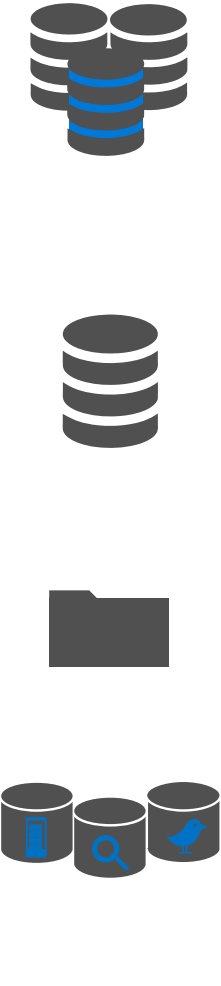# The Modern Data Warehouse
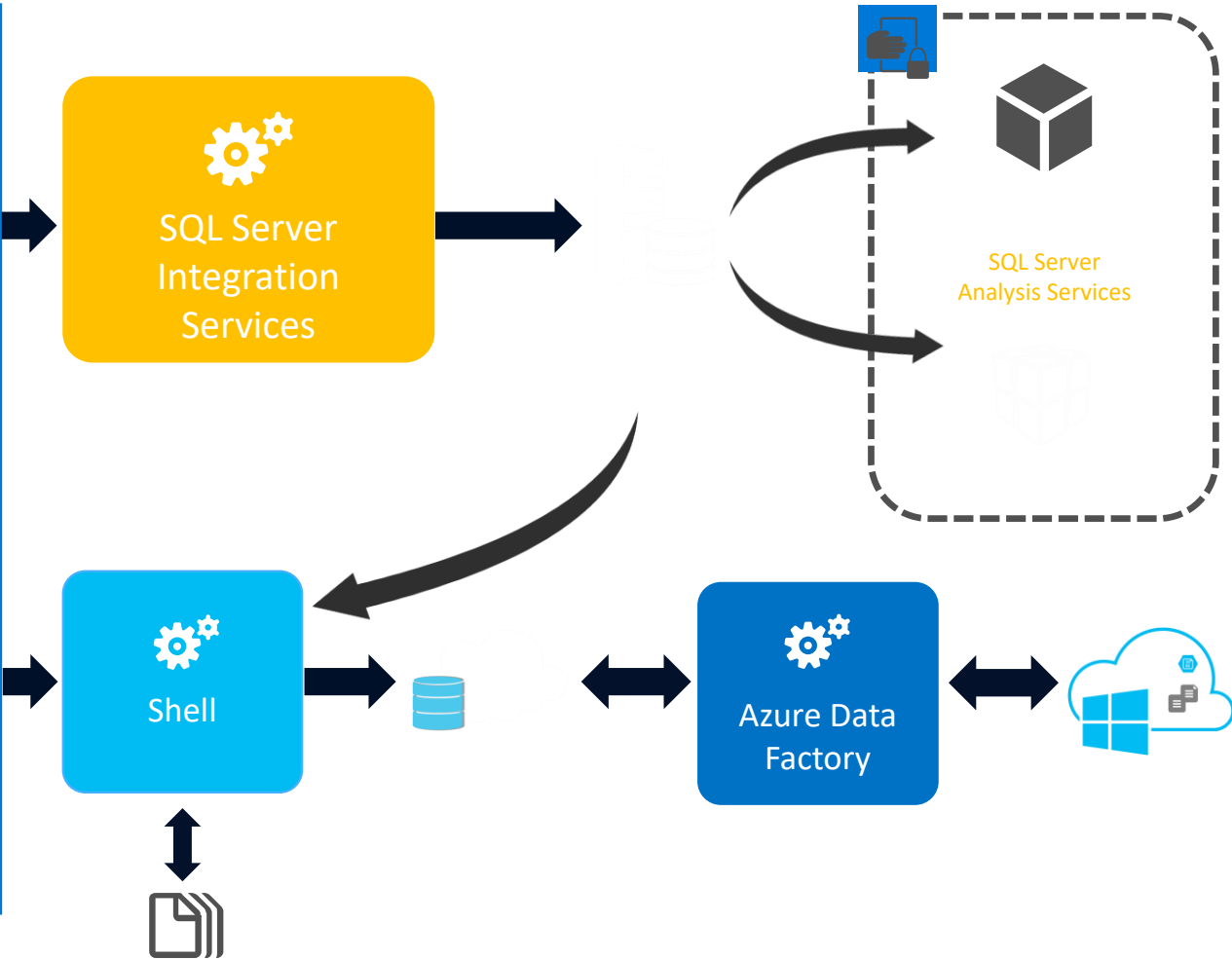
Kellyn Gorman, Azure Platform Architect, Microsoft

Microsoft

Student Analytics Architecture

Data Sources

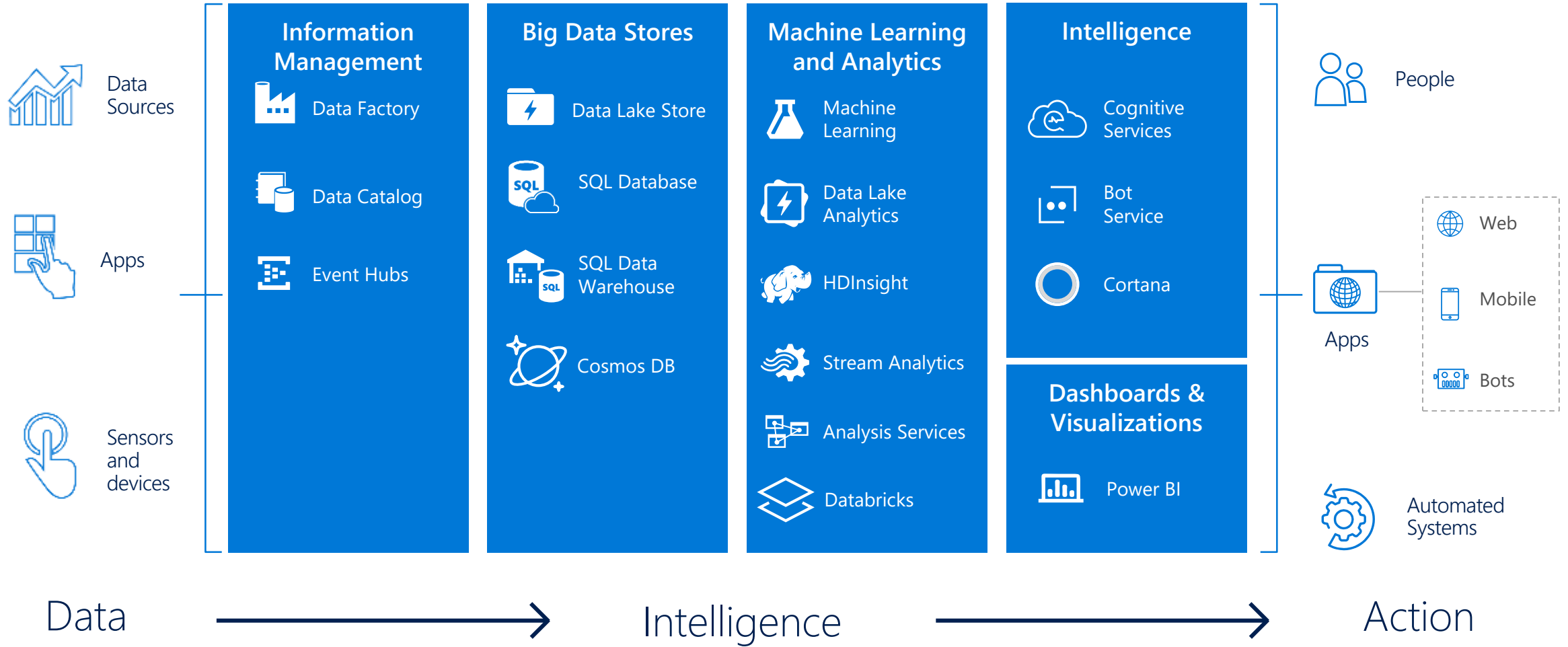**Integration/Analysis/Data**

Presentation

SQL Server Integration Services

SQL Server Analysis Services

Shell

Azure Data Factory

SharePoint

Q&A

Users

How to A to Z with Azure

Cortana Intelligence Suite
Transform data into intelligent action

Data Sources

Apps

Sensors & Devices

Cortana Intelligence

People

Apps

Automated systems

Data → Intelligence → Action

Easily turn data into intelligent action

Data Sources

Apps

Sensors and devices

Information Management
- Data Factory
- Data Catalog
- Event Hubs

Big Data Stores
- Data Lake Store
- SQL Database
- SQL Data Warehouse
- Cosmos DB

Machine Learning and Analytics
- Machine Learning
- Data Lake Analytics
- HDInsight
- Stream Analytics
- Analysis Services
- Databricks

Intelligence
- Cognitive Services
- Bot Service
- Cortana

Dashboards & Visualizations
- Power BI

People

Apps
- Web
- Mobile
- Bots

Automated Systems

Data → Intelligence → Action

Information Management

Information Management

- Data Factory
- Data Catalog
- Event Hubs

Data Sources

Apps

Sensors and devices

Data

Microsoft
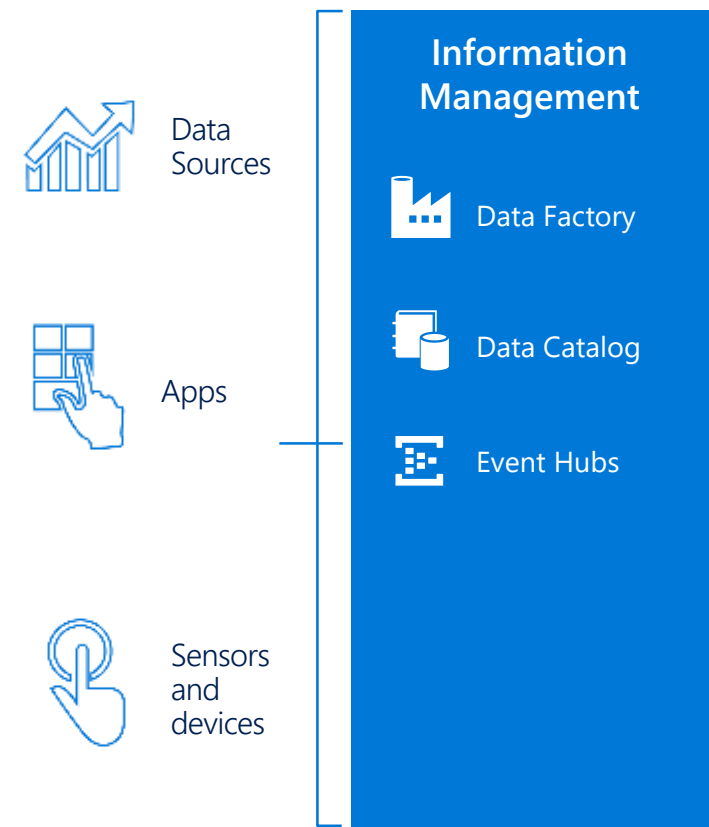
Compose and orchestrate data services at scale
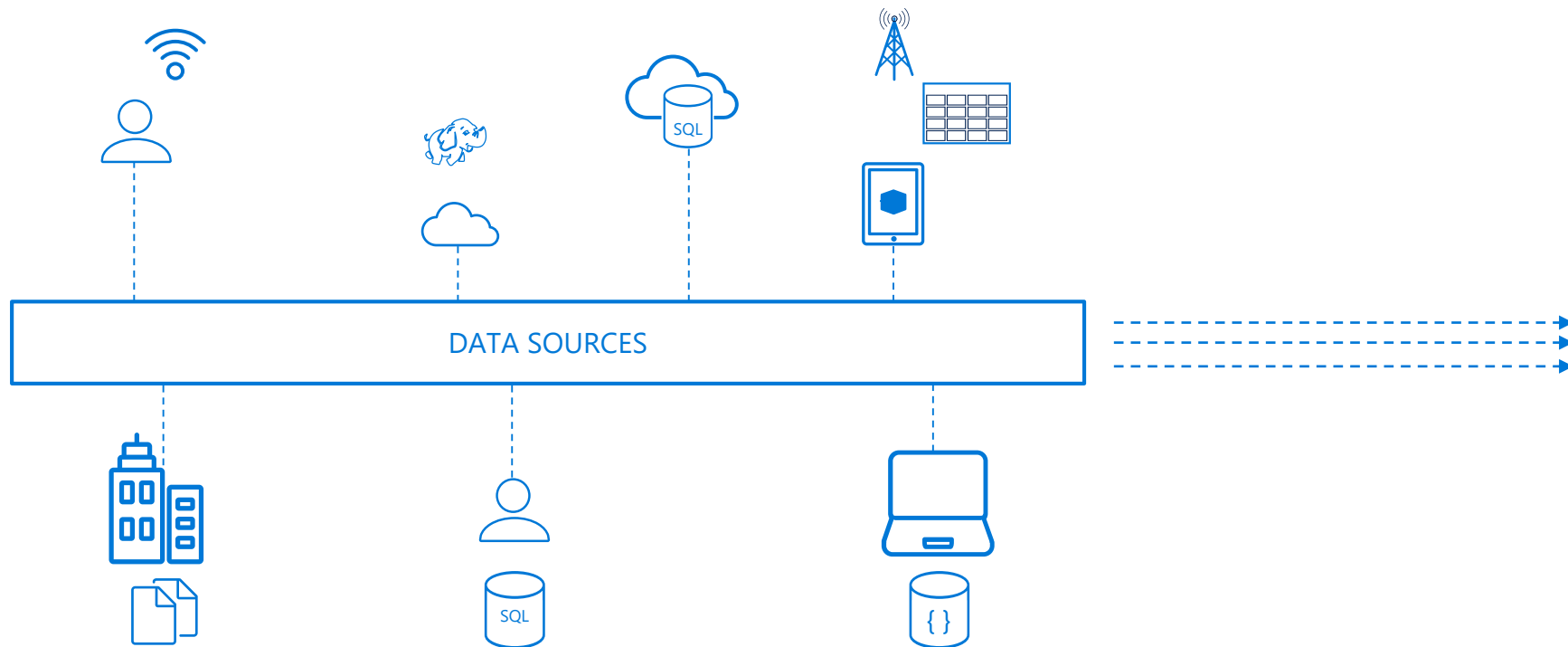
Information Management

Data Factory

Data Catalog

Event Hubs

DATA SOURCES

INGEST

- Create, schedule, orchestrate, and manage data pipelines
- Visualize data lineage
- Connect to on-premises and cloud data sources
- Monitor data pipeline health
- Automate cloud resource management
- Move relational data for Hadoop processing
- Transform with Hive, pig, or custom code

Microsoft

Azure Data Factory

- Hybrid data integration service that simplifies ETL & ELT at scale

- Work efficiently
- Fully accessible visual environment
- Simplified ETL/ELT with automation, visual workflows & data prep, templates
- Continuous integration & delivery (CI/CD)

- Cost-effective integration
- Serverless & fully managed, scales on demand
- Scale-out transformations via Spark
- Reduced overhead, SSIS in the cloud

- Connect with confidence
- All-inclusive connectivity, >80 + connectors
- Trusted, global cloud presence
- Multiple language support for coders

- Documentation
- Product Page
- ADF Videos
- Azure Data Factory Blog

- # Azure Data Factory Mapping Data Flows

  Public Preview @ //build May 2019

  GA Target H2 Calendar Year 2019

- # Azure Data Factory Wrangling Data Flows

  Limited Private Preview @ //build May 2019
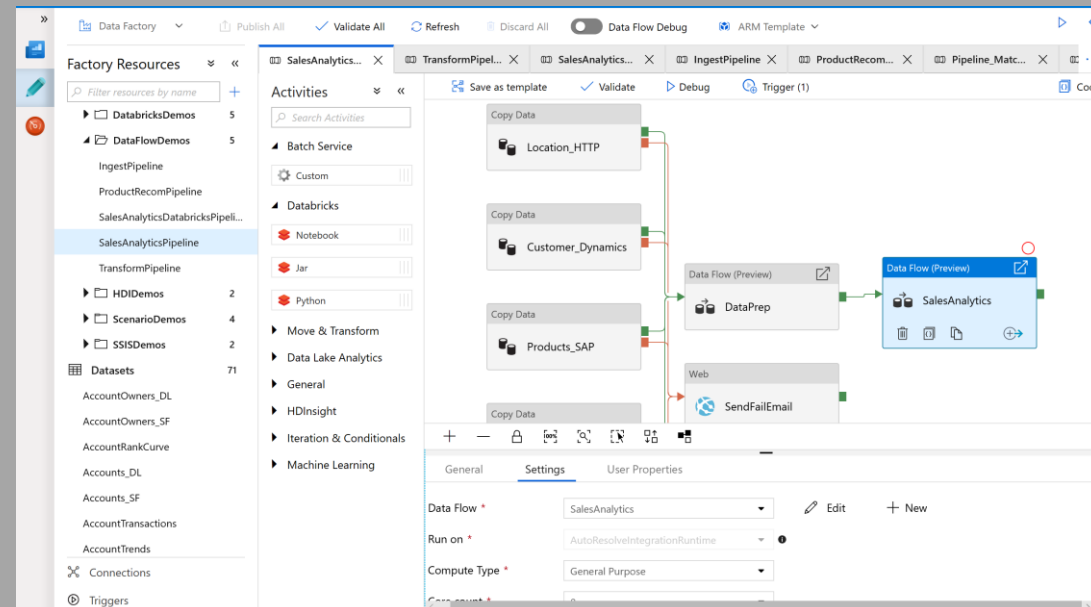
Azure Data Factory—Visual Data Transformation

Visually author, debug & monitor data flows at scale

GA

Update @ //build
May 2019

Transform data at scale

Platform-as-a-Service data transformation

Zero-code environment

No need to lean Spark, Scala, Java, Python ... Focus on business logic

Built-in Resiliency

Schema drift options provide support for flexible/shifting schemas

Elastic Scale

Zero-infrastructure Spark transformations that scale with your project needs
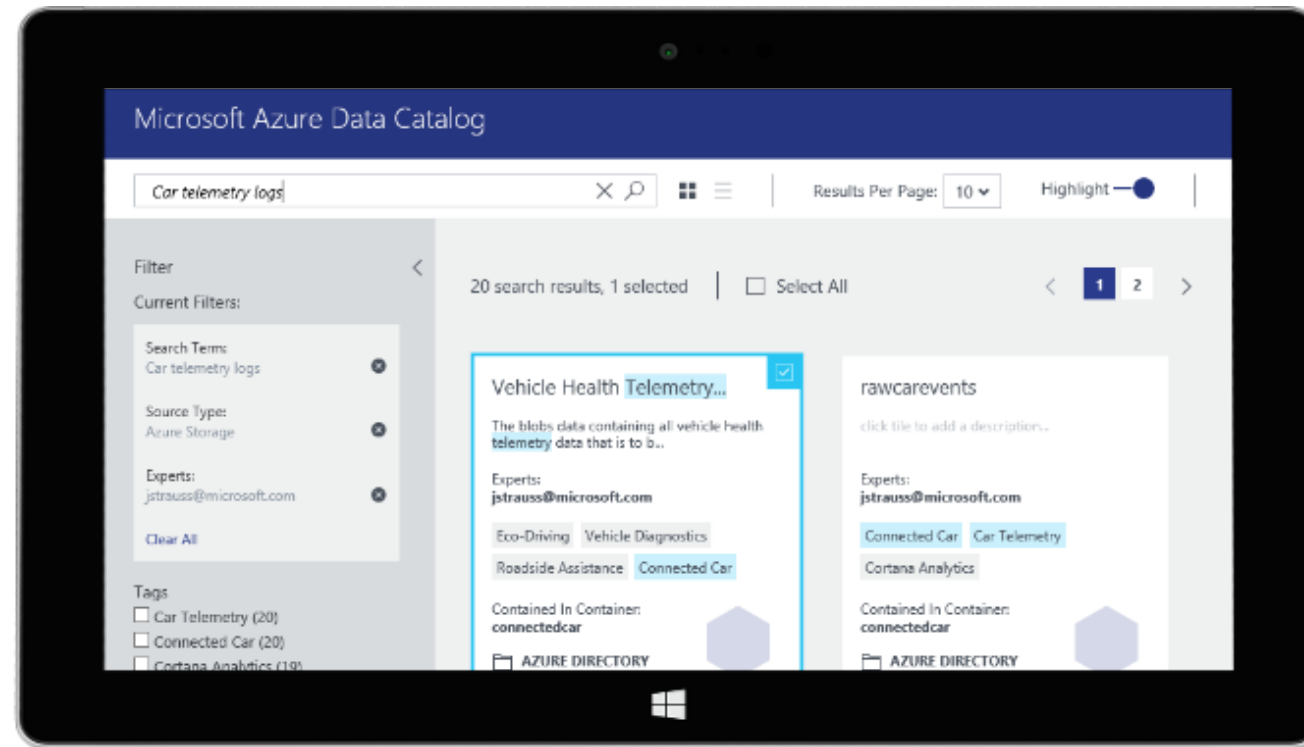
Learn more.

Get more value from your enterprise data assets

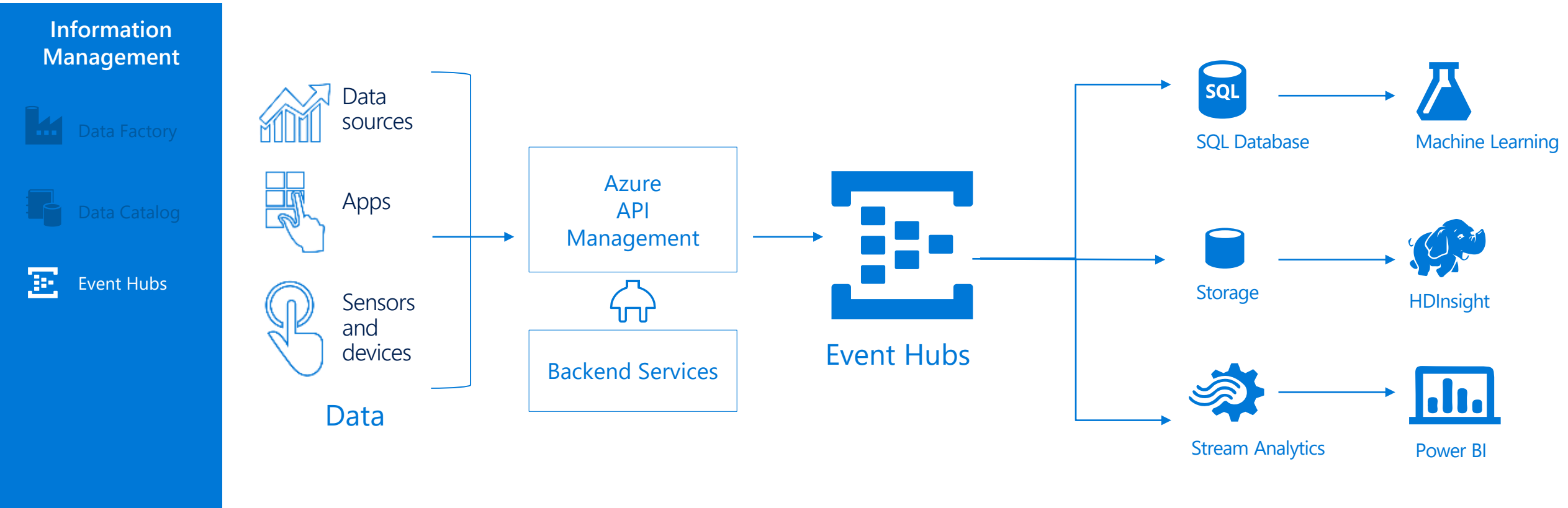

Information Management

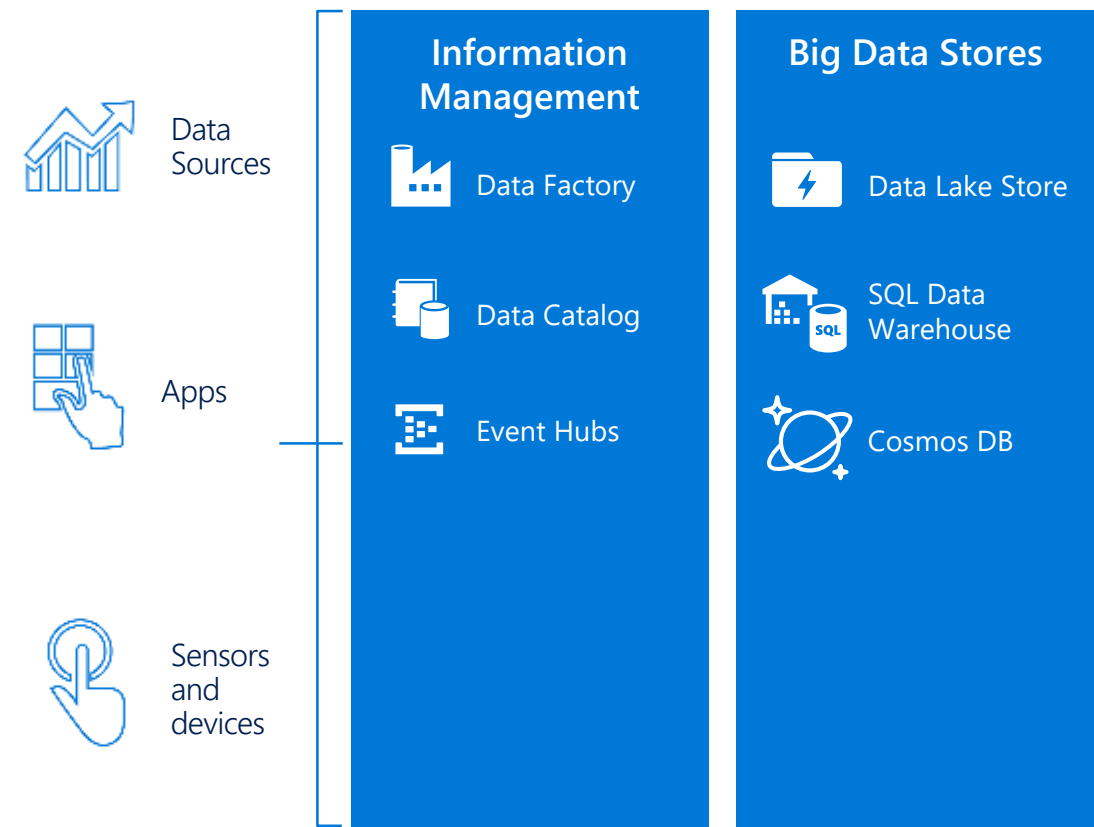Data Factory

Data Catalog

Event Hubs

- Spend less time looking for data, and more time getting value from it
- Register enterprise data sources, discover data assets and unlock their potential, and capture tribal knowledge to make data understandable
- Bridge the gap between IT and the business, allowing everyone to contribute their insights, tags, and descriptions

- Intuitive search and filtering to understand the data sources and their purpose
- Let your data live where you want; connect using tools you choose
- Integrate into existing tools and processes with open REST APIs

Microsoft

Ingest events from websites, apps and devices at cloud scale

Information Management

Data Factory

Data Catalog

Event Hubs

Data

Data sources

Apps

Sensors and devices

Azure API Management

Backend Services

Event Hubs

SQL Database

Machine Learning

Storage

HDInsight

Stream Analytics

Power BI

- Log millions of events per second in near real time
- Connect devices using flexible authorization and throttling
- Use time-based event buffering
- Get a managed service with elastic scale

- Get a managed service with elastic scale
- Reach a broad set of platforms using native client libraries
- Pluggable adapters for other cloud services

Big Data Stores

Data Sources

Apps

Sensors and devices

Information Management

- Data Factory
- Data Catalog
- Event Hubs

Big Data Stores

- Data Lake Store
- SQL Data Warehouse
- Cosmos DB

Data

Microsoft

Comply with industry standards

Compliance is verified by third-party audits and certifications

CCSL (IRAP)	FISC
China GB 18030	HIPAAHITECH
CJIS	ISO 27001
CSA STAR Attestation	ISO 27018
CSA STAR Certification	MeitY
DoD	NZ CC Framework
EU Model Clauses	PCI DS
EU-US Privacy Shield	SOC 1
FedRAMP	SOC 2
FERPA	SOC 3
FIPS 140-2	UK G-Cloud

Find out more about <u>Compliance</u> in the Microsoft Trust Center

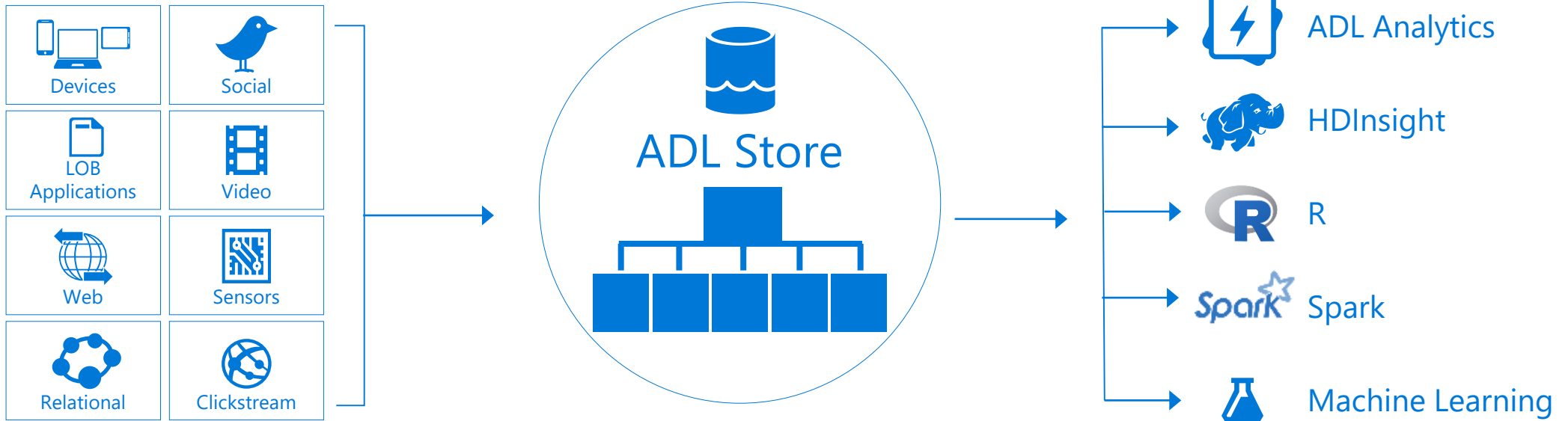# A hyper-scale repository for big data analytics workloads

Big Data Stores

- Data Lake Store
- SQL Data Warehouse
- Cosmos DB

Devices
Social
LOB Applications
Video
Web
Sensors
Relational
Clickstream

ADL Store

ADL Analytics
HDInsight
R
Spark
Machine Learning

- 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
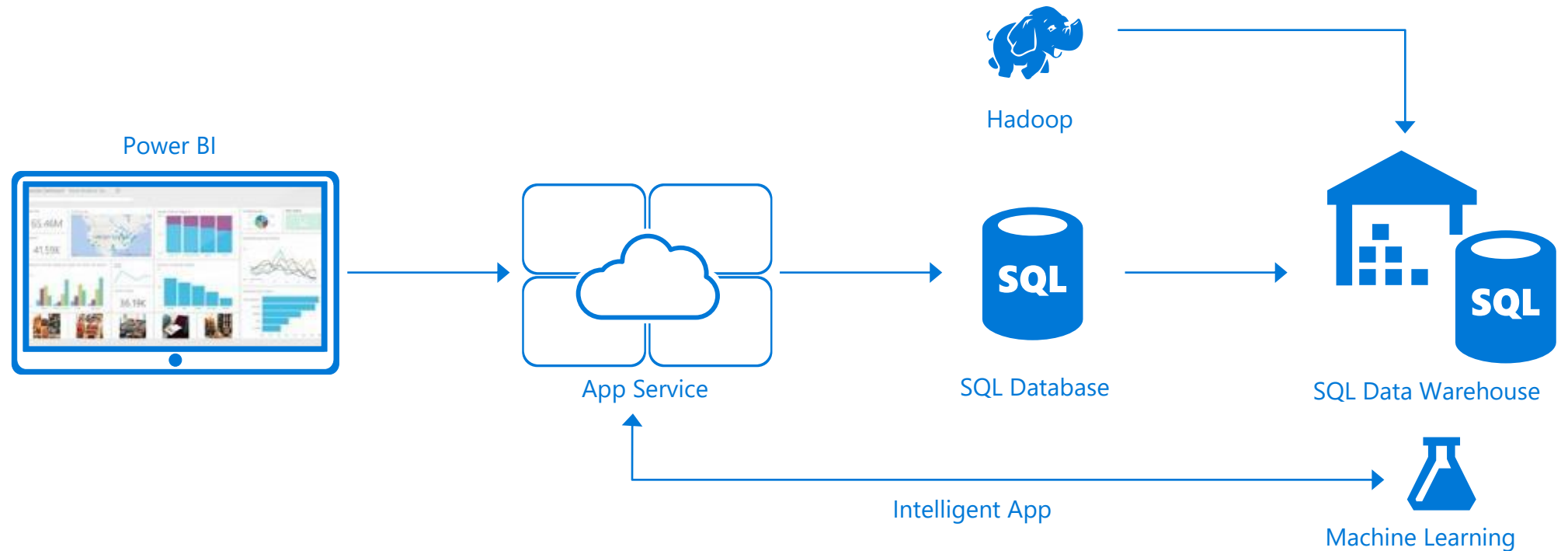- Azure Active Directory access control

Microsoft

Elastic data warehouse as a service with enterprise-class features

Big Data Stores

Data Lake Store

SQL Data Warehouse

Cosmos DB

Power BI

App Service

Hadoop

SQL Database

SQL Data Warehouse

Intelligent App

Machine Learning

- Petabyte scale with massively parallel processing
- Independent scaling of compute and storage—in seconds
- Transact-SQL queries across relational and non-relational data

- Full enterprise-class SQL Server experience
- Works seamlessly with Power BI, Machine Learning, HDInsight, and Data Factory
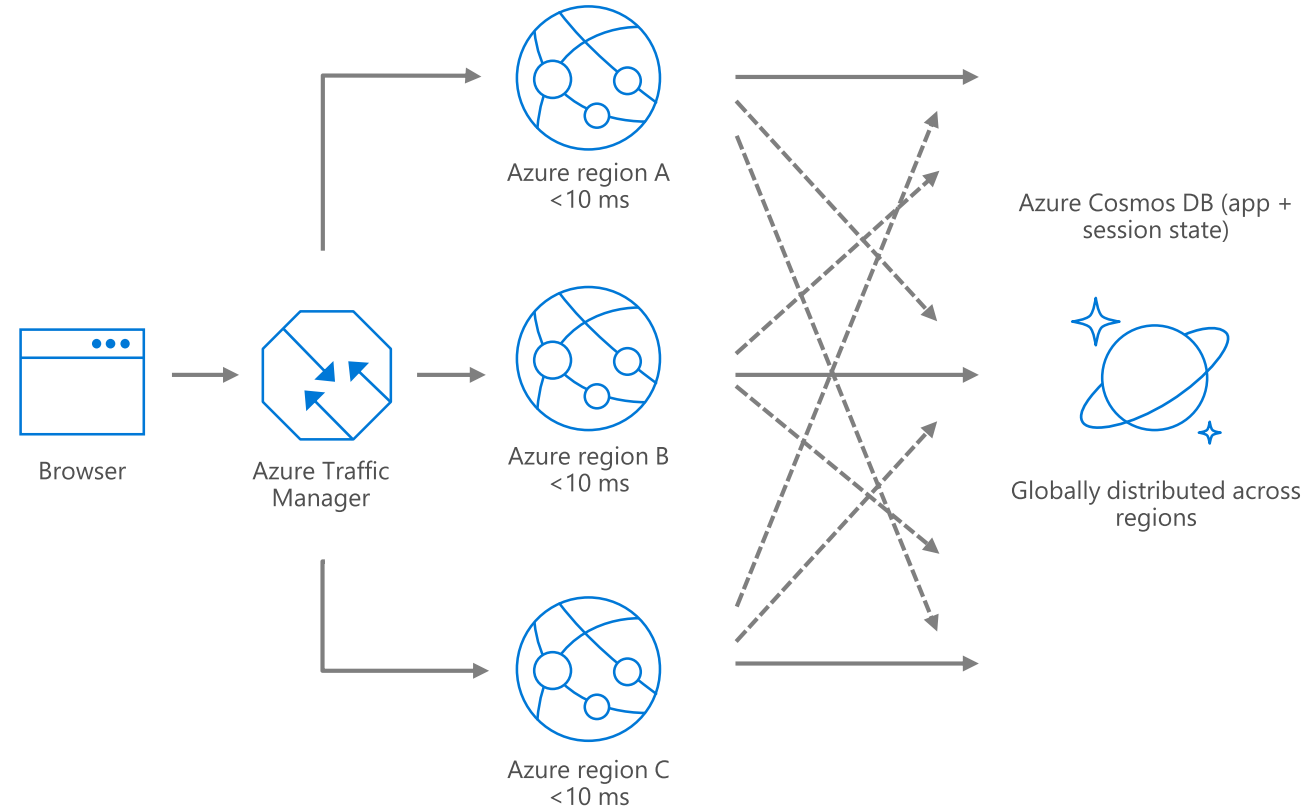
Microsoft

Data distributed, globally available, massive scale, low latency

Big Data Stores

Data Lake Store

SQL Data Warehouse

Cosmos DB

Browser → Azure Traffic Manager

Azure region A
<10 ms

Azure region B
<10 ms

Azure region C
<10 ms

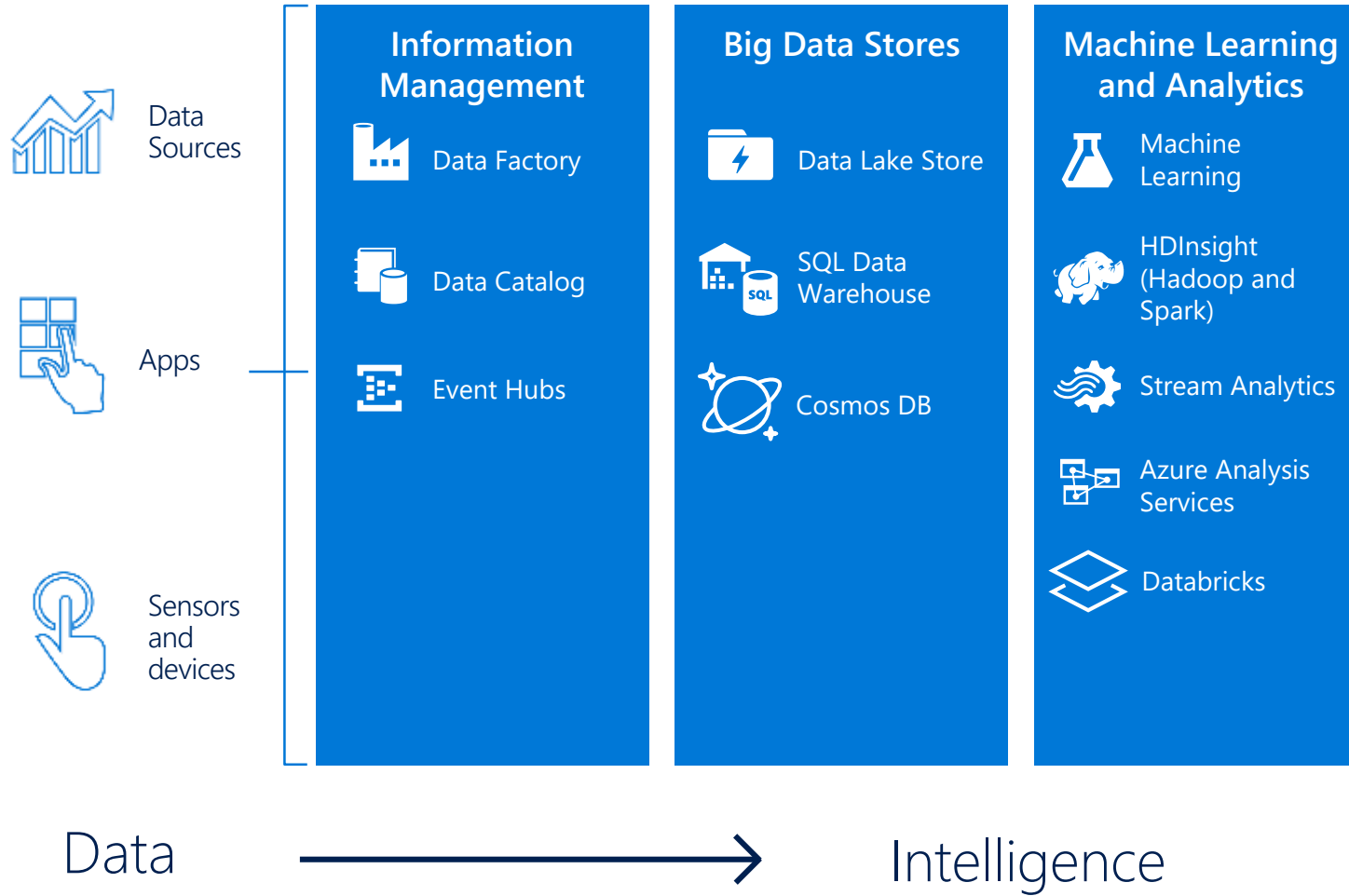Azure Cosmos DB (app + session state)

Globally distributed across regions

- Turkey global data replication across all Azure region
- Guaranteed low-latency experience for global users
- Resiliency for high availability and disaster recovery

- Elastically scale throughput across multiple regions
- Support for requests/sec for different workloads
- Pay only for the throughput and storage you need

Microsoft

Machine Learning and Analytics

Data Sources
- Data Sources
- Apps
- Sensors and devices

Information Management
- Data Factory
- Data Catalog
- Event Hubs

Big Data Stores
- Data Lake Store
- SQL Data Warehouse
- Cosmos DB

Machine Learning and Analytics
- Machine Learning
- HDInsight (Hadoop and Spark)
- Stream Analytics
- Azure Analysis Services
- Databricks

Data → Intelligence

Microsoft

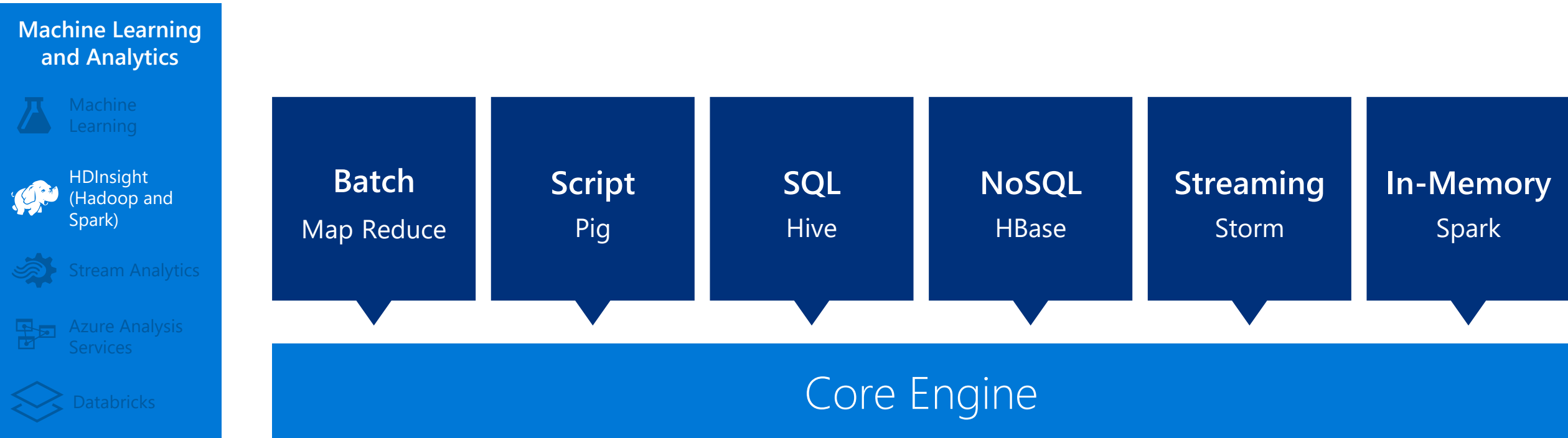# Easily build, deploy, and share predictive analytics solutions

Machine Learning and Analytics

Machine Learning

HDInsight (Hadoop and Spark)

Stream Analytics

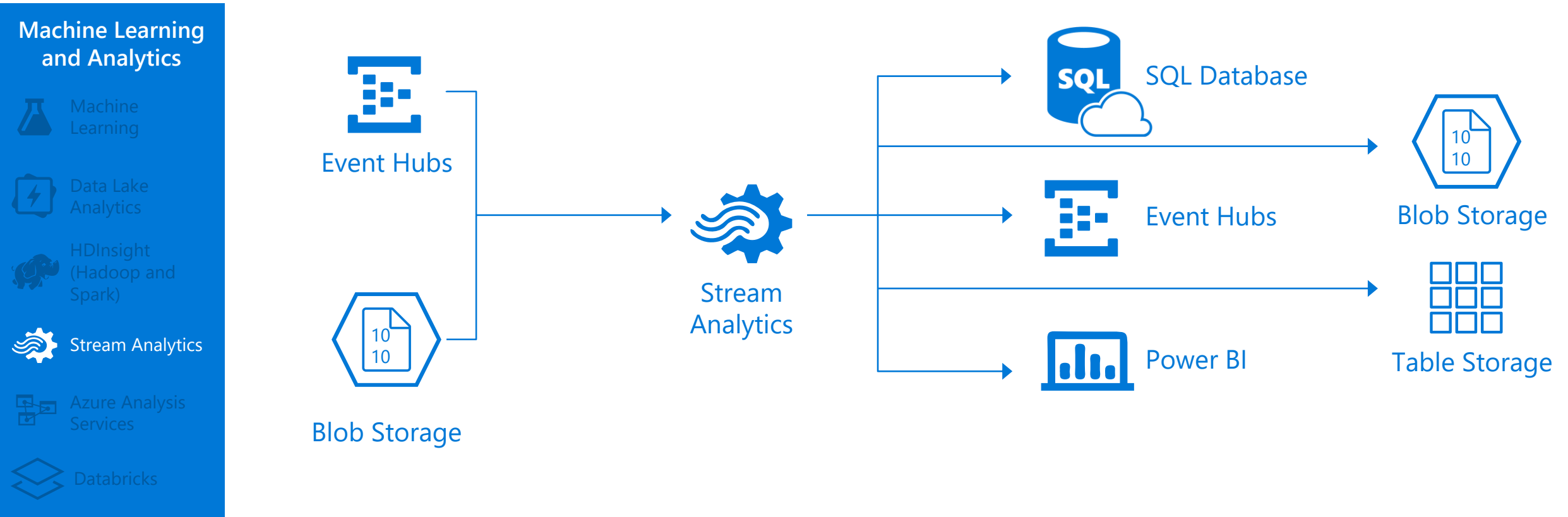Azure Analysis Services

Databricks
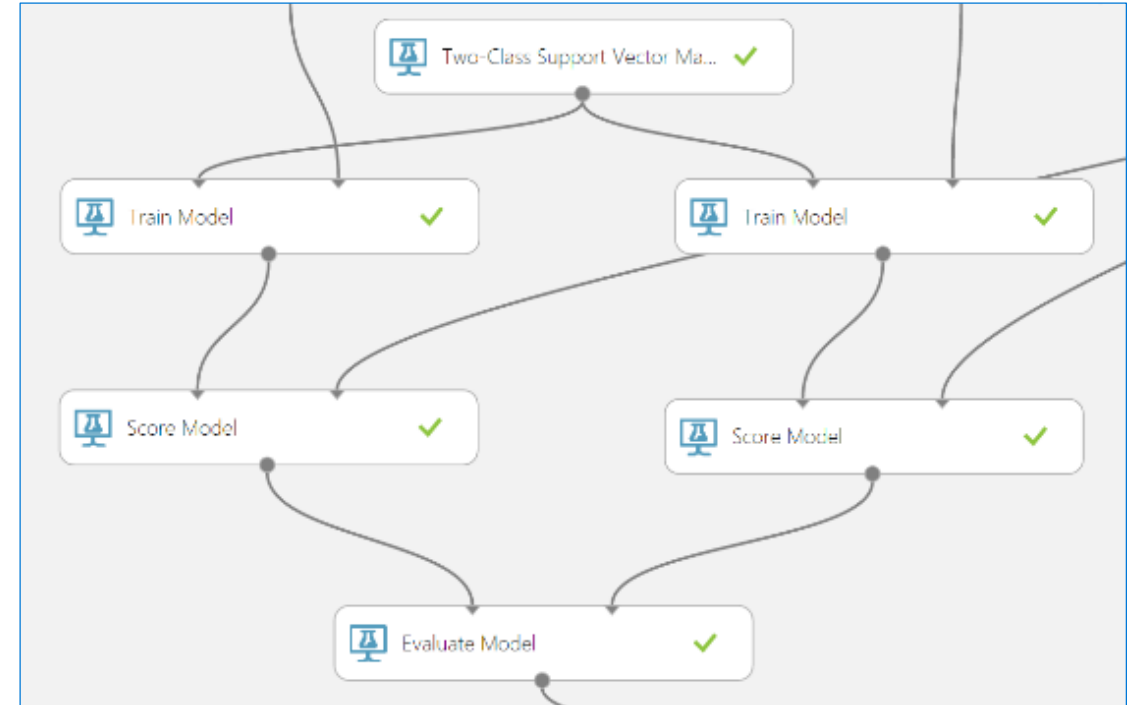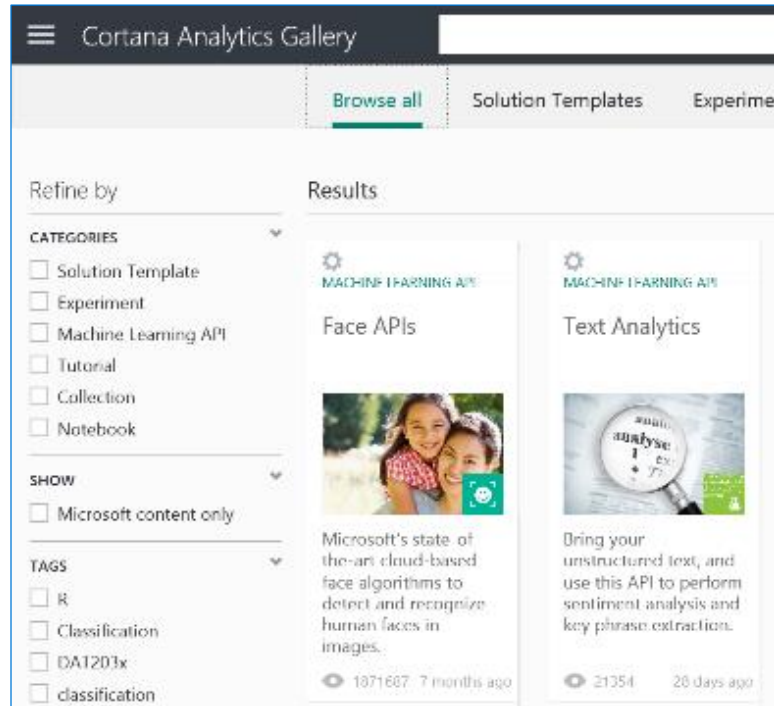
- Simple, scalable, cutting edge. A fully managed cloud service that enables you to easily build, deploy, and share predictive analytics solutions.
- Deploy in minutes. Azure Machine Learning means business. You can deploy your model into production as a web service that can be called from any device, anywhere and that can use any data source.
- Publish, share, monetize. Share your solution with the world in the Gallery or on the Azure Marketplace.

Microsoft

Comprehensive set of managed Apache big data projects

Machine Learning and Analytics

- Machine Learning
- HDInsight (Hadoop and Spark)
- Stream Analytics
- Azure Analysis Services
- Databricks

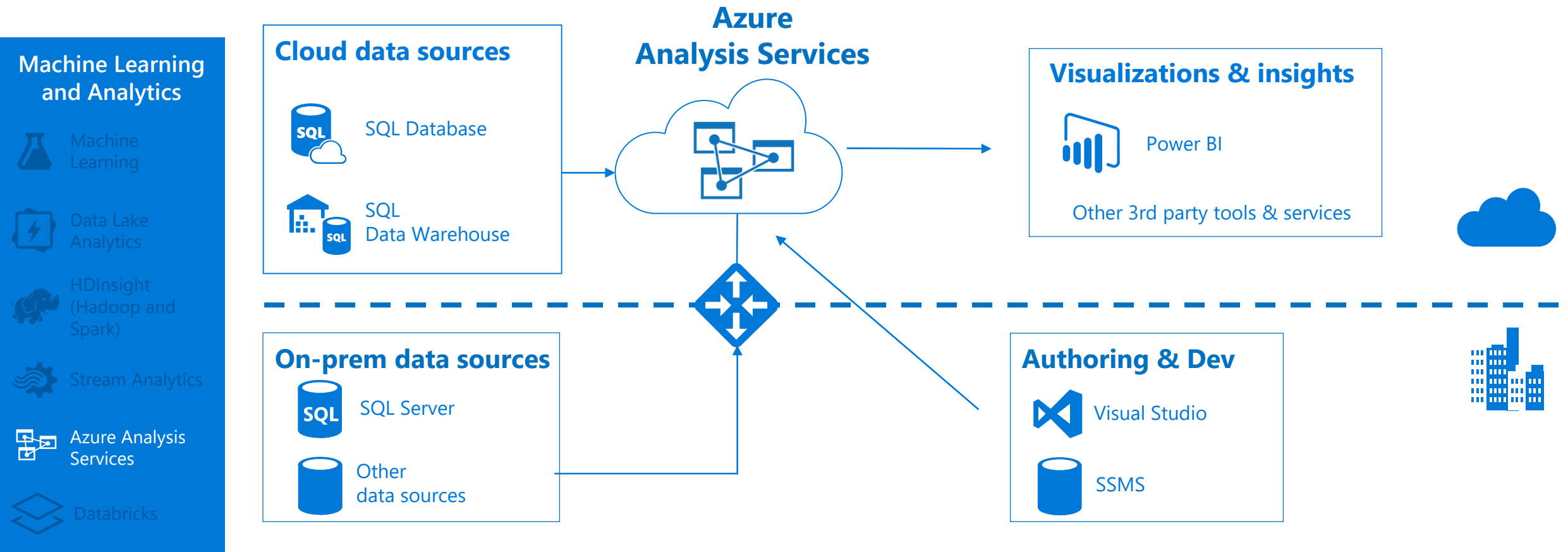| **Batch** Map Reduce	**Script** Pig	**SQL** Hive	**NoSQL** HBase	**Streaming** Storm	**In-Memory** Spark

Core Engine

- Scale to petabytes on demand
- Process unstructured and semi-structured data
- Develop in Java, .NET, and more
- Skip buying and maintaining hardware

- Deploy in Windows or Linux
- Spin up an Apache Hadoop cluster in minutes
- Visualize your Hadoop data in Excel
- Easily integrate on-premises Hadoop clusters

Microsoft

Real-time stream processing in the cloud

Machine Learning and Analytics

- Machine Learning
- Data Lake Analytics
- HDInsight (Hadoop and Spark)
- Stream Analytics
- Azure Analysis Services
- Databricks

Event Hubs

Blob Storage

Stream Analytics

SQL Database

Event Hubs

Power BI

Blob Storage

Table Storage

- Perform real-time analytics for your Internet of Things solutions
- Stream millions of events per second
- Get mission-critical reliability and performance with predictable results

- Create real-time dashboards and alerts over data from devices and applications
- Correlate across multiple streams of data
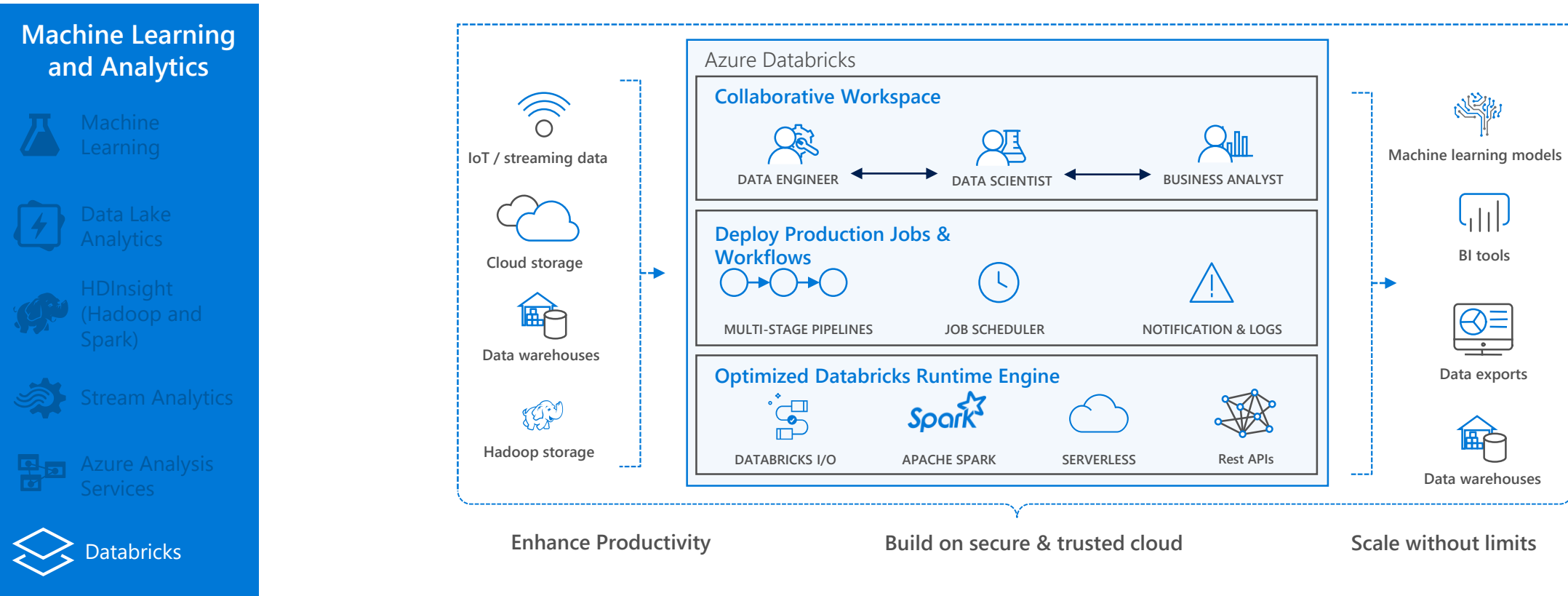- Use familiar SQL-based language for rapid development

Microsoft

Proven analytics engine

Machine Learning and Analytics

Machine Learning

Data Lake Analytics

HDInsight (Hadoop and Spark)

Stream Analytics

Azure Analysis Services

Databricks

Azure Analysis Services

Cloud data sources

SQL Database

SQL Data Warehouse

On-prem data sources

SQL Server

Other data sources

Visualizations & insights

Power BI

Other 3rd party tools & services

Authoring & Dev

Visual Studio

SSMS

- Get started quickly without managing an infrastructure
- Scale resources to match your business needs

- Visualize your data using your favorite data visualization tool
- Govern, deploy, test, and deliver your BI solution with confidence

Microsoft

Fast, easy, & collaborative Apache Spark-based analytics platform

Machine Learning and Analytics

- Machine Learning
- Data Lake Analytics
- HDInsight (Hadoop and Spark)
- Stream Analytics
- Azure Analysis Services

Databricks

IoT / streaming data

Cloud storage

Data warehouses

Hadoop storage

Azure Databricks

Collaborative Workspace

DATA ENGINEER ↔ DATA SCIENTIST ↔ BUSINESS ANALYST

Deploy Production Jobs & Workflows

MULTI-STAGE PIPELINES   JOB SCHEDULER   NOTIFICATION & LOGS

Optimized Databricks Runtime Engine

DATABRICKS I/O   APACHE SPARK   SERVERLESS   Rest APIs

Machine learning models

BI tools

Data exports

Data warehouses

Enhance Productivity   **Build on secure & trusted cloud**   **Scale without limits**

- Designed in collaboration with the founders of Apache Spark
- One-click set up; streamlined workflows
- Interactive workspace that enables collaboration between data scientists, data engineers, and business analysts

- Native integration with Azure services (Power BI, SQL DW, Cosmos, DB, Blob Storage)
- Enterprise-grade Azure security (Active Directory integration, compliance, enterprise-grade SLAs)

Microsoft
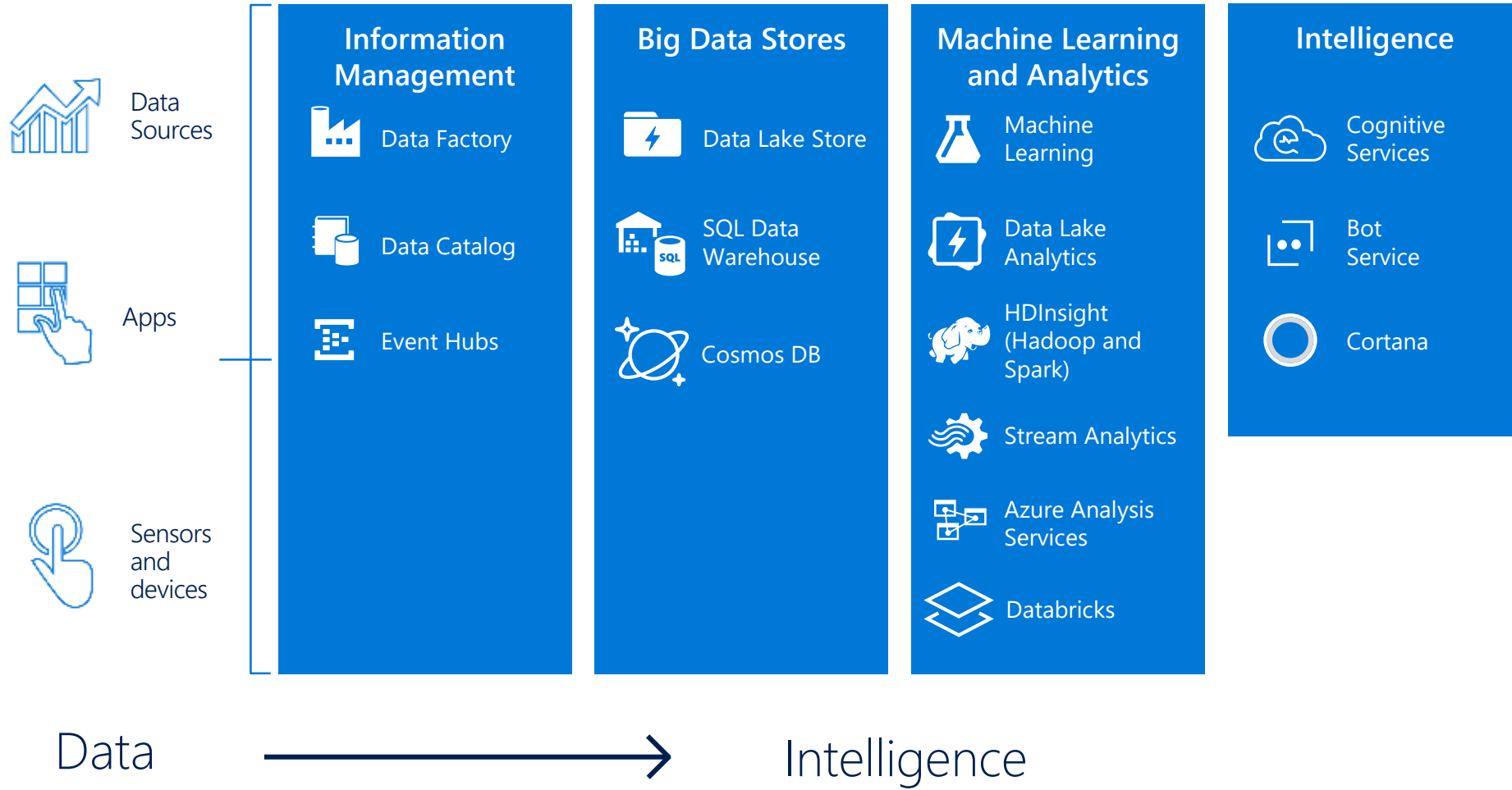
Build applications that understand people

Intelligence	👁 Vision	🎙 Speech	💬 Language	💡 Knowledge	🔍 Search
☁ Cognitive Services	Computer Vision	Speaker Recognition	Text Analytics	Academic	Bing Web Search API
Bot Service	Content Moderator	Bing Speech	Bing Spell Check	Entity Linking	Bing Image Search API
ⵔ Cortana	Face	Custom Speech Service	Web Language Model	Knowledge Exploration	Bing Video Search API
	Emotion		Linguistic Analysis	Recommendations	Bing News Search API
	Video		Language Understanding	QnA Maker	Bing Auto Suggest API
			Translator		

- Faces, images, emotion recognition and video intelligence
- Spoken language processing, speaker recognition, custom speech recognition
- Natural language processing, sentiment and topics analysis, spelling errors

- Complex tasks processing, knowledge exploration, intelligent recommendations
- Bing engine capabilities for Web, Autosuggest, Image, Video and News

Microsoft

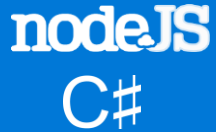# Your bots – wherever your users converse

Intelligence

Cognitive Services

Bot Service

Cortana

Accelerate development cycles

nodeJS
C#

Develop your way

Built in code editor

Quick start templates

Integrated chat window

Enrich your bots

Channel support

Cognitive Services

API

Direct Line support

Embedded web chat

Boost operational efficiencies

Powered by Azure Functions

Continuous deployment

Scale on demand

Reduced dev ops

- Start quickly with built-in templates
- Reach your customers on multiple channels
- Boost the power of bots with Azure services

- No server management or patching needed
- Scale out automatically
- Pay only for what you use

Microsoft

# Get things done in more helpful, proactive and natural ways

Intelligence

Cognitive Services

Bot Service

Cortana

Here are some of the things I can help you with...	Cortana for Consumers (today)	With the Cortana Intelligence
Answers	Public reference data answers – *"How far is it from Los Angeles to San Francisco?"*	Answers from organizational data in Power BI *"What were our biggest deals that closed last month?"*
Predictions	Event predictions – *"Who do you think is going to win the Germany Italy game?"*	Integration with prediction solutions *"Which of our customers are most likely to churn in the next quarter?"*
Monitoring & Alerts	Flight status, traffic conditions, changes in weather, ...	Monitoring KPIs and preemptive alerting *"Alert me if this customer ever has a 90% chance of churn in the next 30 days"*
Task Completion	Setting reminders, scheduling meetings, getting directions, ...	Line of business process integration *Assistance with expense report submission on-time within policy*

Keep a pulse on your business with live, interactive dashboards

Dashboards & Visualizations

Power BI

Stream Analytics

Event Hubs

Machine Learning

Storage
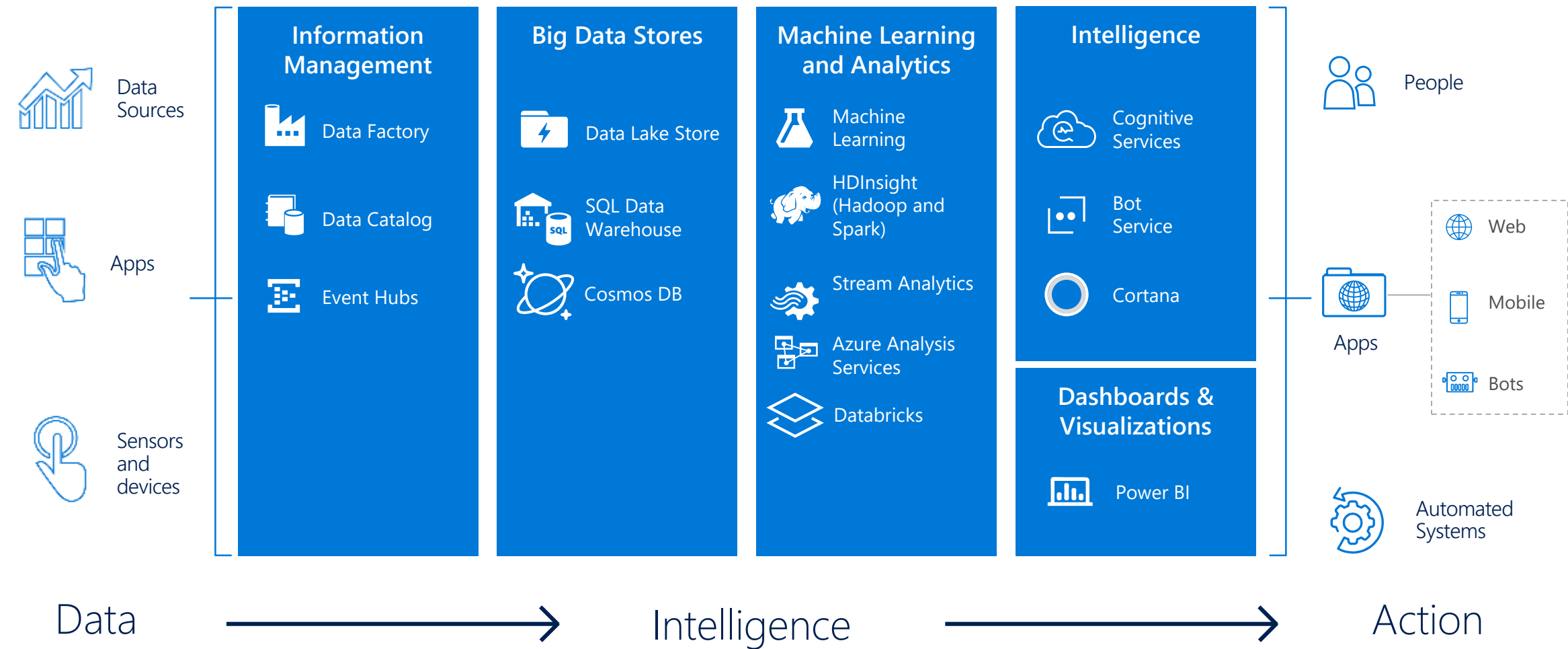
SQL Database

HDInsight

Power BI
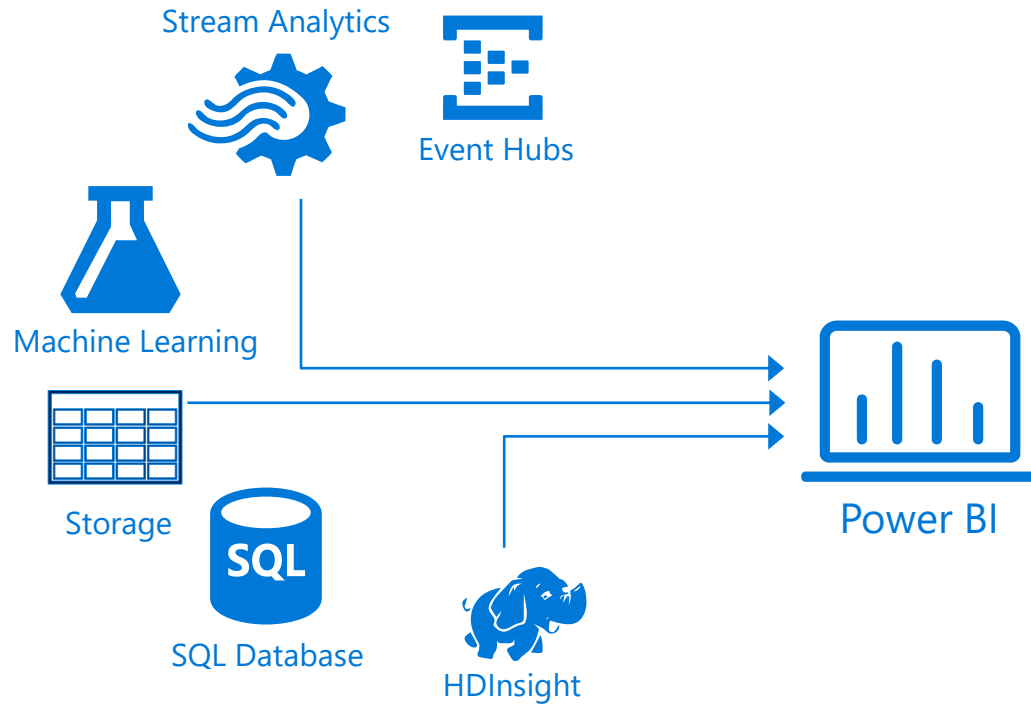
- Analytics for everyone, even non-data experts
- Your whole business on one dashboard
- Create stunning, interactive reports

- Drive consistent analysis across your organization
- Embed visuals in your applications
- Get real-time alerts when things change

Microsoft

Transform data into intelligent action

Data Sources

Apps

Sensors and devices

Information Management
- Data Factory
- Data Catalog
- Event Hubs

Big Data Stores
- Data Lake Store
- SQL Data Warehouse
- Cosmos DB

Machine Learning and Analytics
- Machine Learning
- HDInsight (Hadoop and Spark)
- Stream Analytics
- Azure Analysis Services
- Databricks

Intelligence
- Cognitive Services
- Bot Service
- Cortana

Dashboards & Visualizations
- Power BI

People

Apps
- Web
- Mobile
- Bots

Automated Systems

Data ⟶ Intelligence ⟶ Action

Microsoft

Demo

Azure Data Lake Storage Gen2

The scale and cost-effectiveness of Azure Blob Storage with advanced data lake capabilities:

- Hadoop compatible file system interface for Azure Blob Storage

- Fine grained file and folder permissions (ACLs)

- Atomic file system operations for faster job execution with fewer transactions

- Full support for all Blob features (AAD Integration, Zone Redundant and RA-Geo Redundant Storage)

- Pricing at Blob Storage levels

- Public Preview for Azure Data Lake Storage Gen2 support for Event Grid Notifications. (June)

New
- Public Preview for Blob API Interoperability with Azure Data Lake Storage Gen2 (July):

  - **Manage the same data using Blob or ADLS Gen2 API**

  - **ADLS Gen Partner Ecosystem == Blob Partner Ecosystem**

  - **Better ADLS Gen2 integration with additional Azure First Party services (Event Hub)**

[Learn more](#)

ADLS Gen2 is a "no-compromises" data lake

Secure	Fast	Scalable	Cost effective
• Support for fine-grained ACLs, protecting data at the file and folder level	• Atomic file operations means jobs complete faster	• No limits on data store size	• Object store pricing levels
• Multi-layered protection via at-rest Storage Service encryption and Azure Active Directory integration		• Global footprint (50 regions)	• File system operations minimize transactions required for job completion

**July launch for Blob API Interoperability Public Preview!
Coming Soon: Power BI Desktop Connector for ADLS Gen2**

The MDW Lifecycle in Azure

Azure

1 Data sources
- Relational databases
- File exports
- Big data sources

2 Azure Data Factory
- PaaS hybrid ETL/ELT service
- Can host SSIS packages

4 Azure Data Lake Store
- Hyper-scale repository for big data analytic workloads
- Ingest data of any size, type, velocity

7 Azure SQL Data Warehouse
- Massively Parallel Processing SQL PaaS
- Facts & dimensions
- Petabyte scale
- Pause compute resources
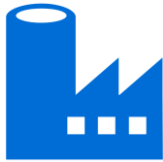
Reference Architecture

Azure

Azure Databricks
- Apache Spark-based analytics platform
- Collaboration between data scientists, data engineers, and business analysts

Data sources
- Relational databases
- File exports
- Big data sources

Azure Data Factory
- PaaS hybrid ETL/ELT service
- Can host SSIS packages

Azure Data Lake Store
- Hyper-scale repository for big data analytic workloads
- Ingest data of any size, type, velocity

Azure SQL Data Warehouse
- Massively Parallel Processing SQL PaaS
- Facts & dimensions
- Petabyte scale
- Pause compute resources

Azure Analysis Services
- PaaS semantic model
- Centralized calculations, hierarchies, KPIs, etc.
- In-memory, compressed

Visualization Tools

Tableau Server
- Tableau Server on Azure infrastructure

Qlik Sense Enterprise
- Qlik Sense on Azure infrastructure

Excel
- Self-service analytics
- Pivot tables & charts

Power BI
- SaaS analytics
- Excel integration
- Seamless, native integration w/ Azure data sources

Reference Architecture

Azure

Azure Monitor
- Monitor cloud & on-premises environments to maintain performance & availability

Azure Data Catalog
- Fully managed cloud metadata repository
- Enable data discovery & capturing team tribal knowledge

Data sources
- Relational databases
- File exports
- Big data sources

Azure Data Factory
- PaaS hybrid ETL/ELT service
- Can move data via Data Factory Pipelines or SSIS packages

Data Lake

Azure Data Lake
- General purpose storage
- Hyper-scale repository for big data analytic workloads
- Capture data of any size, type, velocity

Azure SQL DW (staging)
- Massively Parallel Processing SQL PaaS
- Staging layer

Enterprise Data Warehouse

Azure SQL DW
- Massively Parallel Processing SQL PaaS
- Facts & dimensions

Data Marts & Virtualized Layer

Azure Analysis Services
- PaaS semantic model
- Centralized calculations, hierarchies, KPIs, etc.
- In-memory, compressed

Azure SQL Database
- PaaS SQL database
- Built in DR & HA
- On-demand scale

Data Science Tools

Deep Learning VM
- GPU based VM for training deep learning models

Data Science VM
- Customized pre-configured VM for data science pros

Azure Databricks
- Apache Spark analytics platform
- Build machine learning experiments in notebooks w/ Scala, R, Python, and/or SQL

Azure Machine Learning
- Predictive analytics services for creating, deploying, & managing predictive models

Visualization Tools

Power BI
- SaaS analytics
- Excel integration
- Seamless, native integration w/ Azure data sources

Excel
- Self-service analytics
- Pivot tables & charts

Power BI Report Server
- Paginated reports
- Pixel perfect reports
- Document generation
- Data driven subscriptions

3rd Party Visualization Tools
- Tools connecting to SQL & Analysis Services

Azure Active Directory
- Identity management & authentication across Azure resources

Kellyn Gorman
kellyn.gorman@microsoft.com