Azure IoT
and Advanced Analytics

Ioannis Stavrinides
Cloud Solution Architect
Data Platform, Advanced Analytics and IoT

Microsoft

Table of Contents

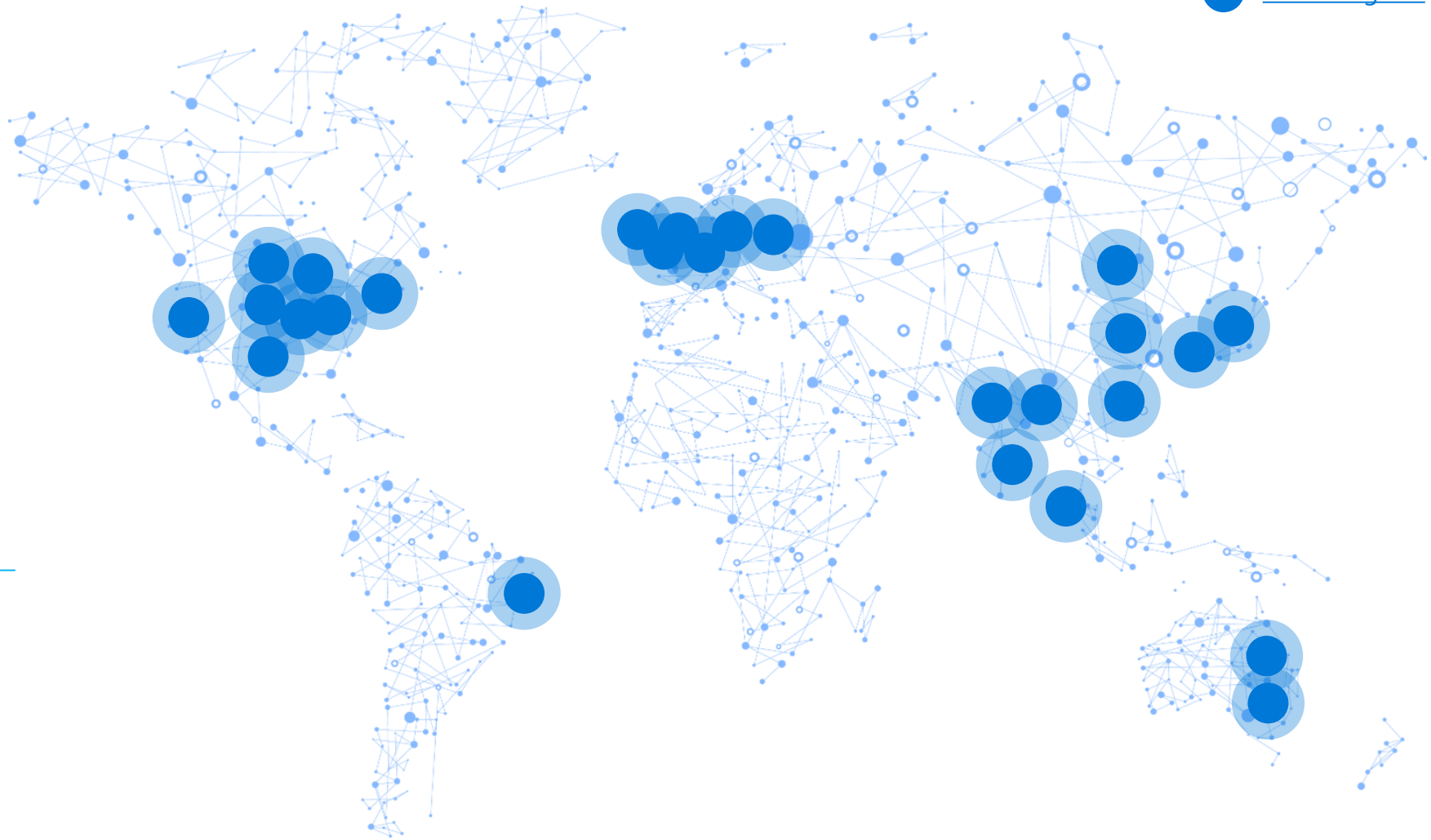
Microsoft Azure

Microsoft Azure

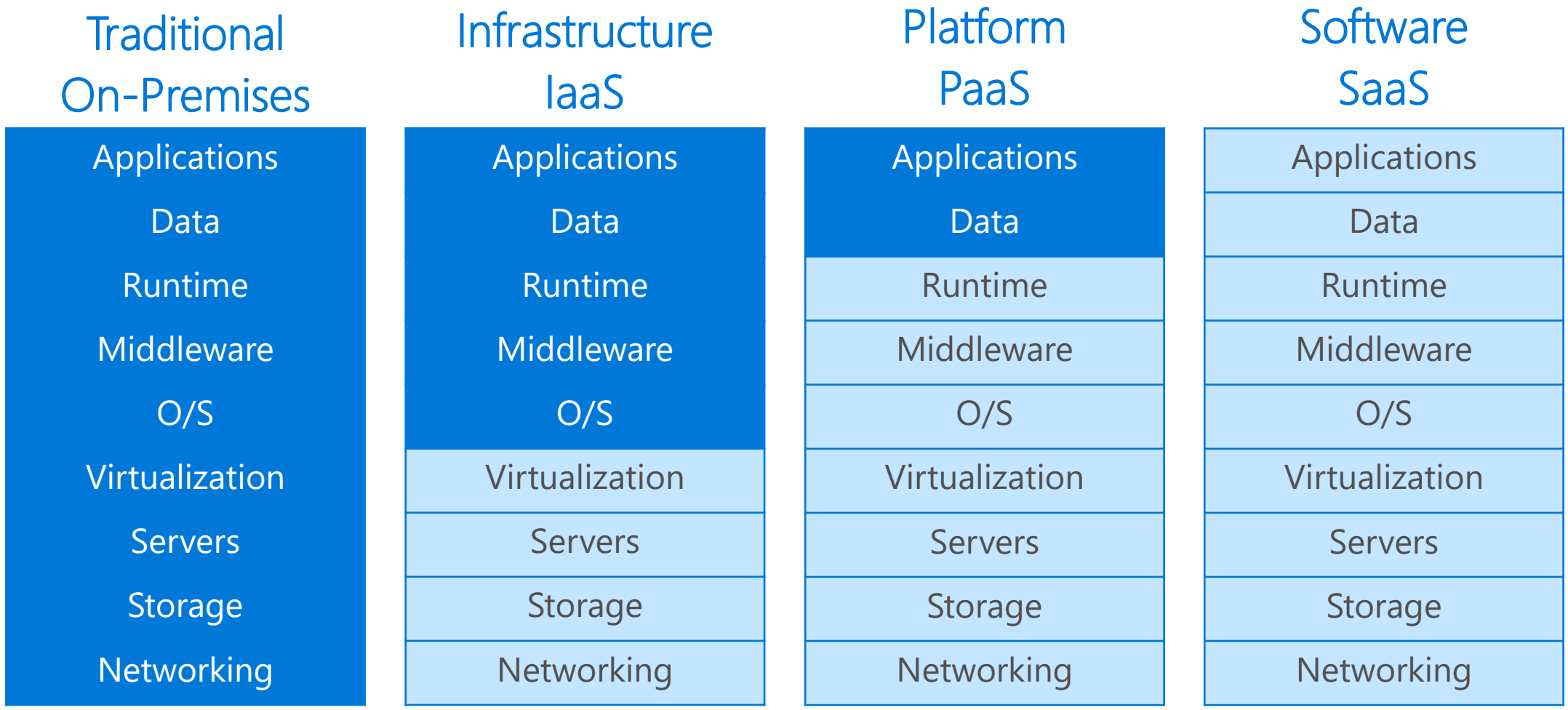
Azure Regions

38

Azure regions announced

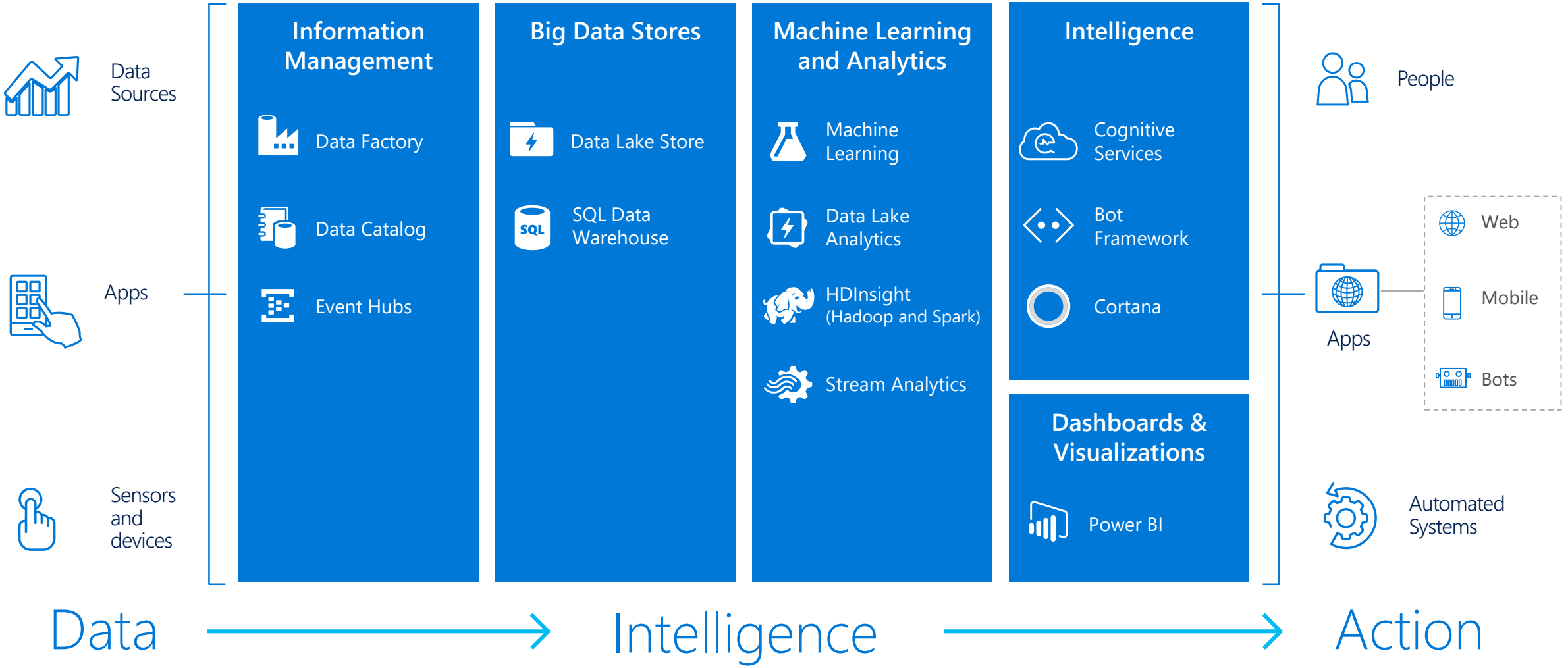


Application deployment patterns



Azure IoT and AA Cortana Intelligence

Cortana Intelligence



Azure Event Hubs

Cloud-scale telemetry ingestion from websites, apps, and any streams of data

Stream **millions** of events per second

Process **real-time and batch** on the same stream

Managed service

Handle **volume**, **variety**, and **velocity**

Azure Messaging Services

Service Bus

Reliable asynchronous message delivery

Event Hubs

Distributed data streaming

Relay

Secure two way communication without changes to your network

Distributed data streaming

Event Hubs

- A streaming service designed to do low latency distributed stream ingress
- A partitioned consumer scale model
- A time retention buffer
- An elastic component in the middle of your chain

Common Event Hubs patterns

Logging / telemetry

- Application logging
- Device / user / performance telemetry
- Dashboarding

Transaction processing (for ex. customer orders/e-commerce)

- Anomaly detection (for ex. fraud/outliers)

Data archival

- Batch processing

Event Hubs features

Archive

Proximity (related data is grouped together)

Order

Consistent playback

Tremendous scale

Azure IoT Hub

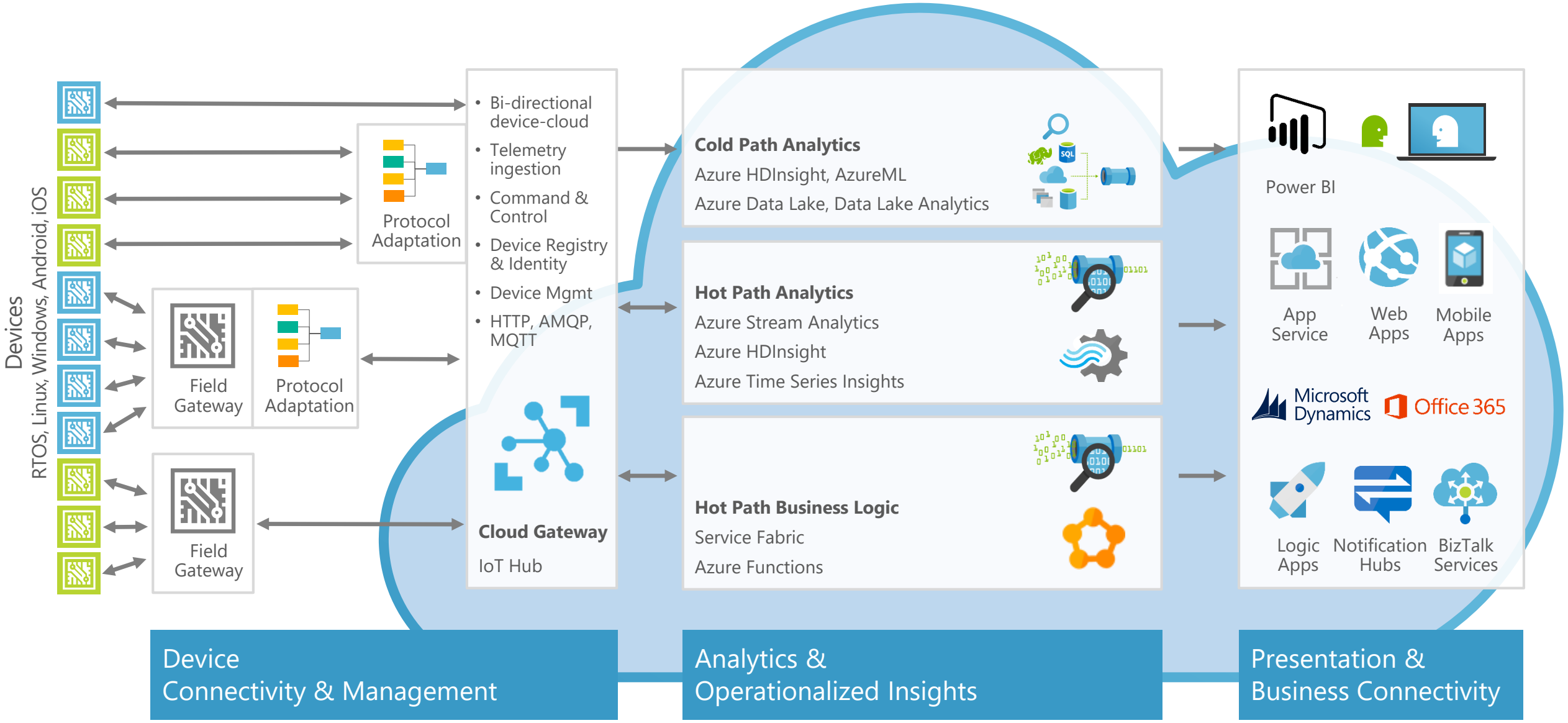
Connect, monitor, and manage billions of IoT assets

Bi-directional communication

Device **authentication** and **registration**

Manage your IoT devices at scale with **device management**

IoT Hub Reference Architecture



Azure IoT Hub

Designed for Scale

- Connect, monitor and manage millions of devices

Designed for Security

- Individual device identities and credentials
- Per-device security tokens
- X.509 via AMQPS/HTTPS/MQTTS
- IP Filter to reject/accept specific IP addresses

Cloud-scale messaging

- D2C, C2D, File transfer & Request/Reply methods
- Durable messages
- Device management: twin/methods/query/jobs
- Delivery receipts, expired messages
- Device communication errors

Flexible & Extensible

- Declarative message routing
- OSS Connectors

Operations Monitoring

- Monitor device connectivity and device identity management events

Connection Multiplexing

- Single device-cloud connection for all communications (C2D, D2C)

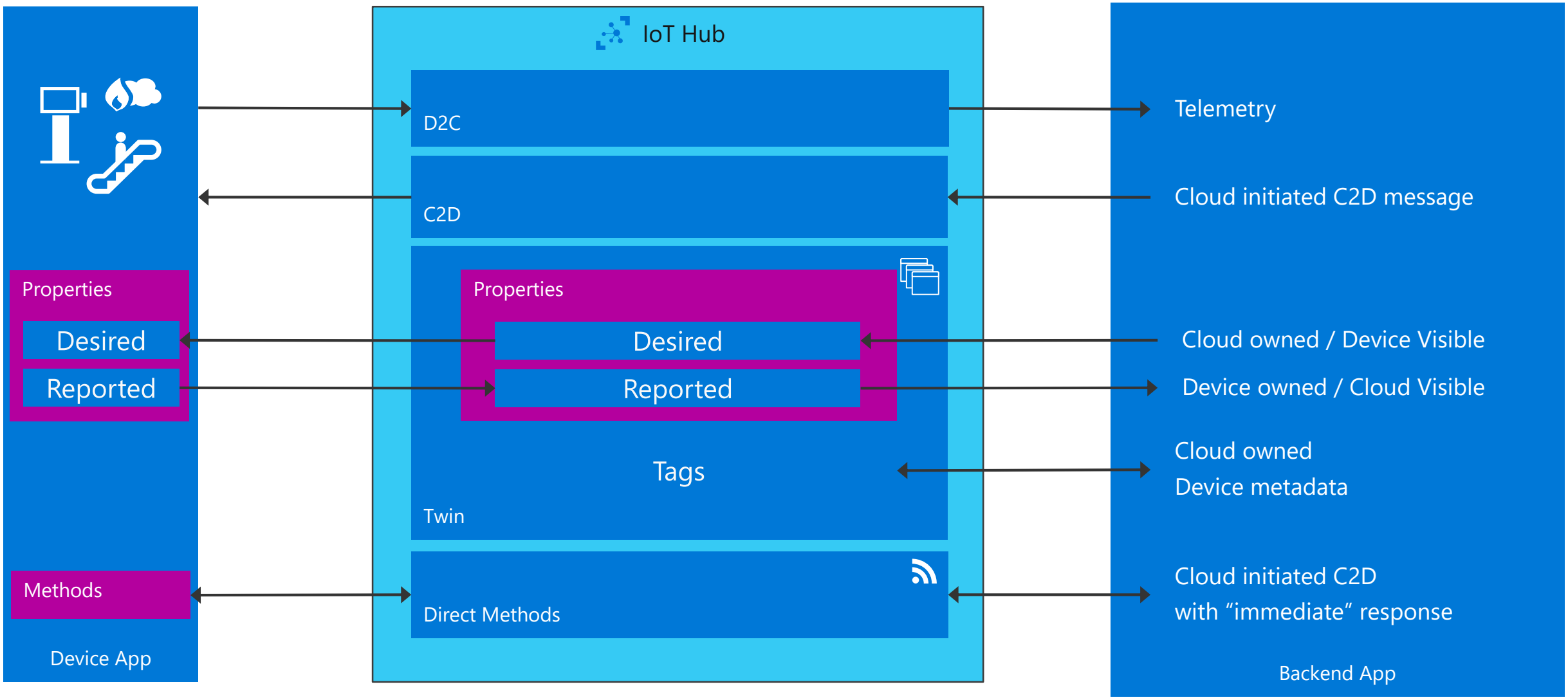
Multi-protocol

- Natively supports AMQP, HTTP, MQTT
- AMQP/MQTT over WebSocket
- Designed for extensibility to custom protocols

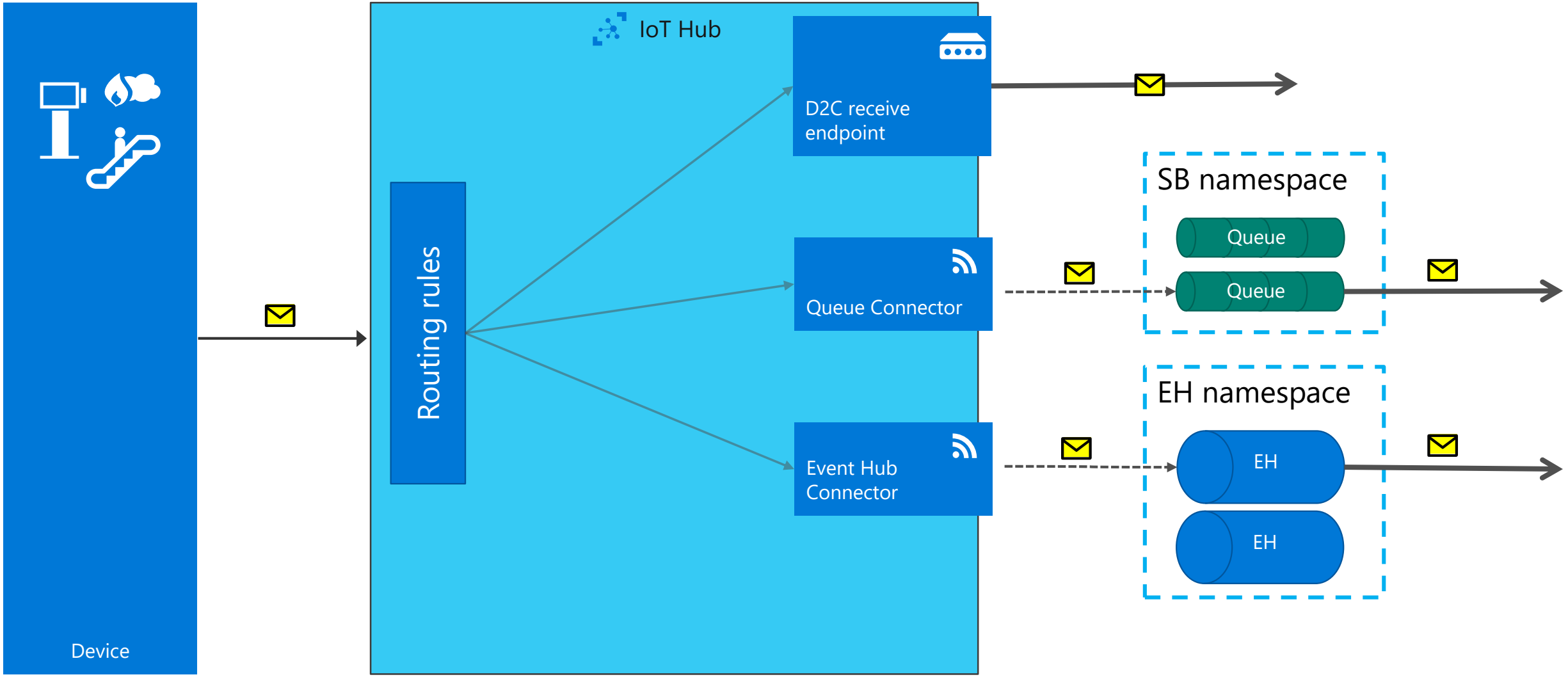
Multi-platform

- Device SDKs available for multiple platforms (e.g. RTOS, Linux, Windows, iOS, Android)
- Multi-platform Service SDK

Manage through Device Twin and Methods

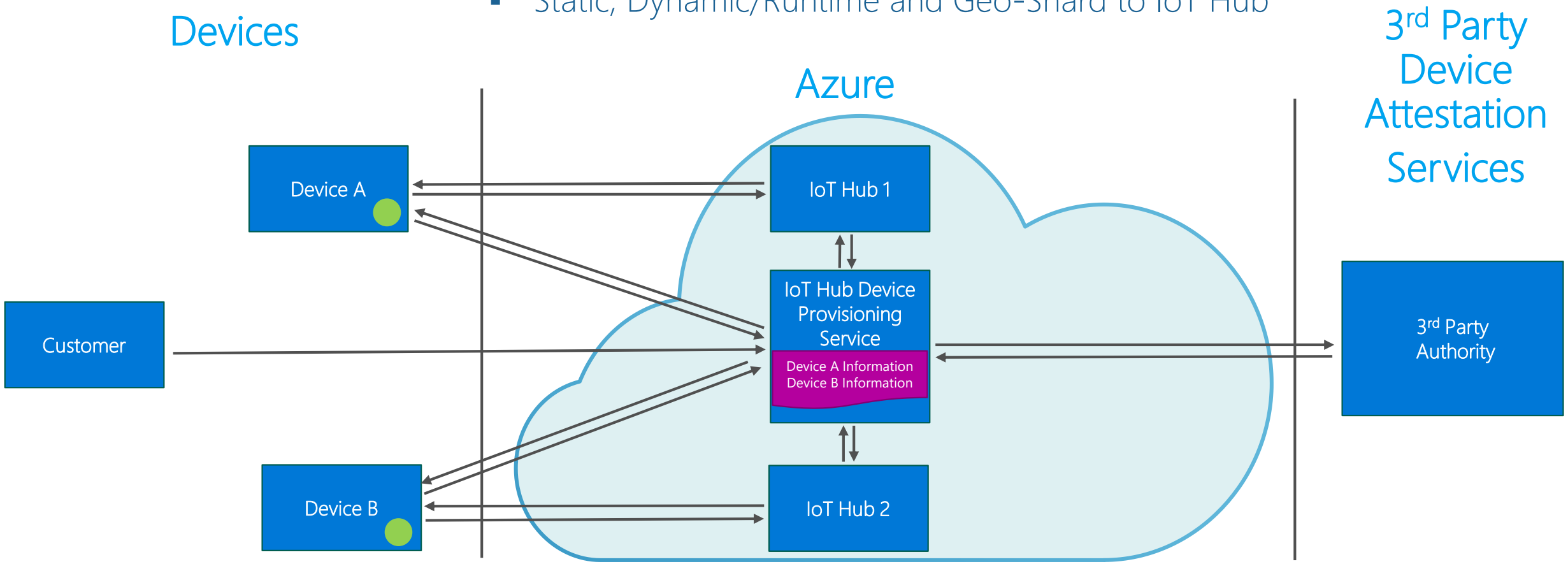


Azure IoT Message Routing



Coming Soon: Azure IoT Hub Device Provisioning Service

- Solve device provisioning & operationalization at scale
- Enable 3rd party hardware attestation services
- Static, Dynamic/Runtime and Geo-Shard to IoT Hub



Azure Stream Analytics

An on-demand real-time analytics service to power intelligent action

Managed Streaming service

No limits to scale

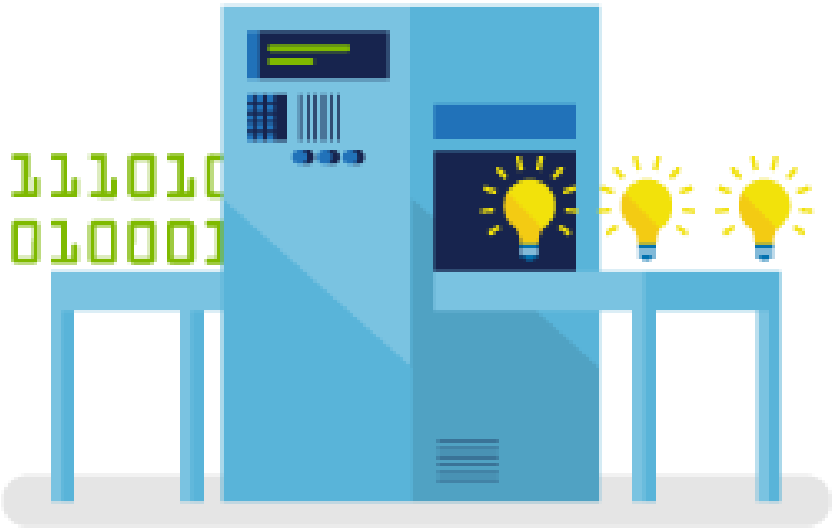
Start **in seconds** and **instantly** analyze data from all your IoT devices and gateways

Develop massively parallel Complex Event Processing pipelines with **simplicity**

Unlocking Real-time Insights

Time to Insight is Critical

Reducing decision latency can unlock business value



Insights are Perishable

Window of opportunity for insights to be actionable

Ask Questions to Data in Motion

Can't wait for data to get to rest before running computation

"Warm Path Analytics" Fills the Gap Between Hot & Cold Path Analytics

Hot Path Analytics: ask questions to data in motion
Cold Path Analytics: ask questions to data in motion

Differentiators

Programmer Productivity

- Declarative SQL like language
- Built-in temporal semantics
- Integrations with sources, sinks, & ML
- Serverless form factor

Lowest Total Cost of Ownership (TCO)

- Fully managed service
- No cluster topology management required
- Seamless scalability
- Usage based pricing

Cloud-Edge Consistency

- Fog computing with Edge Analytics

1,915 lines of code with Apache Storm

```
@ApplicationAnnotation(name="WordCountDemo")
public class Application implements StreamingApplication
{
       protected String fileName =
       "com/datatorrent/demos/wordcount/samplefile.txt";
                    private Locality locality = null;

       @Override  public void populateDAG(DAG dag, Configuration
       conf)
       {
              locality = Locality.CONTAINER_LOCAL;
              WordCountInputOperator input =
              dag.addOperator("wordinput", new
              WordCountInputOperator());
              input.setFileName(fileName);
              UniqueCounter<String> wordCount =
              dag.addOperator("count", new
              UniqueCounter<String>());
              dag.addStream("wordinput-count", input.outputPort,
              wordCount.data).setLocality(locality);
              ConsoleOutputOperator consoleOperator =
              dag.addOperator("console", new
              ConsoleOutputOperator());
              dag.addStream("count-console",wordCount.count,
              consoleOperator.input);
        }
}
```

3 lines of SQL in Azure Stream Analytics

```
SELECT Avg(Purchase), ScoreTollId, Count(*)
FROM GameDataStream
GROUP BY TumblingWindows(5, Minute), Score
```

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

Temporal
Lag
IsFirst
Last
CollectTop

Windowing Extensions
TumblingWindow
HoppingWindow
SlidingWindow

Scaling Extensions
WITH
PARTITION BY
OVER

String
Len
Concat
CharIndex
Substring
Lower, Upper
PatIndex

Mathematical
ABS
CEILING
EXP
FLOOR
POWER
SIGN
SQUARE
SQRT

Geospatial
CreatePoint
CreatePolygon
CreateLineString
ST_DISTANCE
ST_WITHIN
ST_OVERLAPS
ST_INTERSECTS

Mission Critical Reliability

Enterprise Grade SLA

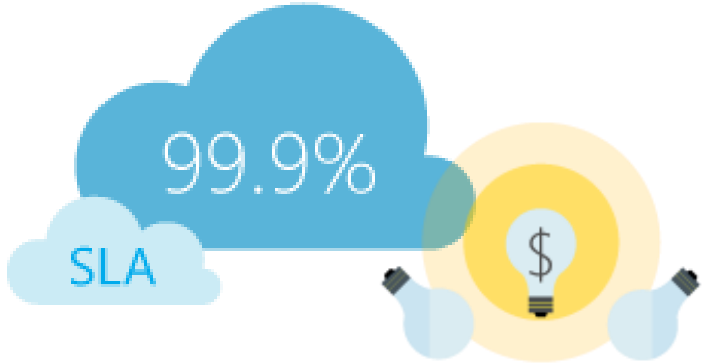
At least three 9s of availability

Business Continuity During Failures

Automatic checkpoint-recovery
Fast restarts

Guaranteed Event Delivery

At-least-once event delivery semantics
No data loss



Other Features

Integration with reference data

Custom code support

JavaScript UDF support

Integration with Azure Machine Learning

Perform real-time scoring on streaming data (Anomaly Detection, Sentiment Analysis etc)

Geospatial capabilities

Azure Data Lake

Azure Data Lake Store

A hyper-scale repository for Big Data analytics workloads

Hadoop File System (HDFS) for the cloud

No limits to scale

Store **any data** in its native format

Enterprise-grade access control, encryption at rest

Optimized for analytic workload **performance**

Azure Data Lake Analytics

A new distributed analytics service

Distributed analytics service built on Apache YARN

Elastic scale per query lets users focus on business goals—not configuring hardware

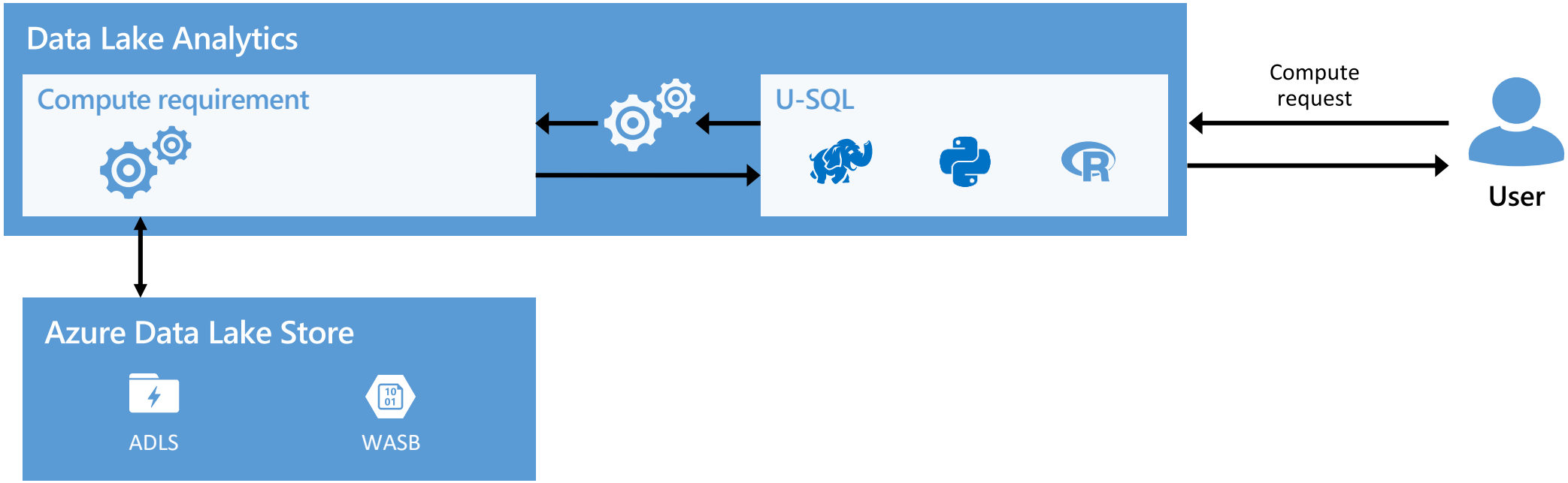
Includes U-SQL—a language that unifies the **benefits of SQL with the expressive power of C#**

Integrates with Visual Studio to develop, debug, and tune code faster

Federated query across Azure data sources

Enterprise-grade **role based access control**

Serverless Architecture



Introducing U-SQL

Familiar syntax to millions of SQL & .NET developers

Unifies

- Declarative nature of SQL with the imperative power of C#
- Processing of structured, semi-structured and unstructured data
- Querying multiple Azure Data Sources (Federated Query)
- Analyzing with Batch, Interactive, Streaming, & Machine Learning in one language

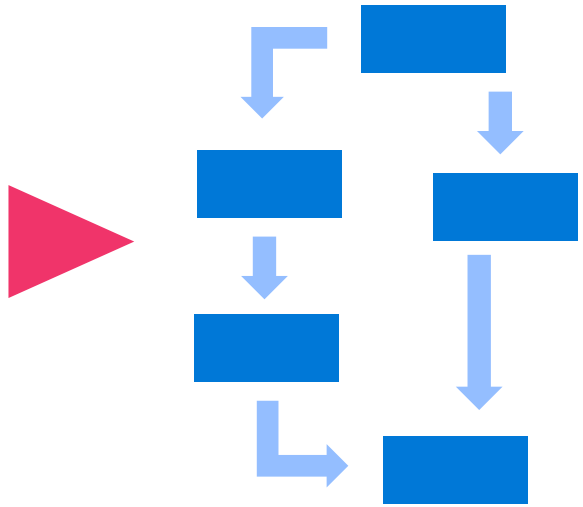
A new language for Big Data

Develop massively parallel programs with simplicity

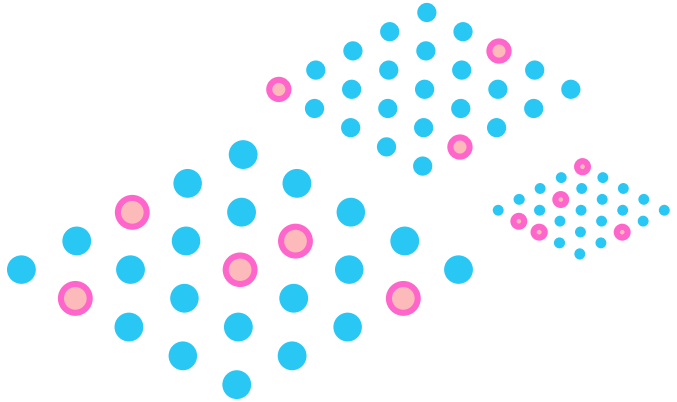
A simple U-SQL script can scale from Gigabytes to Petabytes without learning complex big data programming techniques.

```
@searchlog =
    EXTRACT UserId          int,
            Start           DateTime,
            Region          string,
            Query           string,
            Duration        int,
            Urls            string,
            ClickedUrls     string
    FROM @"/Samples/Data/SearchLog.tsv"
    USING Extractors.Tsv();

OUTPUT @searchlog
    TO @"/Samples/Output/SearchLog_output.tsv"
    USING Outputters.Tsv();
```



U-SQL automatically generates a scaled out and optimized execution plan to handle any amount of data.

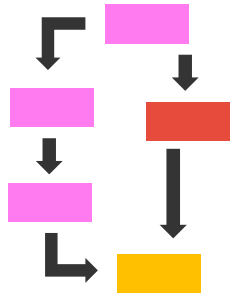


Execution nodes immediately rapidly allocated to run the program.

Error handling, network issues, and runtime optimization are handled automatically.

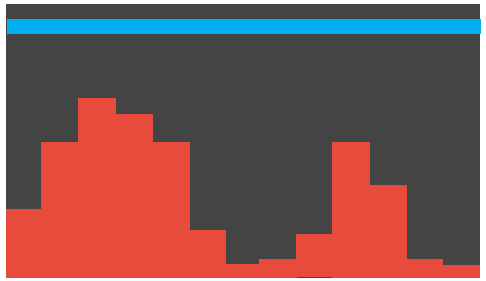
Debug and Optimize your Big Data programs with ease

The execution plan plus detailed logs of the execution nodes are automatically collected and proactively analyzed. Built in views visualize the results in the developer tools.



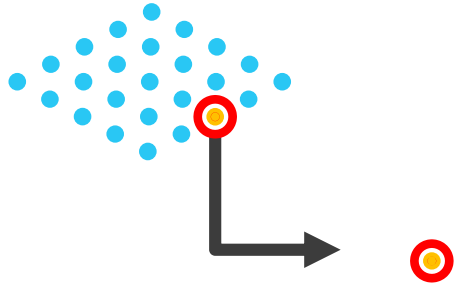
Performance Bottlenecks

Hotspots identified for I/O, execution time, CPU time. Executions plans can be interactively played back for intuitive understanding of performance bottlenecks.



Trade off time versus cost

Efficiency analysis reveals whether the developer has reserved more processing resources than needed. Optimization views estimated number of resources needed to secure the fastest execution time.



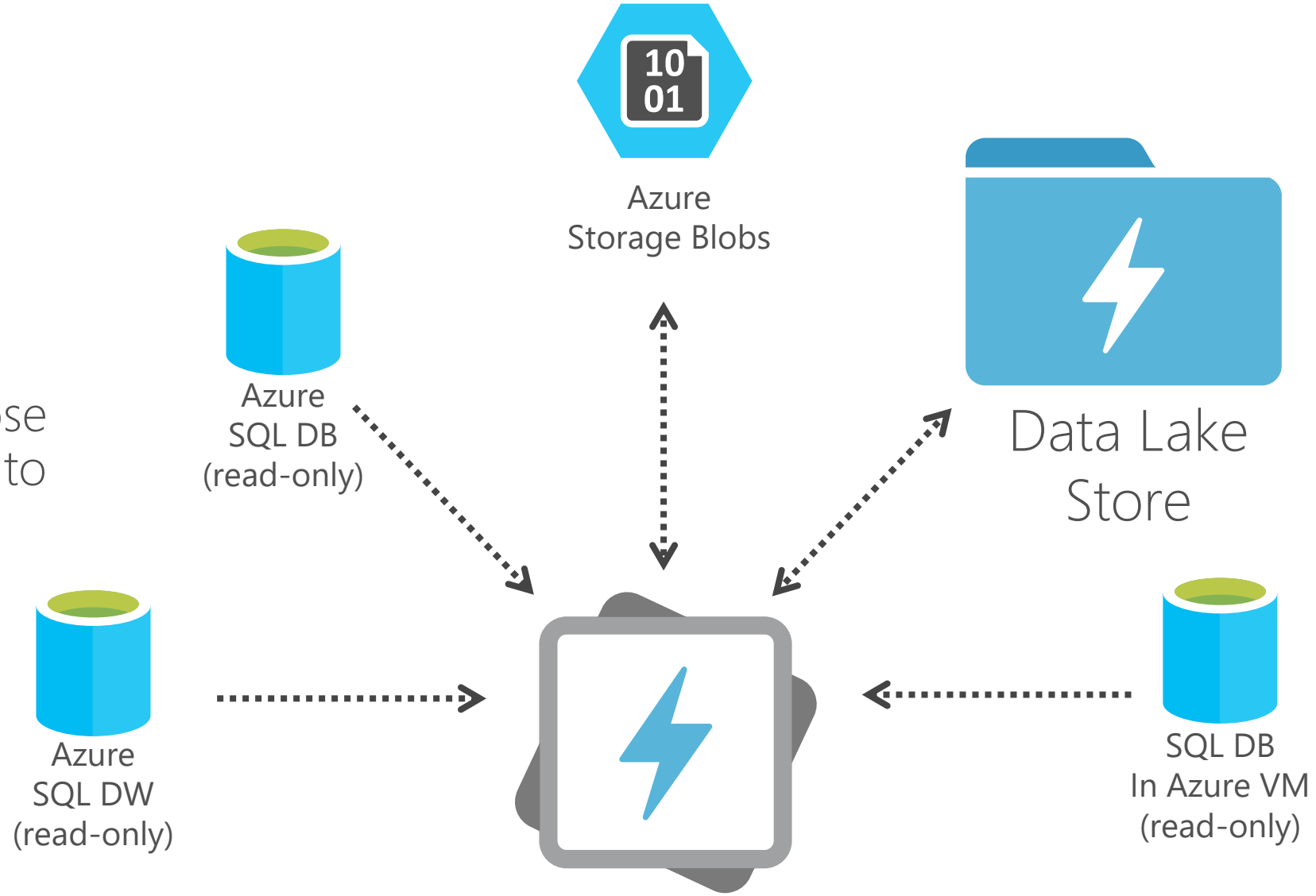
Debug user code errors

When user-defined procedural code is used, failed nodes and input data can be downloaded to the developer workstation for interactive debugging in Visual Studio.

Federated Query

U-SQL can query data from multiple sources in Azure.

Where possible data transformation is pushed close to the remote query engine to minimize data transfer and maximize performance.



Embedded Artificial Intelligence

Host Deep Neural Networks (DNNs)

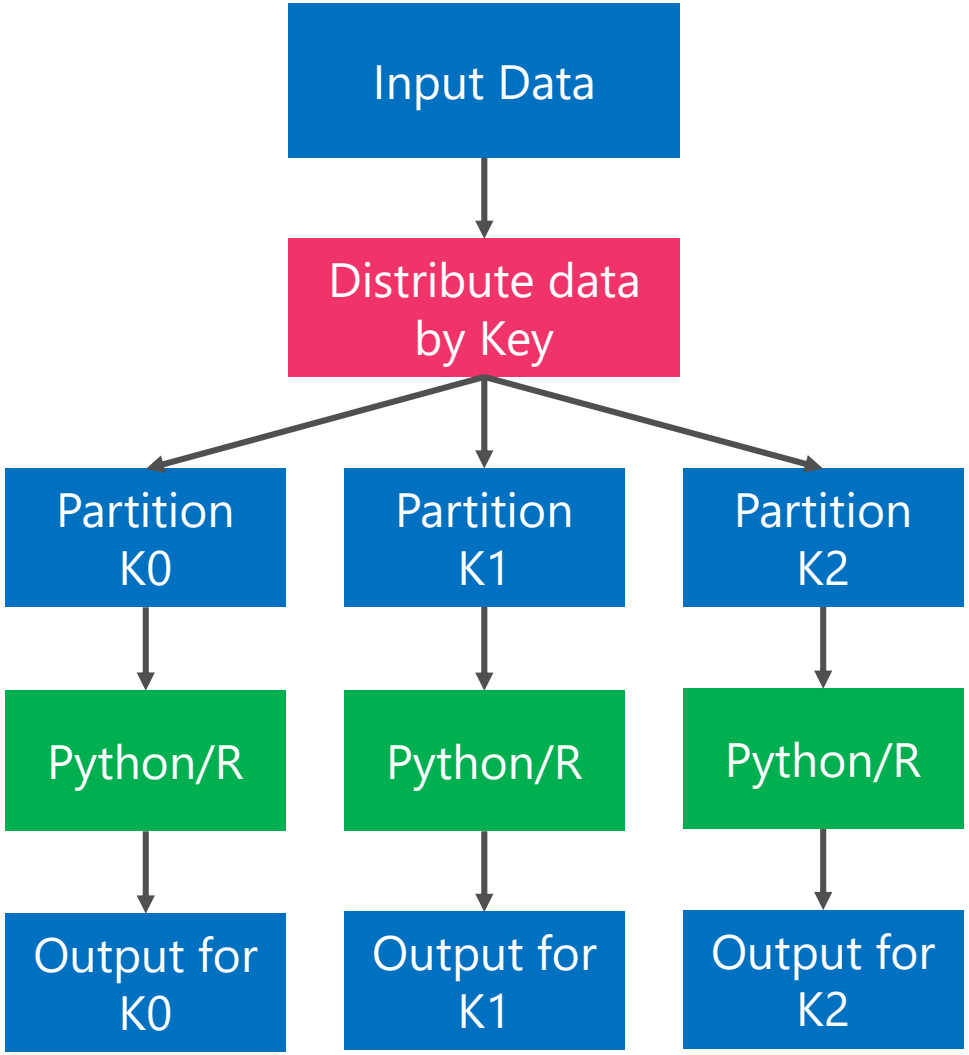
6 Built-in Cognitive Functions

- Face API
- Image Tagging
- Emotion analysis
- OCR
- Text Key Phrase Extraction
- Text Sentiment Analysis

Massively Parallel Programs with Python & R

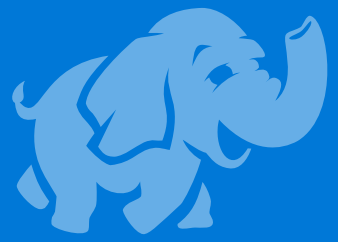
The U-SQL batch query execution system make sit easy to reuse Python and R code on execution nodes.

Reuse Python & R libraries perform massively parallel scoring on thousands of nodes simultaneously.



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**

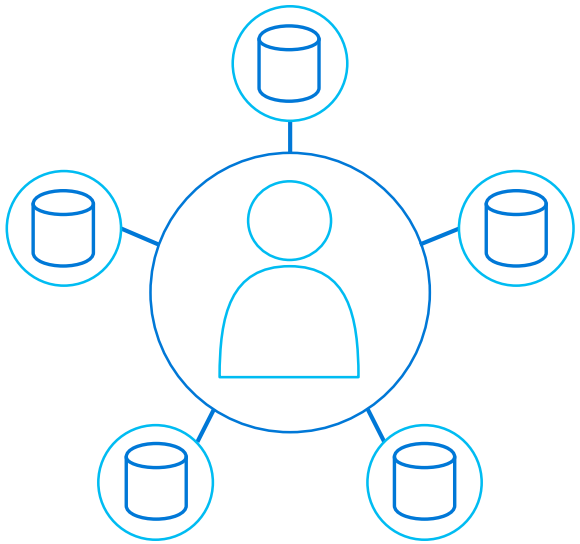
63% lower TCO than deploy your own Hadoop on-premises*

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

Highly Available – Designed for the cloud ground up

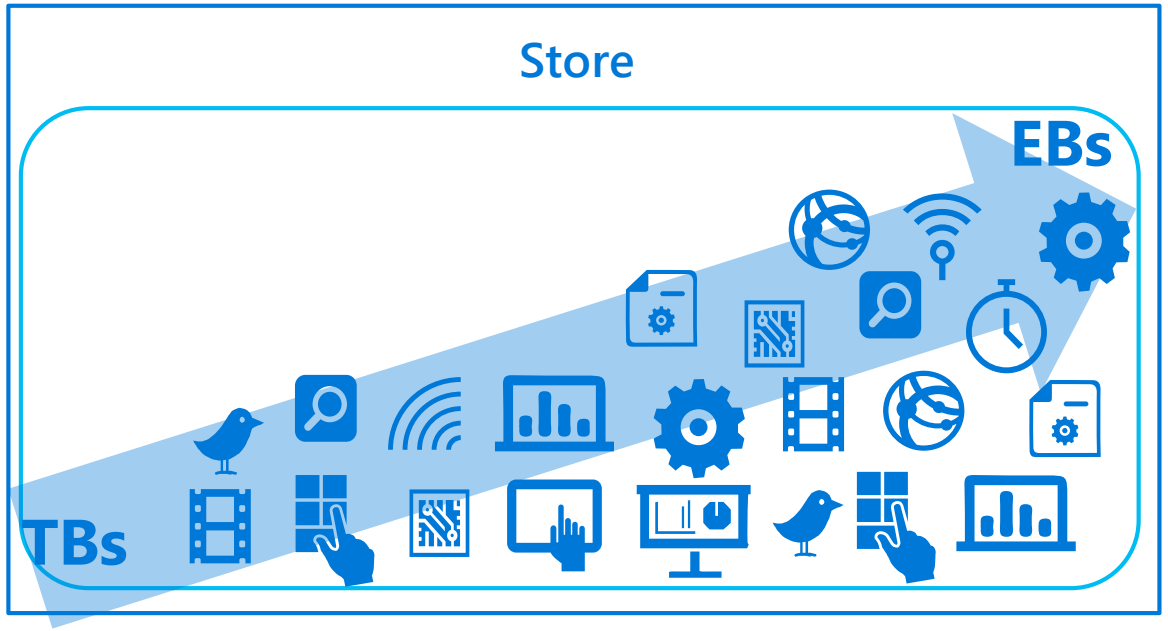
- HDInsight provides primary and secondary headnodes allowing for better reliability
- Have invested in making entire stack including Resource Manager, HiverServer2 HA ready
- HDInsight stack includes Zookeeper nodes at no extra charge to customer
- 99.9% SLA

Always encrypted, Role-based security & Auditing



- Always encrypted; in motion using SSL, and at rest using keys in Azure Key Vault
- Single sign-on, multi-factor authentication and integration of on-premises identities w/Active Directory integration
- Fine-grained ACLs for role-based access controls with Apache Ranger
- Auditing every access / configuration change with Apache Ranger

Petabyte size files and Trillions of objects



- Store data in it's native format
- PB sized files, 200x larger than anyone else
- Scalable throughput for massively parallel analytics
- No need to redesign application or reparation data at higher scale

Backed by Microsoft and Hortonworks

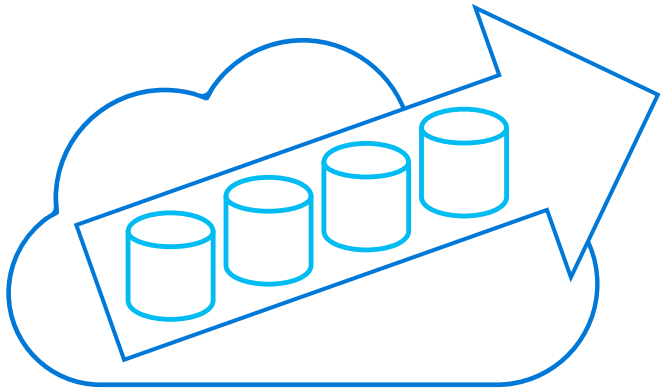
- Microsoft + Hortonworks has **37 committers** for Hadoop Core; more than all managed cloud Hadoop vendors combined
- Uniquely ready to support your deployment
- Can fix and commit code back to Hadoop

Lower total cost of ownership

- No hardware
- Hadoop support included with Azure support
- Pay only for what you use
- Independently scale storage and compute
- No need to hire specialized operations team
- 63% lower total cost of ownership than on-premises*

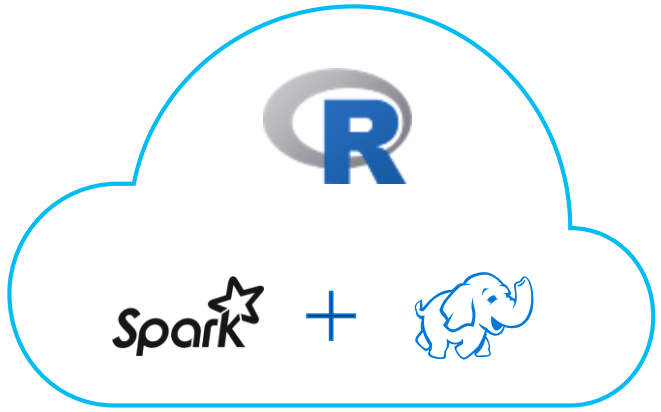
*IDC study "The Business Value and TCO Advantage of Apache Hadoop in the Cloud with Microsoft Azure HDInsight"

Easy for administrators to spin up quickly



- Deploy big data projects in minutes
- No hardware to install, tune, configure or deploy
- No infrastructure or software to manage
- Scale to tens to thousands of machines instantly

Easy for data scientists with familiar R language



*Applies to HDInsight only

R Server for HDInsight

- Largest portable R parallel analytics library
- Terabyte-scale machine learning—1,000x larger than in open source R and up to 100x faster performance using Spark and optimized vector/math libraries
- Deep IDE integration
- Jupyter and Zeppelin notebooks

Workloads

HDFS
MapReduce
Hive
Hbase
Storm
Kafka
Mahout
Spark
R Server

Azure Cosmos Db

Globally distributed, multi-model database service

Turn-key global distribution

Multi-model and **multi-API**

Limitless scale

Well defined **consistency** levels

Industry leading **SLAs**

Cloud first, Mobile first Applications

Mission-critical applications for
a global userbase need …

Global distribution

Elasticity of compute and storage

Fast, Responsive millisecond latency

Durable, Consistent and Highly available

SQL

mongoDB

Gremlin

$G = (V, E)$

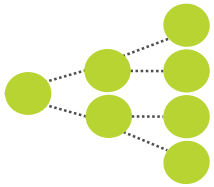
Azure Cosmos DB

... *more coming soon*

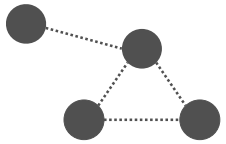
Key-Value

Column-family

Documents



Graph



Global distribution | Elastic scale out | Guaranteed low latency | Five consistency models | Comprehensive SLAs

A globally-distributed, multi-model database service

Design Goals

Elastic scale, Highly responsive, Consistency of data
Always-On from day 1
Reduce the relational “tax”
Developer flexibility
Lowest TCO
Stand behind the tech

Global distribution

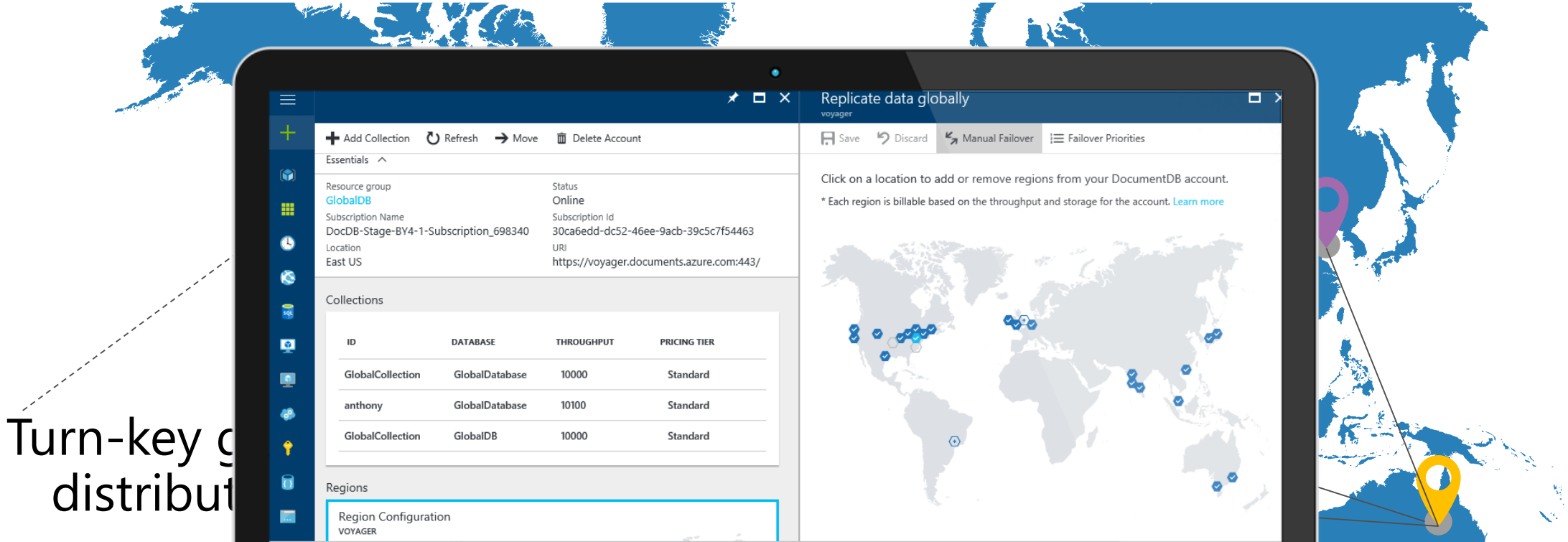
Available in all Azure regions

Multi-homing APIs

Comprehensive SLA

Manual and automatic failover

Automatic & synchronous multi-region replication



Turn-key g
distribut

Limitless scale : storage and throughput

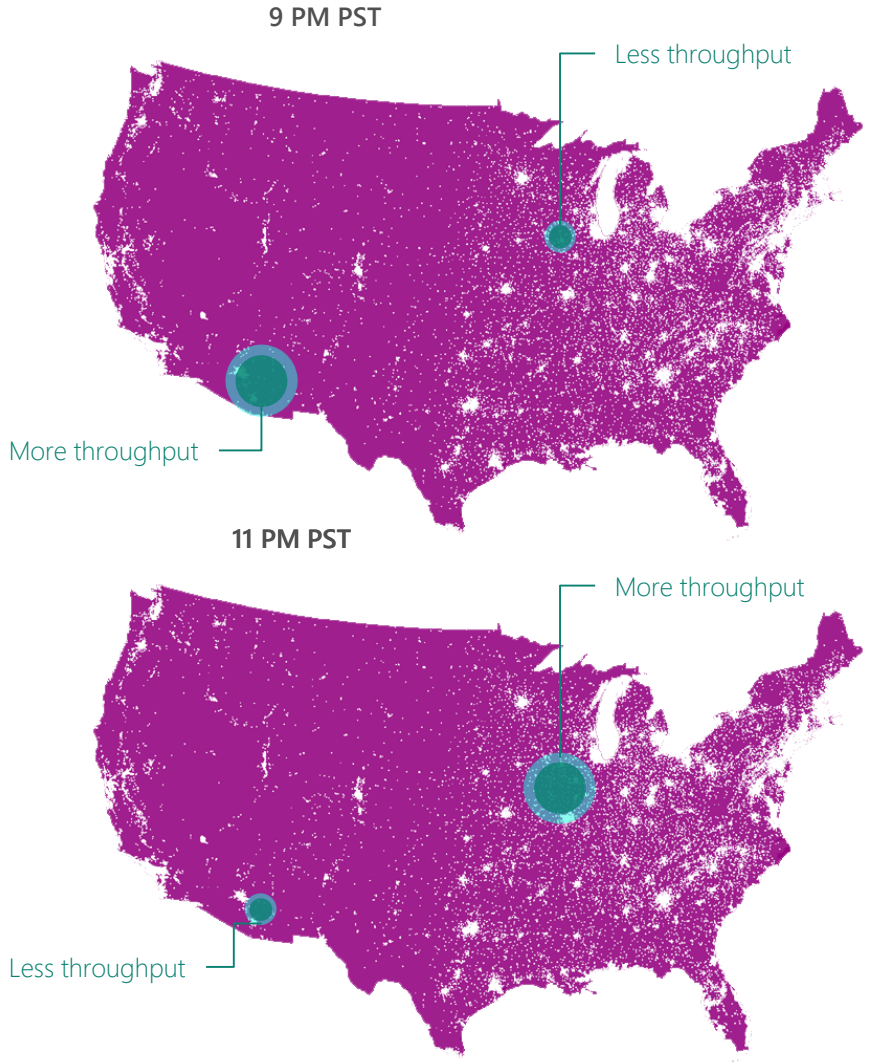
Independent storage and throughput scale

Scale worldwide, on your terms

Storage: Gigabytes to Petabytes

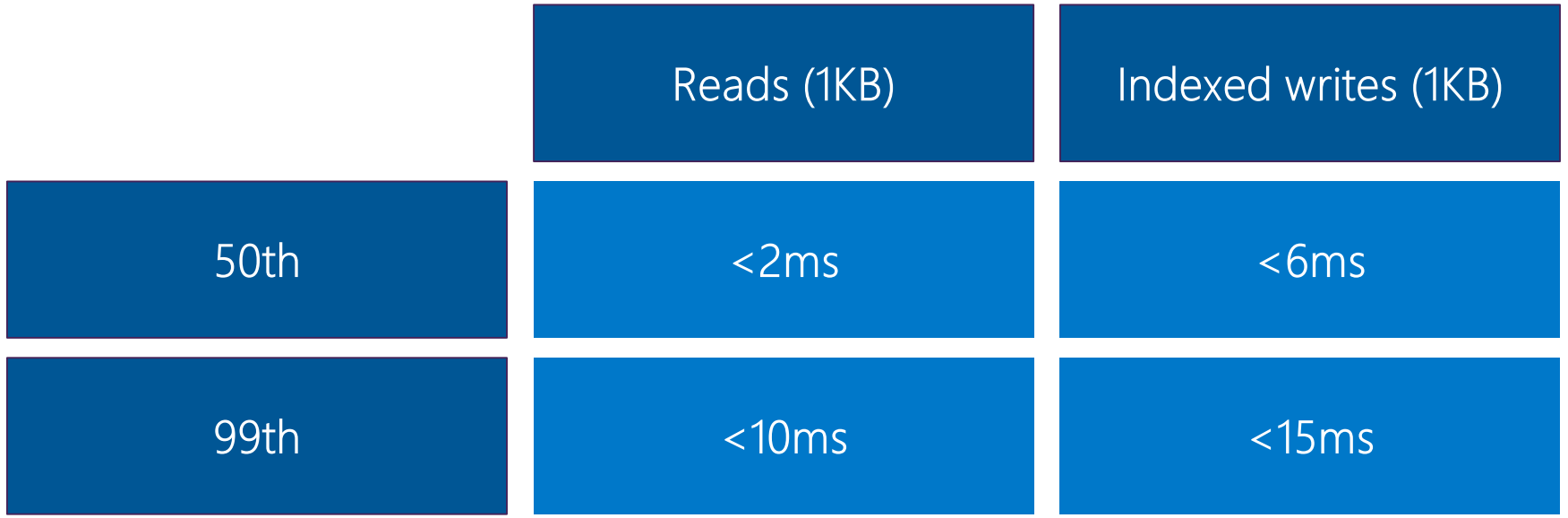
Throughput: 100s to 100s of million requests/sec

Only pay for what you need



Guaranteed low latency

Reads and writes served from local region
Guaranteed millisecond latency worldwide
Write optimized, latch-free database engine
Automatic indexing



Well-defined consistency models

Intuitive programming

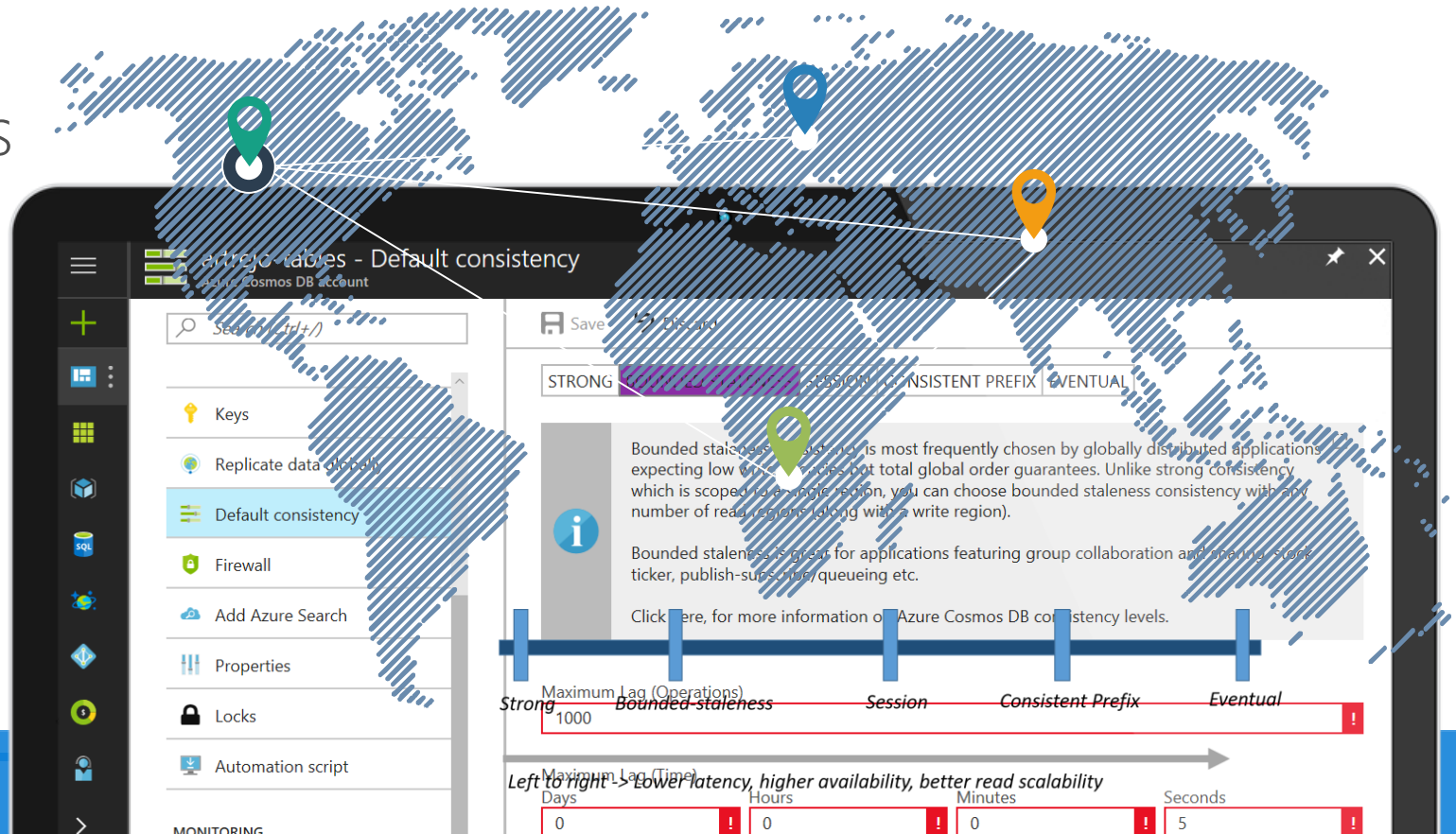
Well-defined, relaxed consistency models

Five consistency levels

Overrides on per-request basis

Navigating CAP theorem

Consistent data worldwide



Clear PACELC tradeoffs

Latency

Availability

Throughput

Azure Machine Learning

Build powerful, cloud-based machine learning applications

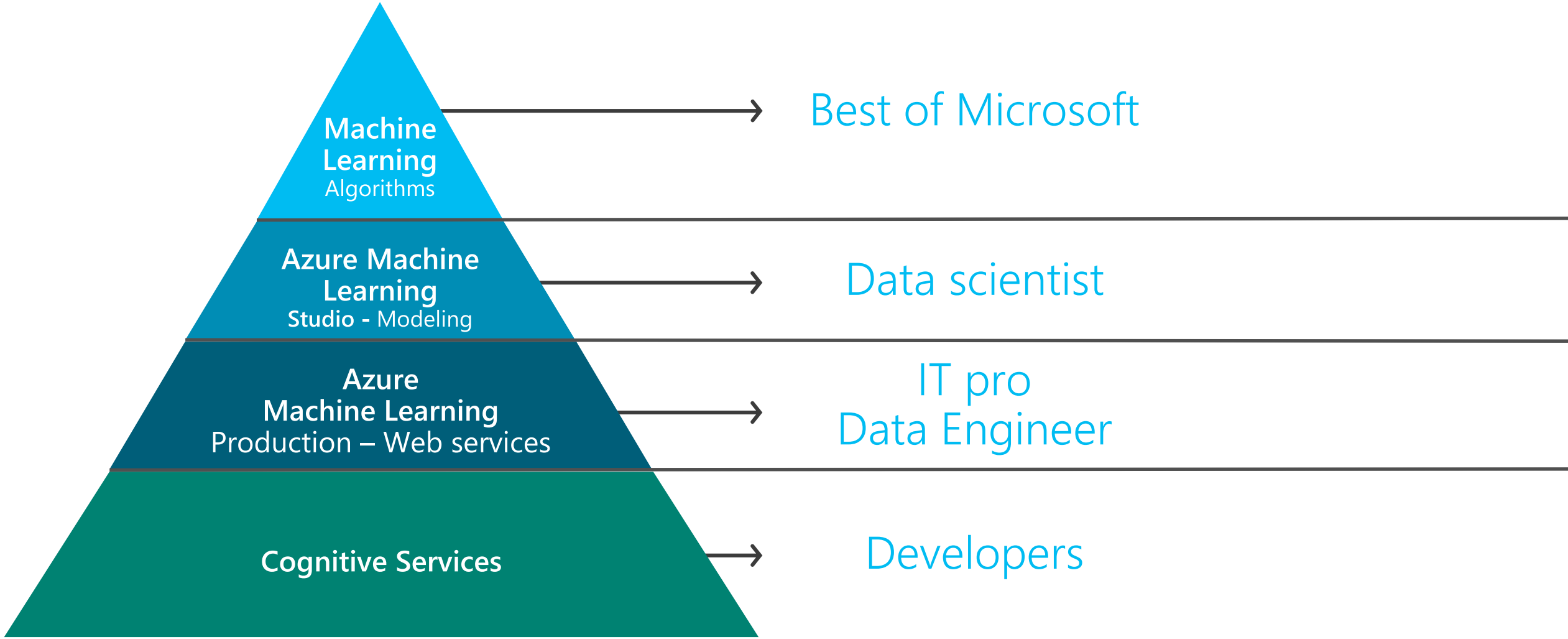
Intuitive modeling experience

Model **deployment** in minutes

R and **Python** support

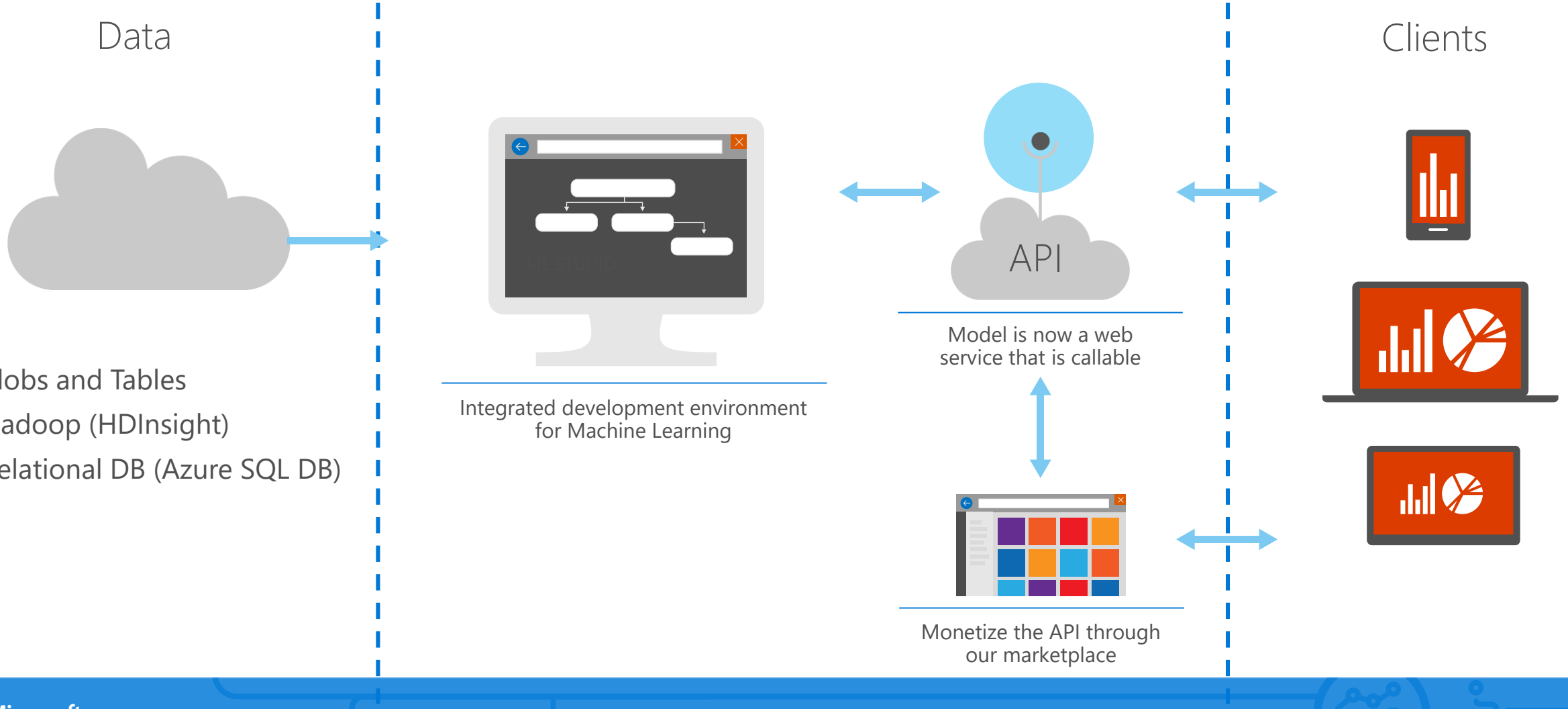
Data Scientist and Developer **friendly**

Machine Learning services in the cloud



Azure Machine Learning Service

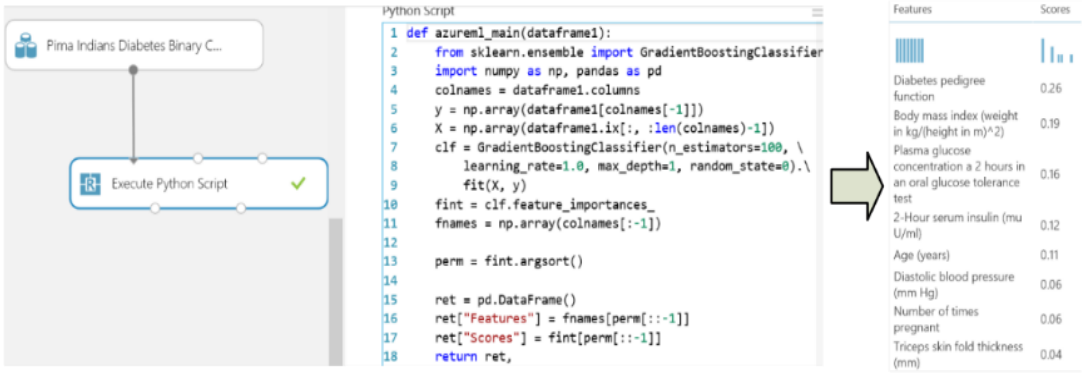
Data -> Predictive model -> Operational web API in minutes



Model Your Way: Open source/our source

Script with R, SQLite or Python

CPython 2.7 support from inside AML Studio
numpy/scipy/panda/scikit-learn/etc.
Anaconda distro pre-installed

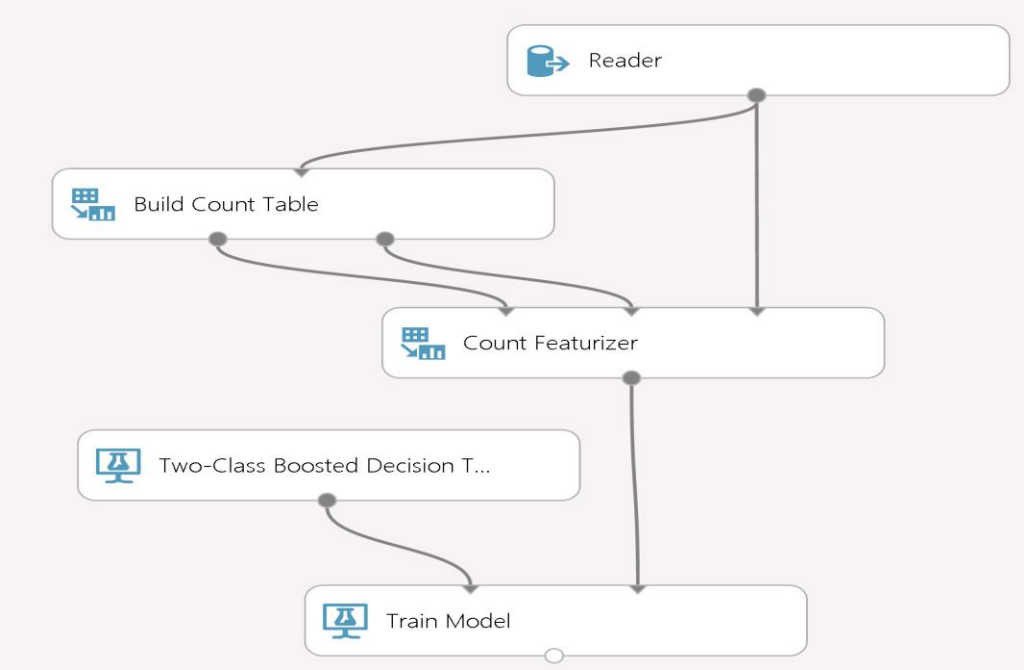


Python client library

Analyze data using Python and its libraries
Use IPython, PTVS, Eclipse to edit/debug

Big learning with counts

TB scale datasets
Modular: tune/monitor/replace in isolation
Monitorable and debuggable



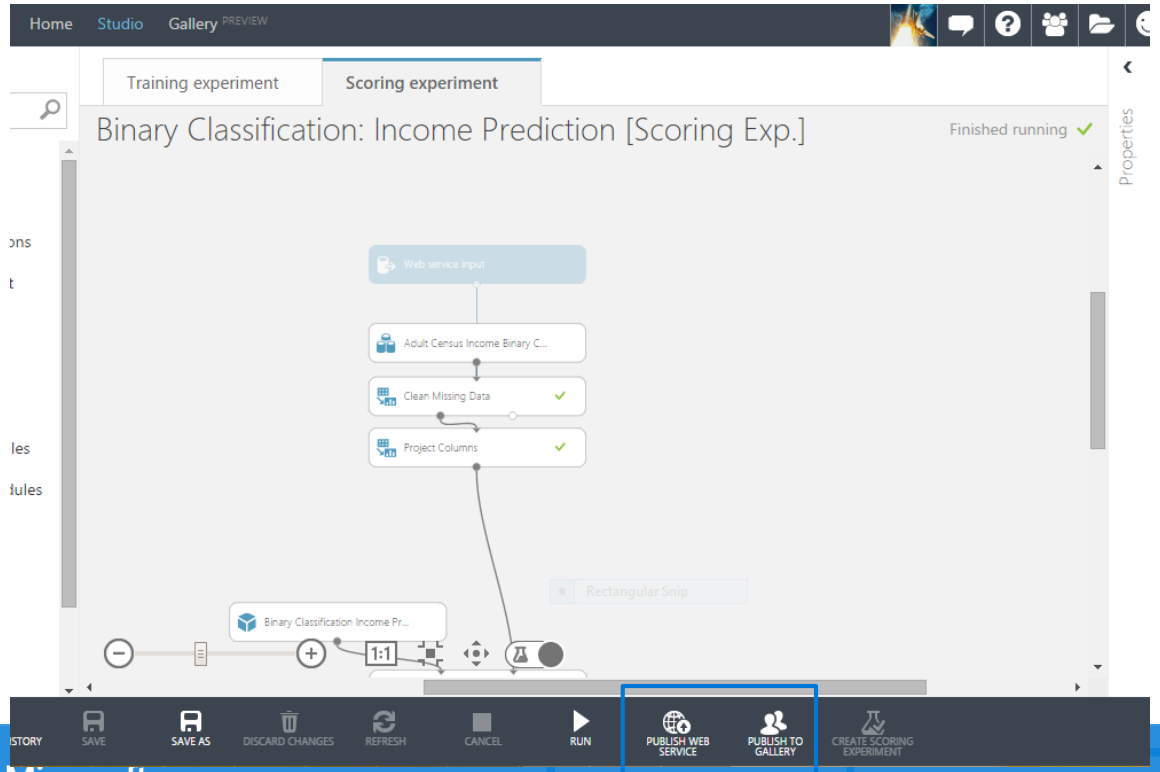
Deploy in Minutes

One click to production

Publish as a Web Service or to Gallery

Continuous updates to streamline process

Stay tuned to our blog for more



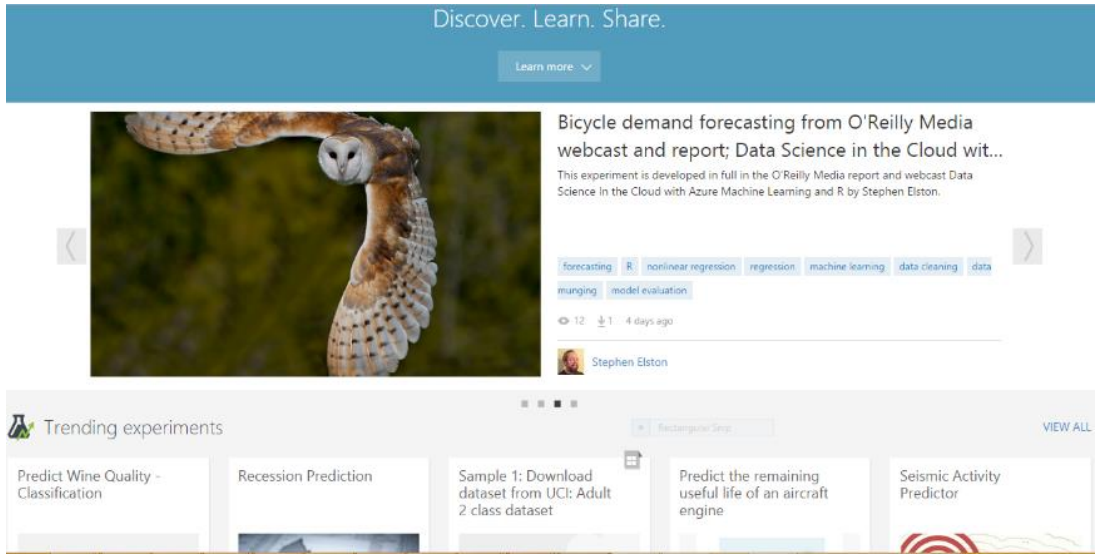
Expand your Reach

New in-product Gallery

Discover what others have built

Learn by dropping these into your workspace

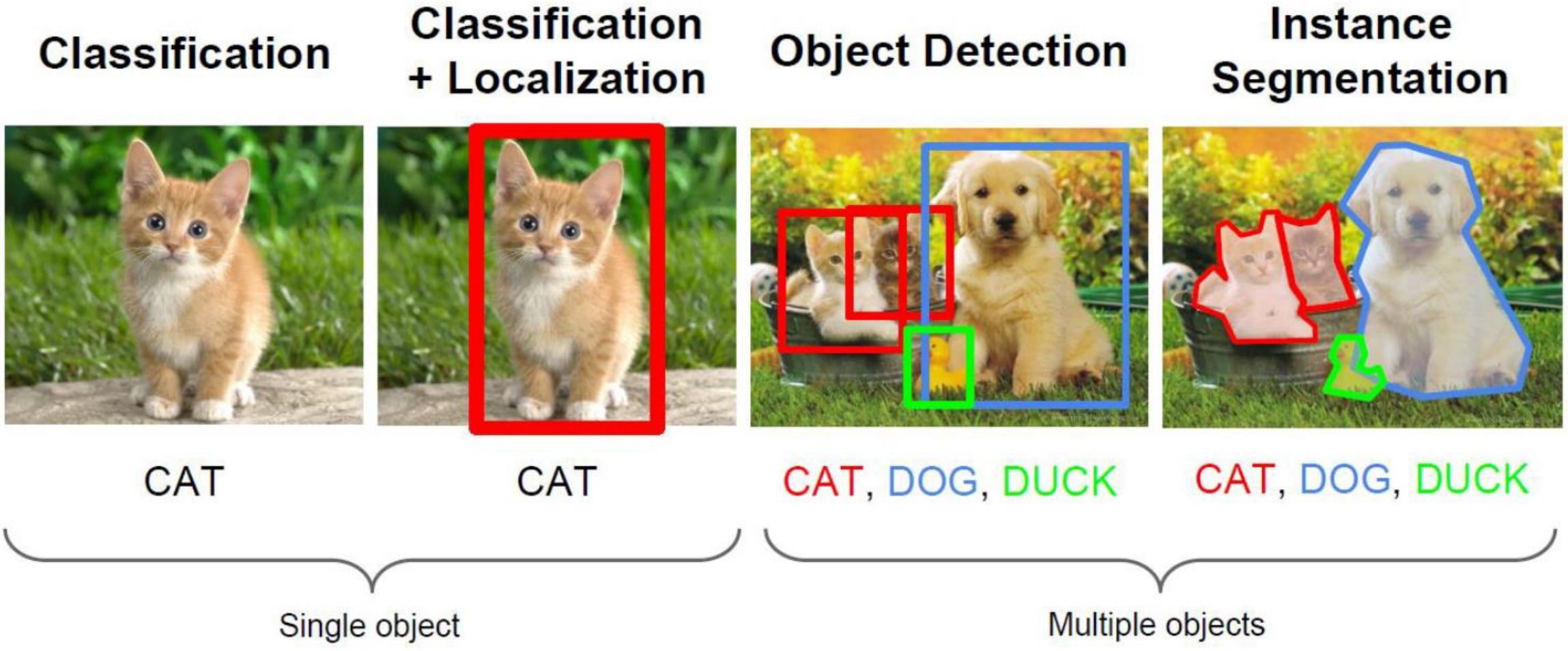
Share your work with others



Microsoft

Microsoft Cognitive Services

Vision

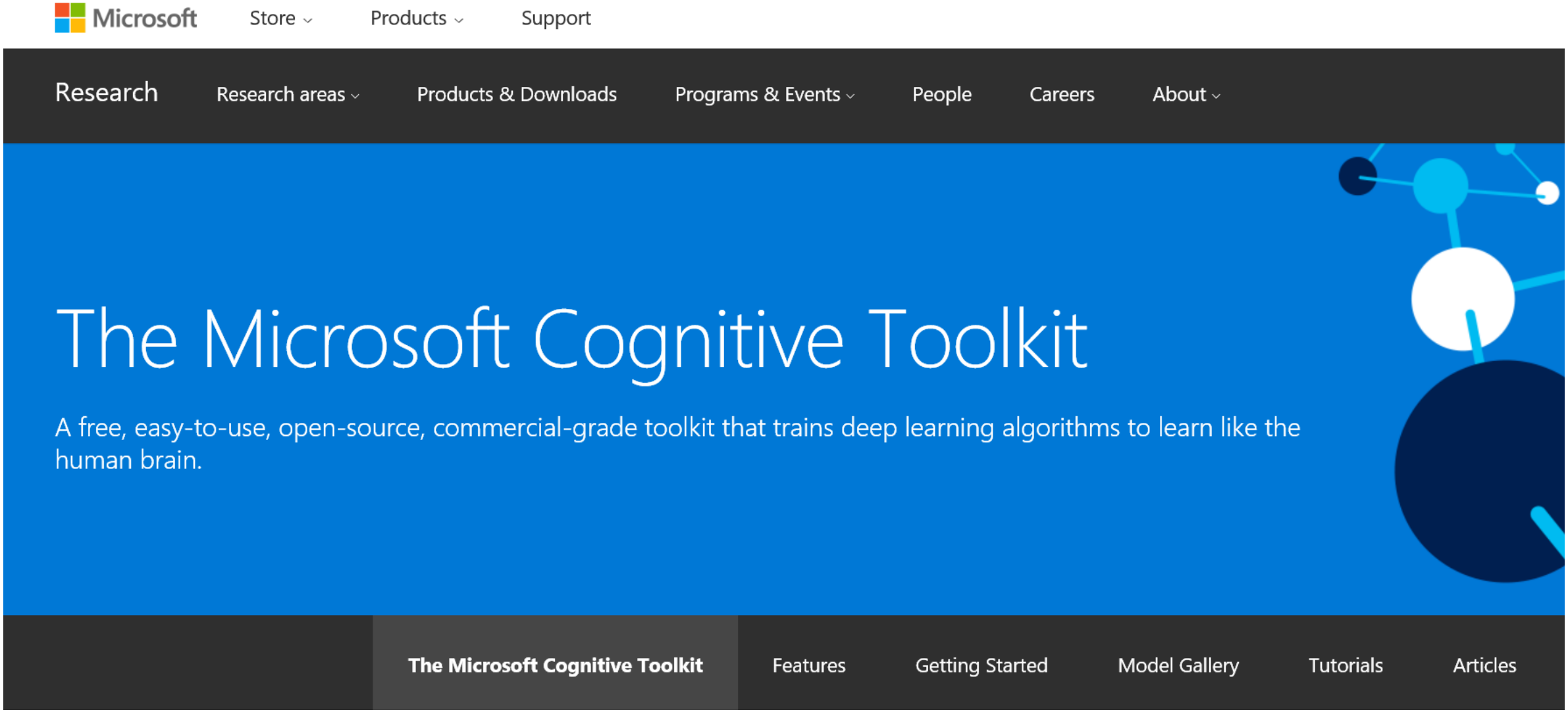

Classification
Classification + Localization
Object Detection
Instance Segmentation
CAT
CAT
CAT, DOG, DUCK
CAT, DOG, DUCK
Single object
Multiple objects

Connected Drone

Microsoft Cognitive Services

VISION	SPEECH	LANGUAGE	KNOWLEDGE	SEARCH
Computer Vision	Custom Speech Service (formerly CRIS)	Bing Spell Check	Academic Knowledge	Bing Autosuggest
Emotion	Speaker Recognition	Language Understanding	Entity Linking	Bing Image Search
Content Moderator	Speech	Linguistic Analysis	Knowledge Exploration	Bing News Search
Face		Text Analytics	Recommendations	Bing Video Search
Video		Translator	QnA Maker	Bing Web Search
		Web Language Model		

Microsoft Cognitive Toolkit



The Microsoft Cognitive Toolkit (CNTK)

- CNTK is Microsoft's **open-source**, **cross-platform** toolkit for learning and evaluating **deep neural networks**
- CNTK expresses (nearly) **arbitrary neural networks** by composing simple building blocks into complex **computational networks**, supporting relevant network types and applications.
- CNTK is **production-ready**: State-of-the-art accuracy, efficient, and scales to multi-GPU/multi-server.

Azure Data Factory

Compose and Manage Data at Scale

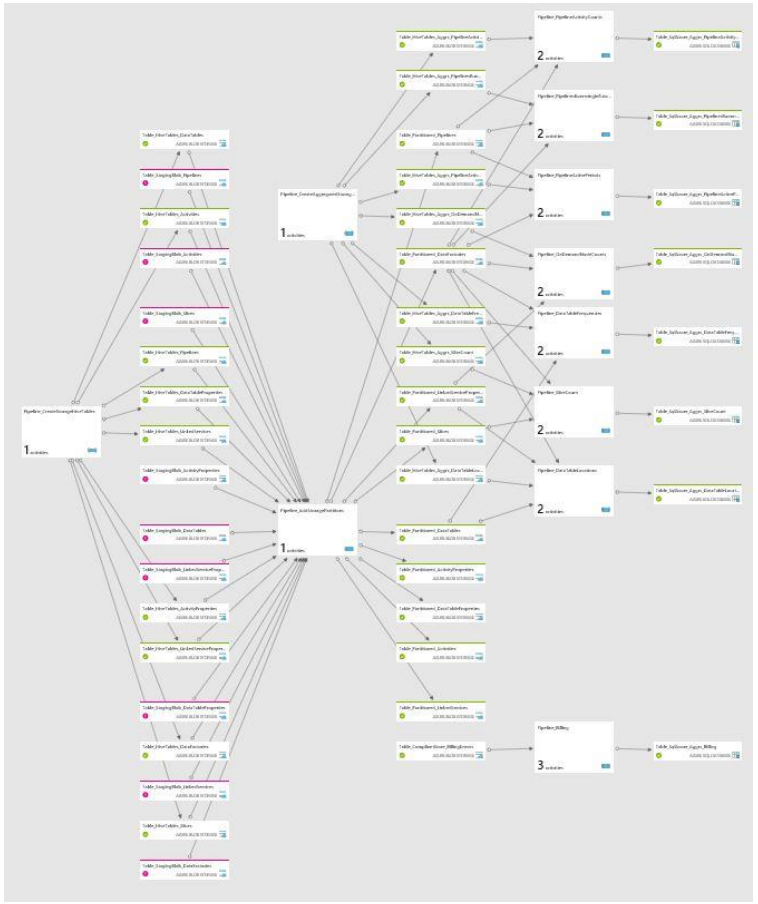
Ingest and **Prepare** data

Transform and **Analyze** data

Orchestrate data movement

Azure Data Factory

Create & manage data pipelines at scale

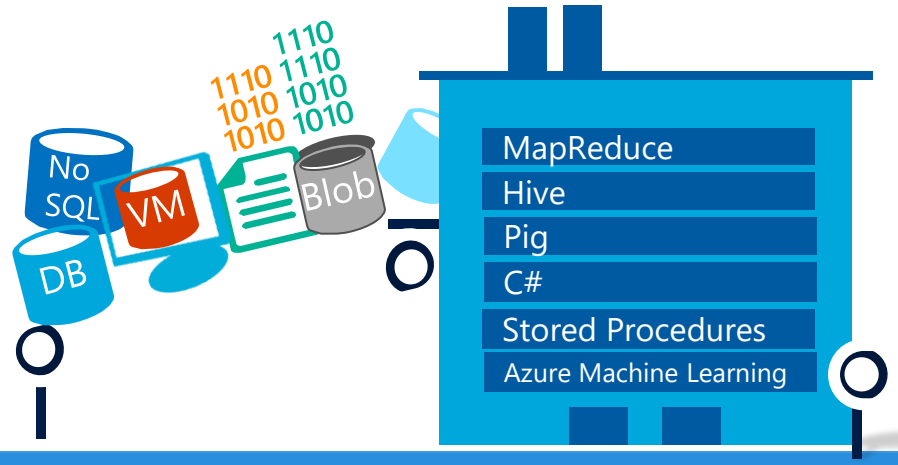


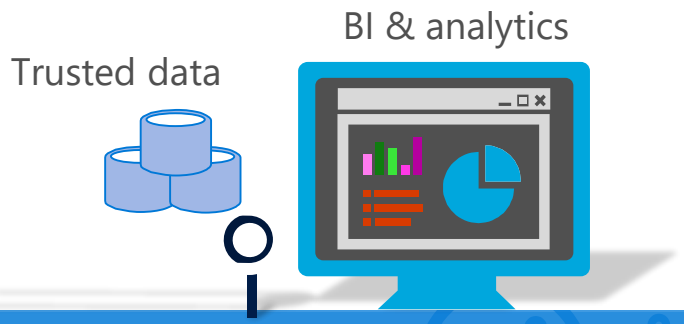
Fully managed service to support orchestration of data movement and transformation

Connect to relational or non-relational data that is on-premises or in the cloud

Single pane of glass to monitor and manage data processing pipelines.

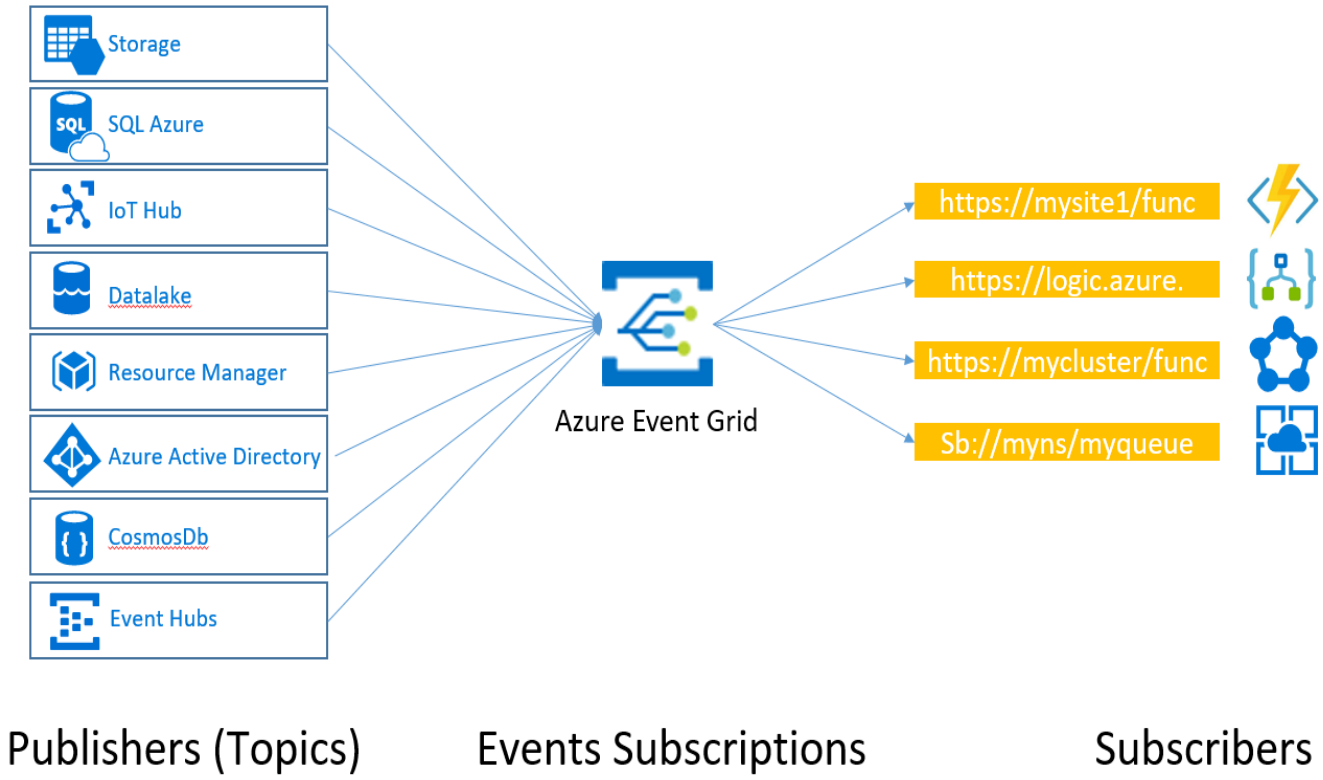
Publish to Power BI





Event Grid

- Eventing backplane
- Enable event based programing with pub/sub semantics
- Reliable distribution & delivery for all Azure services and 3rd parties



- **Events** – user reaction – Create, Read, Update, Delete
- **Event Grid Subscriptions** – user configured entities, direct event from publishers to subscribers

Azure Time Series Insights

Analytics, storage, and visualization service for non-coders

Fast and easy analysis of event data via GUI

No coding required

Global view of IoT scale data

Edge Analytics

Local Execution

Stream analytics runs on ‘edge devices’

Unlocks value of untapped data

Only ~5% of data in industrial processes is sent to the cloud today

Deploy intelligence near the data to unlock the full value of data

Development, deployment, and operational consistency between cloud and edge

Stream analytics jobs run in the cloud and on edge devices

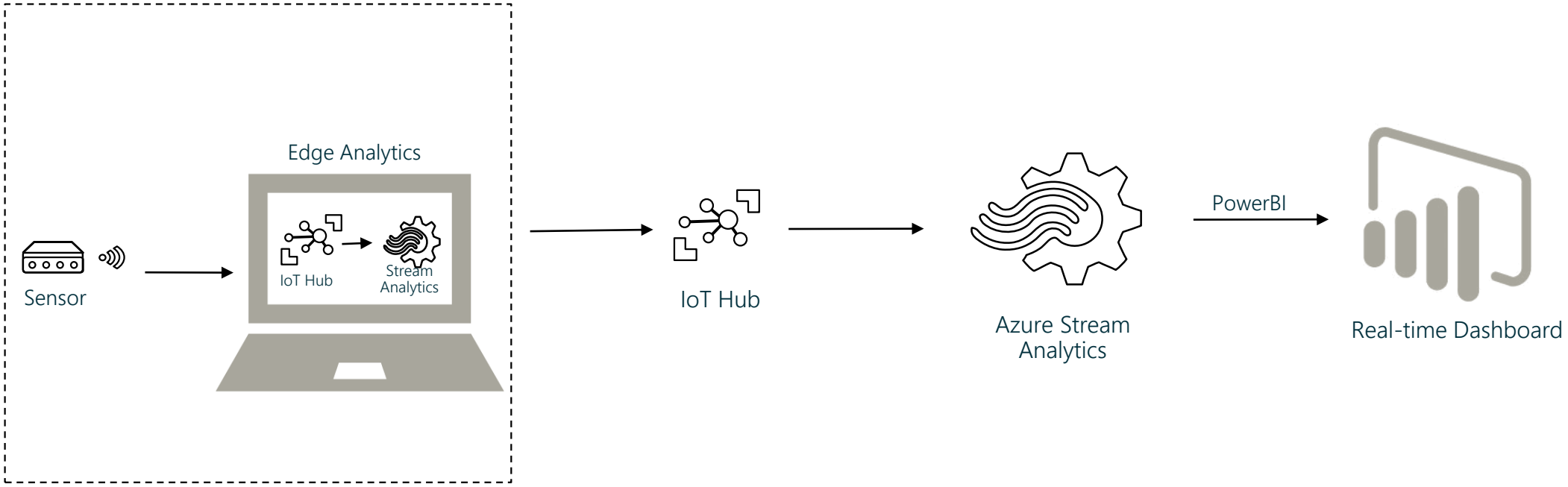
Intelligent action

Deploy situational awareness, custom code, and local execution of ML models on the edge

Industrial IoT Scenarios

- Low-latency command and control
  - Systems such as manufacturing production lines need to analyze and act in real-time to the streams of incoming data
- Sensor fusion on the edge
  - Integrate together sensor from different systems
- Compliance
  - Enables filtering or aggregation to remove PII data before sending it to the cloud
- Intermittent connectivity
  - Need of Resiliency: systems need to operate despite any interruption in the connectivity to the cloud.
- Local data reduction and transformation
  - Transforms raw input from sensors to meaningful information and enables scenarios with large volume of data

Edge Analytics Pipeline



Event Generation ⇨ Event Queuing & Stream Ingestion ⇨ Stream Analytics ⇨ Presentation & Action