walkerscott

Power BI & Azure Data Platform

Farid Jalal | Lead Analytics Consultant

Microsoft

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

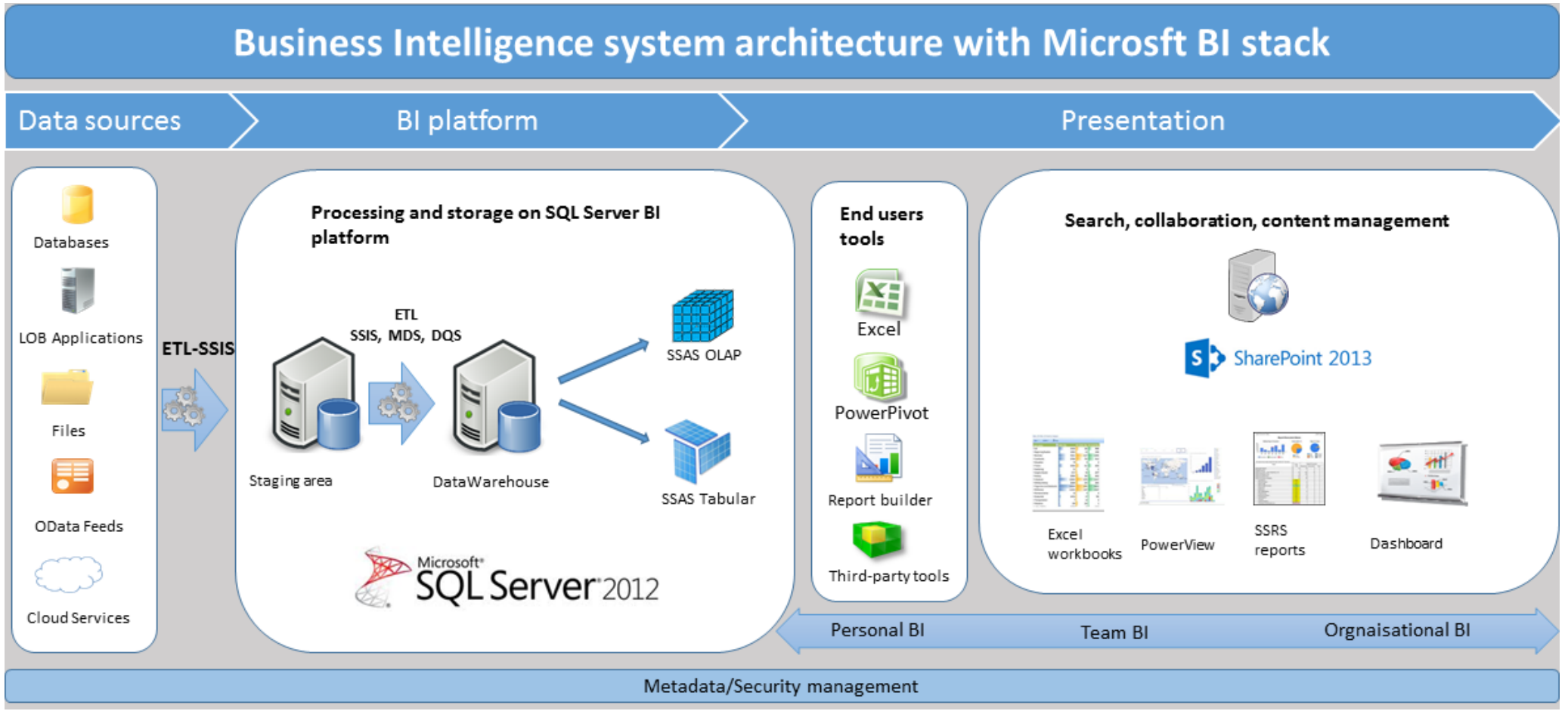
Microsoft Partner
Microsoft
Gold Data Analytics
Gold Cloud Platform

walkerscott

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



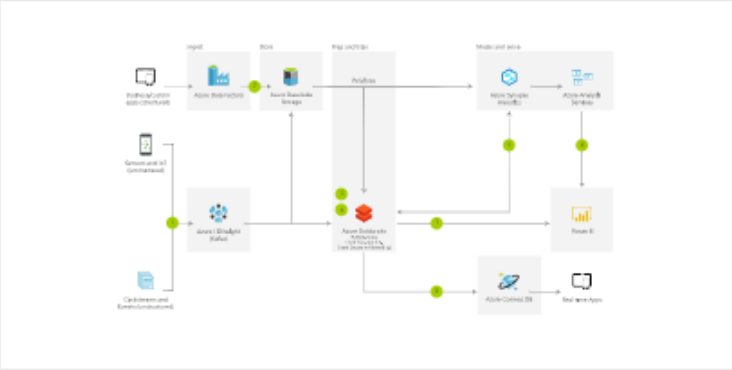
https://bipointblog.wordpress.com/2014/05/28/implementation-of-a-bi-system-using-microsoft-bi-stack-introduction

Modern Azure architecture(s)

Search

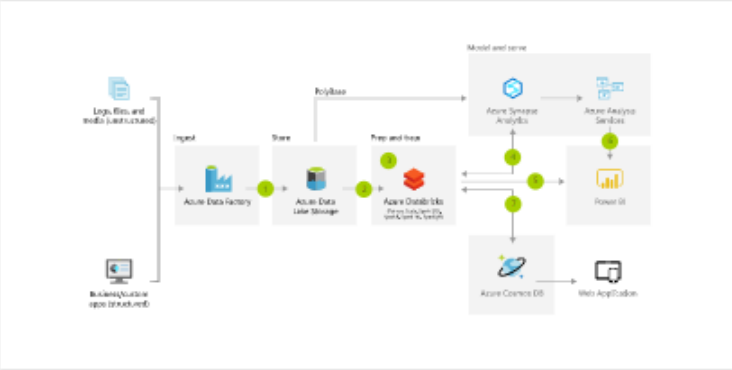
analytics

Solutions: Modern data warehouse | Products: All | Tags: All | Industries: All



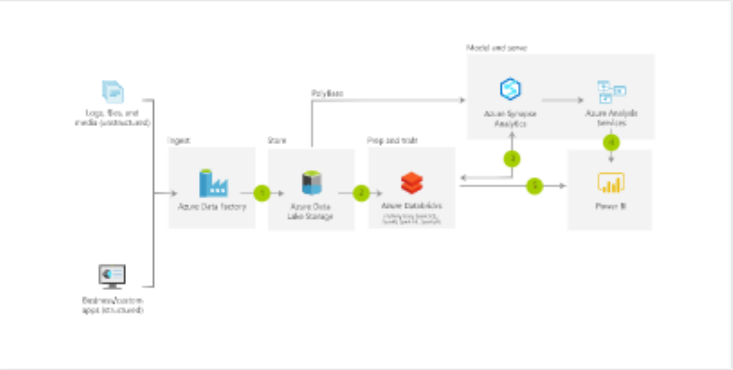
Real-time analytics

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 buil...



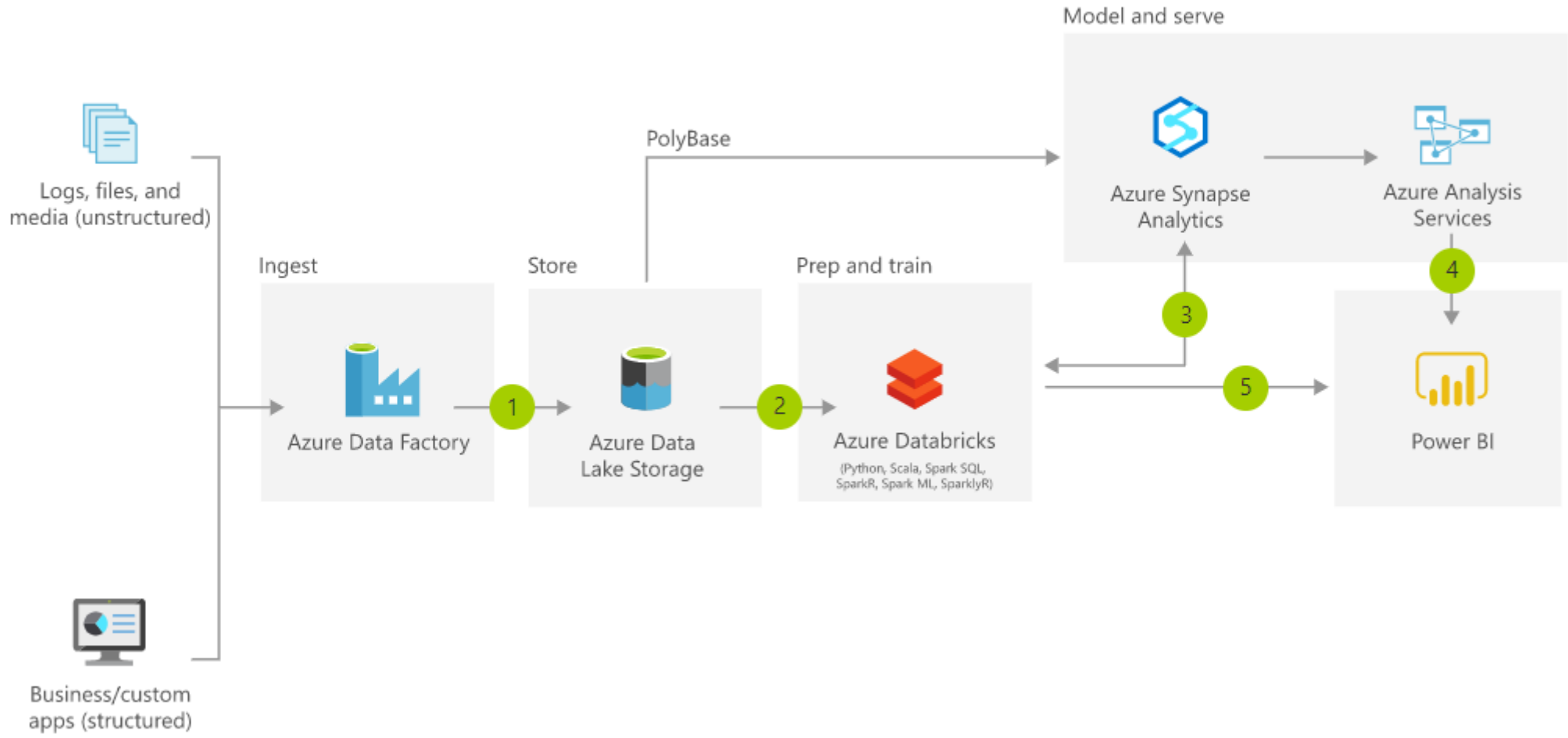
Modern data warehouse

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 ...

walkerscott

https://azure.microsoft.com/en-us/solutions/architecture/?query=analytics&solution=data-warehouse

Modern Data Warehouse on Azure



https://azure.microsoft.com/en-us/solutions/architecture/modern-data-warehouse

Ingest

ETL vs ELT

walkerscott

SSIS (Integration Services)

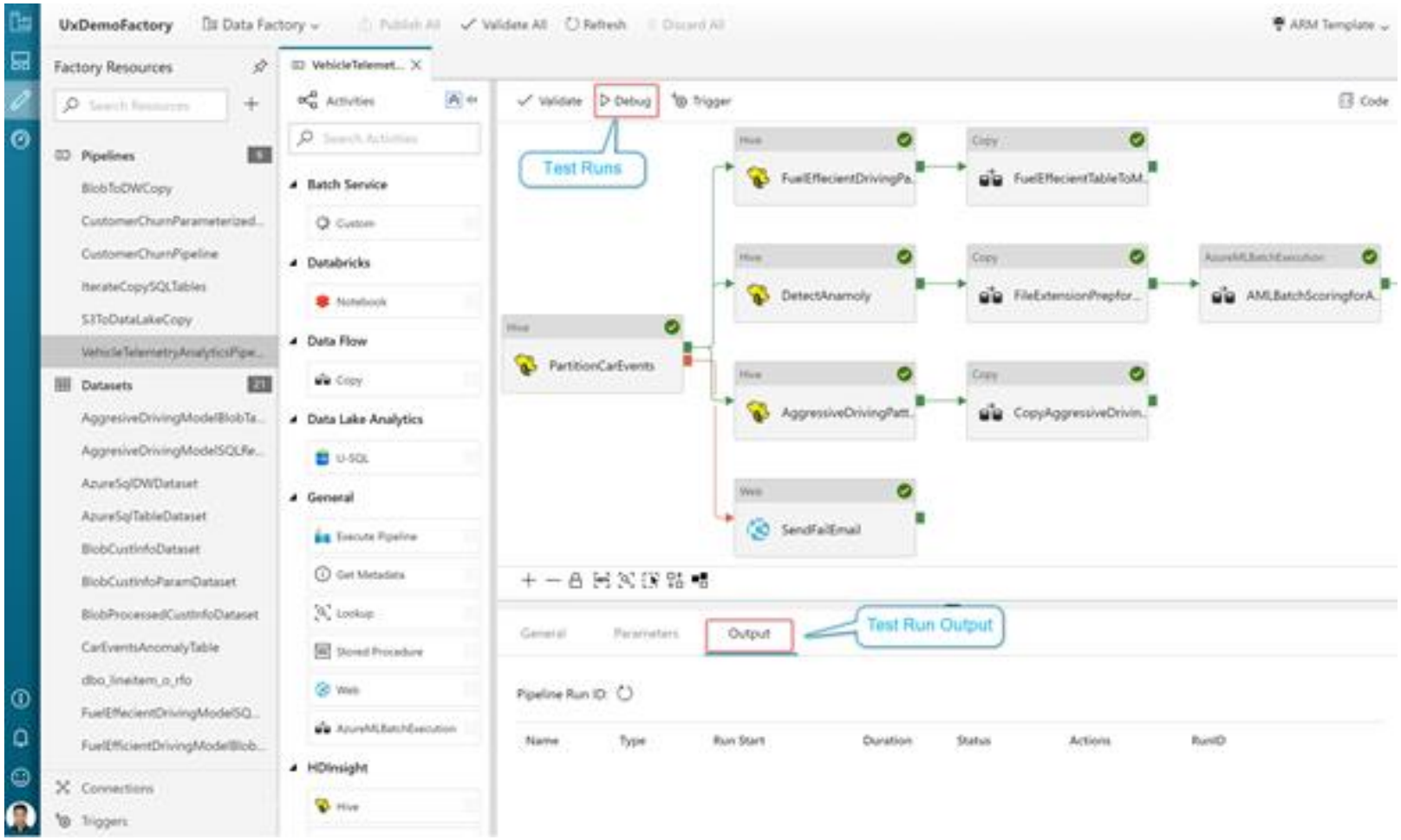
- Still relevant
  - Large developer base and community
  - Both IaaS and PaaS (as part of Data Factory v2)
  - Easy-to-use and stable

- 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



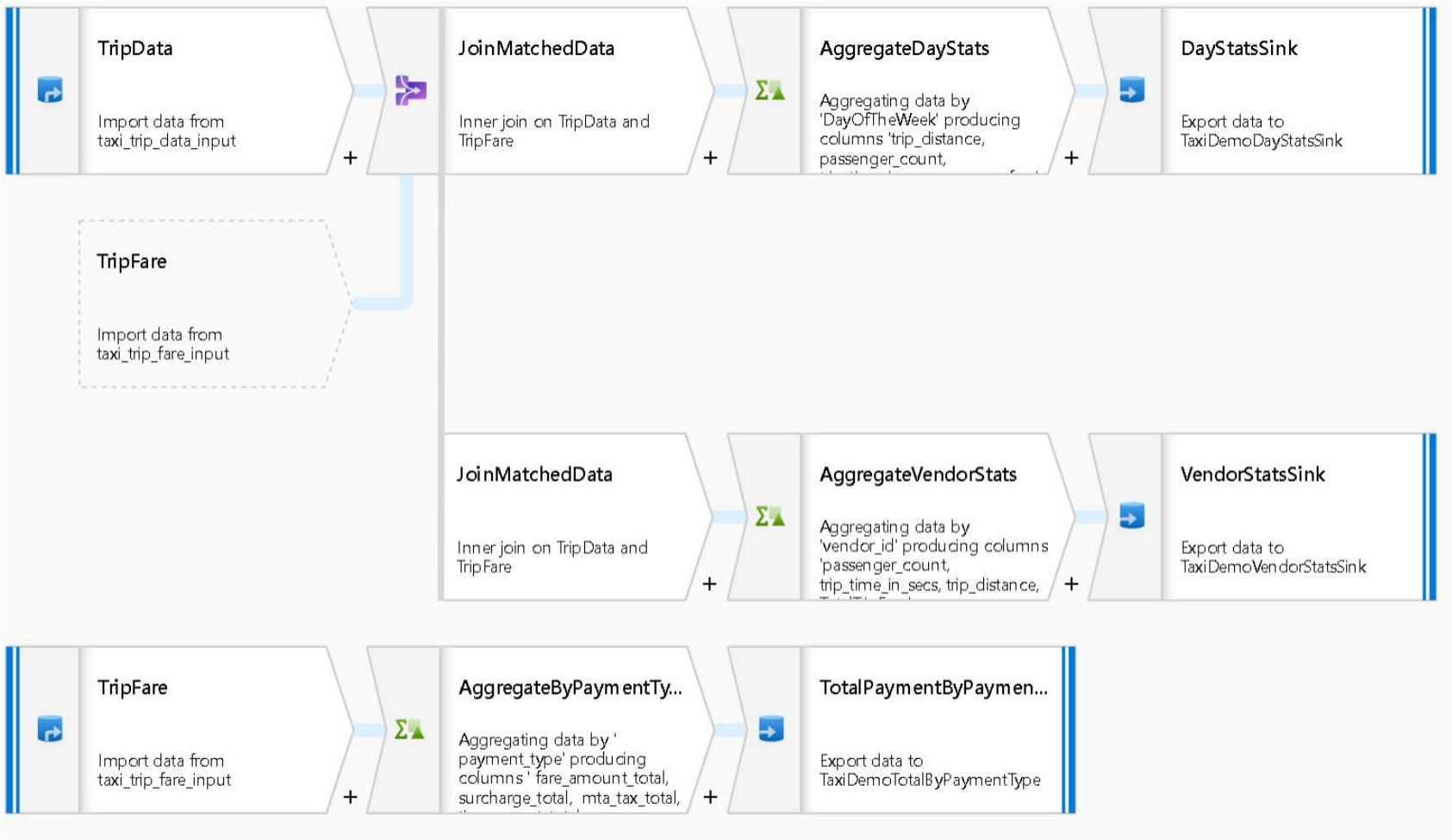
walkerscott

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

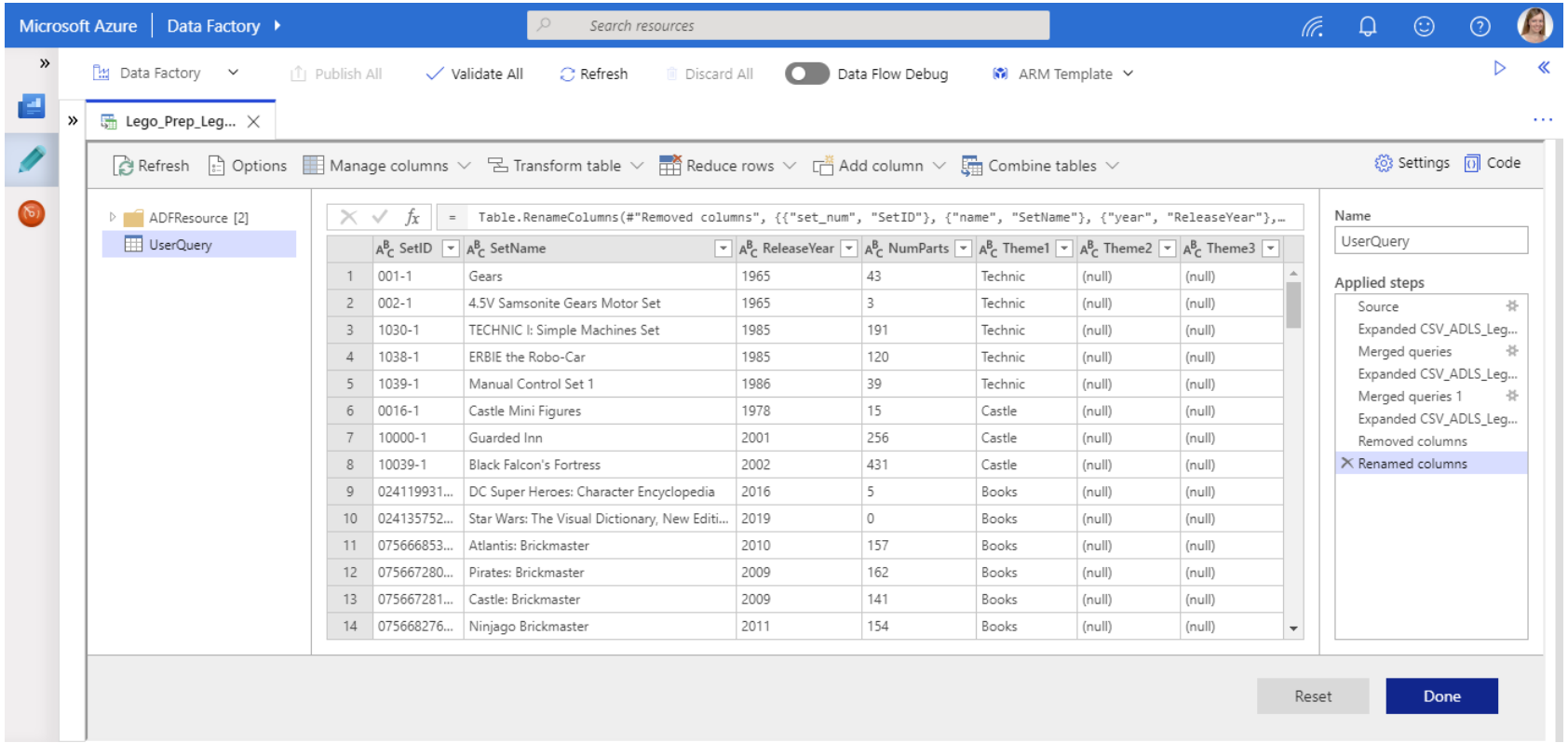
- When to use?
  - Ease of use
  - Scalability



walkerscott

ADF – Wrangling Data Flows (Public Preview)

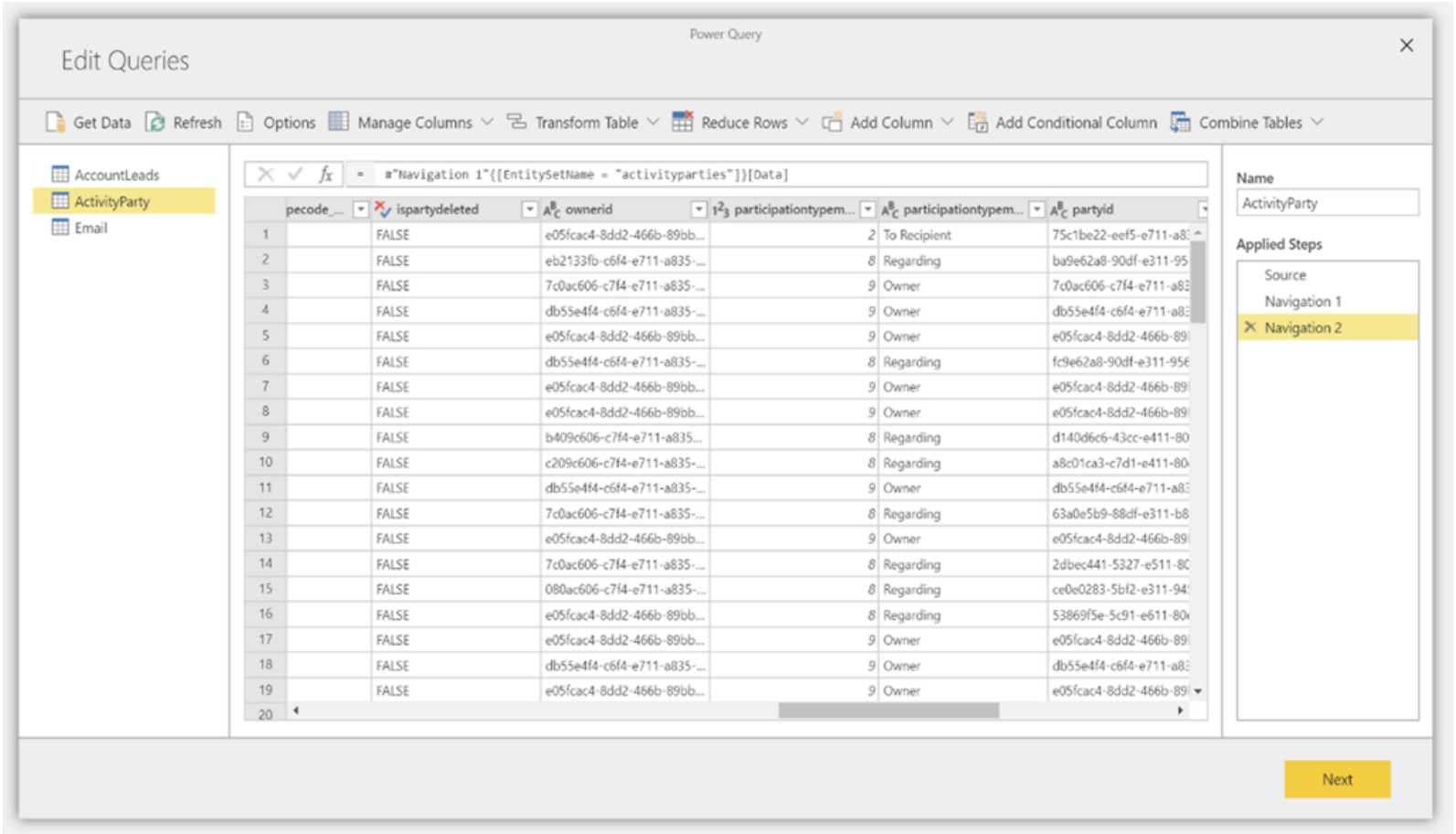
- Visual data preparation in Azure Data Factory
- Think: Power Query in Excel / Power BI
- Pipelines use scaled-out Spark clusters
  - Additional compute is necessary
- When to use?
  - Ease of use
  - Scalability



walkerscott

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



walkerscott

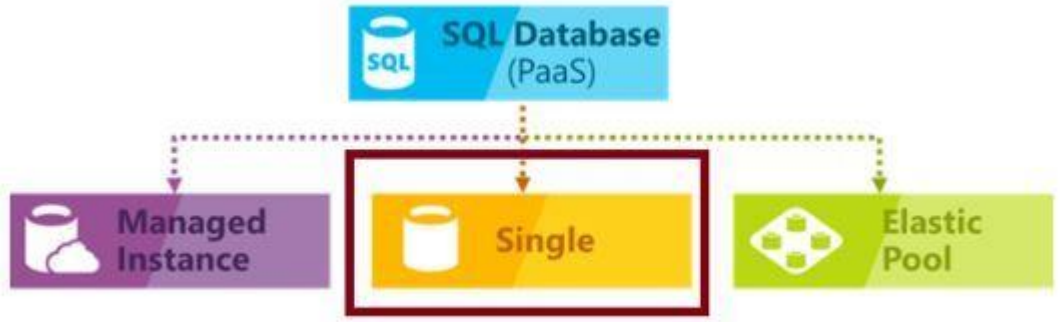
Store

walkerscott

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)

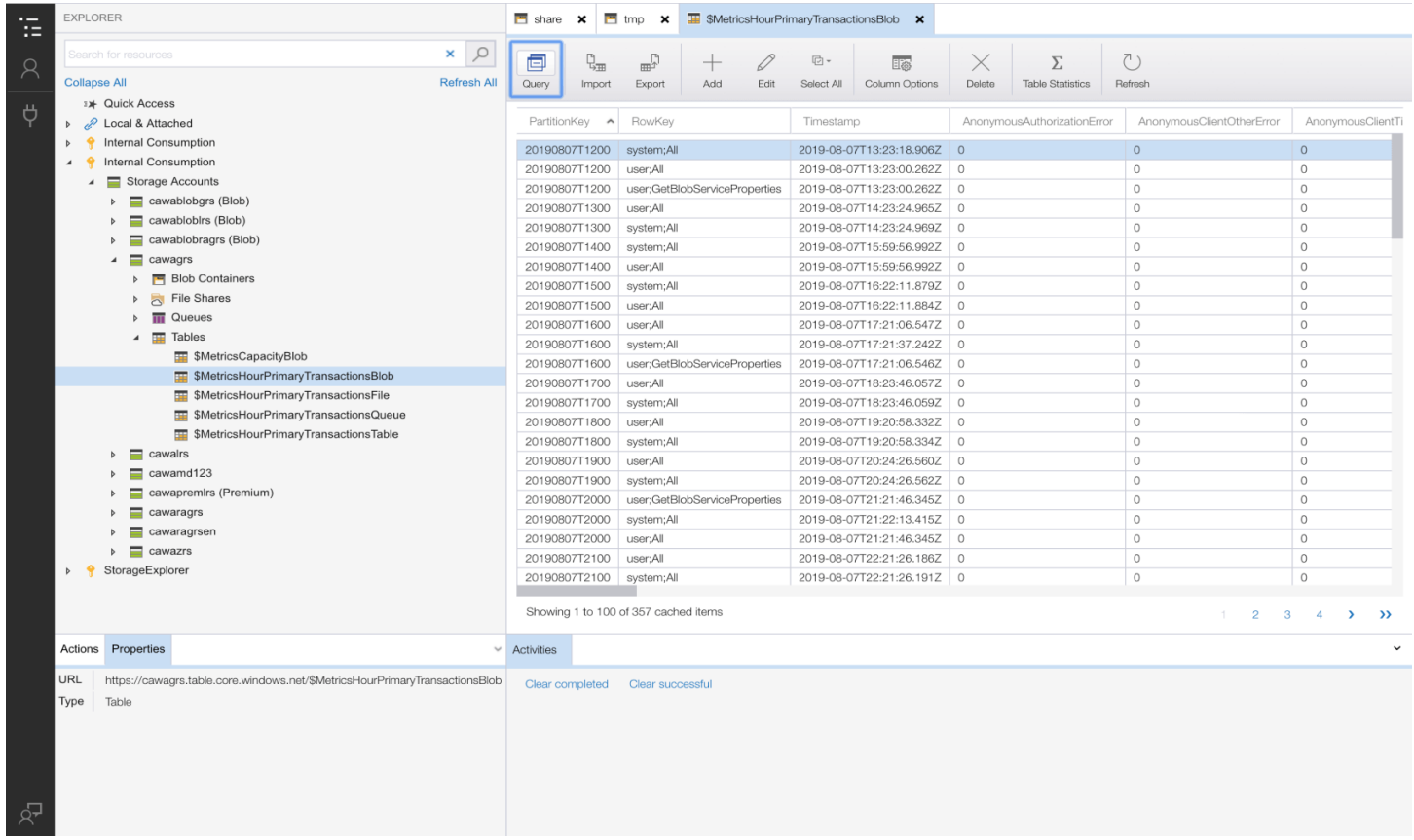
Azure SQL Databases



- 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)

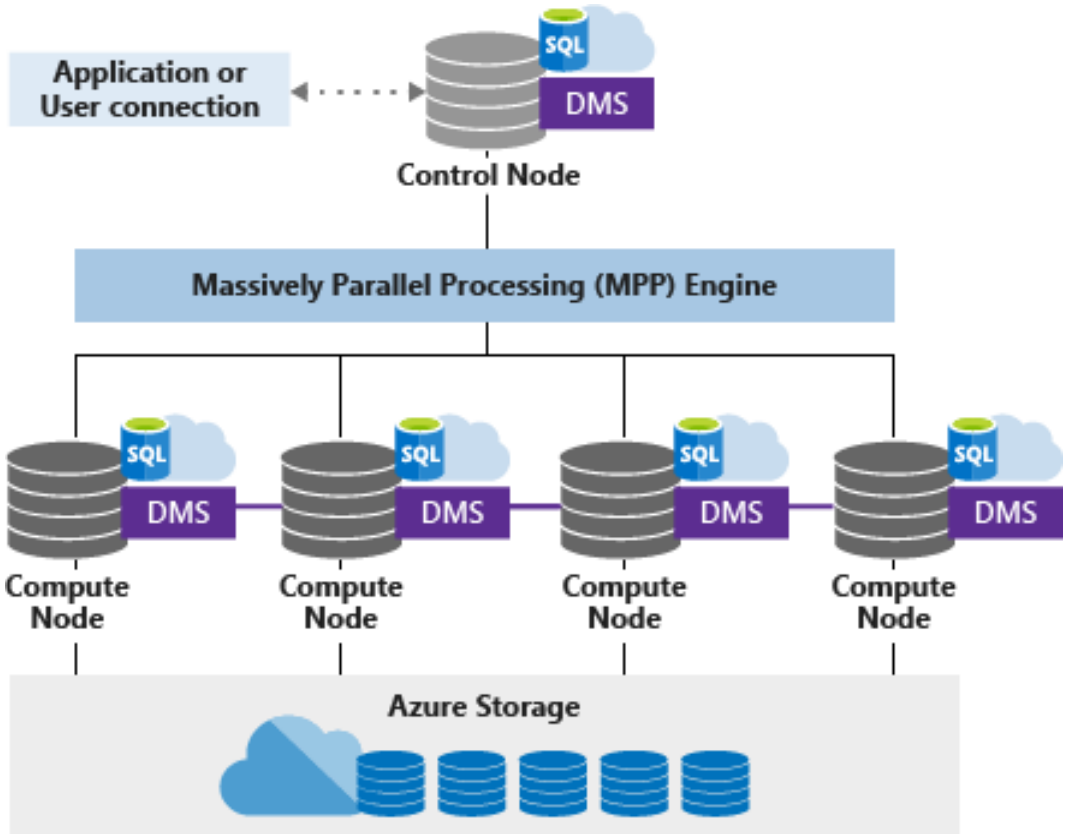
Azure Data Lake Storage Gen 2

- 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



walkerscott

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)

walkerscott

Train & Prep

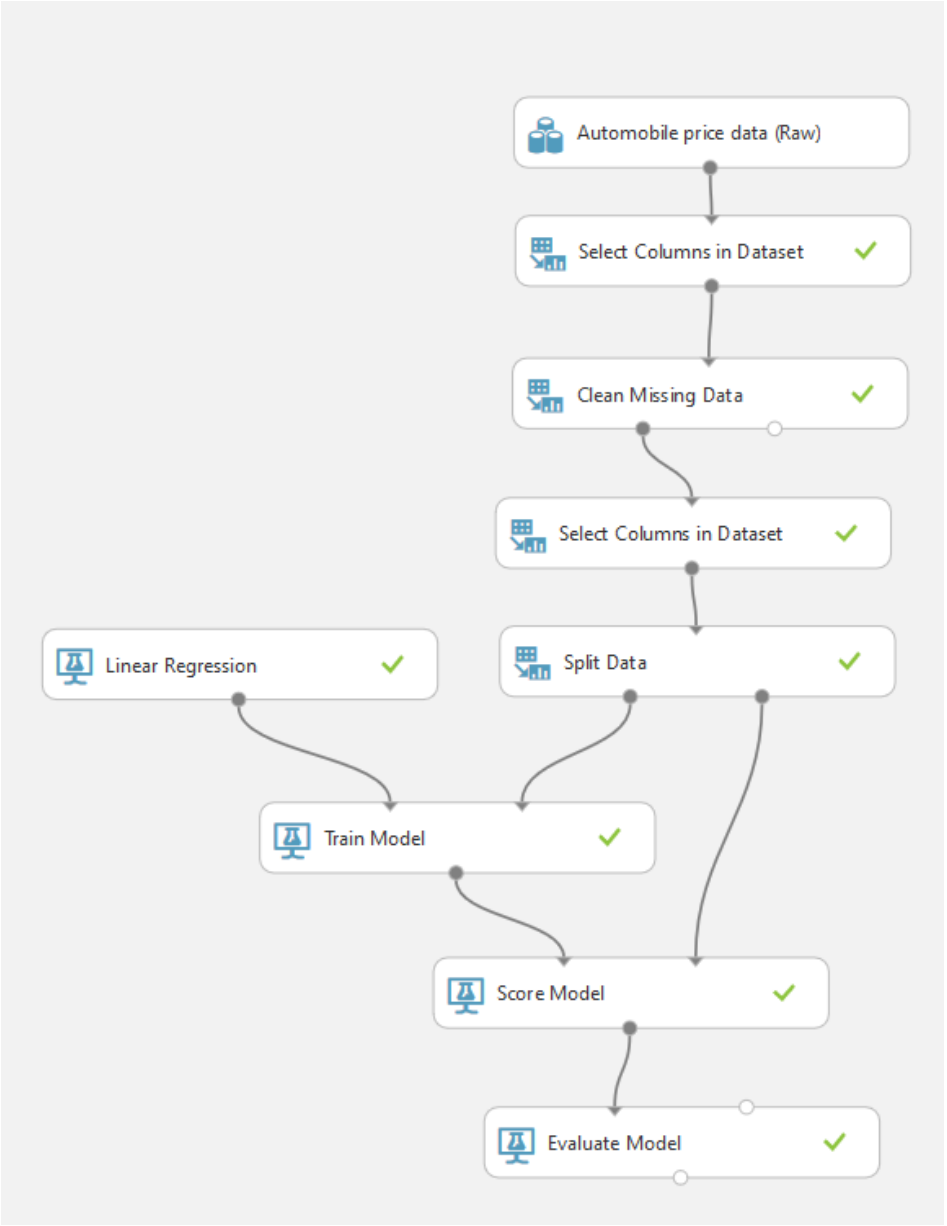
walkerscott

PaaS Services

- Azure Databricks
  - Spark-based analytics platform
  - 1st 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



walkerscott

Model & Serve

walkerscott

Power BI Dataset

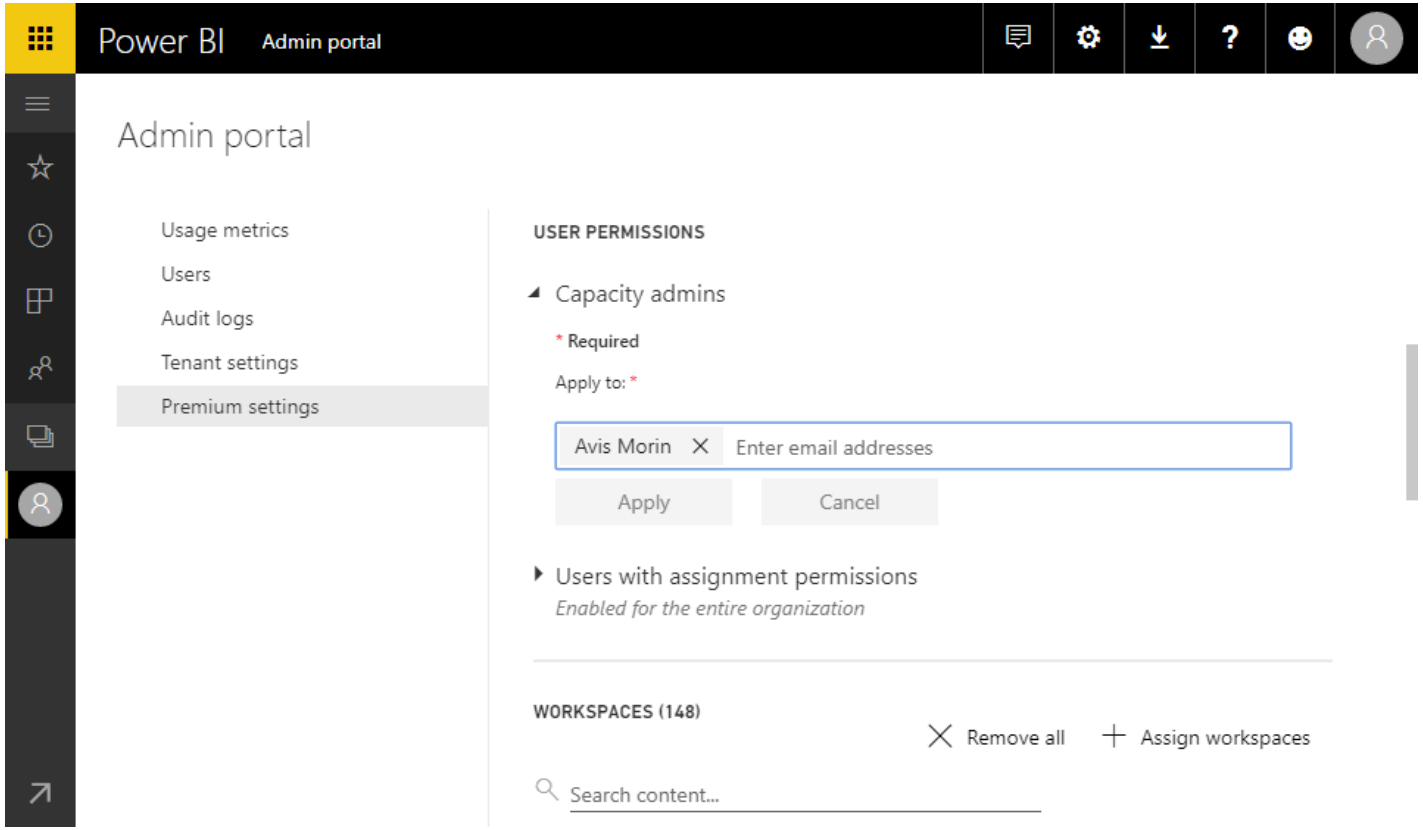
- Power BI Dataset
  - Report published to the Service
  - Pro licensing required to access

- 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

walkerscott

Power BI Premium

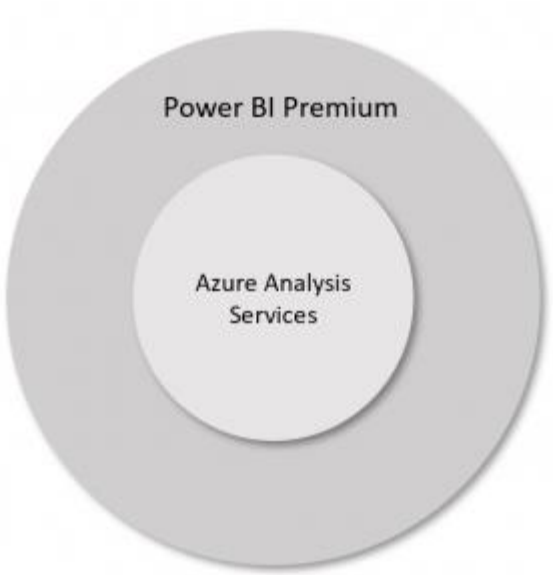
- Power BI Dataset
  - Report published to a Premium Workspace on the Service
  - Pro licensing not required to access
- When to use?
  - Import:
    - 10 GB data size limit
    - 12 GB refresh limit
  - Many considerations including licensing



walkerscott

Azure Analysis Services

- Azure PaaS service
- Feature-parity with SSAS (tabular models >= 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

walkerscott

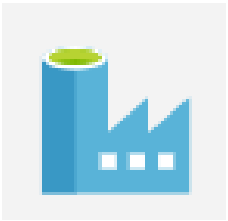
Lift-and-shift, and add PaaS services when required

From

SQL, SSIS,
MDS, SSAS

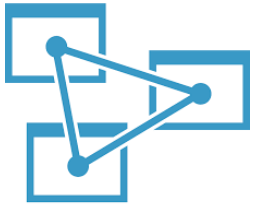
Power BI
Reports

To



Data
Factory

SQL, SSIS,
MDS

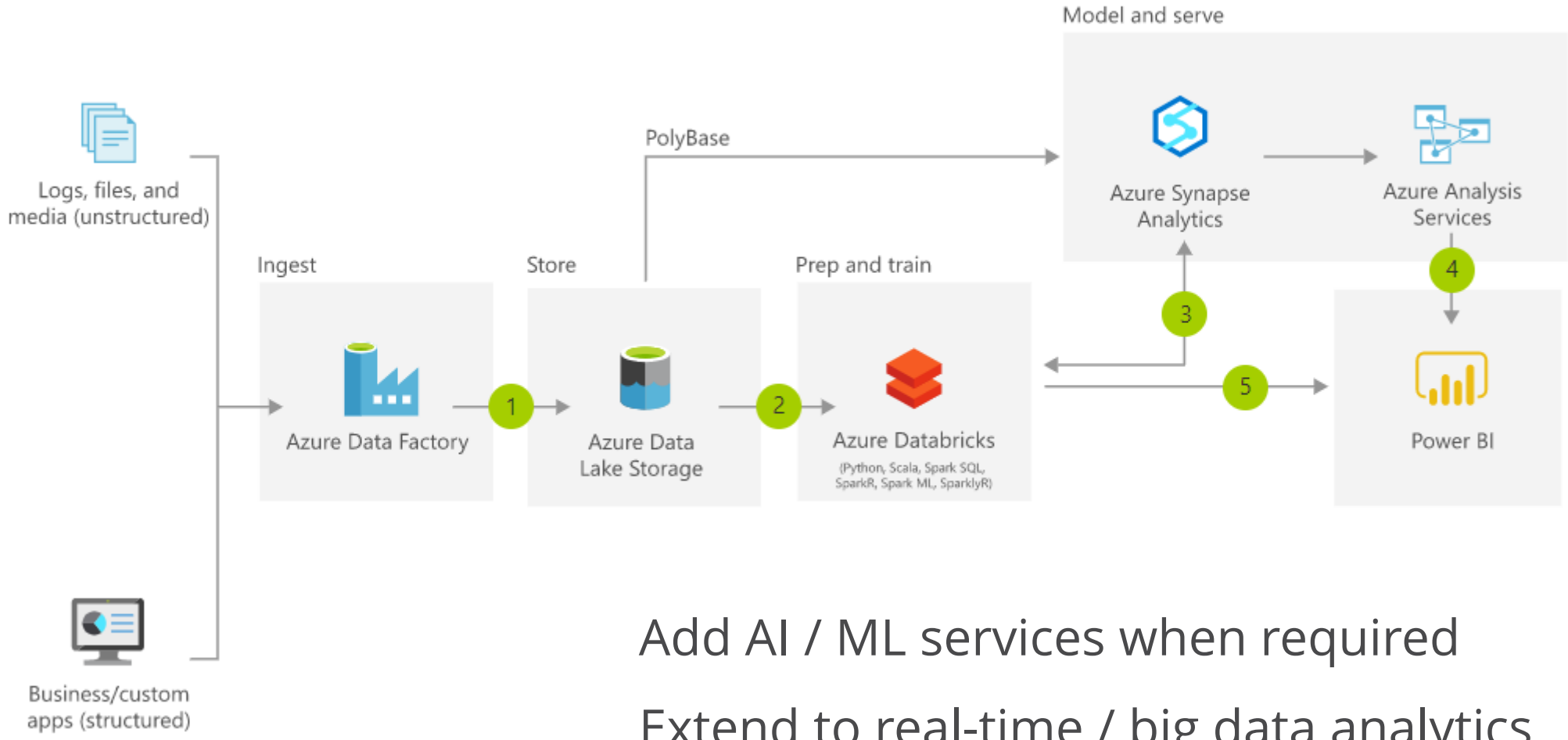


AAS

Power BI
Reports

walkerscott

Refactor / rearchitect modules to a modern DWH



Add AI / ML services when required

Extend to real-time / big data analytics

walkerscott

Standard Frameworks

walkerscott

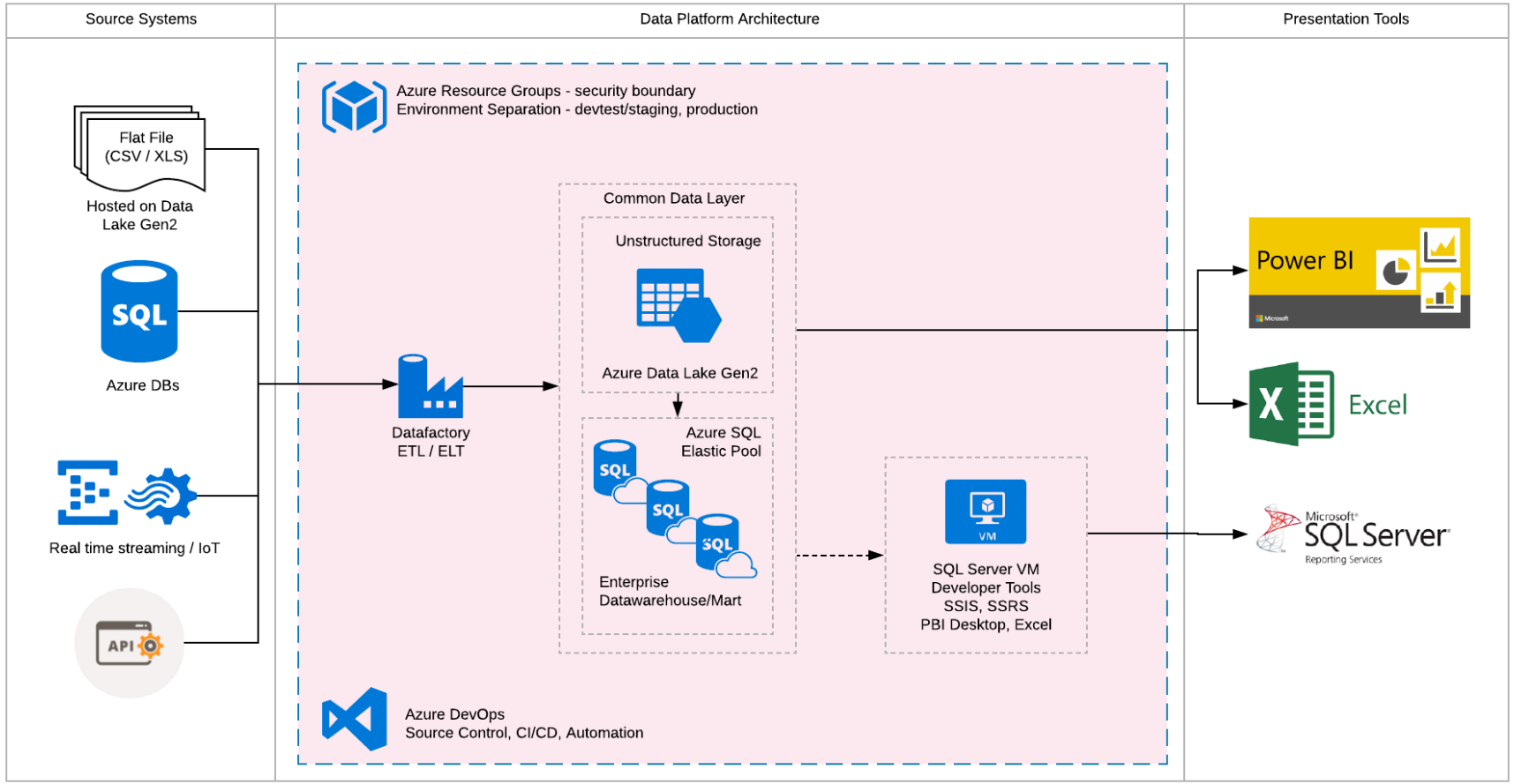
Azure Data Platform: Standard (without VPN & Azure Analysis Services)

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

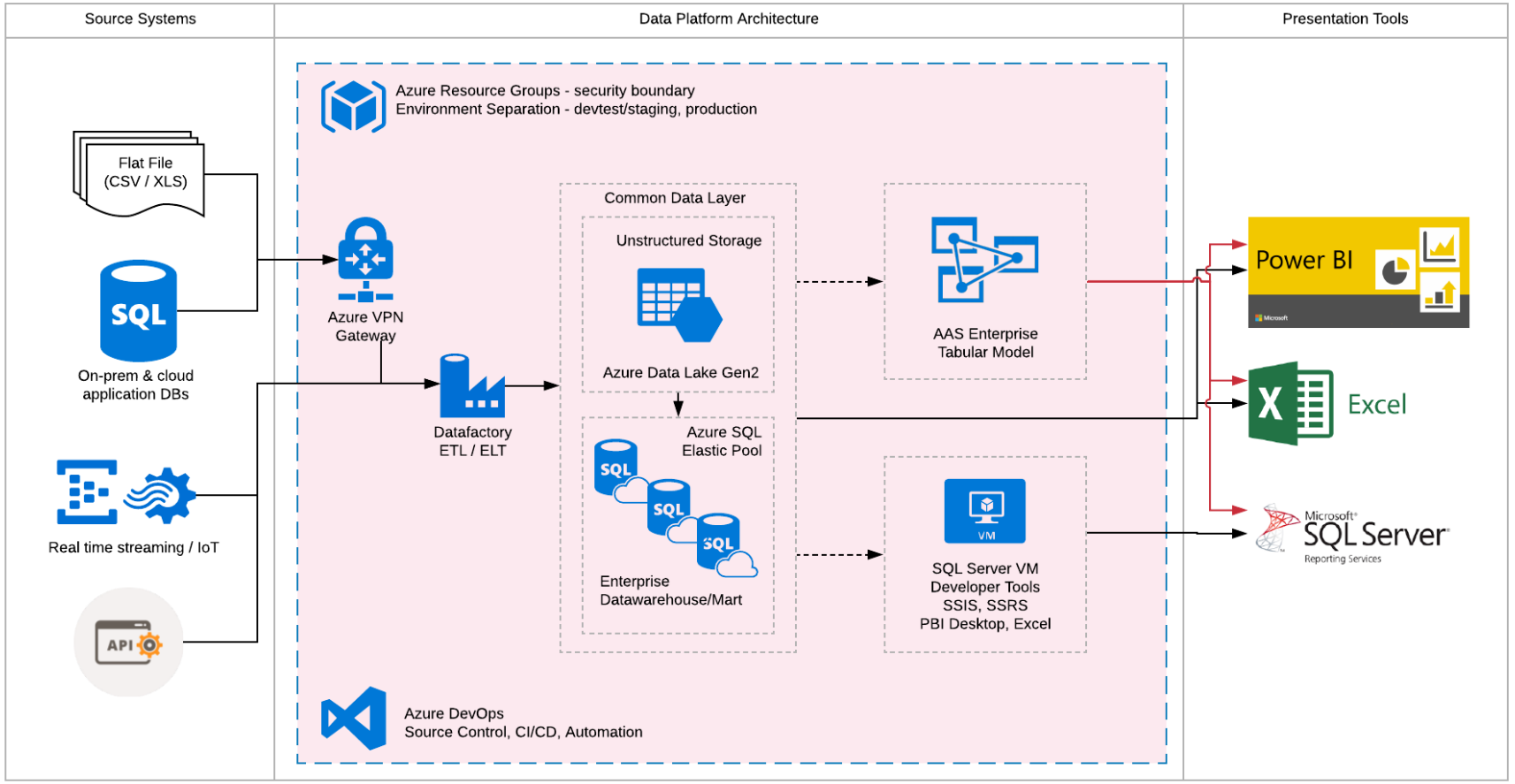
https://azure.com/e/e482b973752d48a2bf61b9233ba16d61

walkerscott

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

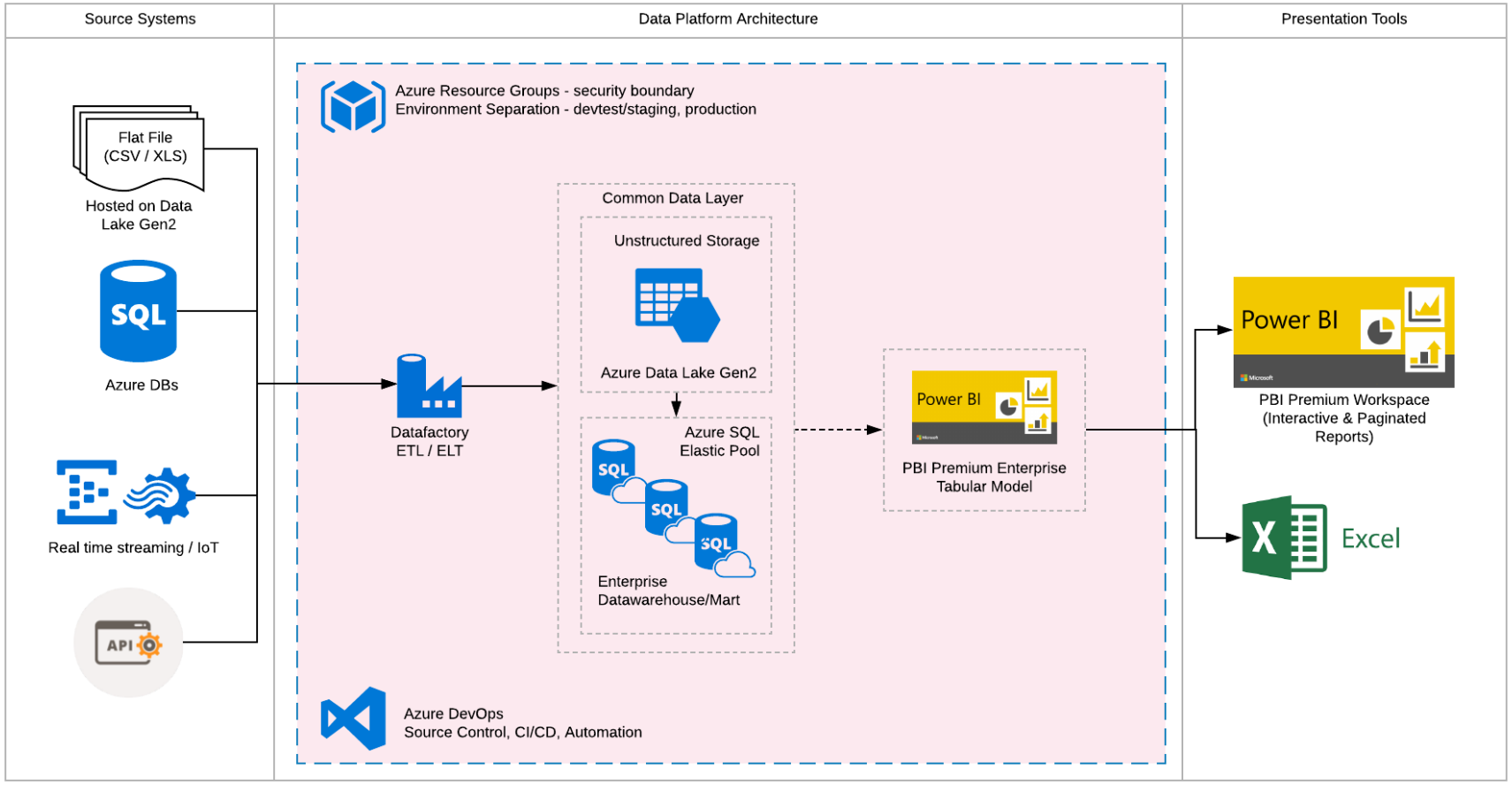
https://azure.com/e/5ac5b4d66ae94080a93a7db772145d7a

walkerscott

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

https://azure.com/e/1dd99925283e462ea3b4ba7bbd428e97

walkerscott

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

Thank you, questions?

Contact Us

NZ

Farid Jalal

farid.jalal@walkerscott.co

027 494 0128

www.walkerscott.co