

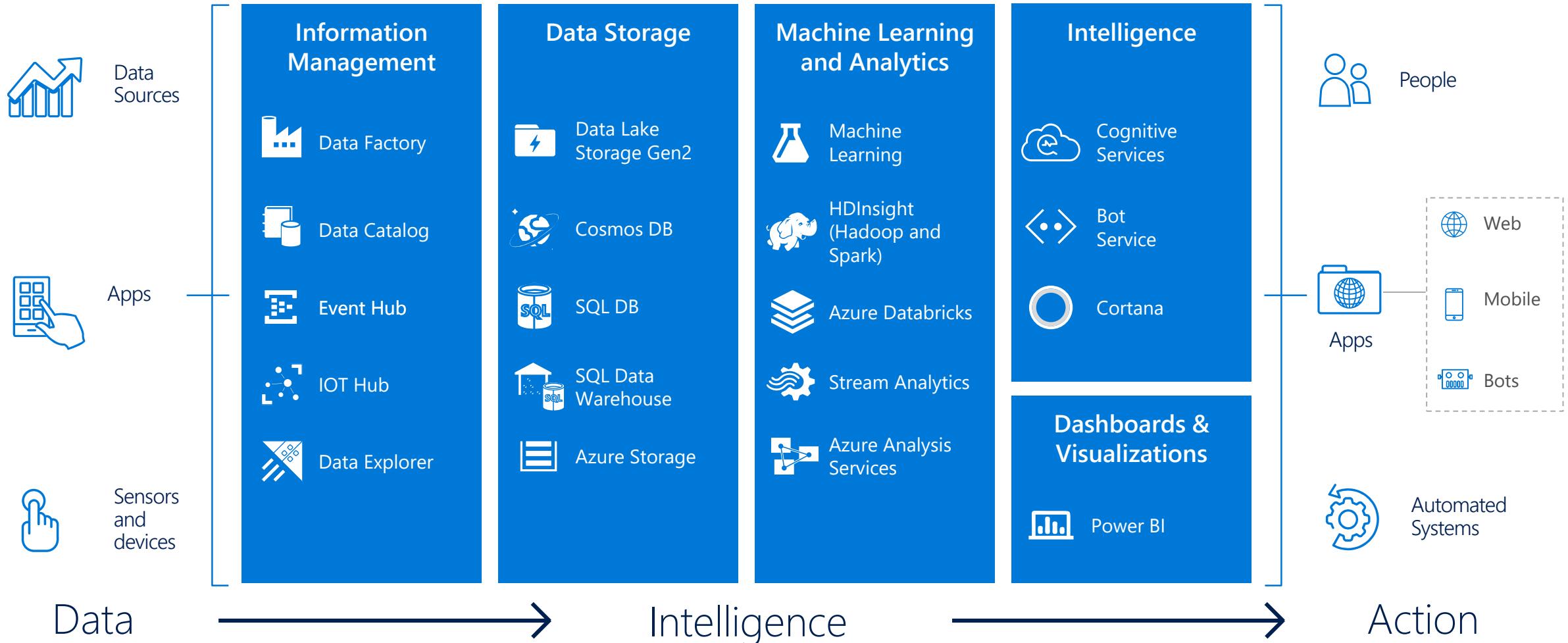


Data & AI Partner Kick-off

Sept 2019

The world is changing

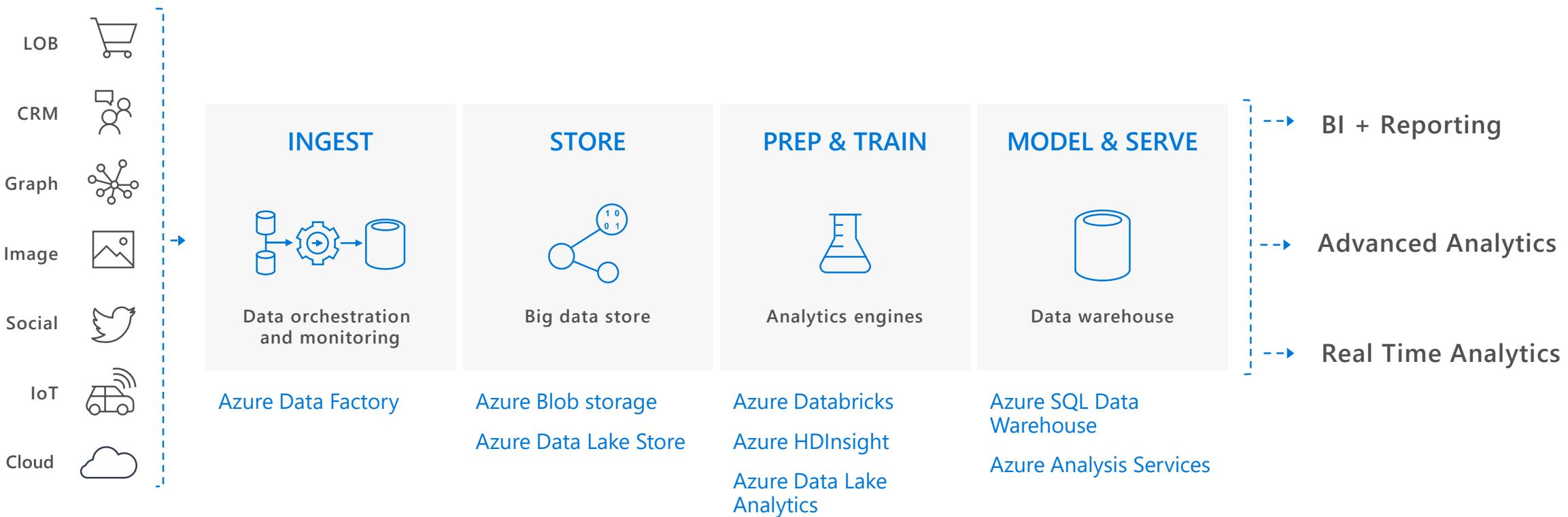
DATA MANAGEMENT SOLUTION FOR INTELLIGENCE IN THE CLOUD



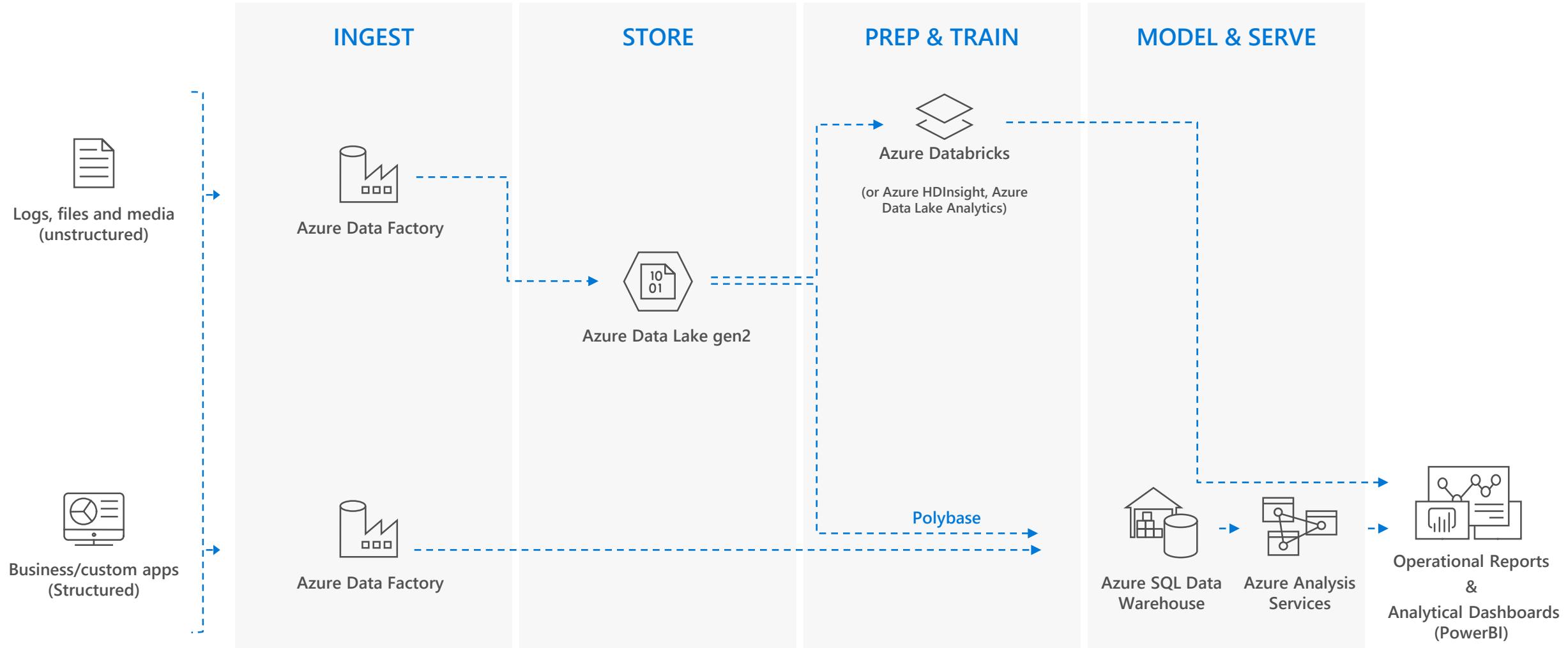
Objective

- Understanding Azure data platform product info
- Telling which product to apply in future use case
- Enabling further deep dive on solution design

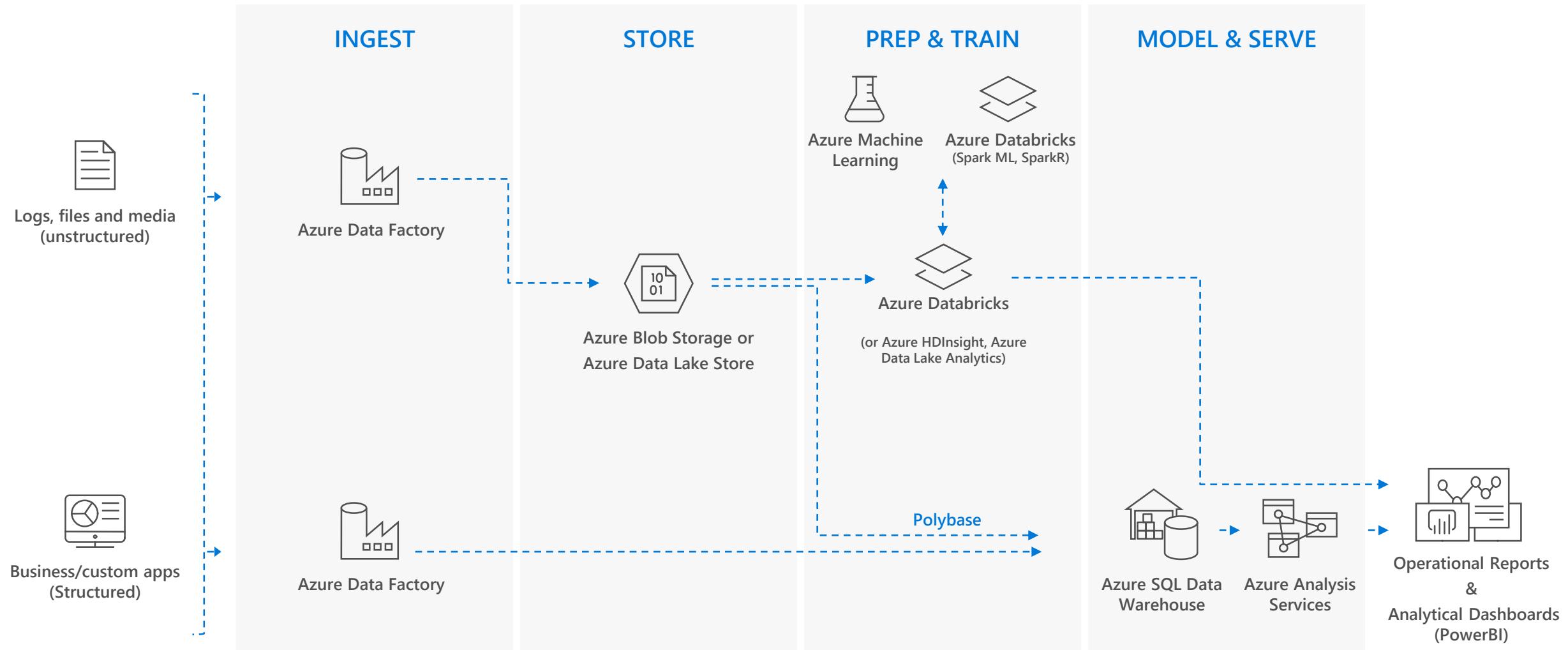
BIG DATA & DATA WAREHOUSE



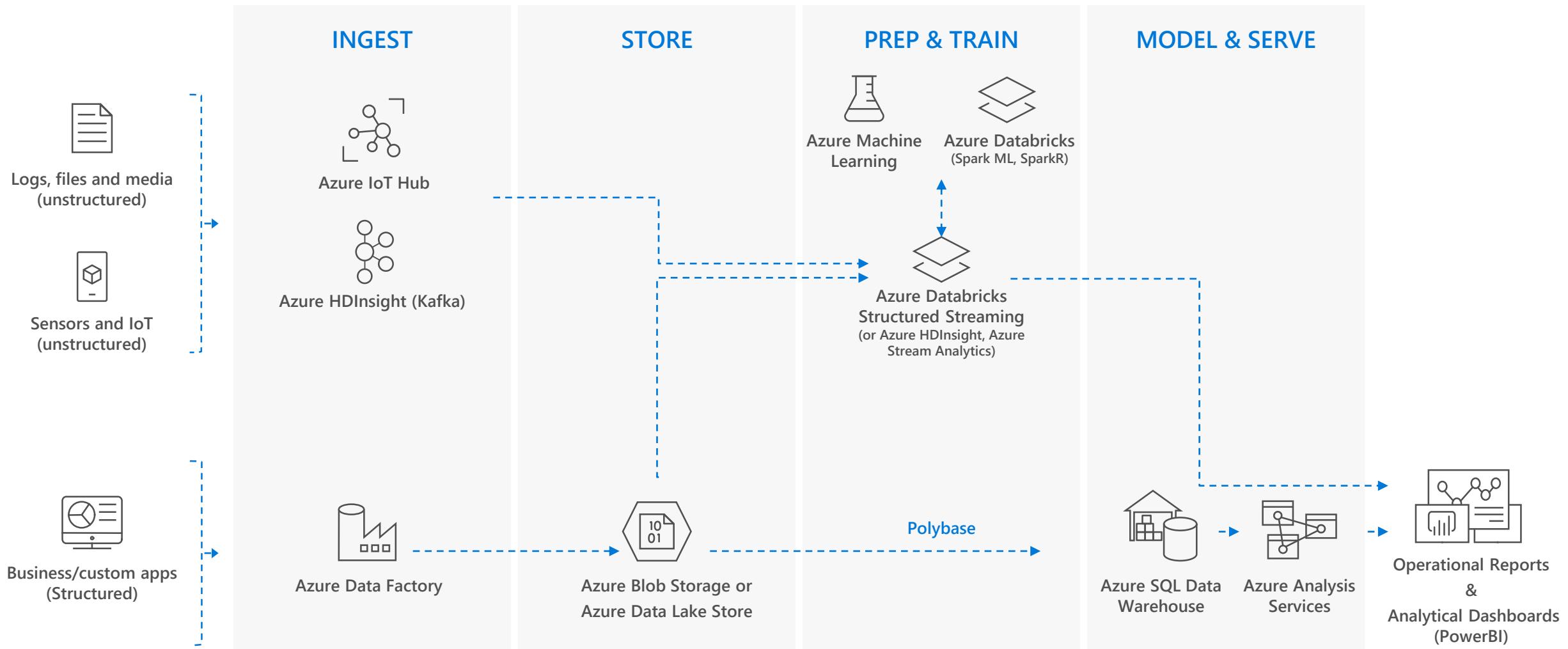
MODERN DATA WAREHOUSE



ADVANCED ANALYTICS

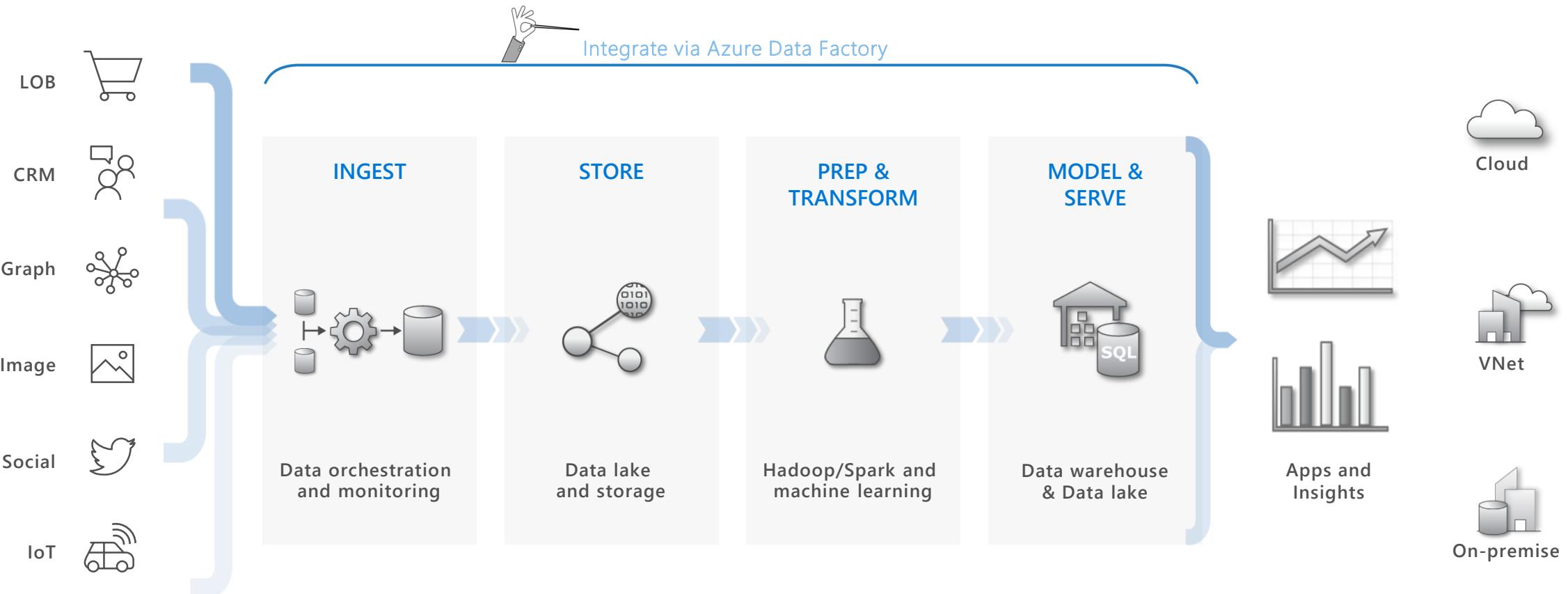


REAL TIME ANALYTICS

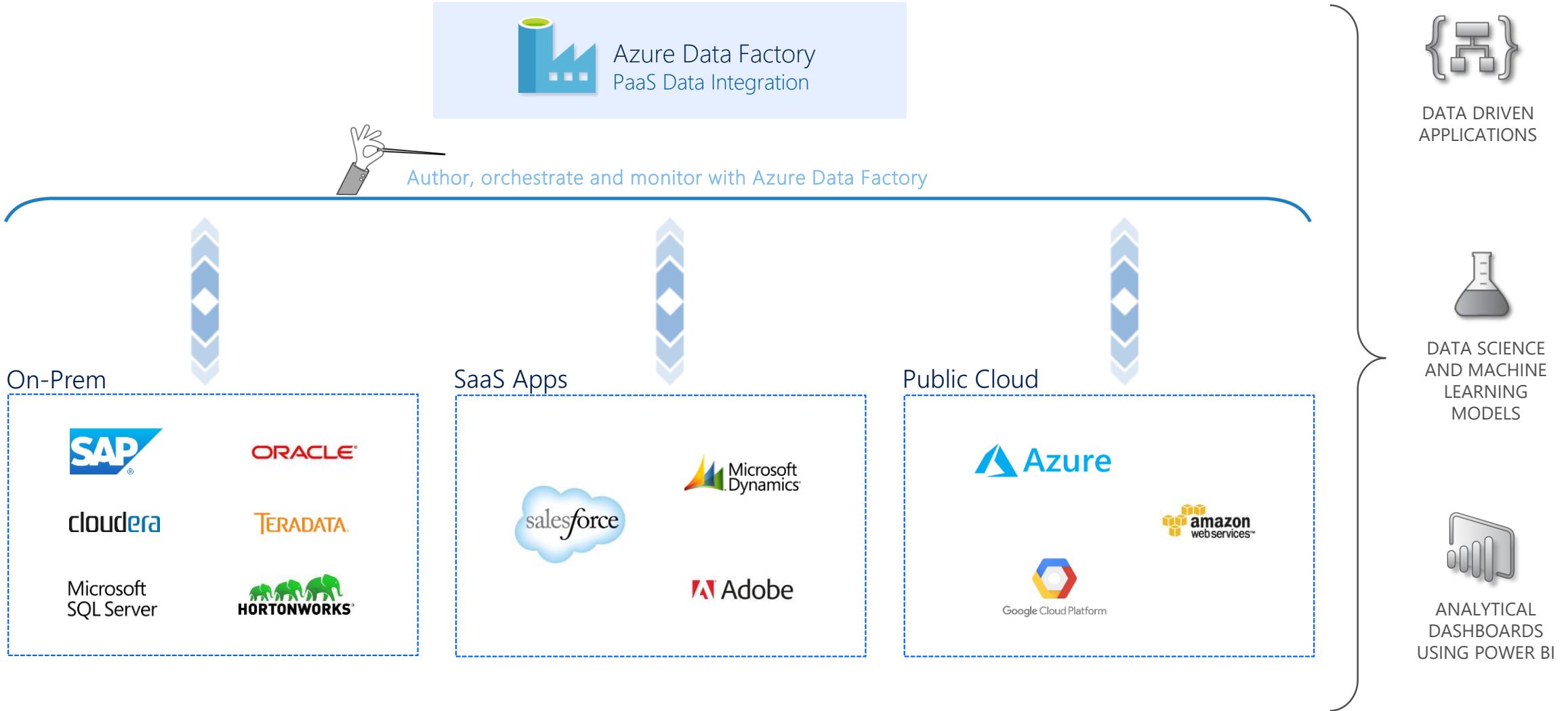


AZURE DATA FACTORY

Hybrid data integration, at global scale

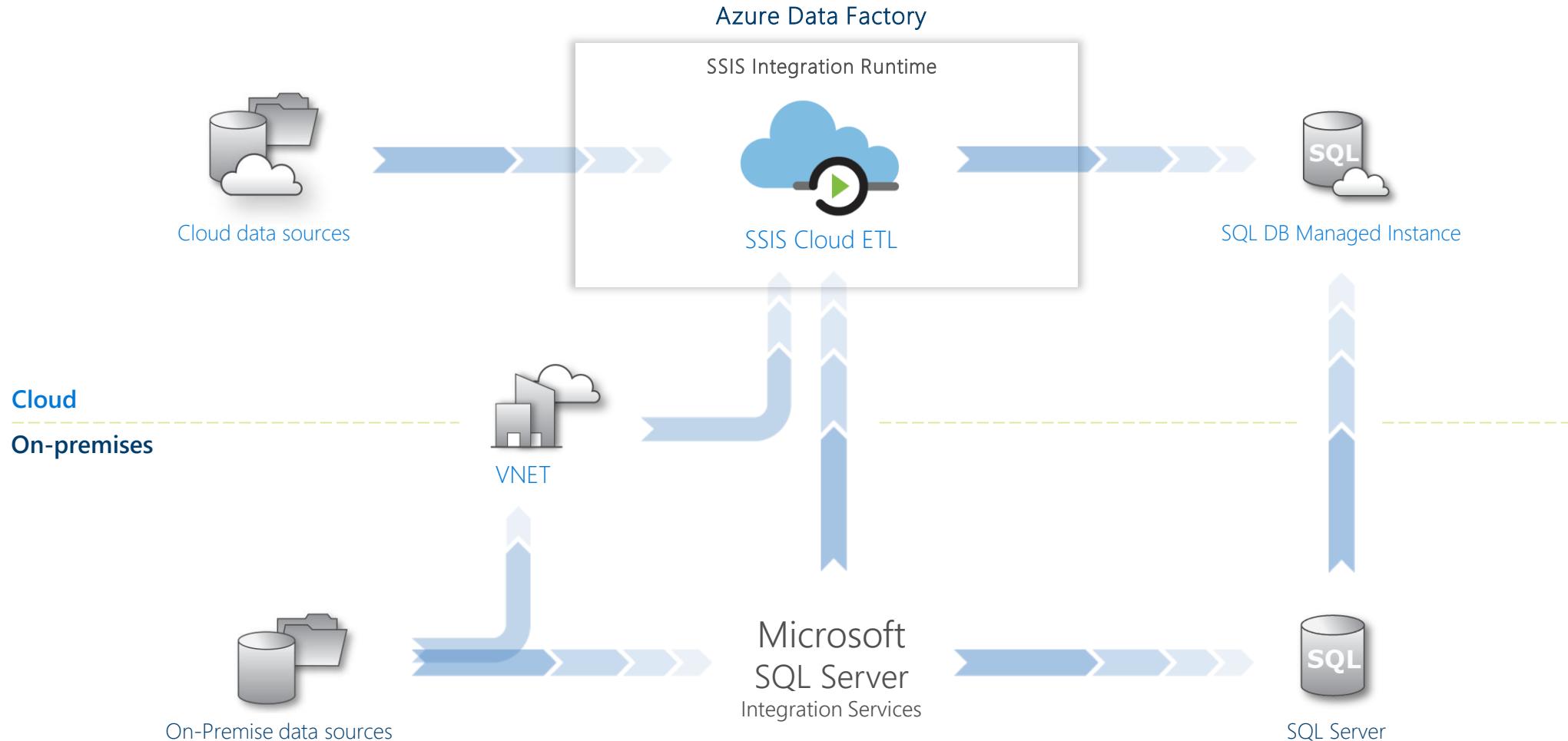


Hybrid and Multi-Cloud Data Integration



AZURE DATA FACTORY

Lift your SQL Server Integration Services (SSIS) packages to Azure



AZURE DATA FACTORY

Visual Data Transformation with Mapping Data Flow (in preview)



- ✓ Zero-code experience for data transformation
- ✓ Visually design, build, and manage transformation processes
- ✓ No understanding of Spark or distributed architecture needed
- ✓ Visual drag and drop interface

[Sign up for the preview of Mapping Data Flow >>>](#)

Access all your data

- 65+ connectors & growing
- Azure IR available in 20 regions
- Hybrid connectivity using self-hosted IR: on-prem & VNet

Azure	Database		File Storage	NoSQL	Services and Apps		Generic
Azure Blob Storage	Amazon Redshift	SQL Server	Amazon S3	Couchbase	Dynamics 365	Salesforce	HTTP
Azure Data Lake Store	Oracle	MySQL	File System	Cassandra	Dynamics CRM	Salesforce Service Cloud	OData
Azure SQL DB	Netezza	PostgreSQL	FTP	MongoDB	SAP C4C	ServiceNow	ODBC
Azure SQL DW	SAP BW	SAP HANA	SFTP	HDFS	Oracle CRM	Hubspot	Marketo
Azure Cosmos DB	Google BigQuery	Informix	Oracle Service Cloud		SAP ECC		
Azure DB for MySQL	Sybase	DB2	Drill	Amazon Marketplace	Zendesk	Oracle Responsys	Oracle Eloqua
Azure DB for PostgreSQL	Greenplum	MariaDB			Zoho CRM	Salesforce ExactTarget	
Azure Search	Microsoft Access	Drill			Magento	Atlassian Jira	
Azure Table Storage	Hive	Phoenix	Impala	PayPal	Concur	QuickBooks Online	Xero
Azure File Storage	Hbase	Presto			Shopify		
	Vertica	Spark			GE Historian	Square	

* Supported file formats: CSV, AVRO, ORC, Parquet, JSON

Performance Reference

		Cloud Sinks										On-prem Sinks**				
		# of <i>cloud DMUs</i> OR <i>gateway nodes</i>	Azure Blob (GRS)	Azure Data Lake Storage Gen1	Azure Data Lake Storage Gen2	Azure SQL Data Warehouse (6000 DWU)		Azure SQL Database (P11)		Azure Cosmos DB (SQL API; 100,000 RU)		Azure Table		On-prem SQL Server		
						PolyBase	Bulk Insert									
Cloud Sources	Azure Blob (GRS)	4 8	56 105	56 105	56 105	1250	*	5	6	*	11	129				
	Azure Data Lake Storage Gen1	4 8	56 120	56 108	56 120		1060	*	5	6	*					
	Azure Data Lake Storage Gen2	4 8	56 120	56 108	56 120	*	*	5	6	*	10	114				
	Azure SQL Data Warehouse (2000 DWU)	4	9	8	9	6	1	8	6	0.3						
	Azure SQL Database (P11)	4	9	8	9	6	1	8	6	0.3	14	*				
	Azure Table	4	2	2	2	*	2	2	1	1	1	1	*			
	Azure Cosmos DB (SQL API; 100,000 RU)	4	2	2	2	*	2	2	2	*	*	*	*			
	Amazon S3	8	107	101	120	69	*	*	*	*	*	*	*			
On-prem Sources	Amazon Redshift	4	*	*	*	7.2	*	*	*	*	*	*	*			
	On-premises SQL Server	1	7	7	7	18	0.4	7	6	0.2	*	*	*			
	On-premises File System	1 4	195 505	192 510	*	102 *	0.3	6	6	0.2	*	*	*			
	On-premise HDFS	1 4	179 500	183 525	*	83 *	0.3	3	6	0.2	*	*	*			

Unit: MB/s (Megabyte per second)

*: The throughput numbers for this source-sink combination will be published later.

**: For copying from cloud sources to on-prem sinks, single Self-hosted Integration Runtime node was used.

Azure Data Factory

Let's get started



Create pipeline



Copy Data



Configure SSIS Integration Runtime

Overview



Overview Video



Introduction to Data Factory



Lift & shift SSIS packages

CustomerChurnFactory

Pipeline 1 X

Activities

Save Run

Search Resources

Search Activities

AzureML

Custom.Net

Data Prep

Hive

Map Reduce

Pig

Stored Procedure

Spark

Data Flow

CustomerChurnPipeline

MarketingCampaignPipeline

ProductUsagePipeline

ProductUsagePipeline2

Data Flows

CustomerChurned

CustomerInfo

CallDataRecords

Datasets

Linked Services

Integration Runtimes

Triggers

Repository Settings

Enter Git repository information to be associated with your Data Factory: makadfv2c

Repository type:

- Visual Studio Team Services Git

Visual Studio Team Services Account:

- markkromer

Project Name:

- MakADFV2

Git repository name:

- Create new
- Use existing

makadfv2c

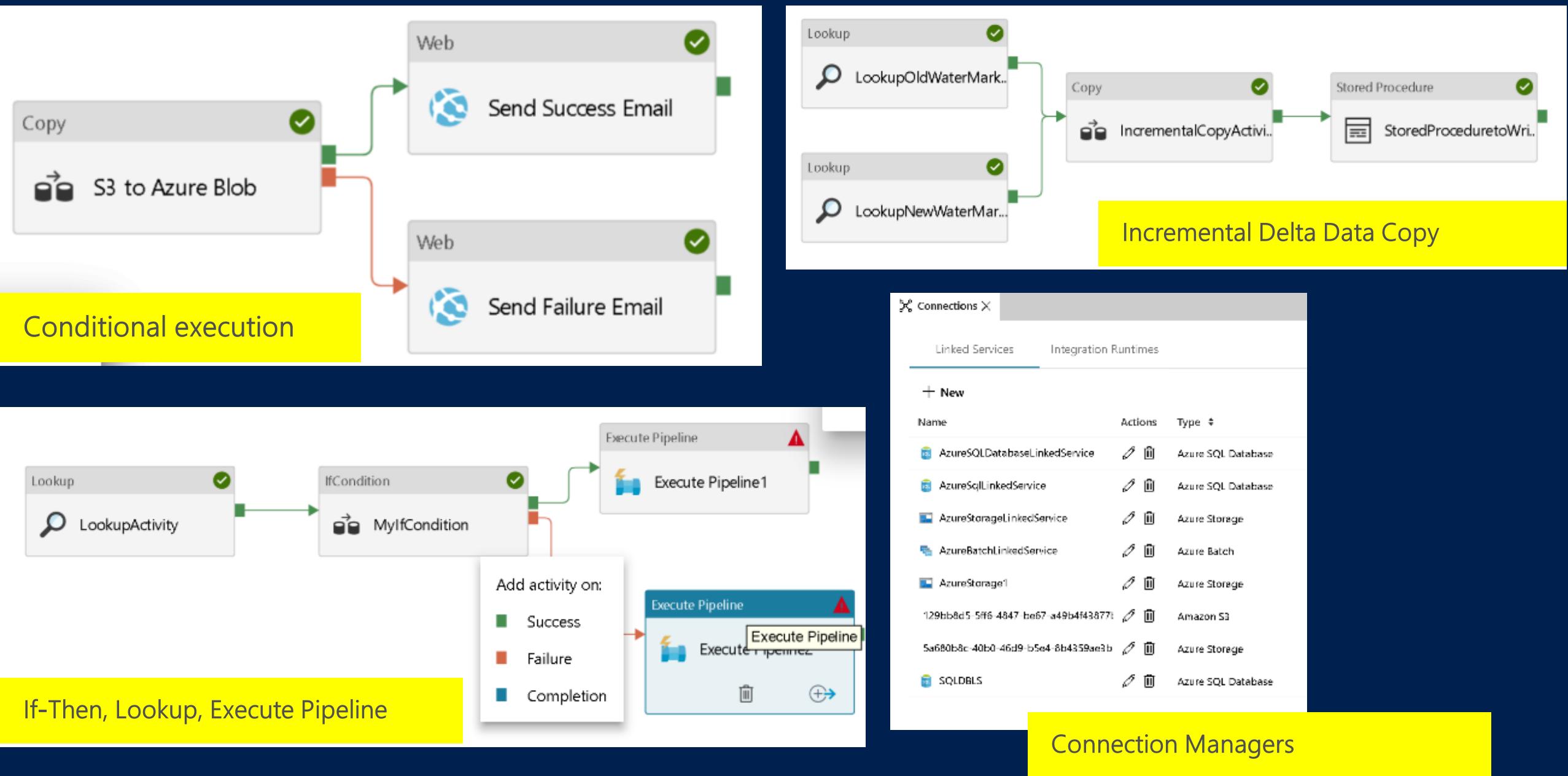
Built-in source control support

Output Log Run Details Errors Configuration

7/12/2017 8am UTC: Started CallDataRecords & CustomerInfo to CustomerChurned activity.

Search output list

```
graph LR; subgraph DF1 [Data Flow]; direction TB; A1[CallDataRecords] --> B1[ProcessCallLogs]; end; subgraph DF2 [Data Flow]; direction TB; C1[CustomerInfo] --> D1[CustomerChurned]; end; B1 --> D1; B1 --> E1[SendEmail];
```



1 Properties
One time copy**2 Source**

- Connection
- Dataset

3 Destination**4 Settings**
Fault tolerance**5 Summary****6 Deployment**

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store. Click [HERE](#) to suggest new copy sources or give comments.

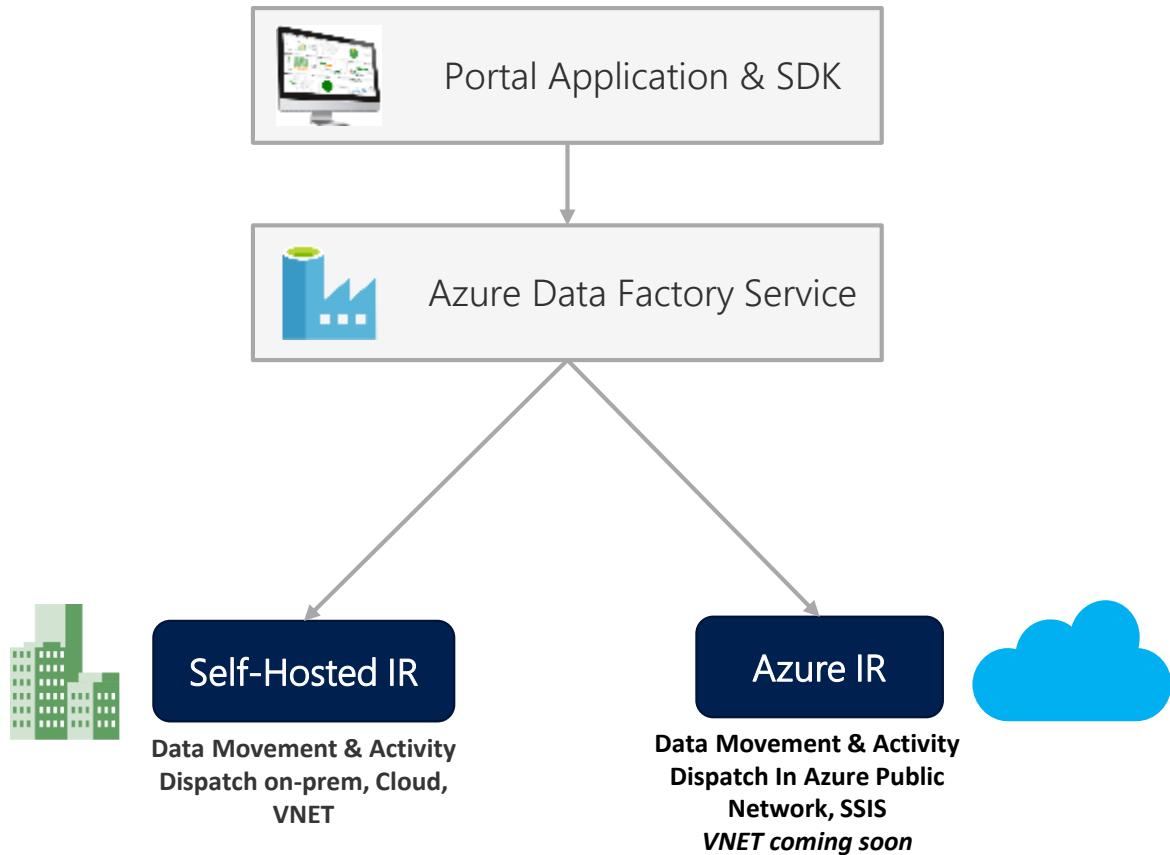
Easy-to-use Wizard for Copying Data at Scale

[FROM EXISTING CONNECTIONS](#)[CONNECT TO A DATA STORE](#)

Amazon Redshift	Amazon S3	Azure Blob Storage	Azure Cosmos DB	Azure Data Lake Store	Azure Database for MySQL
Azure Database for PostgreSQL	Azure File Storage	Azure SQL Data Warehouse	Azure SQL Database	Azure Table Storage	Cassandra

[Previous](#)[Next](#)

ADF Integration Runtime (IR)



- ADF compute environment with multiple capabilities:
 - Activity dispatch & monitoring
 - Data movement
 - SSIS package execution
- To integrate data flow and control flow across the enterprises' hybrid cloud, customer can instantiate multiple IR instances for different network environments:
 - On premises (similar to DMG in ADF V1)
 - In public cloud
 - Inside VNet
- Bring a consistent provision and monitoring experience across the network environments

↔ Command & Control

↔ Data Flow



Azure Cloud



On Premises Apps & Data



TERADATA



cloudera



ORACLE

Cloud Apps, Svcs & Data



Adobe



Demo

- Local CSV sync to Azure SQL DB as table (truncate & load)
- 2 points for demonstrate:
 - GUI base development while Github managed CI/CD
 - DIU (Data Integration Unit) & Degree of copy parallelism
 - DIU only available in cloud runtime
 - Billing happens on same of duration across DIUs
 - License lump sum VS pay-as-you-go



Online retailer uses cloud database to deliver world-class shopping experiences

Challenge

Increasing amount of customer data in various forms and shapes and telemetry from disparate data sources

Building intelligent recommendations and monitoring and debugging transformations on big data compute was complex and time consuming.

Solution

With Azure Data Factory, ASOS can easily ingest data from disparate sources in a serverless manner and is now able to debug, monitor and manage the solutions from a single pane of glass.

ASOS easily integrated custom code to drive intelligent recommendation within the data factory pipeline.

Recognize Any of These Challenges?

Users spend more time **looking** for data, than they do analyzing it

Data is sitting in **multiple sources**, but no insight into which data sits where

Many different **data ecosystems** across the enterprise, but no way to share data artifacts across them

Need data consumption in multiple **different tools**, but no common way of enabling discovery and access to data sources across them

Users are busy **re-producing** data assets that already exist

No way of **tracking usage** of our BI and Analytics assets



What is Azure Data Catalog?

An enterprise-wide directory in Azure that enables self-service discovery of data from *any source*

A metadata repository that allow users to register, enrich, understand, discover, and consume data sources

How is it Different?

Data source discovery



No need for data movement
No latency
No heavy up front IT investment

Data from any source



Structured and unstructured
on premises and in the cloud
Microsoft and non-Microsoft

Consumption through any tool



Enabling publishing, discovery and
consumption of data sources through your
tool of choice

Powered by annotation crowdsourcing



Empowering any user to capture and
share their knowledge about registered
sources

What Can I Do With It?

Publisher

Publish

Register Data Sources

Enrich

Categorize –
Annotate

Consumer

Discover

Browser- Search

Understand

Get context – Identify Intent

IT Admin

Govern

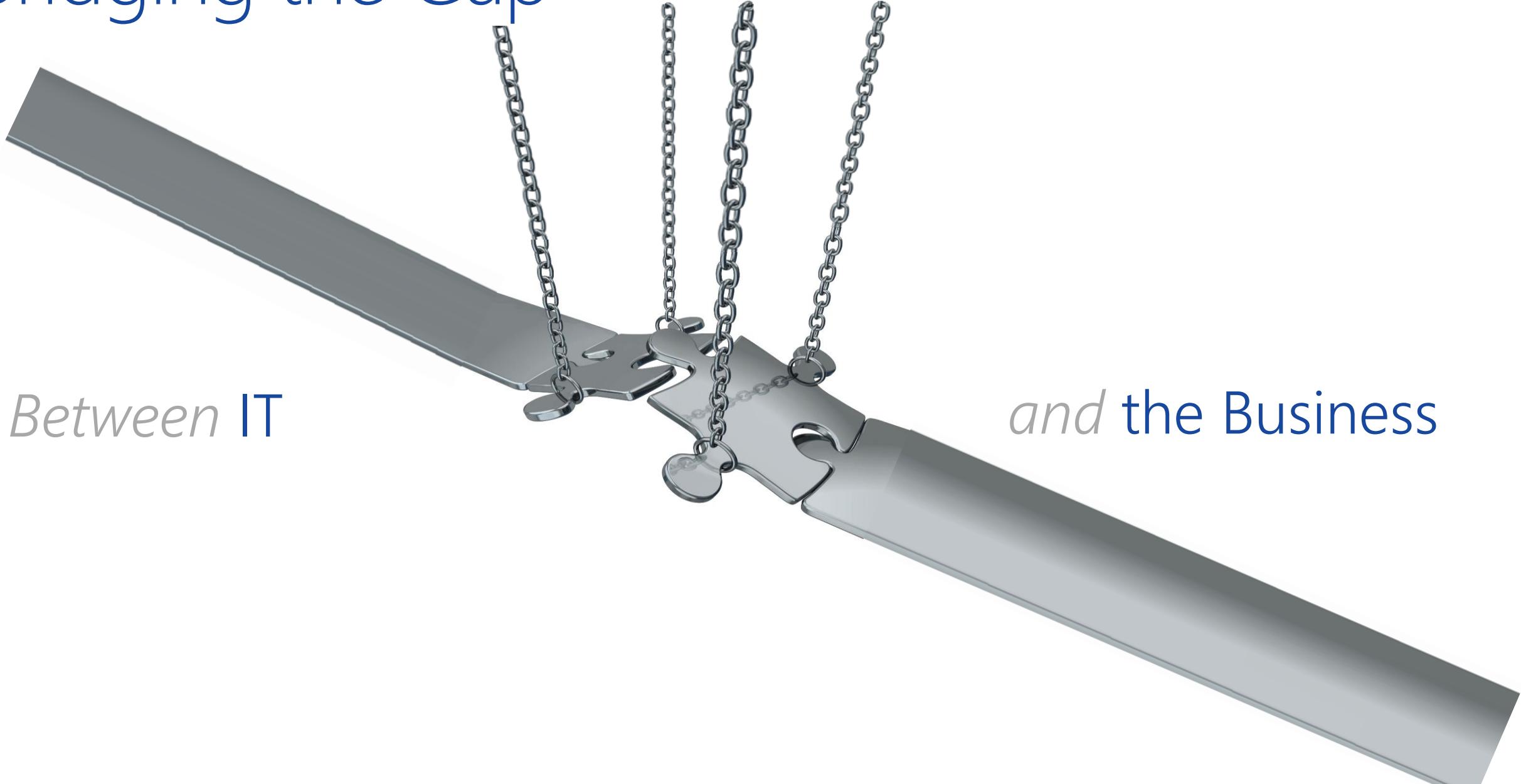
Apply Policies - Control Access

Analyze

Track and monitor usage

Extend

Bridging the Gap



Between IT

and the Business

Azure Data Catalog Programmability

Programmability Scenarios

Incorporating data source discovery into client applications

Future: Defining custom data source and data asset types ("type extensibility")

REST APIs

Register data asset operations: Register Data Asset, Delete Data Asset, and Search Data Asset

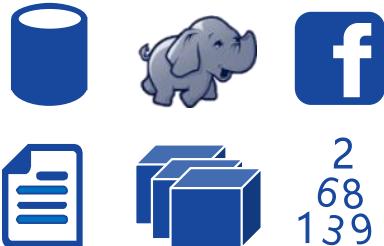
Annotate data asset operations: Annotate Data Asset, Update Annotation, Get Data Asset with Annotations, and Delete Annotation

Search API uses familiar, powerful, search syntax

Enabling the Entire Enterprise Data Ecosystem

Fostering a collaborative virtuous cycle for the Data Value Chain

The Enterprise Data



The Public Data



- Provision
- Configure



- Secure
- Monitor
- Operationalize

DATA PRODUCERS



- Refine
- Combine
- Produce
- Annotate
- Share

DATA STEWARDS



- Discover
- Combine
- Clean
- Annotate
- Publish/Share
- Certify
- Categorize
- View usage analytics and feedback

INFORMATION CONSUMERS



- Discover
- Understand
- Connect
- Consume

Register Data Asset

Microsoft Azure Data Catalog

Select a Data Source:

Search 

 SQL Server	 SQL Data Warehouse	 Analysis Services	 Reporting Services	 Oracle Database
 Tables	 Apache Hadoop	 Apache Spark	 Apache Flink	 Apache Beam

NEXT

© 2016 Microsoft [Privacy + terms](#)

Perform Discovery

Microsoft Azure Data Catalog

← Change Connection / Select Objects

Server Name: markg.database.windows.net

Please select the objects to be registered with Azure Data Catalog. The structural metadata for the selected objects, including object names, attribute names, and data types will be registered in Data Catalog. For data sources that support previews, you can choose to include a 20-record snapshot of the object's data.

Server Hierarchy:

- RLTest
 - dbo
 - SalesLT

Available objects:

Filter Objects

Tables Views Stored Procedures
 Functions

Name	Schema	Type
Address	SalesLT	Table
Customer	SalesLT	Table
CustomerAddress	SalesLT	Table
SalesOrderDetail	SalesLT	Table
SalesOrderHeader	SalesLT	Table
vGetAllCategories	SalesLT	View
vProductAndDescription	SalesLT	View
vProductModelCatalogDesc...	SalesLT	View

Objects to be registered:

Name

Product
ProductCategory
ProductDescription
ProductModel
ProductModelProductDescription

Include Preview
Include Data Profile

Add an expert...
sales, product, azure sql

REGISTER

Important Notice:
All metadata and preview data registered with Azure Data Catalog will be stored in westus.

© 2017 Microsoft Privacy + terms

Data Profile

Preview Columns **Data Profile** Documentation

Table Profile

Number Of Rows	Size	Last Data Update	Last Schema Update
1069	56 MB		8/2/2010

Column Profile

Column Name	Data Type	Null Count	# Distinct Values	Minimum	Maximum
ProductID	int	0	432	1	999
LocationID	smallint	0	14	1	60
Shelf	nvarchar	0	21	A	Y

Data Profiling - Information

A data profile in Azure Data Catalog shows table and column profile information including:

Object data profile

- Number of rows
- Table size
- When the object was last updated

Column data profile

- Column data type
- Number of distinct values
- Number of rows with NULL values
- Minimum, maximum, average, and standard deviation for column values

Tagging and searching by glossary terms

1 search results, 1 selected | Select All

Quarterly Disk Revenue

This asset tracks the quarterly revenue for the Disk business.

Experts:
test@test.com

 Revenue (Sales)  Sales  Computed

Contained In Database:
Southwind

 SQL SERVER TABLE

 Open In ... 

Preview Columns Documentation ^

> Properties

 Preview  Columns  Docs

Add...

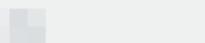
Tags:

 Computed  Revenue (Sales)  Sales  Add...

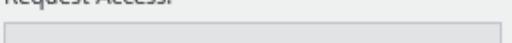
Connection Info:

Server Name: 

Database Name: 

Schema Name: 

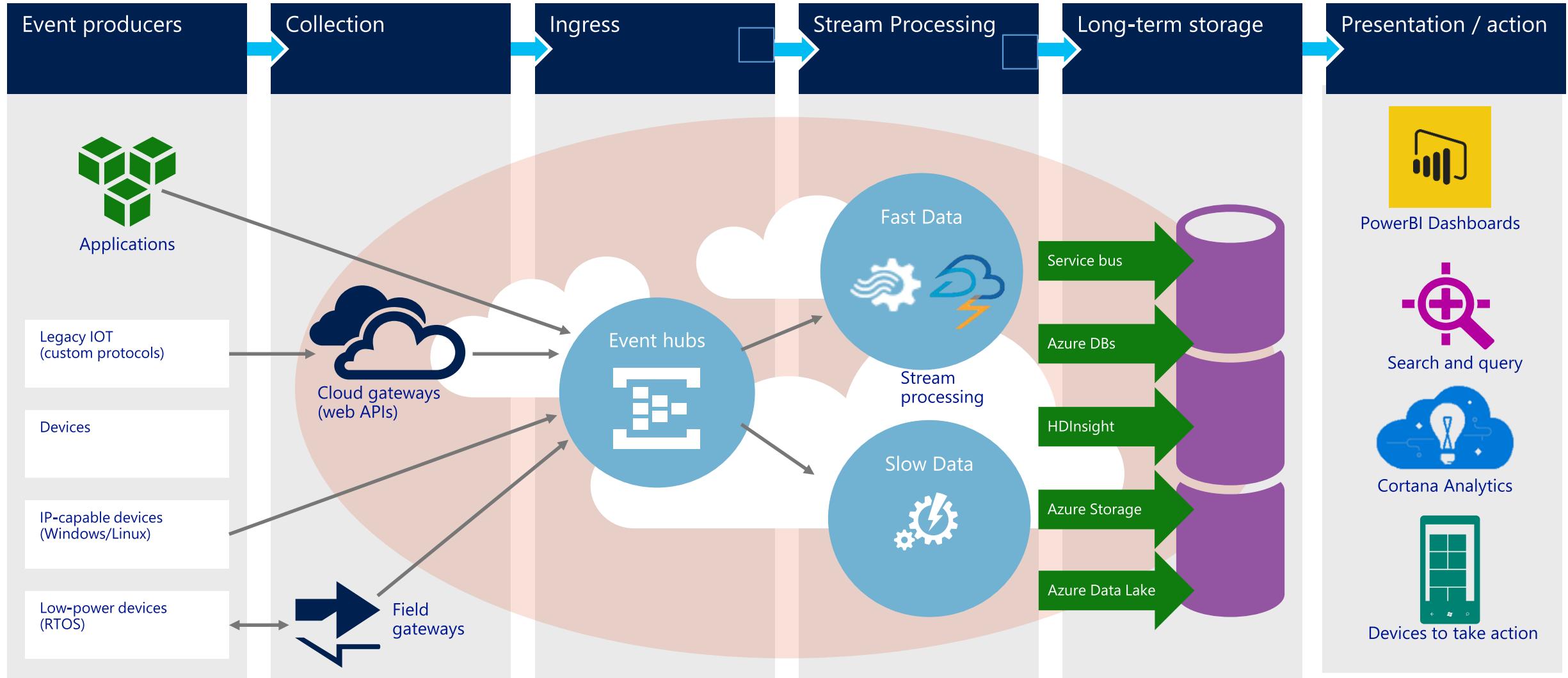
Object Name: 

Request Access: 

add information on how to request data

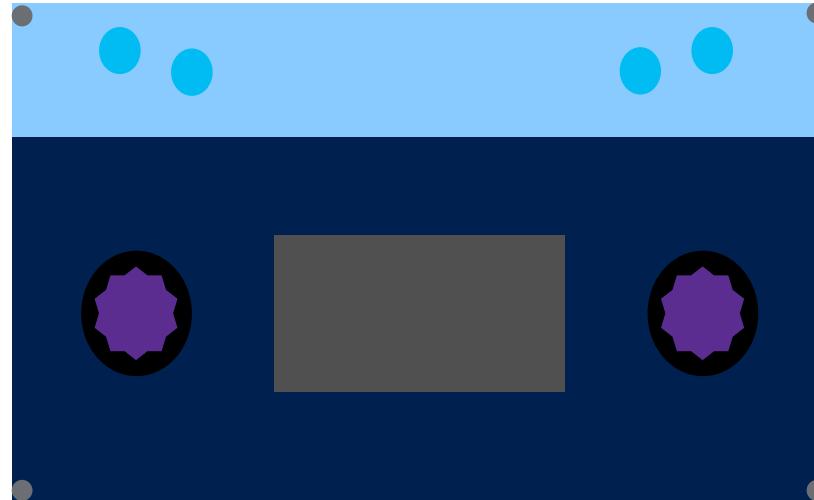
Event Hubs and IoT Hub

Where Telemetry & Streaming fit in



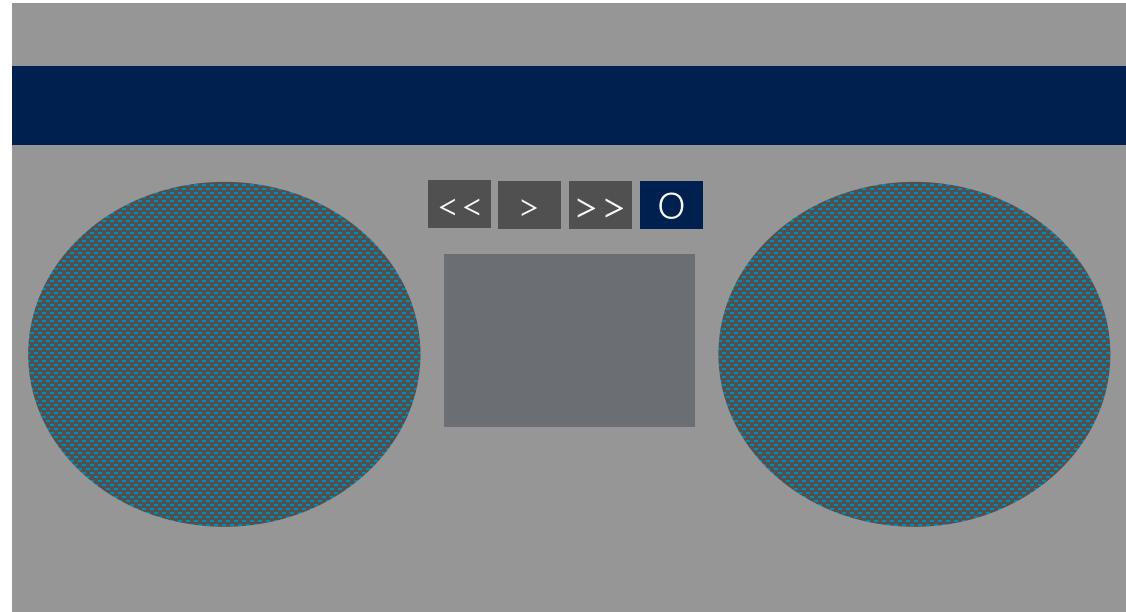
How Event Hubs is different from queues

- Records a stream
- Recording moves forward only
- You can plan the tape over and over again



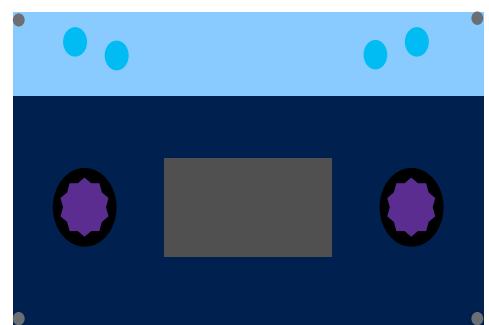
Event Hubs is like a tape deck

- Can play from anywhere on the tape
- FastFwd or Rewind then Play
- Tape loops when its done

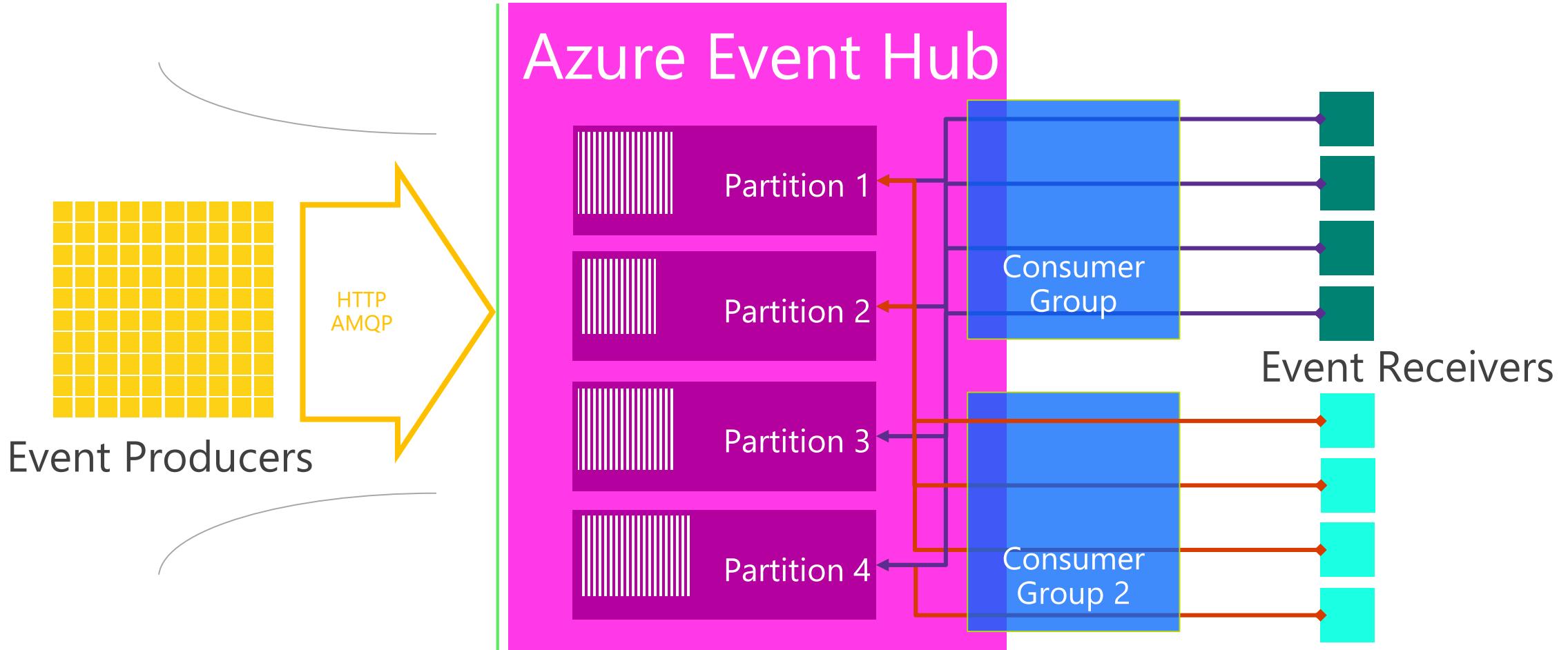


Event Hubs partitions

- A cassette tape actually has Left and Right channels
- When you press record, they both record
- But the data on each channel is different
- The left and right speakers each play one channel
- Event Hubs calls these partitions



Event Hubs conceptual architecture



Event Hubs in the real world: Halo 5

- 80 million requests per minute within 24 hours of release
- All game telemetry and statistics run through Azure Event Hubs, processed, and sent back to console
- 1 Dedicated Capacity cluster (3 CUs)
- Zero administration by Halo team



Azure IoT Hub is ideal service for IoT deployments

Connect, monitor and manage billions of IoT assets. Ideal for IoT deployments proving path forward from POC to full scale, while maximizing IoT security

Establish bi-directional communication



Device to cloud telemetry



Cloud-to-device messages



Durable command messages

Provide enhanced security for IoT solutions



Per device authentication



Individual credentials



Selectively revoke access rights

Provide device lifecycle management



Device Management at scale



Remotely maintain devices from the cloud



Edge intelligence via Gateway SDK

Event Hub is ideal service for telemetry ingestion from websites, apps and streams of big data

Cloud-scale telemetry ingestion service that can log millions of events per second in near real time

Stream millions of events per second



Telemetry and logging



Scale ingestion service



Distributed streaming platform

Process real-time and batch on same stream



Event Hub Archive* to load data to Azure



Batch processing



Real time processing

Handle volume, variety and velocity of data



Fully-manage service



Ingest events with elastic scale



Accommodate variable load profiles

Comparison on IoT Hub and Event Hub

IoT Capability	IoT Hub standard tier	IoT Hub basic tier	Event Hubs
Device-to-cloud messaging	✓	✓	✓
Protocols: HTTPS, AMQP, AMQP over webSockets	✓	✓	✓
Protocols: MQTT, MQTT over webSockets	✓	✓	
Per-device identity	✓	✓	
File upload from devices	✓	✓	
Device Provisioning Service	✓	✓	
Cloud-to-device messaging	✓		
Device twin and device management	✓		
Device streams (preview)	✓		
IoT Edge	✓		

Messaging Protocol

- AMQP
 - Advanced Message Queuing Protocol
 - AMQP is a binary wire protocol which was designed for interoperability between different vendors.
 - JP Morgan use it to process 1 billion messages a day. NASA uses it for Nebula Cloud Computing. Google uses it for complex event processing
- MQTT
 - Message Queue Telemetry Transport
 - Designed for resource-constrained devices and low bandwidth, high latency networks such as dial up lines and satellite links
 - Facebook are using it as part of their mobile applications because it has such a low power draw and is light on network bandwidth

Manufacturing



Selling Packaging as a Service

Connecting machines to collect real-time data has enabled performance and future failure monitoring, allowing Tetra Pak to revolutionize their business model

[LEARN MORE >](#)



Optimizing the factory floor

Used IoT, machine learning, artificial intelligence, and CRM to optimize processes, planning, and predictive maintenance scheduling to avoid downtime

[LEARN MORE >](#)



Filtering the signal from the noise

Used analytics to discover actionable insights around fuel usage, predictive maintenance and stop unscheduled delays

[LEARN MORE >](#)

Transportation



fathym

Enabling officials to keep drivers safe

Using highly localized road condition and weather data, Fathym keeps drivers safe by recommending reroutes, and avoiding unnecessary use of expensive assets

[LEARN MORE >](#)



TEXA

Creating efficiency and agility with IoT

Texa S.P.A saw product orders increase 100% after using IoT to enhance their product offering, support business growth, reduce costs, and increase mechanic efficiency and safety

[LEARN MORE >](#)



 Transport
for London

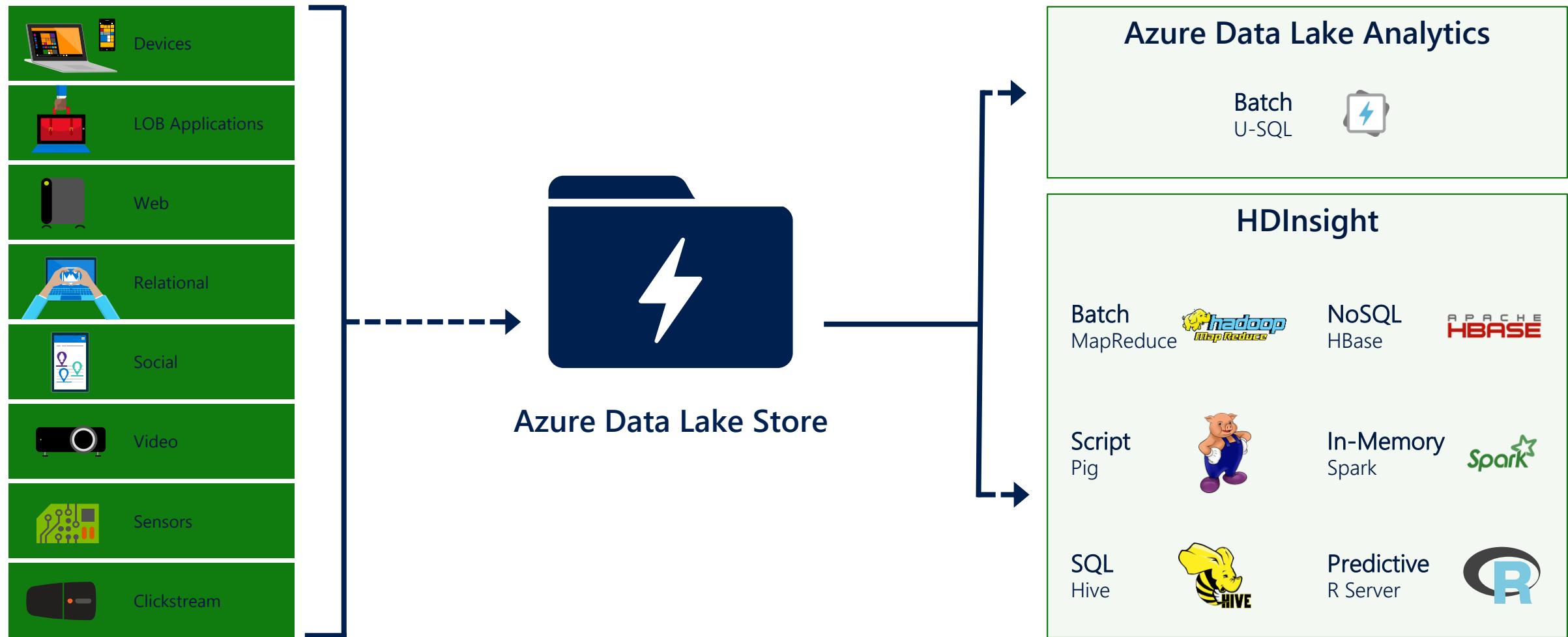
Low cost connectivity for train sensors

Using Azure IoT and TV whitespaces, an infrastructure not built for connected devices now supports delivery of timely insights about traffic conditions and cable car equipment

[LEARN MORE >](#)

Big Data analytics workloads – Gen1

A highly scalable, distributed, parallel file system in the cloud specifically designed to work with a variety of big data analytics workloads



Azure Data Lake Storage Gen2

A “no-compromises” Data Lake: secure, performant, massively-scalable Data Lake storage that brings the cost and scale profile of object storage together with the performance and analytics feature set of data lake storage



SECURE

- ✓ Support for fine-grained ACLs, protecting data at the file and folder level
- ✓ Multi-layered protection via at-rest Storage Service encryption and Azure Active Directory integration



MANAGEABLE

- ✓ Automated Lifecycle Policy Management
- ✓ Object Level tiering



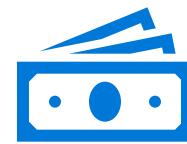
FAST

- ✓ Atomic file operations means jobs complete faster



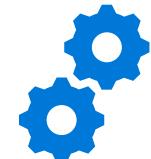
SCALABLE

- ✓ No limits on data store size
- ✓ Global footprint (50 regions)



COST EFFECTIVE

- ✓ Object store pricing levels
- ✓ File system operations minimize transactions required for job completion

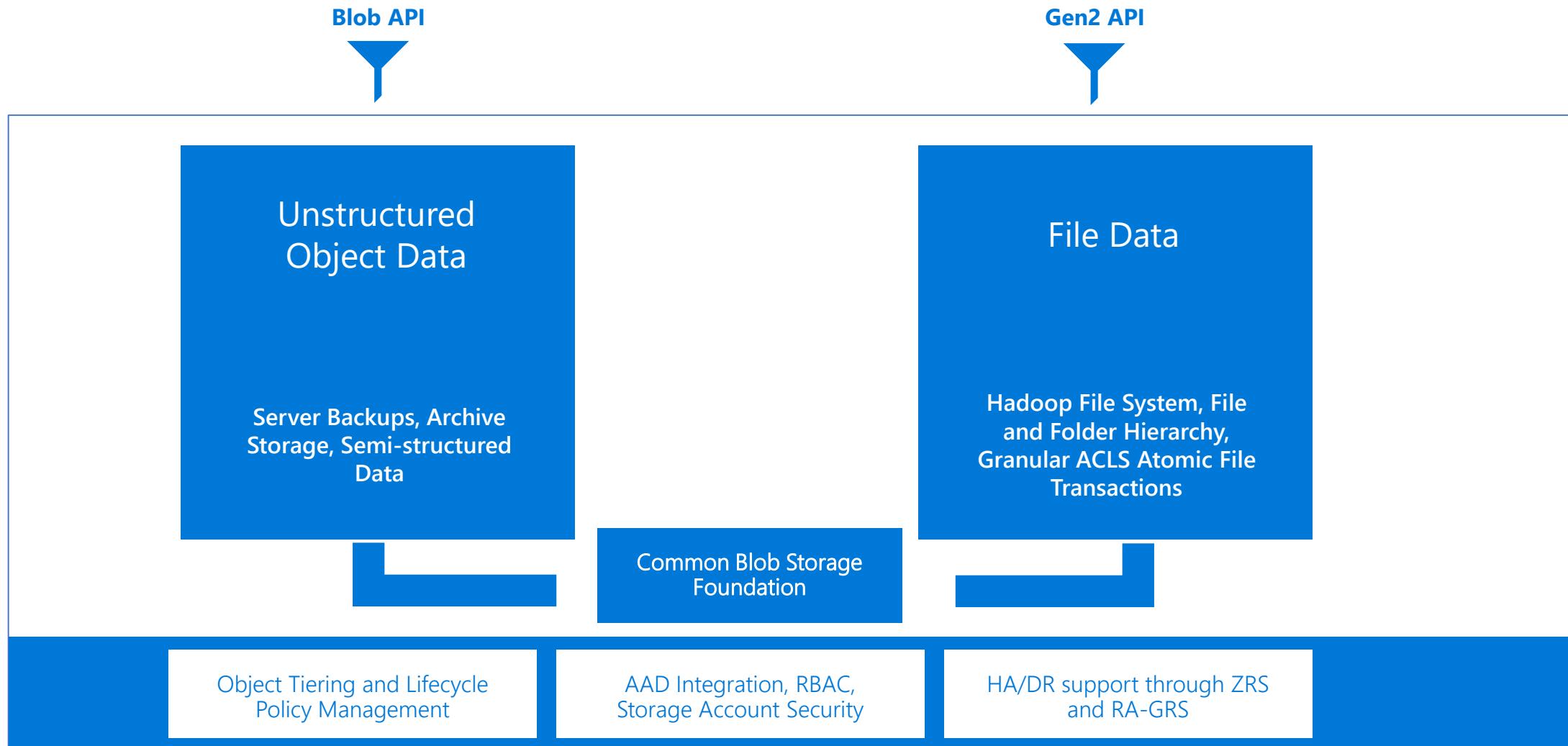


INTEGRATION READY

- ✓ Optimized for Spark and Hadoop Analytic Engines
- ✓ Tightly integrated with Azure end to end analytics solutions

Azure Data Lake Storage Gen2

ADLS Gen2 adds a high performance HDFS Endpoint to Azure Blob Storage and inherits the rich feature set of Azure Blob Storage *



Azure Blob Tiers

Data Lake Storage Pricing Model

Data Lake Storage Pricing			
	 Hot	 Cool	
 PER TB PER MONTH	Frequently accessed data	Less frequently accessed data	Rarely accessed data
	\$18.40	\$10.00	\$2.00

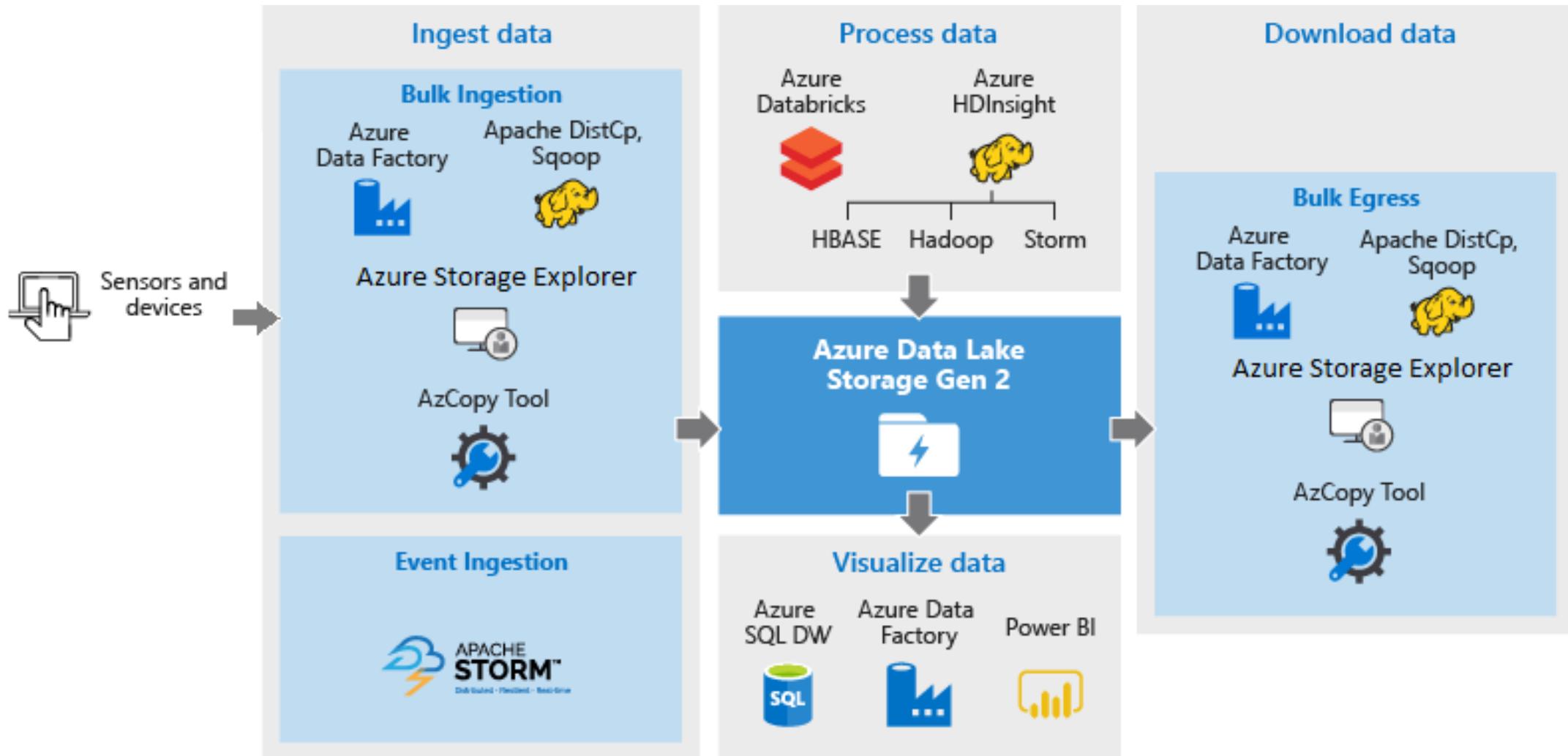
- Same storage pricing as Azure Blob Storage!*
- Supports object level tiering
- Competitive with other cloud providers



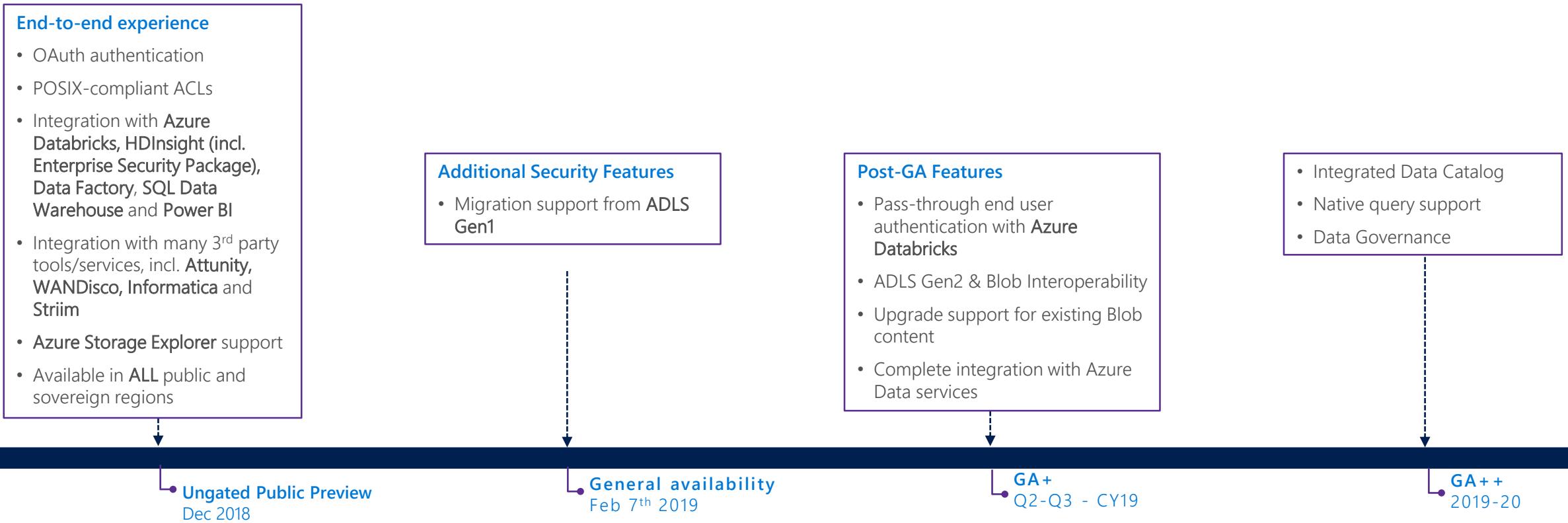
S3 Standard Storage

Big Query

ADLS Gen2 - Interoperability



ADLS Gen2 Roadmap



RICH PARTNER NETWORK

Trusted ecosystem to accelerate time to value

DATA MIGRATION



Informatica



SPARK & HADOOP ENGINES

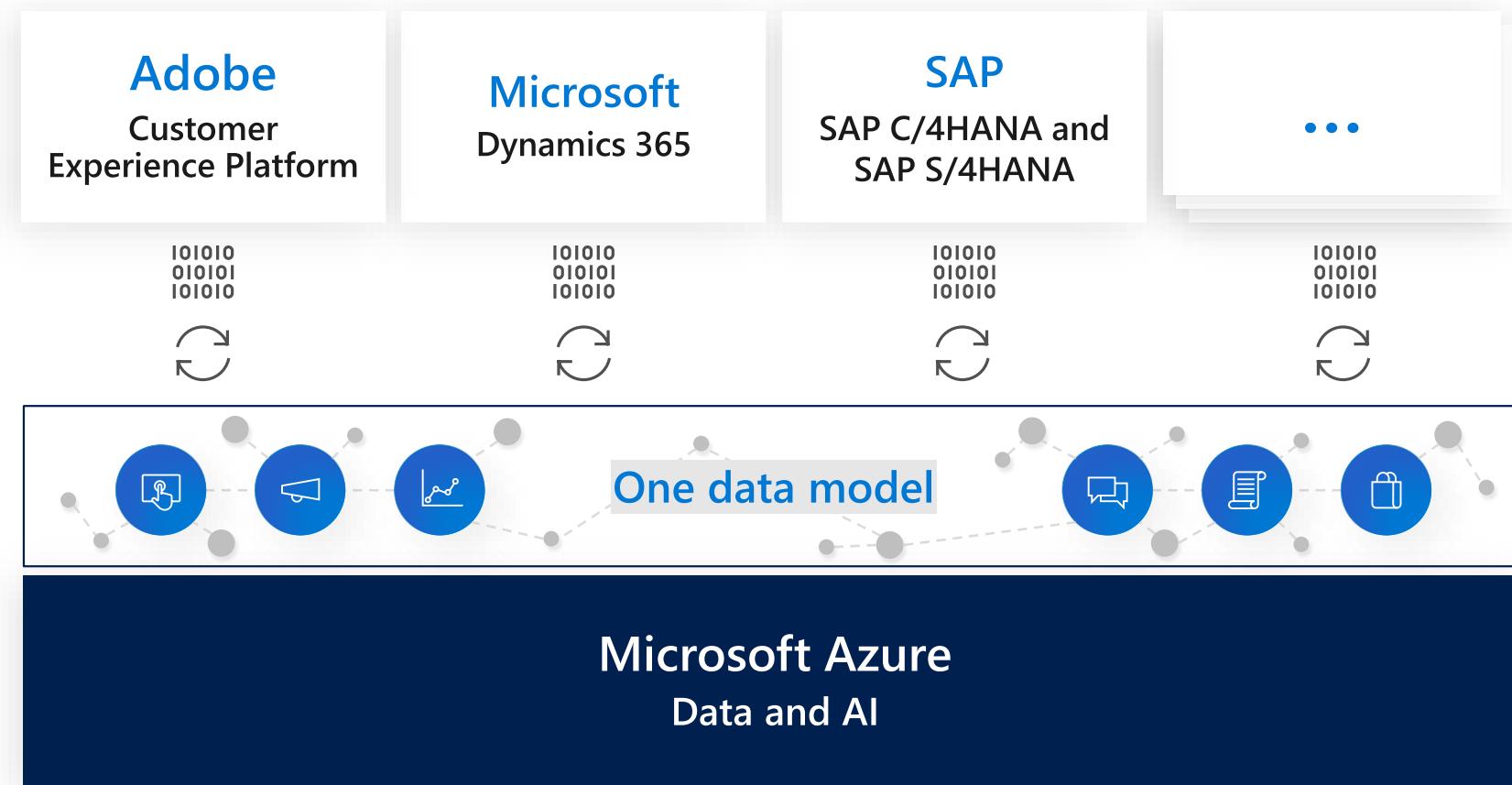


Ignite 2018 | Open Data Initiative announcement



Open Data Initiative, join us

Own, enrich, renew your data



Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service



Table API



cassandra



Key-value



Column-family



Document



Graph



Gremlin

SQL

MongoDB

Turnkey global distribution

Elastic scale out of storage & throughput

Guaranteed low latency at the 99th percentile

Five well-defined consistency models

Comprehensive SLAs

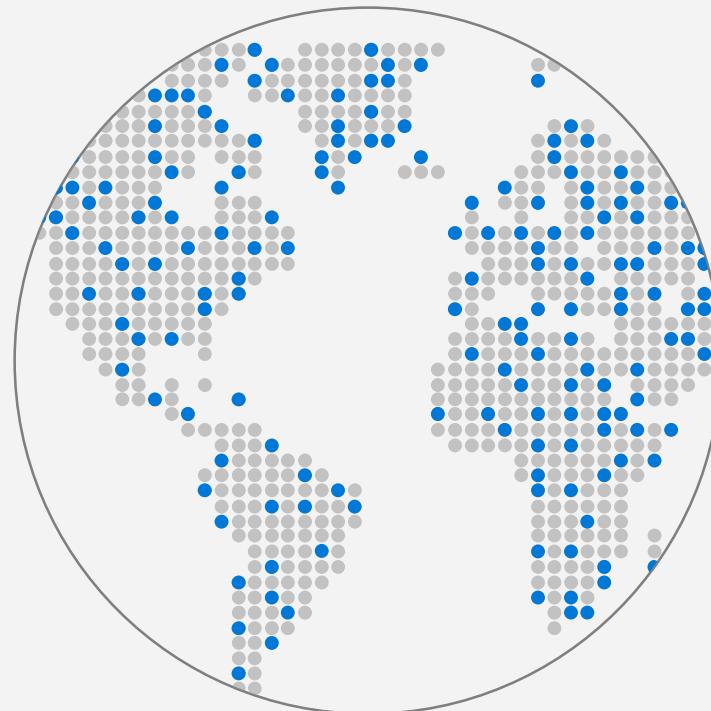


Turnkey global distribution

PUT YOUR DATA WHERE YOUR USERS ARE

Automatically replicate all your data around the world, and across more regions than Amazon and Google combined.

- Available in [all Azure regions](#)
- Manual and automatic failover
- Automatic & synchronous multi-region replication



FIVE WELL-DEFINED CONSISTENCY MODELS

CHOOSE THE BEST CONSISTENCY MODEL FOR YOUR APP

Offers five consistency models

Provides control over performance-consistency tradeoffs,
backed by comprehensive SLAs.

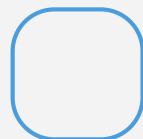
An intuitive programming model offering low latency and
high availability for your planet-scale app.



Strong



Bounded-stateless



Session



Consistent prefix



Eventual



Consistency Models - Breakdown

Consistency Level	Guarantees
Strong	Linearizability (once operation is complete, it will be visible to all)
Bounded Staleness	Consistent Prefix. Reads lag behind writes by at most k prefixes or t interval Similar properties to strong consistency (except within staleness window), while preserving 99.99% availability and low latency.
Session	Consistent Prefix. Within a session: monotonic reads, monotonic writes, read-your-writes, write-follows-reads Predictable consistency for a session, high read throughput + low latency
Consistent Prefix	Reads will never see out of order writes (no gaps).
Eventual	Potential for out of order reads. Lowest cost for reads of all consistency levels.

Multiple data models and apis

USE THE MODEL THAT FITS YOUR REQUIREMENTS, AND
THE APIs, TOOLS, AND FRAMEWORKS YOU PREFER

Cosmos DB offers a multitude of APIs to access and query data including, SQL, various popular OSS APIs, and native support for NoSQL workloads.

Use key-value, tabular, graph, and document data

Data is automatically indexed, with no schema or secondary indexes required

Blazing fast queries with no lag



Table API



cassandra



Key-value



Column-family

SQL



Document

MongoDB



Graph



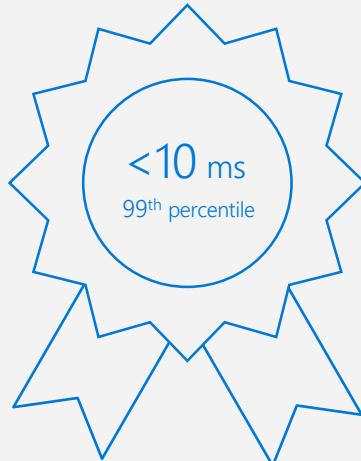
Gremlin

COMPREHENSIVE SLAS

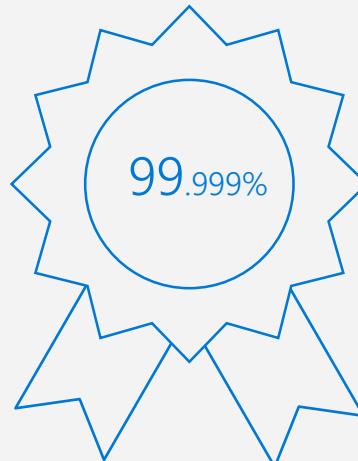
RUN YOUR APP ON WORLD-CLASS INFRASTRUCTURE

Azure Cosmos DB is the only service with financially-backed SLAs for millisecond latency at the 99th percentile, 99.999% HA and guaranteed throughput and consistency

Latency



HA



Throughput



Consistency



Replicating Data Globally

andrl-global - Replicate data globally
Azure Cosmos DB account

Search (Ctrl+/
Save Discard Manual Failover Automatic Failover

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Quick start Data Explorer

SETTINGS

Replicate data globally Default consistency Firewall Keys

Click on a location to add or remove regions from your Azure Cosmos DB account.
* Each region is billable based on the throughput and storage for the account. [Learn more](#)



Replicating Data Globally



```
[amypond@blink:~/docdb$ node testQ2.js
12.736112 milliseconds, 2 RUs, ActivityId: dd3c17b9-1b76-445a-8e27-29b7486bd7e4
4.947605 milliseconds, 2 RUs, ActivityId: e2f4c899-9fb1-4f76-a4ab-e5718fac5742
5.044005 milliseconds, 2 RUs, ActivityId: 0fc5d216-78a0-4d92-a3d0-63efd9dd6552
5.351205 milliseconds, 2 RUs, ActivityId: 861155f0-81ba-4c8a-9933-e50d8708cc21
4.553505 milliseconds, 2 RUs, ActivityId: 3db9641f-70f1-4ef1-84bb-809280bbe1a5
5.427506 milliseconds, 2 RUs, ActivityId: 10d1b2e5-e795-4c77-8655-815f410ba11e
5.900106 milliseconds, 2 RUs, ActivityId: bda1df86-c5ad-45c5-bcfb-93d417c54751
4.895405 milliseconds, 2 RUs, ActivityId: f206d58d-64a2-47f2-9653-145eaf47ff97
5.244306 milliseconds, 2 RUs, ActivityId: 3aa7e177-b1a9-413d-8023-55f5054d1b74
```

Automatic Failover

Automatic Failover



Enable Automatic Failover

ON

OFF

Drag-and-drop read regions items to reorder the failover priorities.

Tip: Drag on the left of the hovered row to reorder the list.

WRITE REGION

Central US

READ REGIONS

PRIORITIES

North Europe

1

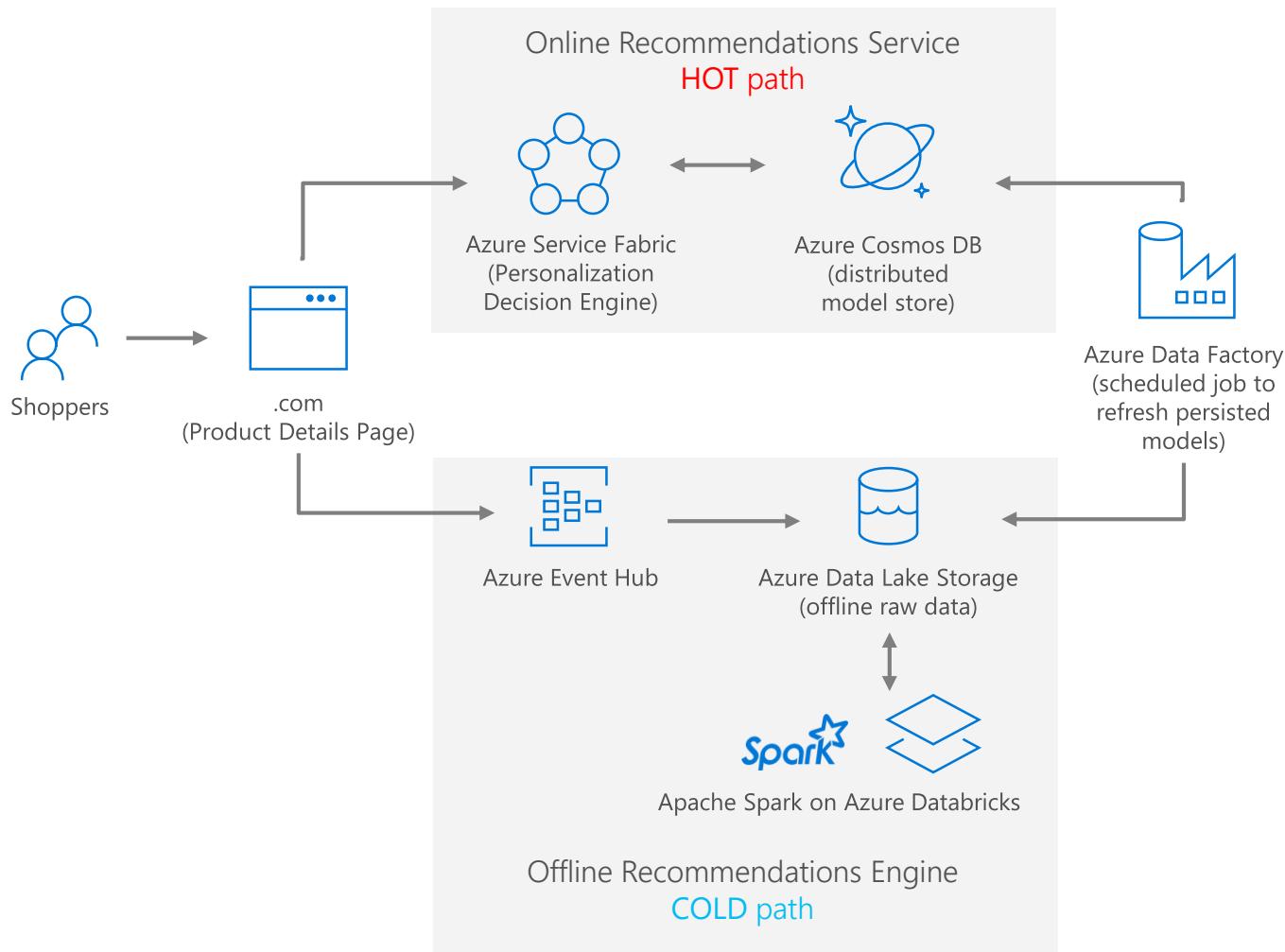
Southeast Asia

2

Build Real-Time Customer experiences

Offer latency-sensitive applications with personalization, bidding, and fraud-detection.

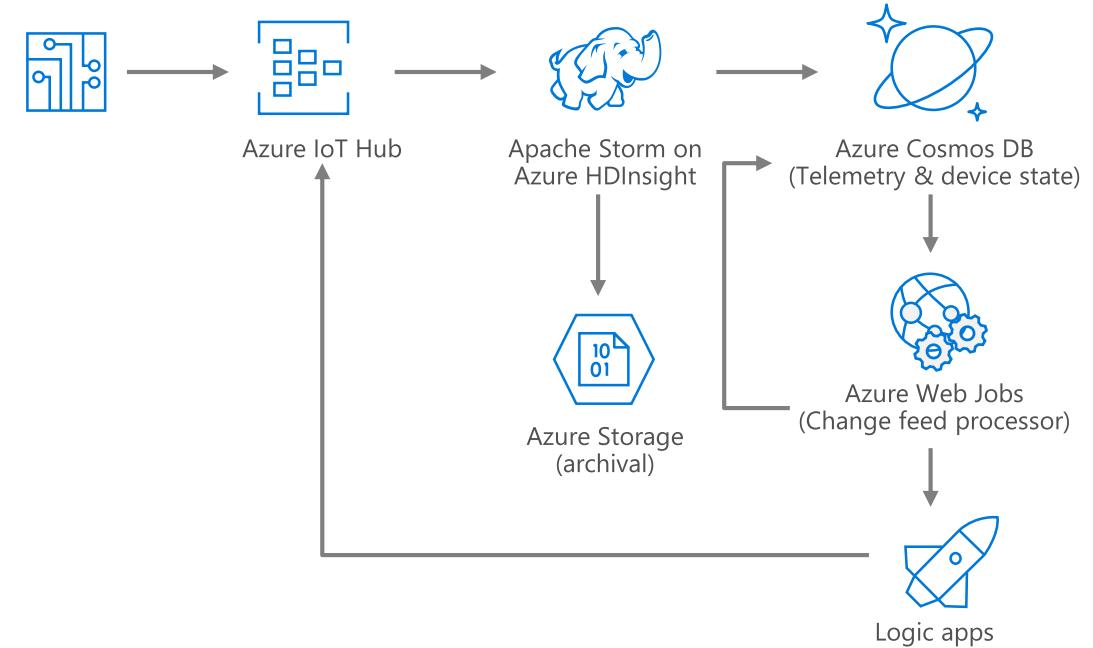
- Machine learning models generate real-time recommendations across product catalogues
- Product analysis in milliseconds
- Low-latency ensures high app performance worldwide
- Tunable consistency models for rapid insight



Massive Scale Telemetry Stores for IoT

Diverse and unpredictable IoT sensor workloads require a responsive data platform

- Seamless handling of any data output or volume
- Data made available immediately, and indexed automatically
- High writes per second, with stable ingestion and query performance



Choosing the right database service

Relational databases

Fully-managed services offering hybrid capabilities, and powerful built-in intelligence and machine learning capabilities to transactional apps.

Service	Great for:
Azure SQL Database 	<ul style="list-style-type: none">• Multi-tenant SaaS applications• OLTP applications• Migrating SQL Server or Oracle workloads to the cloud
Azure Database for PostgreSQL 	<ul style="list-style-type: none">• Secure and scalable apps• Personalized digital marketing• Migrating PostgreSQL or Oracle workloads to the cloud
Azure Database for MySQL and MariaDB 	<ul style="list-style-type: none">• Websites and web apps• Low-latency digital gaming• Simple, fast, read-heavy operations• Migrating MySQL/MariaDB workloads to the cloud
SQL Server on Virtual Machines 	<ul style="list-style-type: none">• Apps using any SQL Server version and edition on any virtual machine size, on Windows and Linux

Non-relational databases

Fully-managed services offering guaranteed low-latency and elastic scaling to real-time web and mobile apps.

Service	Great for:
Azure Cosmos DB 	<ul style="list-style-type: none">• Real-time personalization and analytics• Scalable IoT applications• Handling unpredictable traffic patterns• Migrating Cassandra, MongoDB, and other NoSQL apps and data to the cloud
Azure Cache for Redis 	<ul style="list-style-type: none">• Web apps built on .NET or Java• Routing real-time messages and scaling web communications frameworks

Azure SQL Database

The intelligent, relational cloud database service



- Great for:
- Multi-tenant SaaS applications
- OLTP applications
- Migrating SQL Server or Oracle workloads to the cloud



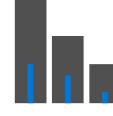
Frictionless migration

Get the broadest SQL Server engine compatibility and native VNet support with [SQL Database Managed Instance](#).



Industry-leading TCO

Realize up to a 212% ROI* through a fully-managed and intelligent service. Save up to 80% versus license included pricing with unique offers.



Enterprise scale and performance

Rapidly scale compute resources and restore databases up to 100TB in minutes, regardless of the size of data operation.



Layers of security

Control access to your database, leave sensitive data encrypted while in use, and monitor your database for potential threats.



Built-in intelligence

Continuously optimize database performance in real time with automatic tuning, machine learning and adaptive technologies.



Familiar tools

Seamlessly enable DevOps by developing in SQL Server containers and deploying in SQL Database with the [easy-to-use tools](#) you already have.

* Source: "The Total Economic Impact™ of Microsoft Azure SQL Database Managed Instance." Forrester Consulting. October 2018

Azure Database for PostgreSQL

Enterprise-ready, fully managed community PostgreSQL



- Great for:
- Secure and scalable apps
- Personalized digital marketing
- Migrating PostgreSQL or Oracle workloads to the cloud
- Vast workloads requiring high-performance horizontal scale-out



Fully managed

Leverage automatic updates, security fixes, and easy migration with Azure Database for PostgreSQL.



Best total cost of ownership

Maintain business continuity at 50% cost savings compared to AWS RDS.



Secure and compliant

Protect your data automatically with up-to-date, enterprise grade security and compliance.



Unmatched Scale

Break free from the limits of single-node Postgres & scale out across multiple nodes with Hyperscale (Citus)



Built-in intelligence

Optimize your database performance and security with customized performance Insights.



Integrated Azure services

Get your app up and running with the built-in connection to App Service and other Azure services.

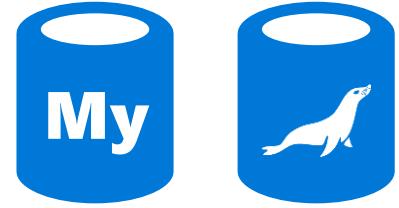
← Easy Migration →

Enterprise-ready

← Developer productivity →

Azure Database for MySQL and MariaDB

Enterprise-ready, fully managed open-source community editions



- Great for:
- Websites and web apps
- Low-latency digital gaming
- Simple, fast, read-heavy operations
- Migrating MySQL/MariaDB workloads to the cloud



Fully managed

Leverage automatic updates, security fixes, and easy migration with Azure Database for MySQL and MariaDB.



Best total cost of ownership

Maintain business continuity at 50% cost savings compared to AWS RDS.



Secure and compliant

Monitor your database around the clock and detects potential malicious activities with Advanced Threat Protection.



Familiar tools and frameworks

Easily manage and provision database for all common OSS frameworks (Drupal and WordPress) and languages.



Integrated Azure services

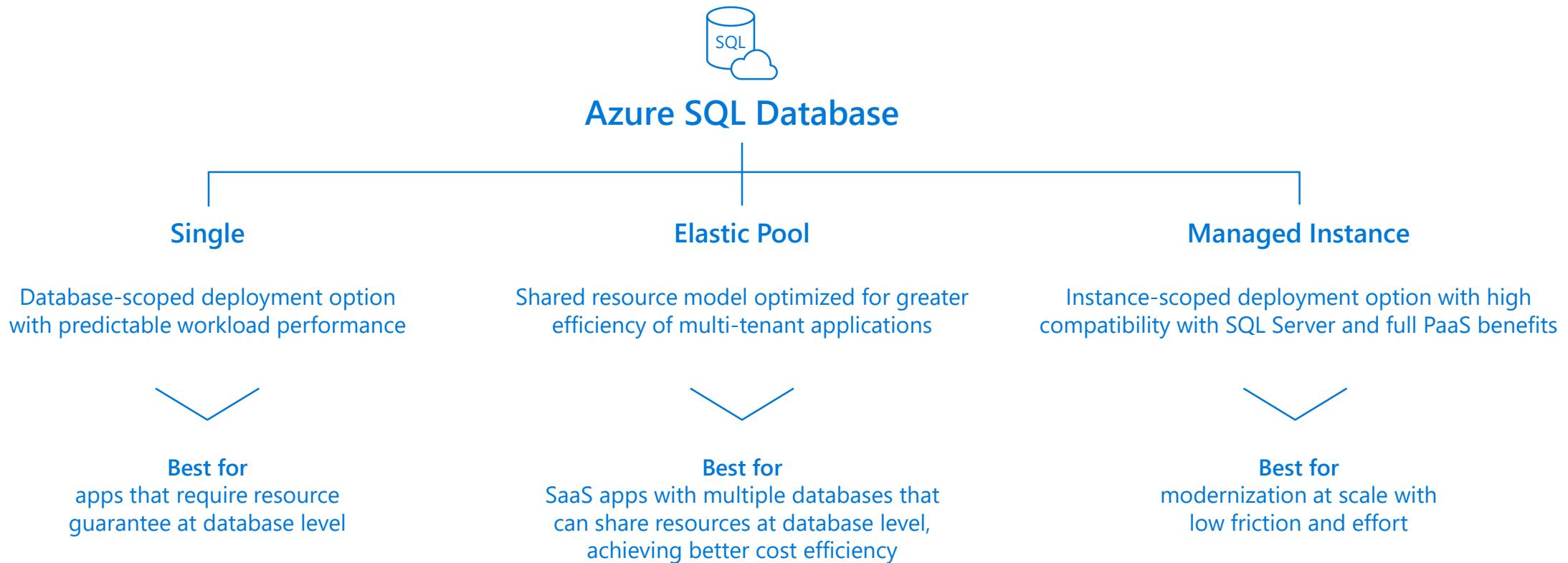
Get your app up and running with the built-in connection to App Service and other Azure services.

← Easy Migration →

← Enterprise-ready →

← Developer productivity →

AZURE SQL DATABASE DEPLOYMENT OPTIONS

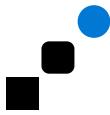


SQL Server on Virtual Machines

Easily migrate your SQL Server workloads to the cloud



- Great for:
- Migrating SQL Server or Oracle workloads to the cloud
- Apps using any SQL Server version and edition on any virtual machine size, on Windows and Linux
- Workloads requiring SQL Server Reporting Services



Seamless migration

Get the broadest SQL Server engine compatibility.



Simplify security

Benefit from automatic application of security patches on the most compliant cloud.



Improve business continuity

Improve business continuity in the event of disaster by placing your SQL Server Always On Availability Group replicas in virtual machines.



Develop and test efficiently

Spin up a new dev environment in minutes, and only pay for what you use. Only with Azure can you have pre-configured Developer edition images.



Linux and Windows

Enjoy the flexibility to run your choice of SQL Server edition in an Azure virtual machine, now on the operating system of your choice.

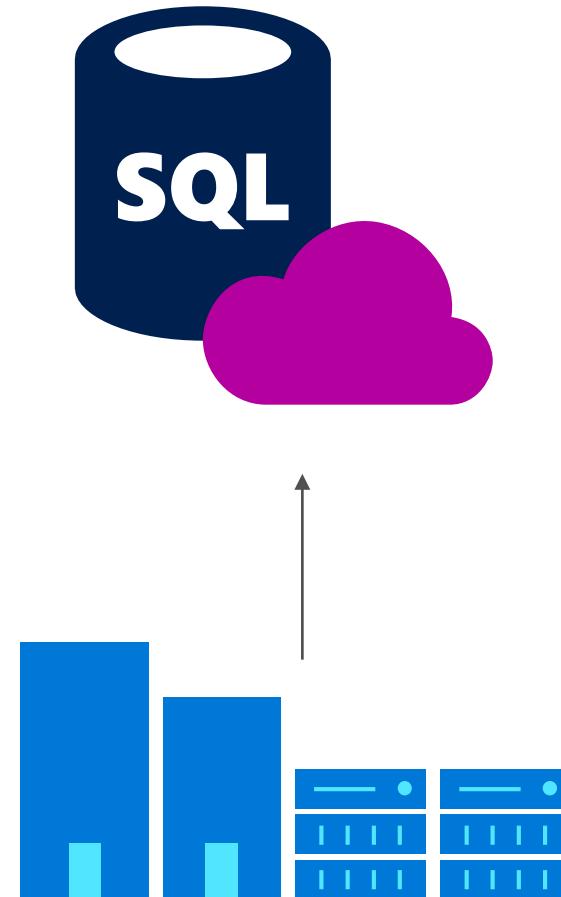


Pay less with Azure

Realize 5 times the cost savings than AWS with Azure for Windows Server and SQL Server. Why run them anywhere else?

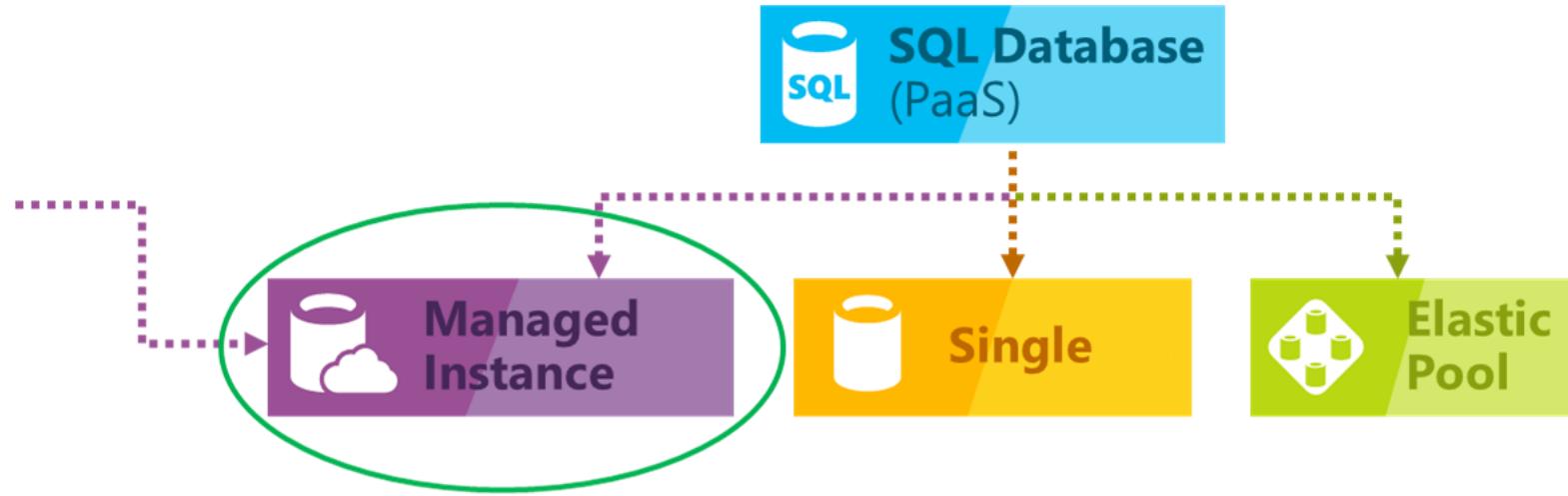
Azure SQL DB managed Instance

Customers looking to **migrate a large number of apps** from on-premise or IaaS, self-built or ISV provided, with **as low migration effort as possible & cost being a crucial factor**



What is SQL Database Managed Instance?

New deployment option that enables frictionless migration for SQL apps and modernization in a fully managed service



Easy lift and shift

- Fully-fledged SQL instance with nearly 100% compat with on-prem

Fully managed PaaS

- Built on the same PaaS service infrastructure
- All PaaS features

Full isolation and security

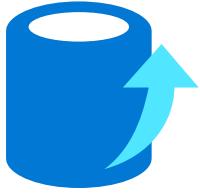
- Native VNET implementation
- Private IP addresses

New business model

- Competitive
- Transparent
- Frictionless

Hyperscale tier

for Single database, Managed Instance



Offers VLDB support without the headaches with a built-for-the-cloud architecture of highly-scalable storage & a multi-layer cache optimized for OLTP and HTAP workloads. Provides low latency and high throughput without size of data operations. Future-proofing on size and data.

Best for HTAP VLDB workloads with highly scalable storage and read-scale requirements

Key Scenarios

Migrating on-premises or cloud VLDB

Real-time analytics & reporting

Data marts running SMP on-premises

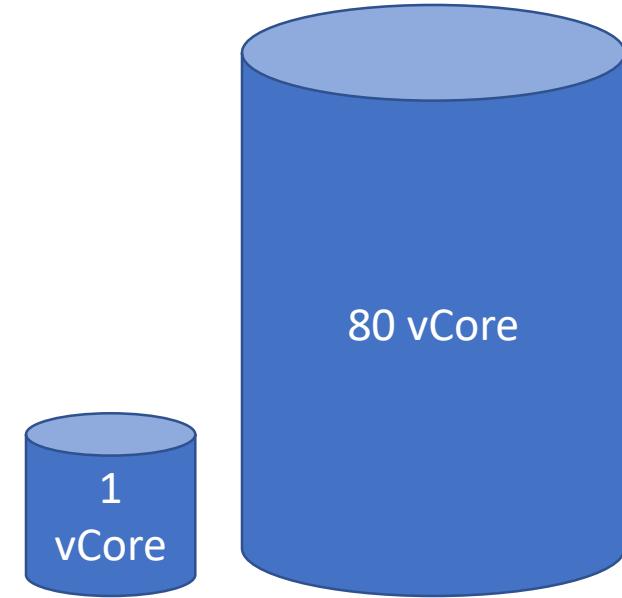
HTAP applications

Resources	
Storage	Local SSD storage for local buffer pool cache/data Azure remote storage as final long-term data store Supports up to 100TB storage
Compute (vCores)	Gen4: 1 to 24 vCore Gen5: 1 to 80 vCore
I/O	TBD
HA	2 replicas (up to 5 at GA)
In-Memory	Not supported

HyperFast & HyperElastic

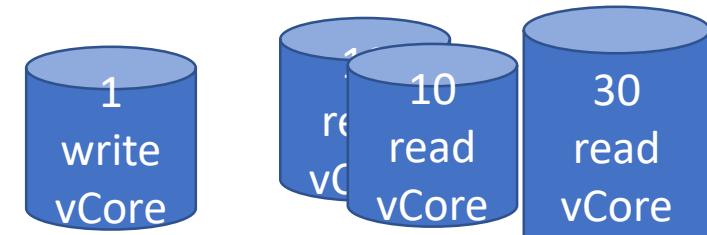
- Scale Up/Down

- Write vCores Scale-Up/Down Time: 5 minutes
 - Impact on running workload : 2s failover. No perf impact.



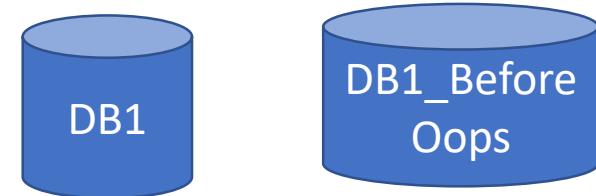
- Read-Scale Out

- Add Read vCores: 5 minutes
 - Maximum Read vCores: ∞ (Validated to $320 = 80 \times 4$)
 - Impact on running workload: 0



- Restore

- Restore to any point in time : 20 minutes
 - Impact of Backup on workload: 0



COMBING AND ANALYZING 150K ARTICLES DAILY

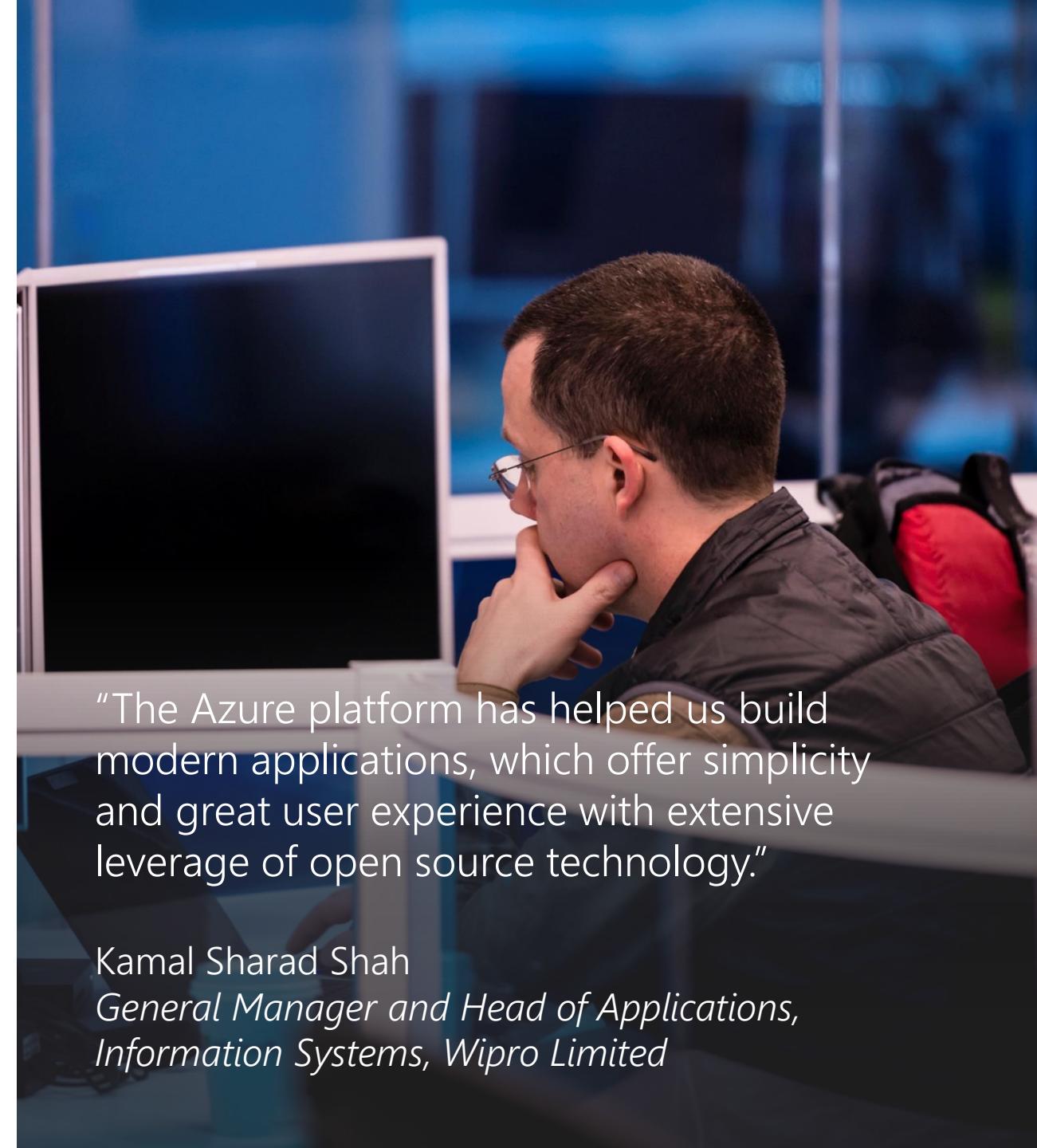
Real-time media monitoring with Azure SQL Database

Challenge

- IT company was seeking a cost-effective, resilient, and enterprise friendly IT infrastructure solution

Impact

- Developed and deployed an entire application in just 6 weeks
- Decreased lines of code by 10x with extensive use of micro services from Azure PaaS
- Estimated 60% cost savings on developing and deploying a new app



"The Azure platform has helped us build modern applications, which offer simplicity and great user experience with extensive leverage of open source technology."

Kamal Sharad Shah
*General Manager and Head of Applications,
Information Systems, Wipro Limited*

Azure SQL Data Warehouse

Fast, flexible, and secure analytics
platform for enterprise



Petabyte-scale data warehouse in the
cloud

Industry leading price/performance

Provision in seconds and scale in minutes

Scale compute and storage independently

Integrate with Azure and SQL ecosystems

Enterprise Grade Cloud Data Warehouse

On-demand

- Automated provisioning in seconds
- Re-size compute in minutes
- Create multiple instances
- Pause and Resume warehouses
- Consume compute and storage independently

Secure

- Virtual Networks
- Transparent Data Encryption
- Object level security
- Azure Active Directory
- Firewalls
- Azure Key Vault
- Auditing and Threat Detection
- Compliant

Scalable

- 4,000+ core compute scale
- Infinite data storage
- Compute scale independent of storage
- 128 active concurrent queries per cluster
- Adaptive caching
- Instant data movement

AZURE SQL DATA WAREHOUSE GEN 2

5X performance
...for the same price

5X compute scale
>4K cores in minutes

4X concurrency
>2x Redshift

∞ storage
with smart caching

Elastic, on-demand compute
with pause/resume capability

COMPUTE LAYER



Database Storage



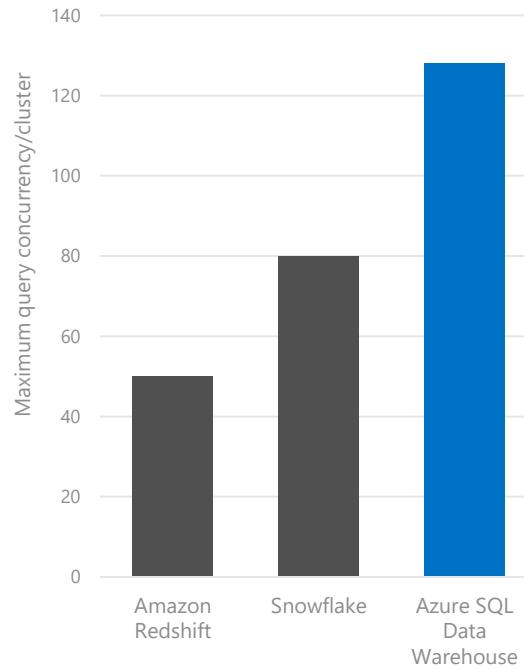
(unlimited columns)



Data Lake
Storage

Azure SQL Data Warehouse overview

Highest query concurrency
for cloud DW



Deep integration with big
data ecosystem

Databricks

Native bi-directional data
movement optimized for Azure
Databricks

Data Factory

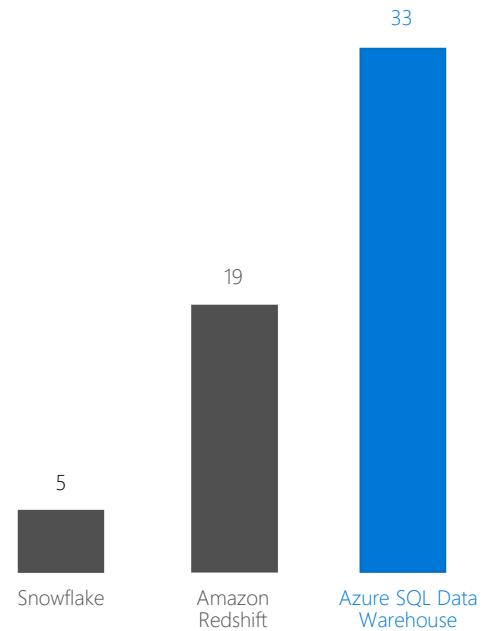
75+ data integration connectors
using Azure Data Factory

Polybase

Native SQL queries over cloud
object store using Polybase

Globally available in
the most regions

Global regions available



Industry leading
price/performance

1.6 X faster
AWS Redshift

Advanced Security

Built-in for zero additional cost

Data warehouse unit (DWU)

Measure of power



Simply buy the query performance you need, not just hardware

Transparency



Quantified by workload objectives: how fast rows are scanned, loaded, copied

On demand



First DW service to offer compute power on demand, independent of storage

100 DWU*

Scan Rate

3.36M row/sec

Loading Rate

130K row/sec

Table Copy Rate

350K row/sec

Scan 1B rows*

100 DWU

= 297 sec

400 DWU

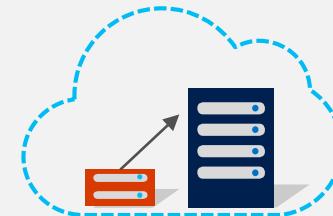
= 74 sec

800 DWU

= 37 sec

1,600 DWU

= 19 sec

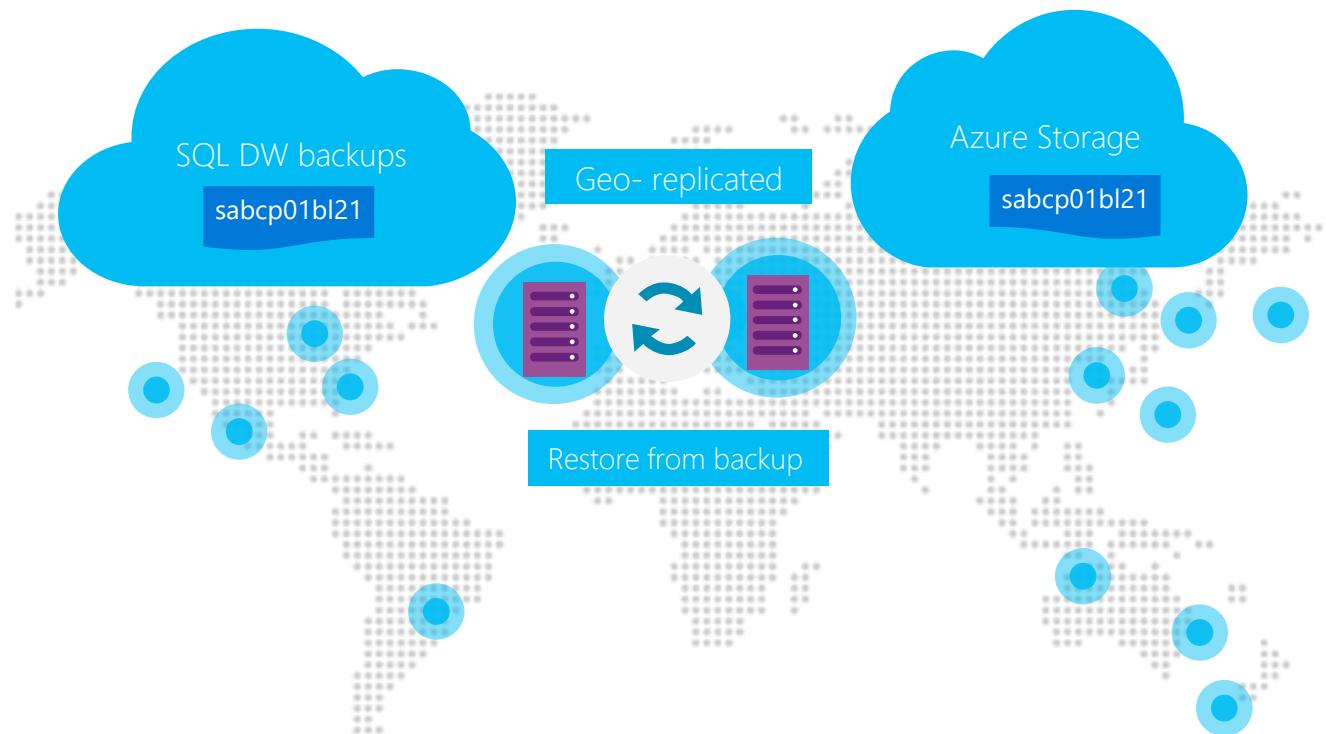


* Preliminary estimates; actual results may change

Automatic backup and geo-restore

Recover from data deletion or alteration or disaster

- Auto backups, every 4 hours
- On-demand backups in Azure Storage
- REST API, PowerShell or Azure Portal
- Scheduled exports
- Near-online backup/restore
- Backups retention policy:
 - Auto backups, up to 35 days
 - On-demand backups retained indefinitely



Data Loading Options

- Large ecosystem of powerful ETL tools
- Directly load from a variety of sources
- Transparently parallelized loads
- Guaranteed consistency and stability



BulkLoad API

Seamless loading to and from
Files/SQL SMP

SSIS

Parity with on-premises
abilities of powerful loading
suite

DWS Loader

Blazingly fast custom loader for
APS/DWS

PolyBase

Advanced data movement and
deep integration with Hadoop

Attunity

Replicate data from 1st/3rd
party storage worldwide

Informatica

Migrate advanced Informatica
packages directly to Azure

Processing invoices 24 times faster

Self-service portal with modern architecture created using Azure

Challenge

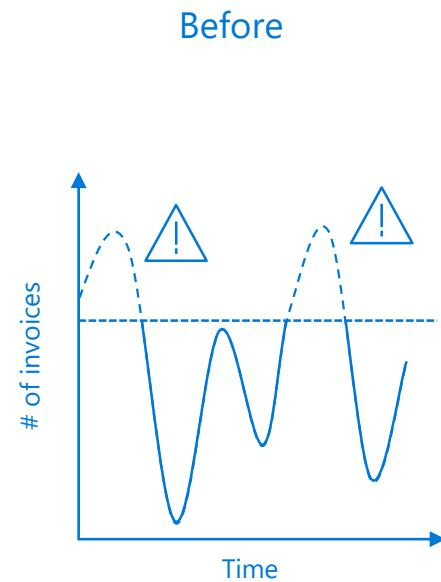
Electronic invoicing solution was unable to scale as needed to accommodate increased invoice submissions during billing cycles

Impact

Processed 2.5 million invoices in the first three months after launching

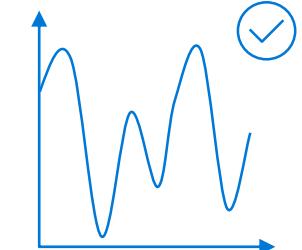
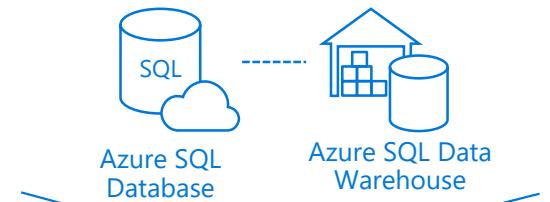
Reduced time to process 28,000 invoices from 2 days to 2 hours

Increased transaction speed to 74 validations per second



Cannot support extremes in invoices

Azure SQL Data Warehouse +
Azure SQL Database



SQL Data Warehouse Case Studies

Customer	Vertical	Location	Blurb/link
Adcorp Group	Professional Services	South Africa	Adcorp uses SQL Data Warehouse to drive advanced analytics
Adobe	Partner Professional Services	United States	Adobe uses global SQL DW for print to create, deliver, and manage better digital experiences
Agoop	Professional Services	Japan	Agoop optimizes data processing with Azure SQL Data Warehouse
Association Chilean de Seguridad	Nonprofit	Chile	Chilean nonprofit uses machine learning to process insurance claims 100 times faster
Bahrain Olympic Committee	Government	Bahrain	Bahrain prepares for digital Olympics with SQL Data Warehouse
Carnival Cruises	Hospitality & Travel	Germany	Carnival Maritime transform resource optimization with SQL Data Warehouse and machine learning
Centrica	Power and Utilities	United Kingdom	Centrica energy trading business innovated faster in the cloud
Coats	Process Mfg & Resources	United States	Coats spins its future in the cloud with SQL Data Warehouse
Coloplast	Health Provider	United States	Medical device provider builds on SQL Data Warehouse to future proof
Comune di Catania	Government	Italy	Italian municipality uses SQL Data Warehouse to help safeguard public data and service continuity
Dallas Zoo	Nonprofit	United States	The Dallas Zoo now knows what its elephants are really doing, thanks to SQL Data Warehouse
Damco	Logistics	Netherlands	Damco enables SQL Data Warehouse and analytics to transform supply chain management
Datometry	Professional Services	United States	Startup virtualizes data warehouse applications to run on SQL DW
DriveTime	Retail and Consumer Goods	United States	DriveTime better predicts loan profitability with SQL Data Warehouse solution
Civil Aviation of Kuwait	Government	Kuwait	Kuwait International Airport uses SQL Data Warehouse and Office 365 to handle growing air and passenger traffic
Evoqua Water Technologies	Discrete Manufacturing	United States	Evoqua Water is a 100-plus-year-old company that has reinvented itself for the digital age with SQL Data Warehouse
Fast Shop	Retail and Consumer Goods	Brazil	Brazilian retailer stands out from the crowd with data analytics platform on SQL Data Warehouse
FunRock	Computer Software	Sweden	FunRock takes mobile strategy games to next level with Azure SQL Data Warehouse
Integral Analytics	Power & Utilities	United States	Integral Analytics switches to Azure SQL Data Warehouse from AWS for high performance and cost-effective data warehousing
Jeju Air	Hospitality & Travel	Korea	Integrating sales data in SQL Data Warehouse allowed Jeju Air to reduce the time to create reporting template
Kansas State University	Education	United States	Kansas State sets sights on top 50 distinction with an enterprise data analytics platform on SQL Data Warehouse
LG Electronics	Discrete Manufacturing	Korea	Realizing Always-on Digital Signage Services with SQL Data Warehouse
Marico	Retail and Consumer Goods	India	Marico Gains Agility in Decision Making by Moving Data Warehouse to Cloud
Maritz Motivation Solutions	Professional Services	United States	Leading rewards program offers better insights with Azure SQL Data Warehouse
MediaBrix	Media & Cable	United States	Digital advertising company gets answers from terabytes of data with Microsoft SQL Data Warehouse
Miami-Dade Water and Sewer	Government	United States	How Miami-Dade Water gets smarter with SQL Data Warehouse
Microsoft Corporation	Discrete Manufacturing	United States	Delivering epic Xbox experiences by analyzing hundreds of billions of game events in SQL Data Warehouse
Microsoft Corporation	Professional Services	United States	By moving retail operations to SQL Data Warehouse, Microsoft Stores provides seamless customer service at 106 locations
Newell Brands	Retail and Consumer Goods	United States	Global Consumer goods company drives business value through SQL Data Warehouse
P:Cubed	Professional Services	South Africa	P:Cubed turns to Azure SQL Data Warehouse for disaster recovery
Precision Diagnostics	Healthcare	United States	Precision Diagnostics used SQL Data Warehouse to aggregate behavioral health
Presence Orb	Discrete Manufacturing	United Kingdom	Wi-Fi analytics firm gains real-time benefits with Azure SQL Data Warehouse
Reckitt Benckiser	Retail and Consumer Goods	United Kingdom	Reckitt Benckiser empowers sales teams with SQL Data Warehouse
Rubikloud	Retail and Consumer Goods	Canada	Rubikloud helps retailers improve customer experience with AI as a Service
Sejong	Discrete Manufacturing	South Korea	Auto parts specialist innovates with Azure SQL Data Warehouse
Sierra Pacific Mortgage	Banking & Capital Markets	United States	Mortgage banker unlocks data with cloud-based analytics service, gains insight, and boosts advantage
SnelStart	Discrete Manufacturing	Netherlands	SnelStart rapidly expanded its business services with SQL Data Warehouse
Targetbase	Professional Services	United States	Targetbase transforms customer engagement with SQL Data Warehouse
The Entertainer Middle East	Media & Telecommunications	United Arab Emirates	UAE consumer offers leader uses SQL Data Warehouse to improve their customers' experience
Thomson Reuters	Media & Telecommunications	United States	Thomson Reuters builds highly extensible and scalable system for customers with SQL Data Warehouse
Toshiba	Professional Services	United States	Toshiba uses IOT data services on SQL Data Warehouse to keep customers up and running
Track Revenue	Discrete Manufacturing	United States	Track Revenue moved from Amazon to Azure SQL Data Warehouse and boosted customer revenue-per-click by 38 percent

Machine Learning on Azure

Domain specific pretrained models

To simplify solution development



Vision



Speech



Language



Search

Familiar Data Science tools

To simplify model development



Visual Studio Code



Azure Notebooks



Jupyter



Command line

Popular frameworks

To build advanced deep learning solutions



PyTorch



TensorFlow



Scikit-Learn



ONNX

Productive services

To empower data science and development teams



Azure
Databricks



Azure Machine
Learning



Machine
Learning VMs

Powerful infrastructure

To accelerate deep learning



CPU



GPU



FPGA

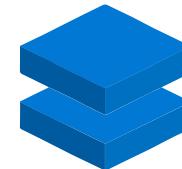


From the Intelligent Cloud to the Intelligent Edge



Productive Services

To empower data science and development teams



Azure Machine Learning

Python-based machine learning service

Develop models faster with automated machine learning

Use any Python environment and ML frameworks

Manage models across the cloud and the edge.

Azure Databricks

Apache Spark-based big-data service

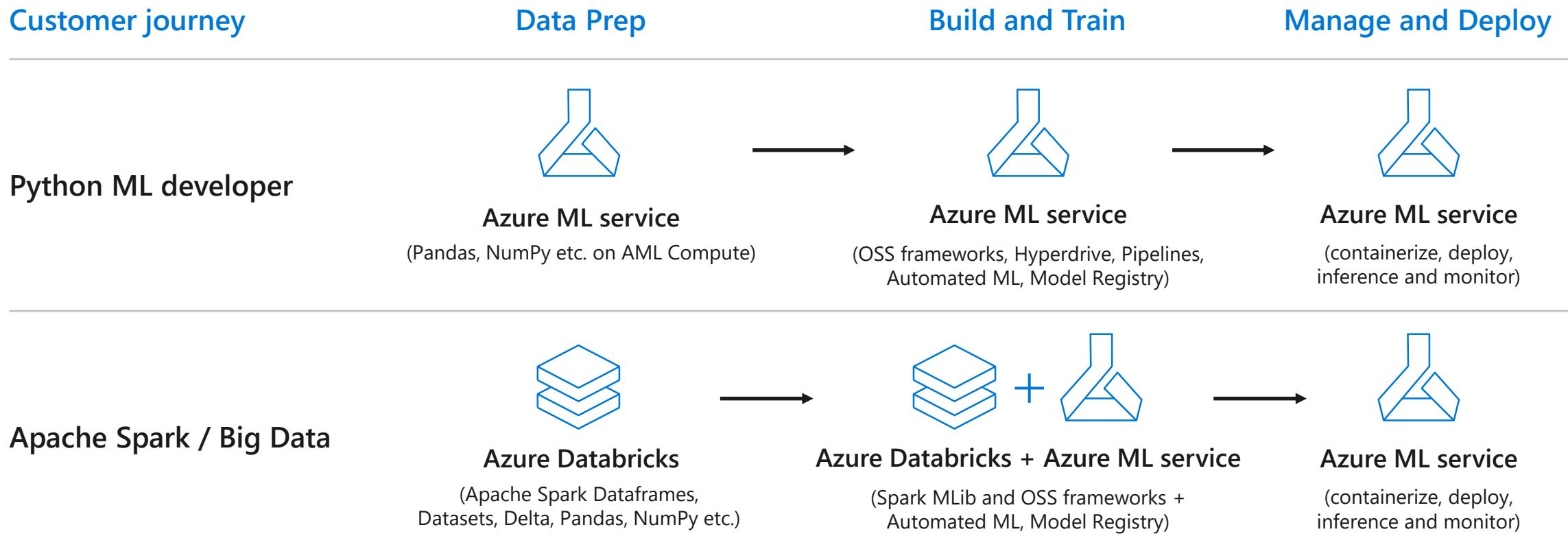
Prepare data clean data at massive scale

Enable collaboration between data scientists and data engineers

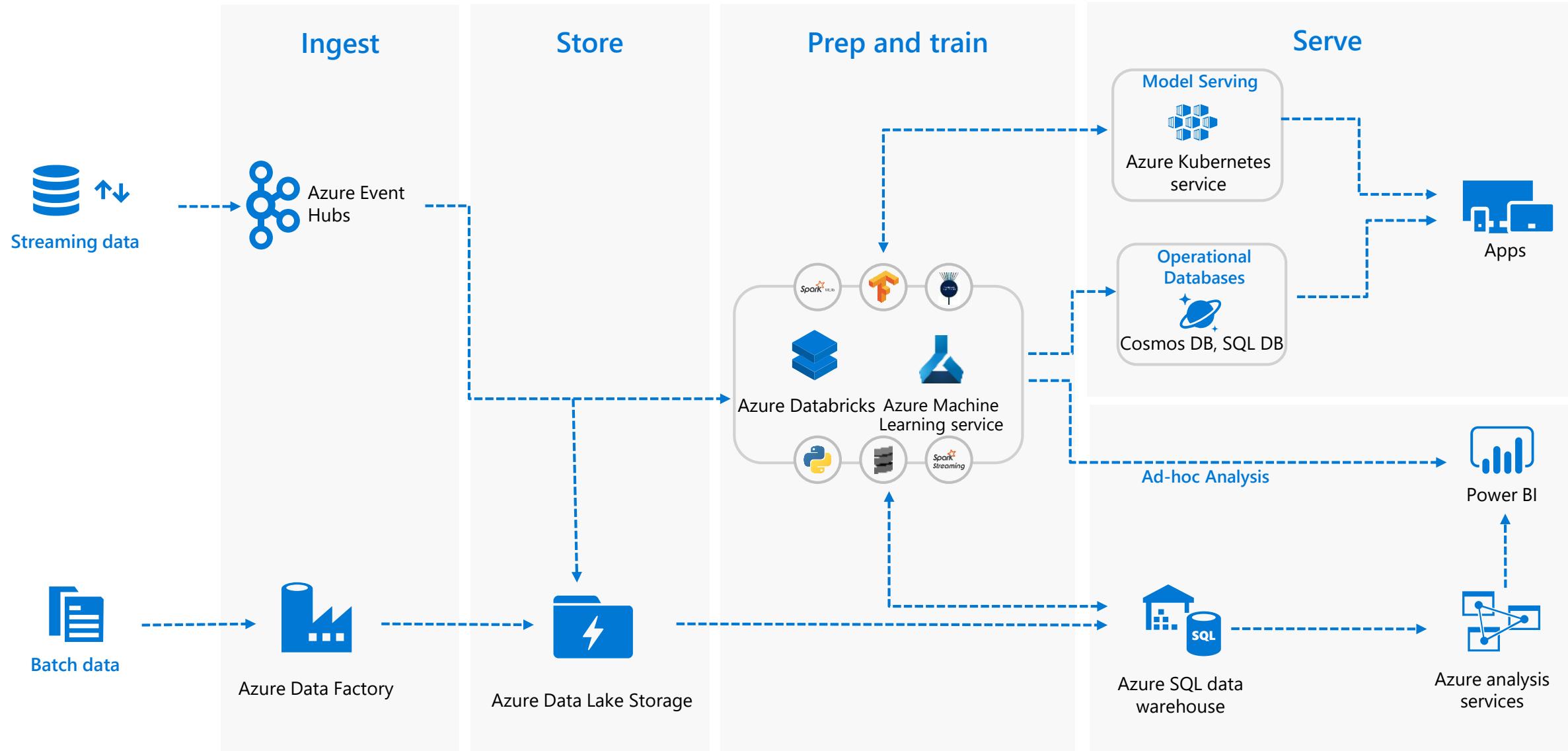
Access machine learning optimized clusters

Productive Services

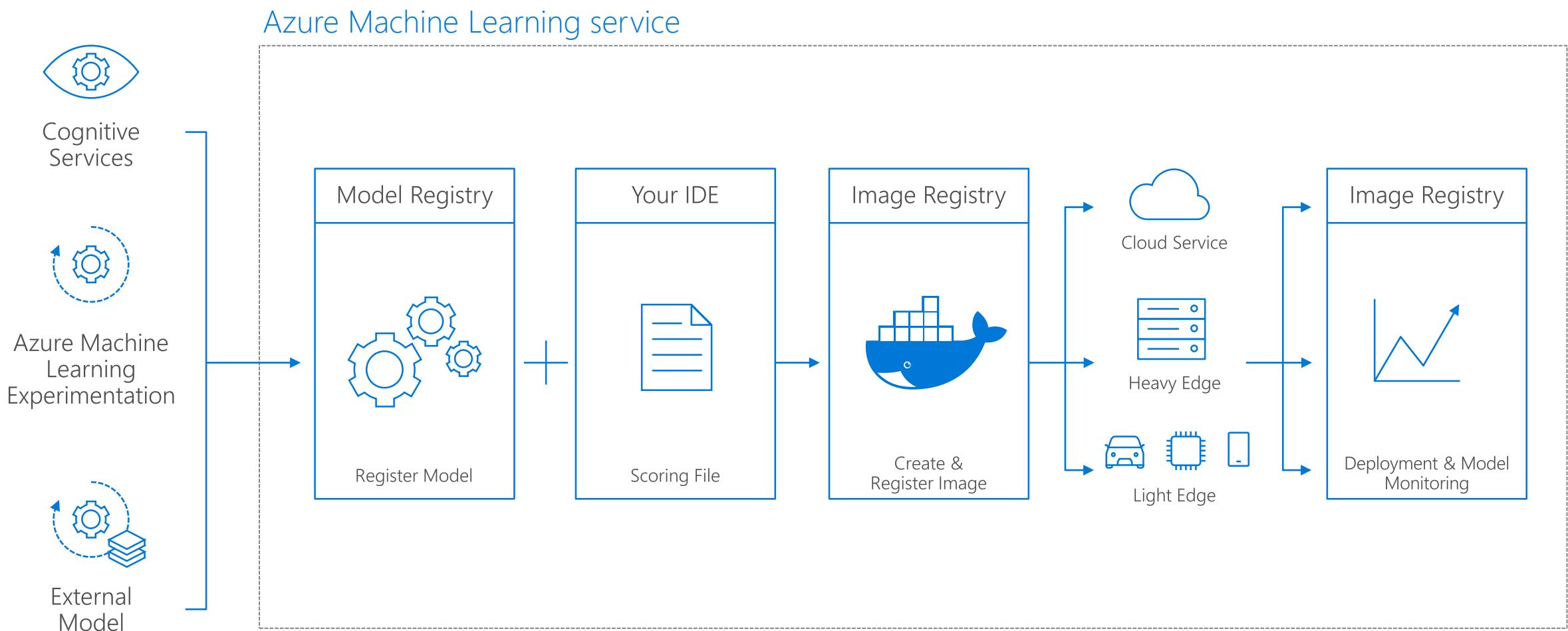
What to use when?



Recommended architecture to build e2e ML solutions



Deploy Azure ML models at scale





Drone-based electric grid inspector powered by deep learning

Challenge

- Traditional power line inspection services are costly
- Demand for low cost image scoring and support for multiple concurrent customers
- Needed powerful AI to execute on a drone solution

Solution

- Deep learning to analyze multiple streaming data feeds
- Azure GPUs support Single Shot multibox detectors
- Reliable, consistent, and highly elastic scalability with Azure Batch Shipyards



Increased match accuracy with image analysis

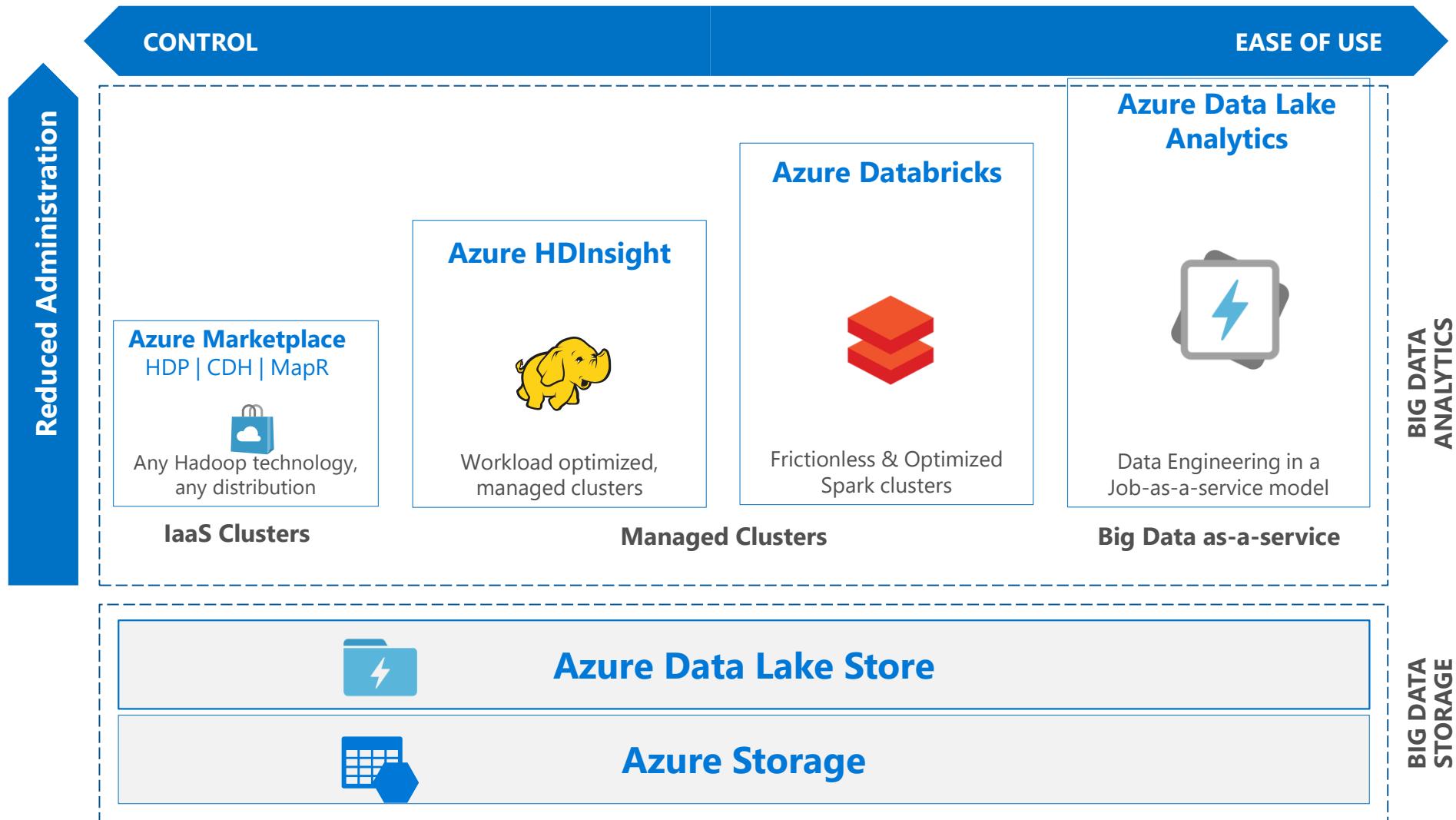
Challenge

- Assist buyer search by providing accurate options of similar clothing items from catalogue
- Need for improved smart-image matching capability based on color, pattern, neck style, etc.

Solution

- Training data created using Bing and domain-specific images
- Leveraged pre-trained ImageNet deep neural network
- Accurate list of most similar clothing items using similarity metrics for apparel
- Match accuracy of 74%

KNOWING THE VARIOUS BIG DATA SOLUTIONS



Microsoft + Hortonworks

Promoting Open Hadoop

Engineering alignment
Corporate alignment
Field alignment



Azure HDInsight (99.9% SLA on each technology)



Quick create Custom (size, settings, apps)**1** Basics
Configure basic settings**2** Storage
Set storage settings**3** Summary
Confirm configurations

This cluster may take up to 20 minutes to create.

Basics

* Cluster name
testHDInsight001.azurehdinsight.net

* Subscription
BDHadoopTeamPMTTestDemo

* Cluster type ⓘ
Configure required settings

* Cluster login username ⓘ
admin

* Cluster login password ⓘ

Secure Shell (SSH) username ⓘ
sshuser

Use same password as cluster login ⓘ

* Resource group
 Create new Use existing

* Location
East US 2

Cluster configuration

i Learn about HDInsight and cluster versions. →

Cluster configuration

* Cluster type ⓘ

- Hadoop
- HBase
- Storm
- Spark
- R Server
- Kafka
- Interactive Query

* Operating system

 Linux Windows

* Version



Quick create Custom (size, settings, apps)

1 Basics
Configure basic settings >

2 Storage
Set storage settings >

3 Summary
Confirm configurations >



This cluster may take up to 20 minutes to create.

* Cluster name
testHDInsight001 .azurehdinsight.net

* Subscription
BDHadoopTeamPMTTestDemo

* Cluster type ⓘ
Configure required settings >

* Cluster login username ⓘ
admin

* Cluster login password ⓘ

Secure Shell (SSH) username ⓘ
sshuser

Use same password as cluster login ⓘ

* Resource group
 Create new Use existing

* Location
East US 2

Click here to view cores usage.

Learn about HDInsight and cluster versions. →

Cluster configuration

* Cluster type ⓘ
Interactive Query

* Operating system
Linux

* Version
Interactive Query 2.1.0... ▾

Enterprise Security Package ⓘ **Adds 0.02 USD per Core-Hour.**

Interactive Query : Build Enterprise Data Warehouse with in-memory analytics using Hive (SQL on Hadoop) and LLAP (Long Live and Process). If you are planning to use other Hadoop components please create a separate Hadoop cluster. Note that this feature requires high memory instances.

Features

* denotes preview feature

+ Apache Ranger* (Enterprise Security Package)

+ Domain joining* (Enterprise Security Package)

+ Secure shell (SSH) access

+ HDInsight applications

+ Custom virtual network

+ Custom Hive metastore

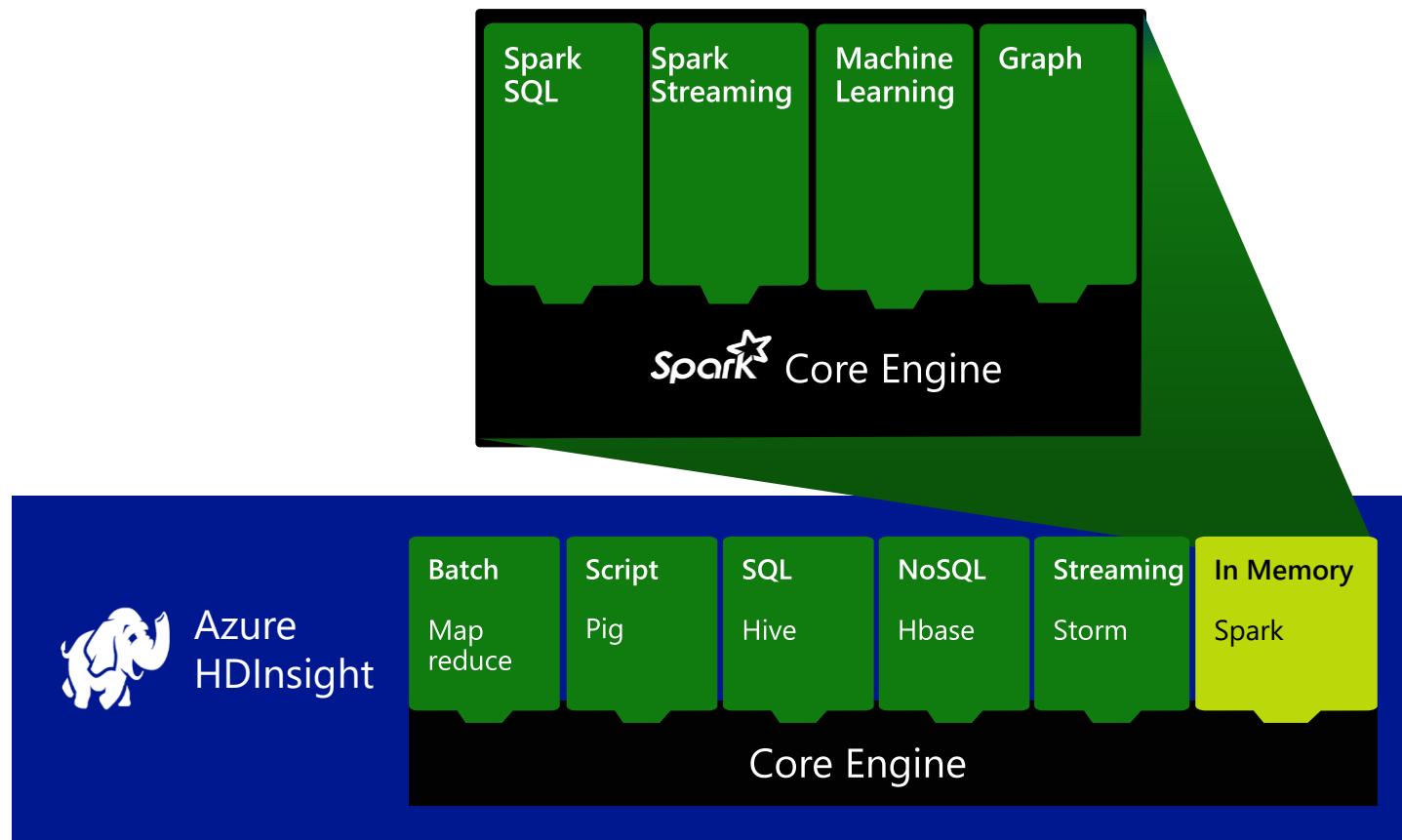
+ Custom Oozie metastore

+ Data Lake Store access

+ Data Lake Store as primary data storage

Spark for Azure HDInsight

In Memory Processing on Multiple Workloads



- Single execution model for multiple tasks
- Processing up to 100x faster performance
- Developer friendly (Java, Python, Scala)
- BI tool of choice (Power BI, Tableau, Qlik, SAP)
- Notebook experience (Jupyter/iPython, Zeppelin)



HDInsight Supports Hive

SQL-like queries on Hadoop data in HDInsight

HDInsight provides easy-to-use graphical query interface for Hive

HiveQL is a SQL-like language (subset of SQL)

Hive structures include well-understood database concepts such as tables, rows, columns, partitions

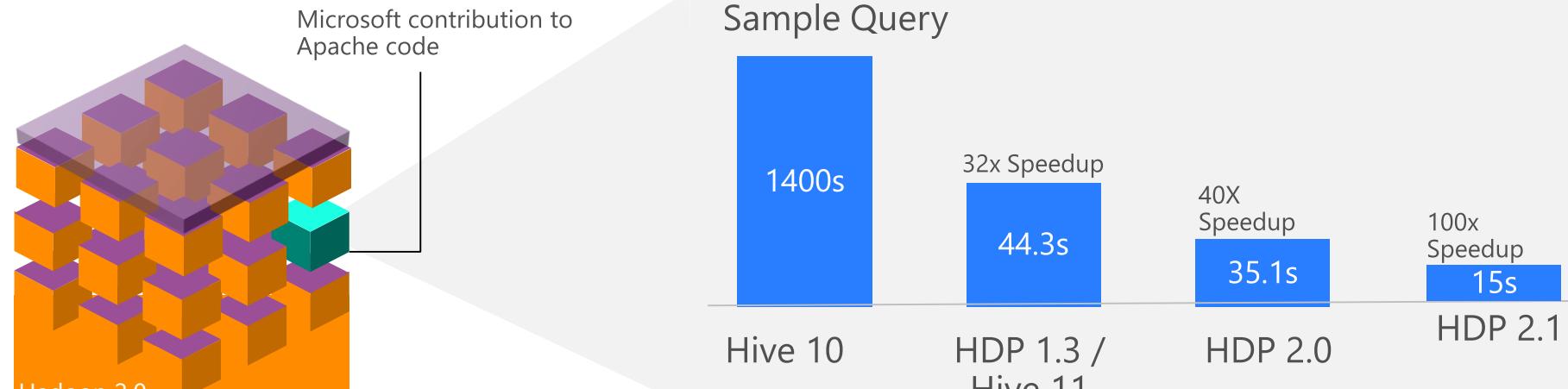
Compiled into MapReduce jobs that are executed on Hadoop

Dramatic performance gains with Stinger/Tez

Stinger is a Microsoft, Hortonworks and OSS driven initiative to bring interactive queries with Hive

Brings query execution engine technology from Microsoft SQL Server to Hive

Performance gains up to 100x



HDInsight Supports Storm



Stream analytics for Near-Real Time Processing

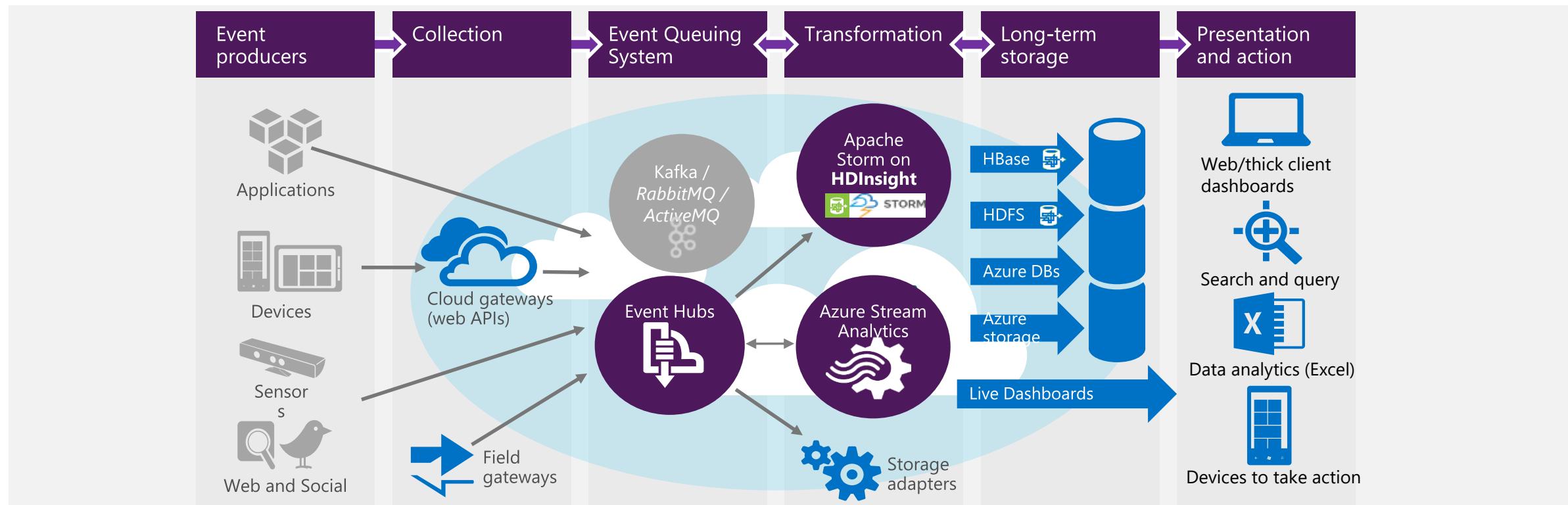
Consumes millions of real-time events from a scalable event broker (ie. Apache Kafka, Azure Event Hub)

Performs time-sensitive computation

Output to persistent stores, dashboards or devices

Customizable with Java + .NET

Deeply integrated to Visual Studio



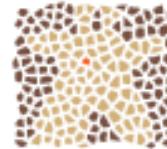
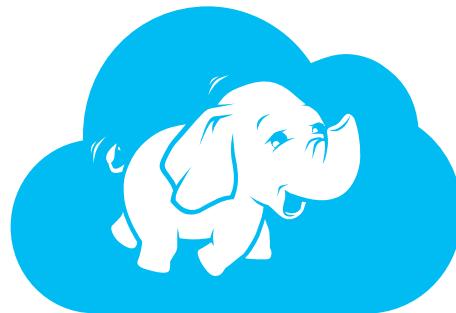
HDInsight Allows You To Add Hadoop Projects

Add Hadoop Projects to HDInsight

Modify HDInsight clusters with custom script

Add Apache Hadoop projects to HDInsight

Documented for Spark, R, Giraph, Solr



...

HDInsight Monitoring options



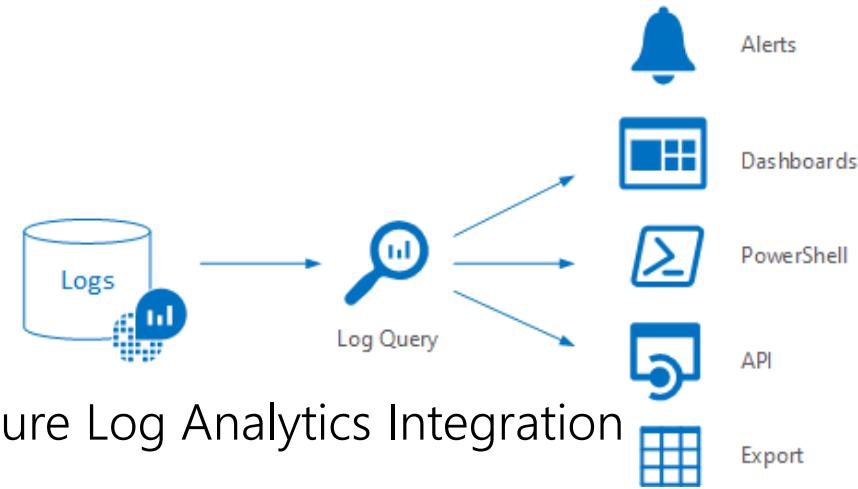
Apache Ambari

- View cluster metrics like CPU, memory, and disk usage at a glance in real time
- Identify malfunctioning components with Ambari alerts
- Monitor queue capacities, jobs, and view associated OSS logs



HDInsight Cluster Metrics

- See gateway requests to monitor cluster stress and cluster size to monitor costs
- Apply filters and chart splitting to extract important data
- Set up alert rules to receive notifications and trigger actions for key metrics



Azure Log Analytics Integration

- Organizes cluster metrics and OSS log records into queryable tables
- Create custom dashboards to surface all the metrics you need from multiple clusters on a single pane of glass

Looking for a big data solution



Customer example: Virginia Tech crunch endless amounts of Genomic data

Scenario

DNA sequencers are generating 15PB of genomic data each year. Virginia Tech needed to process it to foster medical breakthroughs including new cancer treatments. They were evaluating creating a multimillion dollar supercomputer center, but wanted to find a different way to process the data.

Solution

Azure HDInsight (Hadoop-as-a-service) was chosen to process genome data resulting in significant cost savings as they only pay for what they need.

Result

- Significant cost savings with the cloud
- Elastic scale that keeps up with huge data volumes
- Powering the search for cancer treatments



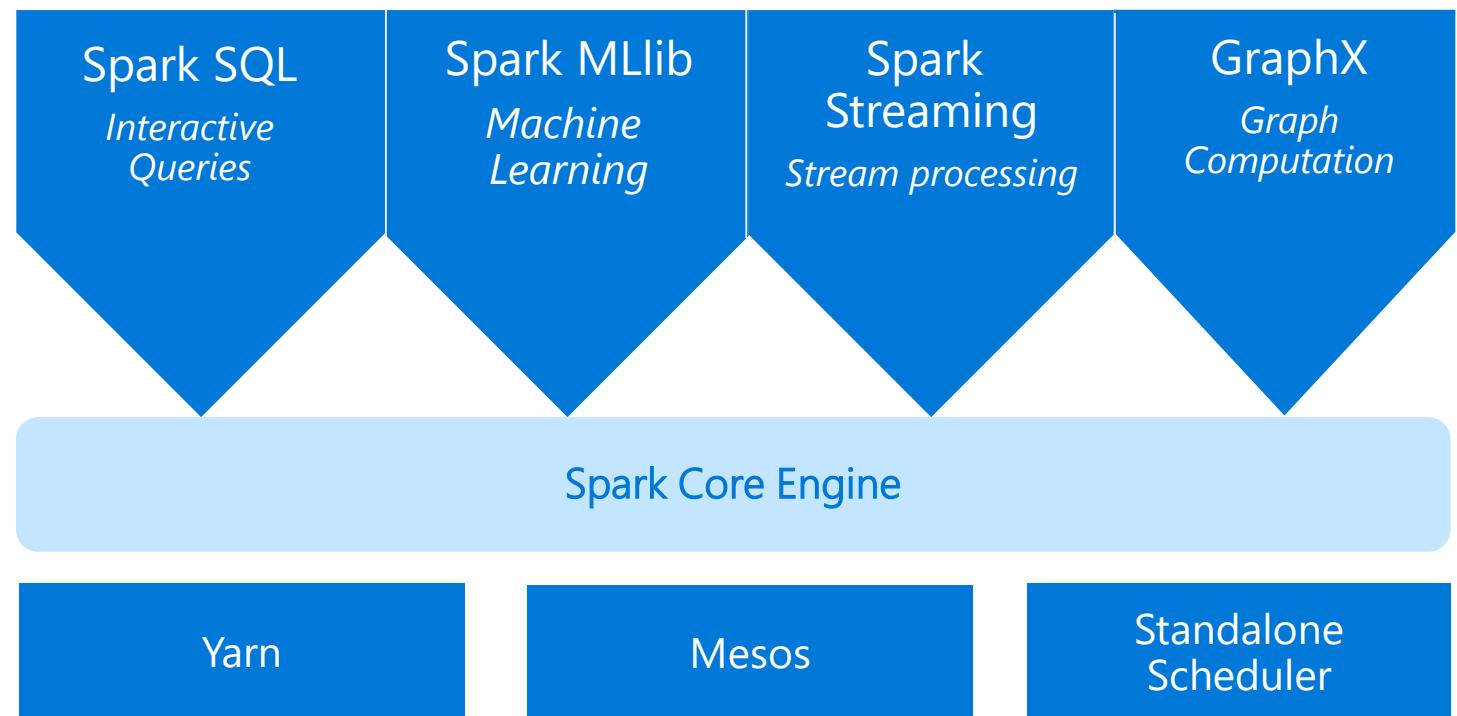
 VirginiaTech
1872

A P A C H E S P A R K

An unified, open source, parallel, data processing framework for Big Data Analytics

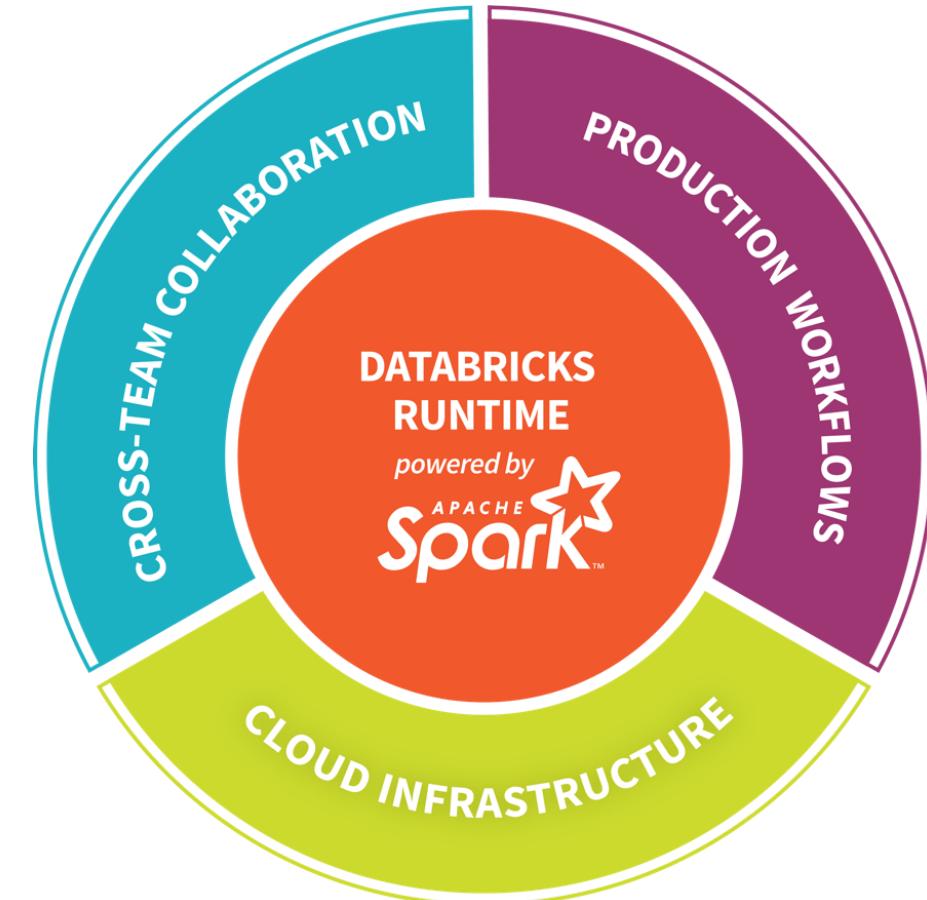
Spark Unifies:

- Batch Processing
- Interactive SQL
- Real-time processing
- Machine Learning
- Deep Learning
- Graph Processing



DATABRICKS - COMPANY OVERVIEW

- Founded in late 2013
- By the creators of Apache Spark, original team from UC Berkeley AMPLab
- Largest code contributor code to Apache Spark
- Level 2/3 support partnership with
 - Hortonworks
 - MapR
 - DataStax
- Provides [certifications](#) such as Databricks Certified Application, Databricks Certified Distribution and Databricks Certified Developer
- Main Product: The [Unified Analytics Platform](#)
- In Oct 2017, introduced [Databricks Delta](#) (currently in private preview).



PROVISIONING AZURE DATABRICKS WORKSPACE

- Azure Databricks is provisioned directly from the Azure Portal like any other Azure service
 - In contrast, with other clouds, it has to be provisioned through the Databricks portal.
 - With Azure Databricks, the Azure Portal offers a unified portal to provision and administer Azure Databricks as well as other Azure services.
- Any Azure user with the appropriate subscription and authorization can provision Azure Databricks service*.
 - There is no need for a separate Databricks account

The image shows two screenshots of the Microsoft Azure Portal. The top screenshot is a modal window titled 'Azure Databricks Service' under 'Azure Databricks (preview)'. It contains fields for 'Workspace name' (set to 'mytestworkspace'), 'Subscription' (set to 'Azure conversion - External'), 'Resource group' (radio button selected for 'Create new', set to 'mytestresgroup'), and 'Location' (set to 'East US 2'). The bottom screenshot shows the 'mytestworkspace' resource group in the Azure Portal. The left sidebar lists various Azure services. The main pane shows the 'Overview' tab for the Databricks service, which includes details like 'Managed Resource Group' (databricks-rg-mytestworkspace-va64qm...), 'Subscription' (Azure conversion - External), 'Subscription ID' (15c5cb6e-191a-40ea-9f69-08207a17fe97), and a large red 'Initialize Workspace' button. To the right of the workspace details, there are four cards: 'Documentations', 'Getting Started', 'Import Data from File', and 'Import Data from Azure Storage'. A caption on the left side of the bottom screenshot reads 'After provisioning the is complete'.

Provisioning the Azure Databricks Service

After provisioning the is complete

* During the current preview phase, the subscription has to be whitelisted.

WORKSPACES

Workspaces enables users to organize—and share—their Notebooks, Libraries and Dashboards

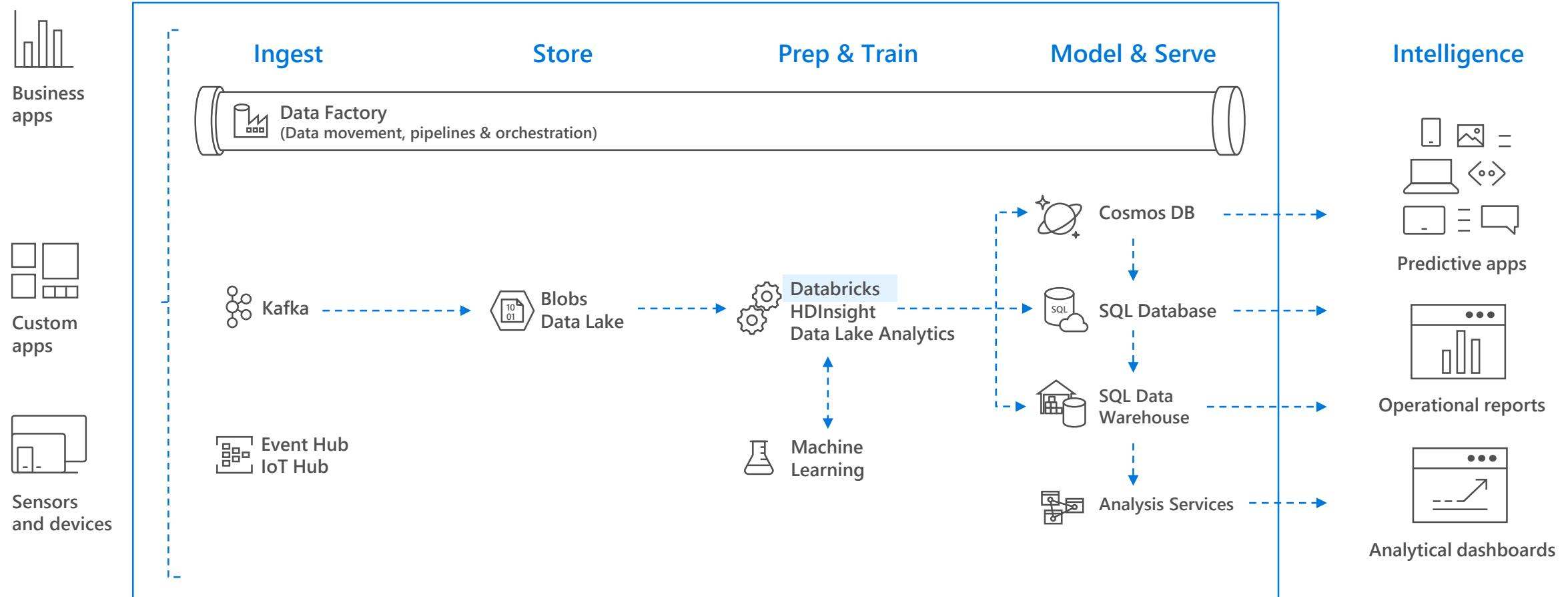
- Workspaces—sort of like Directories—are a convenient way to organize an user's Notebook, Libraries and Dashboards.
- Everything in a workspace is organized into hierarchical folders. Folders can hold Libraries, Notebooks, Dashboard or more (sub) folders.
 - Icons indicate the type of the object contained in a folder
- Every user has one directory that is private and unshared.
 - By default, the workspace and all its contents are available to users.
- Fine grained access control can be defined on workspaces to enable *secure collaboration with colleagues*.

The screenshot shows the Microsoft Azure workspace interface. On the left is a sidebar with icons for Azure Databricks, Home, Workspace (selected), Recent, Data, and Clusters. The main area has a header 'Microsoft Azure' and 'Workspace' with a dropdown set to 'MyTestFolder'. Below is a list of items:

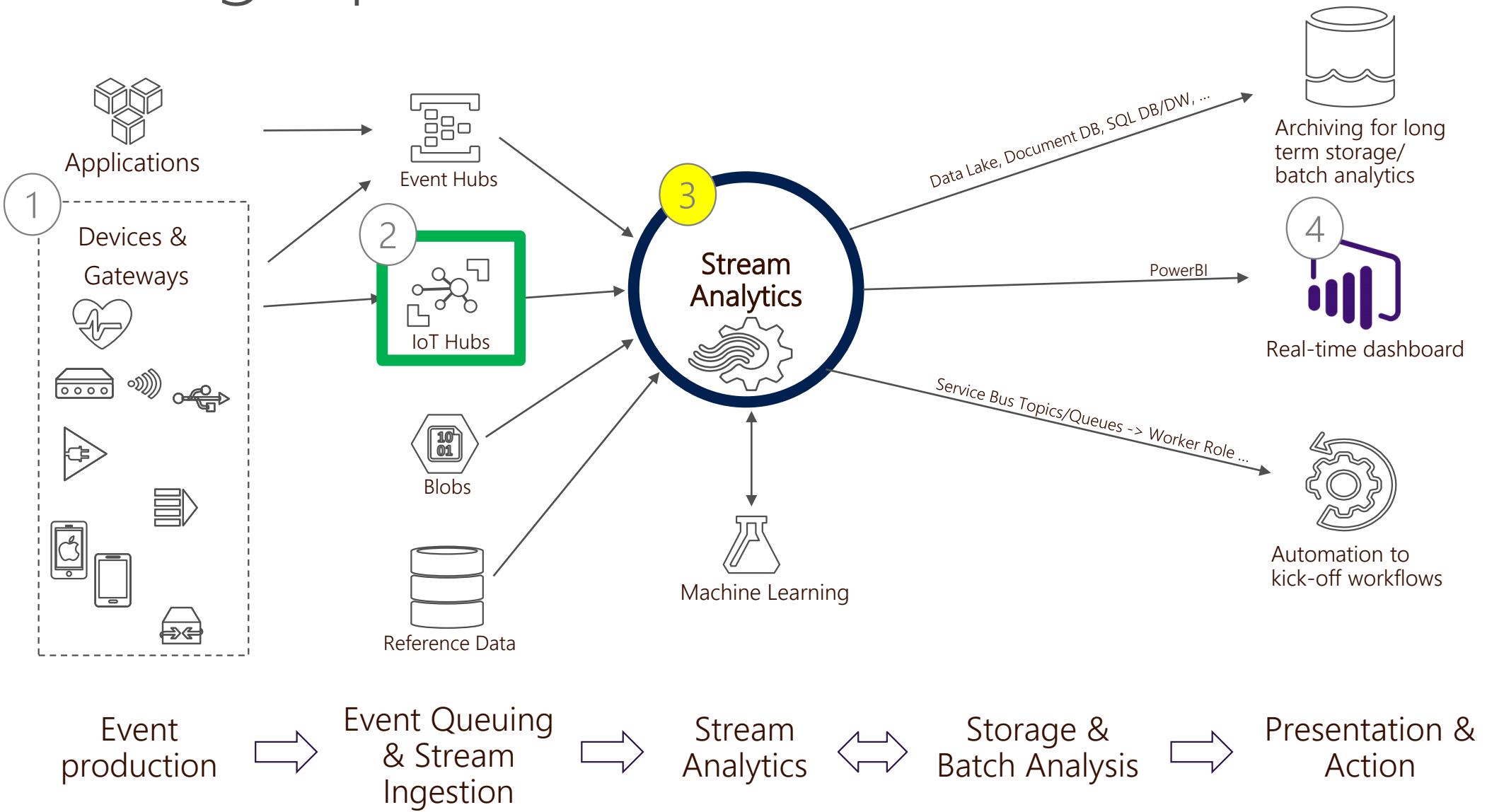
- Documentation
- Release Notes
- Training & Tutorials
- Shared
- Users
- ConfigureKafkaAccess
- ConfigureKafkaAccessNotebook
- InstallCNTK
- InstallCNTKOld
- InstallODBC
- ModelEvaluationNotebook
- MyTestFolder (selected)
- StreamingEvaluation

The screenshot shows the Microsoft Azure workspace interface with a context menu open over the 'Workspace' header. The menu includes options: Create, Import, Export, and Permissions.

BIG DATA & ADVANCED ANALYTICS AT A GLANCE



Streaming Pipeline

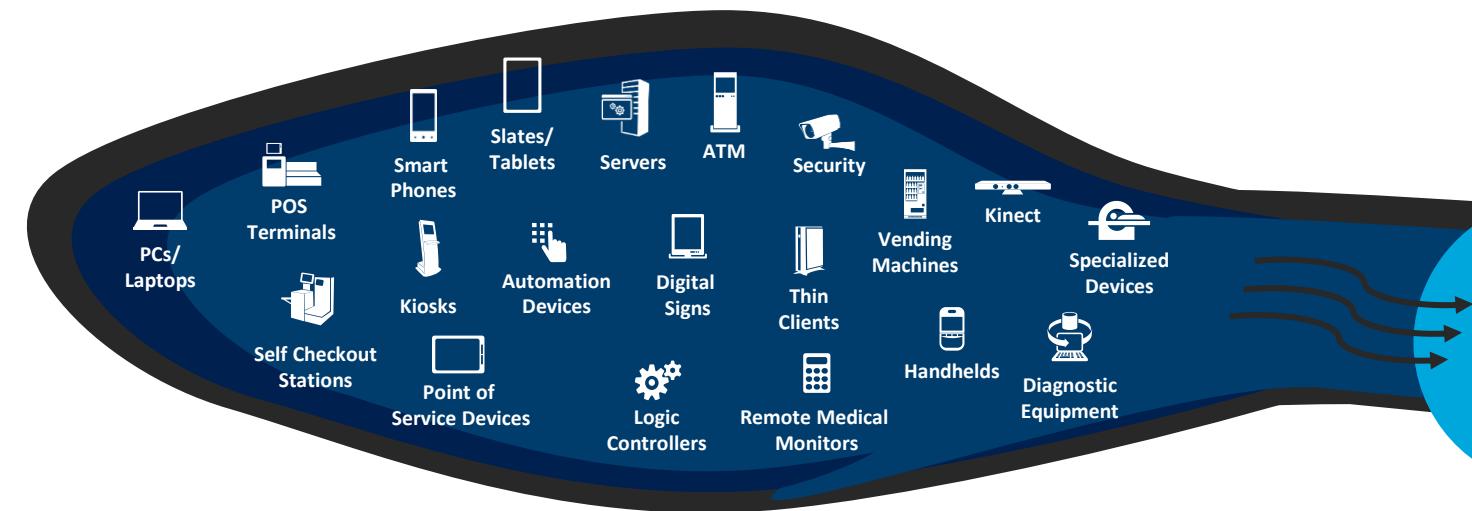


Azure Stream Analytics

Consumes millions of real-time row data, messages, events from Event Hub collected from devices, sensors, infrastructure, and applications

Performs time-sensitive analysis using SQL-like language against multiple real-time streams and reference data

Outputs to persistent stores, dashboards or back to devices



lightsensor PREVIEW

DASHBOARD MONITOR INPUTS **QUERY** OUTPUT SCALE CONFIGURE

query

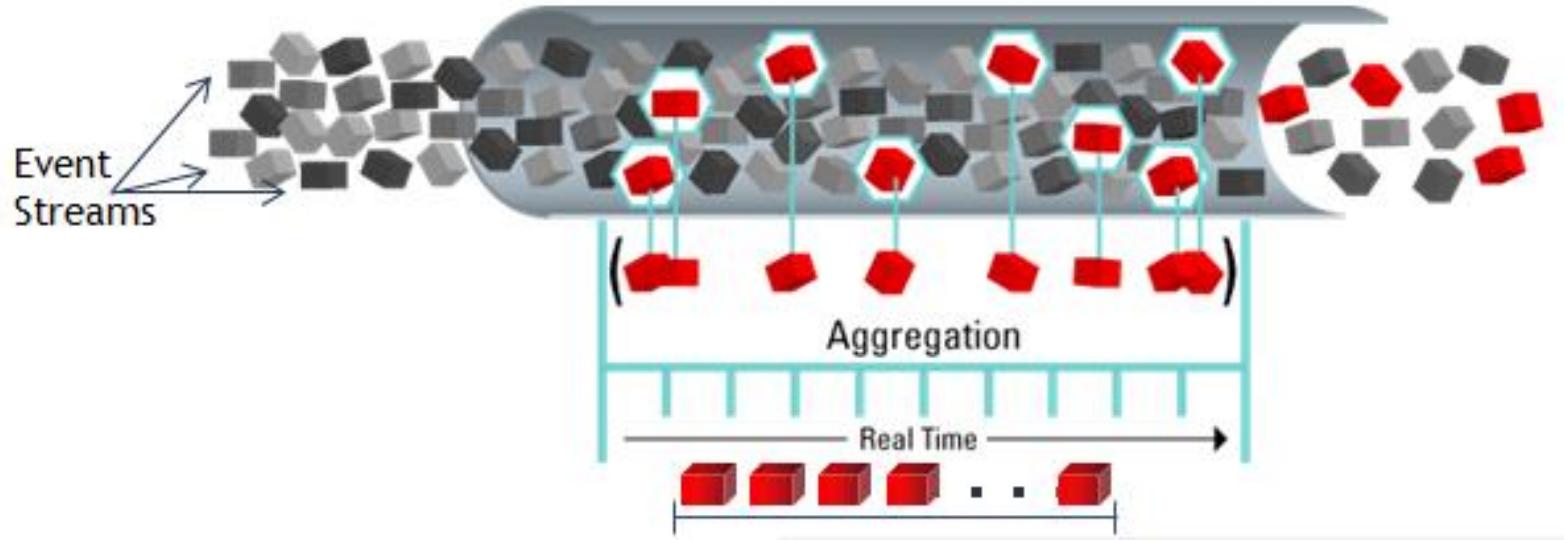
Query can't be edited while a job is running.

```
1 SELECT
2   'LightSensor' AS alerttype,
3   dspl AS dsplalert,
4   'The Light is turned OFF' AS message,
5   MAX(time) AS timestamp
6 FROM
7   StreamInput TIMESTAMP BY time
8 GROUP BY
9   HoppingWindow(DURATION(s, 5), HOP(s, 5)),
10  dspl
11 HAVING
12  AVG(light) < 0.02
13  AND COUNT(*) >= 2
14
15
```

Stream Analytics

A screenshot of the Azure Stream Analytics web interface. At the top, there's a navigation bar with tabs: DASHBOARD, MONITOR, INPUTS, **QUERY**, OUTPUT, SCALE, and CONFIGURE. The **QUERY** tab is currently selected. Below the navigation is a section titled "query" with a note: "Query can't be edited while a job is running." Underneath is a code editor containing a SQL-like query. To the right of the code editor is a monitor showing a map with data points. At the bottom, there's a toolbar with icons for copy, cut, paste, and other operations.

Real-time Stream Processing



https://en.wikipedia.org/wiki/Event-driven_architecture

Simple Event Processing

- Filter
- Transform
- Enrich
- Split
- Route

Event Stream Processing

[Simple event processing] +
Aggregate
Rules

Complex Event Processing

[Event Stream Processing] +
Pattern detection
Time windows
Joins & correlations

Stream Analytics Query Language (SAQL)

Declarative SQL like language to describe transformations

Filters ("Where")

Projections ("Select")

Time-window and property-based aggregates

("Group By")

Time-shifted joins (specifying time bounds within which the joining events must occur)

and all combinations thereof

Data Manipulation

SELECT
FROM
WHERE
HAVING
GROUP BY
CASE WHEN THEN
ELSE
INNER/LEFT OUTER JOIN
UNION
CROSS/OUTER APPLY
CAST INTO
ORDER BY ASC, DSC

Aggregation

SUM
COUNT
AVG
MIN
MAX
STDEV
STDEVP
VAR
VARP
TopOne

Date and Time

DateName
DatePart Day, Month, Year
DateDiff
DateTimeFromParts
DateAdd

String

Len
Concat
CharIndex
Substring
Lower, Upper
PatIndex

Temporal

Lag
IsFirst
Last
CollectTop

Mathematical

ABS
CEILING
EXP
FLOOR
POWER
SIGN
SQUARE
SQRT

Windowing Extensions

TumblingWindow
HoppingWindow
SlidingWindow

Geospatial (preview)

CreatePoint
CreatePolygon
CreateLineString
ST_DISTANCE
ST_WITHIN
ST_OVERLAPS
ST_INTERSECTS

Scaling Extensions

WITH
PARTITION BY
OVER

Stream Analytics Job

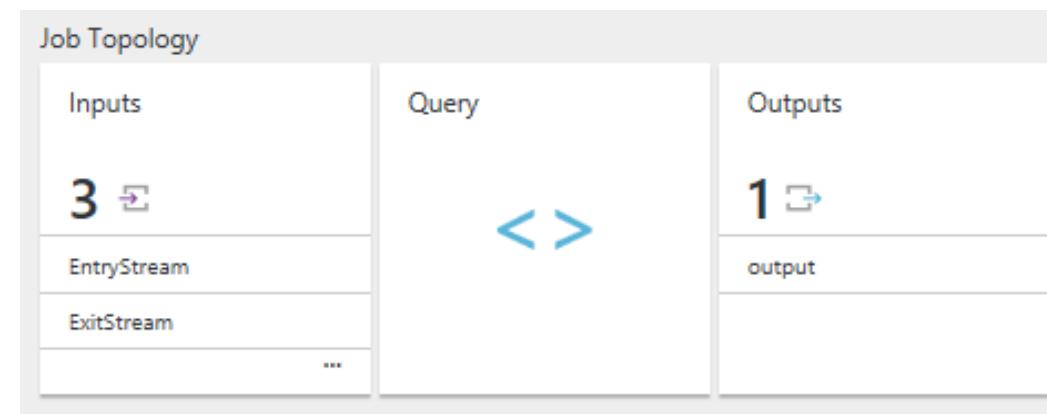
Users construct and deploy jobs to ASA

Job definition includes inputs, a query, and output

Inputs are from where the job reads the data stream

Query runs for perpetuity unless explicitly stopped and transforms the input stream

Output is where the job sends the job results to

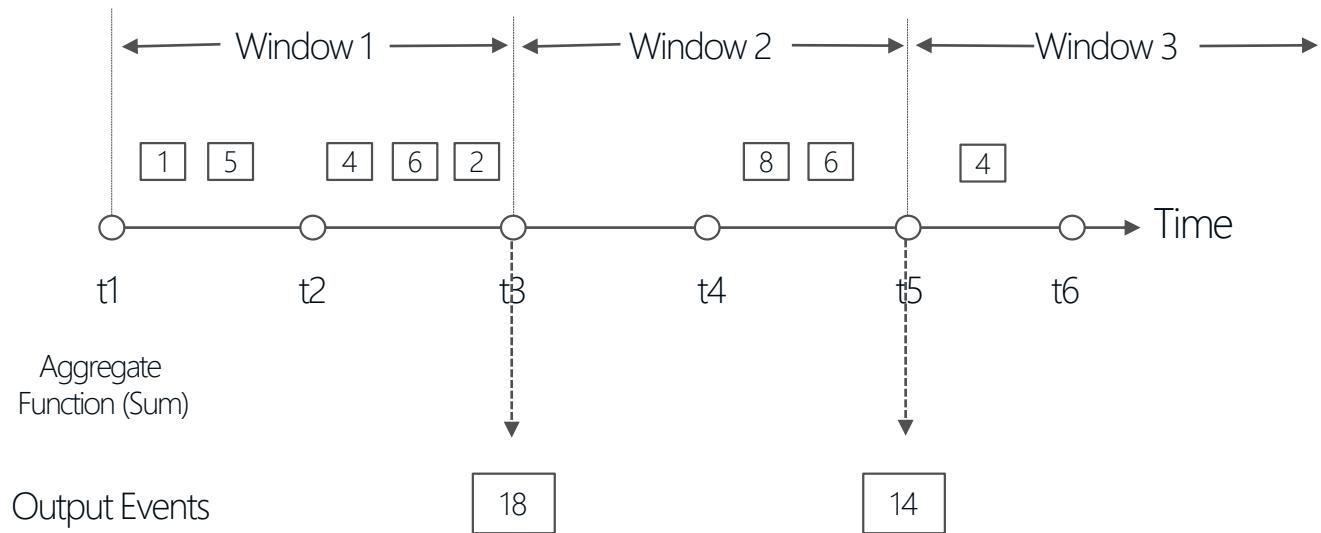


Windowing Concepts

Output at the end of each window

Windows are fixed length

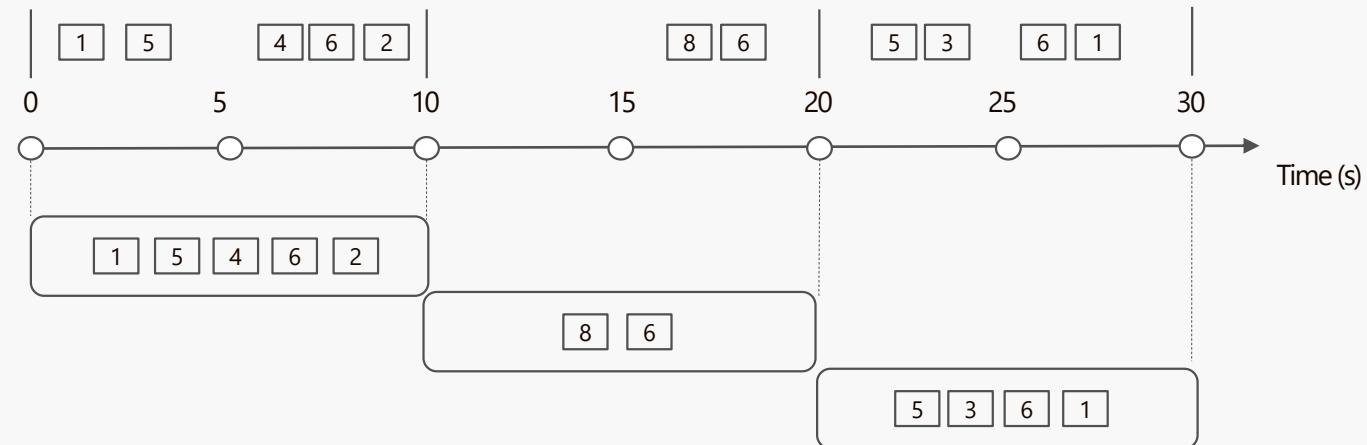
Used in a GROUP BY clause



Tumbling Windows

Every 10 seconds give me the count of vehicles entering each toll booth over the last 10 seconds

A 10-second Tumbling Window

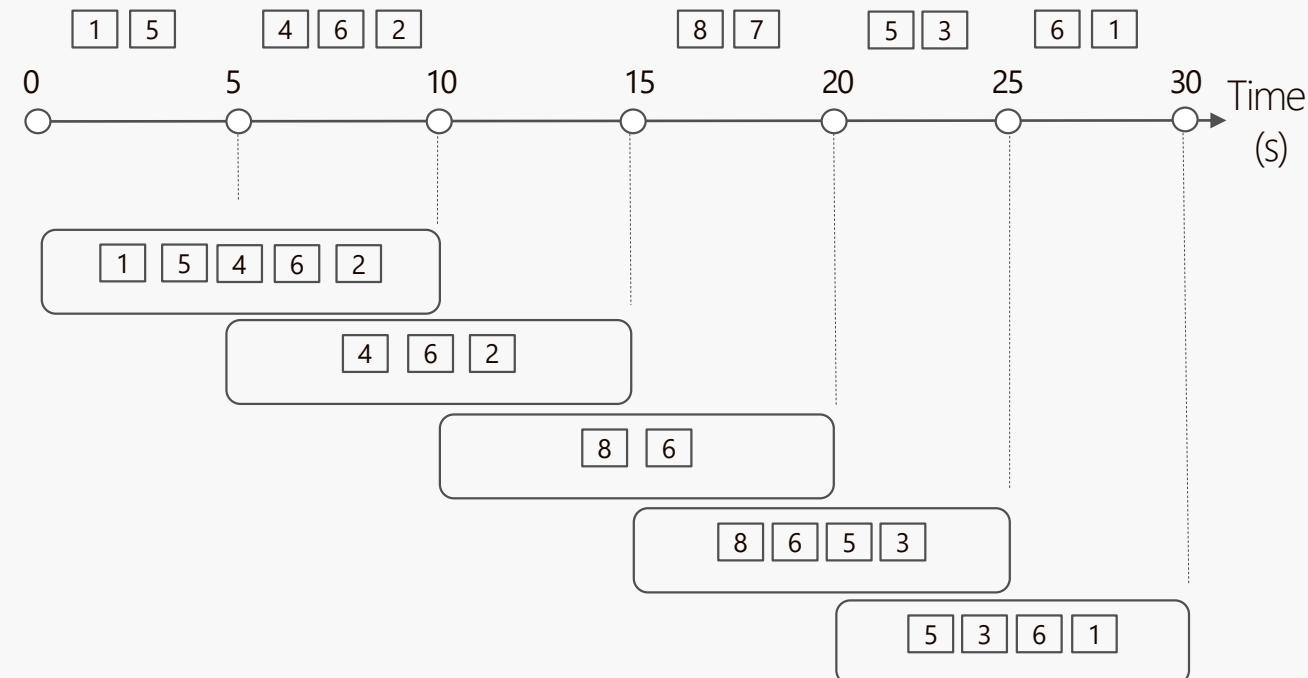


```
SELECT TollId, Count(*)  
FROM EntryStream TIMESTAMP BY EntryTime  
GROUP BY TollId, TumblingWindow(second, 10)
```

Hopping Windows

Every 5 seconds give me the count of vehicles entering each toll booth over the last 10 seconds

A 10 second Hopping Window with a 5 second hop



```
SELECT TollId, Count(*)
```

```
FROM EntryStream TIMESTAMP BY EntryTime
```

```
GROUP BY TollId, HoppingWindow(second, 10, 5)
```

Sliding Windows

Find all toll booths that have
served more than 10 vehicles in
the last 20 seconds

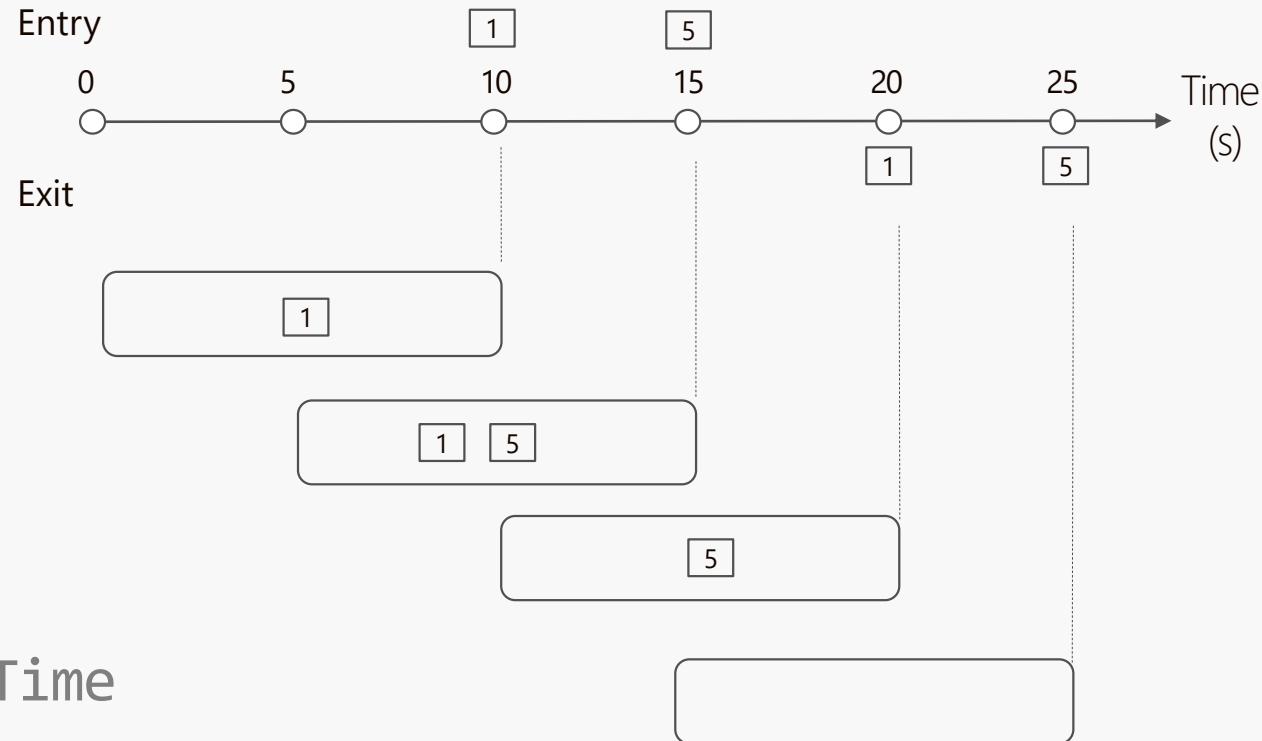
```
SELECT TollId, Count(*)
```

```
FROM EntryStream TIMESTAMP BY EntryTime
```

```
GROUP BY TollId, SlidingWindow(second, 20)
```

```
HAVING Count(*) > 10
```

A 10-second Sliding Window



An output is generated whenever an event either enters/leaves the system

Streaming Units

Represents the computing resources footprint
of a Azure Stream Analytics job

Seamlessly add/remove Streaming Units



Edge Analytics

Local Execution

Stream analytics runs on 'edge devices'

Unlock the Value of Untapped data

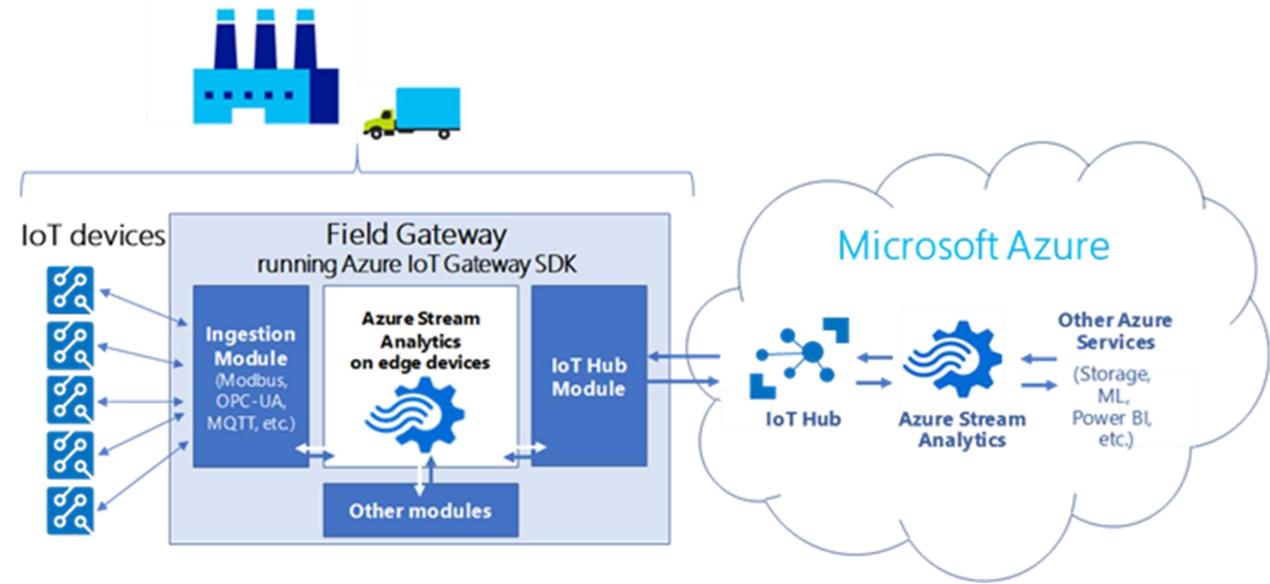
Only ~5% of data in industrial processes is sent to the cloud
Deploy intelligence near the data to unlock the full value of data

Seamless development and operations

Stream analytics jobs run in the cloud and on edge devices

Intelligent actions

Deploy situational awareness, custom code, ML models on the edge



Gaming studio learns what players want and delivers quickly with the cloud

"Our Azure-based analytics pipeline can handle the billion events a month we're throwing at it, and with it, we're improving our games at an even faster cadence."

Oliver Löffler: Founder and Chief Technology Officer



Products and Services
Microsoft Azure
Azure App Service
Azure Data Lake Store
Azure Event Hubs
Azure Stream Analytics

Organization Size
50 Employees

Industry
Media and telecommunications

Country
Germany



Real time analytics enable better business outcomes



A photograph showing a maintenance worker in safety gear sitting on the nacelle of a white wind turbine. He is working on equipment connected by cables. In the background, several other wind turbines are visible on a hillside under a clear sky.

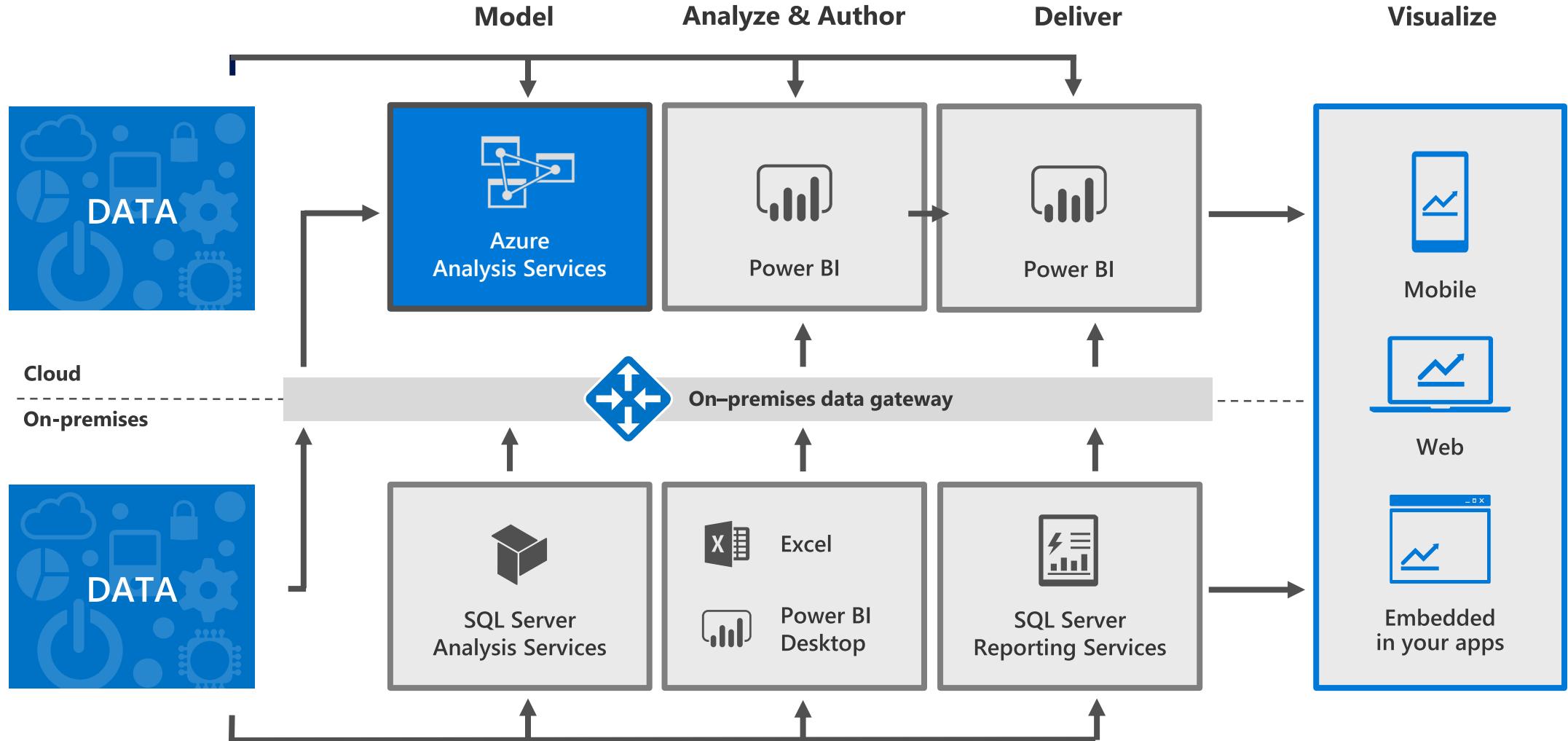
"Plant equipment heat and vibration readings are passed along to asset management teams to ensure our equipment is being maintained correctly. Production output can be tracked and provided to our regulator to ensure compliance, and our commercial teams use this telemetry for billing purposes"

- Kent Weare, Lead Architect

Benefits ▪ Lower equipment failures and downtime ▪ Secure infrastructure ▪ Lower operational costs

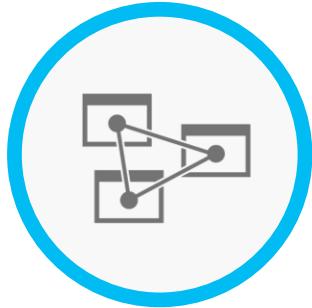
Comprehensive Microsoft BI offerings

Hybrid architecture



Azure Analysis Services

Enterprise grade analytics engine as a service



Build rich semantic models

Transform complex data into business user friendly semantic models



Gain insights at the speed of thought

Gain instant insights with in-memory cache using your preferred visualization tools



Proven technology

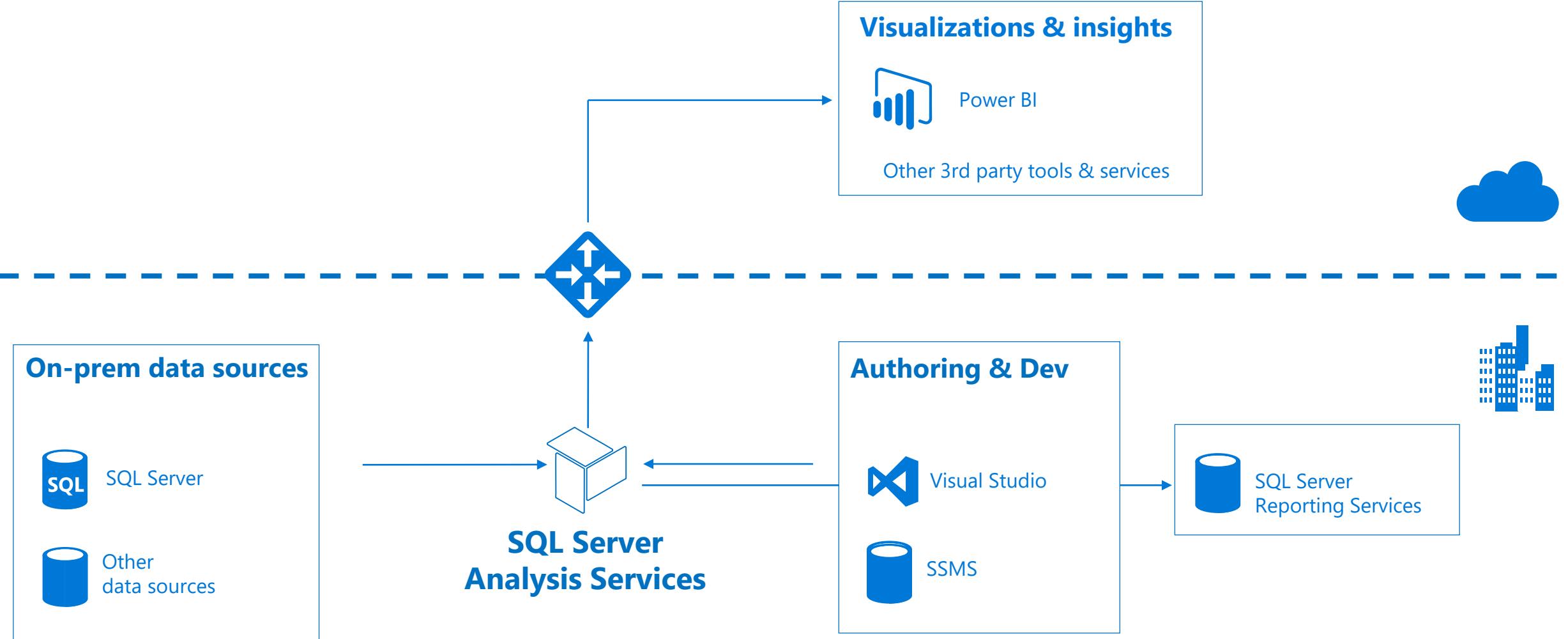
Based on powerful, proven SQL Server Analysis Services



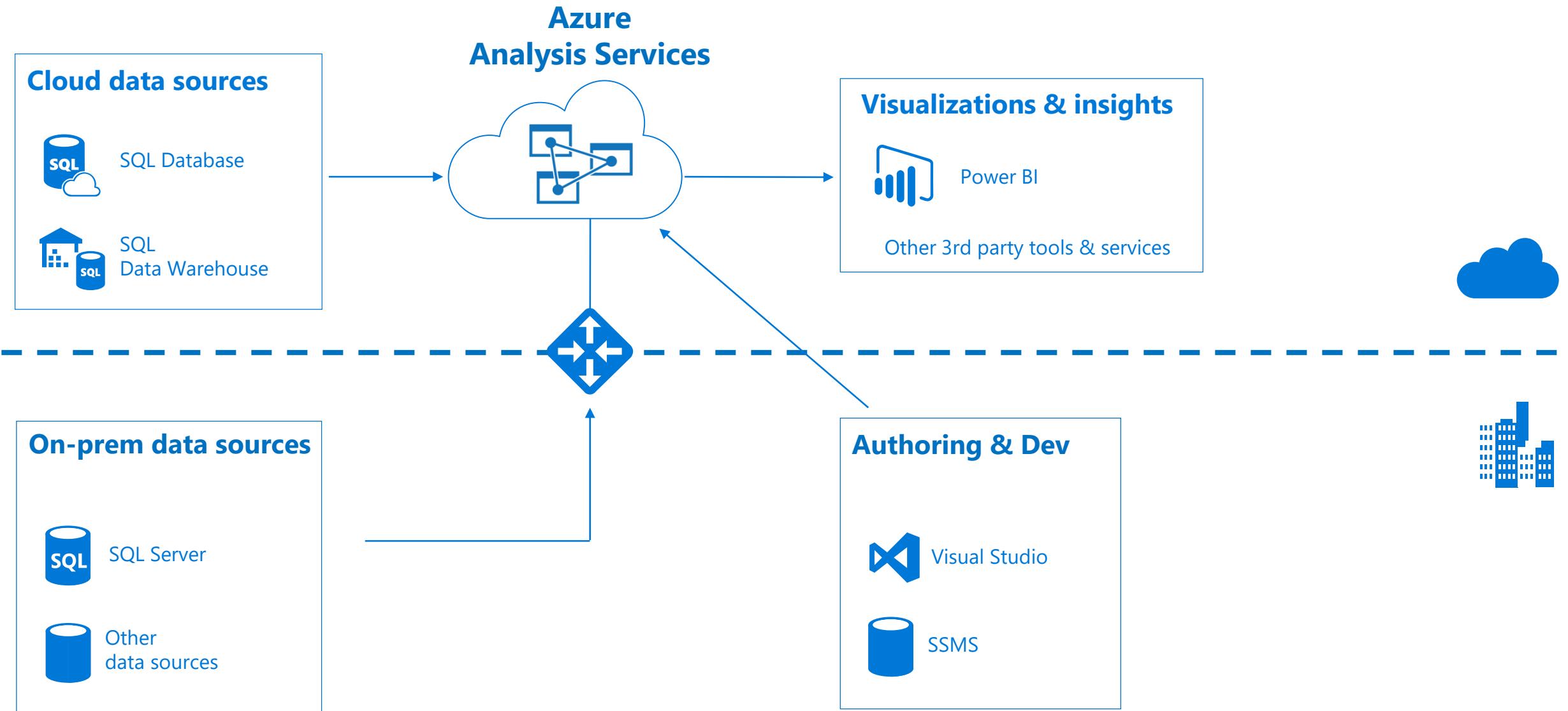
Provision and scale with ease

Easy to deploy, scale, and manage as a platform-as-a-service solution

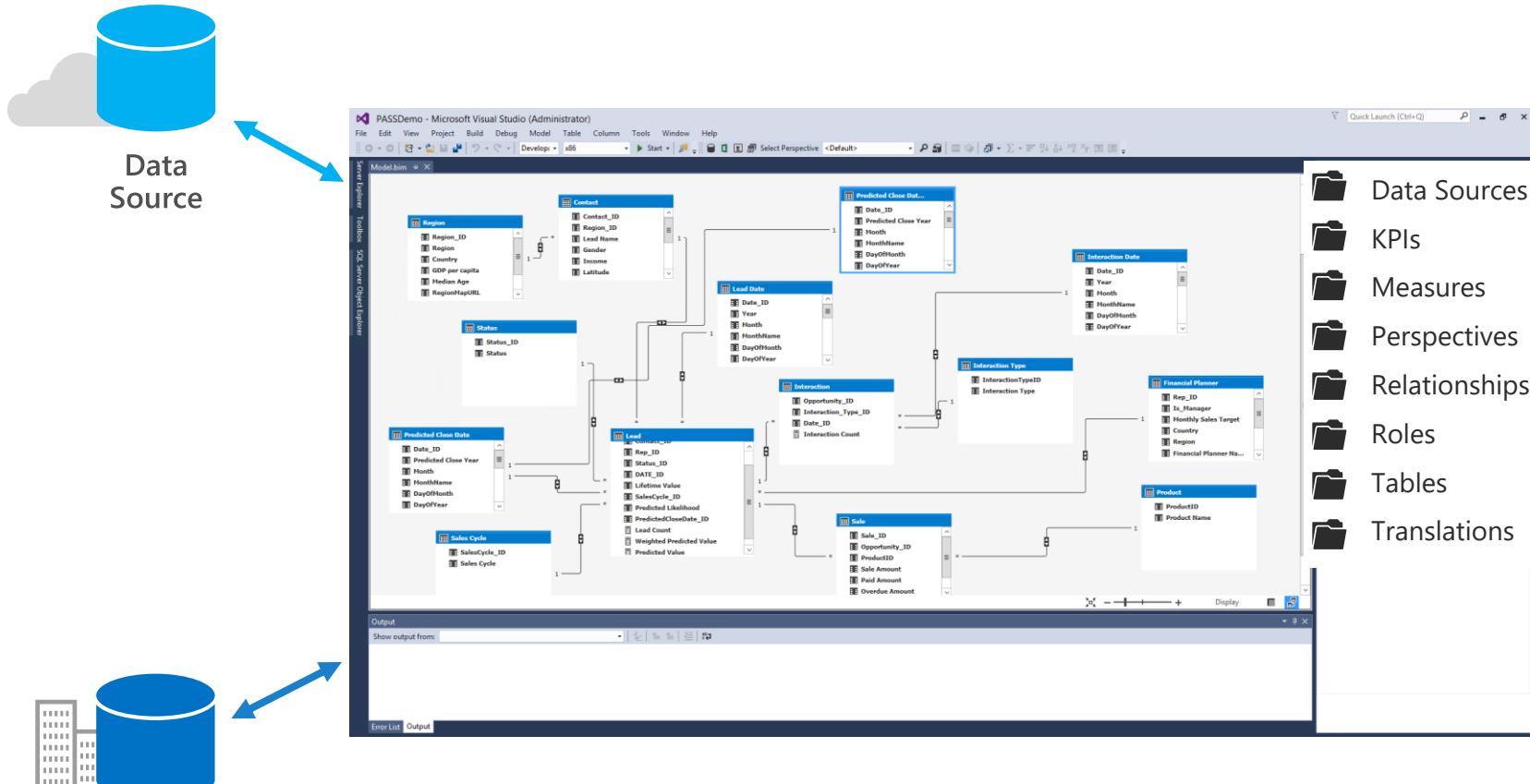
SQL Server Analysis Services



Azure Analysis Services



Rich semantic modelling



Data
Source

→ **Shape, transform, and clean**
data for analysis

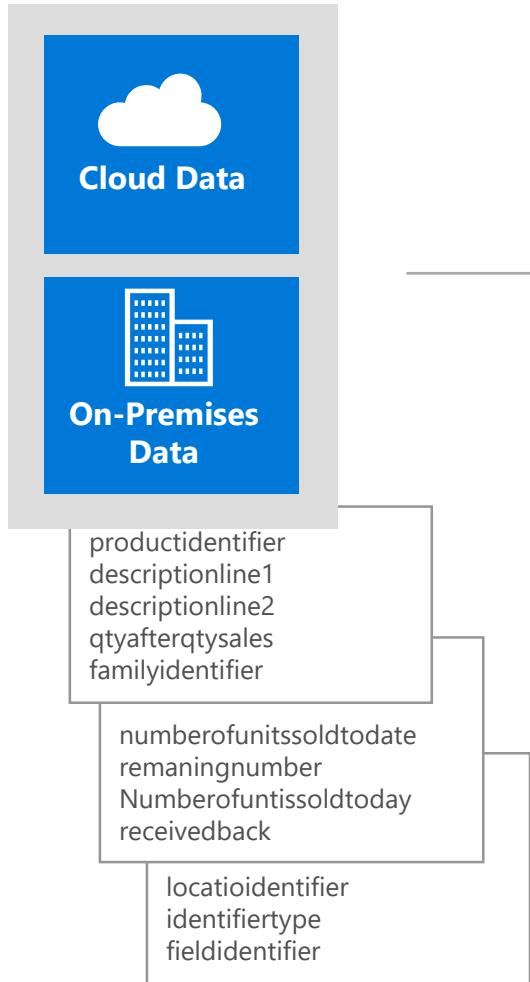
→ **Connect** to a broad range of
data across on-premises and
cloud

→ **Join** and **model** data from
multiple data sources and
different types

Rich semantic modelling

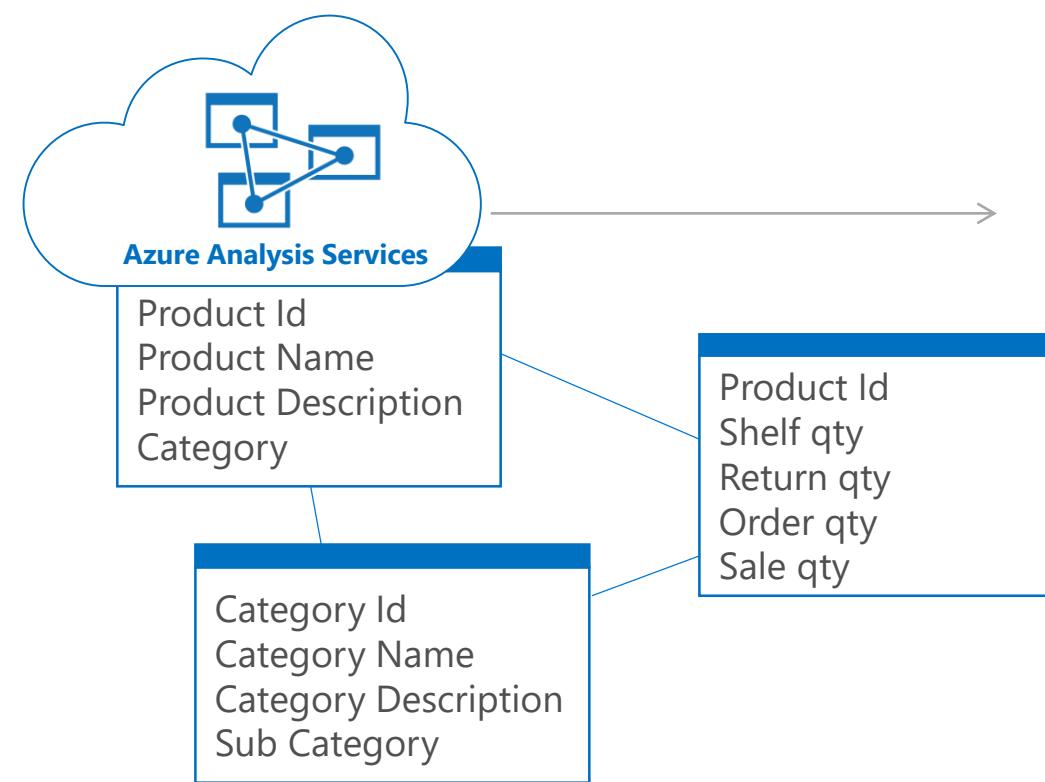
Source Data

Complex raw data optimized for processing



Semantic Model

Rich, business user friendly semantic model



Key capabilities at a glance

Create & Deploy



SQL Server Data Tools 2015 to create & deploy
AAS models (SSDT)
All SQL Server 2016 AS Tabular Enterprise Edition
features

- Tabular models (1200 compatibility)
- Including Direct Query, Partitions,
Perspectives, RLS, Translations
- Cloud data sources
- Gateway for on-prem data sources
- MD not supported at GA

Manage



Fully managed Platform-as-a-Service
Up to 100 GB memory
Deployment Wizard
SSMS & XMLA Management APIs
Azure Resource Manager & Portal integration
Pause & Resume
Elastic scale up/down (GA)
99.9% uptime SLA (GA)

Consume



AAD Auth/Identity
Full DAX & MDX support
PowerBI.com
Power BI Desktop
Excel
3rd party tools
Power BI Embedded (GA)

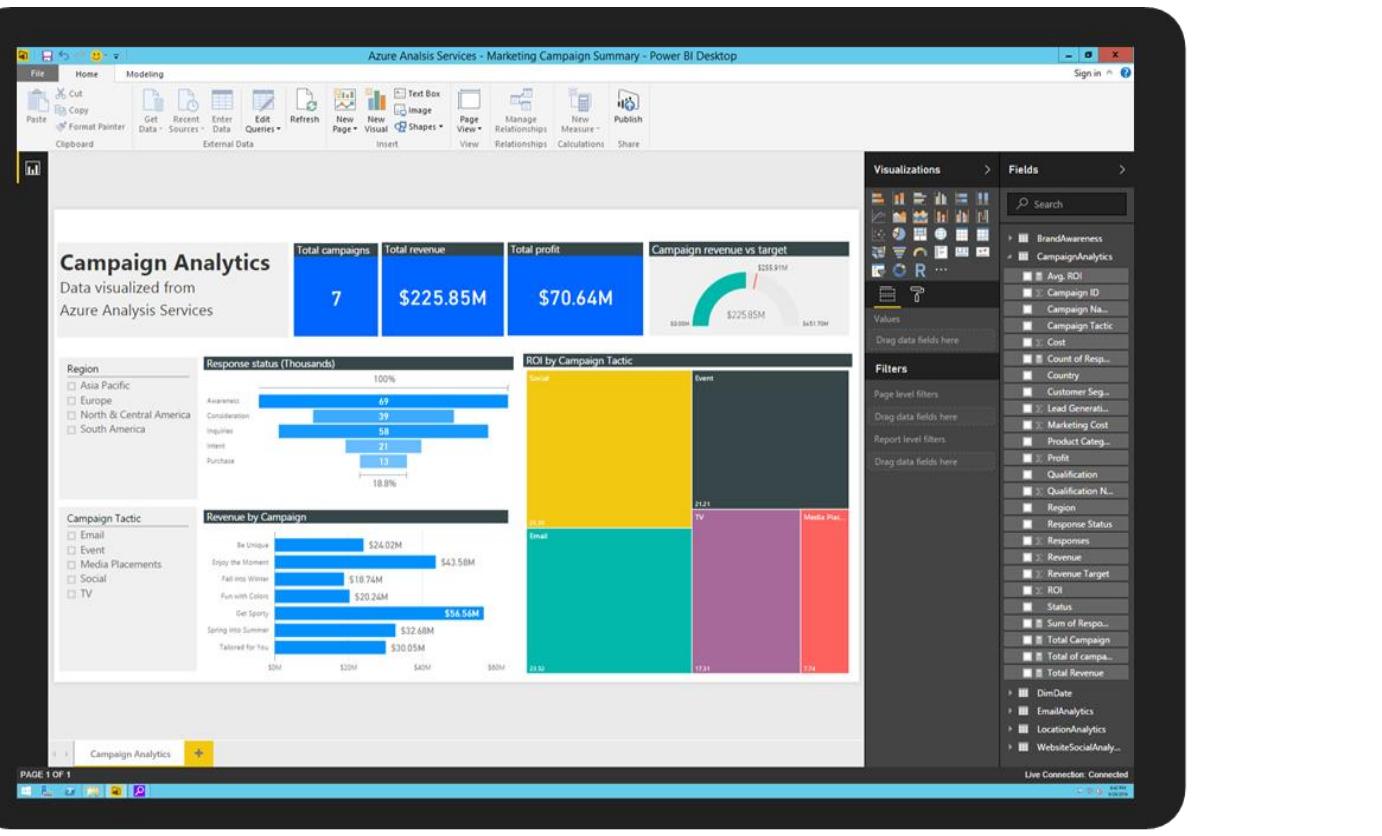
Service

South Central US , West Europe

Note: Check Azure regions for latest information <https://azure.microsoft.com/en-us/regions/services/>

Consume

Enable interactive, self-service tools for insights



Analysis at the speed of thought

Interactive query performance over billions of rows

Simplified view over complex data

Trusted data models

Single model for one version of the truth

Business user friendly

Support for preferred visualization tool

Full MDX and DAX support

Power BI & Power BI Desktop

Excel

Major 3rd party tools

Microsoft Cognitive Services

Give your apps a human side



Vision

From faces to feelings, allow your apps to understand images and video



Speech

Hear and speak to your users by filtering noise, identifying speakers, and understanding intent



Language

Process text and learn how to recognize what users want



Knowledge

Map complex information and data in order to solve specific tasks



Search

Access billions of web pages, images, videos, and news with the power of Bing



Labs

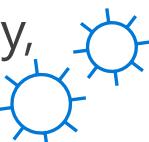
An early look at emerging Cognitive Services technologies: discover, try, and give feedback on new technologies before general availability

Why Microsoft Cognitive Services?

Easy

Roll your own with REST APIs

Simple to add: just a few lines of code required

Get a key,
Build 

Flexible

Integrate into the language and platform of your choice
Breadth of offerings helps you find the right for your app
Bring your own data for your custom experience



Tested

Built by experts in their field from Microsoft Research, Bing, and Azure Machine Learning
Quality documentation, sample code, and community support



Microsoft Cognitive Services updates



Vision

Video Indexer
Computer Vision
Face
Emotion
Content Moderator
Custom Vision



Speech

Speaker Recognition
Bing Speech
Custom Speech
Translator Speech
Unified Speech
Speech to Text w. Custom Speech
Text to Speech w. Custom Voice
Speech Translation w. Custom Translator



Language

Text Analytics
Bing Spell Check
Translator Text
Language Understanding (LUIS)



Knowledge

QnA Maker
Custom Decision



Search

Bing Entity Search
Bing Autosuggest
Bing Search
Web Search
Image Search
News Search
Video Search
Bing Statistics add-in
Bing Visual Search
Bing Custom Search



Labs

Project Gesture
Project Local Insights
Project Academic Knowledge
Project Entity Linking
Project Knowledge Exploration
Project Event Tracking
Project Answer Search
Project URL Preview
Project Anomaly Finder
Project Conversation Learner
Project Personality Chat

Cognitive Services capabilities

Infuse your apps, websites, and bots with human-like intelligence



Vision

- Object, scene, and activity detection
- Face recognition and identification
- Celebrity and landmark recognition
- Emotion recognition
- Text and handwriting recognition (OCR)
- Video metadata, audio, and keyframe extraction and analysis
- Explicit or offensive content moderation
- Custom image recognition



Speech

- Speech transcription (Speech-to-text)
- Speech Synthesis (Text-to-speech)
- Real-time speech translation
- Speaker identification and verification
- Custom Speech models for transcription and translation
- Custom voice



Language

- Language detection
- Text sentiment analysis
- Key phrase extraction
- Entity recognition
- Spell checking
- Explicit or offensive text content moderation, PII detection
- Text translation
- Customizable text translation
- Contextual language understanding



Knowledge

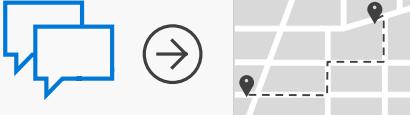
- Q&A extraction from unstructured text
- Knowledge base creation from collections of Q&As
- Semantic matching for knowledge bases
- Customizable content personalization learning



Search

- Ad-free web, news, image, and video search results
- Trends for video, news
- Image identification, classification and knowledge extraction
- Identification of similar images and products
- Named entity recognition and classification
- Knowledge acquisition for named entities
- Search query autosuggest
- Ad-free custom search engine creation

A variety of real-world applications

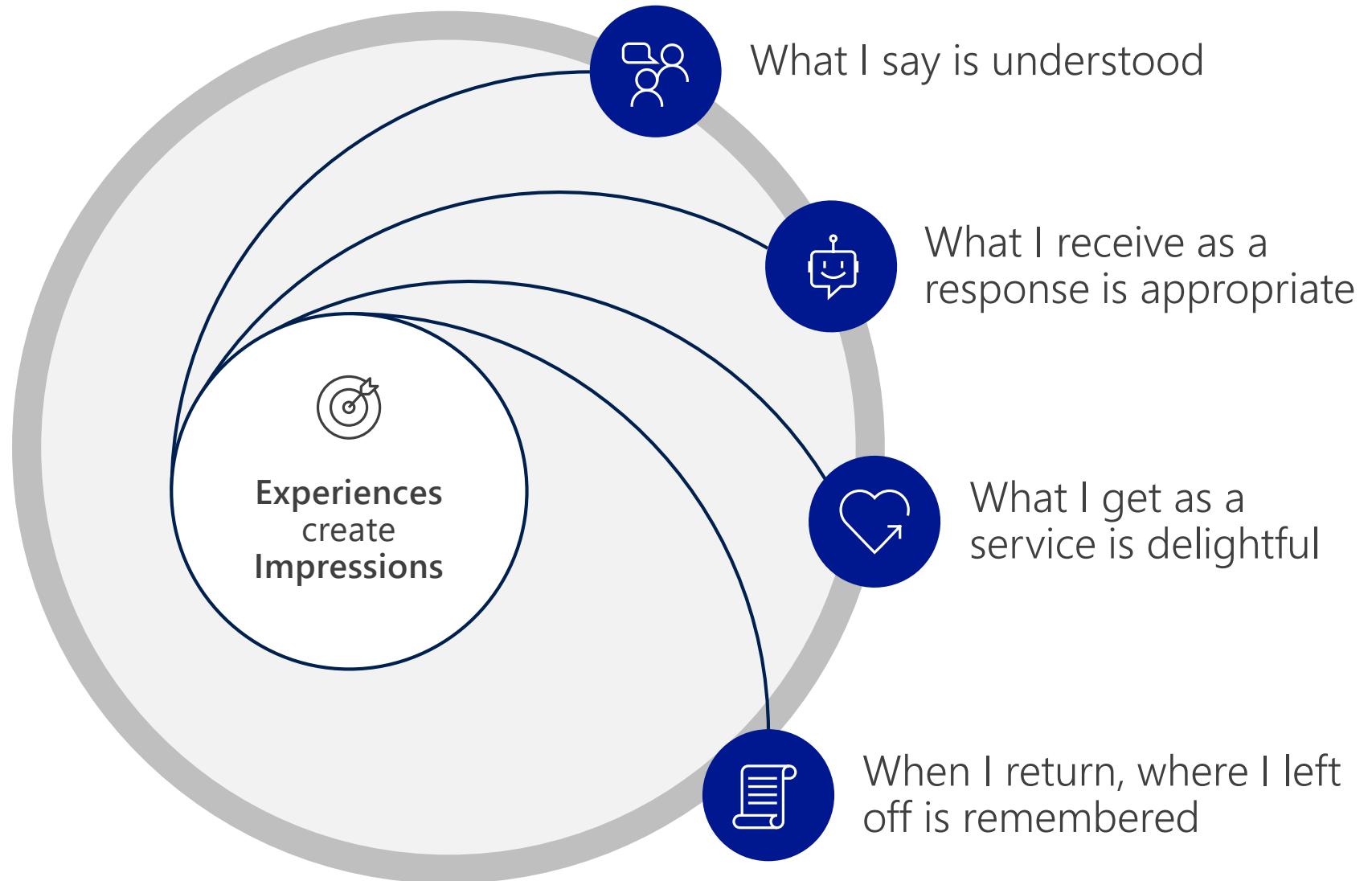
Vision	Speech	Language	Knowledge	Search														
 What is in the image or video? Intelligent Image insights  <table border="1"><tr><td>Category</td><td>People; 5 faces</td></tr><tr><td>Adult/Racy?</td><td>False/False</td></tr><tr><td>Dominant colors</td><td></td></tr><tr><td>Accent color</td><td></td></tr></table> Computer Vision	Category	People; 5 faces	Adult/Racy?	False/False	Dominant colors		Accent color		 Give me directions to the nearest local branch Speech to text  <table border="1"><tr><td>Convert spoken audio to text</td></tr><tr><td>Convert text to spoken audio</td></tr><tr><td>Extract intent of user</td></tr></table> Speech Service	Convert spoken audio to text	Convert text to spoken audio	Extract intent of user	 Play today's customer call recording Natural Language Processing  Intent: PlayCall Content: Customer# Date/Time.date: today Now Playing 11/29/2016 Customer Call Language Understanding	 QnA Pair of this site? Automatic extraction of questions and answers  QnA Maker	 Search for 'fraud prevention' Intelligent web search <table border="1"><tr><td> Information Communications Media Market News It also investigates the top three expected Fraud Detection and Prevention programs, in terms of demand in key markets...</td></tr><tr><td> The Big Question: In-House or Outsourced Fraud Protection? First, let's point out that there is not one absolute answer—there are "pros" and "cons" to each. Those who favor in-house...</td></tr><tr><td> How to Protect Your Business from Online Fraud this Holiday Season Michael heads fraud prevention tool. Online and mobile shopping are expected to continue growing apace...</td></tr></table> Bing News Search	 Information Communications Media Market News It also investigates the top three expected Fraud Detection and Prevention programs, in terms of demand in key markets...	 The Big Question: In-House or Outsourced Fraud Protection? First, let's point out that there is not one absolute answer—there are "pros" and "cons" to each. Those who favor in-house...	 How to Protect Your Business from Online Fraud this Holiday Season Michael heads fraud prevention tool. Online and mobile shopping are expected to continue growing apace...
Category	People; 5 faces																	
Adult/Racy?	False/False																	
Dominant colors																		
Accent color																		
Convert spoken audio to text																		
Convert text to spoken audio																		
Extract intent of user																		
 Information Communications Media Market News It also investigates the top three expected Fraud Detection and Prevention programs, in terms of demand in key markets...																		
 The Big Question: In-House or Outsourced Fraud Protection? First, let's point out that there is not one absolute answer—there are "pros" and "cons" to each. Those who favor in-house...																		
 How to Protect Your Business from Online Fraud this Holiday Season Michael heads fraud prevention tool. Online and mobile shopping are expected to continue growing apace...																		

Kinds of bots

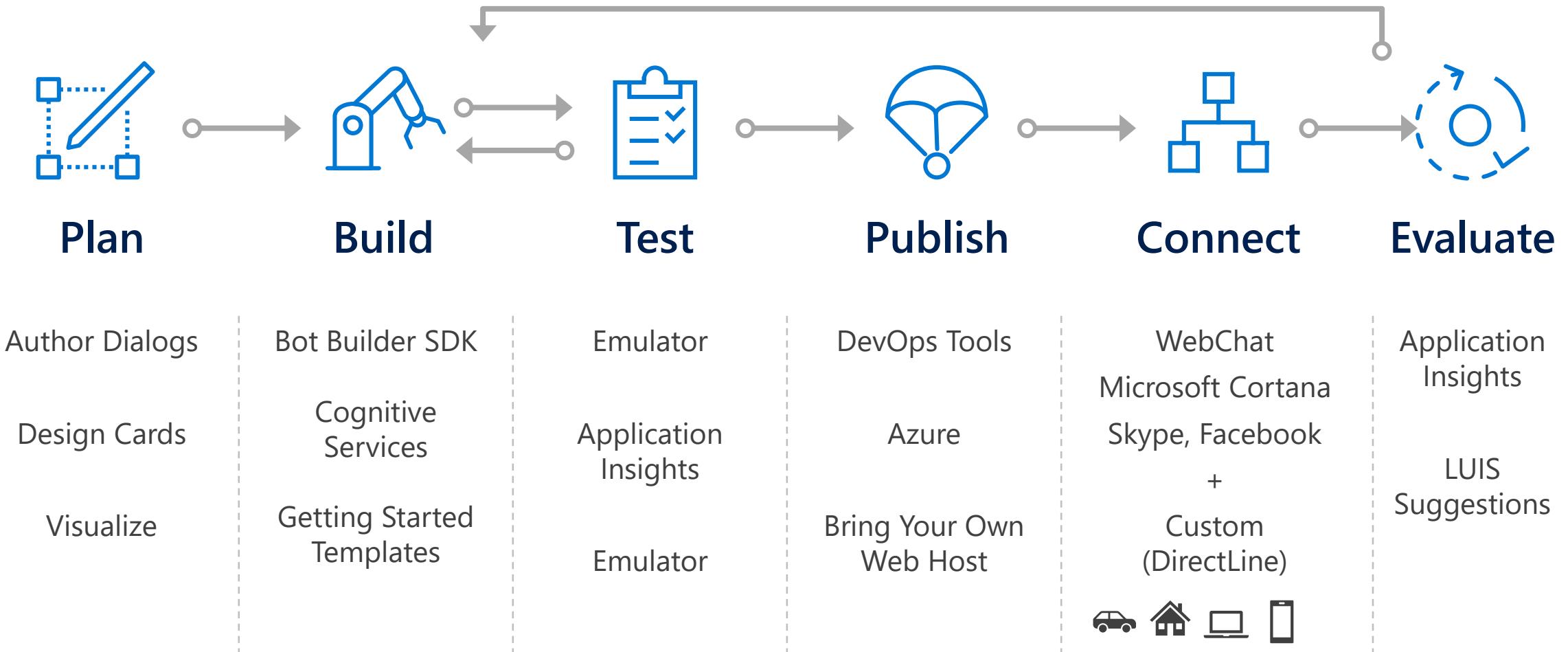
1,000+ companies engaging us

Scenario	Retail	Finance	Insurance	Telecoms	Government	Automotive	Manufacturing	Healthcare	Media	Events
Customer service	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Customer retail	✓	✓	✓	✓				✓	✓	
Audio/speech analysis	✓	✓	✓	✓	✓				✓	
Translation	✓	✓	✓							
Surveillance	✓				✓					
Knowledge extraction	✓	✓	✓	✓			✓			
Video/photo analysis	✓				✓				✓	
Product identification	✓						✓	✓	✓	
Digital assistant						✓				
Footfall analysis	✓						✓			✓
HD maps and object detection							✓			

**Hey Bot.
Here's what I
expect in a
conversation**



The Lifecycle of Your Project



Why Microsoft Services For Conversational

We equip organizations to put innovation into practice using our global reach and ecosystem to deliver strategic business outcomes, maximize the value of cloud technology, and drive success through continual support

Ingenuity

We creatively use technology to bring data assets into a searchable data estate that supports business insights and allows the latest AI services to be combined & applied to transform business processes

"Microsoft Services is instrumental in helping us overcome our biggest challenge, which is adopting a cloud-first mindset."

– Henk Van Driel, Manager, Cloud Competence Center, Rabobank



Expertise

We are the first, in close collaboration with Microsoft Research & Engineering to apply new technology to transform a customer's business and help them disrupt within their industries

"We're first to market with a digital solution that delivers extraordinary customer value. That's one result of a successful partnership that combines our deep industry expertise with the technology expertise of Microsoft."

– Christophe Devins, Founding Partner and CEO, Adents



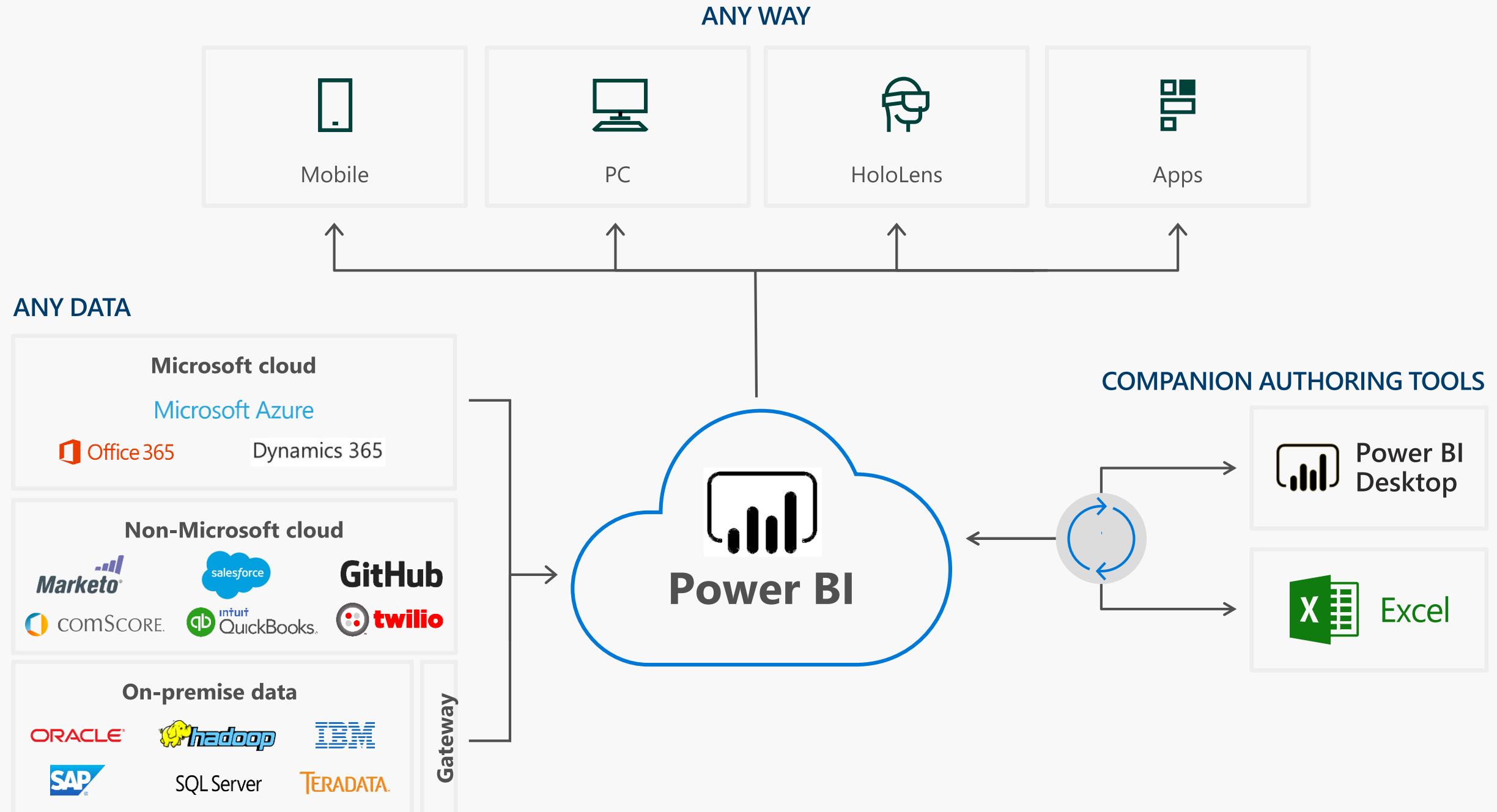
Empowerment

We mobilize organizations for success across the AI Maturity journey, delivering value with exceptional execution and knowledge transfer across strategy, implementation, adoption & change management and support

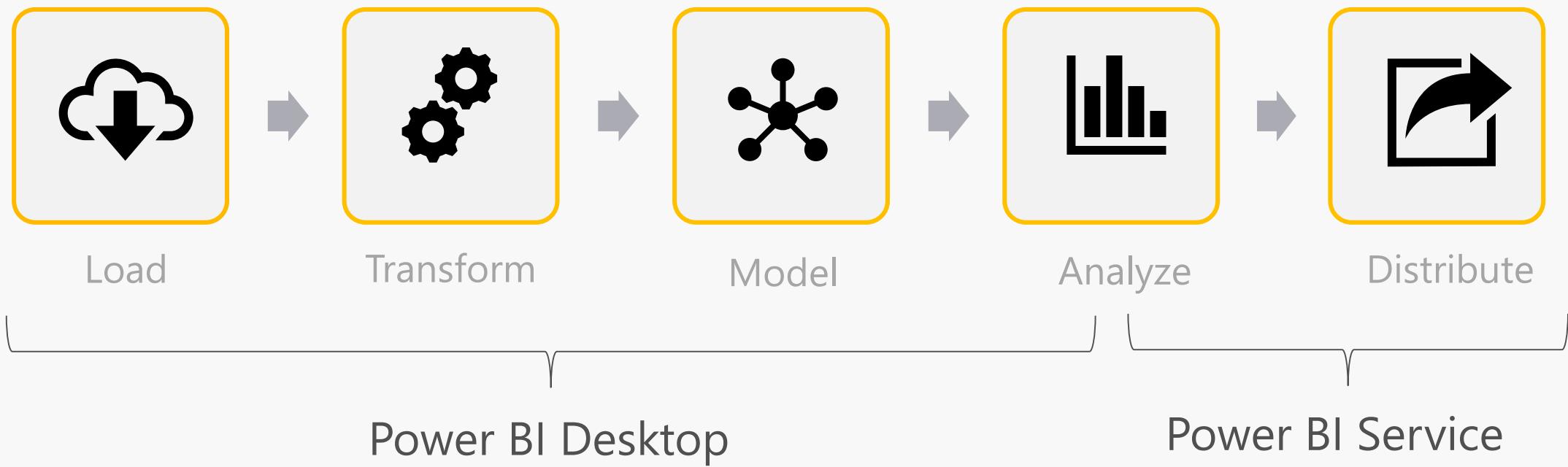
"Far from being a 'mere' technology provider, Microsoft is our close partner, providing advisory services from the envisioning phase and business case development to design, deployment, change management, and more."

– Joakim Plate, Director, Services Market Toyota Material Handling Europe





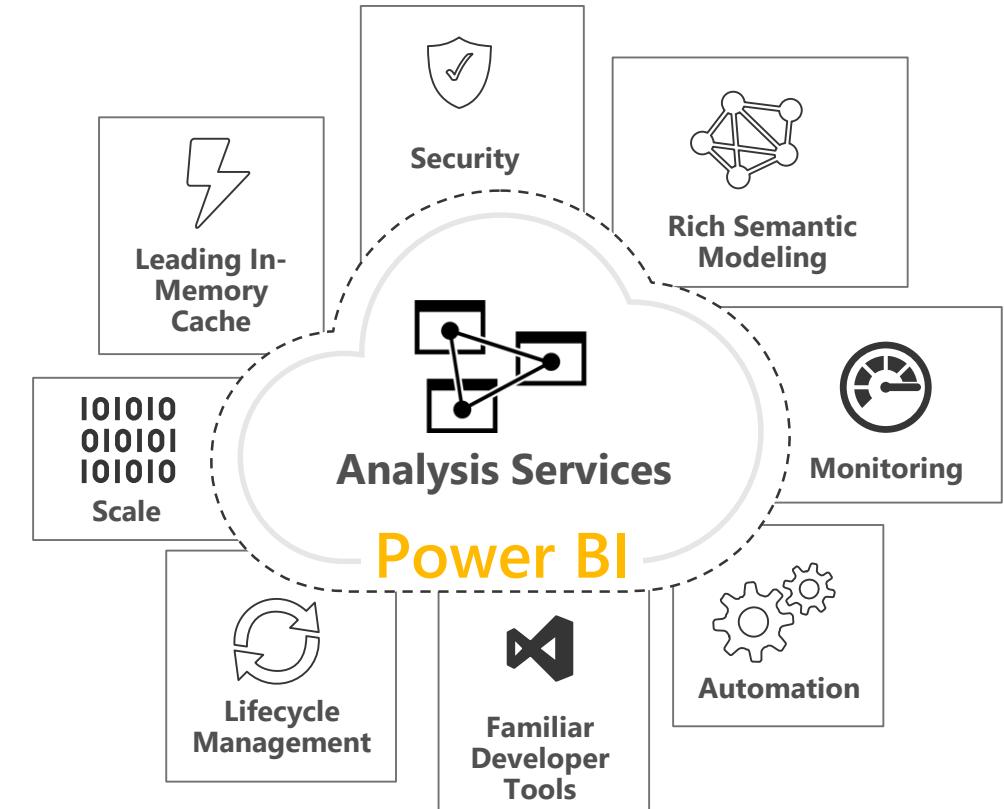
BI workflow



Analysis Services integrated into Power BI

Most Used, Most Battle-Tested Analytics Engine Technology

- Enterprise grade analytics engine as a service
- Leading query performance over massive data volumes
- Rich semantic modeling
- Proven DevOps experience

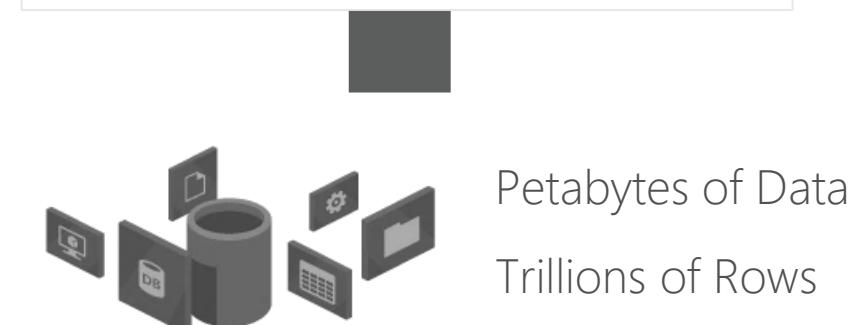
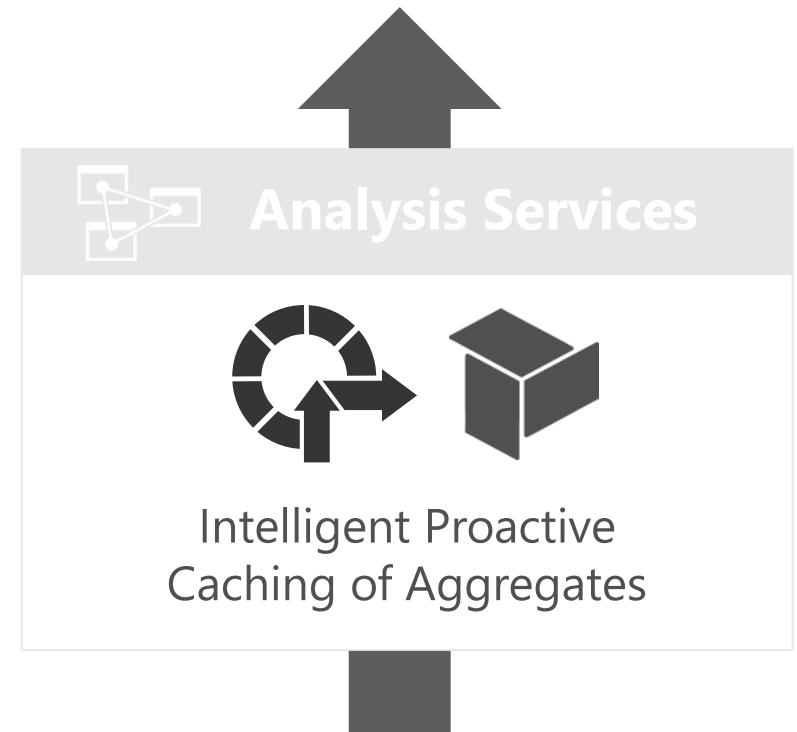


Analysis Services integrated into Power BI

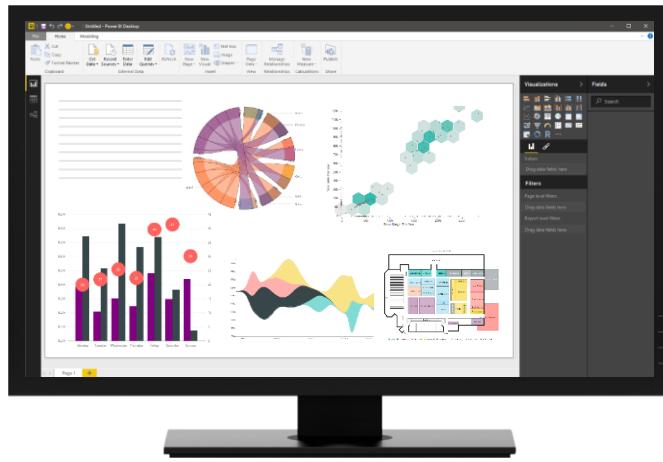
Enabling BI at any scale

- Interactive Exploration over even **TRILLIONS** of rows of data
- Enabled by intelligent proactive caching of aggregates for big data direct query sources
- Native capability of Analysis Services

Sub-Second Interactive Queries

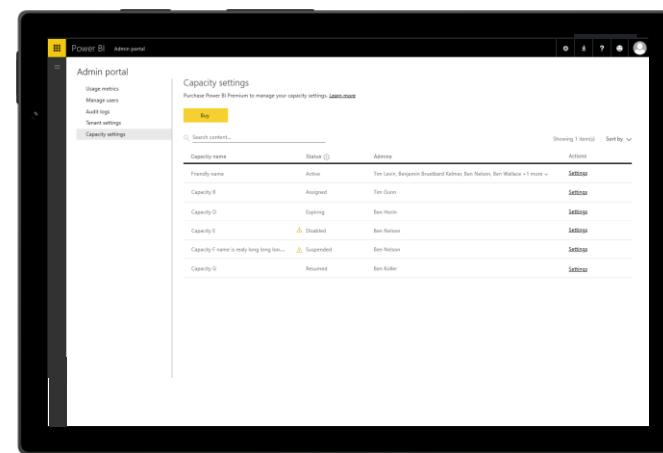


Create



Power BI Desktop
Free

Develop



Power BI Pro
\$10/month

Deploy



Power BI Embedded
Start for \$1/hour

Power BI or Power BI Embedded?

Power BI is right for you if:

- You are an **enterprise** looking for a complete BI solution
- You want a single view of what's happening **for your organization, partners, customers, and suppliers**
- You prefer a finished solution and APIs to customize your deployment
- You want to **help your organization** make decisions

Power BI Embedded is right for you if:

- You are a **software vendor or developer** and are building applications
- You want to **embed** visuals into your **applications**
- You want to leverage the **familiar environment of Azure**
- You want to **help your customers** make decisions



GE Healthcare

"Power BI Embedded performed much better ... for the data volumes that we're working with.

The visualization ... including the look, feel, and usability—is crisp and intuitive."

*Vamsee Rangavajhala
Senior Staff Digital Product Manager*

One-line Description

Data Factory

- Hybrid ETL with GUI + code base at scale

Data Catalog

- Cloud sourced Metadata repository

Event hub / IoT hub

- Message hub enabling big data analysis on telemetry signals



© 2018 Microsoft Corporation. All rights reserved. Microsoft, Windows, and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.