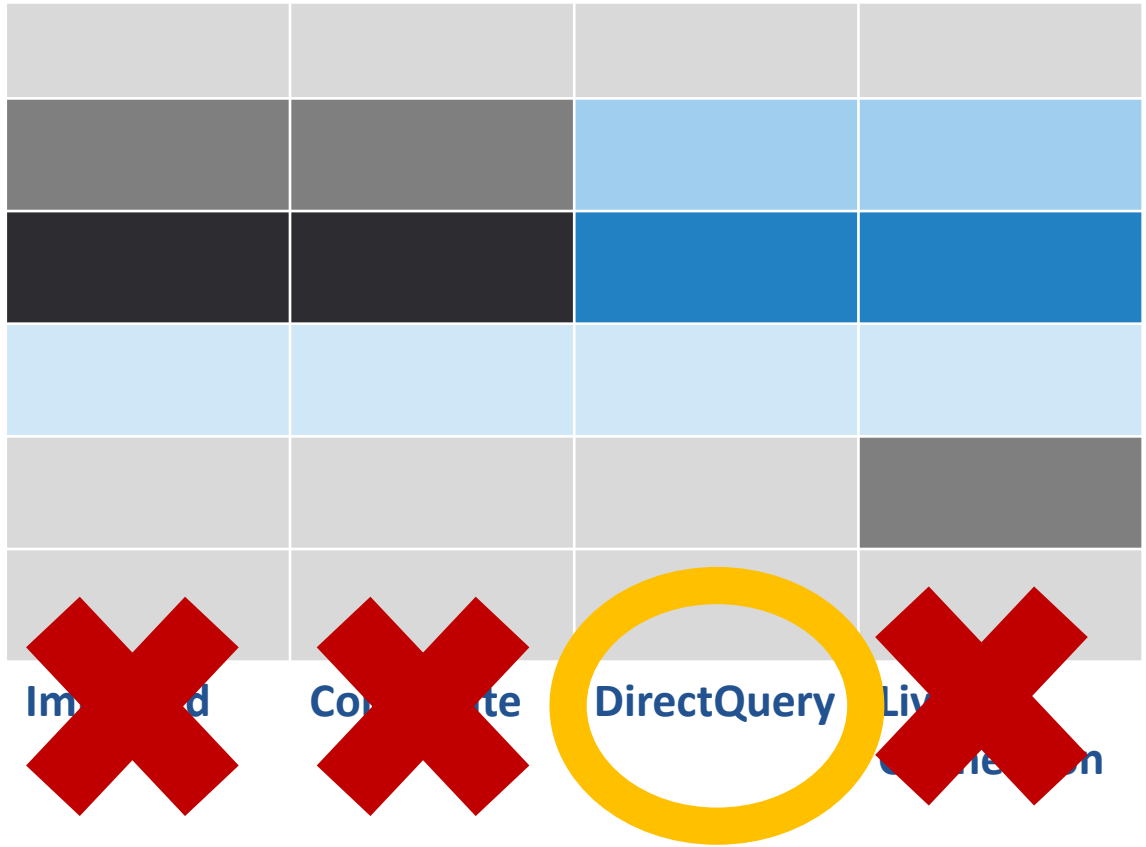
500 GB source SQL database

Need data within 10 minutes of receipt

Maintained by IT

No existing Analysis Services

Summarized data with drillthroughs



Final Thoughts

Default to imported and deviate when there is a reason

Data size and required latency will indicate if DirectQuery is needed

Try to reuse existing assets, but don’t force it where it doesn’t fit

Find Your Balance

Intentionally manage tradeoffs between

- Control
- Complexity
- Performance
- Security
- Cost

Questions?

Meagan Longoria
@Mmarie
DataSavvy.Me
Meagan@dcac.com

Denny Cherry
& Associates Consulting

Your Data, Our Expertise
www.dcac.com