



PASS



# Argentina Local Group PASS

DATA Summit 2019

# Modern BI azure

27 de mayo 2019

**Maximiliano D. Accotto**

Owner Triggerdb Consulting SRL.

[www.triggerdb.com](http://www.triggerdb.com)

[www.triggerdb.com/blog](http://www.triggerdb.com/blog)

@maxiaccotto



# Acerca de mi

**Microsoft MVP Data Platform y speaker desde el año 2005.**

**Fundador de TriggerDB Consulting SRL.**

**Technical Solution Specialist Data Platform & BI con mas de 15 años de experiencia.**



<https://www.triggerdb.com>



[@maxiaccotto](https://twitter.com/maxiaccotto)



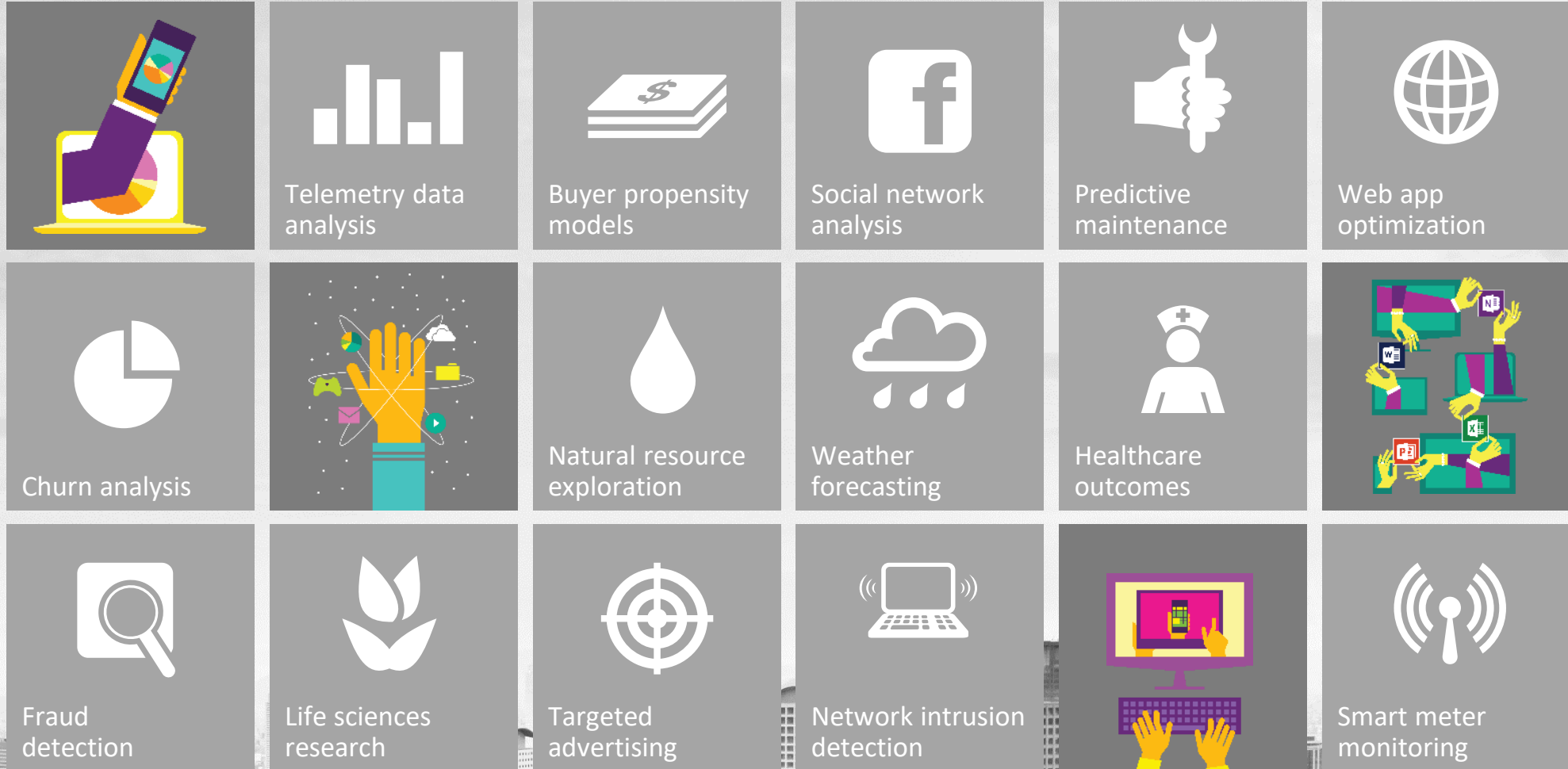
<https://www.linkedin.com/in/maxiaccotto>



[maxi@triggerdb.com](mailto:maxi@triggerdb.com)



# Cualquier tipo y de diferentes orígenes





# BI en la era de AI



IT



Científico de Datos

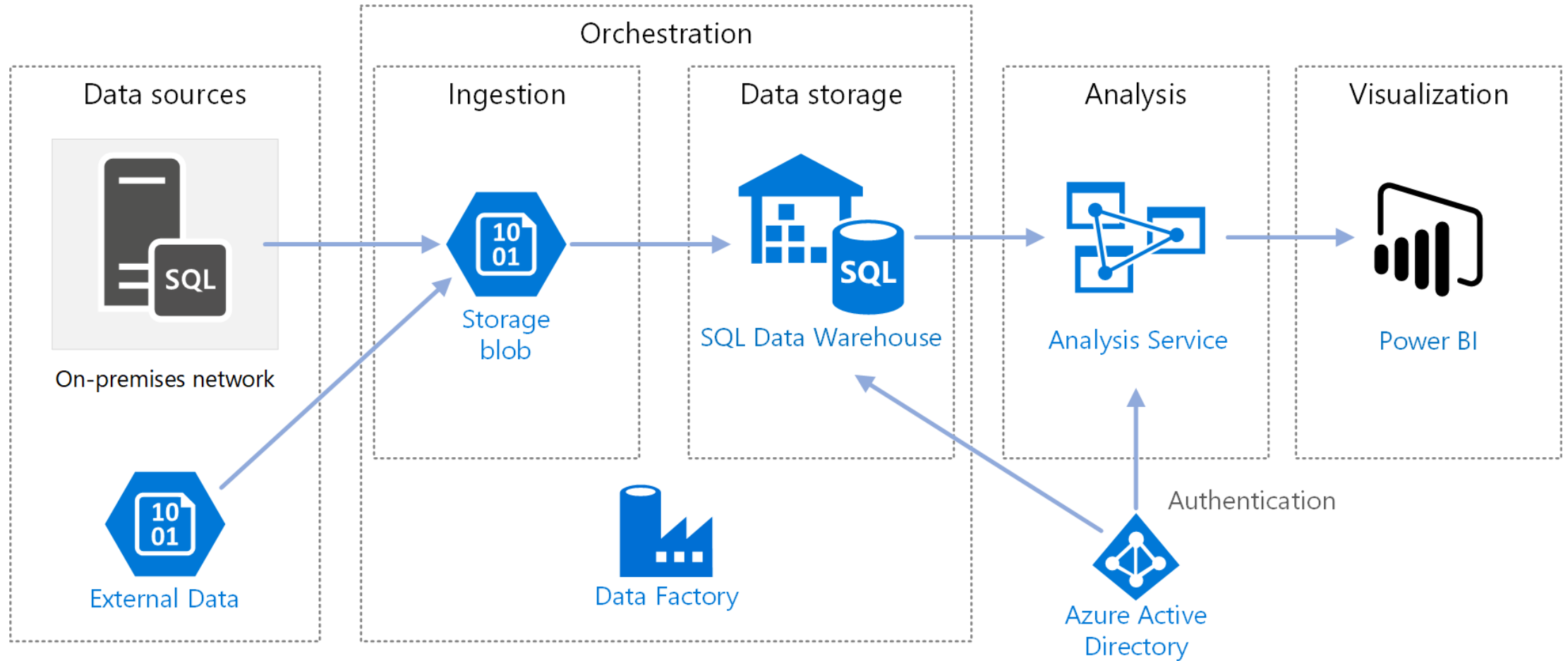


Analista

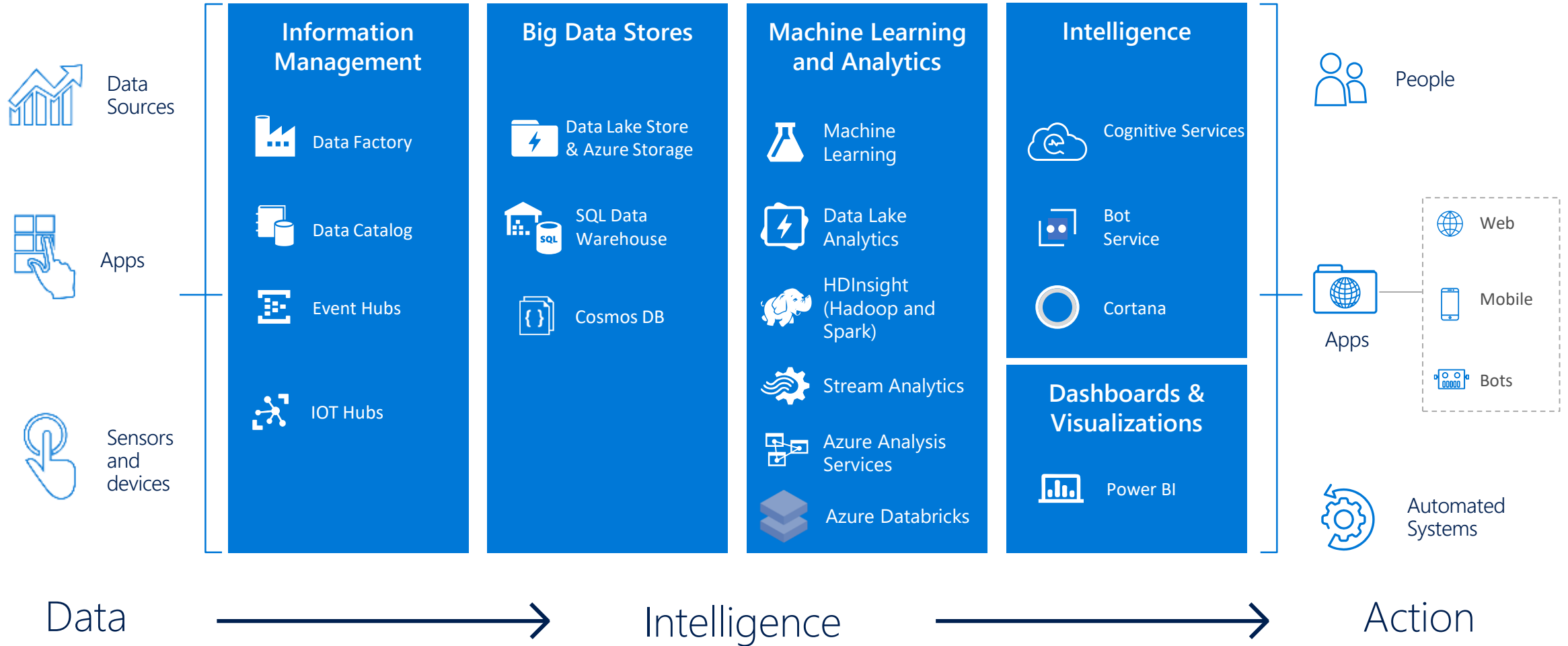


Usuario Final

# Arquitectura de BI moderna clásica



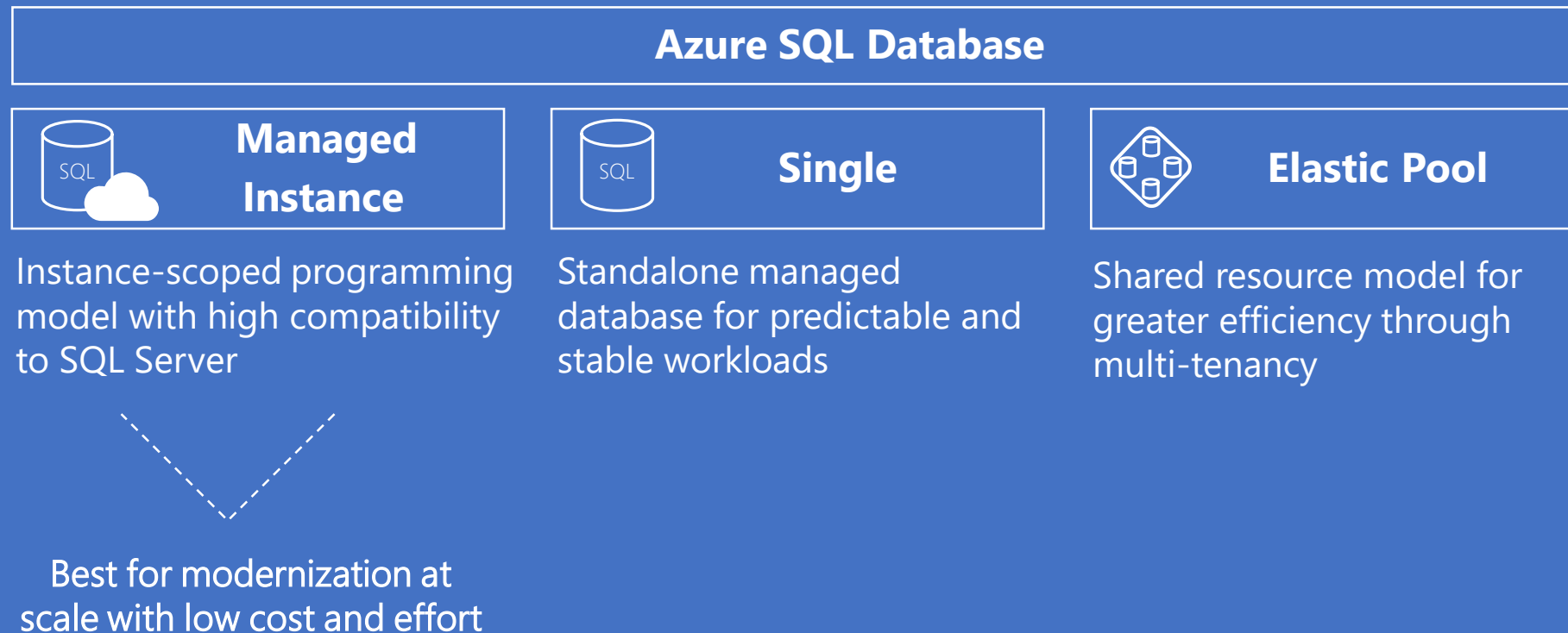
# Data Warehouse Modern BI: Components



# Azure SQL Database

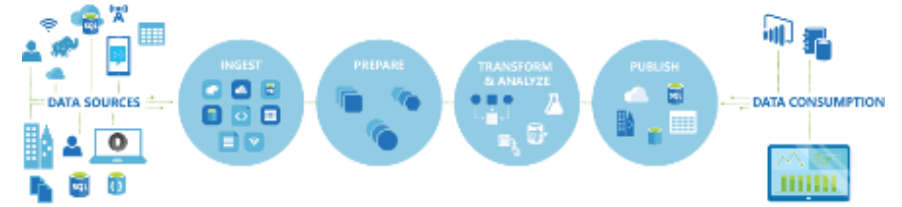


# AZURE SQL DATABASE



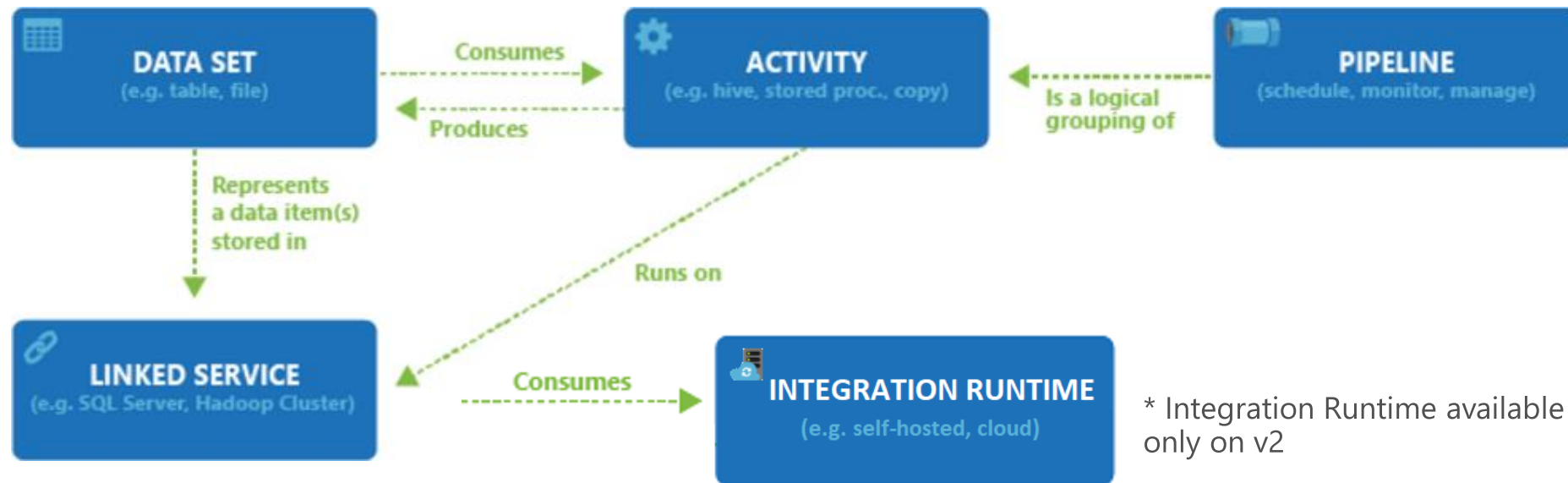
# Azure Data Factory

# Azure Data Factory



- **Orchestrate, monitor & schedule** - Compose data processing, storage & movement services (on premises & cloud)
- **Automatic infrastructure management**
  - Combine pipeline intent w/ resource allocation & mgmt
  - Globally available data movement capability
- **Single pane of glass** - One place to manage and monitor your data pipelines

- ADF is Microsoft's unified platform for ETL/ELT services in the cloud
- ADF allows you to build data pipelines and execute them/schedule their runs
- Data pipeline is a chain/group of activities to be performed on your data, e.g. data movements/transformations
- Some activities are powered by services with data store/compute resources allocated outside ADF, e.g. HDI/ML/etc.





# Azure SQL Data Warehouse

# Azure SQL Data Warehouse

PB scale with MPP and independent compute/ storage scale

Grow, shrink, and pause in seconds

Compute-optimized for demanding workloads

Unlimited columnar storage



# Fully managed PaaS

The screenshot displays the Azure portal interface for a ContosoRetailDW SQL Data Warehouse. The interface is divided into several sections:

- Essentials:** Displays key information about the resource group and server.
  - Resource group: [jrjwestusrg](#)
  - Location: West US
  - Subscription name: [ElasticScaleDev\\_657854](#)
  - Server name: [jrjwestus.database.windows.net](#)
  - Status: Online
  - Connection strings: [Show database connection strings](#)
- Monitoring:** Features a 'Query Activity' chart showing query execution over time (12 PM to 6 AM on May 3). The chart shows a peak in activity around 6 PM. Below the chart, a summary shows 48 SUCCESS and 1 FAIL queries.
- Settings:** A sidebar menu for configuring the data warehouse.
  - SUPPORT + TROUBLESHOOTING: Troubleshoot, Audit logs, Resource health, New support request.
  - RESOURCE MANAGEMENT: Locks, Export template.
  - GENERAL: Properties, Scale, Auditing & Threat detection (selected), Transparent data encryption.
- Auditing & Threat detection:** A panel for managing auditing and threat detection settings.
  - Inherit settings from server:** Checked.
  - Auditing:** ON/OFF toggle.
  - Storage Details:** [Storage Details](#) (locked), [Storage Details](#) (locked).
  - Audited Events:** All (locked).
  - Threat detection (preview):** ON/OFF toggle.
  - Threat detection types:** All (locked).
  - Send alerts to:** Email addresses (input field).
  - Email service and co-administrators:** Checked.

# Target workload: Analytics

- Store large volumes of data
- Consolidate disparate data into a single location
- Shape, model, transform and aggregate data
- Perform query analysis across large datasets
- Ad-hoc reporting across large data volumes
- All using simple SQL constructs

“SQL on SQL”



# Scaling and Pausing



Compute



Control



Storage



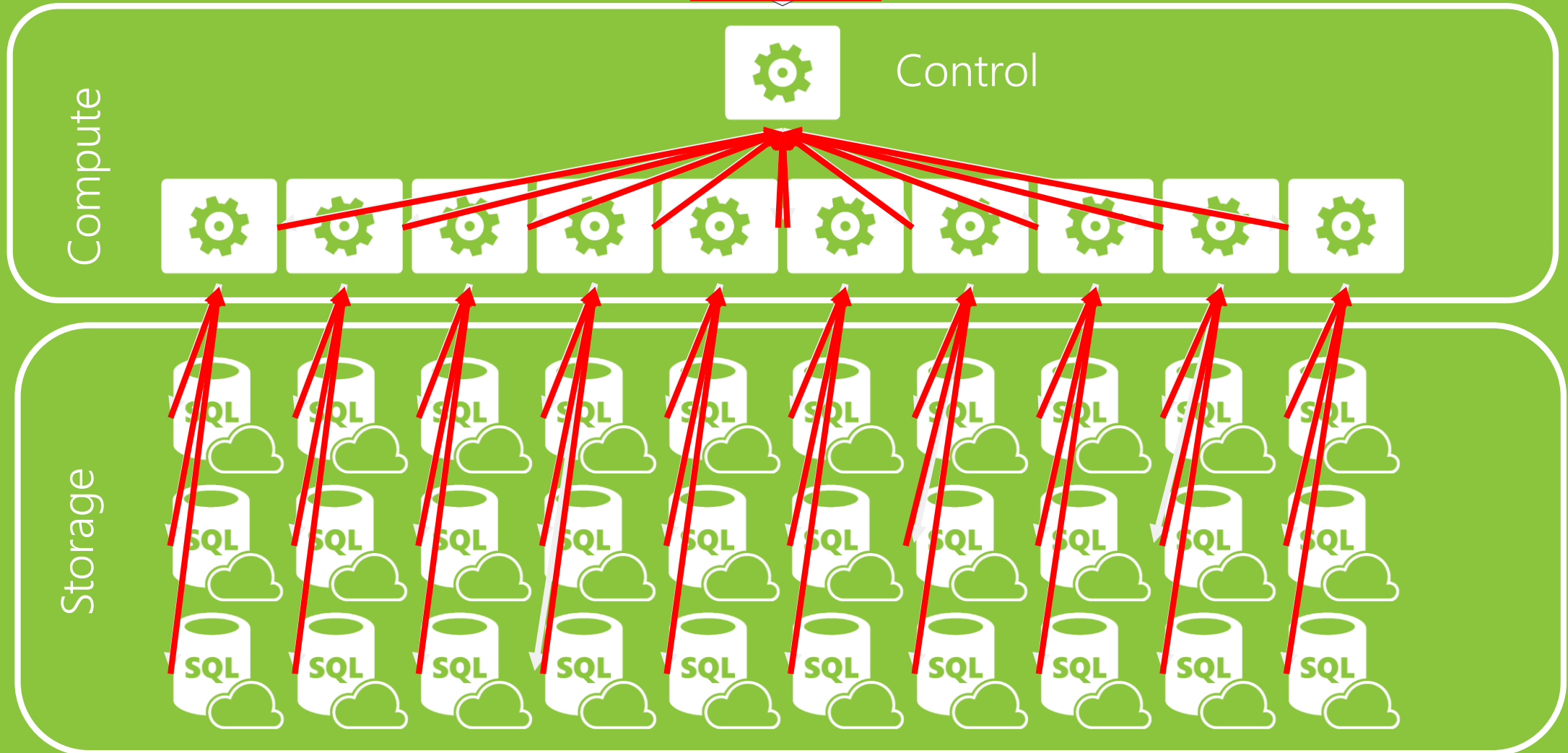
# Distributing Data

```
CREATE TABLE [build].[FactOnlineSales]
(
    [OnlineSalesKey]    int      NOT NULL
,   [DateKey]          datetime NOT NULL
,   [StoreKey]         int      NOT NULL
,   [ProductKey]       int      NOT NULL
,   [PromotionKey]     int      NOT NULL
,   [CurrencyKey]      int      NOT NULL
,   [CustomerKey]      int      NOT NULL
,   [SalesOrderNumber] nvarchar(20) NOT NULL
,   [SalesOrderLineNumber] int      NULL
,   [SalesQuantity]    int      NOT NULL
,   [SalesAmount]      money    NOT NULL
)
WITH
(
    DISTRIBUTION = HASH([ProductKey])
)
;
```

# Query Execution



Result

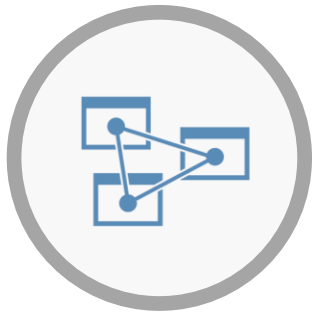


# Azure Analysis Services



# 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

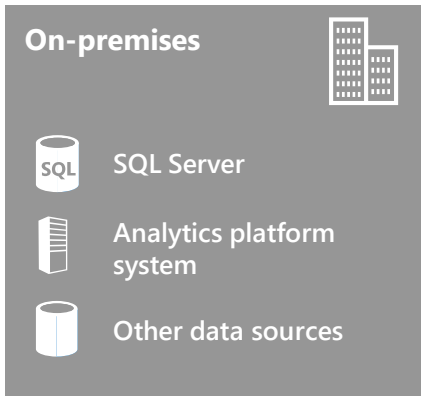
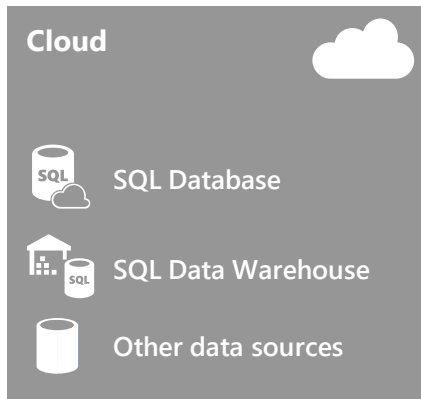


## Provision and scale with ease

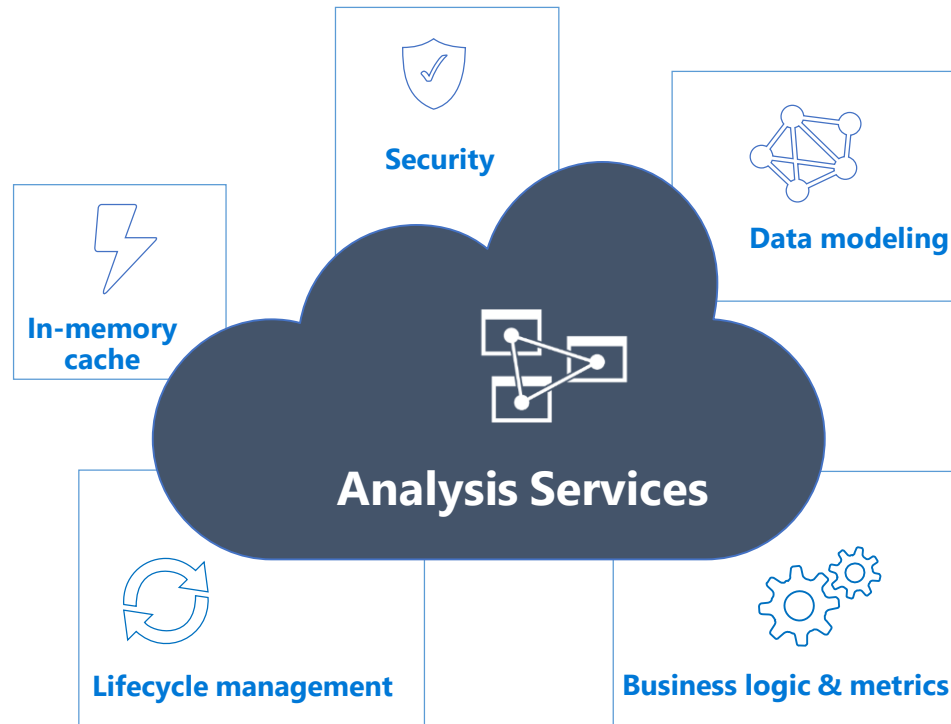
Easy to deploy, scale, and manage as a platform-as-a-service solution

# Analysis Services

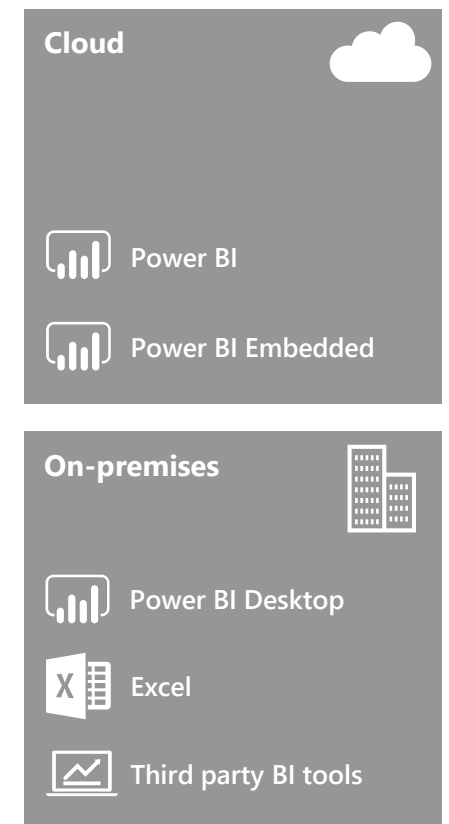
## Data sources



## BI semantic model



## Client tools



# Azure Analysis Services

## Azure Analysis Services

### Cloud data sources



SQL Database



SQL Data Warehouse

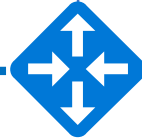


### Visualizations & insights



Power BI

Other 3rd party tools & services



### On-prem data sources



SQL Server



Other data sources



### Authoring & Dev



Visual Studio



SSMS



# Big Data Store





Data Sources



Apps



Sensors and devices

## Information Management



Data Factory



Data Catalog



Event Hubs



IOT Hubs

## Big Data Stores



Data Lake Store



SQL Data Warehouse



Cosmos DB

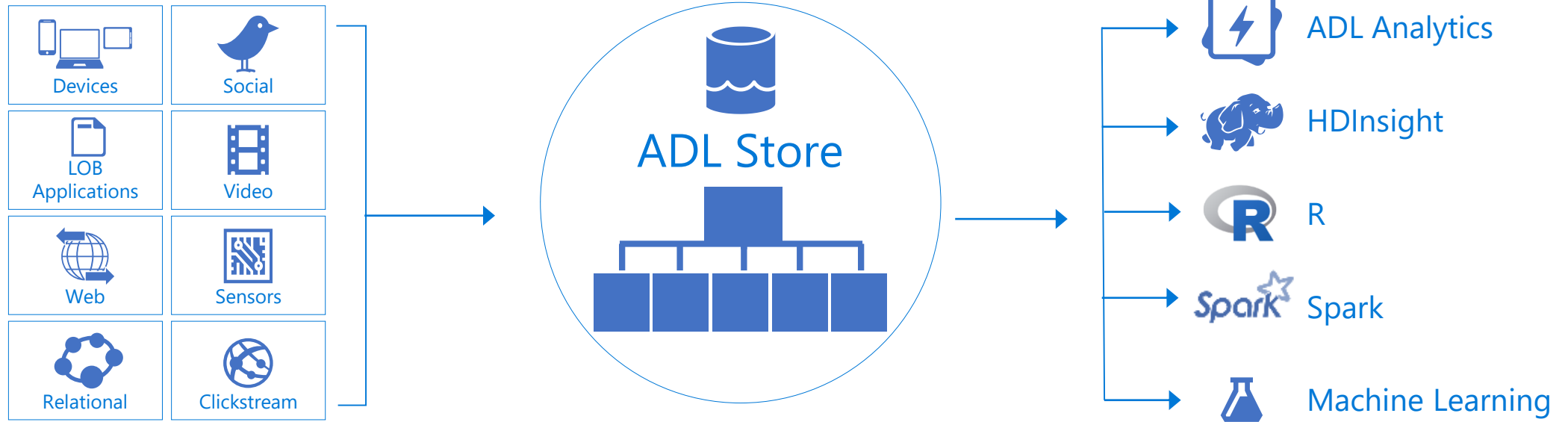
Data

## Big Data Stores

 Data Lake Store

 SQL Data Warehouse

 Cosmos DB



- A Hadoop Distributed File System for the cloud
- No fixed limits on file size
- No fixed limits on account size
- Unstructured and structured data in their native format

- Massive throughput to increase analytic performance
- High durability, availability, and reliability
- Azure Active Directory access control

# Data lake is the center of a big data solution

A storage repository, usually Hadoop, that holds a vast amount of raw data in its native format until it is needed.

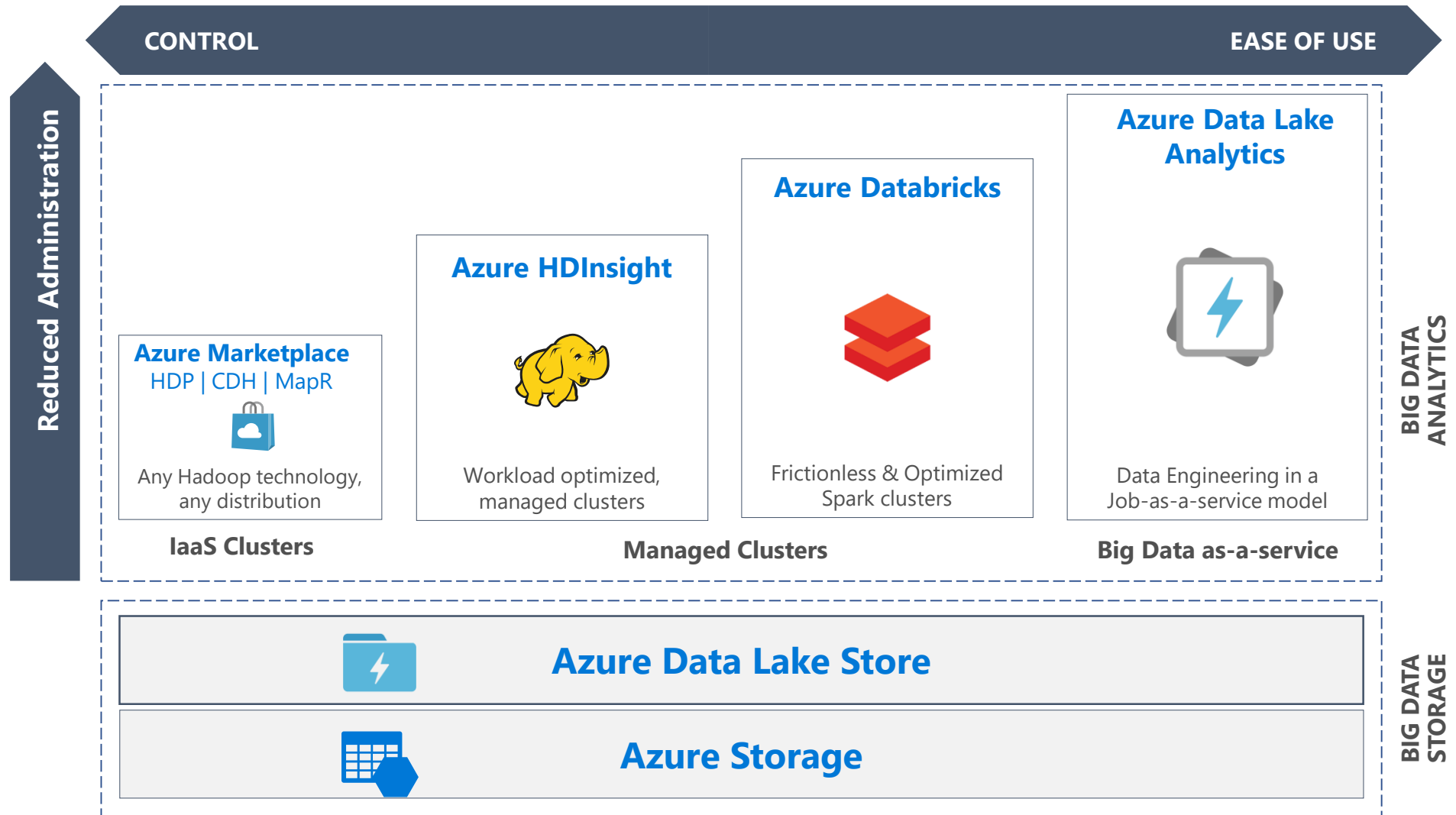
- Inexpensively store unlimited data
- Collect all data “just in case”
- Store data with no modeling – “Schema on read”
- Complements EDW
- Frees up expensive EDW resources
- Quick user access to data
- ETL Hadoop tools
- Easily scalable
- With Hadoop, high availability built in

# ADL Store vs Blob Store

	Azure Data Lake Store	Azure Blob Storage
Purpose	Optimized for big data analytics	General purpose bulk storage
Use Cases	Batch, Interactive, Streaming	App backend, backup data, media storage for streaming
Units of Storage	Accounts / Folders / Files	Accounts / Containers / Blobs
Structure	Hierarchical File System	Flat namespace
WebHDFS	Implements WebHDFS	No (WASB)
Security	AD	SAS keys
Storage	Auto Shared/Files chunked	Manually manage expansion/Files intact
Size Limits	No limits on account size, file size, # files	500TB account, 4.75TB file
Service State	Generally Available	Generally Available
Billing	Pay for data stored and for I/O	Pay for data stored and for I/O
Region Availability	Two US regions (East, Central) & North Europe (Other regions coming soon)	All Azure Regions

# Big Data

# KNOWING THE VARIOUS BIG DATA SOLUTIONS



# Azure HDInsight

Hadoop and Spark  
as a Service on Azure



**Fully-managed** Hadoop and Spark  
for the cloud

**100% Open Source** Hortonworks  
data platform

Clusters up and **running in minutes**

Managed, monitored and supported  
by Microsoft with the **industry's best SLA**

Familiar **BI tools for analysis**, or open source  
notebooks for **interactive data science**

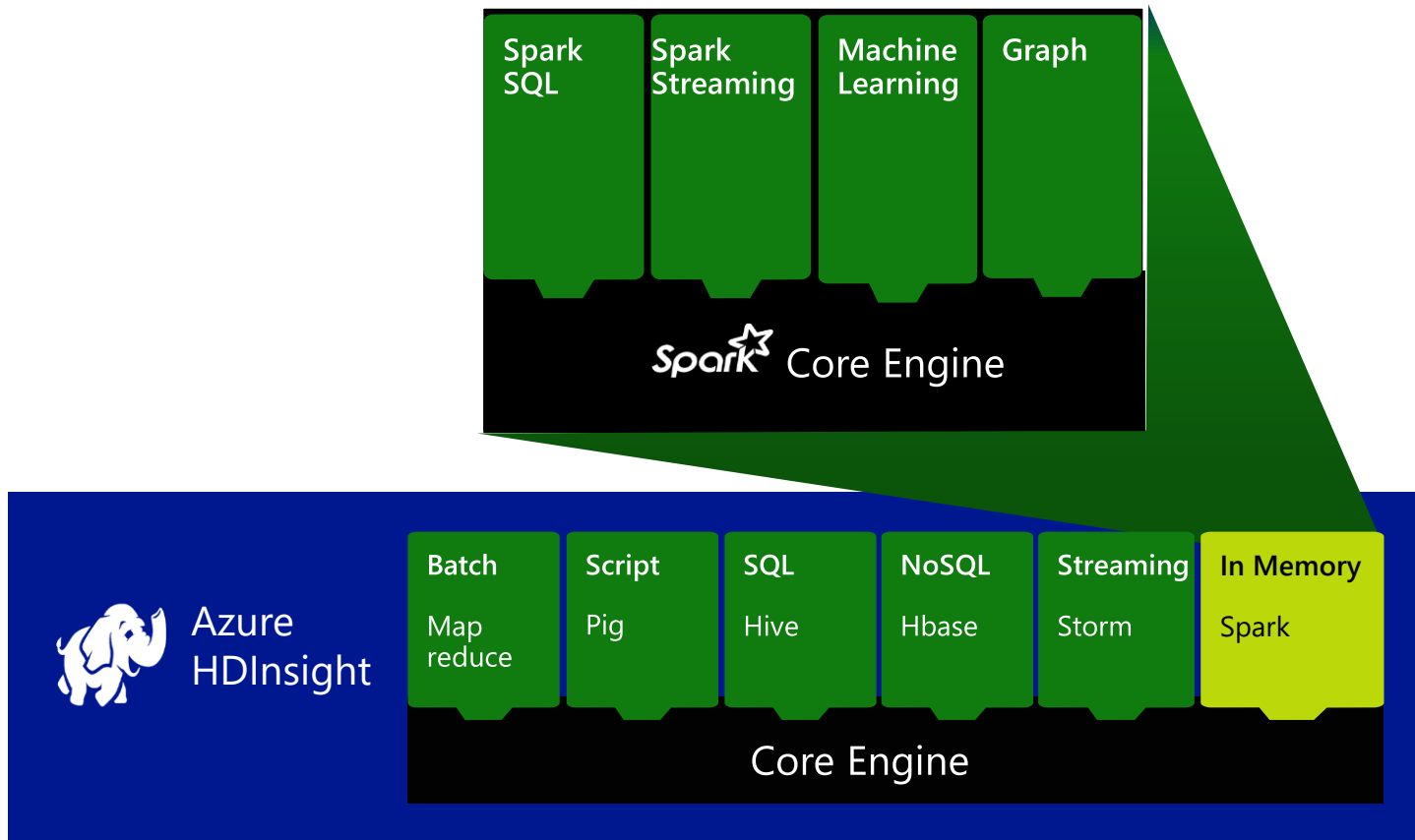
**63% lower TCO** than deploy your own  
Hadoop on-premises\*

\*IDC study "The Business Value and TCO Advantage of Apache Hadoop in the Cloud with Microsoft Azure HDInsight"



# Spark for Azure HDInsight

## In Memory Processing on Multiple Workloads



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

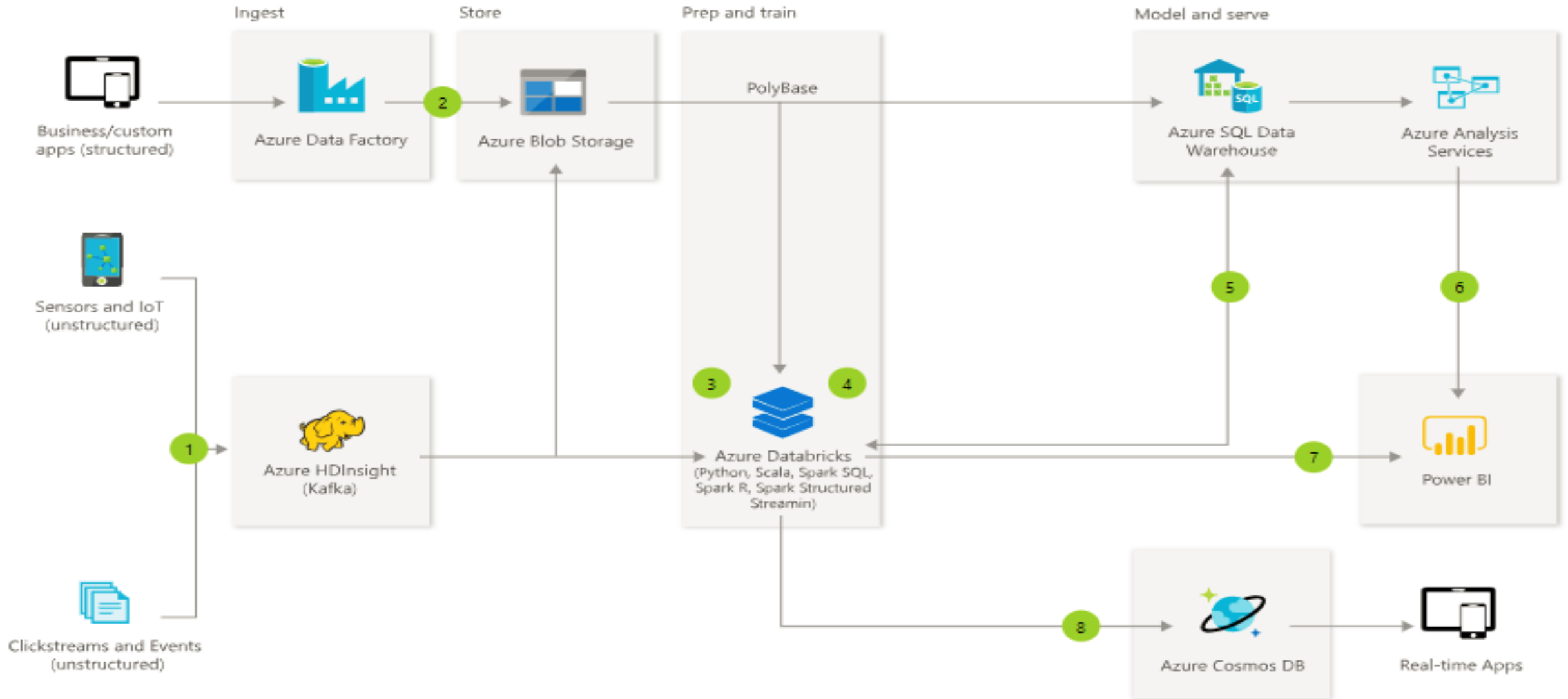
# Machine Learning Server for HDInsight



- **Familiarity of R or Python**
- **Scalability of Hadoop and Spark**
- **Up to 7x faster using Spark engine**
- **Train and run ML models on datasets of any size**
- **Cloud managed solution (easy setup, elastic, SLA)**

Only managed, cloud solution for doing R or Python

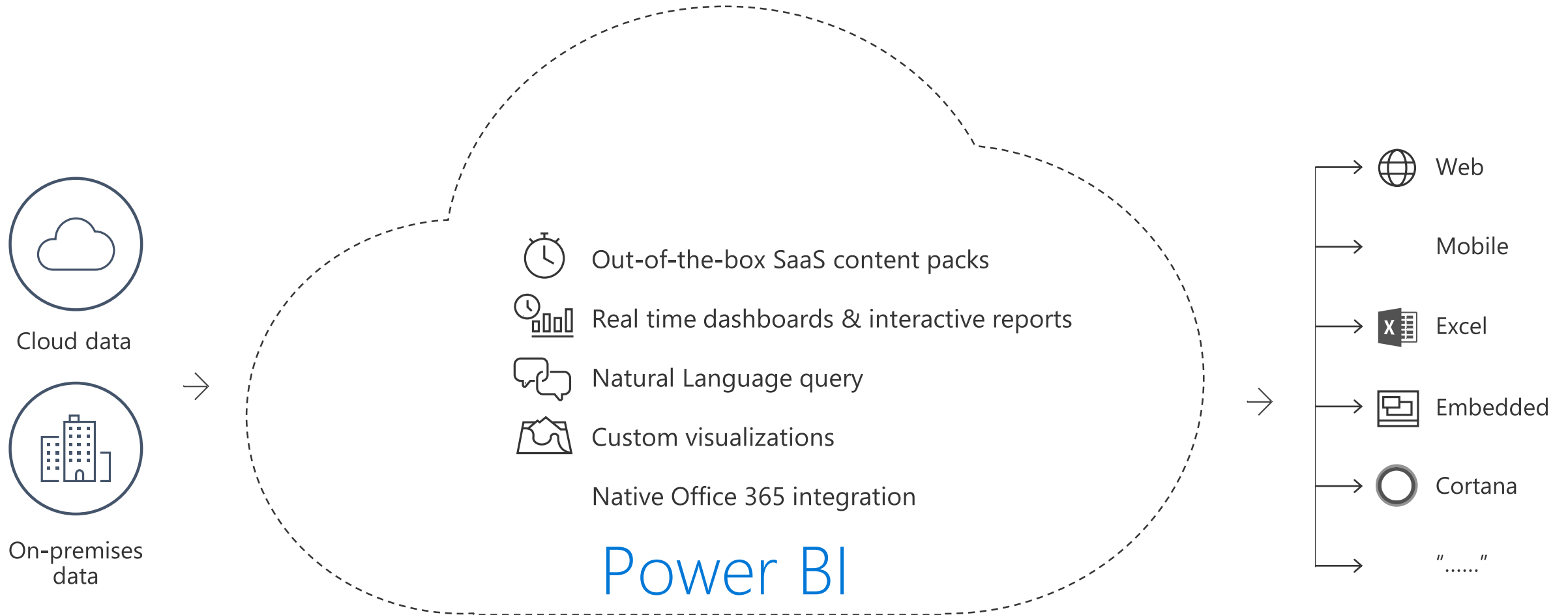
# Real-time analytics



PowerBI

# Power BI: experience your data

Any data, any way, anywhere



# Power BI product portfolio

Author



Power BI  
Desktop

Free data analysis  
and report authoring tool

Share and collaborate



Power BI  
service

Cloud-based modern  
business analytics solution

Large scale deployments



Power BI  
Premium

Dedicated capacity for  
increased performance

Share and collaborate



Power BI  
Report Server

On-premises report server

App dev



Power BI  
Embedded

Visual analytics embedded  
in your applications

Demo



# Para Conectarnos

Registrarse para una membresía gratuita y asociarse al grupo Local SQL Argentina:

pass.org  
Sqlargentina.pass.org



<https://www.facebook.com/groups/SQLArgentina/>



@sqlargentina



<https://www.linkedin.com/groups/3664741>



<https://www.youtube.com/channel/UCGa-ldNDwP8nAVfQwqoeYHw>