



Microsofts Big Data & AI Solution Patterns

Jon Jahren
Data & AI Architect
Microsoft



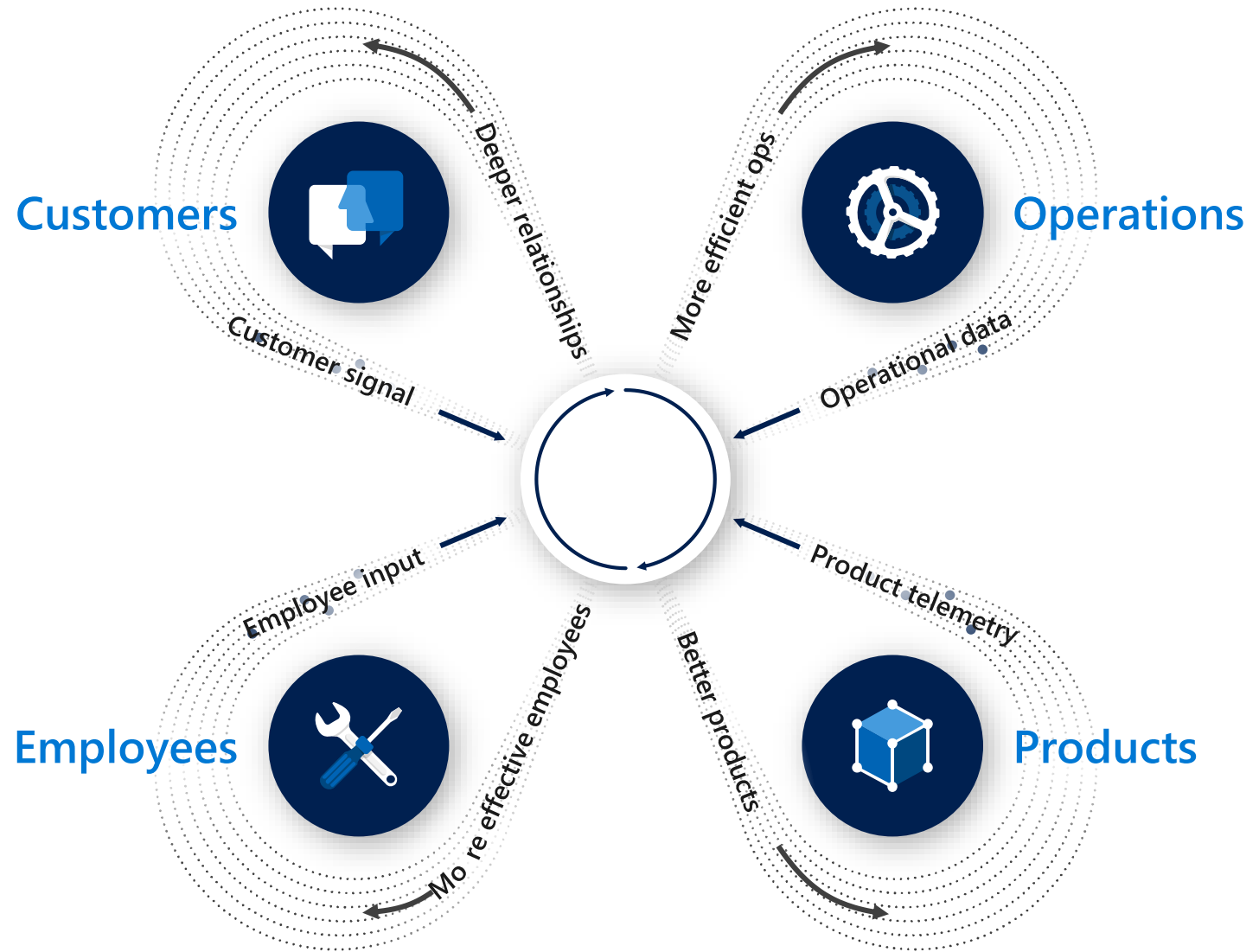
January 2019

A Leader in
Analytics & BI
Platforms



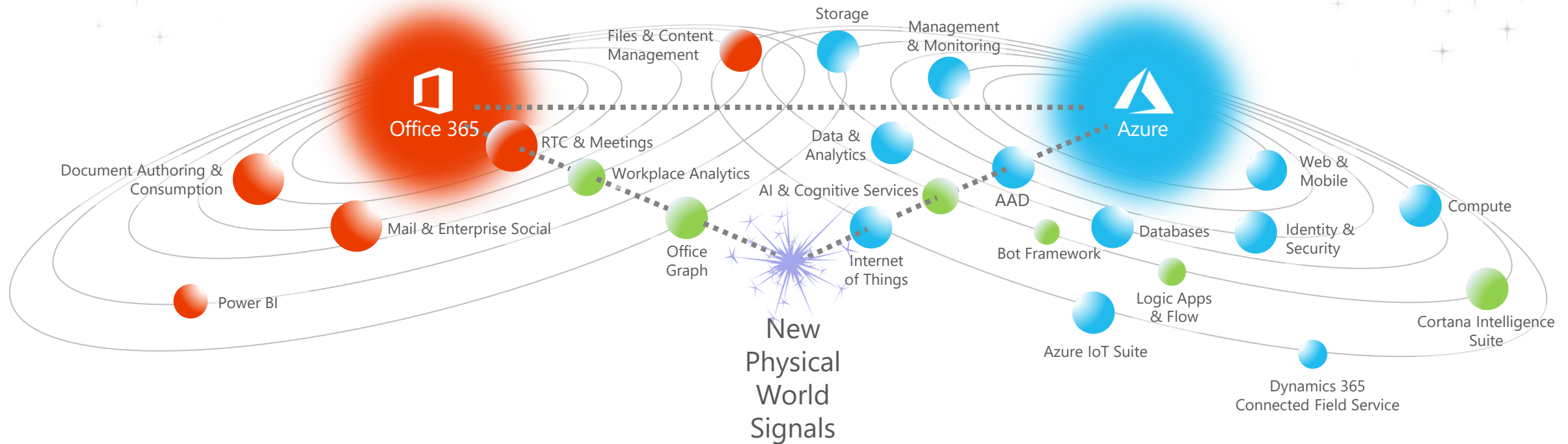
Digital transformation creates a virtuous cycle

- 1 Data: Capture digital signal across business
- 2 Insight: Connect and synthesize data
- 3 Action: Improve business outcomes

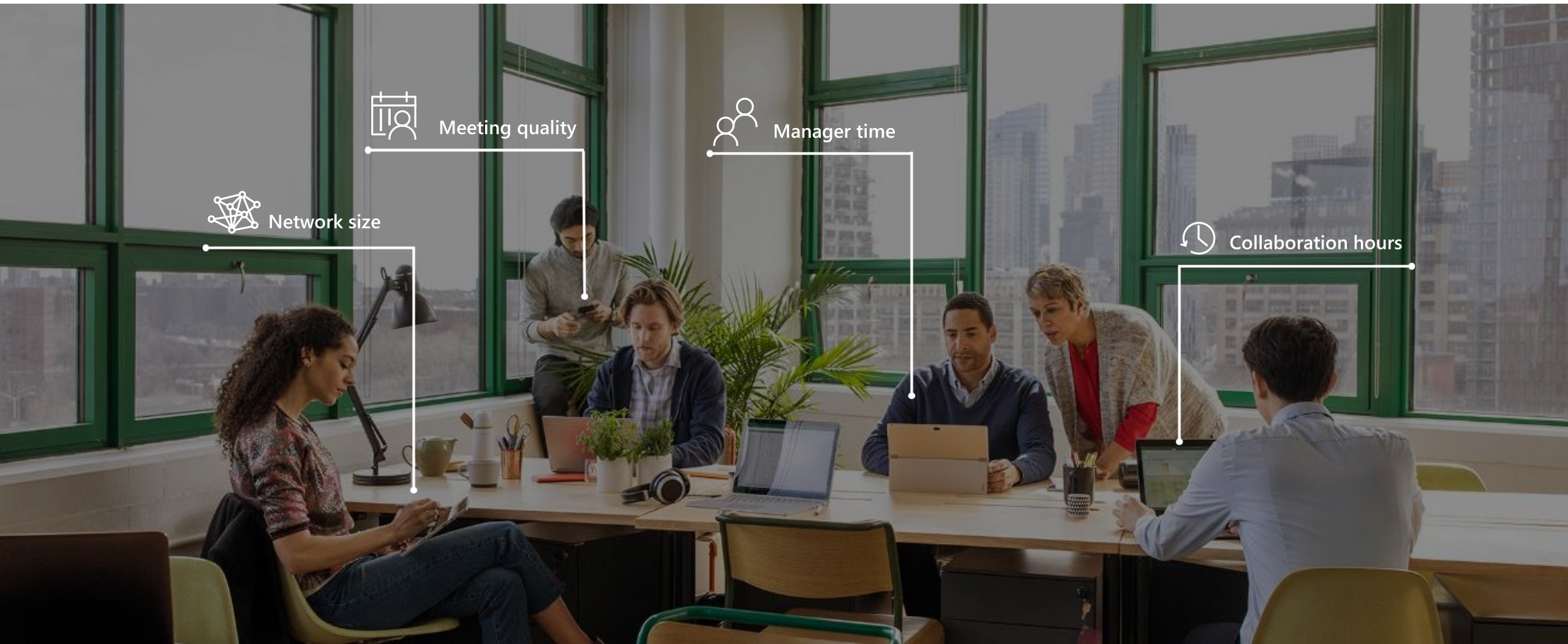


WORKPLACE 2.0 & IOT

As the physical world and digital worlds fuse into one, we have a historic opportunity to redefine the future of work.



With behavioral data, we can shift culture from soft to science through data-driven cultural transformation



How does it work?



Native Office 365 data

Start with aggregated behavioral data from everyday work in collaboration tools

+

Organizational context

Map behavioral data to organizational attributes to reveal collaboration patterns

+

Meaningful outcomes

Bring in other data sources to connect collaboration patterns to business outcomes

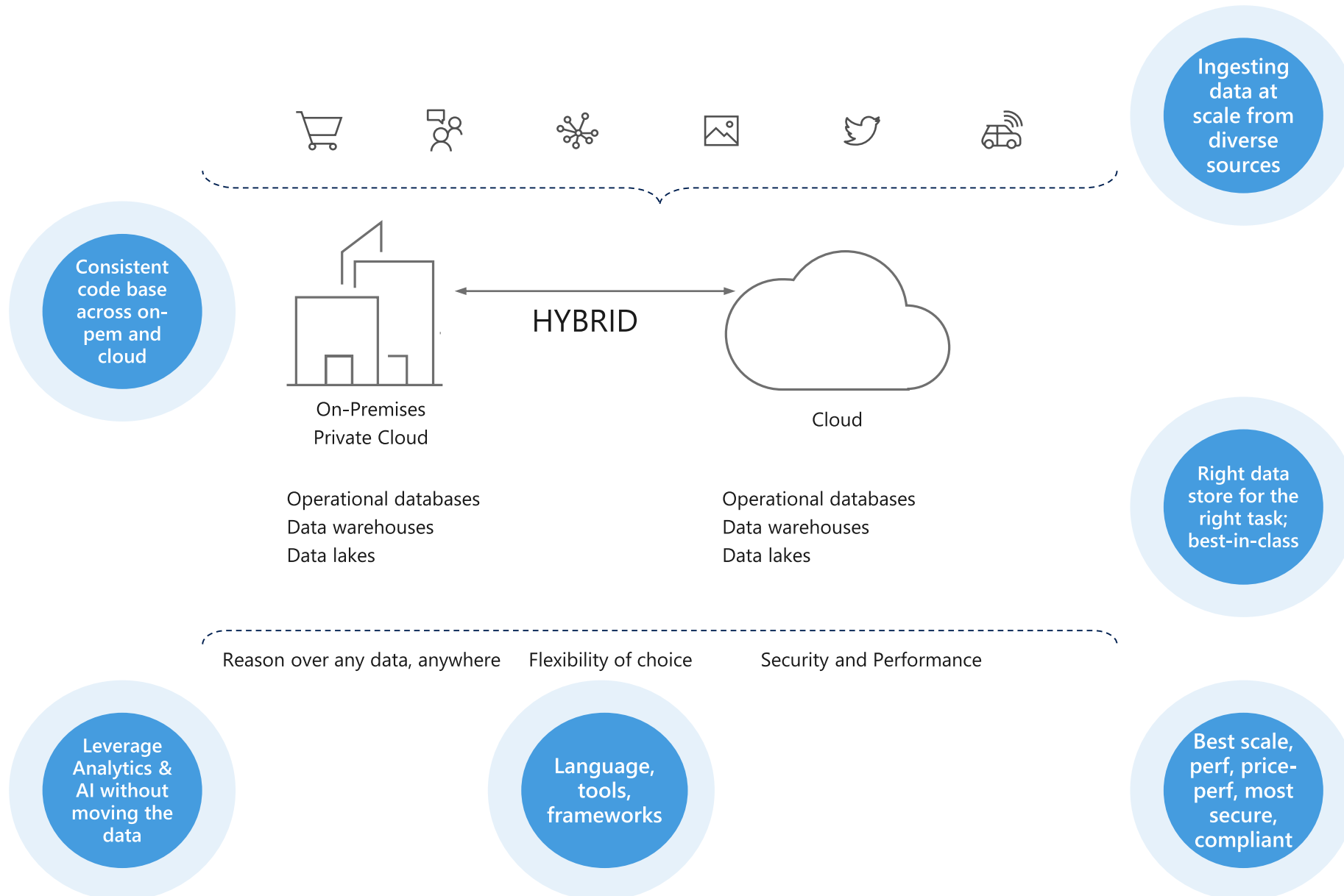
=

Actionable insights

Use powerful insights to make better business decisions and transform your organization

Workplace Analytics provides flexible privacy controls for data access and usage and compliance to help you protect your data.

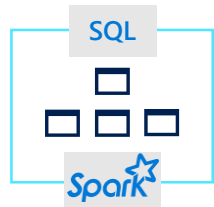
What is **The Modern Data Estate**?



Modernize on premises with SQL Server 2019

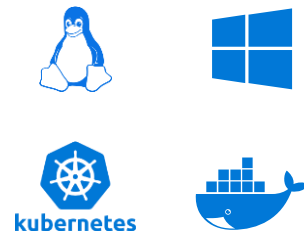
Now with big data clusters

Intelligence over any data



Analytics over structured and unstructured data with the power of SQL and Apache Spark

Choice of platform and language



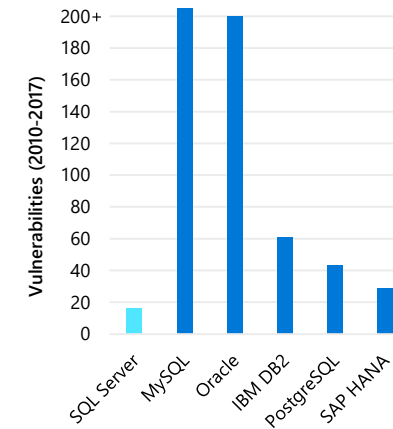
T-SQL	SCALA	Python
Java	Node.js	Ruby
C/C++	C#/VB.NET	.NET core

Industry-leading performance



#1 OLTP performance¹
#1 DW performance on 1TB², 10TB³, and 30TB⁴

Most secure over the last 8 years⁵



Insights in minutes and rich reports



The best of Power BI and SQL Server Reporting Services in Power BI Report Server



Private cloud

In-memory across all workloads

Most consistent data platform

1/10th the cost of Oracle



Public cloud

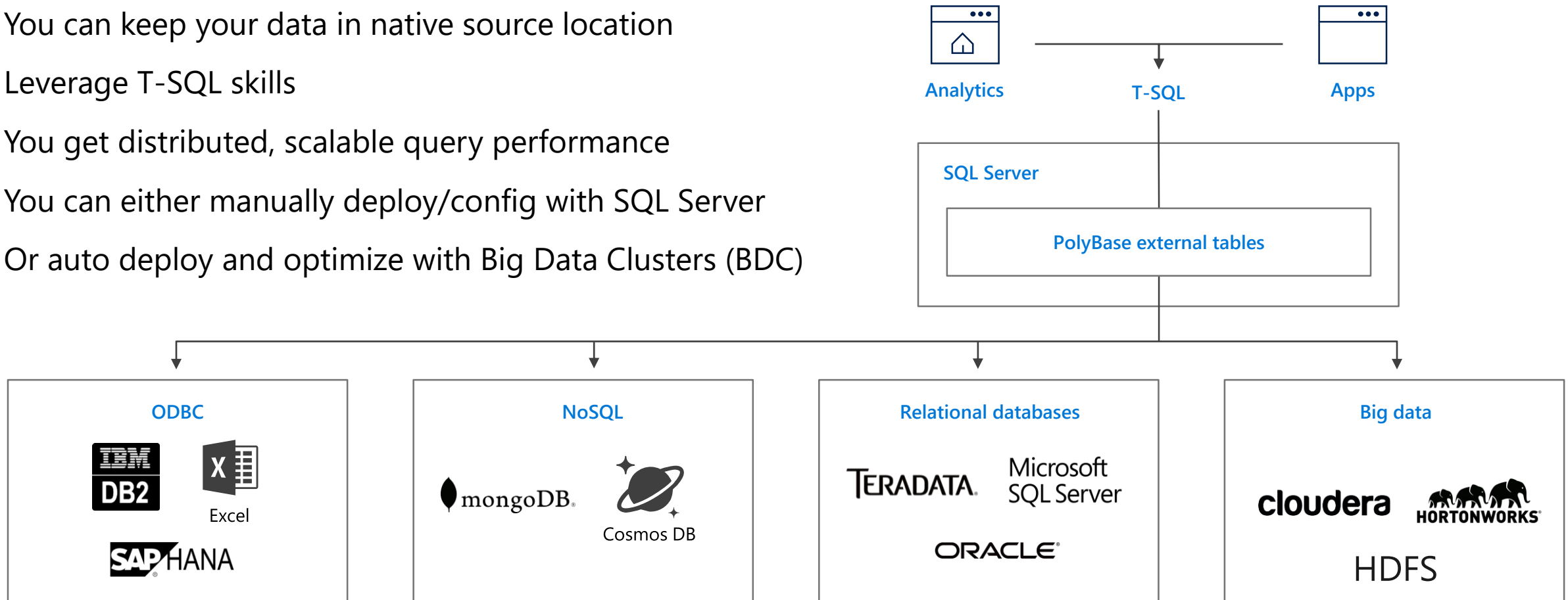
All TPC Claims as of 1/19/2018.

¹ <http://www.tpc.org/4081>; ² <http://www.tpc.org/3331>; ³ <http://www.tpc.org/3326>; ⁴ <http://www.tpc.org/3321>; ⁵ National Institute of Standards and Technology Comprehensive Vulnerability Database

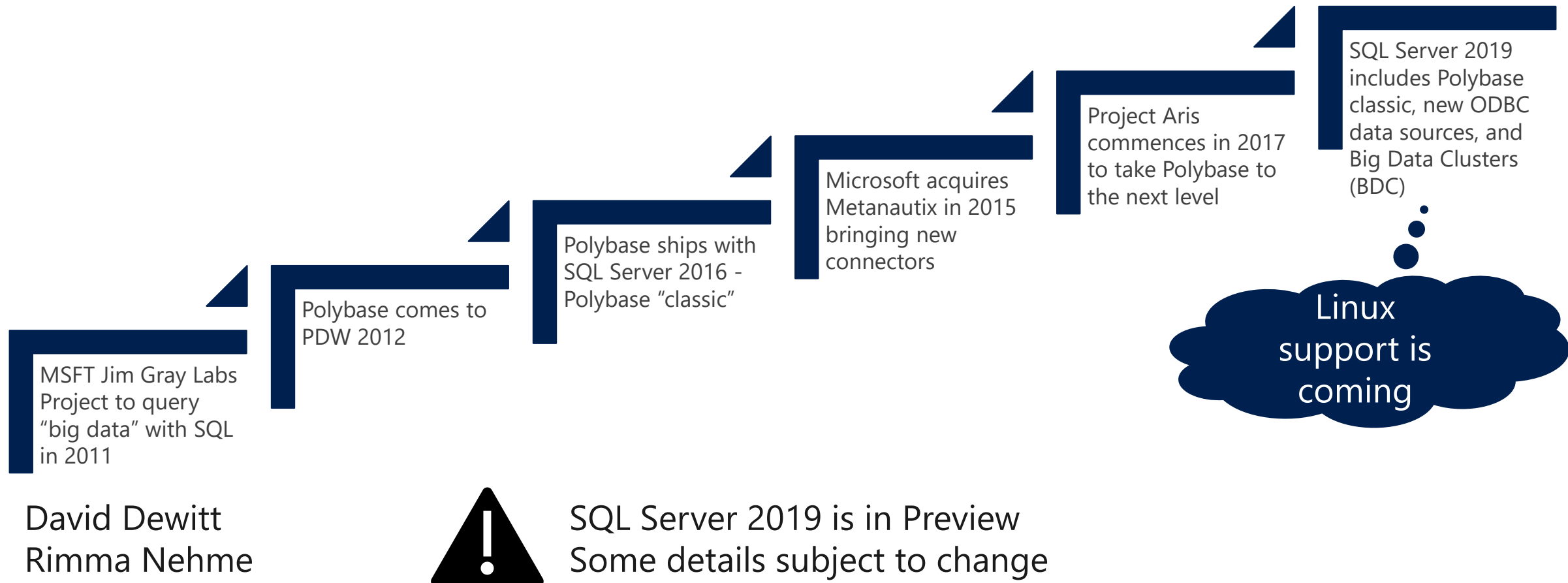
What is SQL Server Polybase?

“It’s all about
Data Virtualization”

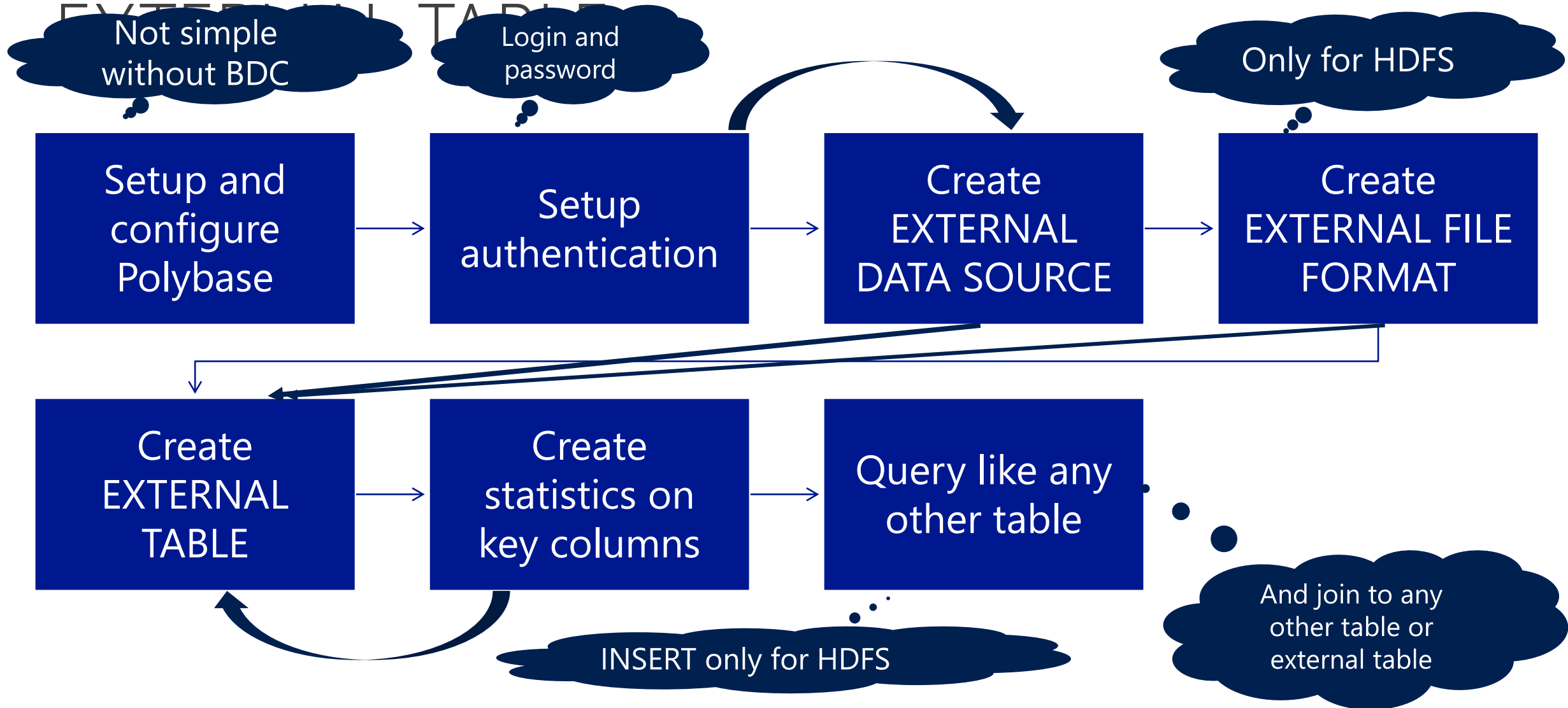
- ✓ A distributed compute engine integrated with SQL Server
- ✓ You can easily combine relational and non-relational data stores
- ✓ You can keep your data in native source location
- ✓ Leverage T-SQL skills
- ✓ You get distributed, scalable query performance
- ✓ You can either manually deploy/config with SQL Server
- ✓ Or auto deploy and optimize with Big Data Clusters (BDC)



The Journey to Data Virtualization in SQL Server 2019



Using Polybase in SQL Server: T-SQL



The SQL Server Polybase Architecture

control and execution

Data flow

EE only

All editions

All editions

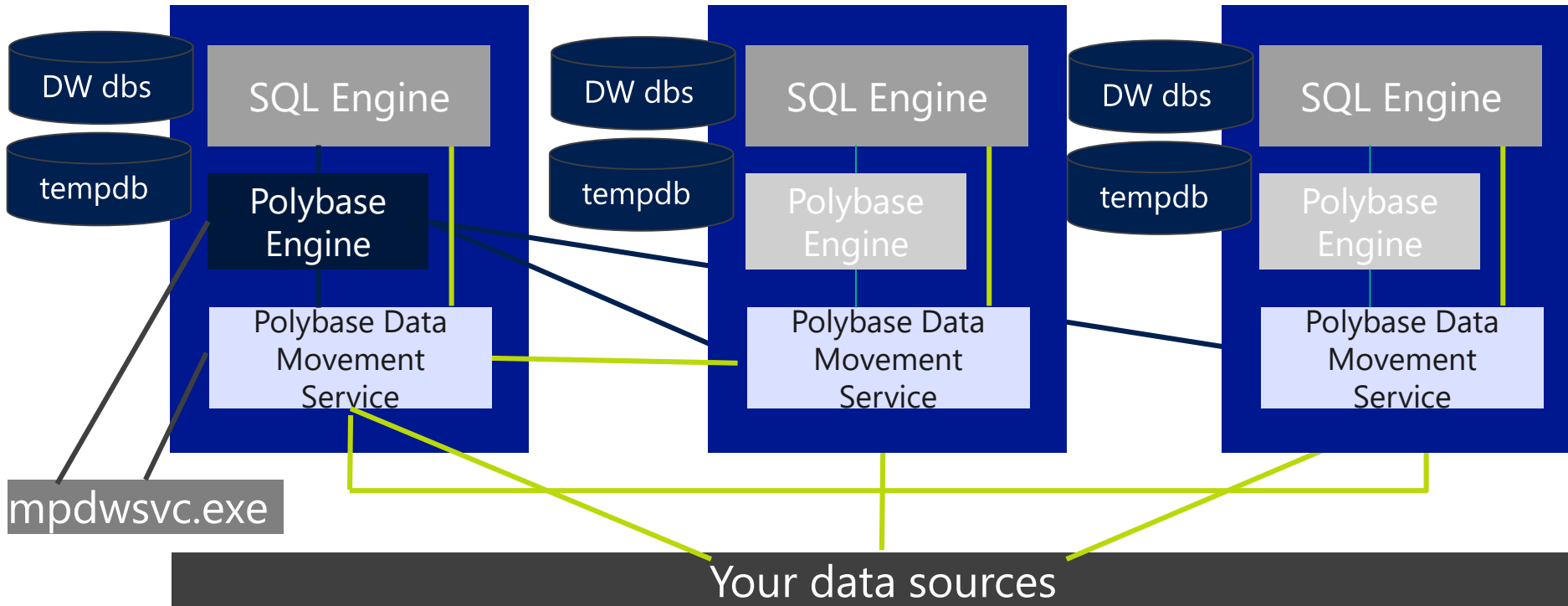
"Head" node

"Compute" node

"Compute" node

Need more
scale?
Add compute
nodes

Scan or pushdown
Partition and shuffle



SAP HANA

Cosmos DB

TERADATA

Microsoft
SQL Server

ORACLE

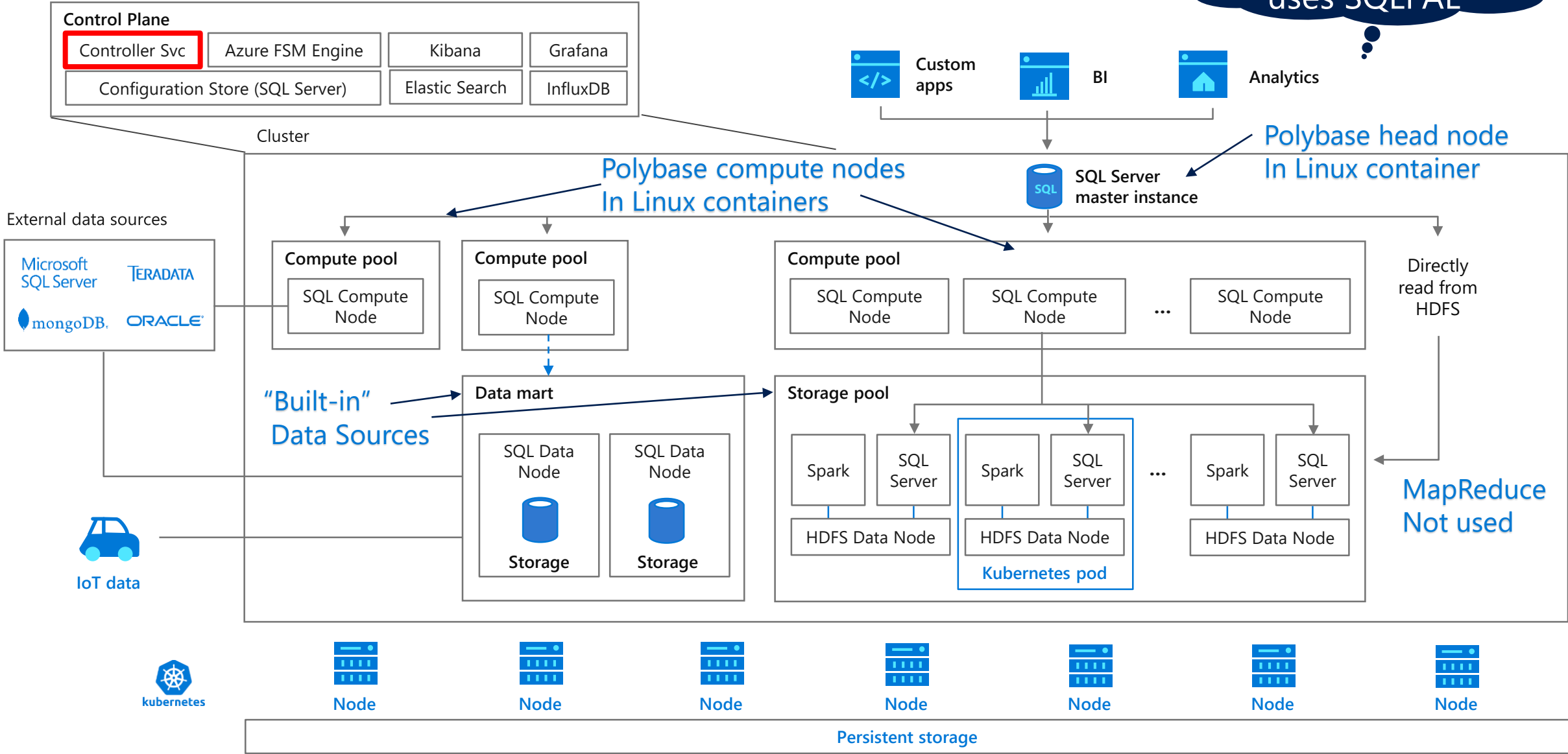
cloudera

HORTONWORKS

HDFS

Polybase in SQL Server 2019 Big Data Clusters

mpdwsvc.exe uses SQLPAL



In principle, IoT solutions are easy...



Things
(Generate Data)

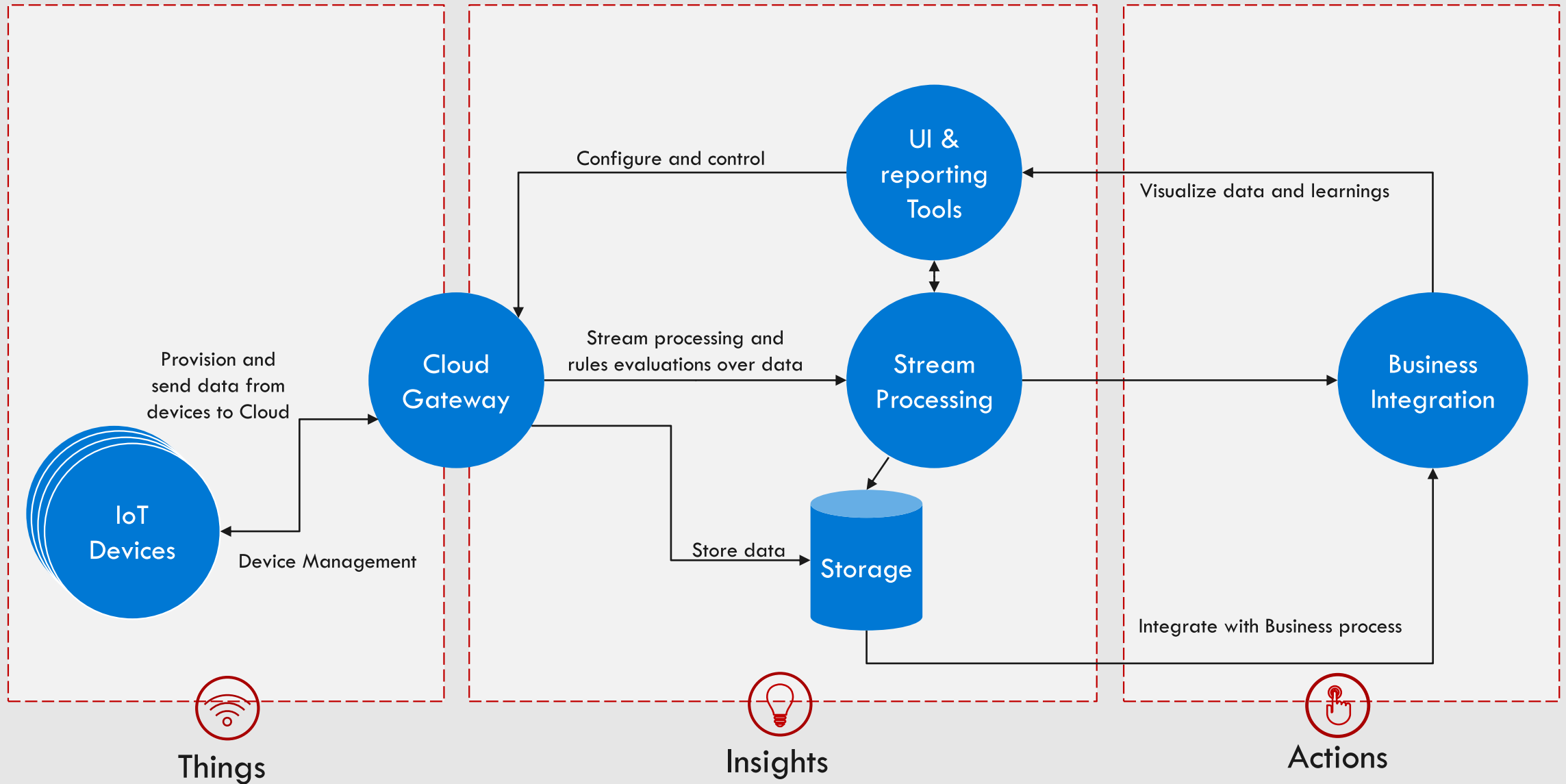


Insights
(Based on Data)

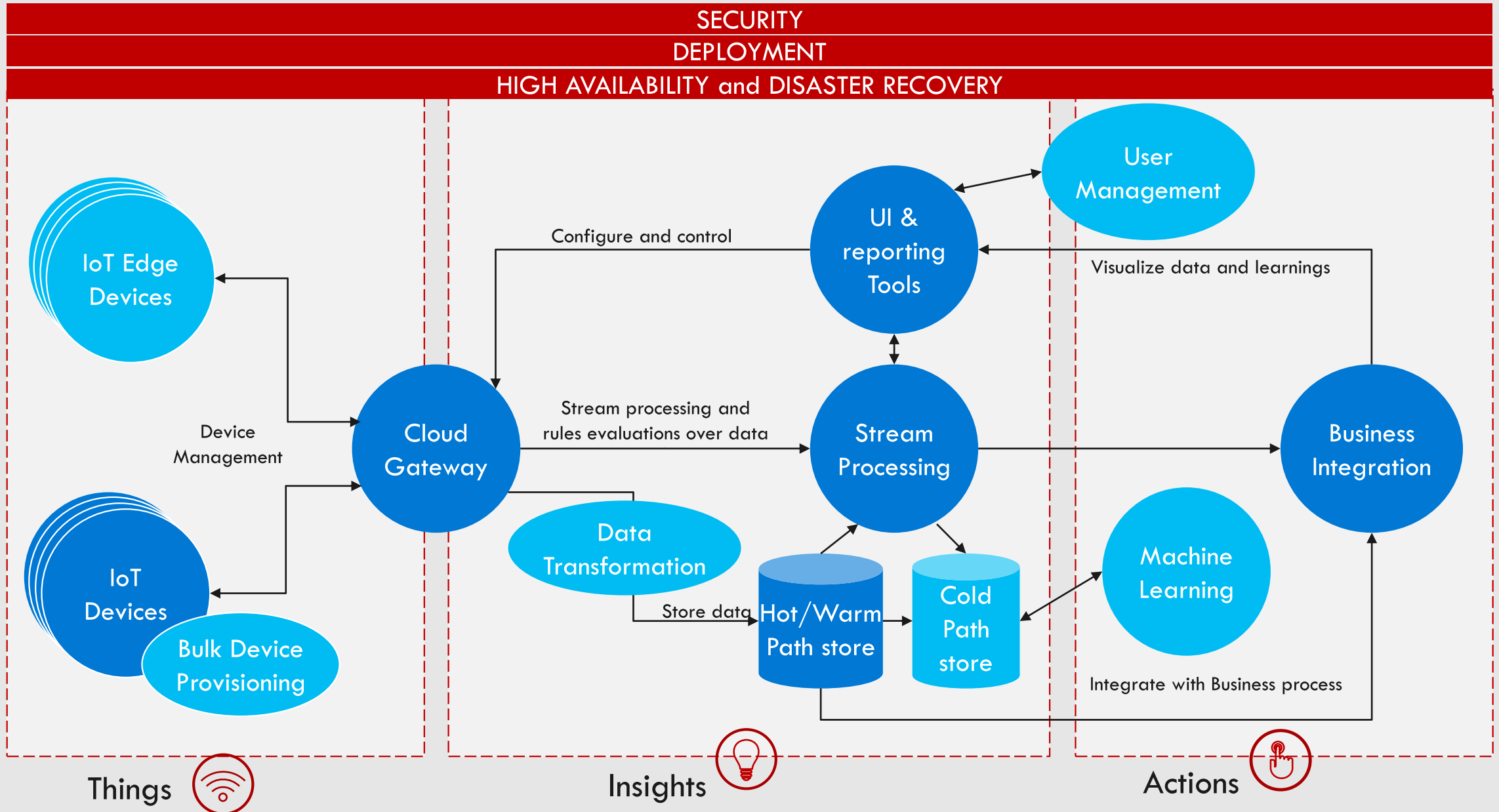


Actions
(Based on Insights)

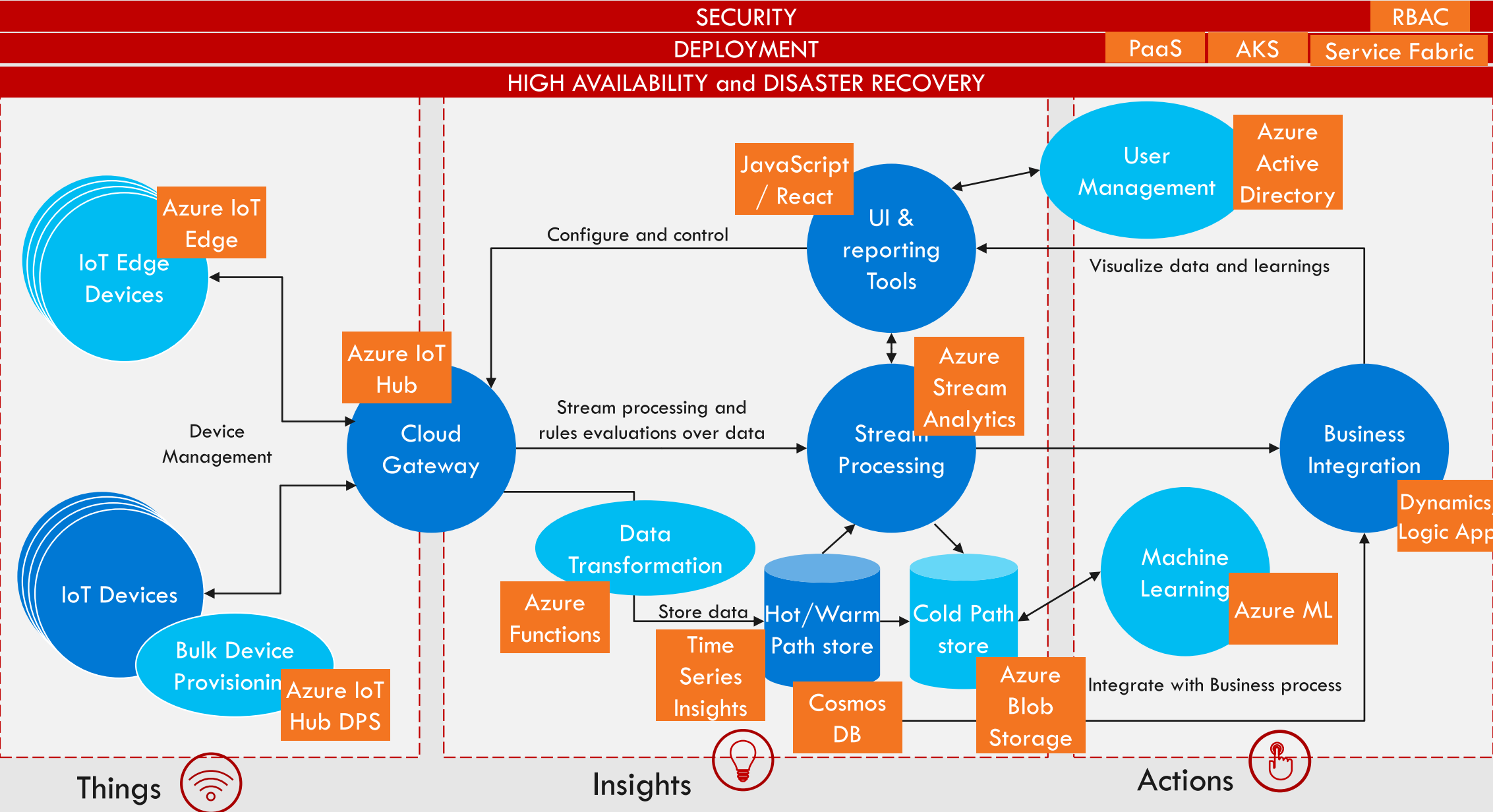
Core Subsystems



All Subsystems & Cross Cutting Needs



Tech choices

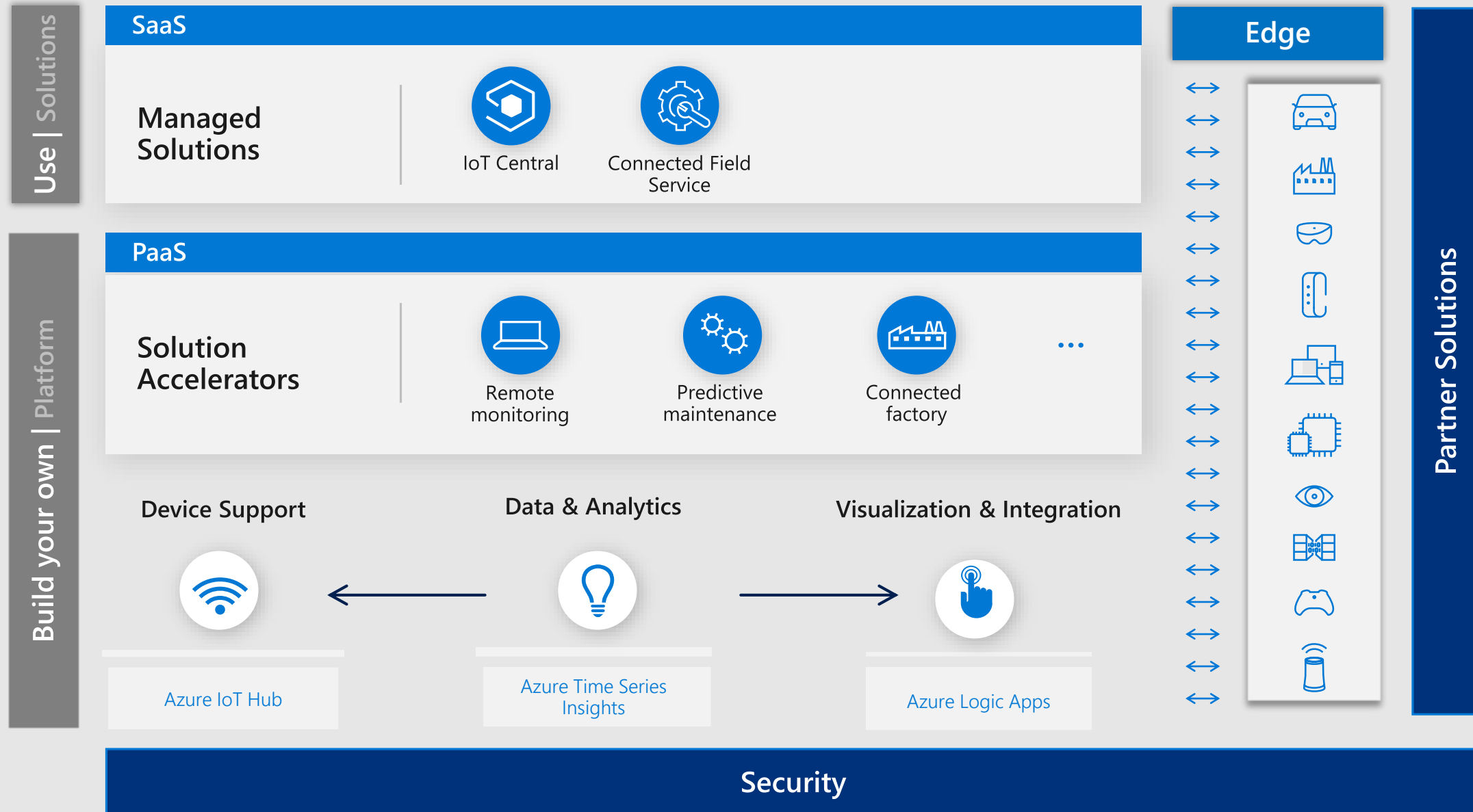


Azure IoT Reference Architecture

<https://aka.ms/iotrefarchitecture>

<https://iotschool.microsoft.com>

The Azure Portfolio of offers for IoT



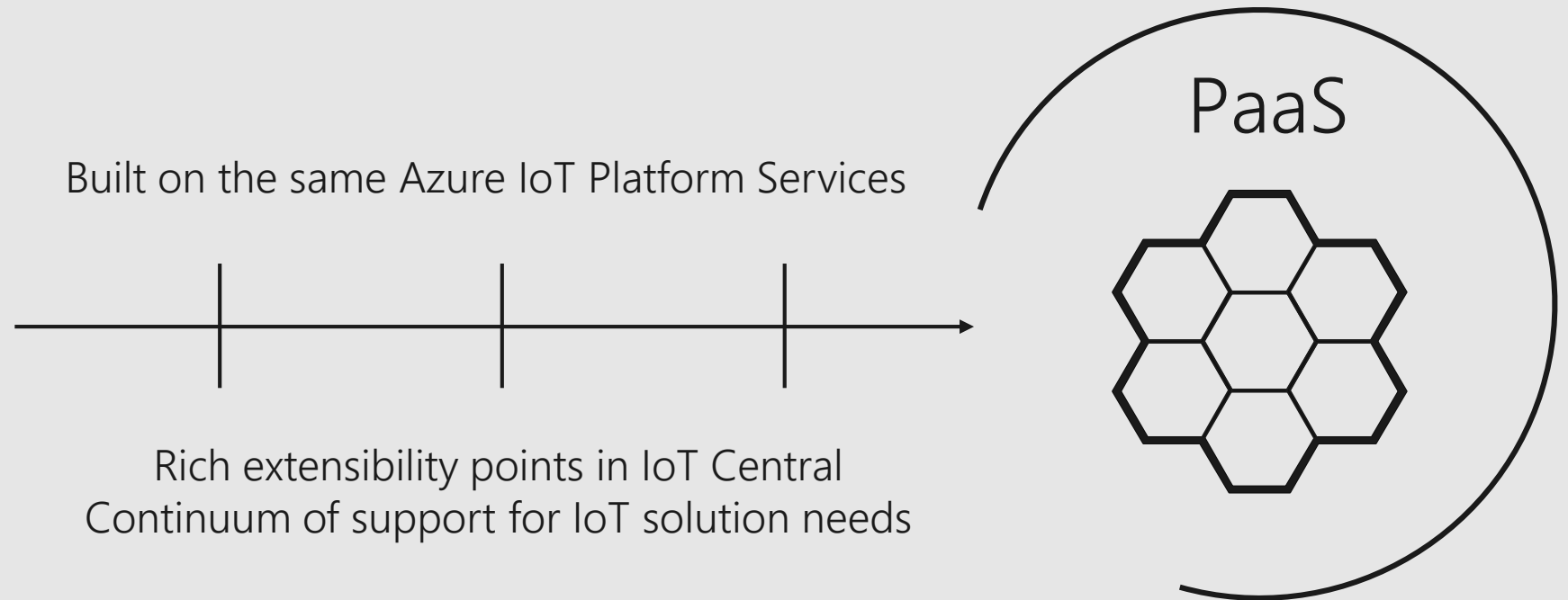
Azure IoT Central

Fully managed IoT SaaS
No cloud solution expertise required



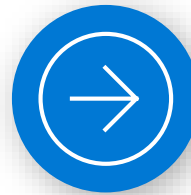
Azure IoT Solution Accelerators

Solution accelerators for customers with cloud
solution expertise and the need to fully customize



Accelerate time to value

Start quickly for
common IoT scenarios



Finish with your
IoT application



Get started in minutes

Modify existing rules and alerts

Add your devices and begin tailor to your needs

Fine-tuned to specific assets and processes

Highly visual for your real-time operational data

Integrate with back-end systems

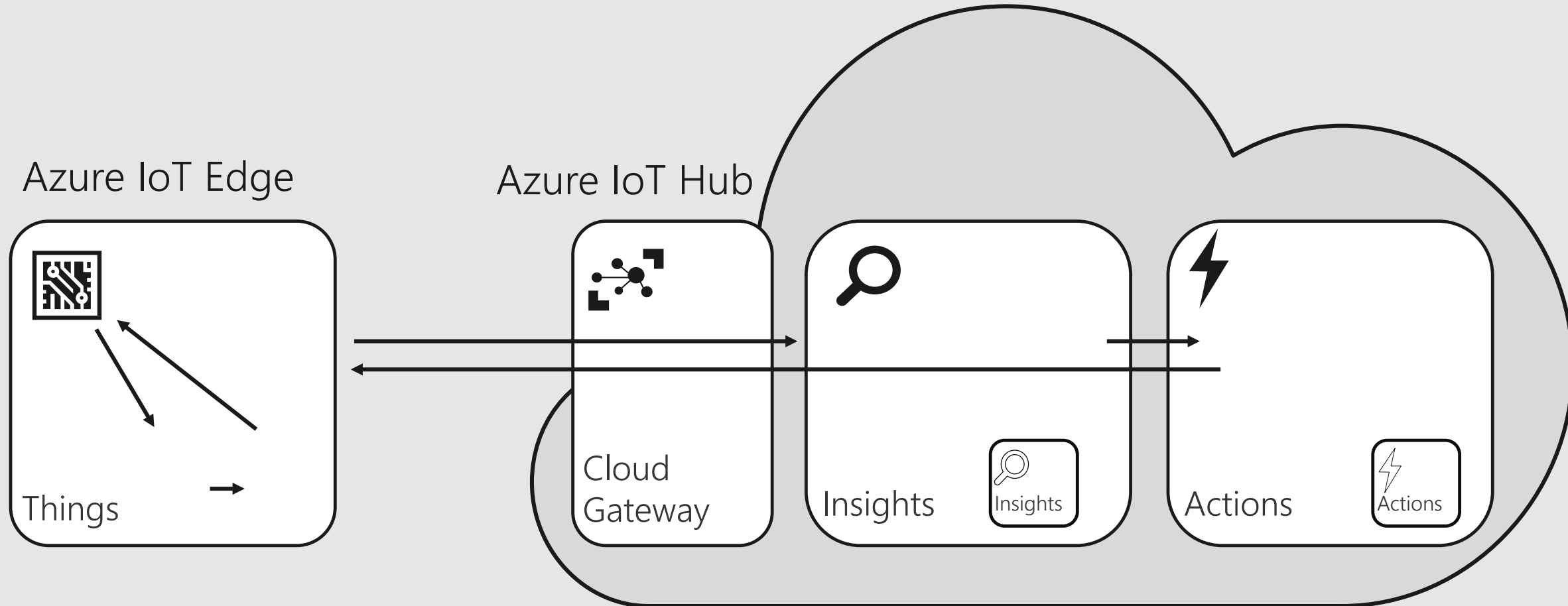
Prediktor Spektron D20



Online NIR Spectroscopy
& Realtime Cloud Analytics



IoT Pattern + Edge



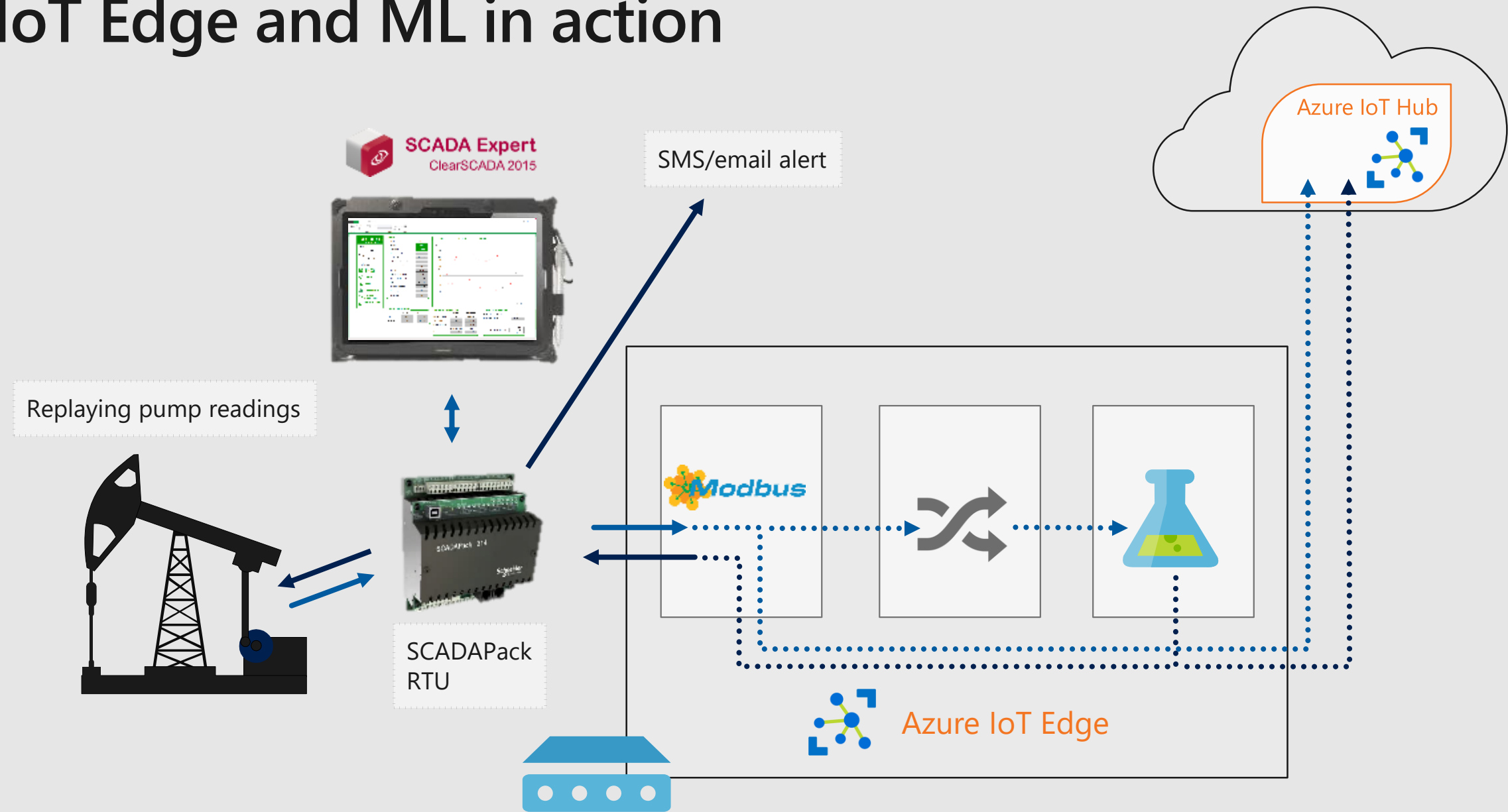


Edge analytics makes remote asset management easier

Schneider Electric was interested in enhancing their remote management solution to do more than just respond to real-time situations. It has now the flexibility to run machine learning in the cloud or at the edge.

[LEARN MORE >](#)

IoT Edge and ML in action



Shipping soon

Azure IoT Edge Vision AI Developer Kit



Purchase now: <https://visionaidevkit.com/>



The Evolution of IoT

- Early IoT solutions were based on lists of devices connected to Azure and monitoring them or predicting their maintenance needs
- We are seeing the beginning of a new trend in IoT solutions
- Customers want to model a physical environment first, and then keep the model up to date with IoT data
- We're finding this is a more natural approach for building IoT solutions



Azure Digital Twins

Public Preview releases
after Ignite (Oct '18)
Expected GA: Summer 2019

Use a digital twin to build solutions that transform spaces into smart environments

The digital twin graph enables organizations to virtually represent their physical world by modelling the relationships between people, places and devices.

Accelerate and simplify the creation of smart spaces solutions and securely scale them across multiple tenants

Use pre-defined device protocols and data schema aligned to a solution's needs to accelerate and simplify the creation of smart spaces solutions.

Multi and nested tenancy allows built solutions to be securely deployed across tenants without needing to re-build solutions.

Customize functions that generate actions in a physical space to improve efficiency and occupant experience

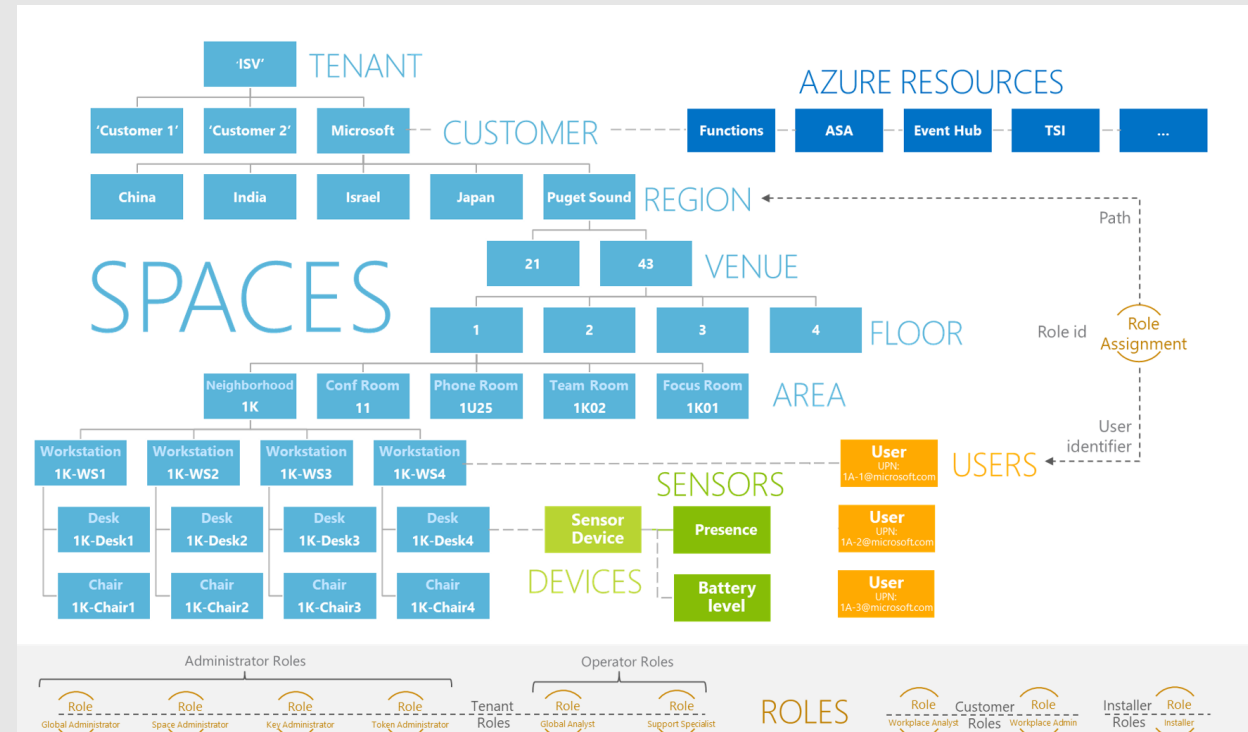
Define and execute functions that can generate notifications or events to specific endpoints within a space based on incoming telemetry messages.

[Documentation](#)

[Product Page](#)

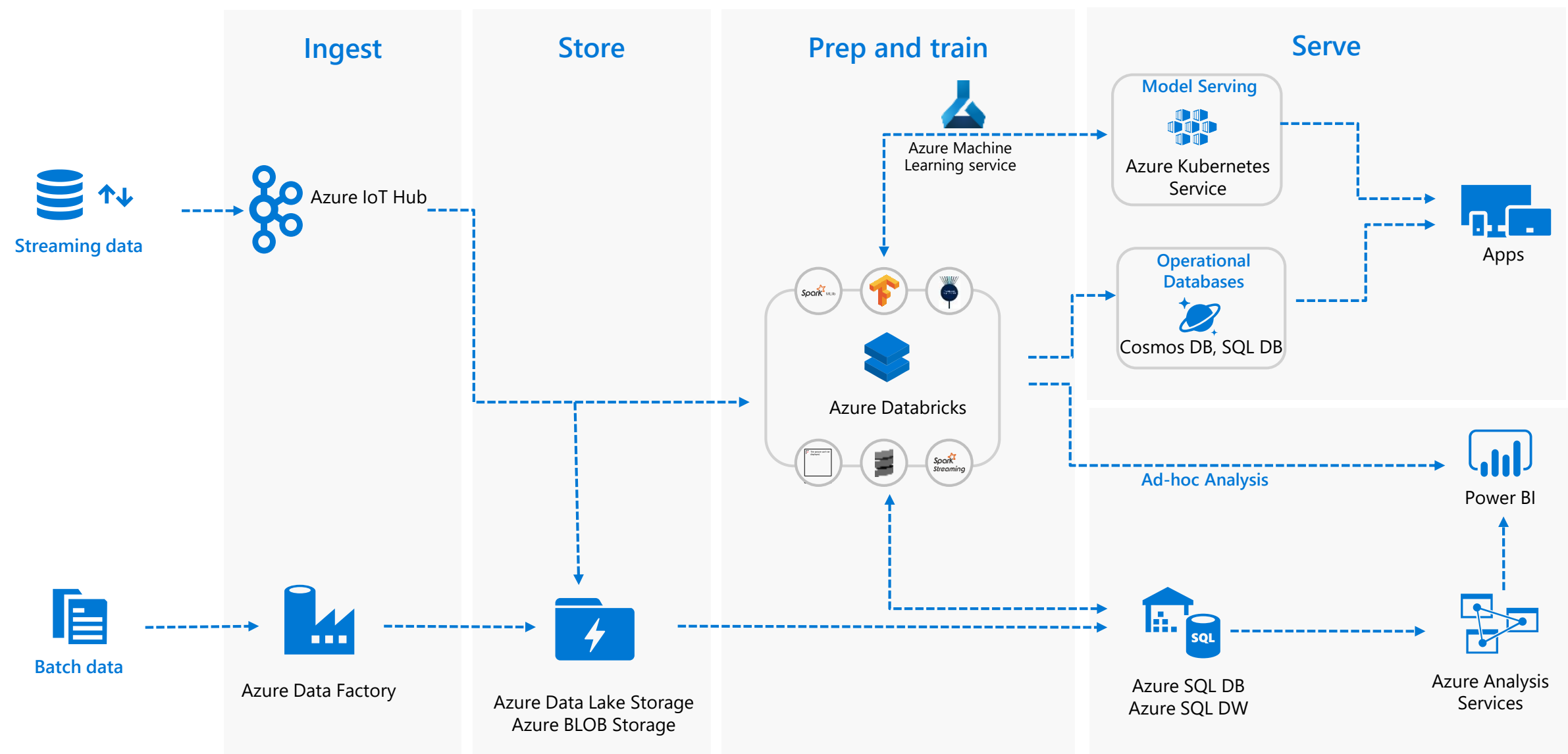
Channel 9 [Video](#)

<https://aka.ms/digitaltwinsblog>

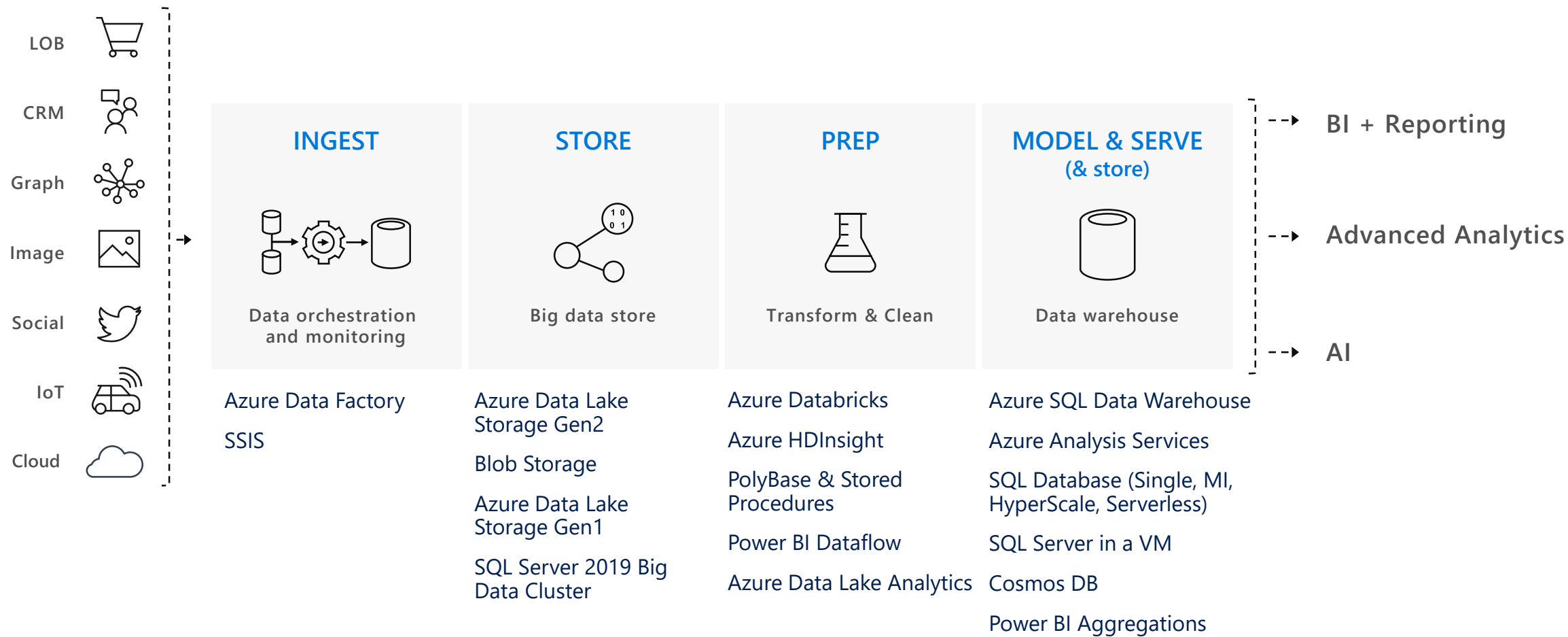


- An illustrative example of a digital twin graph used in an ISV solution that enables a virtual representation the physical world.
- In the case shown the solution is a Smart Building solution and includes customers, regions, buildings, floors and other topography layers.

Example architecture for e2e ML solutions



Modern Data Warehouse (possible products by four areas)



Note: Those products that span more than one area are listed in there primary area

MODERN DATA WAREHOUSING



AZURE DATA FACTORY



Ingest from 75+ sources,
up to 1 Gbps



Hybrid data integration
with drag-n-drop UI



Execute SQL Server
Integration Services packages



AZURE DATABRICKS



Fast, easy, collaborative Apache®
Spark™-based analytics platform



Native integration
with Azure services



Enterprise-grade Azure security



AZURE SQL DATA WAREHOUSE



Unlimited scale and performance



Flexibility to fit your needs



Secure, trusted and compliant



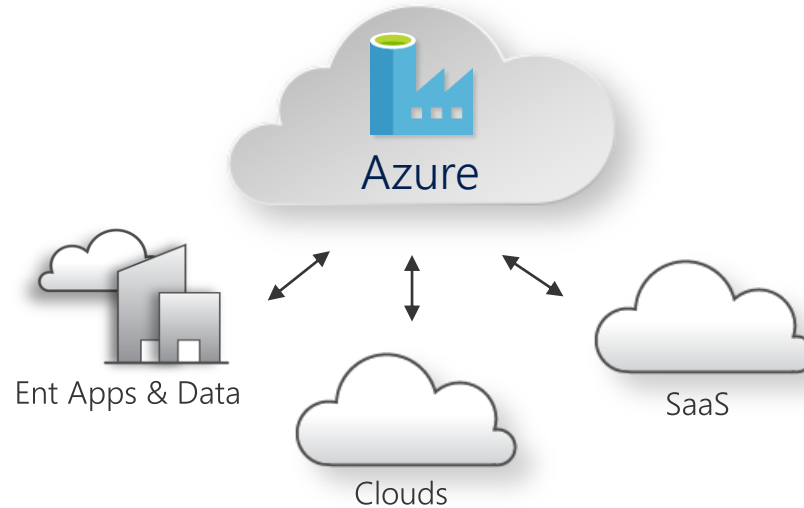
AZURE DATA LAKE STORAGE

GEN 2

- ✓ Combines Hierarchical File and Object Store
- ✓ Optimized for Analytics Workloads
- ✓ Azure Active Directory
- ✓ Exabyte Scale

Azure Data Factory

Data Integration Service: Serverless, Scalable, Hybrid



Hybrid Pipeline Model

Seamlessly span: on prem, Azure, other clouds & SaaS
Run on-demand, scheduled, data-availability or on event

SSIS Package Execution

Lift existing SQL Server ETL to Azure
Use existing tools (SSMS, SSDT)

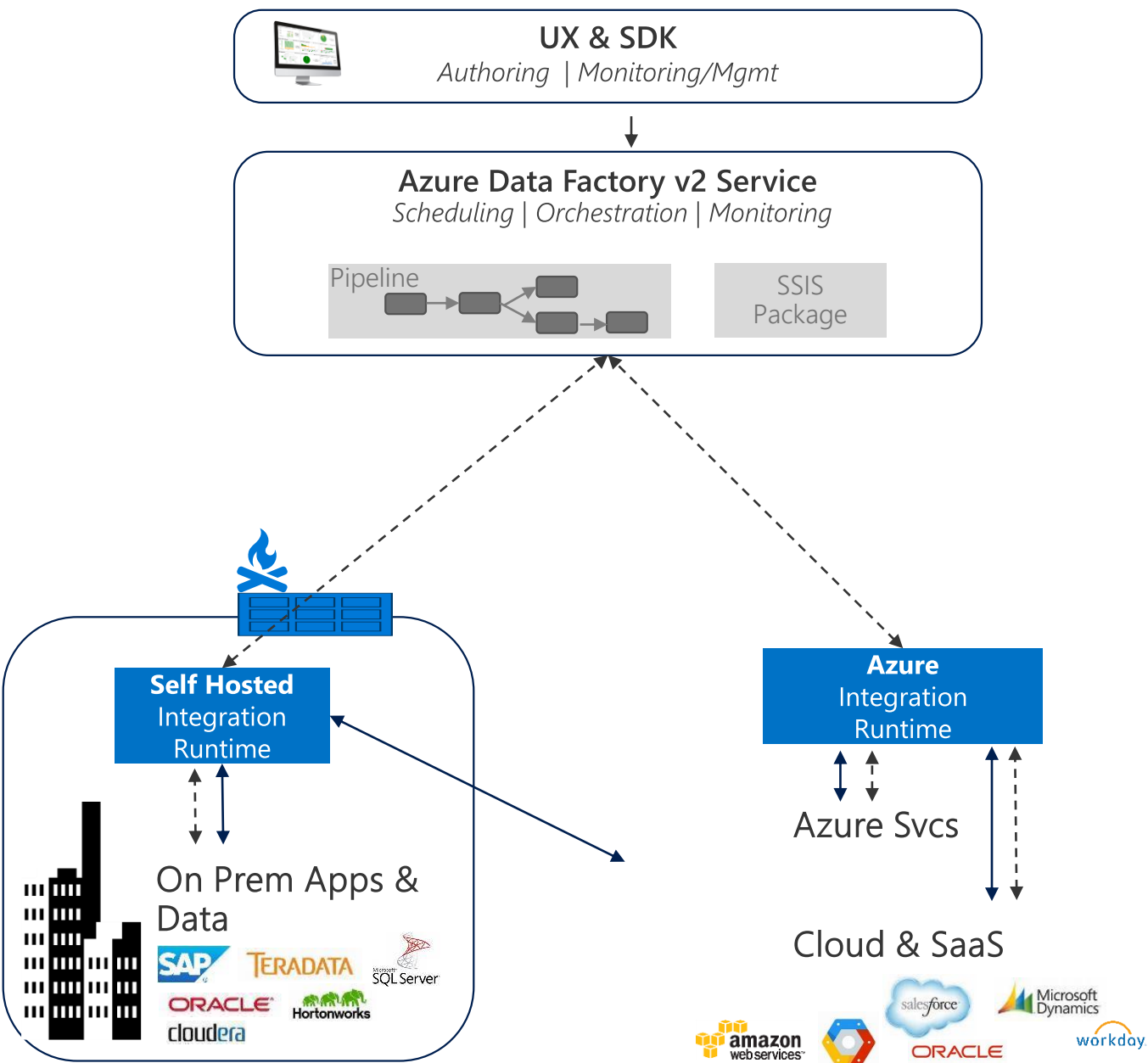
Data Movement @Scale

Cloud & Hybrid w/ 75+ connectors provided
Up to 1 GB/s

Author & Monitor

Programmability w/ multi-language SDK
Visual Tools

←---→ Command and Control
←→ Data



Data Factory

A data integration account.
Location of orchestration, service metadata

Integration Runtime (IR)

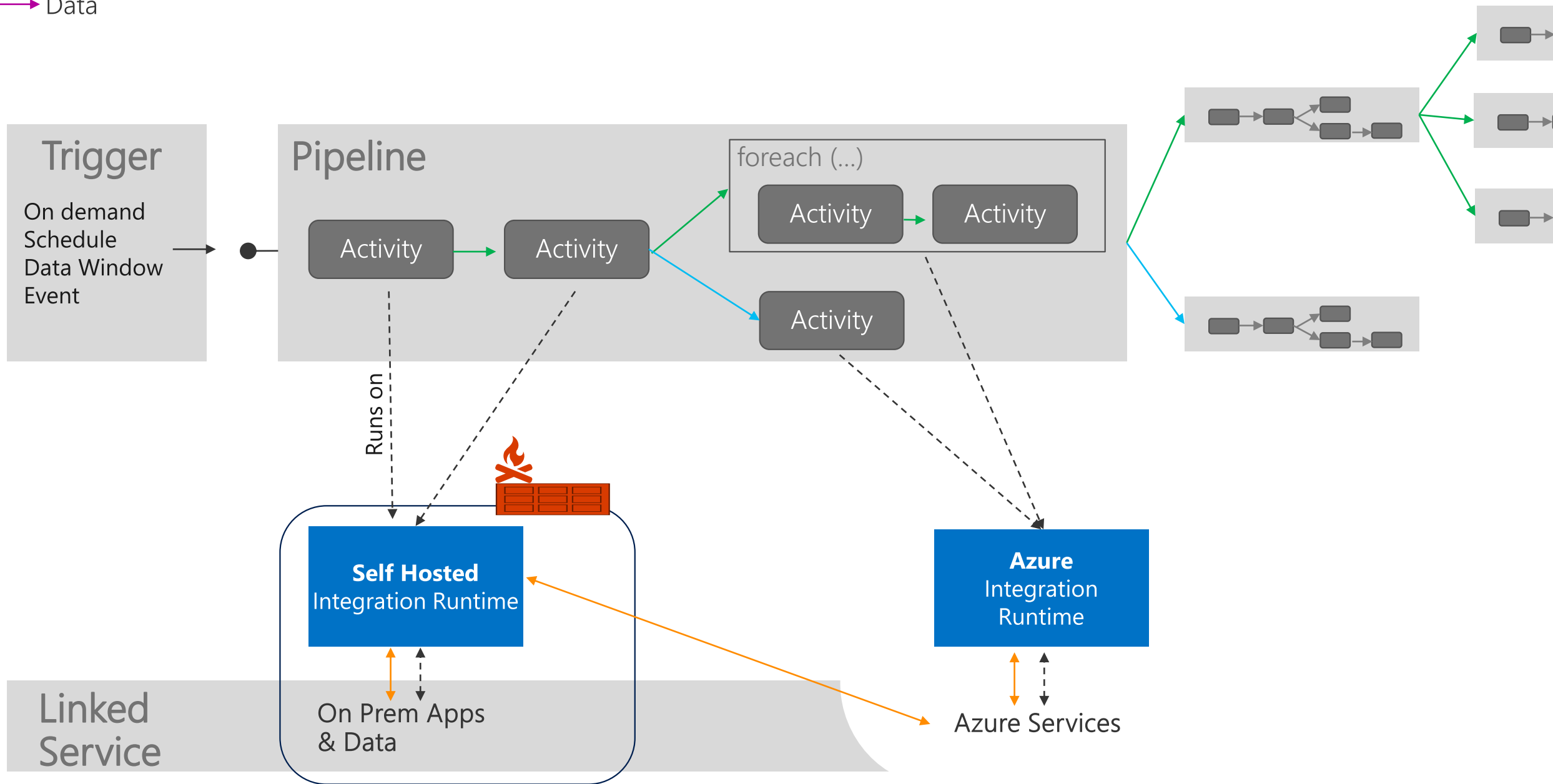
ADF's execution engine

Three core capabilities:

- data movement
- pipeline activity execution
- SSIS package execution

←---→ Command and Control

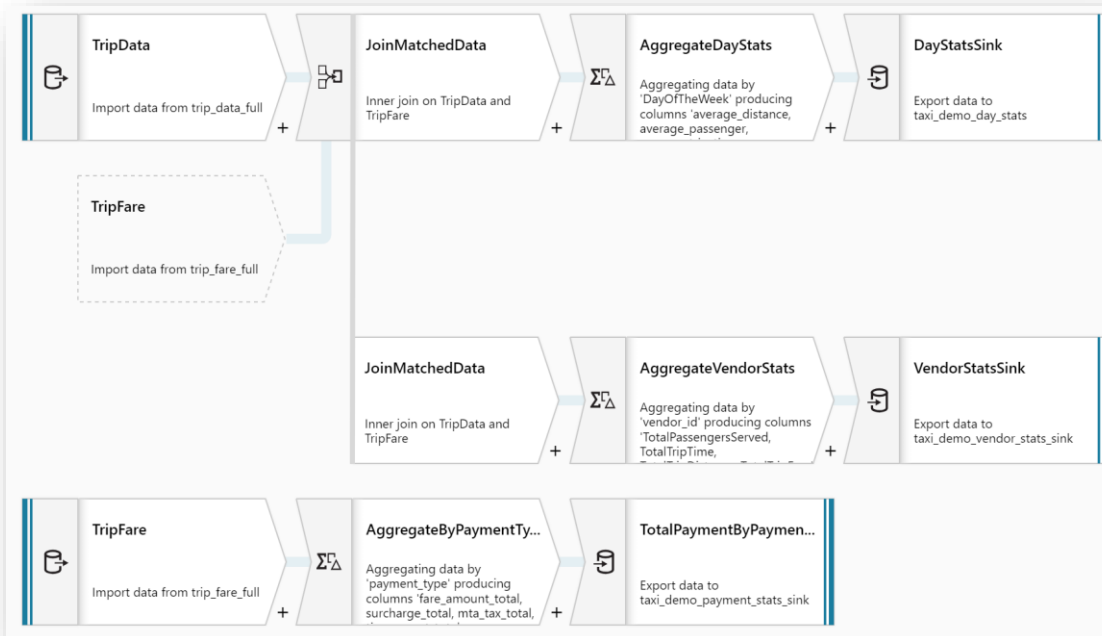
↔ Data



Mapping Dataflow

No Code Data Transformation @Scale

- Data cleansing, transformation, aggregation, conversion, etc.
- Cloud scale via Spark execution
- Easily build resilient data flows



... not

```
MovieRecommender2Demo.txt
1
2 HDI Cluster Details:
3 Adfhd1.azurehdinsight.net
4 Admin
5 Adf@123456
6
7 Storage:
8 adfhd1storage
9 /anyPw661j7T811BmmlSo/YGdJyGT4d+51Ja+5m57bJg954706uUOK0ksZ19UXso40x2o8x1K9hd9WwQ==
10
11 Cluster Remote Login Details:
12 Adf
13 India@1234
14
15 HiveQuery:
16 DROP TABLE IF EXISTS MovieRatings;
17 CREATE EXTERNAL TABLE MovieRatings
18 (
19     UserID int,
20     MovieID int,
21     Rating int,
22     Timestamp string
23 ) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE LOCATION '${hiveconf:MovieRatings}';
24
25 DROP TABLE IF EXISTS MovieTitles;
26 CREATE EXTERNAL TABLE MovieTitles
27 (
28     MovieID int,
29     MovieName string
30 ) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE LOCATION '${hiveconf:MovieTitles}';
31
```


Guided experience to build data flows

Microsoft Azure | Data Factory | datafloweast

Search resources

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

Factory Resources

- Pipelines: 21
- Datasets: 73
- Data Flows (Preview): 26

cleanAddresses

dataflow2

dedupe

flightDF

PivotProductCostsbyColor

regexflow

Baseball: 3

Bug Bashes: 2

Demos: 11

Misc: 2

Params: 1

Soccer: 1

Connections

Triggers

soccerETL

Debug | Validate

Code

SpecifySchemaExtracts

Import data from soccer_events

DictionaryMapping1

Creating/updating the columns 'id_odsp, id_event, sort_order, time, text, event_type, event_type2, side, event_team,

JoinStringDataViaLookup

Columns: 25 total

ColumnSelectionNaming

Renaming JoinStringDataViaLookup to ColumnSelectionNaming with columns id_odsp, id_event

TimeBins

Aggregates data based on a window and joins with original data

Lookup Settings

Optimize

Inspect

Data Preview

Output stream name * JoinStringDataViaLookup

Primary stream * DictionaryMapping1

Lookup stream * locationMap

Lookup conditions * Left: DictionaryMapping1's column abc location == Right: location abc Type

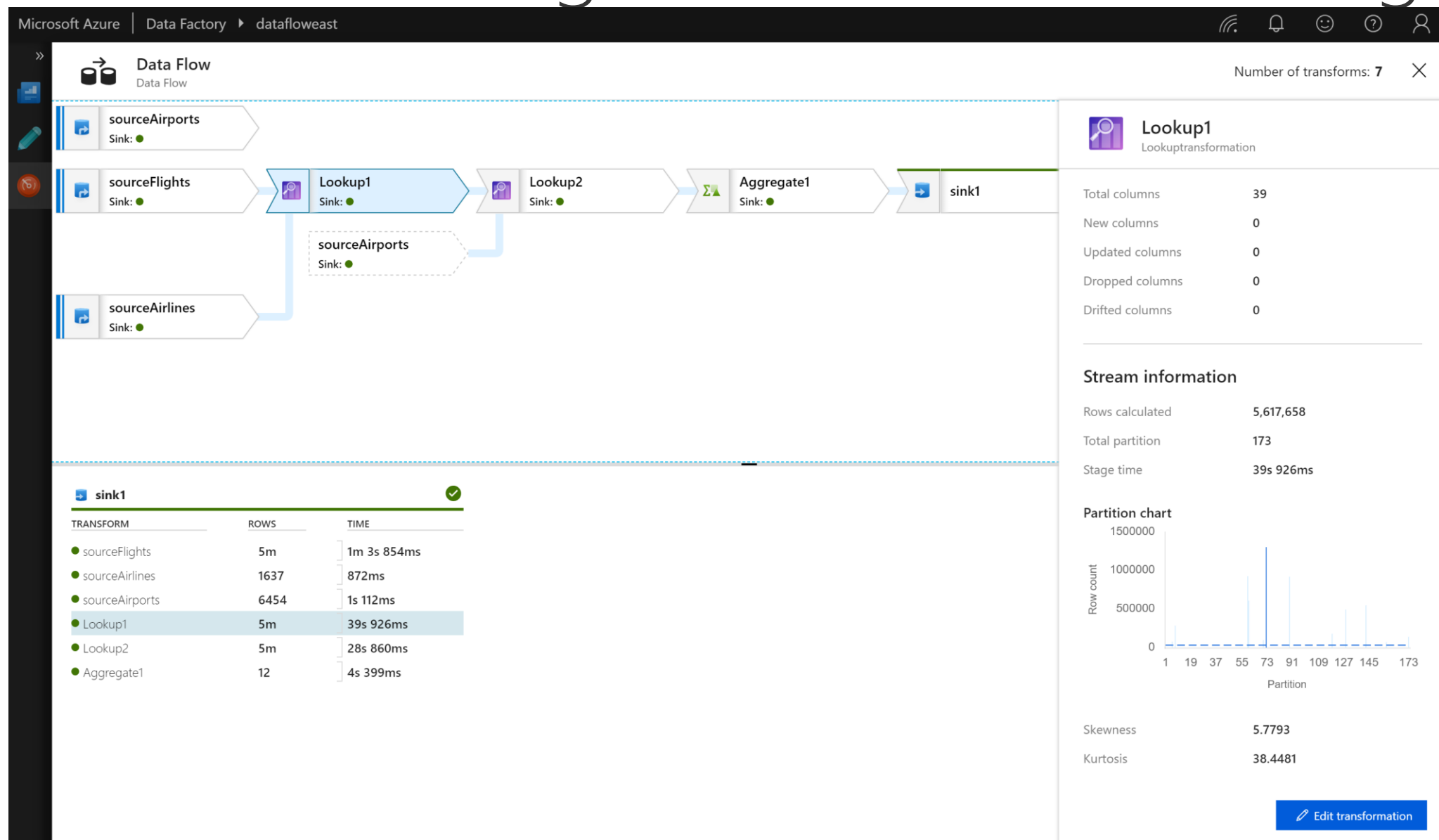
Multiple inputs/outputs

- New Branch
- Join
- Conditional Split
- Union
- Lookup

Schema modifier

- Derived Column
- Aggregate
- Surrogate Key
- Pivot
- Unpivot

Monitor & Manage Dataflows from a single



AI Platform

AI SERVICES

CONVERSATIONAL AI

Azure Bot Service

TRAINED SERVICES

Cognitive Services

CUSTOM SERVICES

Azure Machine Learning

AI INFRASTRUCTURE

AI ON DATA

Cosmos
DB

SQL
DB

SQL
DW

Data
Lake

AI COMPUTE

Spark

DSVM

Batch
AI

ACS

Edge

CPU, GPU, FPGA

AI TOOLS

CODING & MANAGEMENT TOOLS

VS/VS Code
Tools for AI

Azure ML
Studio

Azure ML
Workbench

Others (Pycharm, Jupyter notebooks...)

DEEP LEARNING FRAMEWORKS

Cognitive
Toolkit

TensorFlow

Caffe

Others (Scikit-learn, MXNet, Keras, ...)

Open AI Ecosystem

Gluon


ONNX



AI Lab

AI.lab

Experience, learn and code the latest
breakthrough AI innovations by Microsoft.

 Scroll down to view projects

ailab.microsoft.com

Solution gallery

Getting started easily with solution templates for the most common AI Enterprise scenarios.

Reusable architectures

Common business scenarios: predictive maintenance, fraud detection, demand forecasting, personalized offerings...

Customize and deploy

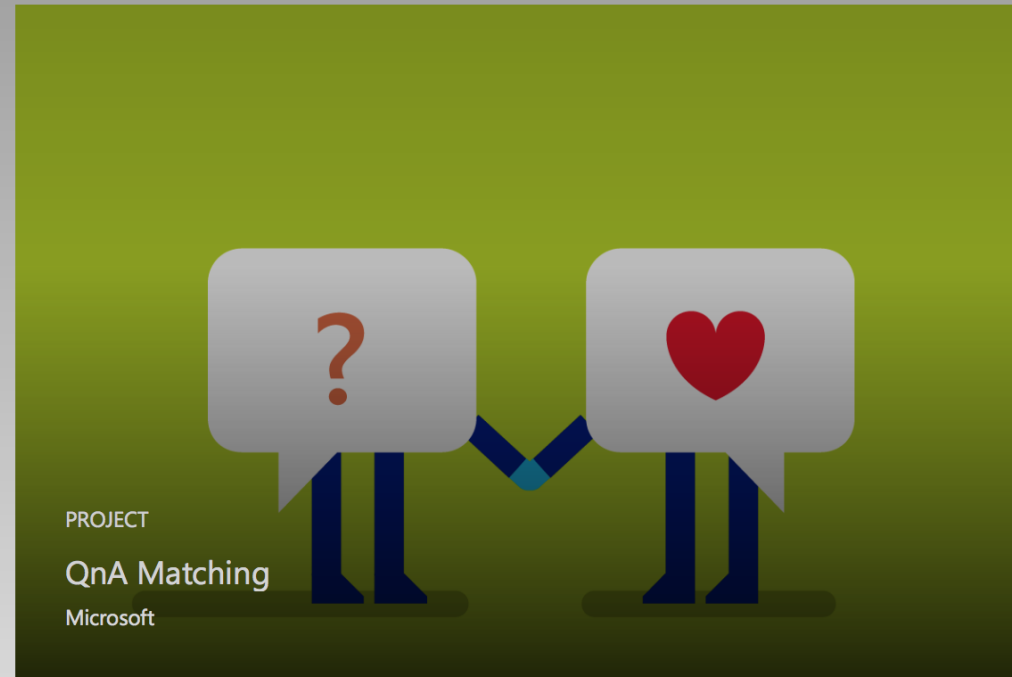
Models and architectures to get started, operationalize in one click.

Trusted

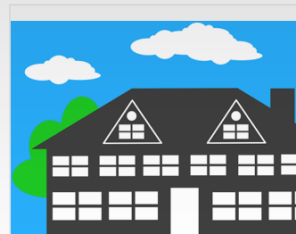
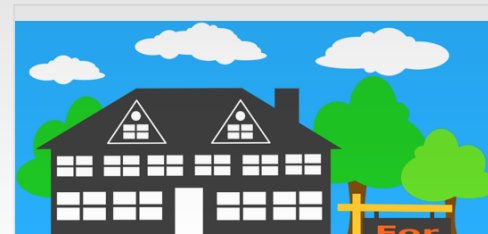
Trained partners and technical guidance.

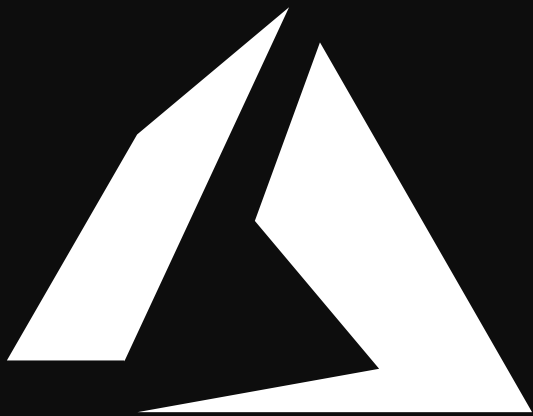
gallery.azure.ai

Azure AI Gallery enables our growing community of developers and data scientists to build and share AI solutions. [Learn how to contribute.](#)



Recently added





Azure is the best cloud for AI