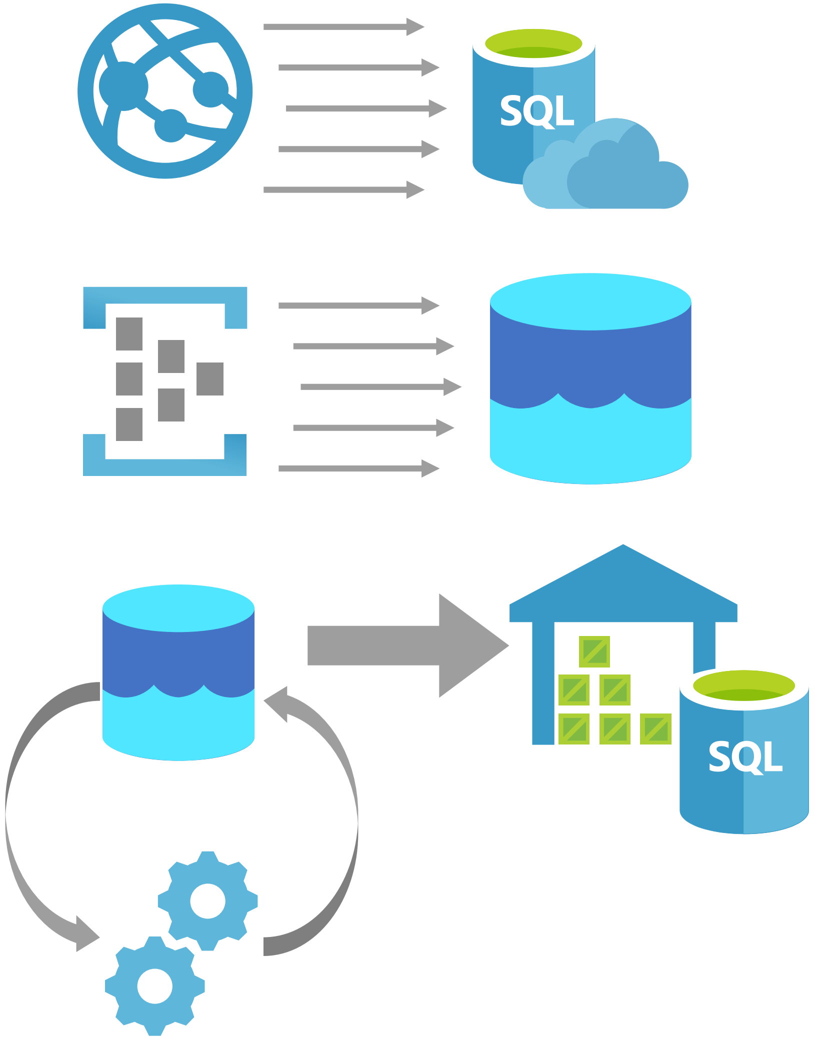
What workloads are NOT suitable?

Operational workloads (OLTP)

- High frequency reads and writes.
- Large numbers of singleton selects.
- High volumes of single row inserts.

Data Preparations

- Row by row processing needs.
- Incompatible formats (XML).



What Workloads are Suitable?

Analytics

Store large volumes of data.

Consolidate disparate data into a single location.

Shape, model, transform and aggregate data.

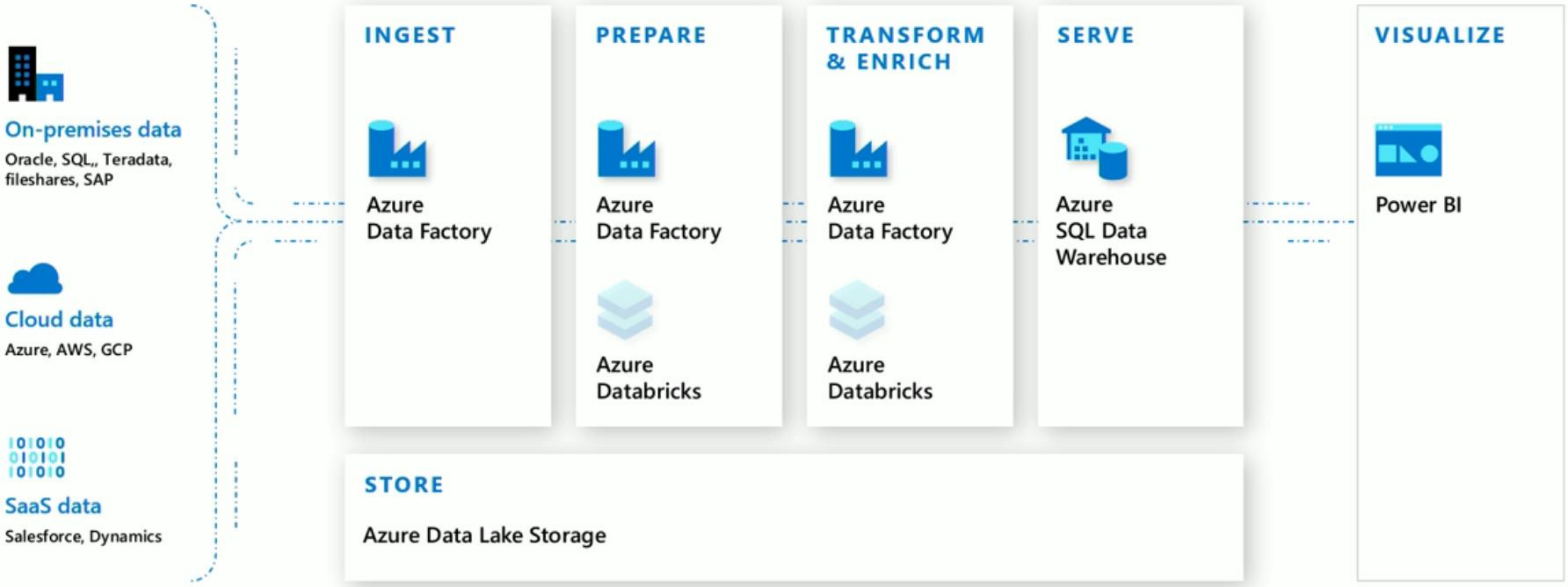
Batch/Micro-batch loads.

Perform query analysis across large datasets.

Ad-hoc reporting across large data volumes.

All using simple SQL constructs.

Modern Data Warehouse



On-premises data

Oracle, SQL,, Teradata, fileshares, SAP

Cloud data

Azure, AWS, GCP

SaaS data

Salesforce, Dynamics

INGEST

Azure Data Factory

PREPARE

Azure Data Factory

Azure Databricks

TRANSFORM & ENRICH

Azure Data Factory

Azure Databricks

SERVE

Azure SQL Data Warehouse

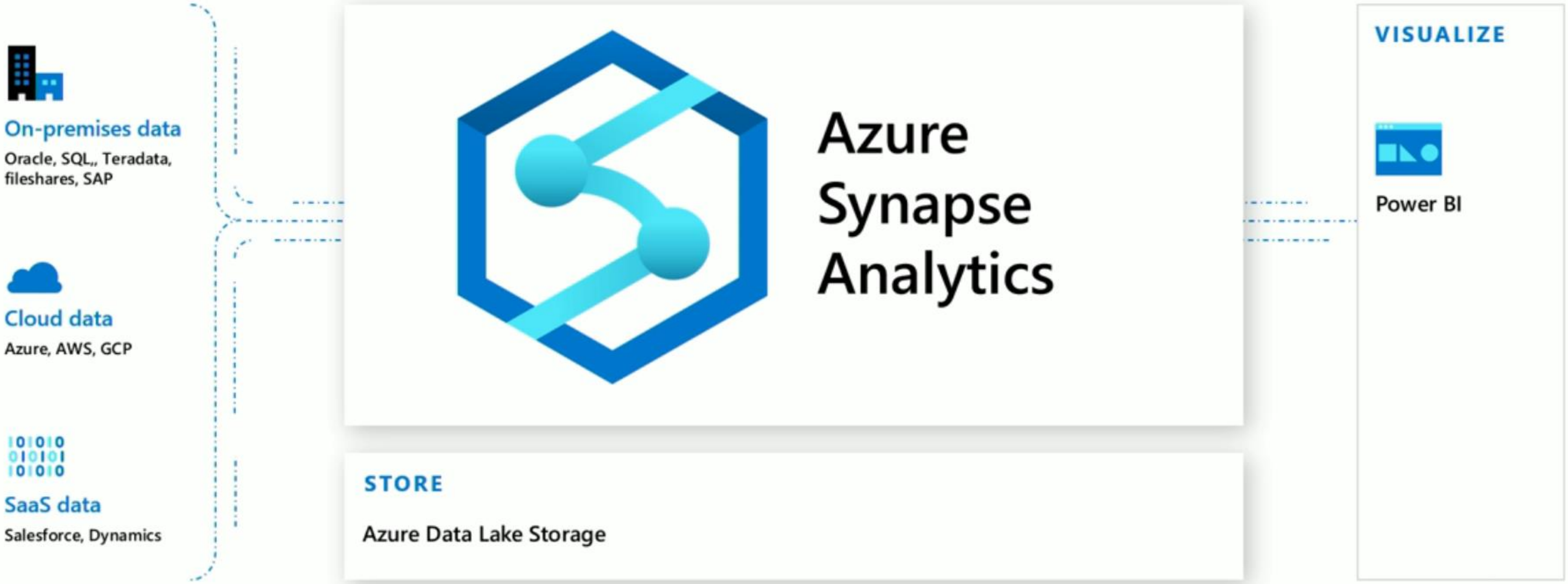
STORE

Azure Data Lake Storage

VISUALIZE

Power BI

Azure Synapse Analytics - *Data Lakehouse*



On-premises data

Oracle, SQL,, Teradata, fileshares, SAP

Cloud data

Azure, AWS, GCP

SaaS data

Salesforce, Dynamics

Azure Synapse Analytics

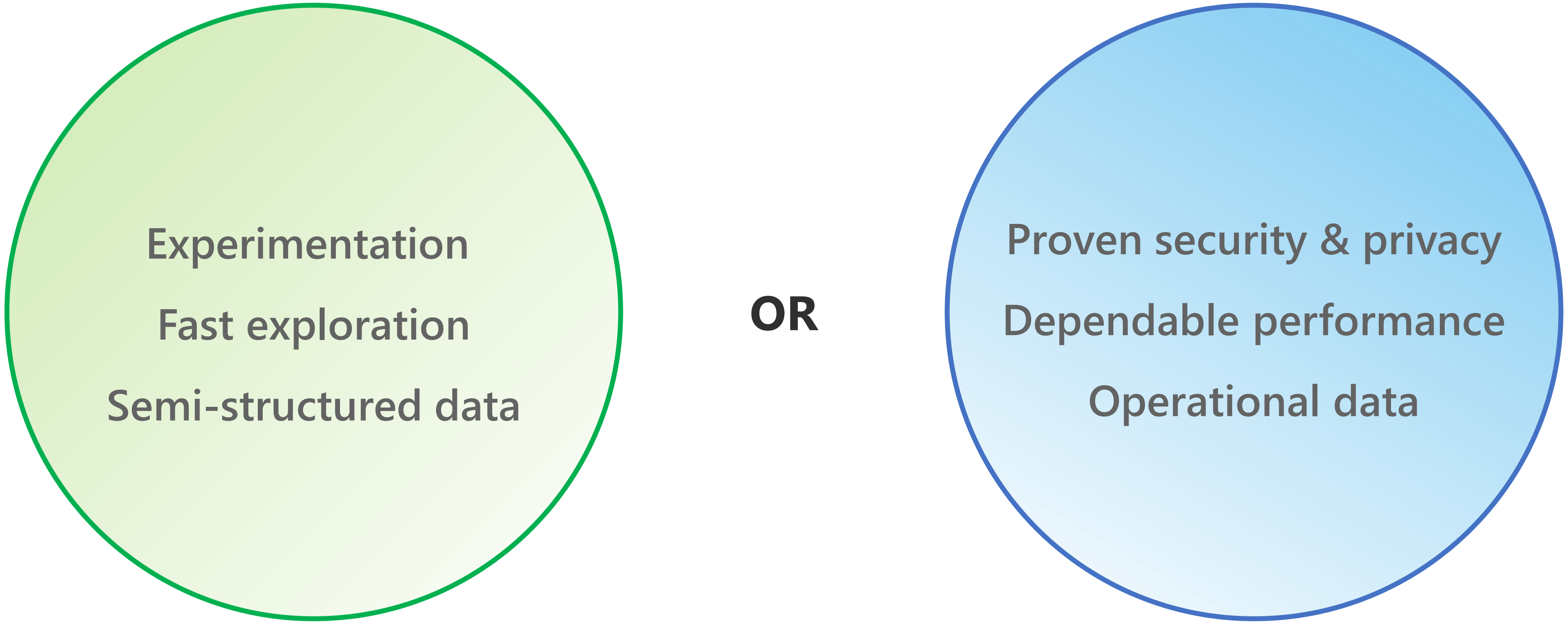
STORE

Azure Data Lake Storage

VISUALIZE

Power BI

Multiple analytics platforms



Big Data

Experimentation

Fast exploration

Semi-structured data

Data
Lake

OR

Relational Data

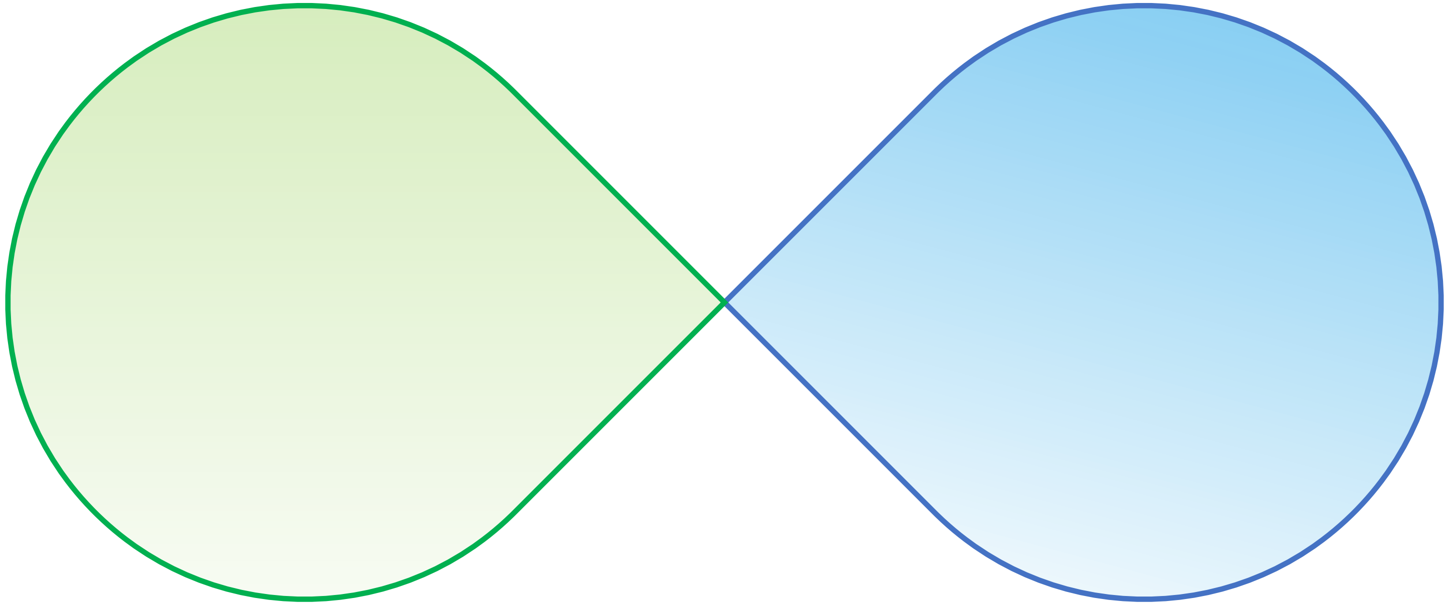
Proven security & privacy

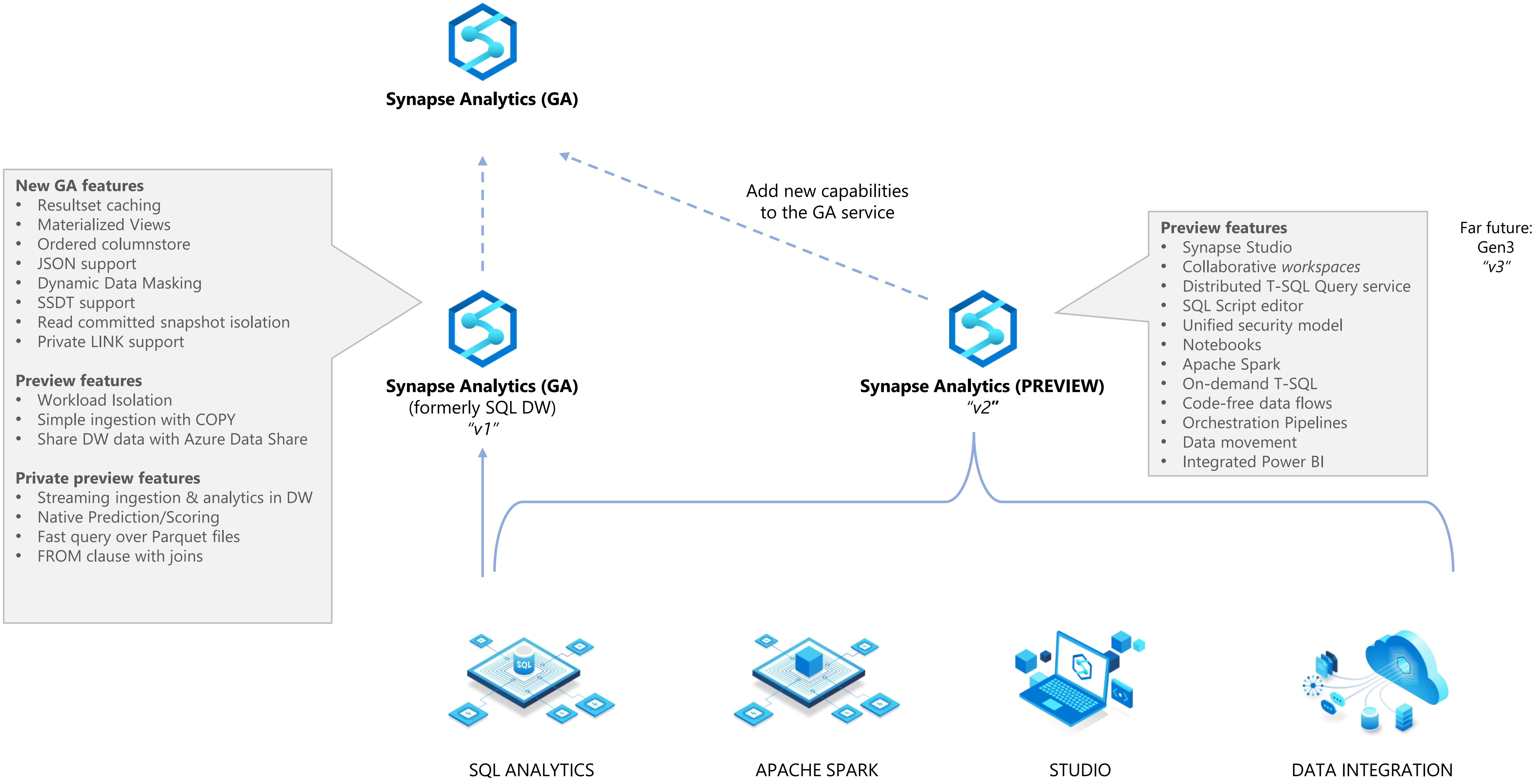
Dependable performance

Operational data

Data
Warehouse

Azure Azure Synapse Analytics
these two worlds together





Synapse Analytics (GA)

Add new capabilities
to the GA service

New GA features

- Resultset caching
- Materialized Views
- Ordered columnstore
- JSON support
- Dynamic Data Masking
- SSDT support
- Read committed snapshot isolation
- Private LINK support

Preview features

- Workload Isolation
- Simple ingestion with COPY
- Share DW data with Azure Data Share

Private preview features

- Streaming ingestion & analytics in DW
- Native Prediction/Scoring
- Fast query over Parquet files
- FROM clause with joins

Synapse Analytics (GA)
(formerly SQL DW)
"v1"

Synapse Analytics (PREVIEW)
"v2"

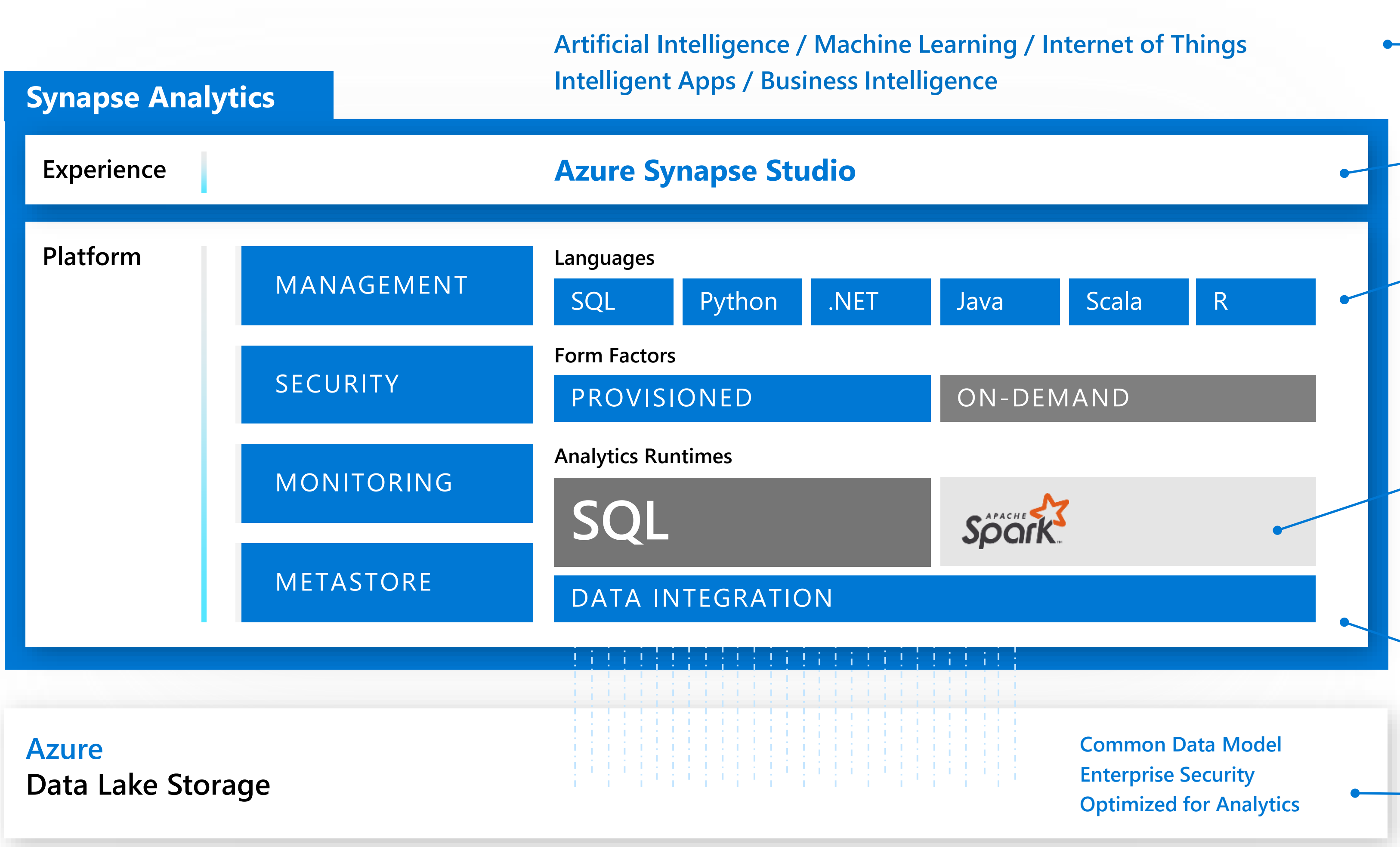
Preview features

- Synapse Studio
- Collaborative *workspaces*
- Distributed T-SQL Query service
- SQL Script editor
- Unified security model
- Notebooks
- Apache Spark
- On-demand T-SQL
- Code-free data flows
- Orchestration Pipelines
- Data movement
- Integrated Power BI

Far future:
Gen3
"v3"

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Designed for analytics **workloads at any scale**

SaaS **developer experiences** for code free and code first

Multiple **languages** suited to different analytics workloads

Integrated analytics runtimes available provisioned and serverless on-demand

SQL Analytics offering T-SQL for batch, streaming and interactive processing

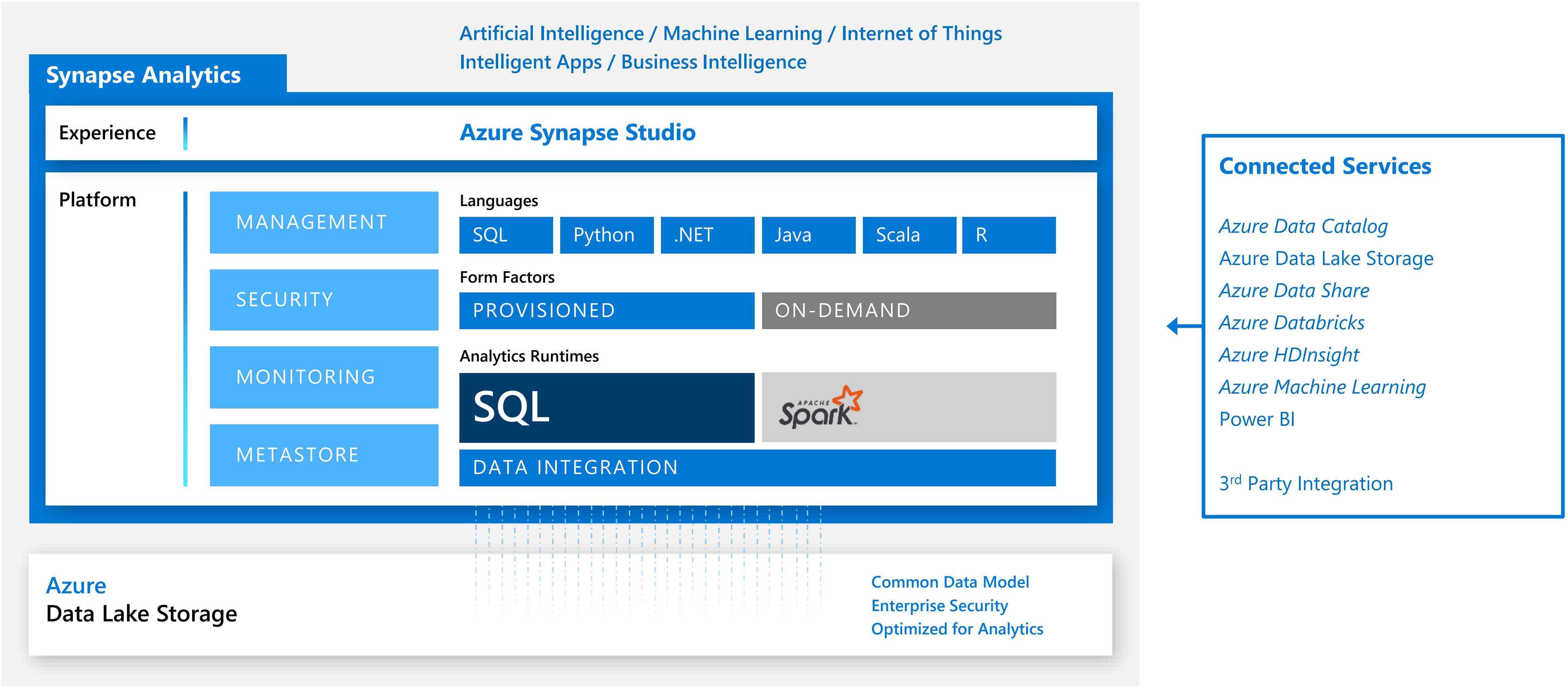
Spark for big data processing with Python, Scala, R and .NET

Integrated **platform services** for, management, security, monitoring, and metastore

Data **lake integrated** and Common Data Model aware

Azure Synapse Analytics

Integrated data platform for BI, AI and continuous intelligence



Artificial Intelligence / Machine Learning / Internet of Things
Intelligent Apps / Business Intelligence

Synapse Analytics

Experience | **Azure Synapse Studio**

Platform

MANAGEMENT

SECURITY

MONITORING

METASTORE

Languages

SQL | Python | .NET | Java | Scala | R

Form Factors

PROVISIONED | ON-DEMAND

Analytics Runtimes

SQL | APACHE Spark

DATA INTEGRATION

Azure
Data Lake Storage

Common Data Model
Enterprise Security
Optimized for Analytics

Connected Services

Azure Data Catalog
Azure Data Lake Storage
Azure Data Share
Azure Databricks
Azure HDInsight
Azure Machine Learning
Power BI

3rd Party Integration

New Products/Features

- Azure Synapse Analytics – Umbrella name. For now just includes SQL DW. In preview adds a Synapse Workspace which includes SQL DW and all the new product/features below
- SQL pool or SQL analytics pool – Really just SQL Data Warehouse (SQL DW) which includes compute and storage
- Azure Synapse Studio – New product. Single pain of glass that is a web-based experience. Collaborative workspaces. Access SQL Databases, Spark tables, SQL Scripts, notebooks (supports multiple languages), Data flows (Data Integration), pipelines (Data Integration), monitoring, security. Has links to ADLS Gen2 and Power BI workspace
- Data Integration – Really just Azure Data Factory (ADF). They use the same code base. Note in Synapse Studio Data Flows are under "Develop", Pipelines are under "Orchestrate", and Datasets are under "Data" (In ADF they are all under "Author")
- Spark – Including Apache Spark is new. Similar to Spark in SQL Server 2019 BDC
- On-demand T-SQL – New feature. Was code-named Starlight Query
- T-SQL over ADLS Gen2 – New feature. Was code-named Starlight Query
- New SQL DW features (see next slides) – Some are GA now and some are in preview
- Multiple query options (see next slides) – Some are GA now and some are in preview
- Distributed Query Processor (see next slides) – some in preview or Gen3

New Synapse Features

GA features:

- Performance: Result-set caching
- Performance: Materialized Views
- Performance: Ordered clustered columnstore index
- Heterogeneous data: JSON support
- Trustworthy computation: Dynamic Data Masking
- Continuous integration & deployment: SSDT support
- Language: Read committed snapshot isolation

Public Preview features:

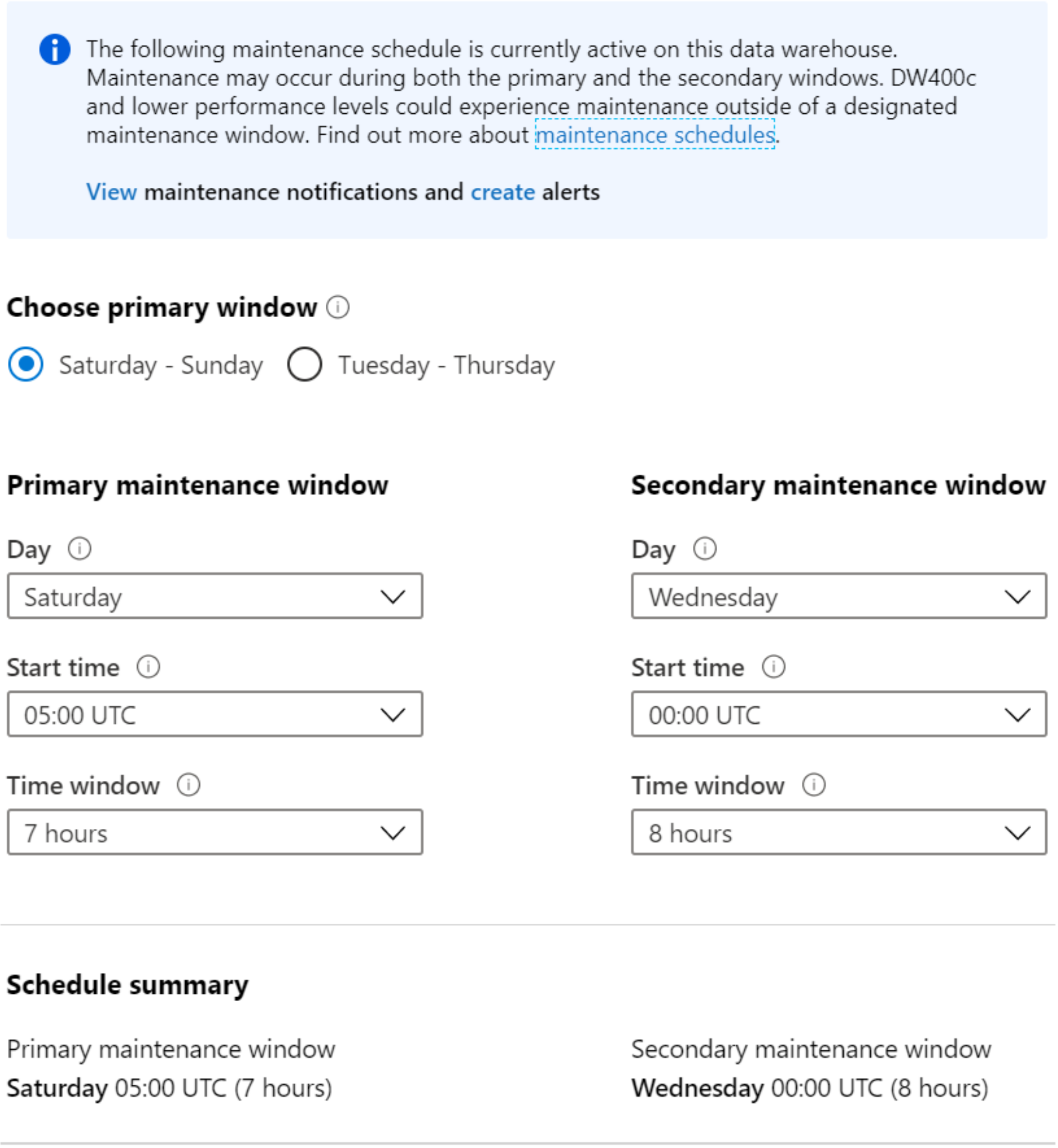
- Workload management: Workload Isolation
- Data ingestion: Simple ingestion with COPY
- Data Sharing: Share DW data with Azure Data Share
- Trustworthy computation: Private LINK support

Private Preview features:

- Data ingestion: Streaming ingestion & analytics in DW
- Built-in ML: Native Prediction/Scoring
- Data lake enabled: Fast query over Parquet files
- Language: Updateable distribution column
- Language: FROM clause with joins
- Language: Multi-column distribution support
- Security: Column-level Encryption

Note: private preview features require whitelisting.

Maintenance Schedule (preview)



The following maintenance schedule is currently active on this data warehouse. Maintenance may occur during both the primary and the secondary windows. DW400c and lower performance levels could experience maintenance outside of a designated maintenance window. Find out more about maintenance schedules.

View maintenance notifications and create alerts

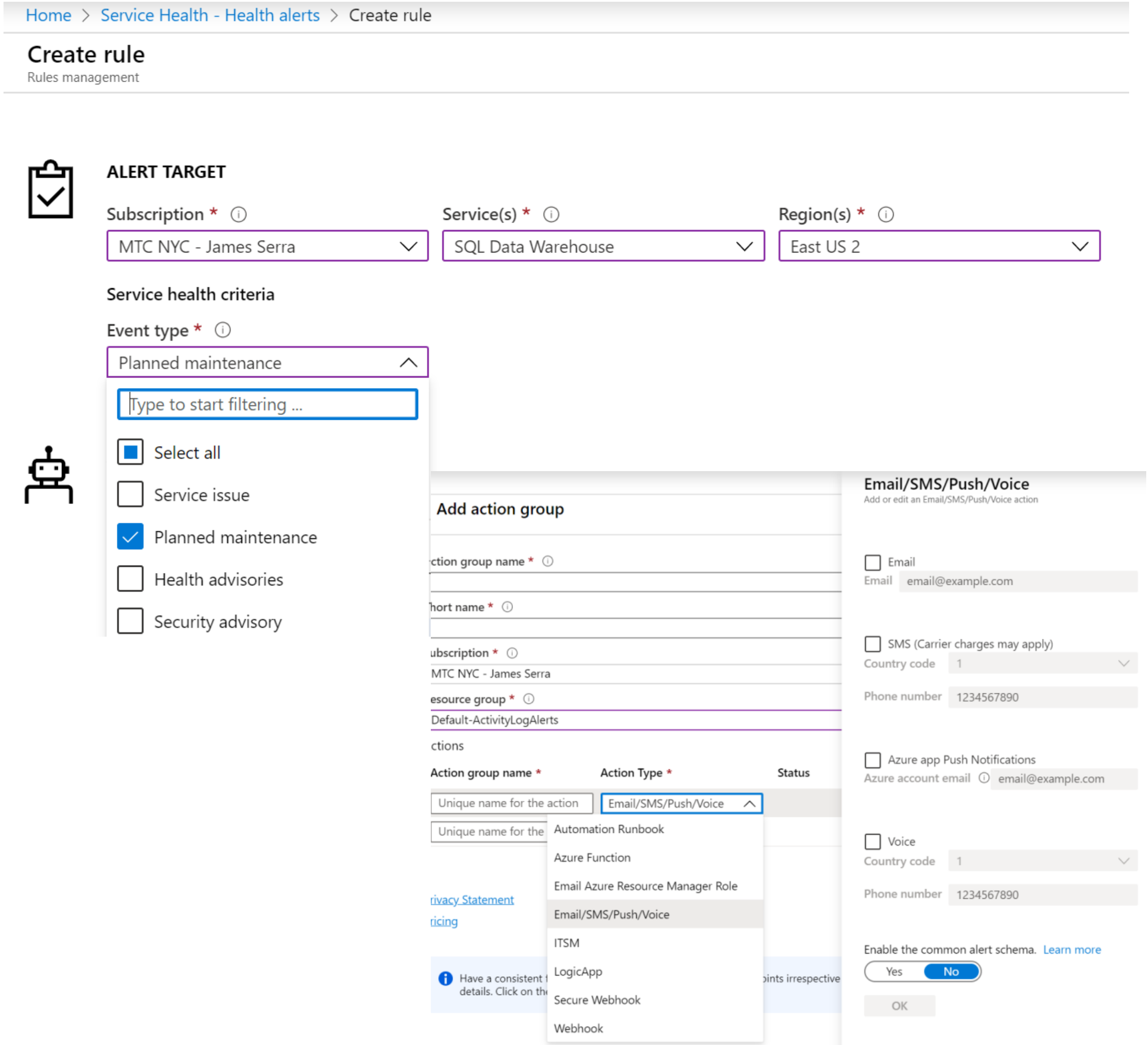
Choose primary window

● Saturday - Sunday ○ Tuesday - Thursday

Primary maintenance window	**Secondary maintenance window**
Day: Saturday	Day: Wednesday
Start time: 05:00 UTC	Start time: 00:00 UTC
Time window: 7 hours	Time window: 8 hours

Schedule summary

Primary maintenance window	Secondary maintenance window
Saturday 05:00 UTC (7 hours)	**Wednesday** 00:00 UTC (8 hours)



Home > Service Health - Health alerts > Create rule

Create rule

Rules management

ALERT TARGET

Subscription *: MTC NYC - James Serra

Service(s) *: SQL Data Warehouse

Region(s) *: East US 2

Service health criteria

Event type *: Planned maintenance

Type to start filtering ...

■ Select all

☐ Service issue

☑ Planned maintenance

☐ Health advisories

☐ Security advisory

Add action group

ction group name *

hort name *

ubscription *: MTC NYC - James Serra

esource group *: Default-ActivityLogAlerts

ctions

Action group name *	Action Type *	Status
Unique name for the action	Email/SMS/Push/Voice	
Unique name for the		

Have a consistent … details. Click on th… …oints irrespective

Email/SMS/Push/Voice

Add or edit an Email/SMS/Push/Voice action

☐ Email

Email: email@example.com

☐ SMS (Carrier charges may apply)

Country code: 1

Phone number: 1234567890

☐ Azure app Push Notifications

Azure account email: email@example.com

☐ Voice

Country code: 1

Phone number: 1234567890

Enable the common alert schema. Learn more

Yes [✓ No]

OK

Query Options

1. Provisioned SQL over relational database – Traditional SQL DW [existing]
2. Provisioned SQL over ADLS Gen2 – via external tables or openrowset [existing via external tables in PolyBase, openrowset not yet in preview]
3. *On-demand SQL over relational database - dependency on the flexible data model (data cells) over columnstore data [new, not yet in preview: the ability to query a SQL relational database (and other types of data sources) will come later but not in H1 2020]*
4. On-demand SQL over ADLS Gen2 – via external tables or openrowset [new in preview]
5. Provisioned Spark over relational database – [new in preview]
6. Provisioned Spark over ADLS Gen2 [new in preview]
7. *On-demand Spark over relational database - On-demand Spark is not supported (but provisioned Spark can auto-pause)*
8. *On-demand Spark over ADLS Gen2 – On-demand Spark is not supported (but provisioned Spark can auto-pause)*

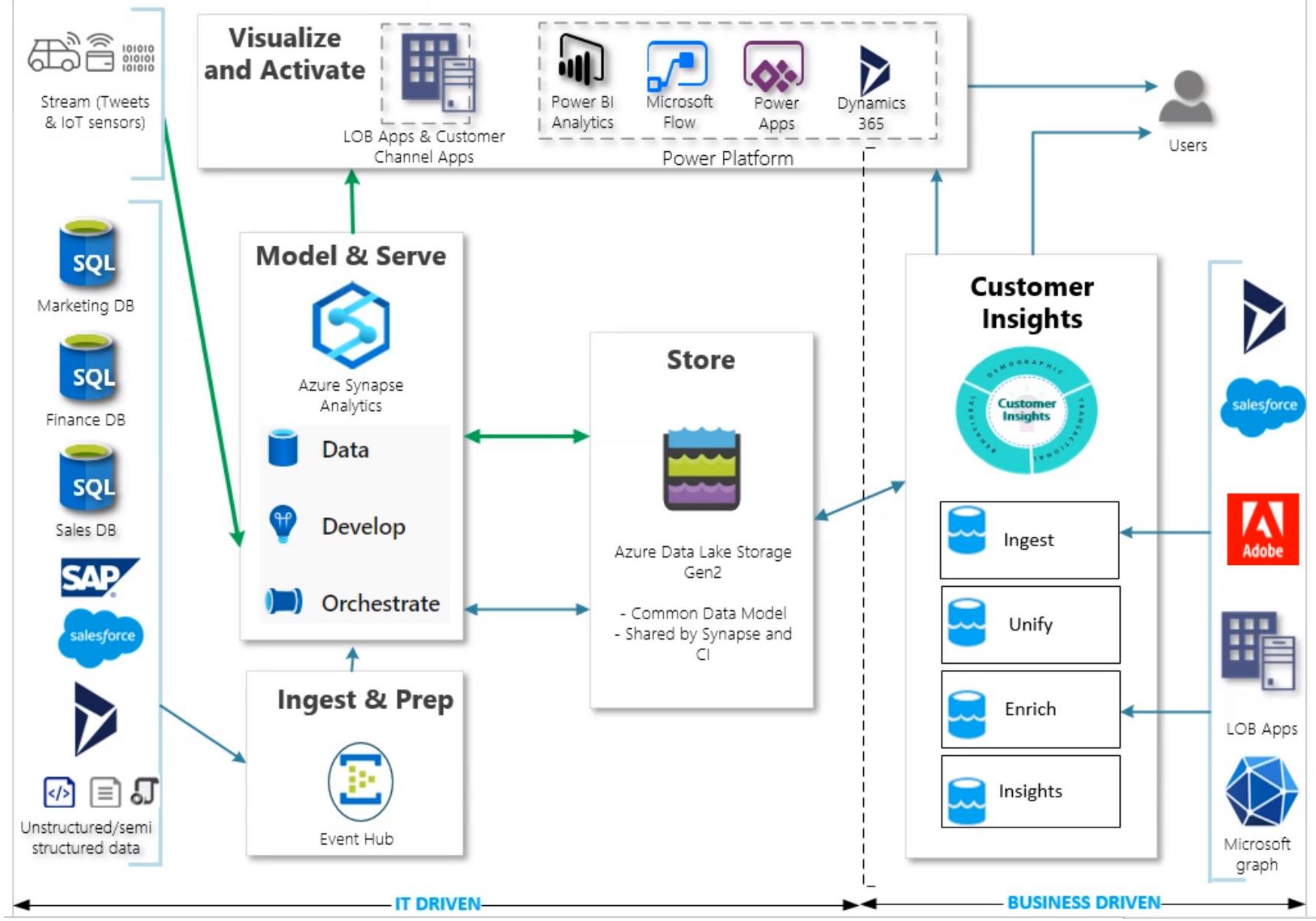
Notes:

- Separation of state (data, metadata and transactional logs) and compute
- Queries against data loaded into SQL Analytics tables are 2-3X faster compared to queries over external tables
- Copy statement: Improved performance compared to PolyBase. PolyBase is not used, but functional aspects are supported
- Warm-up for first on-demand SQL query takes about 30-40 seconds
- If you create a Spark Table, that table will be created as an external table in SQL Pool or SQL On-Demand without having to keep a Spark cluster up and running
- Currently one on-demand SQL pool but by GA will support many
- Provisioned SQL may give you better and more predictable performance due to resource reservation
- Existing PolyBase via external tables is not pushdown (#2), but #4 will be pushdown (SQL on-demand will push down queries from the front-end to back-end nodes)
- Supported file formats are parquet, csv, json
- ***Each SQL pool can currently only access tables created within its pool (there is one database per pool), while on-demand SQL can not yet query a database***

Distributed Query Processor (DQP) - Preview

- **Auto-scale compute nodes (on-demand SQL in preview, provisioned SQL in Gen3)** - Instruct the underlying fabric the need for more compute power to adjust to peaks during the workload. If compute power is granted, the DQP will re-distribute tasks leveraging the new compute container. Note that in-flight tasks in the previous topology continue running, while new queries get the new compute power with the new re-balancing
- **Compute node fault tolerance (on-demand SQL in preview, provisioned SQL in Gen3)** - Recover from faulty nodes while a query is running. If a node fails the DQP re-schedules the tasks in the faulted node through the remainder of the healthy topology
- **Compute node hot spot: rebalance queries or scale out nodes (on-demand SQL in preview, provisioned SQL in Gen3)** - Can detect hot spots in the existing topology. That is, overloaded compute nodes due to data skew. In the advent of a compute node running hot because of skewed tasks, the DQP can decide to re-schedule some of the tasks assigned to that compute node amongst others where the load is less
- **Multi-master cluster (provisioned SQL only in Gen3)** - User workloads can operate over the same shareable relational data set while having independent clusters to serve those various workloads. Allows for very high concurrency. So you could have multiple SQL pools all accessing the same database. Databases are not tied to a pool
- **Cross-database queries (provisioned SQL only in Gen3)** – A query can specify multiple databases. This is because the data of the databases are not in the pool. Rather, each pool just has the metadata of all the databases and the data of the databases are in a separate sharable storage layer
- **Query scheduler (provisioned SQL only in Gen3)** – New way of executing queries within the data warehouse using a scheduler and resource manager/estimator. When a query is submitted, estimates how many resources are needed to complete request and schedules it (and can use workload importance/isolation). Will completely remove the need for concurrency limits. This is how SQL Server works today

Common Data Platform Vision Demo Architecture



https://microsoft.sharepoint.com/sites/Infopedia/pages/docset-viewer.aspx?did=G01KC-2-4992

https://admin.cloudlabs.ai

Query Demo

	Relational Data	ADLS Gen2
Provisioned SQL	3	5 (external table)
On-demand SQL	X	1
Spark	4	2

Supported file formats are parquet, csv, json