Big Data and Power BI

Darwin Schweitzer

Contact Info

Darwin Schweitzer
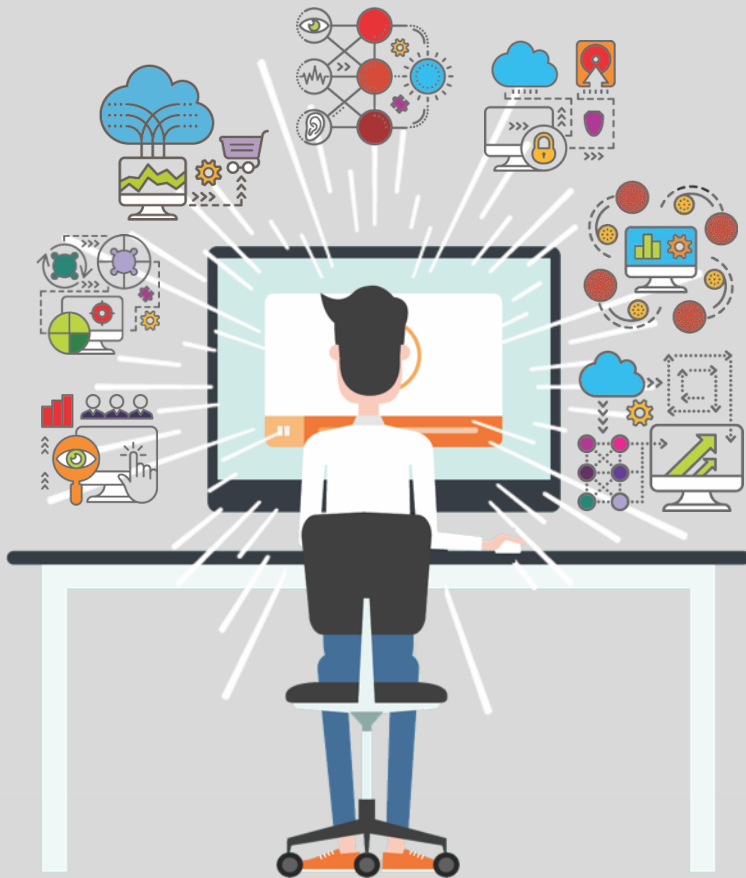
Email: darsch@microsoft.com

Github: datasnowman

Twitter: @DataSnowman

LinkedIn: Cortana Intelligence Solutions

Big data and analytics has huge potential

90% of business leaders expect big data to dramatically change how they do business, *having a landmark affect similar to the development of the Internet.* Source: Accenture, 2016

5X improvement in decision making by using advanced analytics

Source: Bain & Co., 2016

6-8% productivity gains that result from investment in data analytics. *" This is a higher rate of return than other recent technologies have yielded, surpassing even the computer investment cycle in the 1980s"*

Source: McKinsey & Co., 2016

#devdaysasia2017
#azure
#azureiot
#Microsoft

#cloudcomputing
#bigdata
@MSPowerBI
@MSAdvAnalytics
#iot

Microsoft

REDMOND
INTEROPERABILITY PLUGFEST
JUNE 19-23, 2017

https://gallery.cortanaintelligence.com/Tutorial/Time-Series-Insights-using-Twitter

Time Series Insights using Twitter
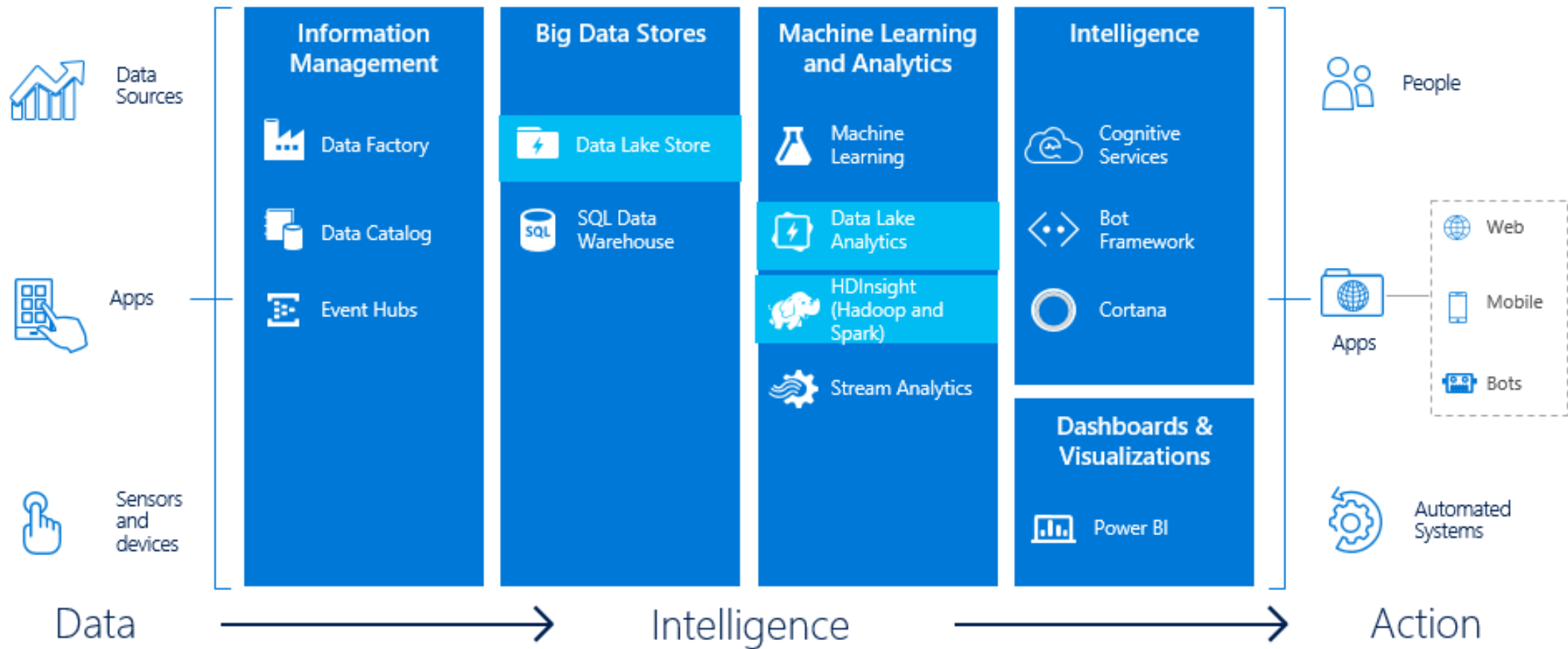
Demo

https://gallery.cortanaintelligence.com/Tutorial/Time-Series-Insights-using-Twitter
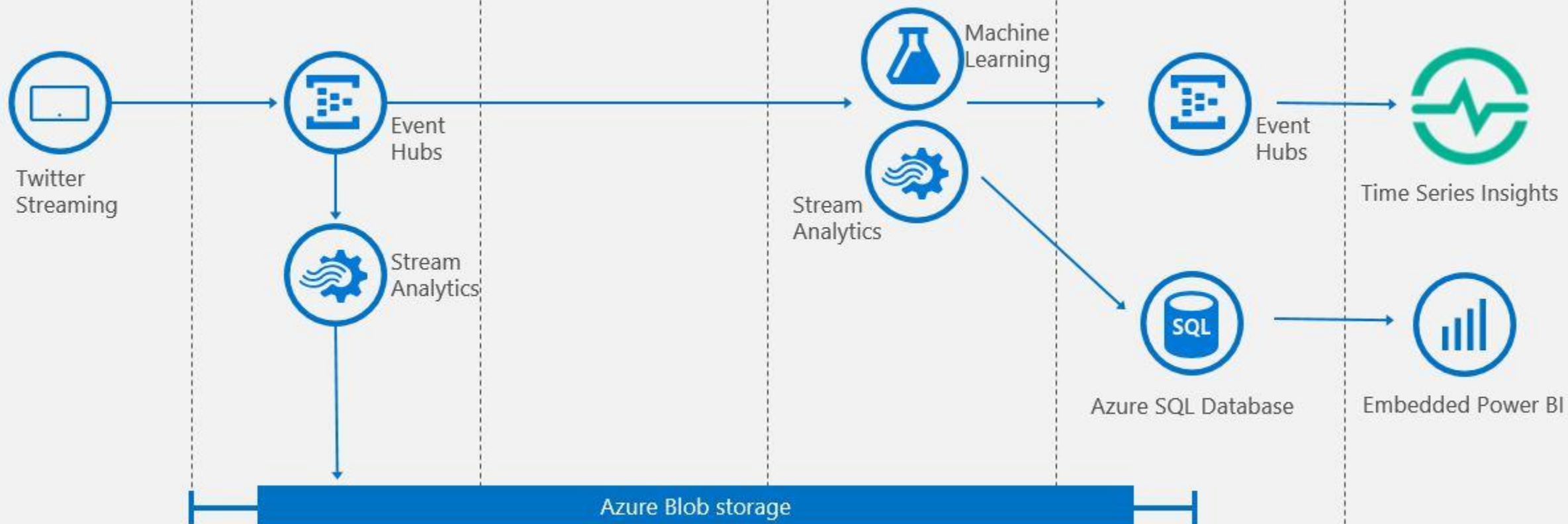
Big Data is driving transformative changes

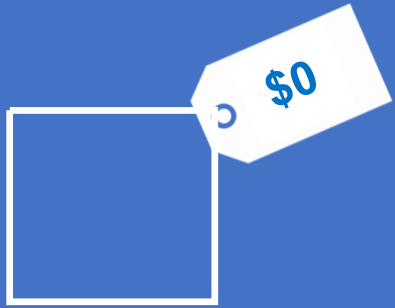|	Traditional	Big Data
Data characteristics	Relational data with highly modeled schema	All data with schema agility
Costs	Specialized HW	Commodity HW
Culture	Operational reporting Focus on rear-view analysis	Experimentation leading to intelligent action With machine learning, graph, a/b testing

Microsoft

Big Data as a cornerstone of Cortana Intelligence

Data Sources

Apps

Sensors and devices

Information Management

- Data Factory
- Data Catalog
- Event Hubs

Big Data Stores

- Data Lake Store
- SQL Data Warehouse

Machine Learning and Analytics

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

Intelligence

- Cognitive Services
- Bot Framework
- Cortana

Dashboards & Visualizations

- Power BI

People

Web

Mobile

Apps

Bots

Automated Systems

Data

Intelligence

Action

Microsoft

Microsoft Big Data

HDInsight

Data Lake Store

Data Lake Analytics

YARN

HDFS

YARN

T-SQL

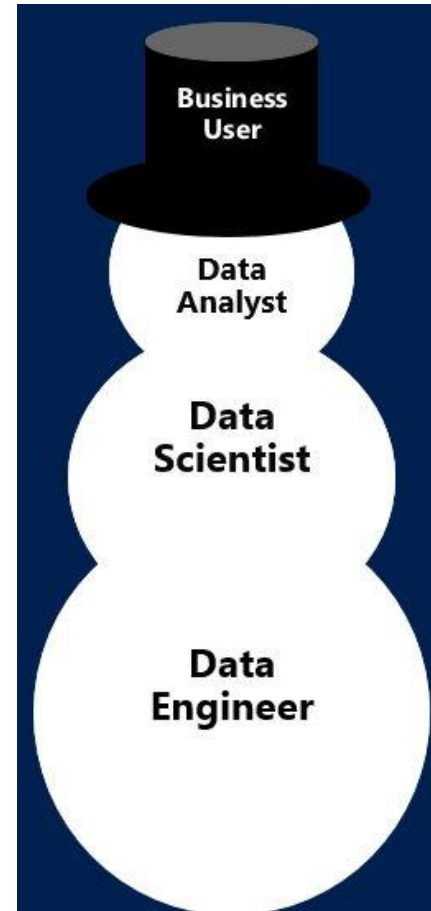# Why Hadoop in the Cloud?

No HW costs

Unlimited scale

Pay what you need

Deployed in minutes

Microsoft

# Something for all roles

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"

Microsoft

Easy for business analysts with interactive reports over big data
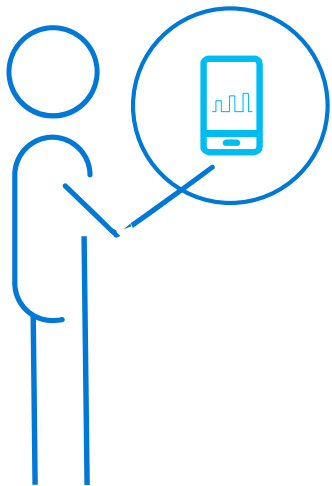
- Interactive BI with big data

- Spark 2.0 integration

- Interactive Hive with LLAP-keeps data compressed running in-memory 25x faster

- ODBC driver to use Power BI or third party tools (Tableau, SAP, Qlik, etc.)

Microsoft

# Easy for data scientists with familiar R language
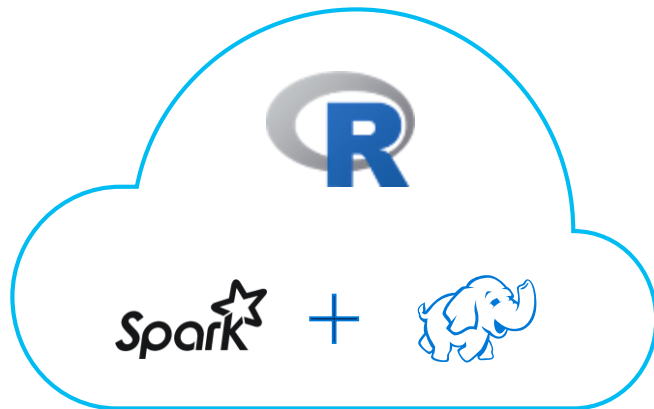
*Applies to HDInsight only

R Server for HDInsight

- Largest portable R parallel analytics library

- Terabyte-scale machine learning—1,000x larger than in open source R

- Up to 100x faster performance using Spark and optimized vector/math libraries

- Enterprise-grade security and support

Microsoft

Easy notebook experience for data engineers
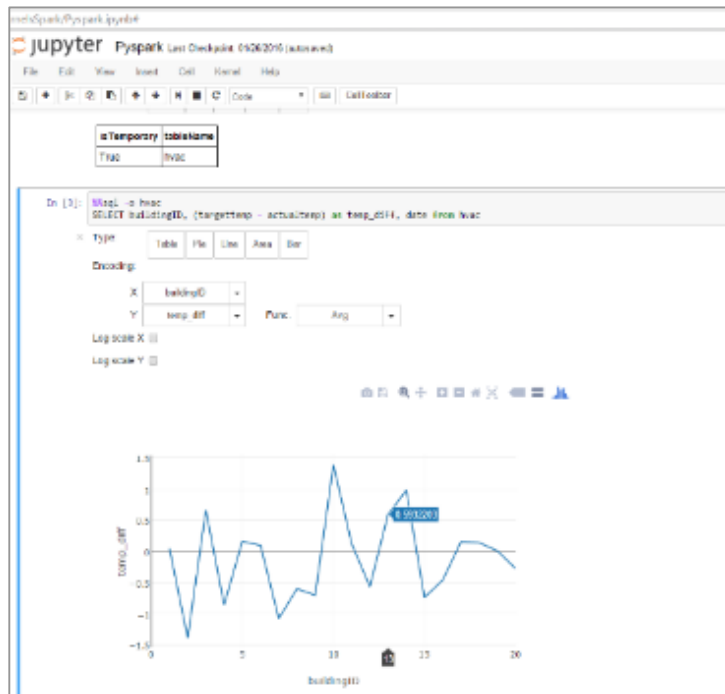
- Most popular notebooks, Jupyter and Zeppelin out-of-the-box

- Combine code, statistical equations and visualizations

- Worked w/ Jupyter community to enhance kernel to allow Spark execution through REST endpoint
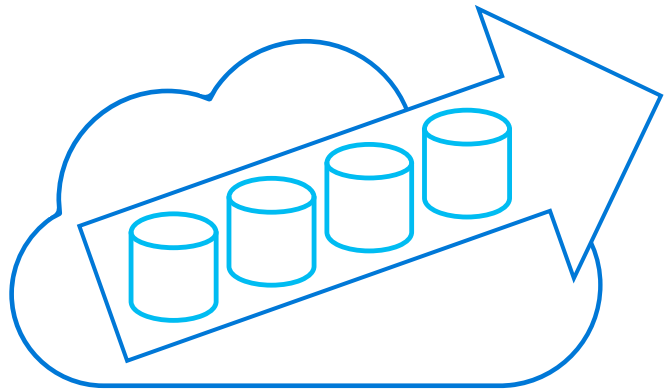
Microsoft

Debug and Optimize your Big Data programs with ease

- Deep integration with IDEs for developer productivity: Visual Studio, Eclipse, & IntelliJ

- Integrated with Hive, Pig, Storm, and Spark

- Visually see execution of Hive jobs ran by the Tez execution engine

- Full Intellisense

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
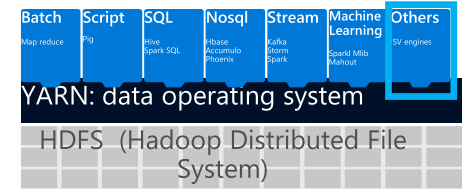
Recognized by top analysts

FORRESTER®

Forrester Wave for Big Data Hadoop Cloud

- Named industry leader by Forrester with the most comprehensive, scalable, and integrated platforms*

- Recognized for its cloud-first strategy that is paying off*

*The Forrester WaveTM: Big Data Hadoop Cloud Solutions, Q2 2016.

Microsoft

ISV Integration

Batch
Map reduce

Script
Pig

SQL
Hive
Spark SQL

Nosql
Hbase
Accumulo
Phoenix

Stream
Kafka
Storm
Spark

Machine Learning
Sparkl Mlib
Mahout

Others
SV engines

YARN: data operating system
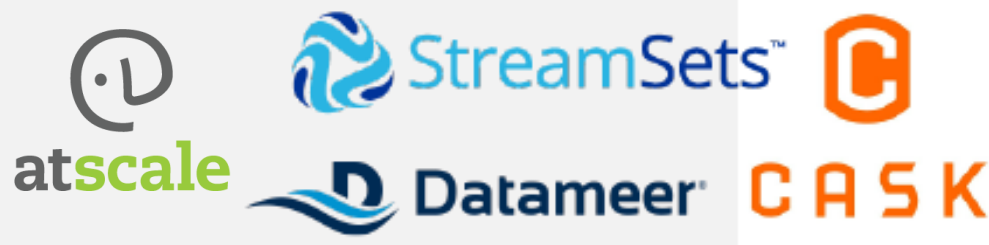
HDFS (Hadoop Distributed File System)

Integration with leading productivity applications

Spin up Hadoop and Spark clusters pre-integrated and pre-tuned with ISV applications out-of-the-box

Runs on the HDInsight clusters; does not require separate VMs

Fast and easy way to spin up applications

Azure
Data Lake Store

A No limits Data Lake that powers Big Data Analytics

Petabyte size files and **Trillions of objects**

Scalable throughput for **massively parallel analytics**

HDFS for the cloud

Always encrypted, role-based security & auditing

Enterprise-grade support

Azure
Data Lake Analytics

A No limits Analytics Job Service to power intelligent action
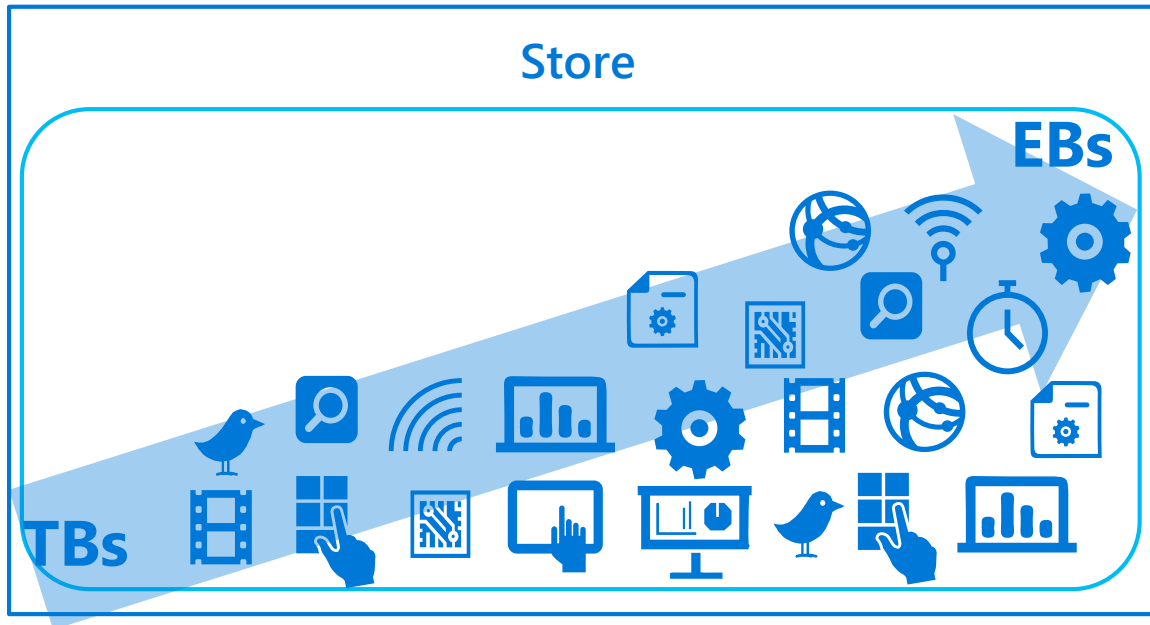
Start in seconds, scale instantly, pay per job

Develop massively parallel programs **with simplicity**

Debug and **optimize** your big data programs **with ease**
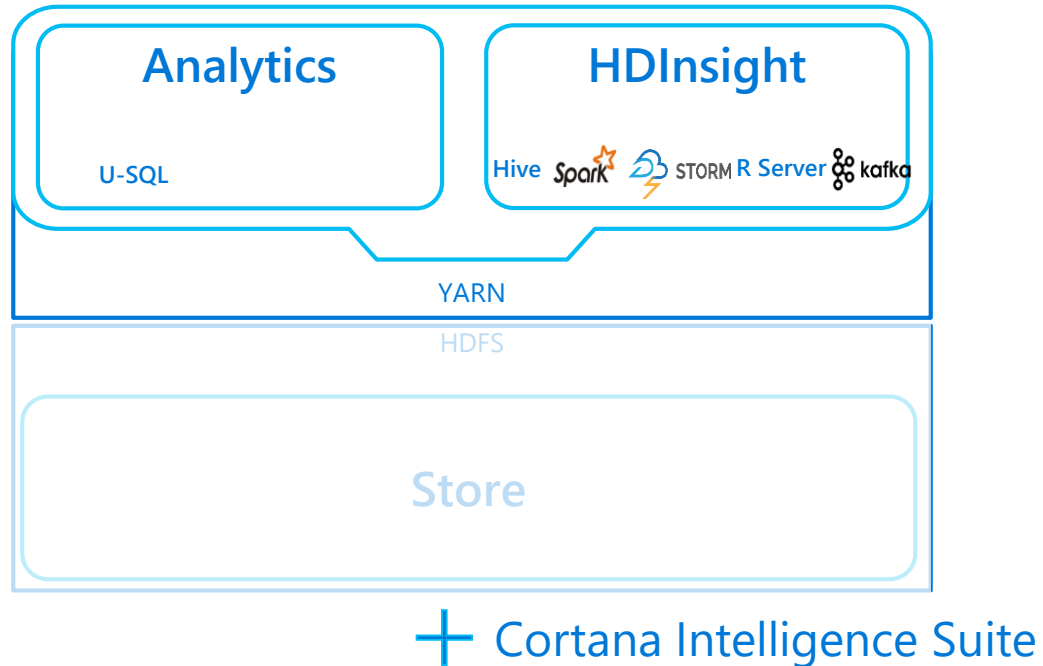
Virtualize your analytics

Enterprise-grade security, auditing and **support**

# Petabyte size files and Trillions of objects

Store

EBs

TBs

- 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

Any type of analytics

Analytics

U-SQL

HDInsight

Hive Spark STORM R Server kafka

YARN

HDFS

Store

+ Cortana Intelligence Suite

- Batch, interactive, streaming, machine learning

- Allows for exploratory analytics over data

- Analyze with Hadoop and Microsoft solutions

Microsoft

Start in seconds, Scale instantly, Pay per job
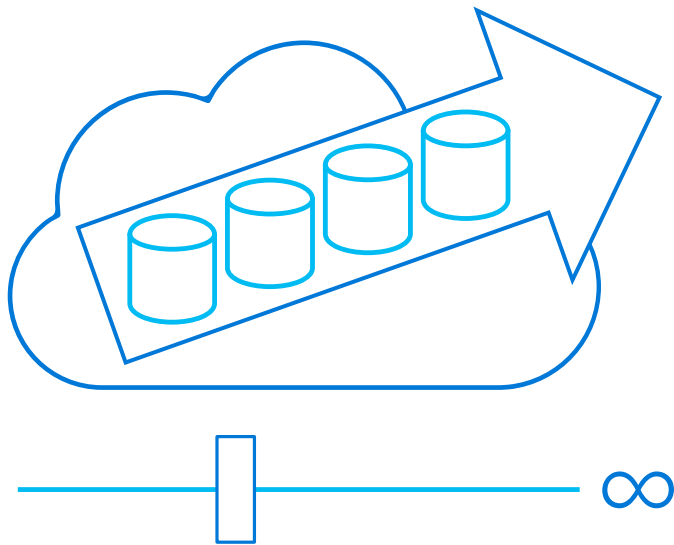
- Process big data jobs in 30 seconds
- No infrastructure to worry about (no servers, no VMs, no clusters)
- Instantly scale analytic units up or down (processing power)
- Architected for cloud scale and performance
- Frees you up to focus only on your business logic

∞

Develop massively parallel programs with simplicity

```
@t = EXTRACT date string
           , time string
           , author string
           , tweet string
    FROM "/Input/MyTwitterHistory.csv"
    USING Extractors.Csv();

@res = SELECT author AS author
           , COUNT(*) AS tweetcount
       FROM @t
       GROUP BY author;

OUTPUT @res TO "/Output/MyTwitterAnalysis.csv"
ORDER BY tweetcount DESC
USING Outputters.Csv();
```

- U-SQL: a simple and powerful language that's familiar and easily extensible

- Unifies the declarative nature of SQL with expressive power of C#

- Leverage existing libraries in .NET languages, R and Python

- Massively parallelize code on diverse workloads (ETL, ML, image tagging, facial detection)

Microsoft

Azure SQL Data Warehouse

A relational **data warehouse-as-a-service**, fully managed by Microsoft.
Industries first **elastic** cloud data warehouse with proven SQL Server capabilities.
Support your **smallest to your largest** data storage needs.
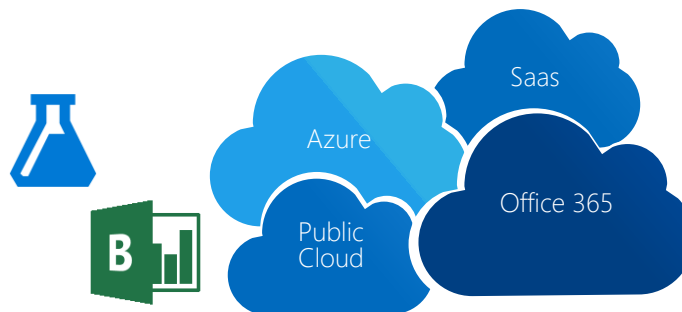
Elastic scale & performance

10101
0101
00100
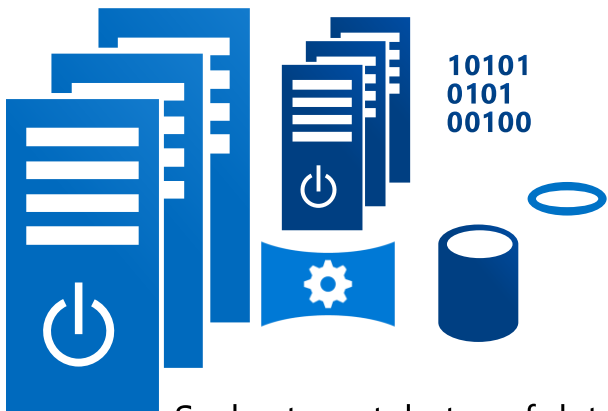
Scales to petabytes of data

Massively Parallel Processing

Instant-on compute scales in seconds

Query Relational / Non-Relational

Powered by the Cloud

Get started in minutes

Integrated with Azure ML, PowerBI & ADF

Saas

Azure

Public Cloud

Office 365

Market Leading Price & Performance

Azure

Simple billing compute & storage

Pay for what you need, when you need it with dynamic pause

A fully managed Platform-as-a-Service

- Azure cloud data warehouse service
- Elastic scale
- Separate storage and compute
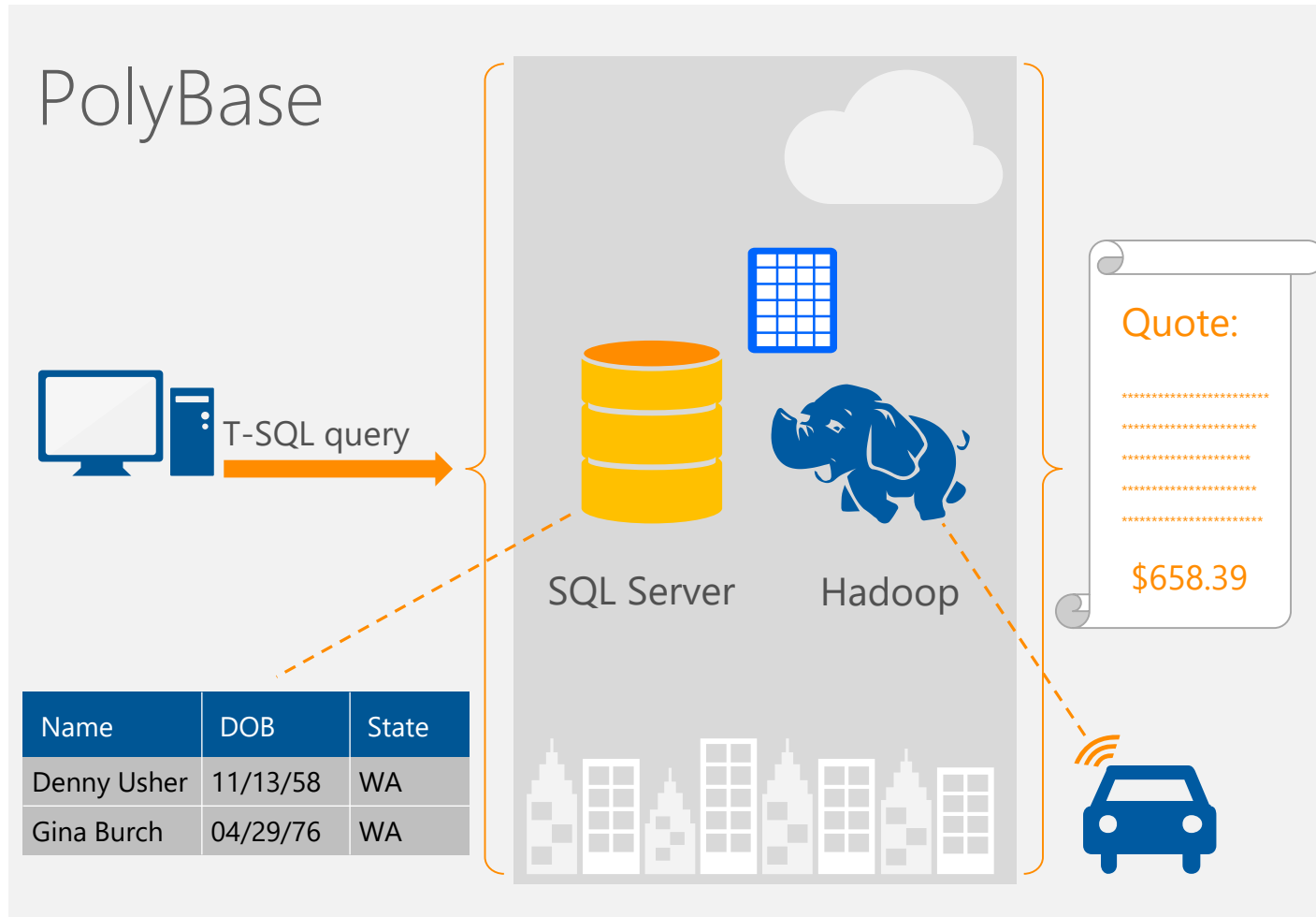- Use existing tools and skills
- Deploy and use in minutes!

Remove the complexity of big data

T-SQL over Hadoop

PolyBase

T-SQL query

SQL Server   Hadoop

Name	DOB	State
Denny Usher	11/13/58	WA
Gina Burch	04/29/76	WA

Quote:

****
****
****
****

$658.39

Leverage existing T-SQL skills and BI tools to gain insights from different data stores

Key Benefits

- Reduce development costs performing query in structured and non-structured data using T-SQL

- Combine SQL with Hadoop enabling database for modern applications

- Simplified architecture with built-in hybrid integration

- Leverage existing T-SQL skills and BI tools to gain insights from different data stores

Data Warehousing and Modern BI on Azure

Demo

https://gallery.cortanaintelligence.com/Tutorial/Data-Warehousing-and-Modern-BI-on-Azure

Microsoft

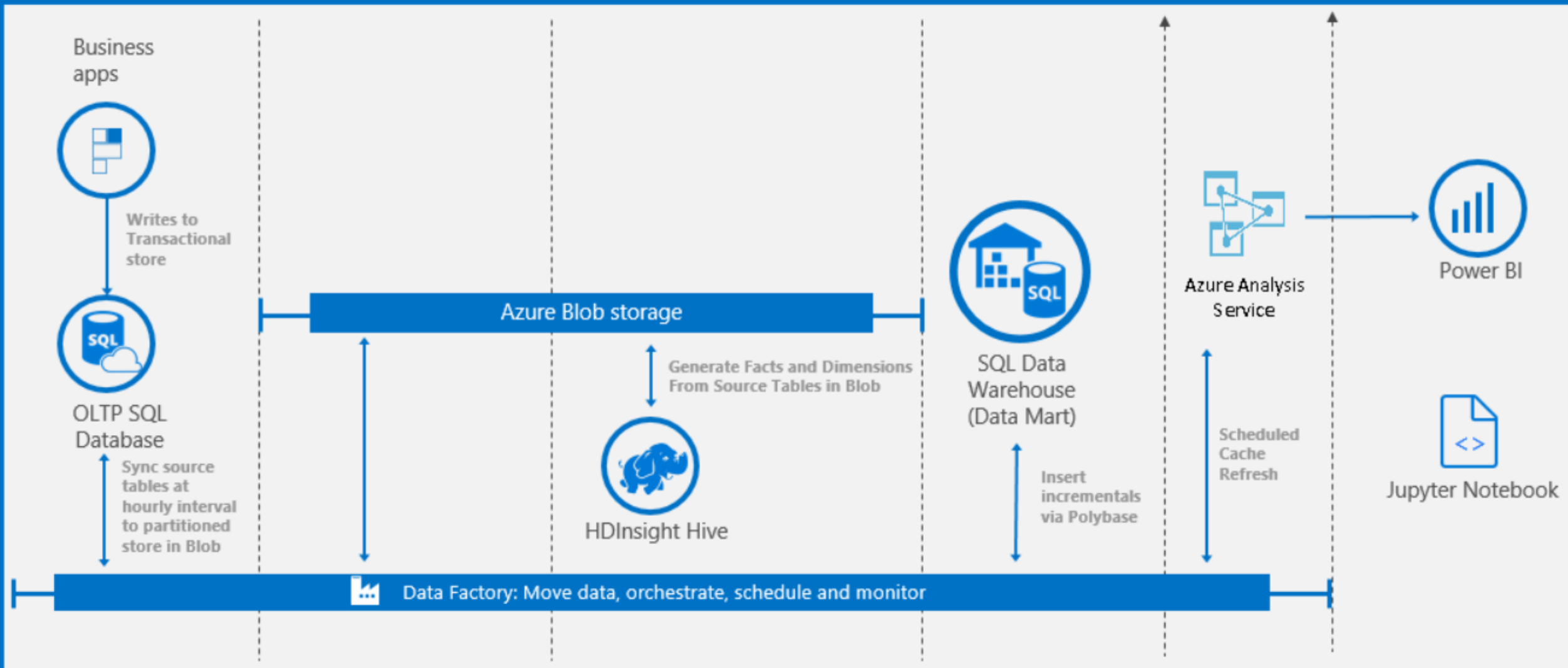https://gallery.cortanaintelligence.com/Tutorial/Data-Warehousing-and-Modern-BI-on-Azure

| Event Producers | Collection | Stream ingestion | Transformation & analytics | Stream ingestion | Presentation and Action |

Applications

Devices

Sensors

Web and Social

Cloud gateways (web APIs)

Field gateways

Kafka for HDInsight

Azure ML

HDInsight

Spark
Spark for HDInsight

STORM
Storm for HDInsight

Storage Adapters
Stream processing

Live Dashboards

APACHE HBASE
HBase for HDInsight

DocumentDB
MonoDB SQL
Solr Azure
Data Lake Store
Search

Web/thick client dashboards

Search and query

Data Analytics

Event hub
Devices to take action

Microsoft

https://azure.microsoft.com/en-us/services/hdinsight/apache-kafka/

Power BI: experience your data

Any data, any way, anywhere

Cloud data

On-premises data

Out-of-the-box SaaS content packs

Real time dashboards & interactive reports

Natural Language query

Custom visualizations

Native Office 365 integration

Power BI
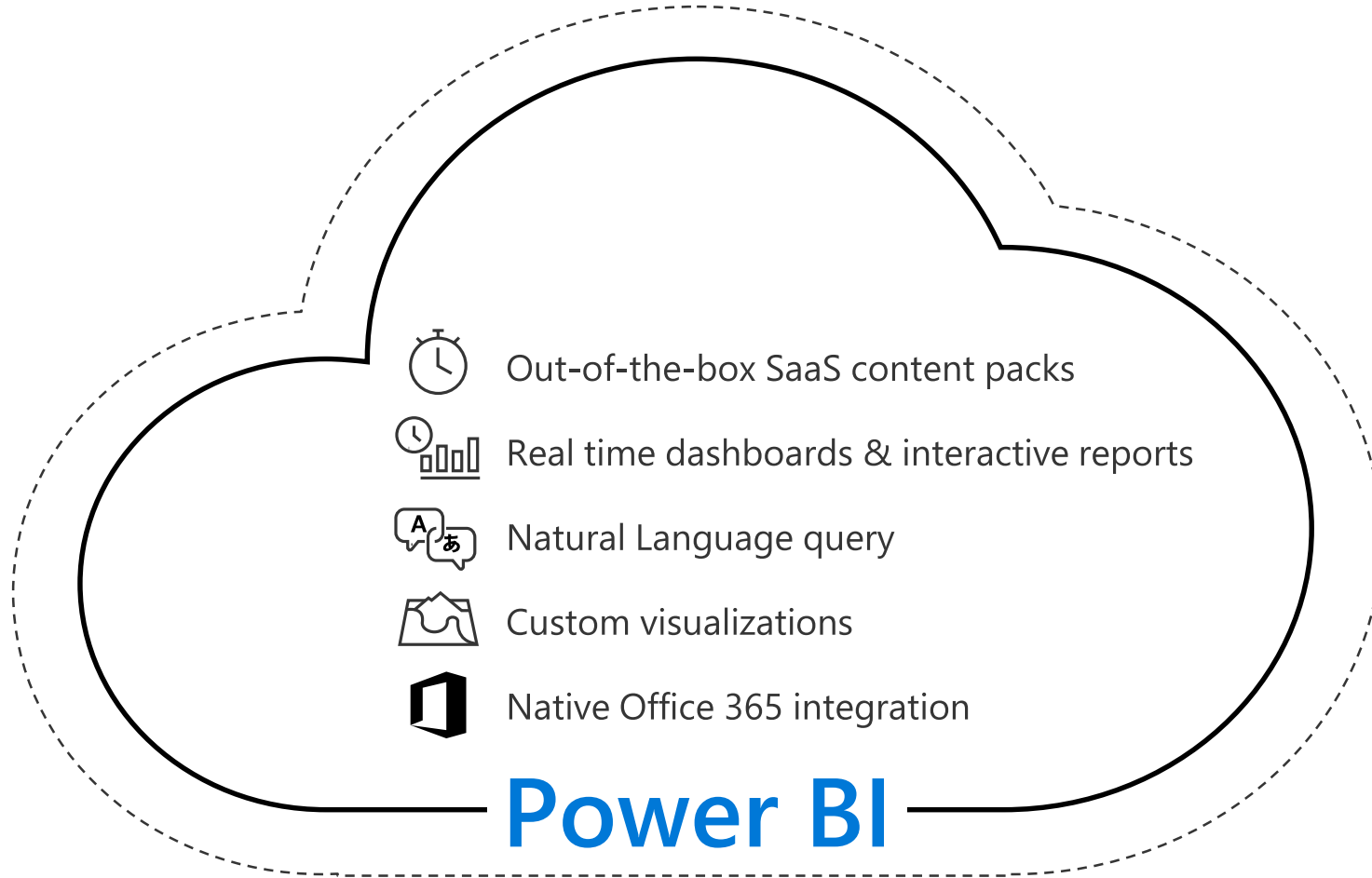
Web

Mobile

Excel

Embedded

Cortana

"......"

Worldwide reach

ISO 27001 — Information Security Management System Certified

cloud security alliance®
CSA

HIPAA

GOV.UK

PCI Security Standards Council ™

FedRAMP

EU Model Clauses — European Commission

irap
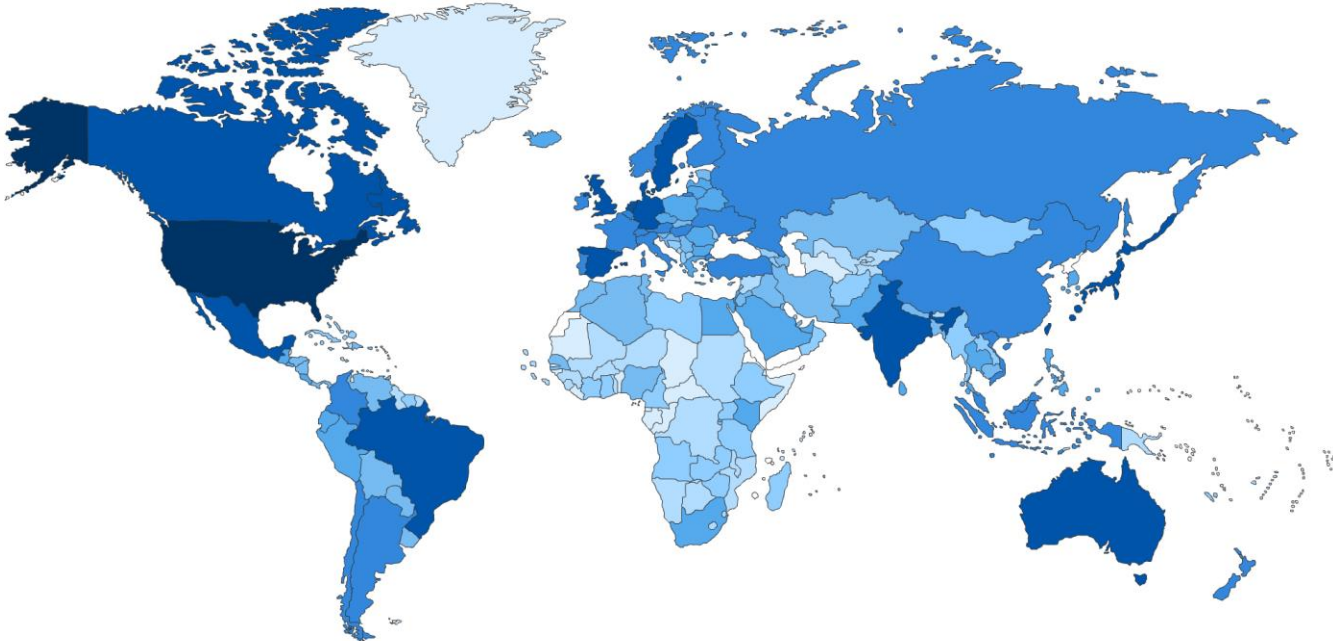
18,000
Cities globally access the service daily

235
Countries & territories

43
Languages

Global coverage
19 datacenters worldwide

National Clouds
US Government
China
Germany

Federal Office for Information Security

ESI
China GB 18030
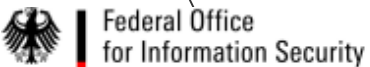
Microsoft

Shaded areas show countries with Power BI usage. The darker the color the greater the number of users

...ERABILITY PLUGFEST
JUNE 19-23, 2017

MAY 2017
Editor's choice

Gartner®

February 2017
A Leader in the Magic Quadrant for BI & Analytics*

Microsoft

REDMOND
INTEROPERABILITY PLUGFEST
JUNE 19-23, 2017

Power BI Premium

Power BI Premium

Sharing without boundaries

Unrestricted scale and performance

Flexibility to deploy in the cloud or on-premises

Dedicated resources in the cloud

Cortana Intelligence Solutions
Improve business processes and increase competitiveness

Intelligent

Interactive dashboards that help highlight actionable insights from data

Proven solution architectures that enable transformation of data into insights

Flexible

Deploy and operationalize a complex big data architecture with ease.

Scale and customize the solution to your needs
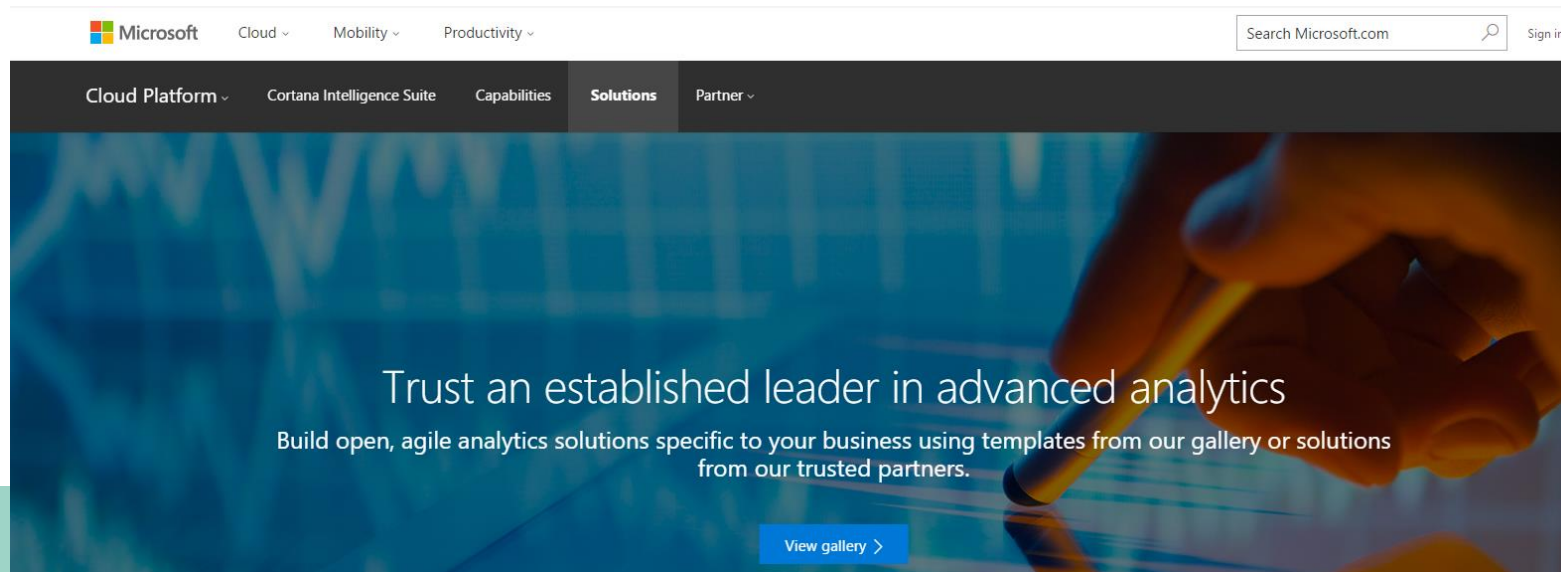
Trusted

Qualified partners to help with proof-of-concepts and operationalization

Rich technical guides to build an enterprise-grade solution

Cortana Intelligence Solutions

- Solution can be discovered from Microsoft.com
  https://www.microsoft.com/en-us/cloud-platform/cortana-intelligence-suite-solutions
- Solutions and Tutorials from the Cortana Intelligence Gallery
  https://gallery.cortanaintelligence.com/solutions
- From aka.ms short urls https://aka.ms/cisolutions and https://aka.ms/citutorials
- AppSource
  https://appsource.microsoft.com/en-us/marketplace/apps?product=cortana-intelligence&page=1

Solutions Implementation Partners

Implementation Partners we are training on these solutions

Thank You!

Questions?

Microsoft