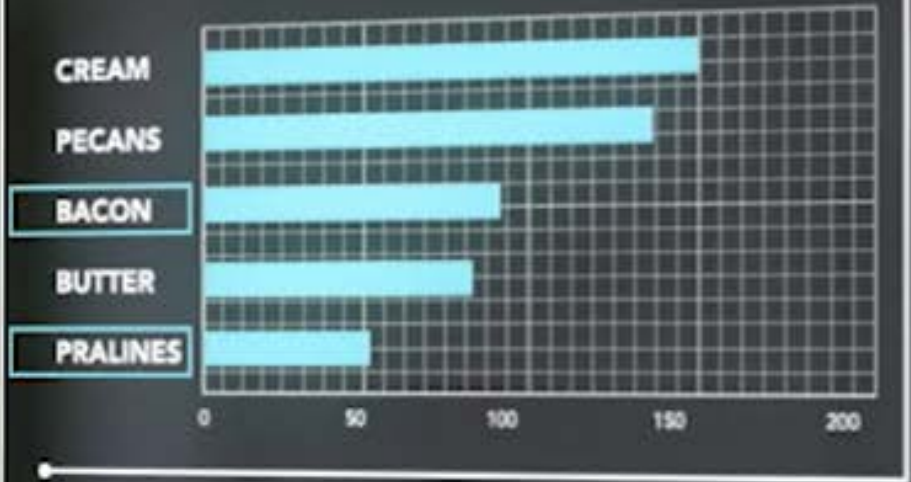


**BEST SELLER:  
PECANS & CREAM**

▣ SOCIAL AFFINITY SEARCH



**SENTIMENT ANALYSIS: BACON + PRALINES**




# Big Data in Azure


Erik Gullbring  
[ergullbr@microsoft.com](mailto:ergullbr@microsoft.com)

# The New Microsoft - Open Source

**InfoWorld**  
FROM IDG INSIDER

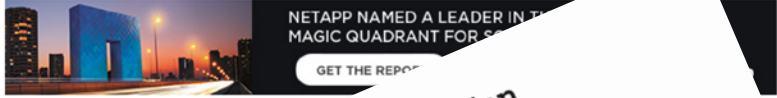
Home > Software > Open Source

 **INFOWORLD TECH WATCH**  
By [Matt Asay](#) | [Follow](#)

About  Informed news analysis every weekday

## Who's No. 1 in open source? Microsoft!

When the heart of your business is about delivering cloud services, it's easy to love open source



InfoWorld | Sep 21, 2016

[Twitter](#) [Facebook](#) [LinkedIn](#) [Google+](#) [Reddit](#) [StumbleUpon](#) [Email](#)

www.infoworld.com

## Google joins Microsoft's .NET Foundation

EMIL PROTALINSKI NOVEMBER 16, 2016 7:55 AM  
TAGS: .NET, .NET FOUNDATION, GOOGLE, MICROSOFT, MICROSOFT CONNECT




**.NET** foundation

**GeekWire** NEWS ▾ JOBS ▾ EVENTS ▾ RESOURCES ▾ DEALS ▾ ABOUT ▾ Search


## Microsoft joins Linux Foundation in another step toward greater openness

BY DAN RICHMAN on November 16, 2016 at 7:30 am

9 Comments [Share](#) 623 [Tweet](#) [Share](#) 269 [Reddit](#) [Email](#)

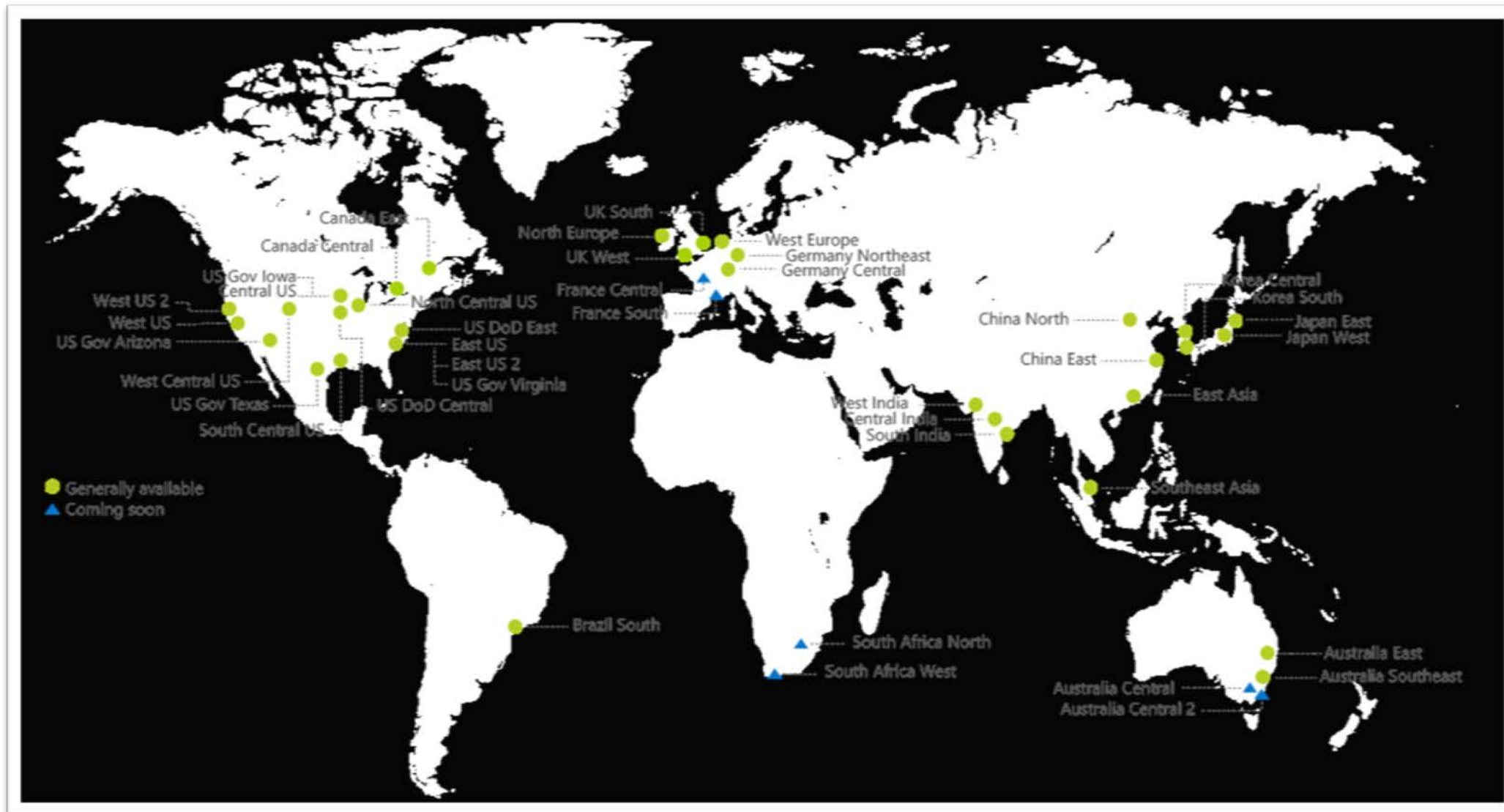


GeekWire Gala tickets on sale now



CPU - MEM - STORAGE - NETWORK  
**FREE VIRTUAL MONITORING**  
ESX/VMWARE HYPER-V XENSERVER - RED HAT

# Data centers at 165 locations



# Why does Microsoft build sub-marines?





# Azure

## Security & Management

- Security Center
- Portal
- Azure Active Directory
- Azure AD B2C
- Multi-Factor Authentication
- Automation
- Scheduler
- Key Vault
- Store/Marketplace
- VM Image Gallery & VM Depot

## Media & CDN

- Media Services
- Media Analytics
- Content Delivery Network

## Integration

- API Management
- BizTalk Services
- Logic Apps
- Service Bus

## Compute Services

- Container Service
- VM Scale Sets
- Batch
- RemoteApp
- Dev/Test Lab

## Platform Services

### Application Platform

- Web Apps
- Mobile Apps
- API Apps
- Cloud Services
- Service Fabric
- Notification Hubs
- Functions

### Developer Services

- Visual Studio
- Mobile Engagement
- VS Team Services
- Xamarin
- Application Insights
- HockeyApp

### Data

- SQL Database
- SQL Data Warehouse
- DocumentDB
- SQL Server Stretch Database
- Redis Cache
- Storage Tables
- Azure Search

### Intelligence

- Cognitive Services
- Bot Framework
- Cortana

### Analytics & IoT

- HDInsight
- Machine Learning
- Stream Analytics
- Data Catalog
- Data Lake Analytics Service
- Data Lake Store
- IoT Hub
- Event Hubs
- Data Factory
- Power BI Embedded

## Hybrid Cloud

- Azure AD Health Monitoring
- AD Privileged Identity Management
- Domain Services
- Backup
- Operational Analytics
- Import/Export
- Azure Site Recovery
- StorSimple

## Infrastructure Services

### Compute

- Virtual Machines
- Containers

### Storage

- Blob
- Queues
- Files
- Disks

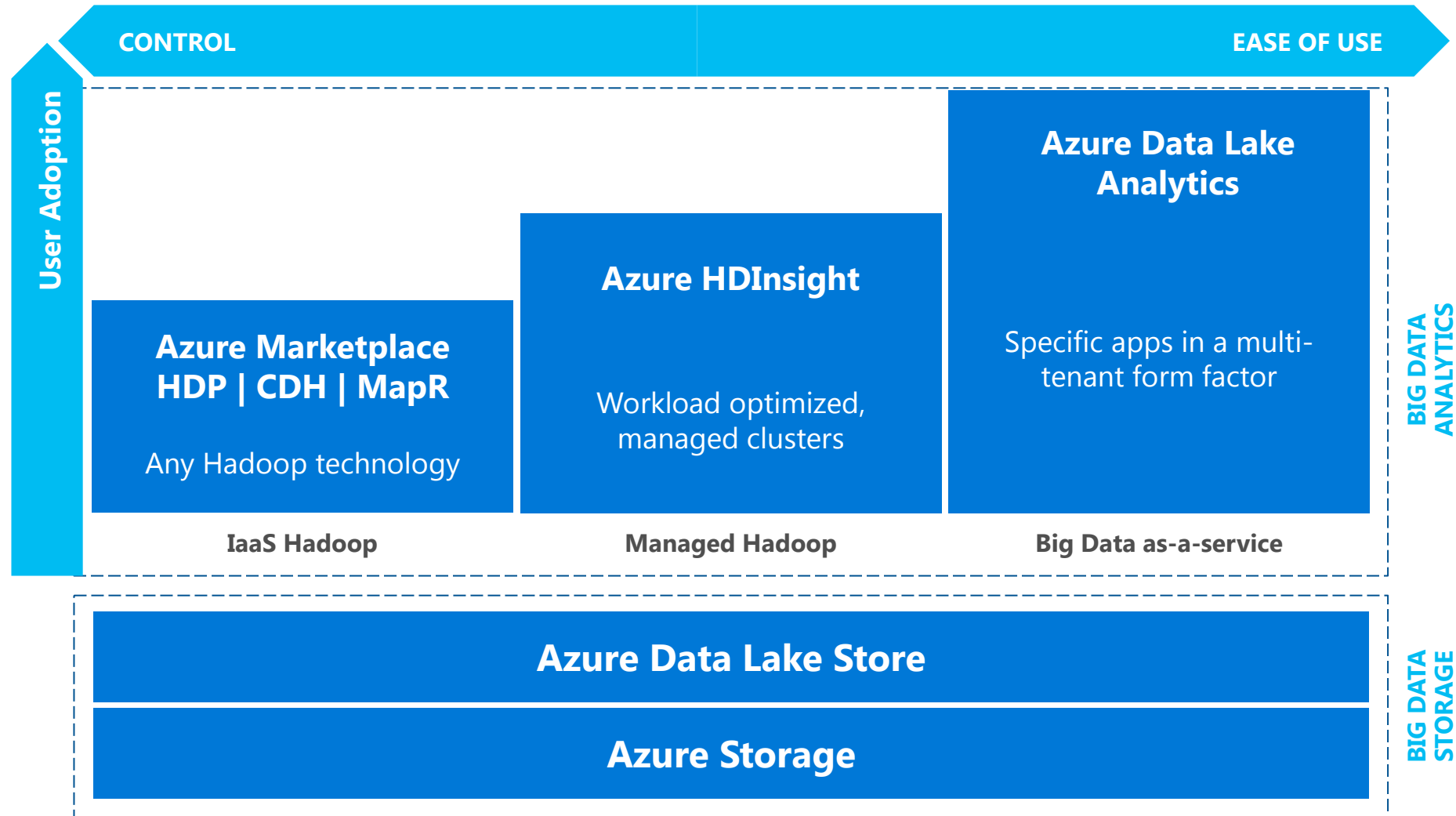
### Networking

- Virtual Network
- Load Balancer
- DNS
- Express Route
- Traffic Manager
- VPN Gateway
- App Gateway

## Datacenter Infrastructure (40 Regions)



# Positioning the various big data solutions



# Big Data Architecture

Big Data defined by

Volume

TB, PB, EB

Variety

Varying types

Varying schemas within same type

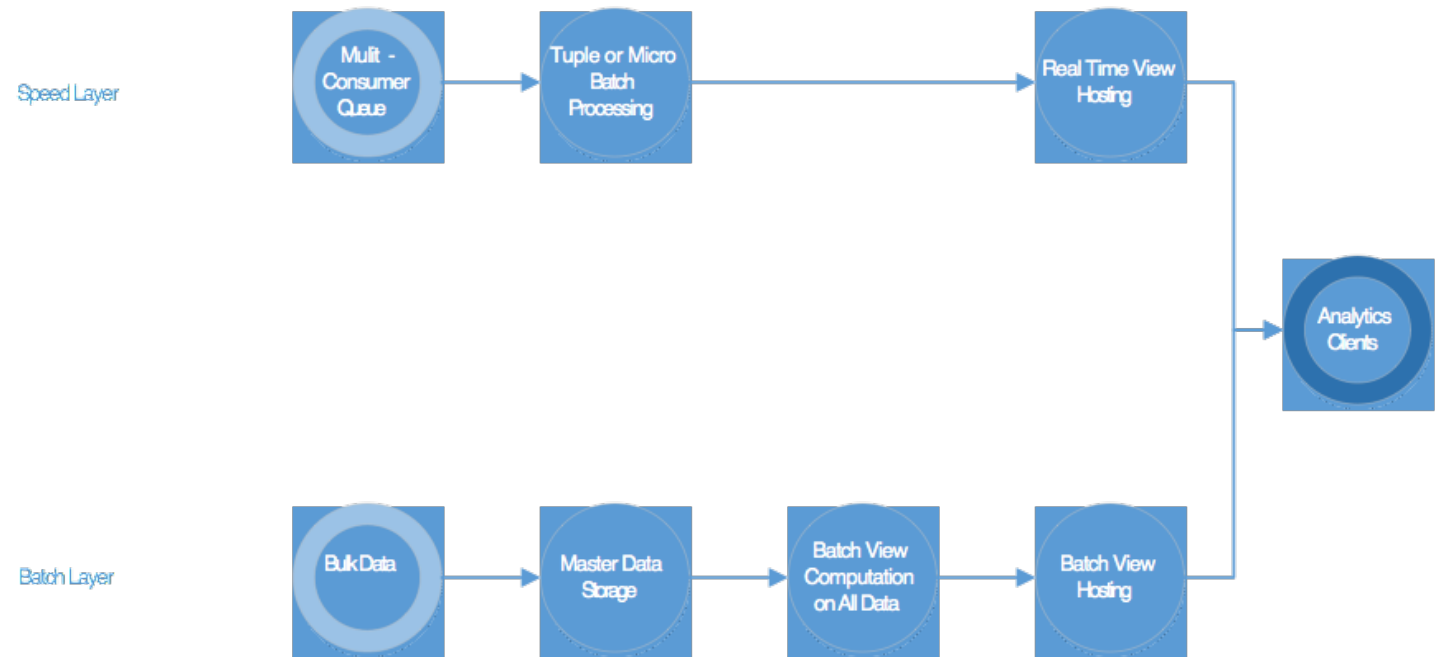
Unstructured

Binary

Velocity

Up to Real-time events ingestion

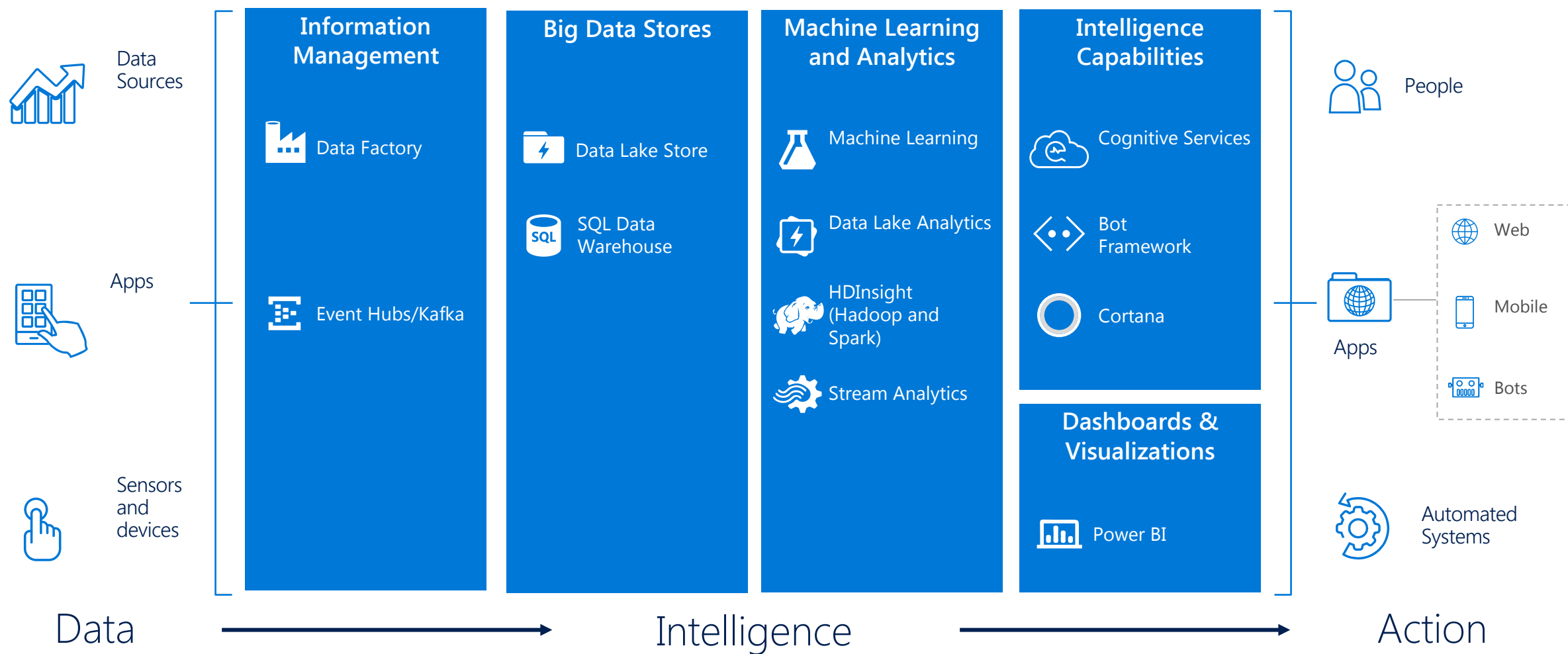
## LAMBDA ARCHITECTURE



*Lambda architecture brings together streaming and batch paths to deliver holistic analytics*



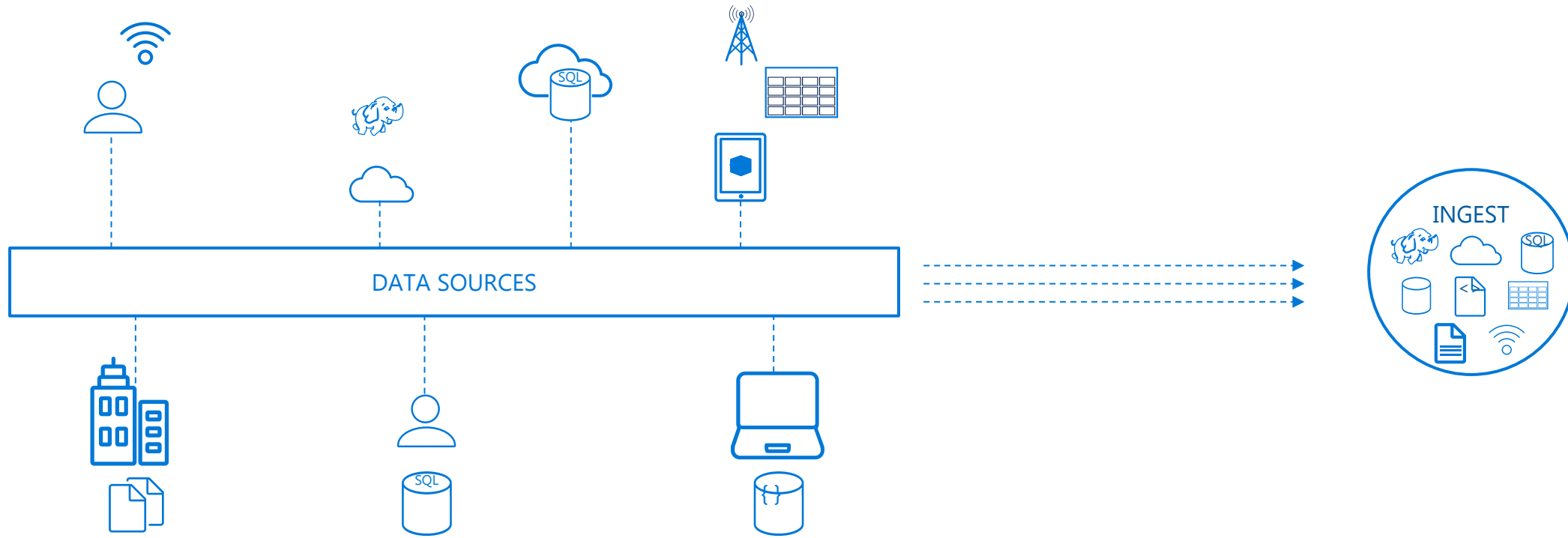
# Microsoft Data solution Overview (CIS)



# Compose and orchestrate data services at scale

**Information Management**

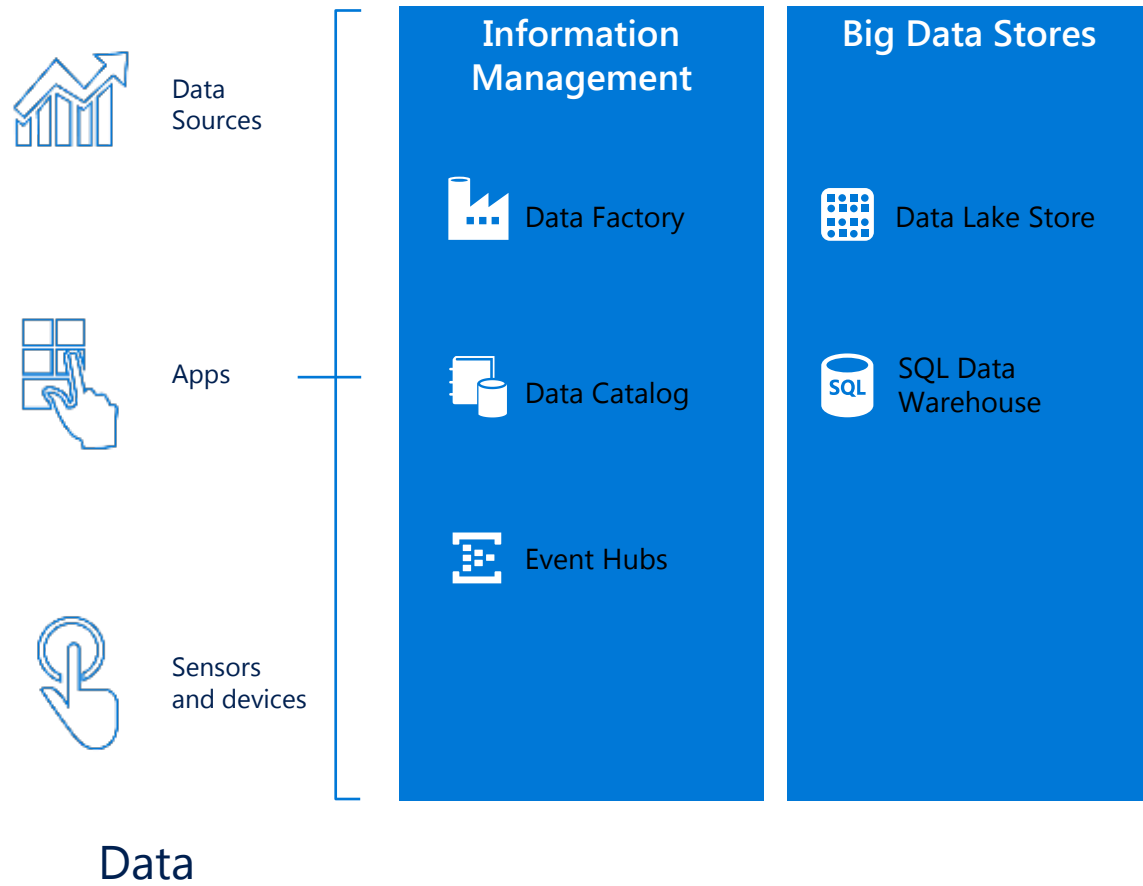
- Data Factory
- Data Catalog
- Event Hubs



- 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

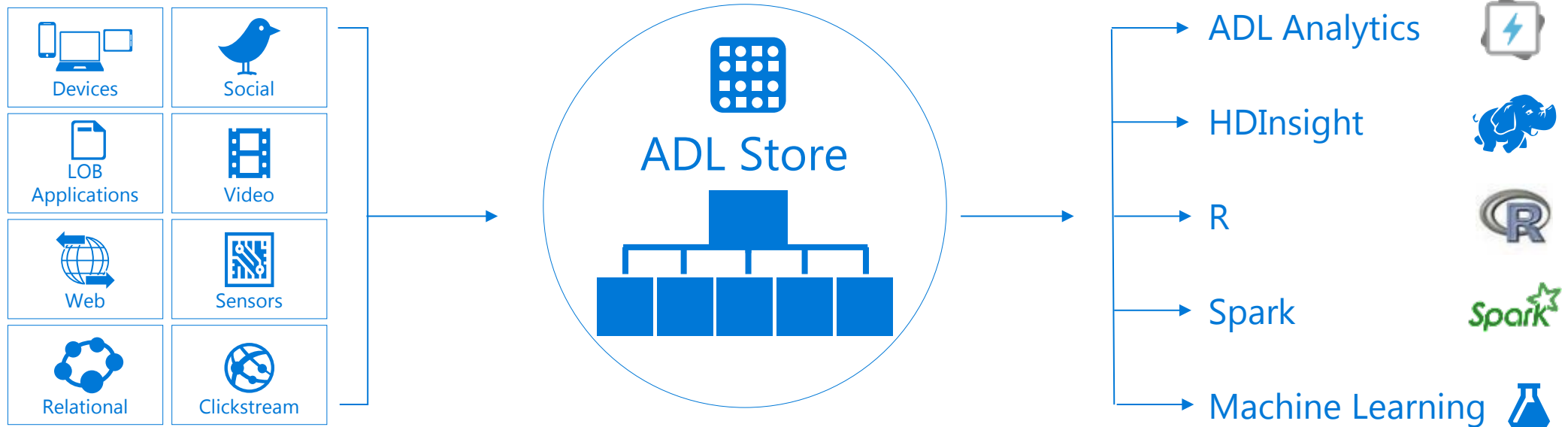
# Big Data Stores



# Unlimited HDFS storage for big data analytics workloads

**Big Data Stores**

-  Data Lake Store
-  SQL Data Warehouse



- 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

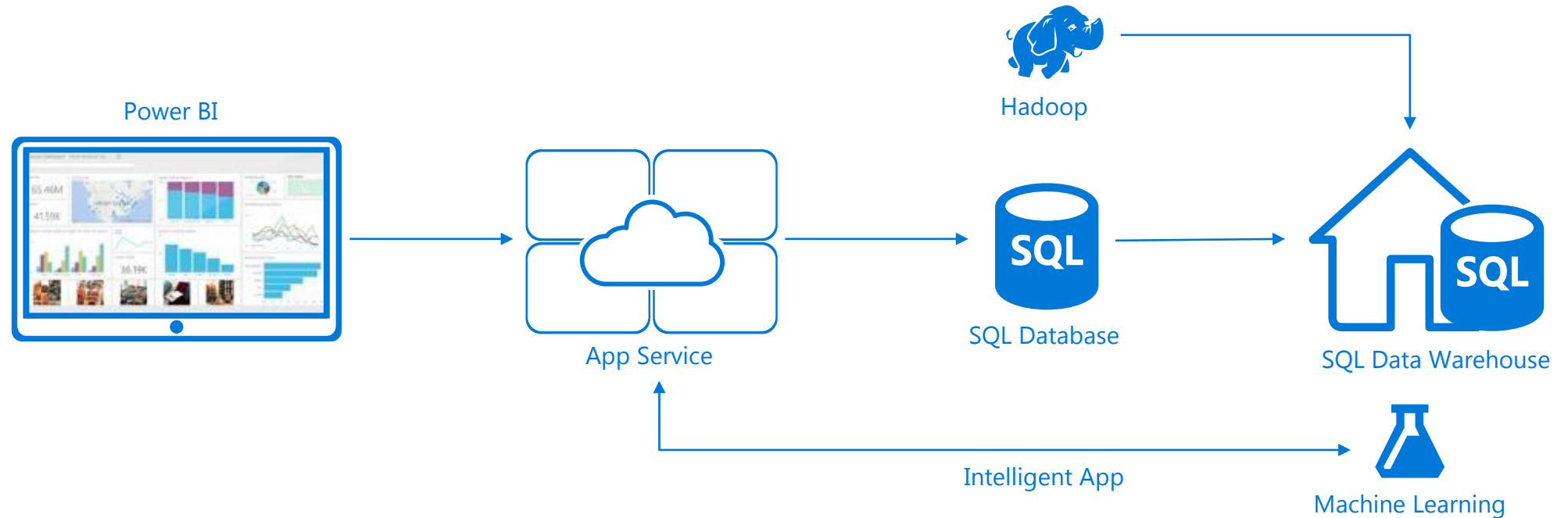
- Massive throughput to increase analytic performance
- High durability, availability, and reliability
- Azure Active Directory access control

# Elastic data warehouse as a service with enterprise-class features

## Big Data Stores

 Data Lake Store

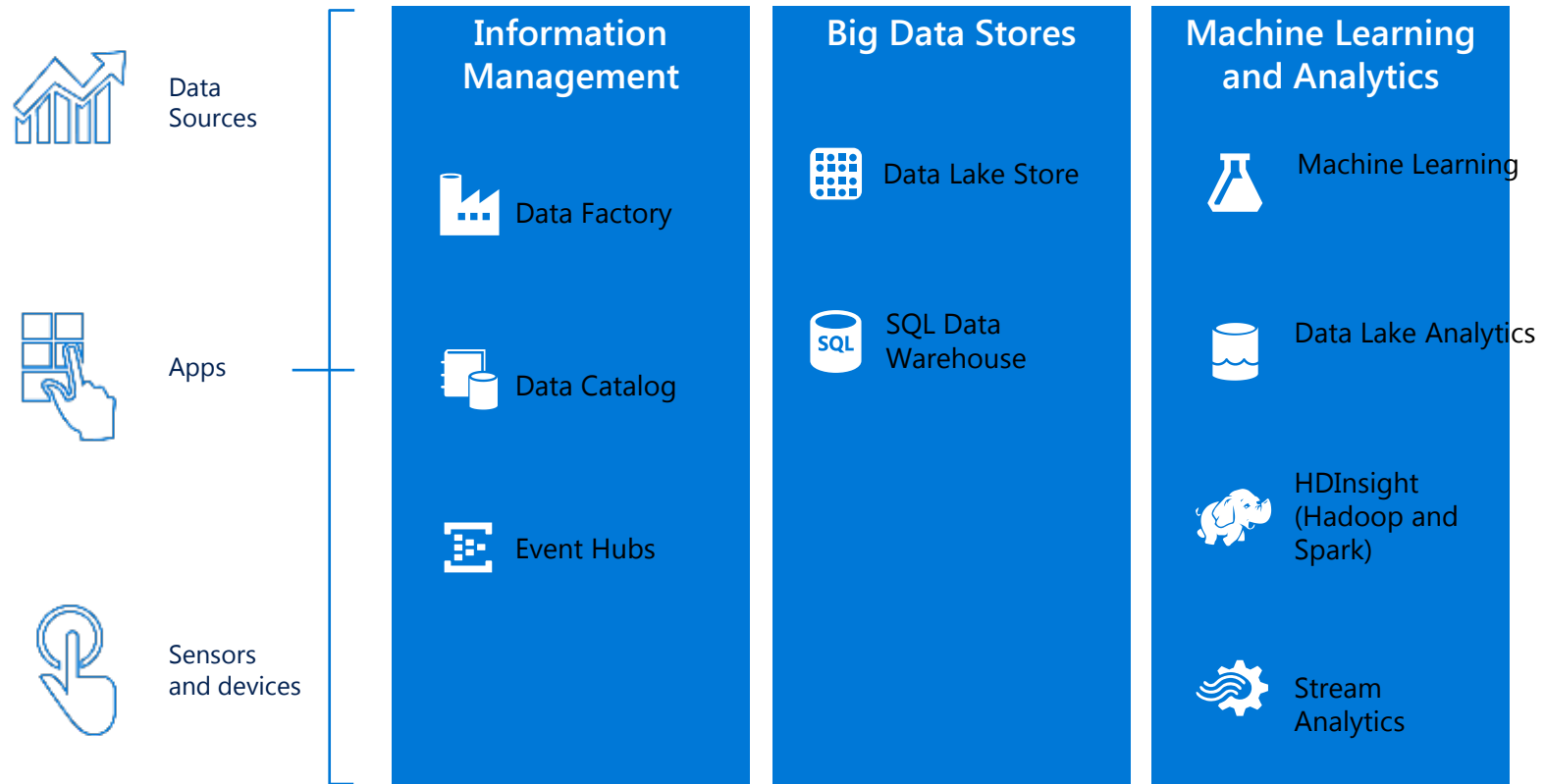
 SQL Data Warehouse



- 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

# Machine Learning and Analytics



a

Intelligence

# Azure Studio ML

## Easily build, deploy, and share predictive analytics solutions

### Machine Learning and Analytics



Machine Learning



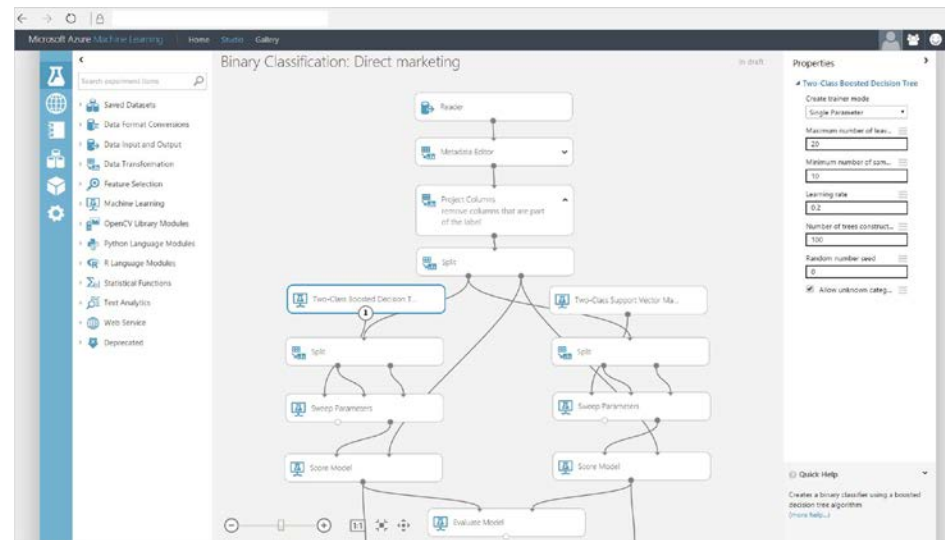
Data Lake Analytics



HDInsight  
(Hadoop and Spark)



Stream Analytics



- 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.

# Demo Machine Learning Studio



# Big data analytics made easy

**Machine Learning and Analytics**

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



Data Lake Analytics



- Analyze data of any kind and size
- Develop faster, debug and optimize smarter
- Interactively explore patterns in your data
- No learning curve—use U-SQL, Spark, Hive, HBase and Storm

- Managed and supported with an enterprise-grade SLA
- Dynamically scales to match your business priorities
- Enterprise-grade security with Azure Active Directory
- Built on YARN, designed for the cloud

# HDInsight

Microsoft + Hortonworks  
Promoting open Hadoop

Engineering alignment

Corporate alignment

Field alignment

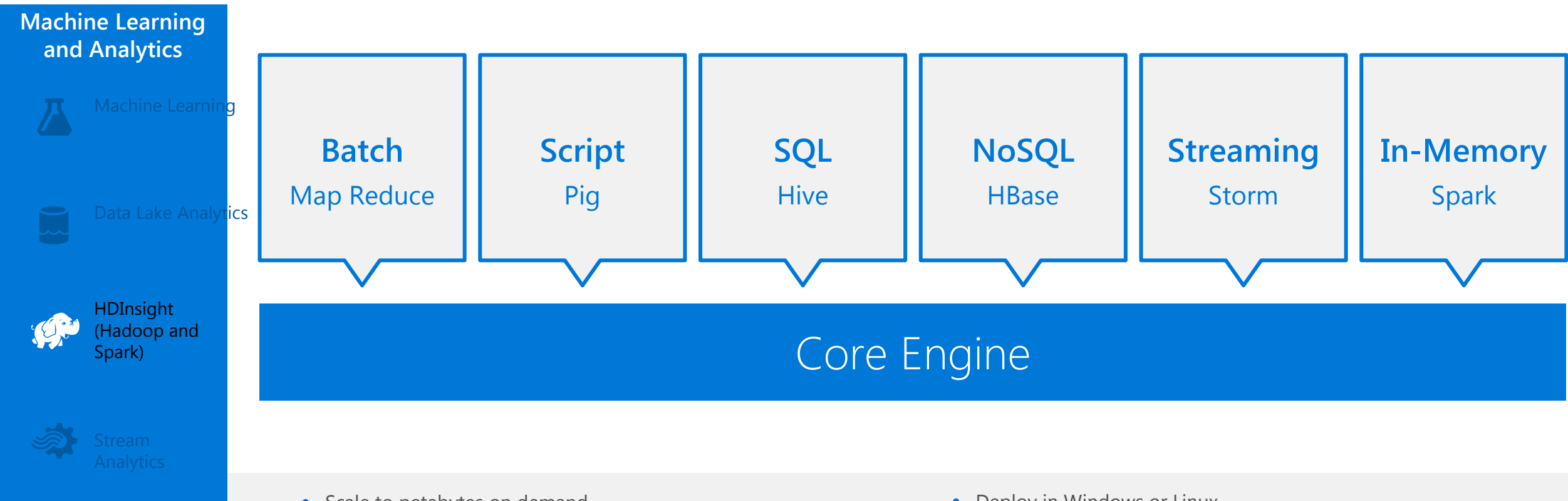


Microsoft



Hortonworks

# Comprehensive set of managed Apache big data projects

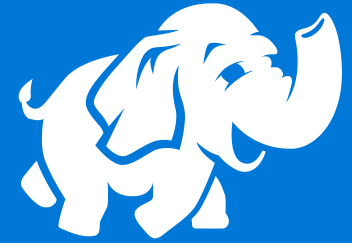


- Scale to petabytes on demand
- Process unstructured and semi-structured data
- Develop in Java, .NET, and more
- Skip buying and maintaining hardware

- 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

# Azure HDInsight

Hadoop as a Service in Cloud

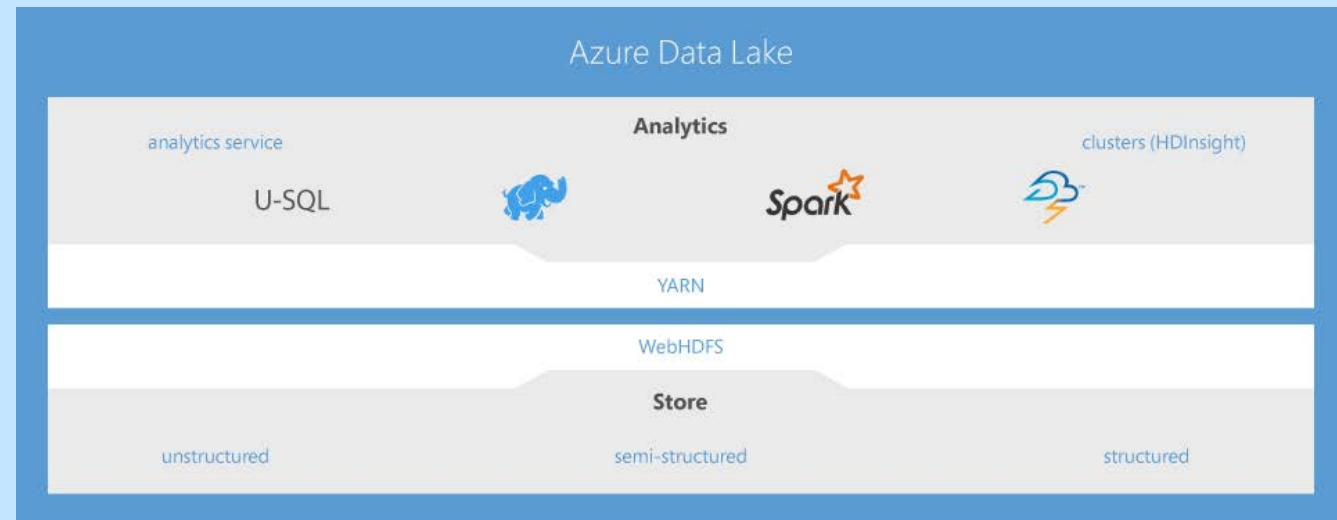


**Fully Managed** HIVE, Spark, HBase, Storm, Kafka

Works on **Azure Storage** or **Data Lake Store**

**100% Open Source** Apache Hadoop

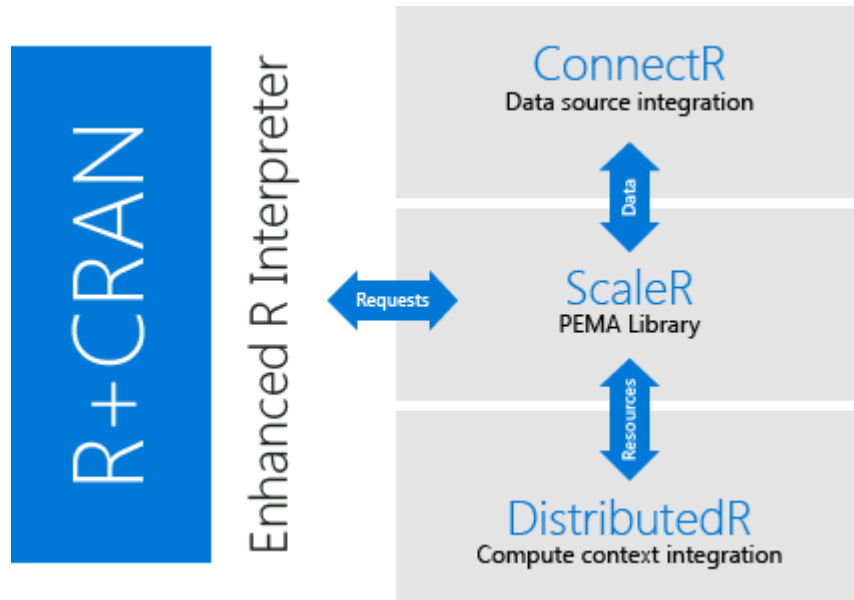
Clusters up and running in minutes





# Microsoft R server

## Coding in R on Big Data



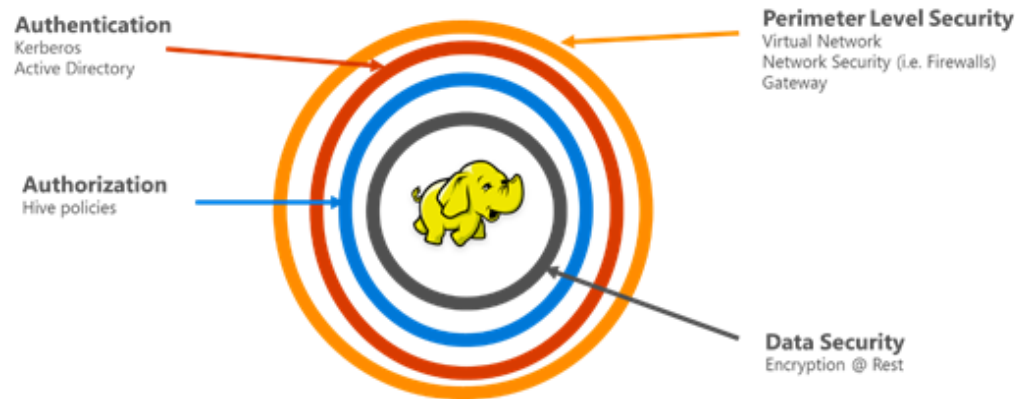
- Handles 1000 x more data
- 50 x faster with Spark
- Runs standard R-script

Demo

HDInsight

# HDInsight security

Enterprise grade capabilities like active directory based authentication, multi-user support, and role based access control and domain joining\*



Enterprise Security contains four big pillars:

- Perimeter Security
- Authentication
- Authorization (apache Ranger policies)
- Encryption

*\*To date available for Hadoop/Hive, Spark cluster*



# Data Science Virtual Machine

Linux or Window

## **Pre-installed with for instance:**

- R-server
- R-studio
- Julia
- CNTK
- Tensorflow
- Caffe & Caffe2
- Relational db
- Local Spark
- Local Hadoop
- ...

<https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-data-science-virtual-machine-overview>



THANKS