



# Power BI & Azure Data Platform

Farid Jalal | Lead Analytics Consultant



# Making business better

We are action oriented and passionate about helping our clients leverage Microsoft technology



## Diverse Team

Over 45 Awesome People



## Cloud First



## Gold Competency

Microsoft  
Partner



Gold Data Analytics  
Gold Cloud Platform

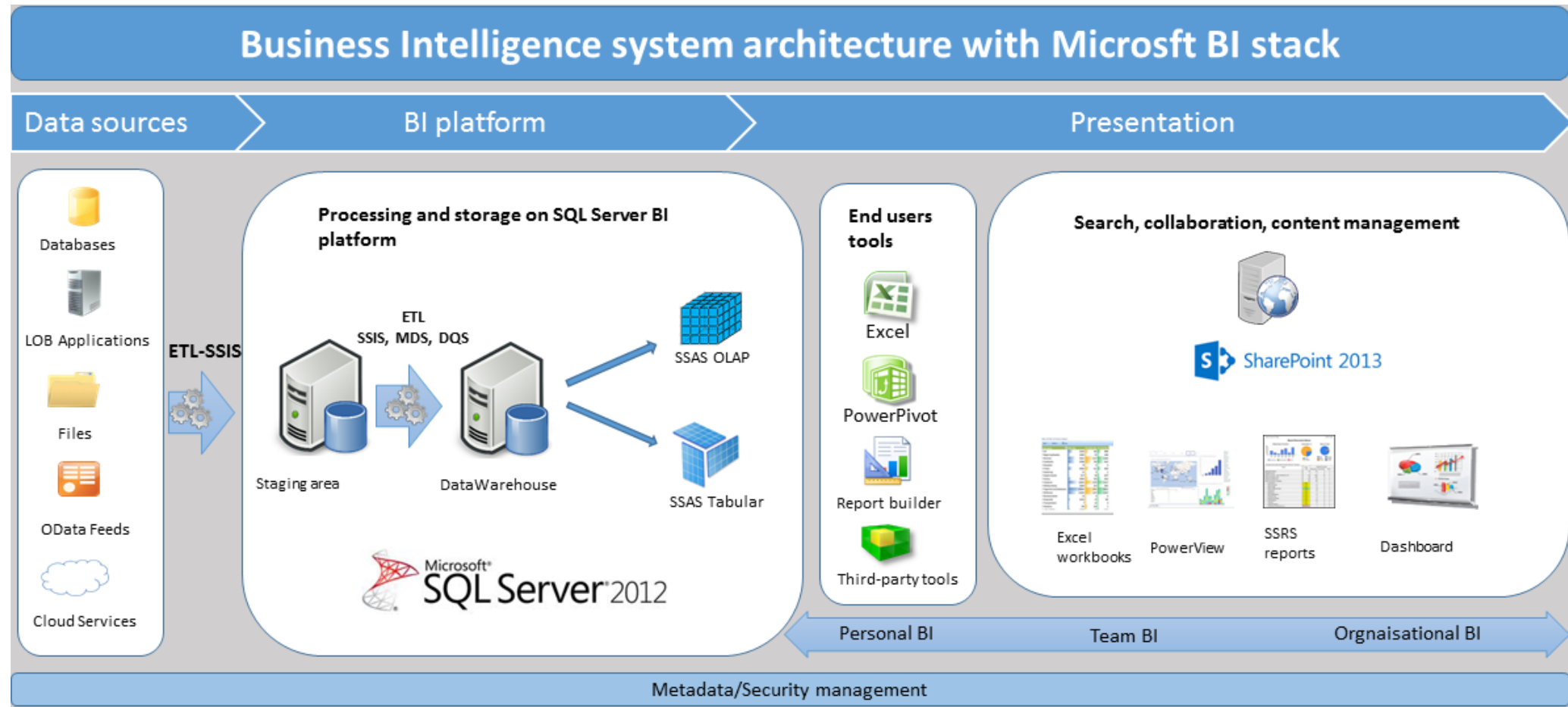


# Objective: To present

- Modern data warehouse reference architecture in Azure
  - Ingest
  - Store
  - Train & Prep
  - Model & Serve
- Typical (unaided) cloud journey
- Standard frameworks to get started



# Traditional Microsoft BI architecture





# Modern Azure architecture(s)

Search

analytics



Solutions

Modern data warehouse

Products

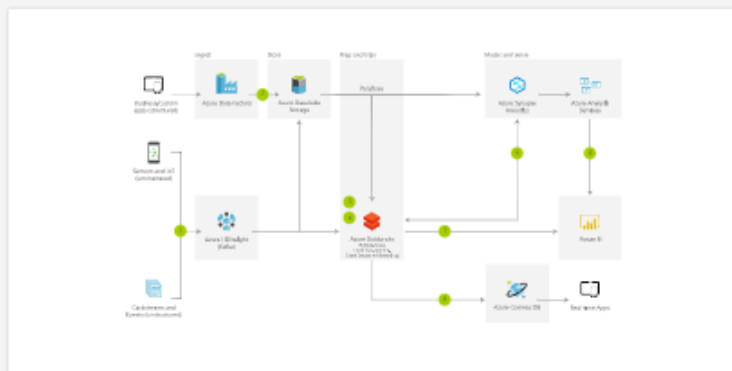
All

Tags

All

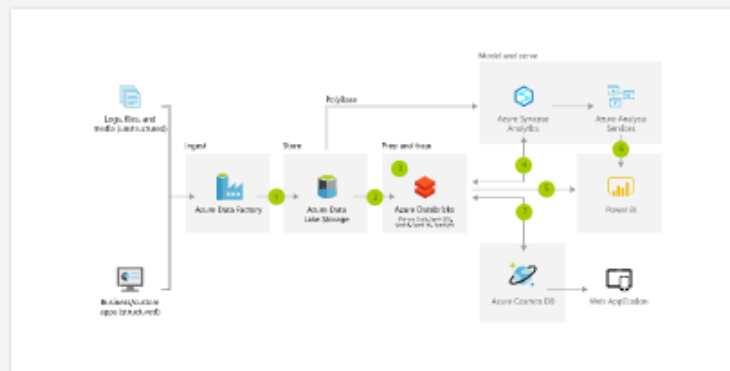
Industries

All



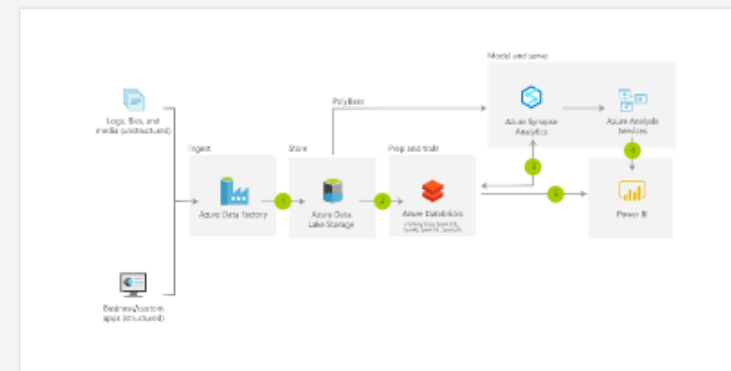
## Real-time analytics

Get insights from live, streaming data with ease. Capture data continuously from any IoT device or logs from website clickstreams and process it in near-real time.



## Advanced analytics on big data

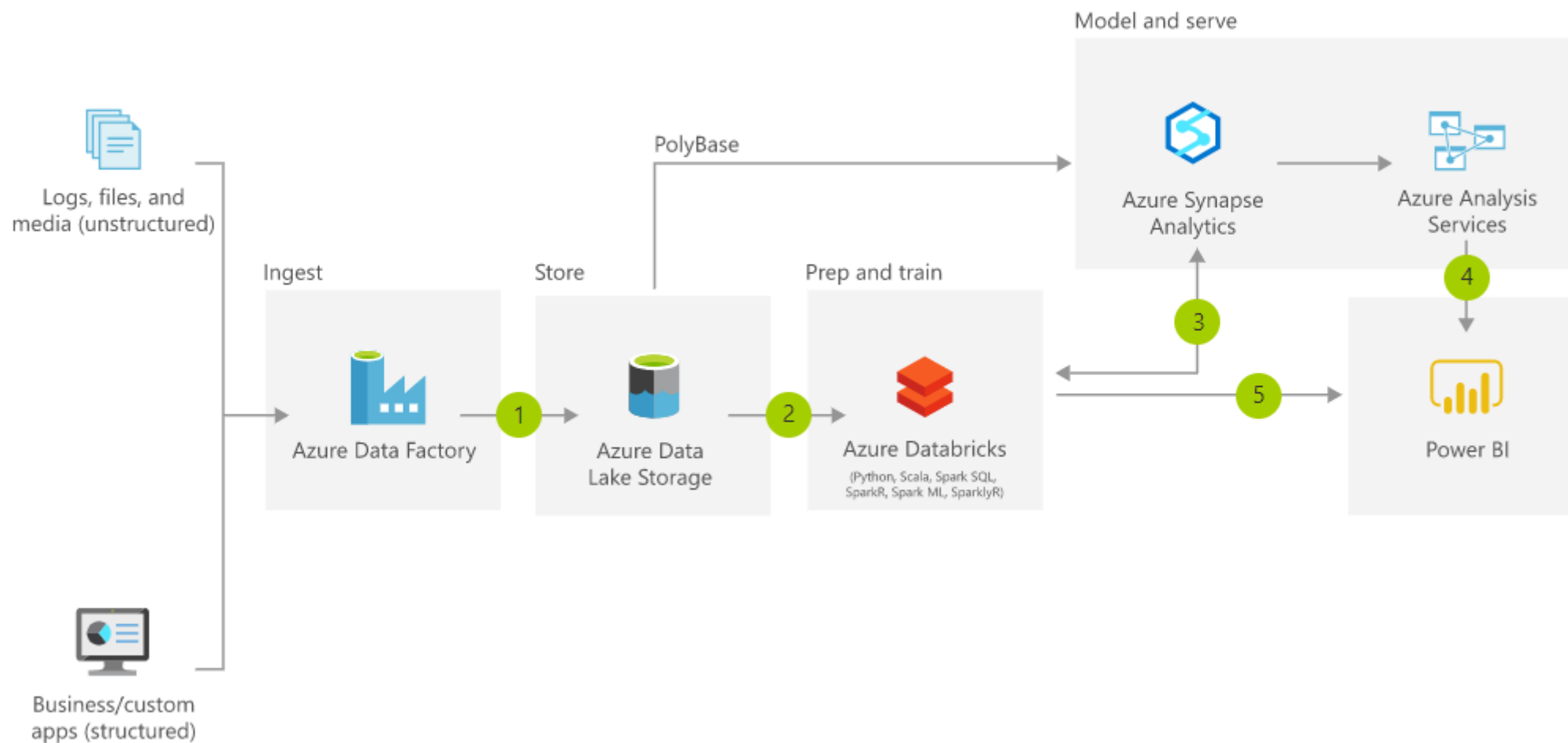
Transform your data into actionable insights using the best in class machine learning tools. This architecture allows you to combine any data at any scale, and to build...



## Modern data warehouse

A modern data warehouse lets you bring together all your data at any scale easily, and to get insights through analytical dashboards, operational reports, or advanced ...

# Modern Data Warehouse on Azure



# Ingest

*ETL vs ELT*



# SSIS (Integration Services)

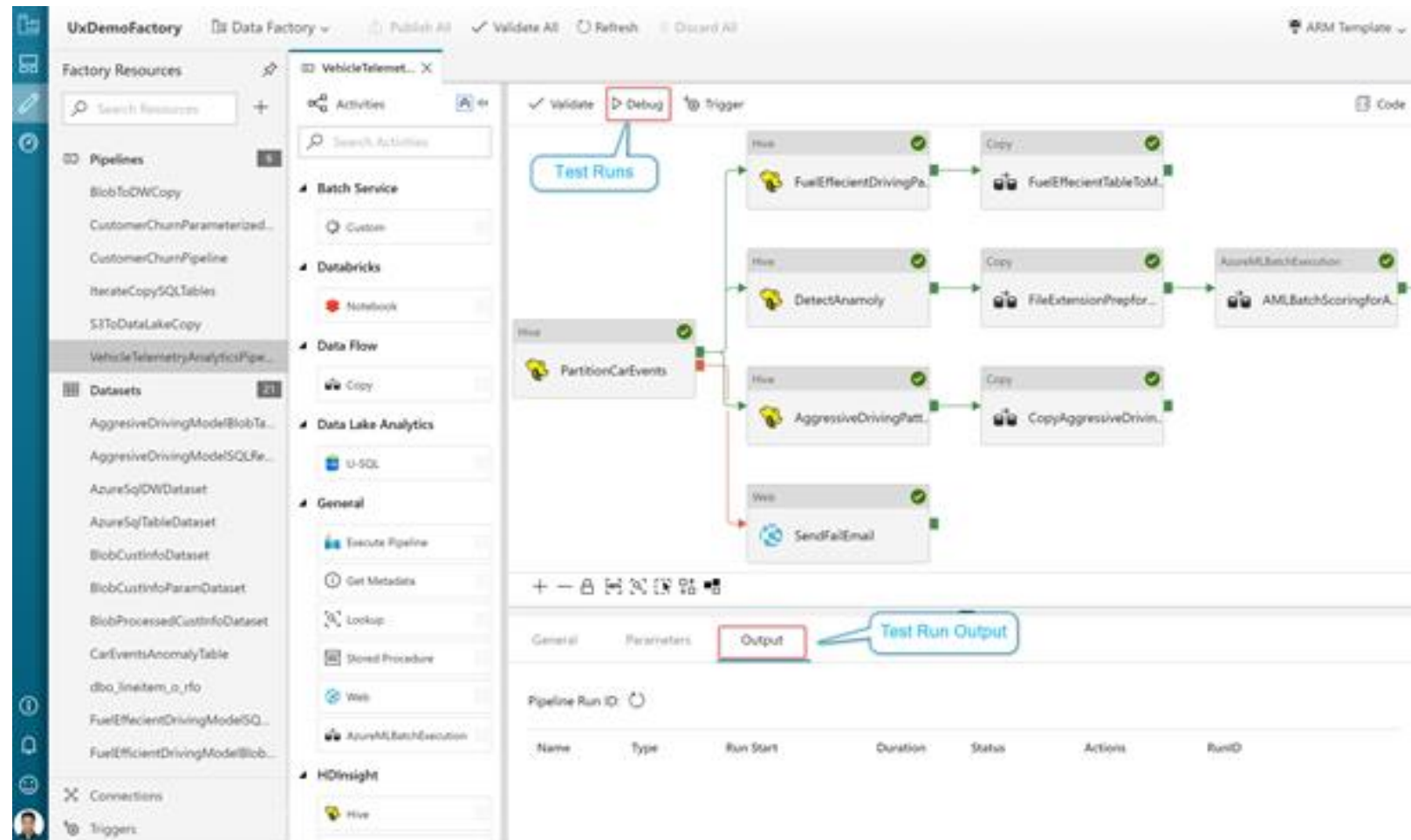
- Still relevant
  - Large developer base and community
  - Both IaaS and PaaS (as part of Data Factory v2)
  - Easy-to-use and stable
- However
  - Typically management framework is required – monitoring, dependencies
  - Cloud connectors not natively available
  - Does not suit well for self-service scenarios
- When to use?
  - Migration of SSIS workloads to the cloud
  - Existing SSIS skills & traditional data sources





# ADF (Azure Data Factory) v2

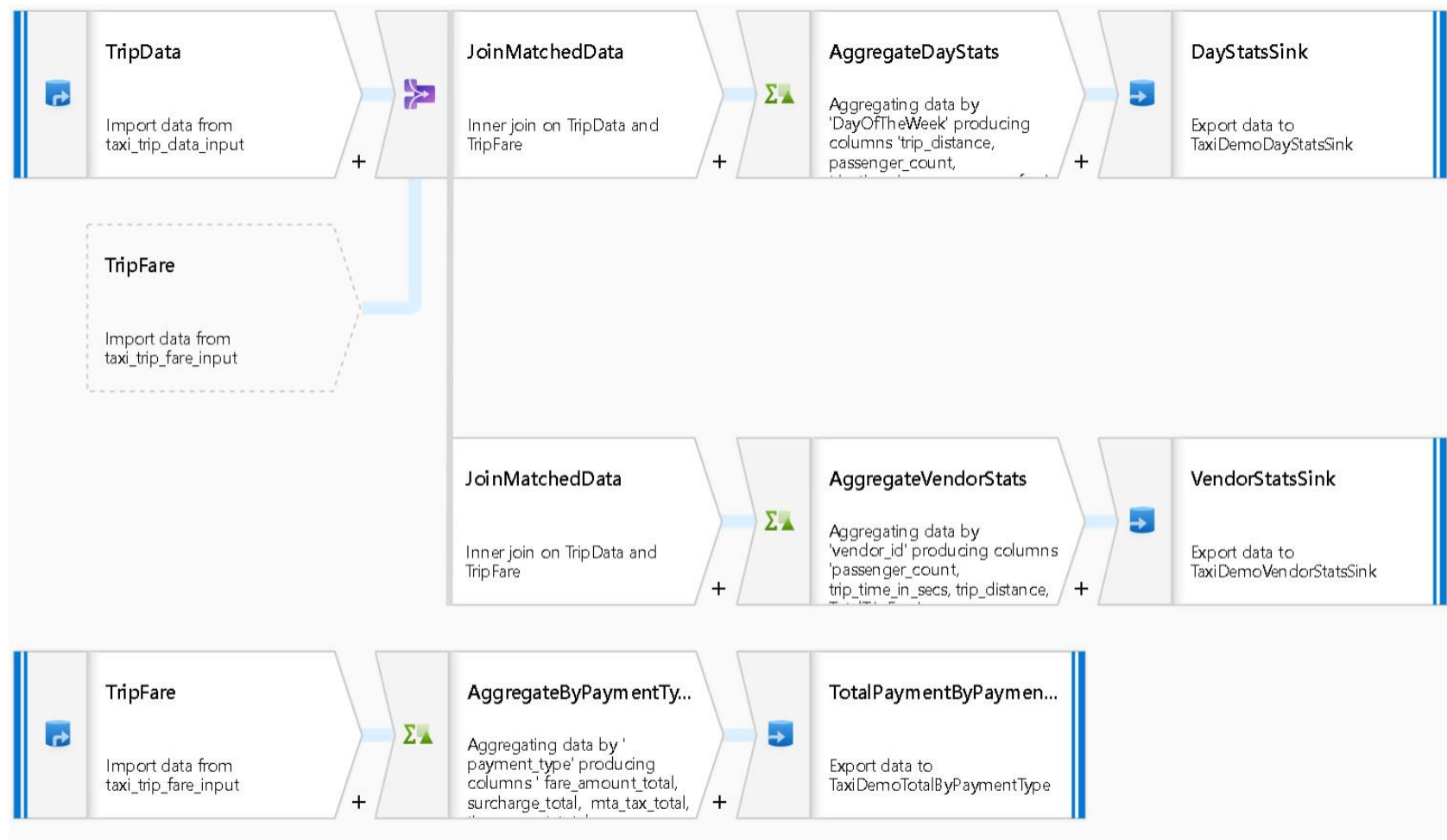
- Azure PaaS service
- 90+ connectors
- Think: Control Flow
- Simple graphical UI
  - Development
  - Monitoring
- Native/SSIS/Spark
  - Additional compute not necessary
- When to use?
  - New cloud sources





# ADF – Mapping Data Flows

- Visual data transformation in Azure Data Factory
- Think: DFT in SSIS
- Pipelines use scaled-out Spark clusters
  - Additional compute is necessary
- When to use?
  - Ease of use
  - Scalability





# ADF – Wrangling Data Flows (Public Preview)

- Visual data preparation in Azure Data Factory
- Think: Power Query in Excel / Power BI
- Pipelines use scaled-out Spark clusters
  - Additional compute is necessary
- When to use?
  - Ease of use
  - Scalability

Microsoft Azure | Data Factory

Search resources

Data Factory | Publish All | Validate All | Refresh | Discard All | Data Flow Debug | ARM Template

Lego\_Prep\_Leg... X

Refresh | Options | Manage columns | Transform table | Reduce rows | Add column | Combine tables | Settings | Code

ADFRsource [2]  
UserQuery

Table.RenameColumns("#Removed columns", {{ "set\_num", "SetID"}, {"name", "SetName"}, {"year", "ReleaseYear"}, ...

	SetID	SetName	ReleaseYear	NumParts	Theme1	Theme2	Theme3
1	001-1	Gears	1965	43	Technic	(null)	(null)
2	002-1	4.5V Samsonite Gears Motor Set	1965	3	Technic	(null)	(null)
3	1030-1	TECHNIC I: Simple Machines Set	1985	191	Technic	(null)	(null)
4	1038-1	ERBIE the Robo-Car	1985	120	Technic	(null)	(null)
5	1039-1	Manual Control Set 1	1986	39	Technic	(null)	(null)
6	0016-1	Castle Mini Figures	1978	15	Castle	(null)	(null)
7	10000-1	Guarded Inn	2001	256	Castle	(null)	(null)
8	10039-1	Black Falcon's Fortress	2002	431	Castle	(null)	(null)
9	024119931...	DC Super Heroes: Character Encyclopedia	2016	5	Books	(null)	(null)
10	024135752...	Star Wars: The Visual Dictionary, New Editi...	2019	0	Books	(null)	(null)
11	075666853...	Atlantis: Brickmaster	2010	157	Books	(null)	(null)
12	075667280...	Pirates: Brickmaster	2009	162	Books	(null)	(null)
13	075667281...	Castle: Brickmaster	2009	141	Books	(null)	(null)
14	075668276...	Ninjago Brickmaster	2011	154	Books	(null)	(null)

Applied steps

- Source
- Expanded CSV\_ADLS\_Leg...
- Merged queries
- Expanded CSV\_ADLS\_Leg...
- Merged queries 1
- Expanded CSV\_ADLS\_Leg...
- Removed columns
- X Renamed columns

Reset Done



# Power BI Dataflows

- Self-service data prep
  - Power Query is everywhere!
- Power BI Premium licensing required
- Uses ADLS Gen2 as storage layer
  - BYO ADLS Gen2 account to Power BI
  - Attach CDM folders created by other services to Power BI as dataflows

Power Query

Edit Queries

Get Data Refresh Options Manage Columns Transform Table Reduce Rows Add Column Add Conditional Column Combine Tables

AccountLeads  
ActivityParty  
Email

Navigation 1: [[EntitySetName = "activityparties"]][Data]

pecode...	ispartydeleted	ownerid	participationtypem...	participationtypem...	partyid
1	FALSE	e05fcac4-8dd2-466b-89bb...	2	To Recipient	75c1be22-eef5-e711-a8...
2	FALSE	eb2133fb-c6f4-e711-a835...	8	Regarding	ba9e62a8-90df-e311-95...
3	FALSE	7c0ac606-c7f4-e711-a835...	9	Owner	7c0ac606-c7f4-e711-a8...
4	FALSE	db55e4f4-c6f4-e711-a835...	9	Owner	db55e4f4-c6f4-e711-a8...
5	FALSE	e05fcac4-8dd2-466b-89bb...	9	Owner	e05fcac4-8dd2-466b-89...
6	FALSE	db55e4f4-c6f4-e711-a835...	8	Regarding	fc9e62a8-90df-e311-95...
7	FALSE	e05fcac4-8dd2-466b-89bb...	9	Owner	e05fcac4-8dd2-466b-89...
8	FALSE	e05fcac4-8dd2-466b-89bb...	9	Owner	e05fcac4-8dd2-466b-89...
9	FALSE	b409c606-c7f4-e711-a835...	8	Regarding	d140d6c6-43cc-e411-80...
10	FALSE	c209c606-c7f4-e711-a835...	8	Regarding	a8c01ca3-c7d1-e411-80...
11	FALSE	db55e4f4-c6f4-e711-a835...	9	Owner	db55e4f4-c6f4-e711-a8...
12	FALSE	7c0ac606-c7f4-e711-a835...	8	Regarding	63a0e5b9-88df-e311-b8...
13	FALSE	e05fcac4-8dd2-466b-89bb...	9	Owner	e05fcac4-8dd2-466b-89...
14	FALSE	7c0ac606-c7f4-e711-a835...	8	Regarding	2dbec441-5327-e511-8c...
15	FALSE	080ac606-c7f4-e711-a835...	8	Regarding	ce0e0283-5bf2-e311-94...
16	FALSE	e05fcac4-8dd2-466b-89bb...	8	Regarding	53869f5e-5c91-e611-80...
17	FALSE	e05fcac4-8dd2-466b-89bb...	9	Owner	e05fcac4-8dd2-466b-89...
18	FALSE	db55e4f4-c6f4-e711-a835...	9	Owner	db55e4f4-c6f4-e711-a8...
19	FALSE	e05fcac4-8dd2-466b-89bb...	9	Owner	e05fcac4-8dd2-466b-89...
20					

Name: ActivityParty

Applied Steps: Source, Navigation 1, Navigation 2

Next

# Store

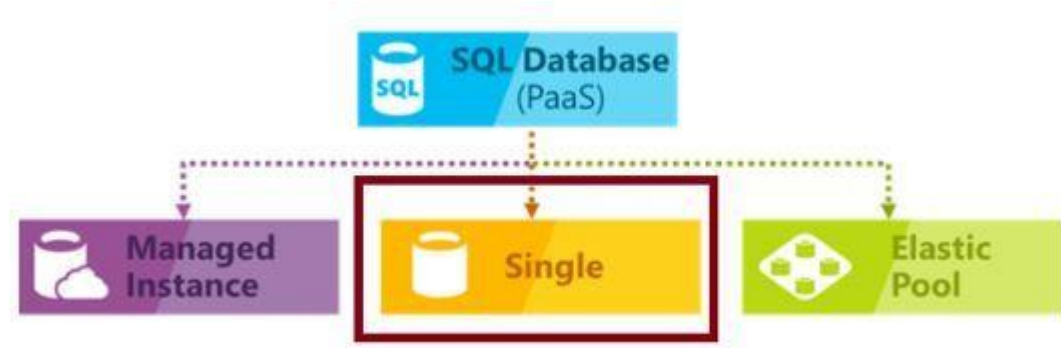


# IaaS – SQL Server VM

- Still relevant
  - Flexible, many options available
  - Easily configured & well understood
  - VMs can be paused to reduce costs
  - Free SQL Server Dev licensing available
- But associated admin overhead & not built to take advantage of PaaS
- When to use?
  - Workloads that aren't PaaS - SSRS, MDS, DQS
  - Cost considerations – a single VM for all services?
  - Lift-and-shift (and features like cross DB queries required)



# Azure SQL Databases



- Single
  - Several provisioned compute and serverless compute tier choices for SQL applications created in the cloud
  - Decoupled storage and compute but no pause option, only scale down
- Elastic Pool
  - Shared resource model - all the DBs share predefined resources
  - Elastic pool is best for new SaaS apps, or modernizing existing apps to SaaS
- Managed Instance (MI)
  - Near 100% compatibility with on-prem while fully PaaS
  - Seamless migration (and features like cross DB queries required)



# Azure Data Lake Storage Gen 2

- Best of breed, low-cost
  - Optimized for unstructured data
  - With hierarchical namespaces
  - Not necessarily ACID compliant
- Inbuilt Dynamics integration
  - Real-time, transactional data
  - D365: Customer Engagement
  - D365: Finance & Operations
- Power BI dataflows uses ADLS Gen2 as storage layer
- When to use?
  - Staging layer (cloud sources, ad-hoc analysis) and more

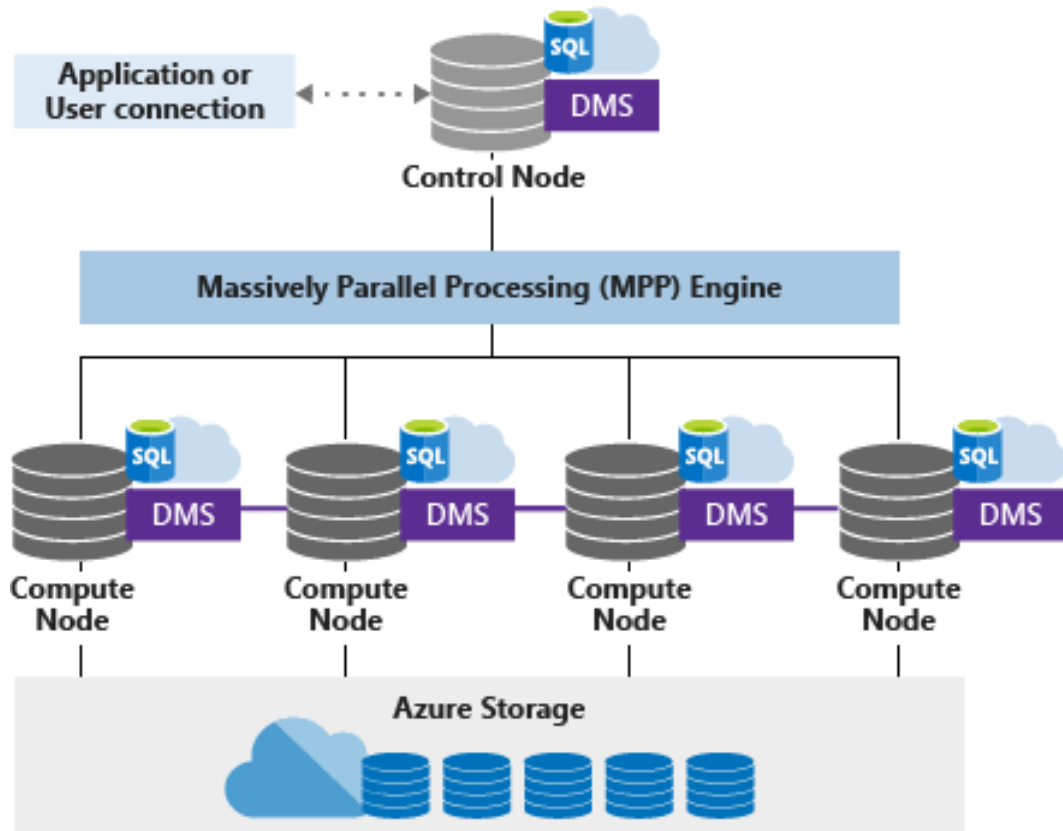
The screenshot displays the Azure Data Explorer interface. On the left, the 'EXPLORER' pane shows a hierarchical view of storage resources, including 'Storage Accounts', 'Blob Containers', 'File Shares', 'Queues', and 'Tables'. The 'Tables' section is expanded, showing a list of tables including '\$MetricsCapacityBlob', '\$MetricsHourPrimaryTransactionsBlob', '\$MetricsHourPrimaryTransactionsFile', '\$MetricsHourPrimaryTransactionsQueue', and '\$MetricsHourPrimaryTransactionsTable'. The '\$MetricsHourPrimaryTransactionsTable' is selected. The main pane shows a table with the following columns: 'PartitionKey', 'RowKey', 'Timestamp', 'AnonymousAuthorizationError', 'AnonymousClientOtherError', and 'AnonymousClientTI'. The table contains 20 rows of data, with timestamps ranging from 2019-08-07T1200 to 2019-08-07T2100. The bottom pane shows the 'Properties' tab for the selected table, displaying the URL 'https://cawagrs.table.core.windows.net/\$MetricsHourPrimaryTransactionsBlob' and the type 'Table'. The 'Activities' tab shows a 'Clear completed' and 'Clear successful' message.

PartitionKey	RowKey	Timestamp	AnonymousAuthorizationError	AnonymousClientOtherError	AnonymousClientTI
20190807T1200	system:All	2019-08-07T13:23:18.906Z	0	0	0
20190807T1200	user:All	2019-08-07T13:23:00.262Z	0	0	0
20190807T1200	user:GetBlobServiceProperties	2019-08-07T13:23:00.262Z	0	0	0
20190807T1300	user:All	2019-08-07T14:23:24.965Z	0	0	0
20190807T1300	system:All	2019-08-07T14:23:24.969Z	0	0	0
20190807T1400	system:All	2019-08-07T15:59:56.992Z	0	0	0
20190807T1400	user:All	2019-08-07T15:59:56.992Z	0	0	0
20190807T1500	system:All	2019-08-07T16:22:11.879Z	0	0	0
20190807T1500	user:All	2019-08-07T16:22:11.884Z	0	0	0
20190807T1600	user:All	2019-08-07T17:21:06.547Z	0	0	0
20190807T1600	system:All	2019-08-07T17:21:37.242Z	0	0	0
20190807T1600	user:GetBlobServiceProperties	2019-08-07T17:21:06.546Z	0	0	0
20190807T1700	user:All	2019-08-07T18:23:46.057Z	0	0	0
20190807T1700	system:All	2019-08-07T18:23:46.059Z	0	0	0
20190807T1800	user:All	2019-08-07T19:20:58.332Z	0	0	0
20190807T1800	system:All	2019-08-07T19:20:58.334Z	0	0	0
20190807T1900	user:All	2019-08-07T20:24:26.560Z	0	0	0
20190807T1900	system:All	2019-08-07T20:24:26.562Z	0	0	0
20190807T2000	user:GetBlobServiceProperties	2019-08-07T21:21:46.345Z	0	0	0
20190807T2000	system:All	2019-08-07T21:22:13.415Z	0	0	0
20190807T2000	user:All	2019-08-07T21:21:46.345Z	0	0	0
20190807T2100	user:All	2019-08-07T22:21:26.186Z	0	0	0
20190807T2100	system:All	2019-08-07T22:21:26.191Z	0	0	0





# Azure SQL Data Warehouse / Synapse (?)



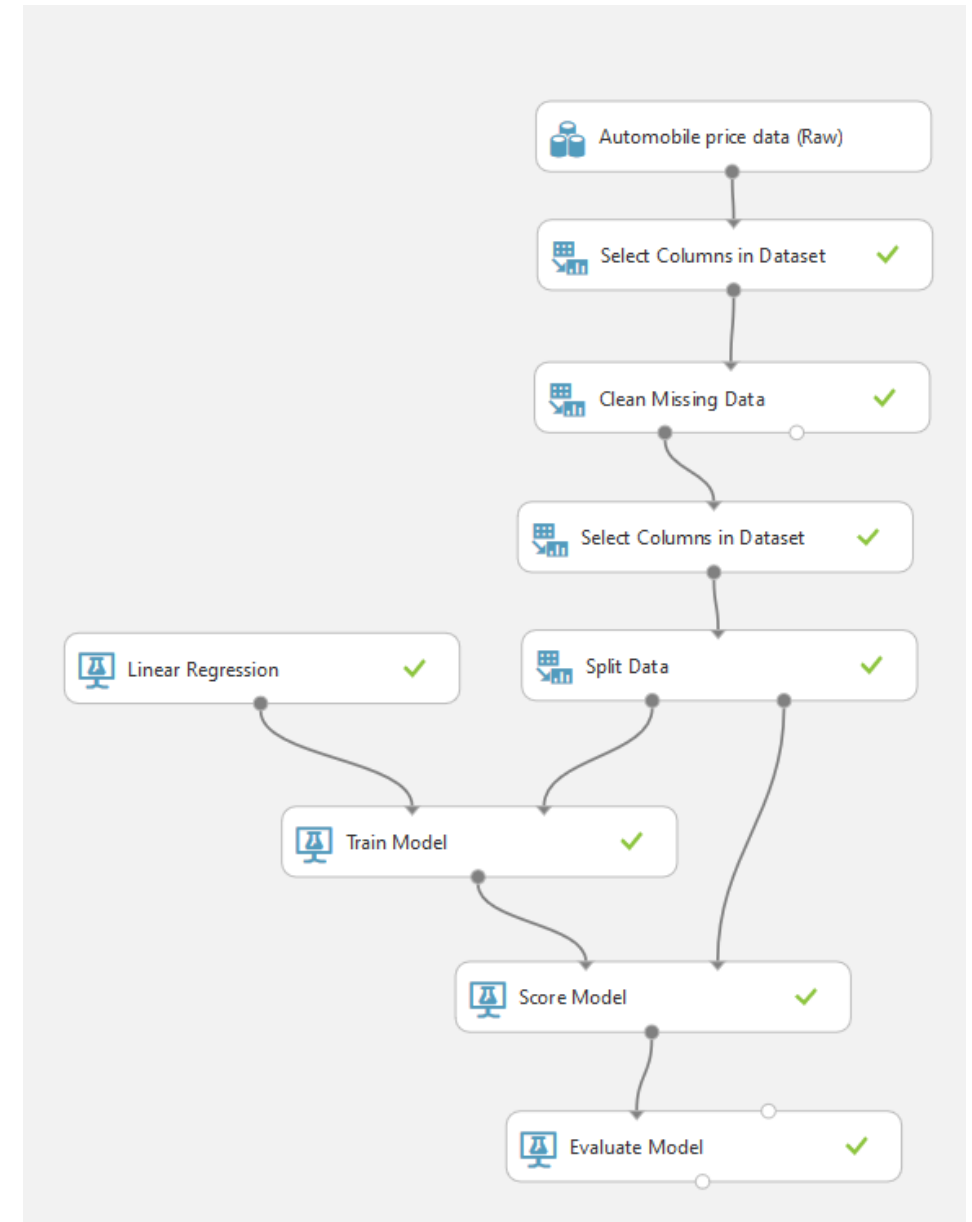
- SQL Server designed for analytical querying on big data
  - with Columnstore compression
  - MPP – Massively Parallel Processing
  - Not “Store” – rather “Model & Serve”
- Decoupled storage and compute
  - Only storage cost billed when compute is paused
- PolyBase with SQL/SSIS/ADF to load data
  - Cross database queries not supported
- When to use?
  - Large analytical workloads (> 50 TB)
  - Dev skills inhouse (Distributed tables)

# **Train & Prep**



# PaaS Services

- Azure Databricks
  - Spark-based analytics platform
  - 1<sup>st</sup> party service fully integrated with Azure
  - Dynamically scalable clusters
  - Notebook based UI
  - Scala, R, Python, SQL, Java
- Azure Machine Learning
  - Studio – Think: SSIS for Machine Learning
  - Automated ML UI
  - MLOps or DevOps for ML
  - Registry for model management
  - Deploy as cloud service
- AI/ML in Power BI



# Model & Serve



# Power BI Dataset

- Power BI Dataset
  - Report published to the Service
  - Pro licensing required to access
- When to use?
- Model Size vs Data Freshness vs Query Performance
  - Import - 1GB data size limit
  - DirectQuery - no size limit, limited sources
  - Composite Models & Aggregations
    - DirectQuery + Import in a dataset
  - Live Connections – existing models
    - Tabular / Multidimensional





# Power BI Premium

- Power BI Dataset
  - Report published to a Premium Workspace on the Service
  - Pro licensing not required to access
- When to use?
  - Import:
    - 10 GB data size limit
    - 12 GB refresh limit
  - Many considerations including licensing

The screenshot displays the Power BI Admin portal. The top navigation bar includes the Power BI logo, 'Admin portal', and icons for chat, settings, download, help, and user profile. A left sidebar contains icons for menu, favorites, recent items, workspace, users, and a user profile. The main content area is titled 'Admin portal' and lists options: Usage metrics, Users, Audit logs, Tenant settings, and Premium settings (which is selected). The 'Premium settings' section is expanded, showing 'USER PERMISSIONS'. Under 'Capacity admins', there is a 'Required' status and an 'Apply to:' field with a dropdown menu showing 'Avis Morin' and 'Enter email addresses'. Below this are 'Apply' and 'Cancel' buttons. Further down, 'Users with assignment permissions' is listed as 'Enabled for the entire organization'. At the bottom, 'WORKSPACES (148)' is shown with 'Remove all' and 'Assign workspaces' options, and a search bar labeled 'Search content...'.

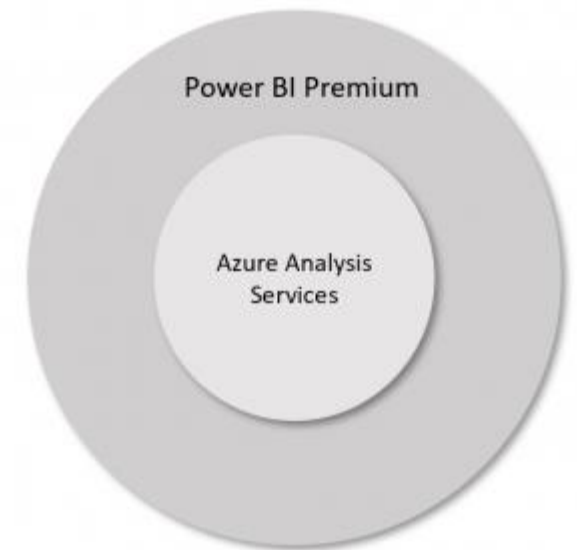


# Azure Analysis Services

- Azure PaaS service
- Feature-parity with SSAS (tabular models  $\geq$  1200 compatibility levels)
  - Scale up\down, out, pause, and resume as required
- Multidimensional models not supported
- Some admin overhead, for example:
  - PowerShell scripts required to process data models
  - Automation Account required to run PS scripts
- Power BI Premium datasets feature gap reducing

## SQL Server Analysis Services

- IaaS – SQL Server VM
- Multidimensional models



# Typical Cloud Journey





# Lift-and-shift, and add PaaS services when required

From

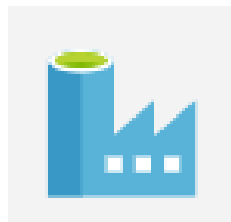


SQL, SSIS,  
MDS, SSAS



Power BI  
Reports

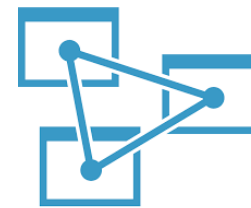
To



Data  
Factory



SQL, SSIS,  
MDS



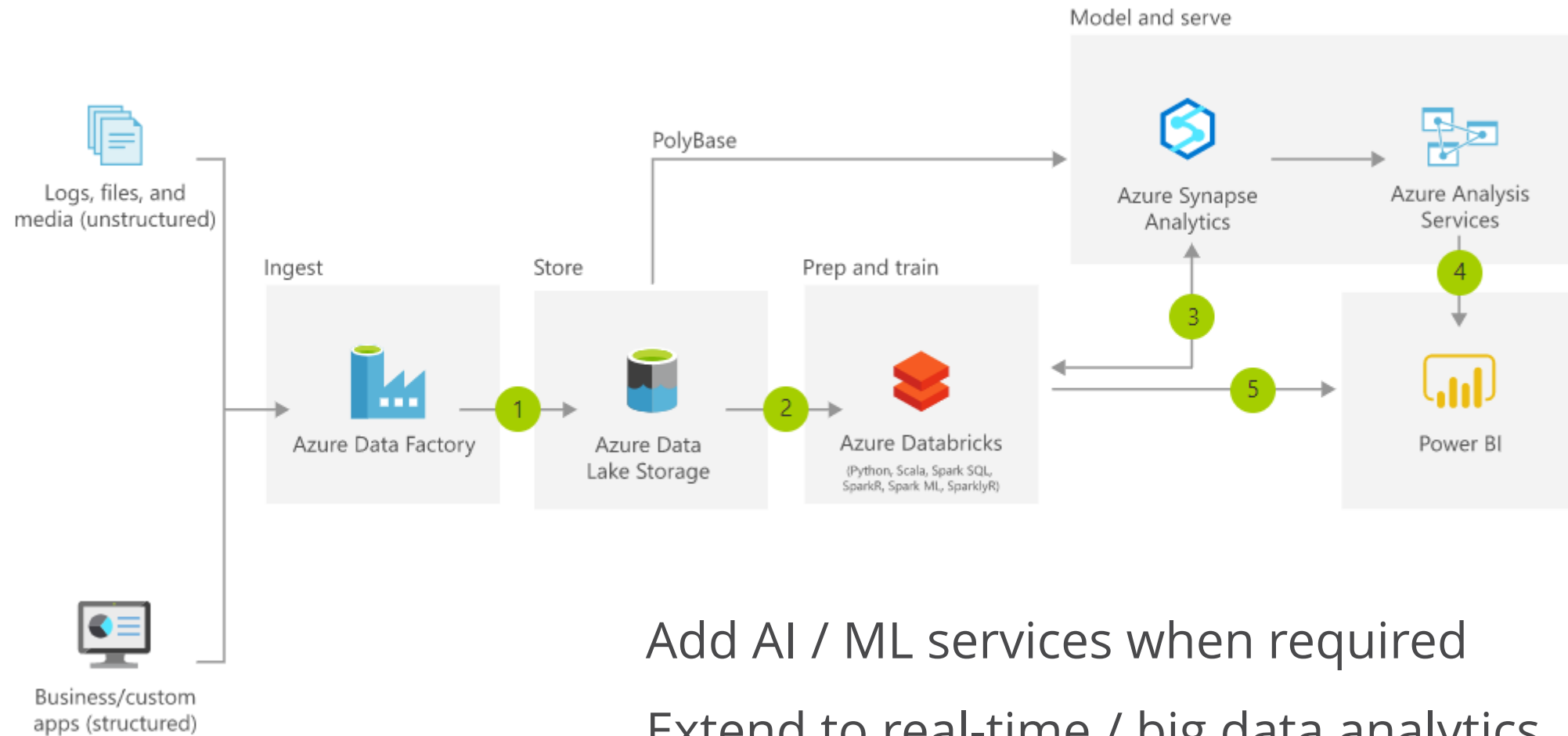
AAS



Power BI  
Reports



# Refactor / rearchitect modules to a modern DWH



# Standard Frameworks



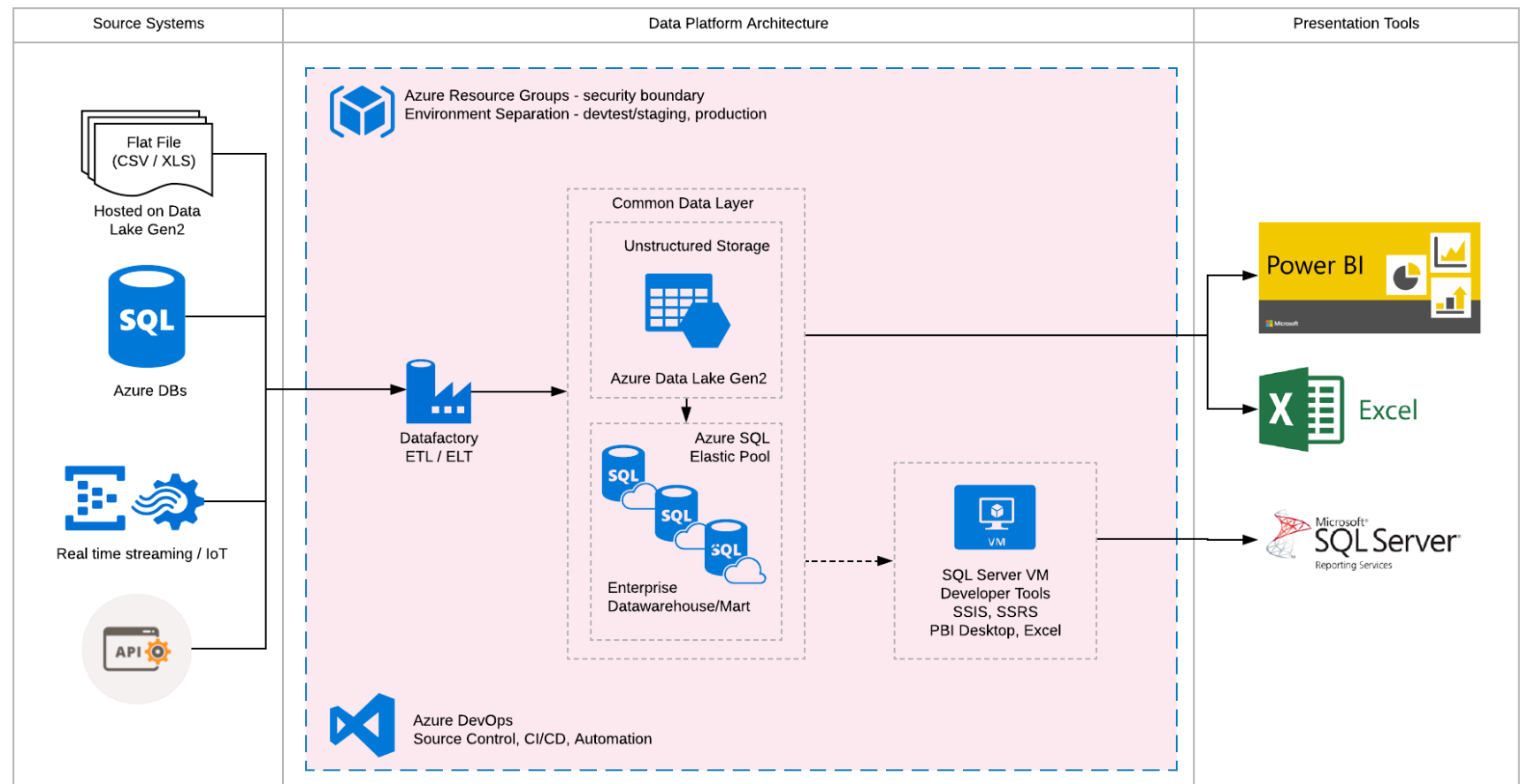
# Azure Data Platform: Standard (without VPN & Azure Analysis Services)

## Data Model

- Power BI dataset (SSRS reports do not consume this model)
  - Shared capacity; data model size, refresh performance & feature limitations

or

- Analysis Services on VM (SSRS reports consume this model, but IaaS - 2 VMs required)



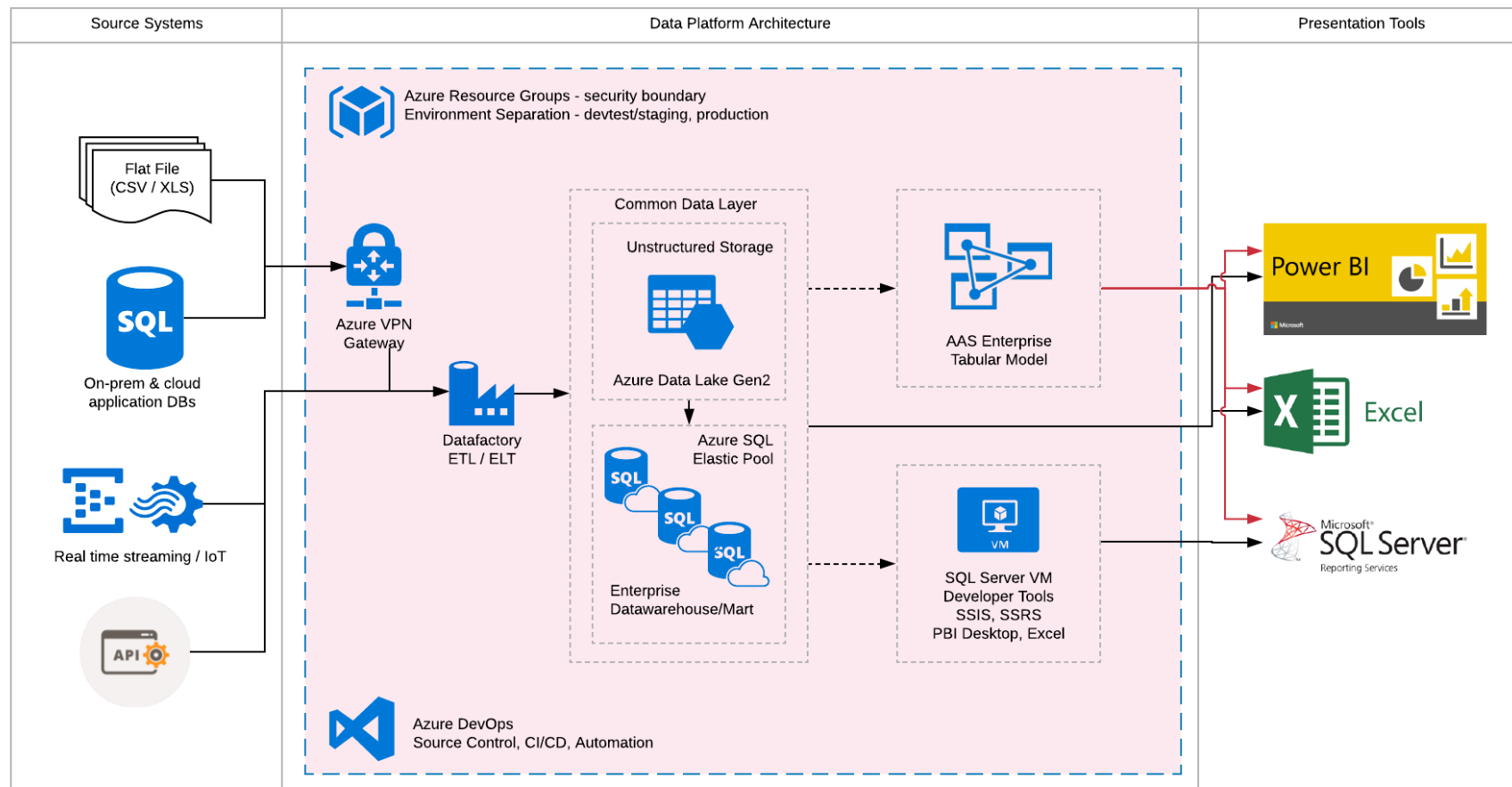
US \$2,100 pm



# Azure Data Platform: Custom (with VPN & Azure Analysis Services)

## Data Model

- Azure Analysis Services - S1 SKU (SSRS reports do not consume this model)
- Dedicated capacity and large data model; size & performance scalable at a cost



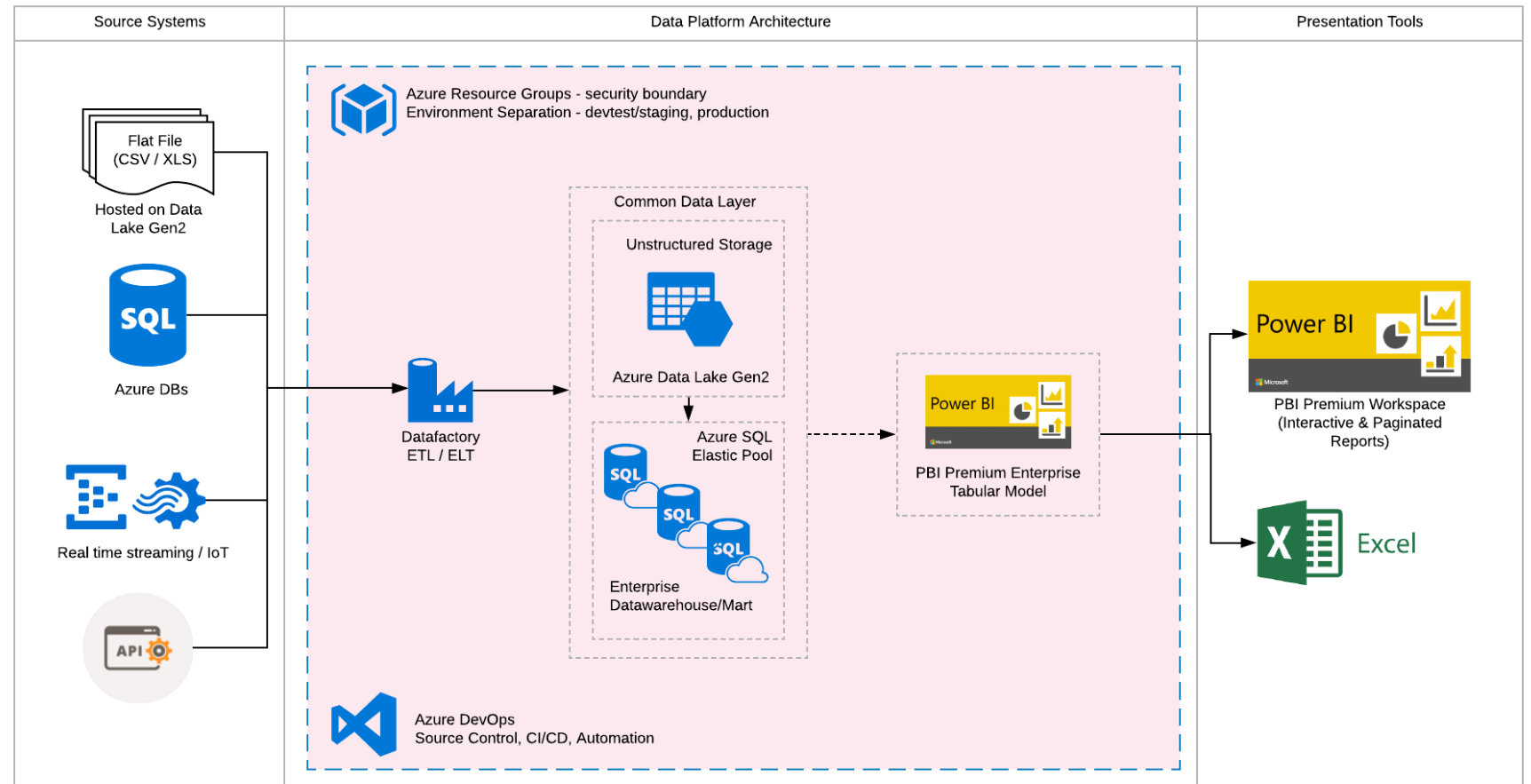
US \$4,400 pm



# Azure Data Platform with Power BI Premium (without VPN, Azure Analysis Services, VMs)

## Data Model

- Power BI dataset – P1 SKU (Power BI RS reports consume this model)
- Dedicated capacity and large data model; size & performance scalable at a cost



US \$1,000 pm + US \$5,000 pm per dedicated Power BI Premium resource



# Azure Data Platform with Power BI Premium (additional Premium goodness)

- Granular control of compute resource allocation
- Multi-geo support for workspaces (datasets, reports and dashboards)
- Incremental Refresh of datasets (with easily configurable partitioning)
- Paginated (.rdl) reports
  - Single pane of glass for interactive and paginated reports
  - Single data source (Power BI dataset) for interactive and paginated reports
- Power BI Dataflows
  - Self-service data prep experience with easily configurable data staging
- Azure Cognitive Services and Azure ML
  - Integration of AI into Power BI dataflows
  - Models created in Power BI can be exported to Azure ML
- Distribution by Pro users without recipients requiring Pro licenses to view content
  - Reports & dashboards in Power BI Service
  - Embed fully interactive reports in
    - Teams, SharePoint, Dynamics
    - Custom applications
- Connectivity - MS & third-party client applications & tools using XMLA endpoints
- Power BI Report Server (on-prem version included, if required)

walkerscott<sup>o</sup>

**Thank you, questions?**

## Contact Us

NZ

**Farid Jalal**

farid.jalal@walkerscott.co

027 494 0128

[www.walkerscott.co](http://www.walkerscott.co)

