

# Building Blocks of Cortana Intelligence Suite

Melissa Coates  
Analytics Architect  
SentryOne



Blog: [sqlchick.com](http://sqlchick.com)  
Twitter: @sqlchick

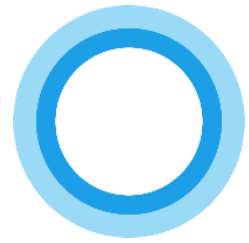


# Building Blocks of Cortana Intelligence Suite

---

## Agenda

1. Azure Overview
2. Tour of Cortana Intelligence Suite:  
Purpose, Use Cases, Building Blocks
3. Resources for Samples & Tutorials



Cortana Intelligence Suite

---

Azure Overview

# Microsoft Azure Platform

Microsoft Azure is a cloud computing platform and infrastructure for building, deploying, and managing applications and services through a global network of Microsoft-managed datacenters.



Azure provides services and supports many different programming languages, tools and frameworks, including both Microsoft-specific and third-party software and systems.

The screenshot displays the Microsoft Azure portal dashboard, organized into several horizontal sections:

- Developer Services:** Visual Studio Team Services, Azure DevTest Labs\*, VS Application Insights\*, HockeyApp, Developer Tools.
- Management & Security:** Azure Portal, Scheduler, Automation, Log Analytics, Key Vault, Security Center\*.
- Compute:** Virtual Machines, Virtual Machine Scale Sets, Cloud Services, Batch, RemoteApp, Service Fabric, Azure Container Service.
- Web & Mobile:** Web Apps, Mobile Apps, Logic Apps\*, API Apps, API Management, Notification Hubs, Mobile Engagement, Functions\*.
- Data & Storage:** SQL Database, DocumentDB, Redis Cache, Storage: Blobs, Tables, Queues, Files and Disks, StorSimple, Search, SQL Data Warehouse\*, SQL Server Stretch Database\*.
- Analytics:** Data Lake Analytics\*, Data Lake Store\*, HDInsight, Machine Learning, Stream Analytics, Data Factory, Data Catalog, Power BI Embedded\*.
- Internet of Things & Intelligence:** Azure IoT Suite, Azure IoT Hub, Event Hubs, Cortana Intelligence Suite, Cognitive Services\*.
- Media & CDN:** Media Services, Content Delivery Network.
- Identity & Access Management:** Azure Active Directory, B2C\*, Domain Services\*, Multi-Factor Authentication.
- Hybrid Integration:** BizTalk Services, Service Bus, Backup, Site Recovery.
- Networking:** Virtual Network, ExpressRoute, Traffic Manager, Load Balancer, Azure DNS\*, VPN Gateway, Application Gateway.

<http://azureplatform.azurewebsites.net/>

# Azure Objectives

Extend data center infrastructure

Scalability

Separate compute from storage



Shared infrastructure = reduced cost of ownership

Simplified management = reduced cost of ownership

Simplified development = faster time to value

Self-service provisioning

Open source interoperability

Integration of services

Built-in high availability & disaster recovery

Shared code base with (some) on-premises resources

# Types of Cloud Deployments

## IaaS

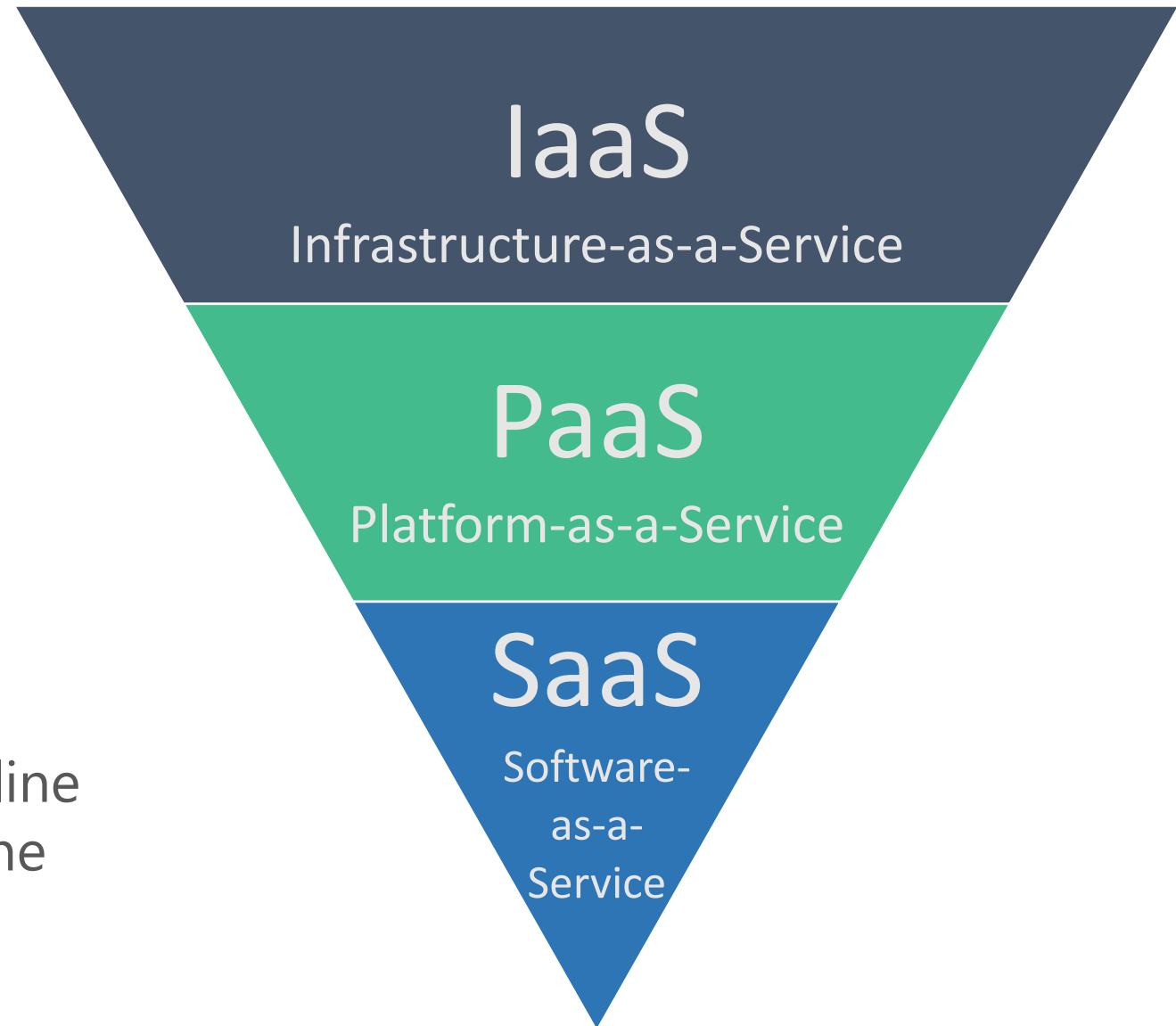
- ✓ Azure Virtual Machines

## PaaS

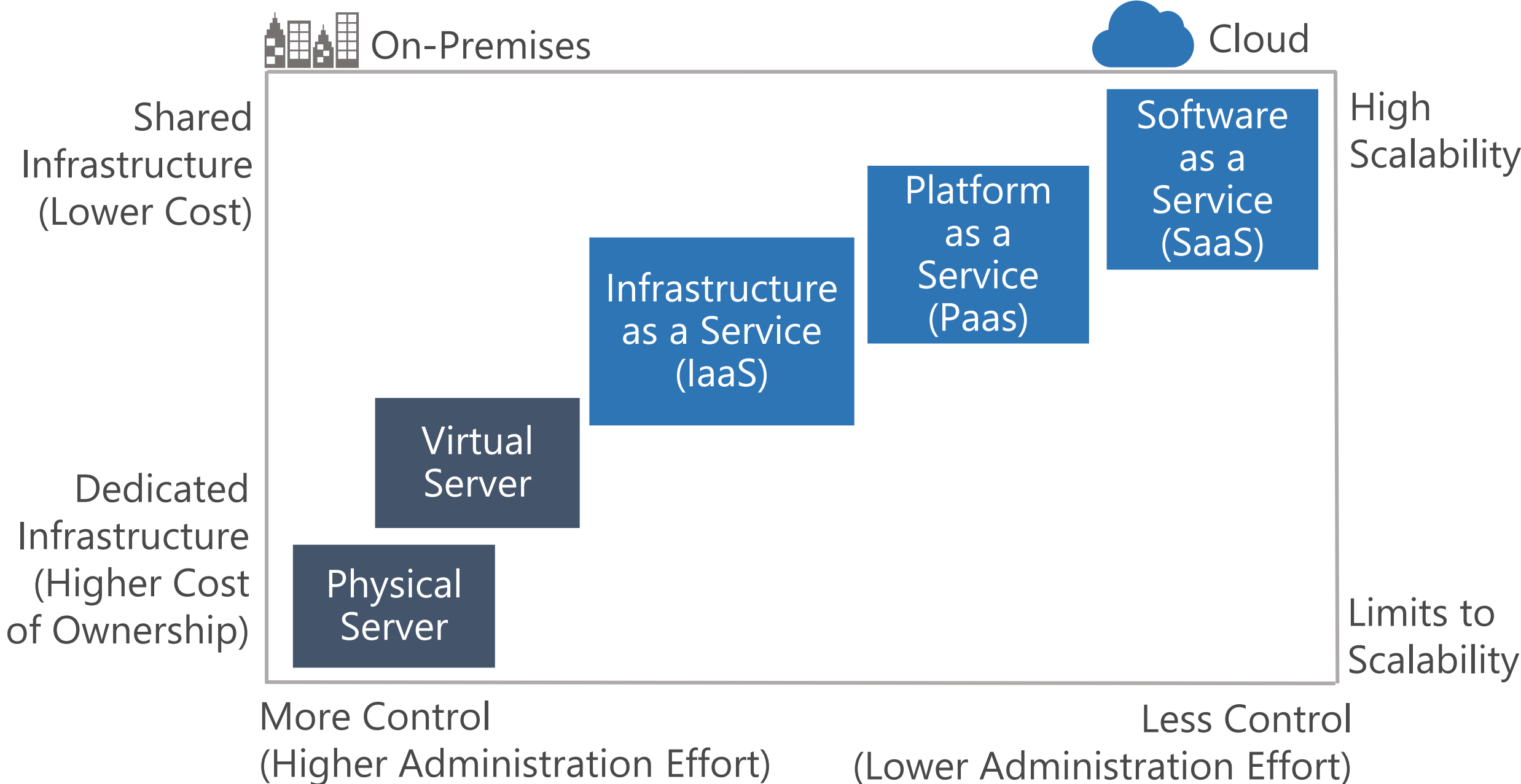
- ✓ Azure SQL Data Warehouse
- ✓ Azure SQL Database
- ✓ Azure HDInsight
- ✓ Azure Data Lake Store
- ✓ Azure Stream Analytics

## SaaS

- ✓ Power BI
- ✓ Office 365
- ✓ Azure Data Catalog
- ✓ Azure Data Factory
- ✓ Azure Machine Learning
- ✓ SharePoint Online
- ✓ Exchange Online

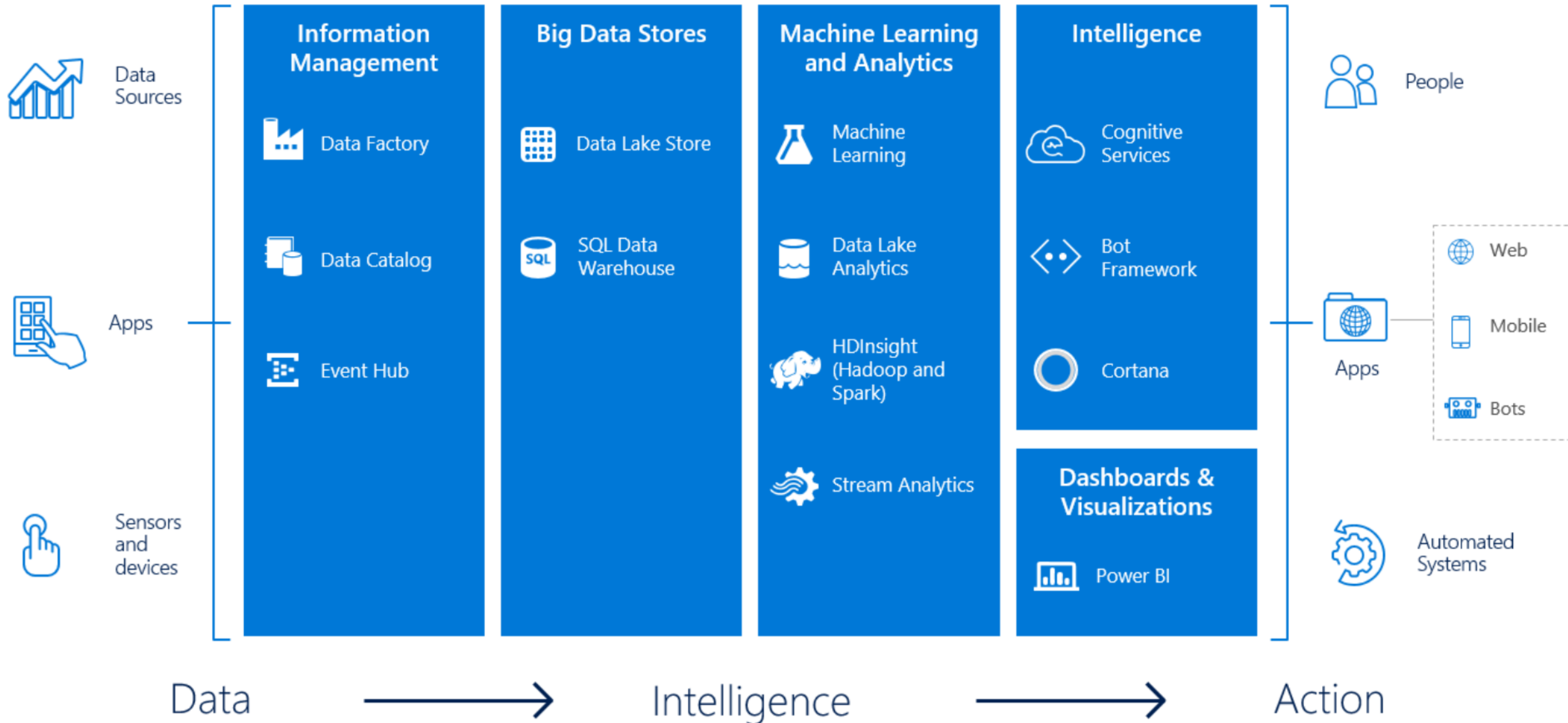


# Types of Cloud Deployments



# Cortana Intelligence Suite

*Formerly known as Cortana Analytics Suite*





# Cortana Intelligence in a Sentence

Collection of  
Azure services

"Cortana Intelligence  
is a platform and a process  
to perform advanced analytics  
from start to finish"

Team Data  
Science Process  
(TDSP)

Source: Chris Testa O'Neill  
Data Science Team at Microsoft

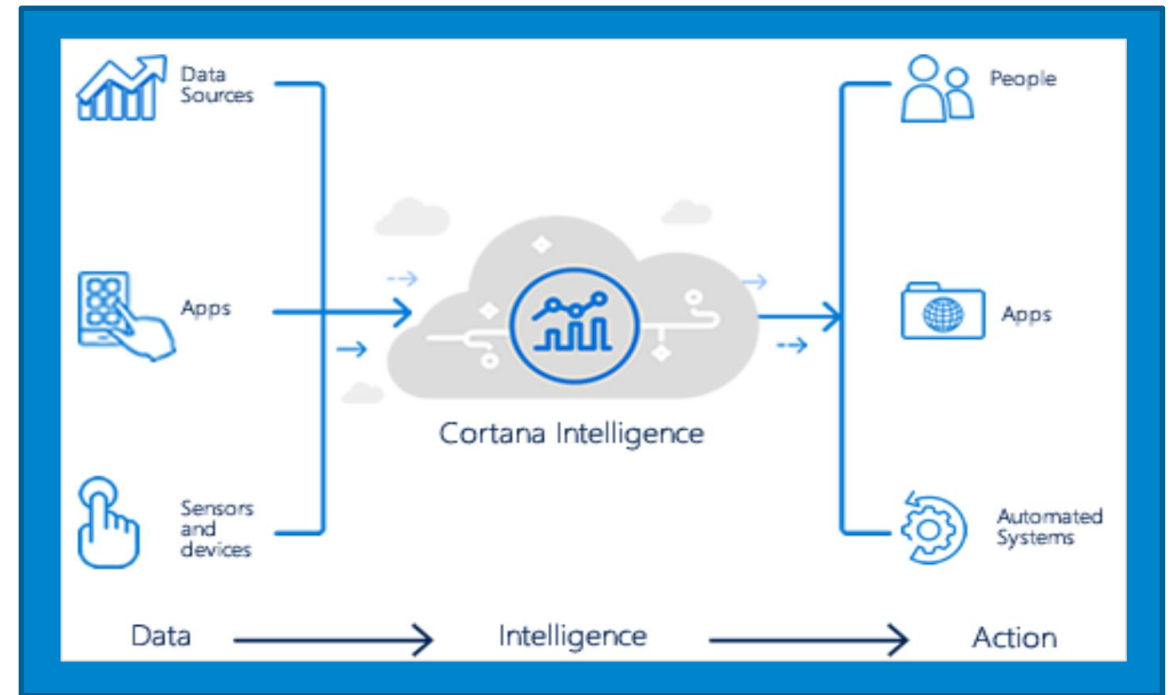
# Cortana Intelligence Suite Objectives

Big Data and Advanced Analytics with less cost and effort

“Intelligent action” from people or automated systems

Enable opportunities for automation and innovation:

- ✓ Templates
- ✓ Preconfigured solutions
- ✓ Interoperability
- ✓ Easier to operationalize solutions
- ✓ Open standards

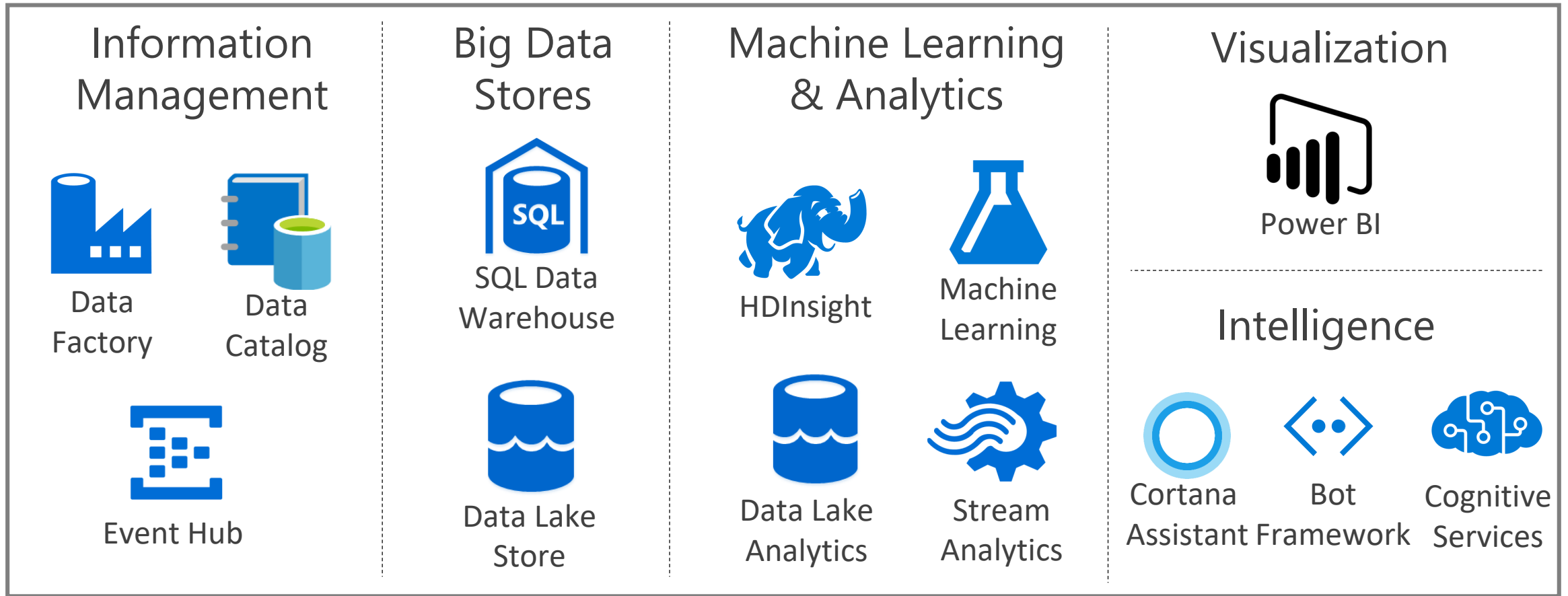


*Cortana Intelligence Suite = a marketing term for a bundle of integrated services*

# Cortana Intelligence Suite Sample Scenarios

Industry	Sales & marketing	Finance & risk	Customer & channel	Operations & workforce
<b>Retail</b>	<ul style="list-style-type: none"><li>• Demand forecasting</li><li>• Loyalty programs</li><li>• Cross-sell &amp; upsell</li><li>• Customer acquisition</li></ul>	<ul style="list-style-type: none"><li>• Fraud detection</li><li>• Pricing strategy</li></ul>	<ul style="list-style-type: none"><li>• Personalization</li><li>• Lifetime customer value</li><li>• Product segmentation</li></ul>	<ul style="list-style-type: none"><li>• Store location demographics</li><li>• Supply chain management</li><li>• Inventory management</li></ul>
<b>Financial services</b>	<ul style="list-style-type: none"><li>• Customer churn</li><li>• Loyalty programs</li><li>• Cross-sell &amp; upsell</li><li>• Customer acquisition</li></ul>	<ul style="list-style-type: none"><li>• Fraud detection</li><li>• Risk &amp; compliance</li><li>• Loan defaults</li></ul>	<ul style="list-style-type: none"><li>• Personalization</li><li>• Lifetime customer value</li></ul>	<ul style="list-style-type: none"><li>• Call center optimization</li><li>• Pay for performance</li></ul>
<b>Healthcare</b>	<ul style="list-style-type: none"><li>• Marketing mix optimization</li><li>• Patient acquisition</li></ul>	<ul style="list-style-type: none"><li>• Fraud detection</li><li>• Bill collection</li></ul>	<ul style="list-style-type: none"><li>• Population health</li><li>• Patient demographics</li></ul>	<ul style="list-style-type: none"><li>• Operational efficiency</li><li>• Pay for performance</li></ul>
<b>Manufacturing</b>	<ul style="list-style-type: none"><li>• Demand forecasting</li><li>• Marketing mix optimization</li></ul>	<ul style="list-style-type: none"><li>• Pricing strategy</li><li>• Performance risk management</li></ul>	<ul style="list-style-type: none"><li>• Supply chain optimization</li><li>• Personalization</li></ul>	<ul style="list-style-type: none"><li>• Remote monitoring</li><li>• Predictive maintenance</li><li>• Asset management</li></ul>

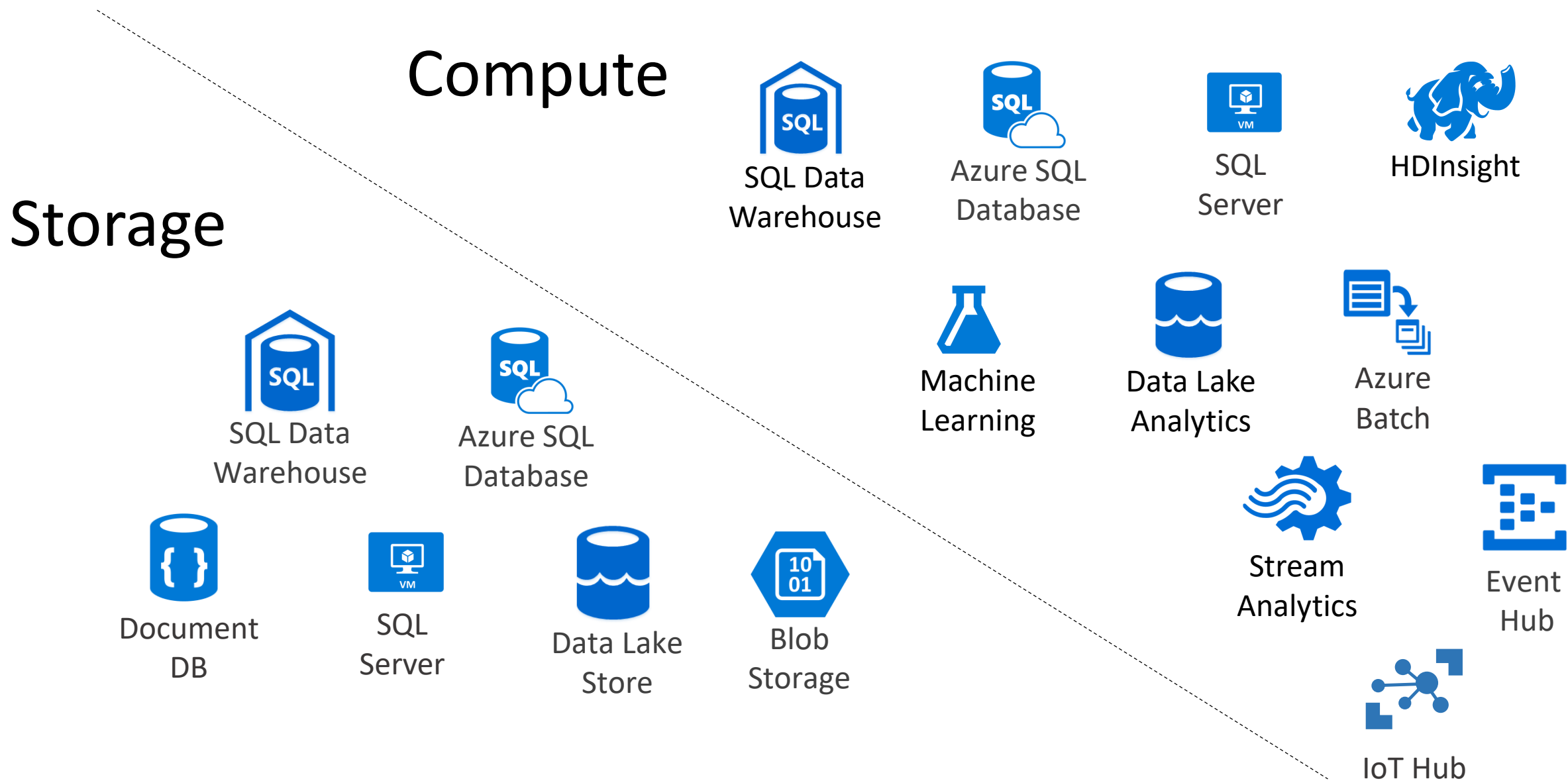
# Cortana Intelligence Components

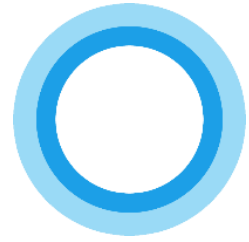


Plus other solution components such as:



# Storage Layers vs. Compute Services





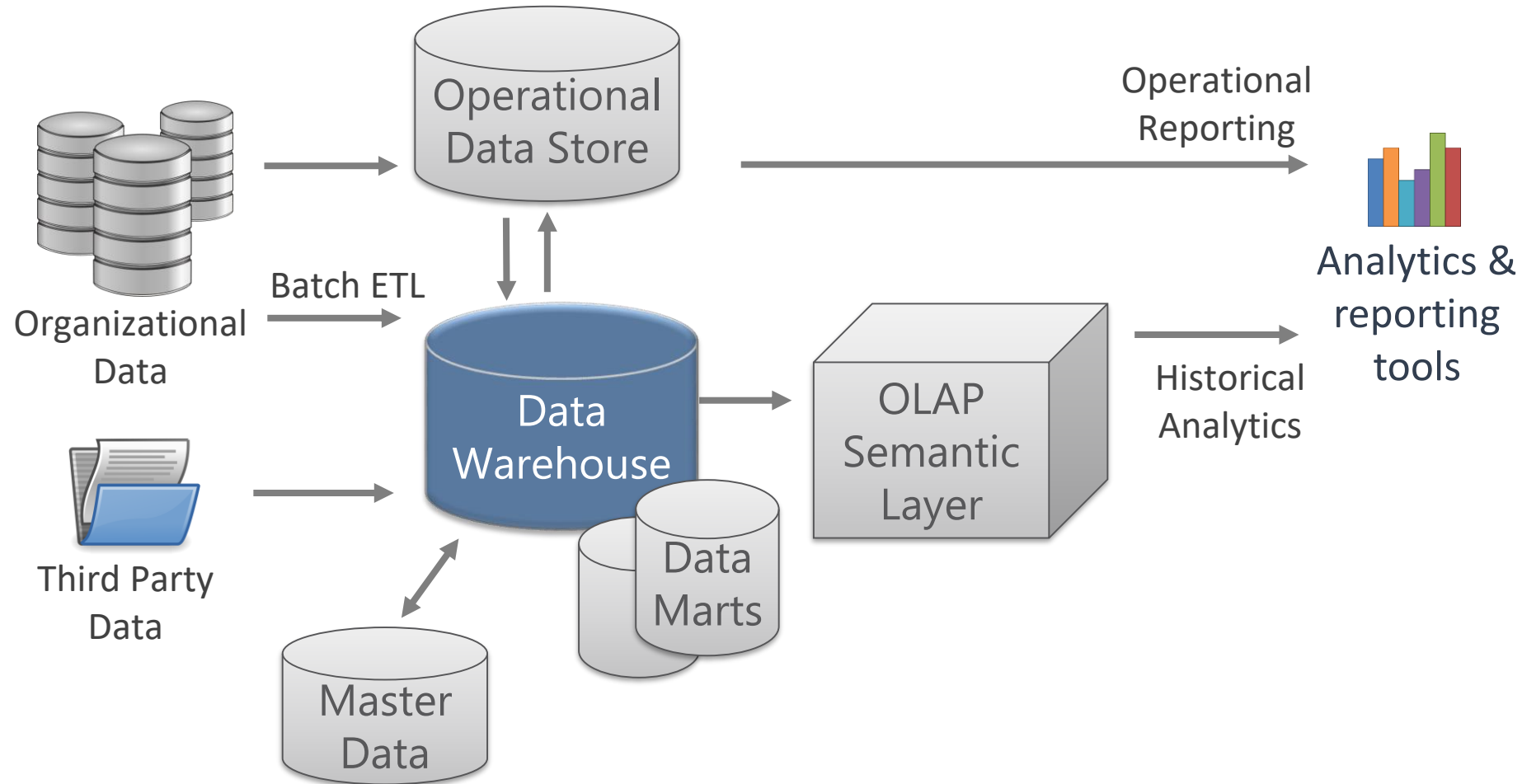
# Cortana Intelligence Suite

---

Scenario:

Loan Repayments

# Loan Repayment Scenario



## Current State:

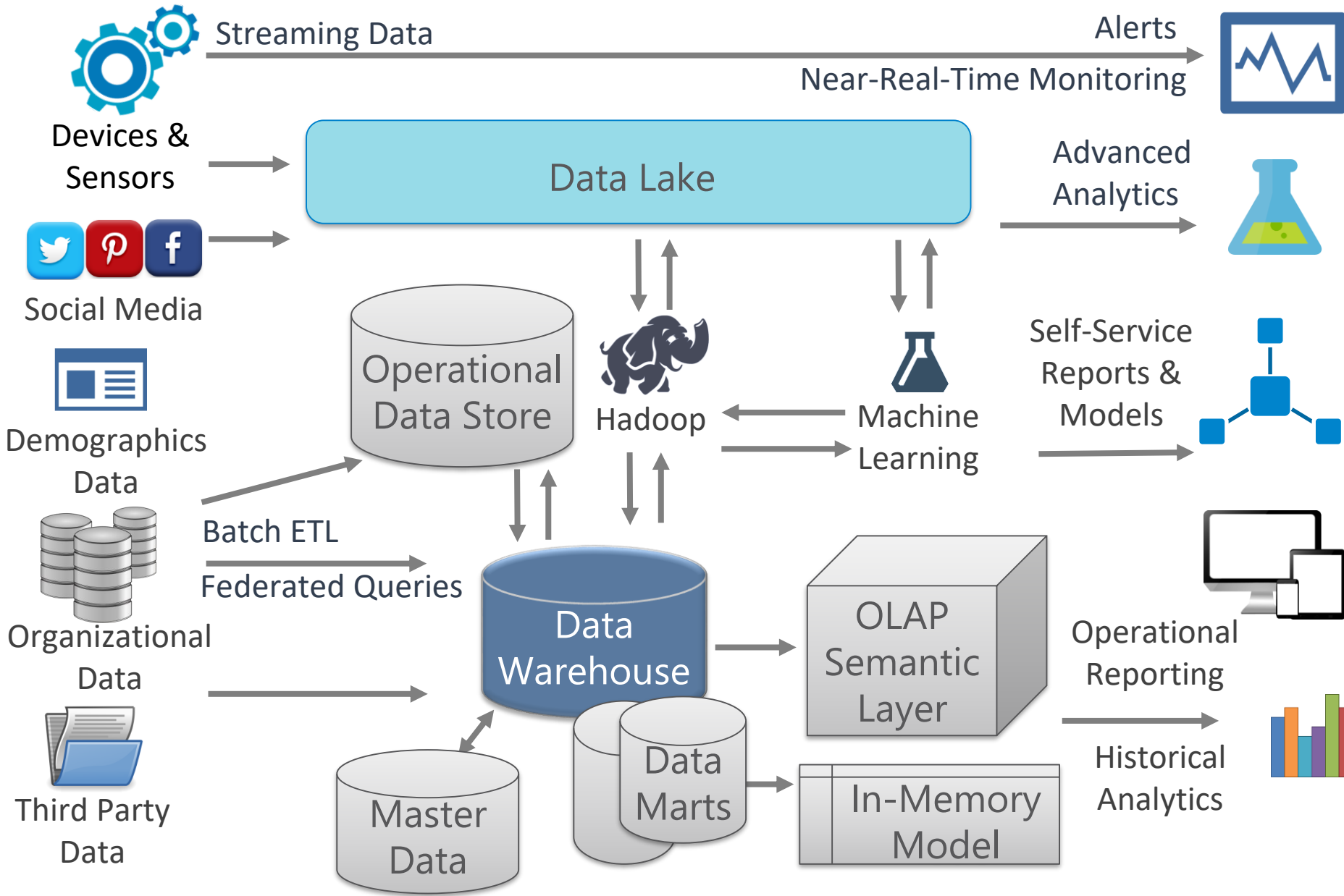
### Organizational data

- Customer application (income, assets)
- Loan history
- Payment activity

### Third party data

- Credit bureau history

# Loan Repayment Scenario



## Desired State:

**Predictive Analytics**  
model to predict  
repayment ability

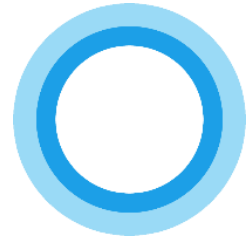
**Fraud alerts** re: line of  
credit withdrawals

**Sentiment analysis** for  
phone records

**Text analytics** for e-mail  
records

Analyze comments made  
on **social media**

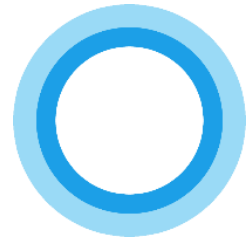




Cortana Intelligence Suite

---

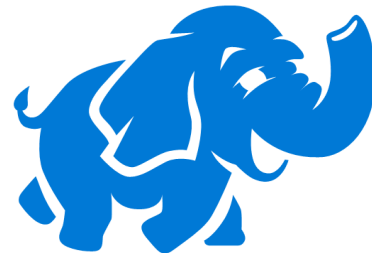
Tour of  
Cortana Intelligence Suite



# Cortana Intelligence Suite

---

## Azure Data Lake



ADLS + ADLA: Generally Available as of November 2016

HDInsight: Generally Available as of October 2013

# Azure Data Lake

A collection of 3 services:



Data Lake Analytics

Big Data queries-as-a-service



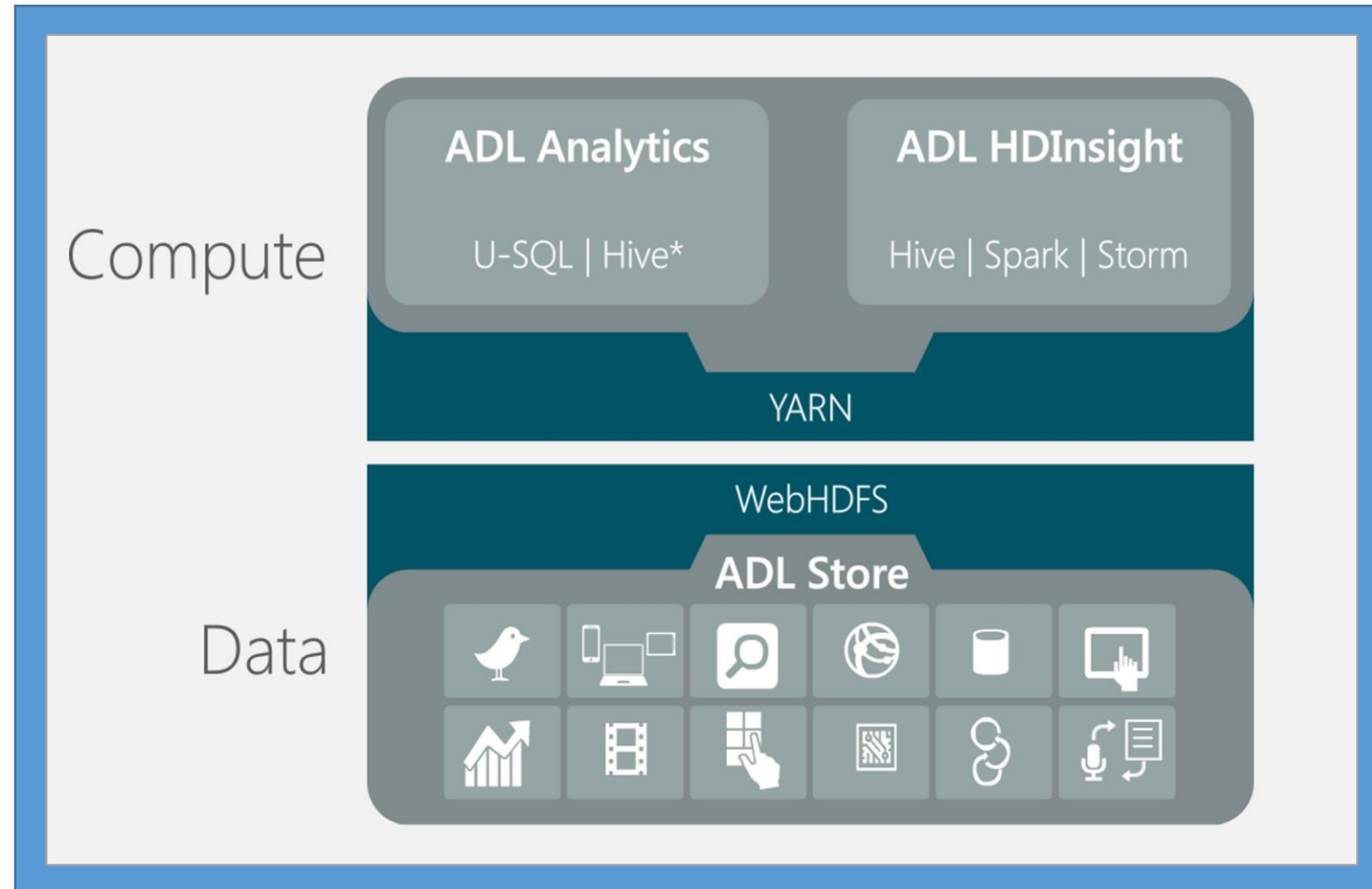
HDInsight

Big Data cluster-as-a-service

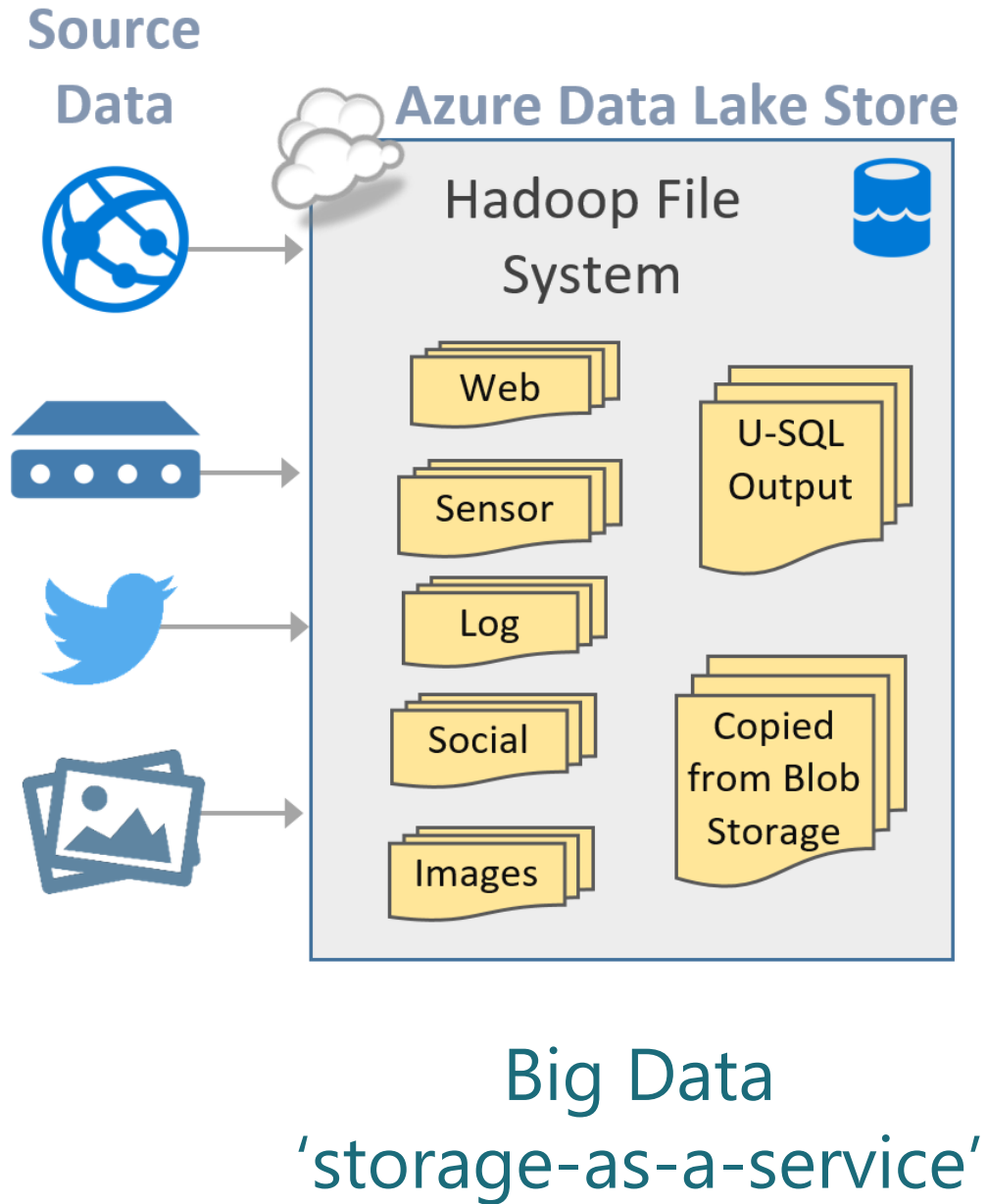


Data Lake Store

Big Data storage as-a-service



# Azure Data Lake Store



A **repository** for analyzing large quantities of disparate sources of data in its native format

One **architectural platform** to house all types of data:

- Machine-generated data (ex: IoT, logs)
- Human-generated data (ex: tweets, e-mail)
- Traditional operational data (ex: sales, inventory)

## Purpose

### Big Data Storage

- ✓ Storage to support analytic applications
- ✓ HDFS (Hadoop Distributed File System) for the cloud, with no size limitations
- ✓ Stores data in its native format: objective is to not reformat



### File System Optimized for Analytics

- ✓ An alternative to general purpose Azure Blob Storage
  - ✓ Parallel read scans
  - ✓ Scaled out over multiple machines
  - ✓ Low latency writes
  - ✓ Large file sizes

### WebHDFS-Compatible

- ✓ Accessible to all HDFS-compliant projects  
*(If integrated with HDInsight)*

## Common Use Cases



### Big Data Analytic Workloads

- ✓ Agility: reduce up-front effort for ingestion of data (defer work to 'schematize')
- ✓ Optimized to work with ADL Analytics
- ✓ Also supports HDInsight (Hadoop)

### Influx of Data

- ✓ Acquire multi-structured data
- ✓ Persist data in its native format
- ✓ No limits on account/file sizes

### Active Archive

- ✓ Rarely used data

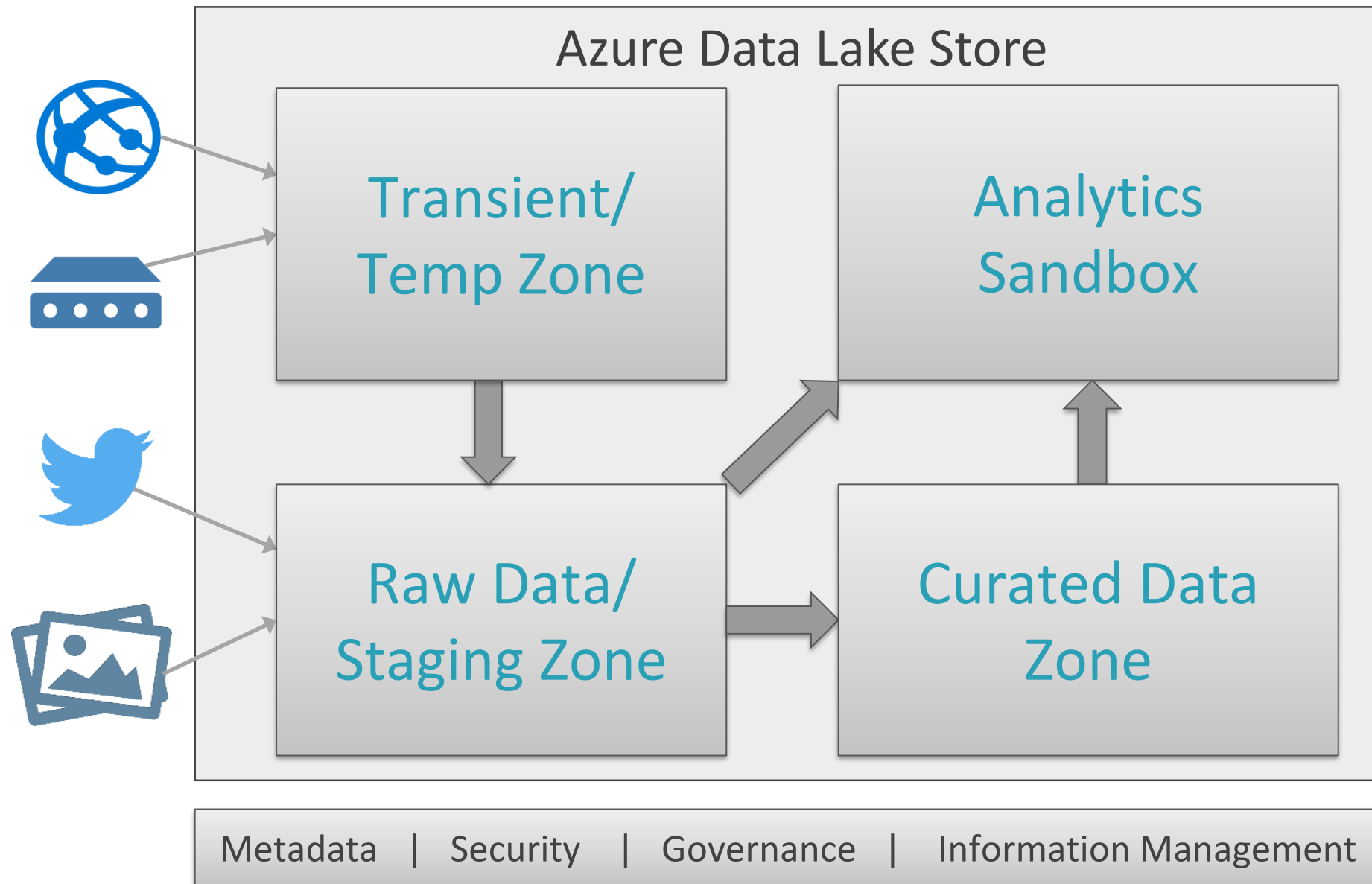
### Data Warehouse Support

- ✓ Staging area for DW

### Data Science

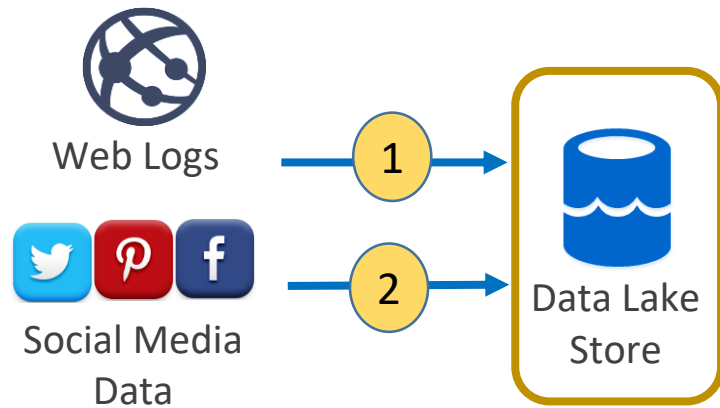
- ✓ Analytic sandbox
- ✓ Raw & curated data zones

## Structuring a Data Lake



## Building Blocks

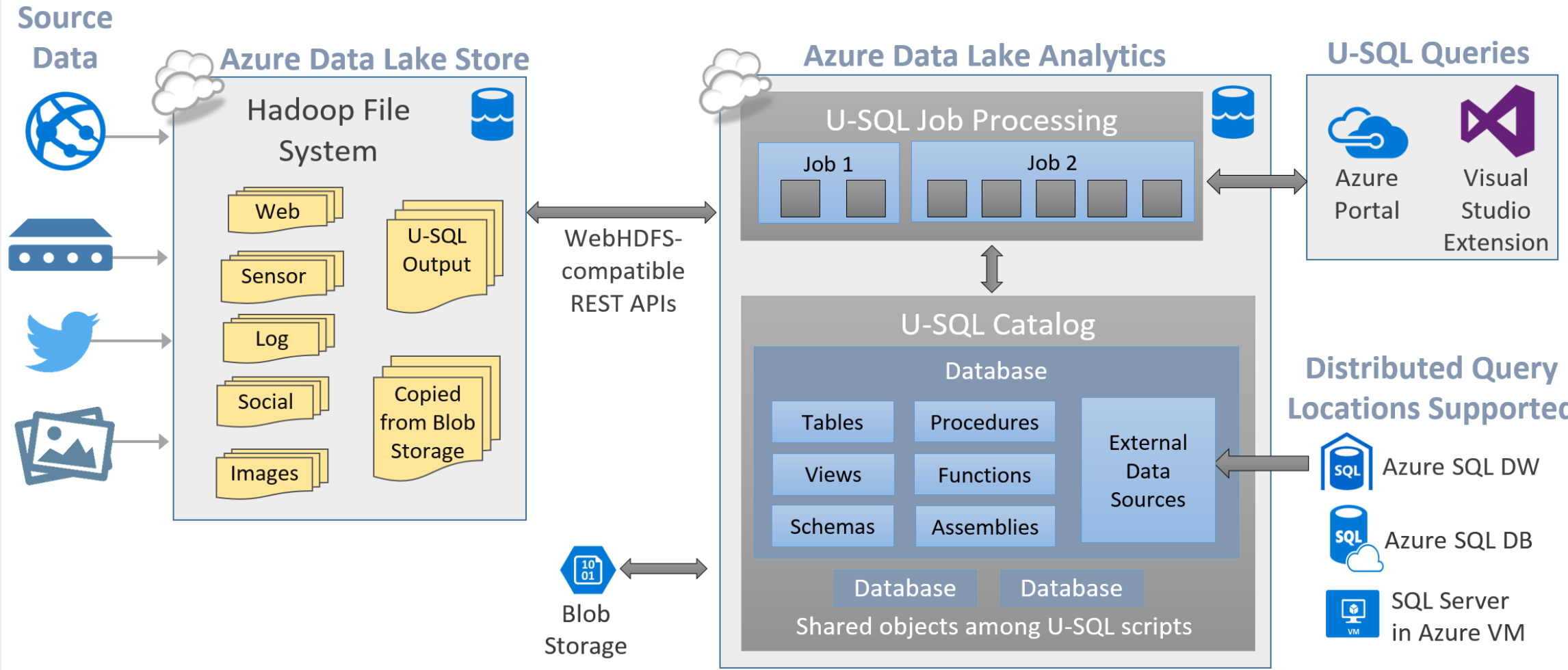
Repository for Ingestion of New Types of Data



- 1 Ingest web logs into the data lake for purpose of analyzing website visits
- 2 Ingest social media data into the data lake for purpose of analyzing comments



# Azure Data Lake Analytics



Big Data 'queries-as-a-service'

## Purpose

### Big Data Processing

- ✓ Ability to process any data, regardless of size or structure
- ✓ Query scalability: resources allocated for each query
- ✓ YARN application built on open standards
- ✓ Optimized to work with Azure Data Lake Store



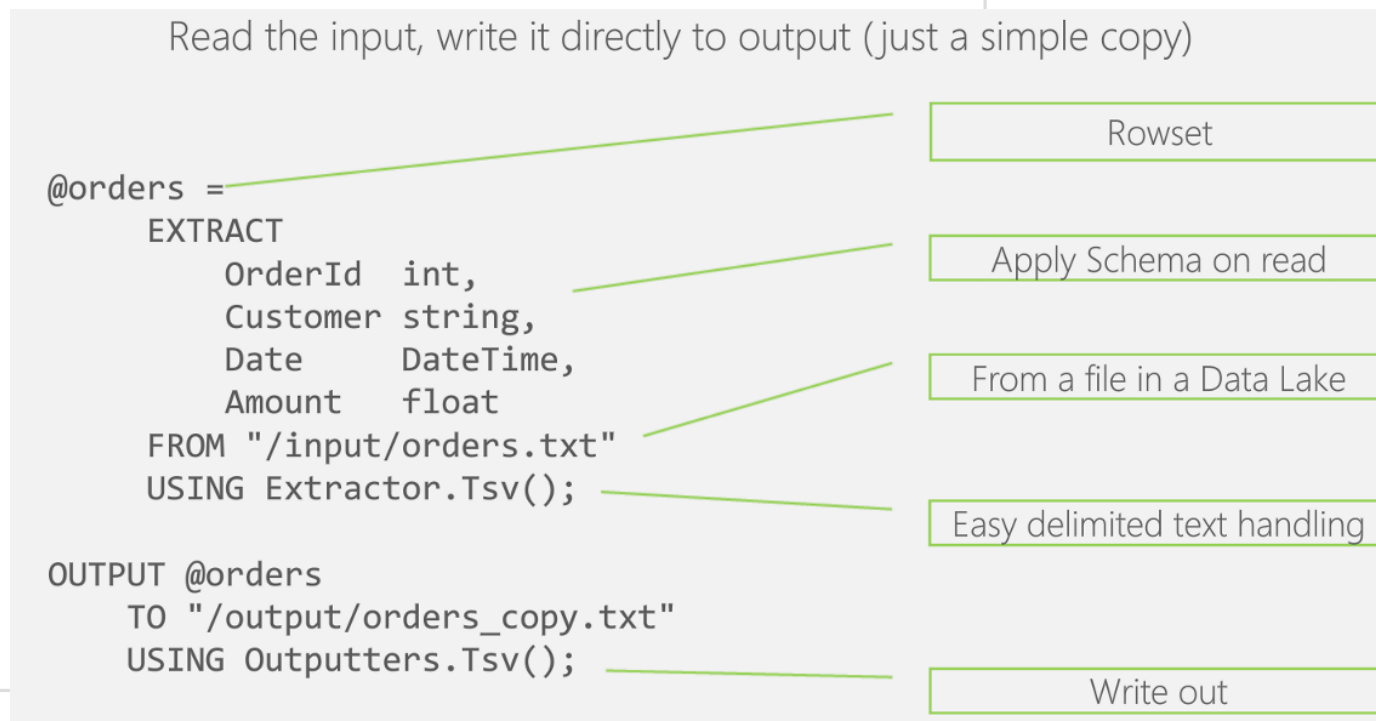
### Simplification

- ✓ Abstracts away the cluster nodes → focuses on convenience, efficiency, and scalability
- ✓ U-SQL = familiar SQL and C# to reduce learning curve
- ✓ Separation of ADL Analytics from ADL Store: easier to manage, debug, and optimize

## U-SQL

### SQL + C#

- ✓ New big data query processing language
- ✓ Applies 'schema on read' logic
- ✓ Mix of multiple SQL dialects (T-SQL and ANSI SQL)
- ✓ Native extensibility of user code written in C#
  - ✓ Full C# expressions
  - ✓ Reuse in assemblies
  - ✓ Define custom types, functions, etc.
- ✓ Automatically scales and parallelizes across nodes



## Common Use Cases

### Focus on Business Logic

- ✓ Focus on jobs rather than on infrastructure for a cluster
- ✓ Abstracts away the cluster nodes and focuses on convenience, efficiency, and scalability



### File Management

- ✓ Scheduled batch processes to manage ADLS or Blob storage  
*(U-SQL is \*not\* currently suitable for ad hoc query workloads)*

### Various Size Workloads

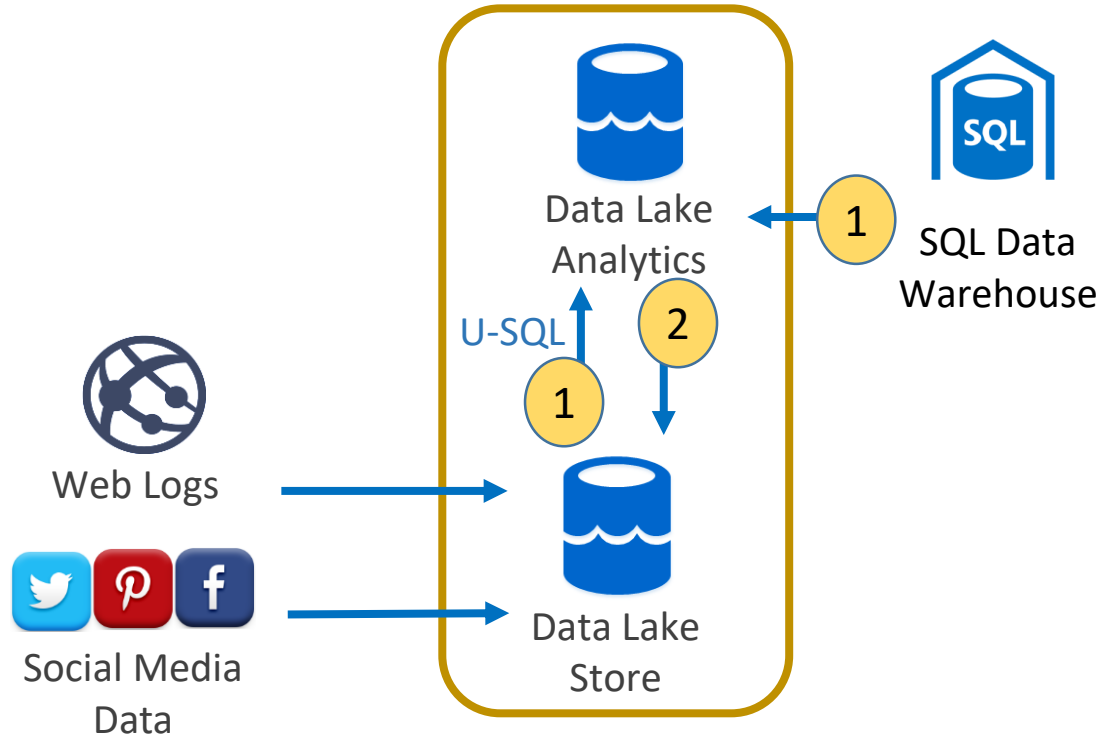
- ✓ Scalability on an individual job basis
- ✓ Objective is to not reserve capacity that's not needed

### U-SQL is a Fit

- ✓ Skillsets and preferences are a fit (U-SQL = SQL + C#)

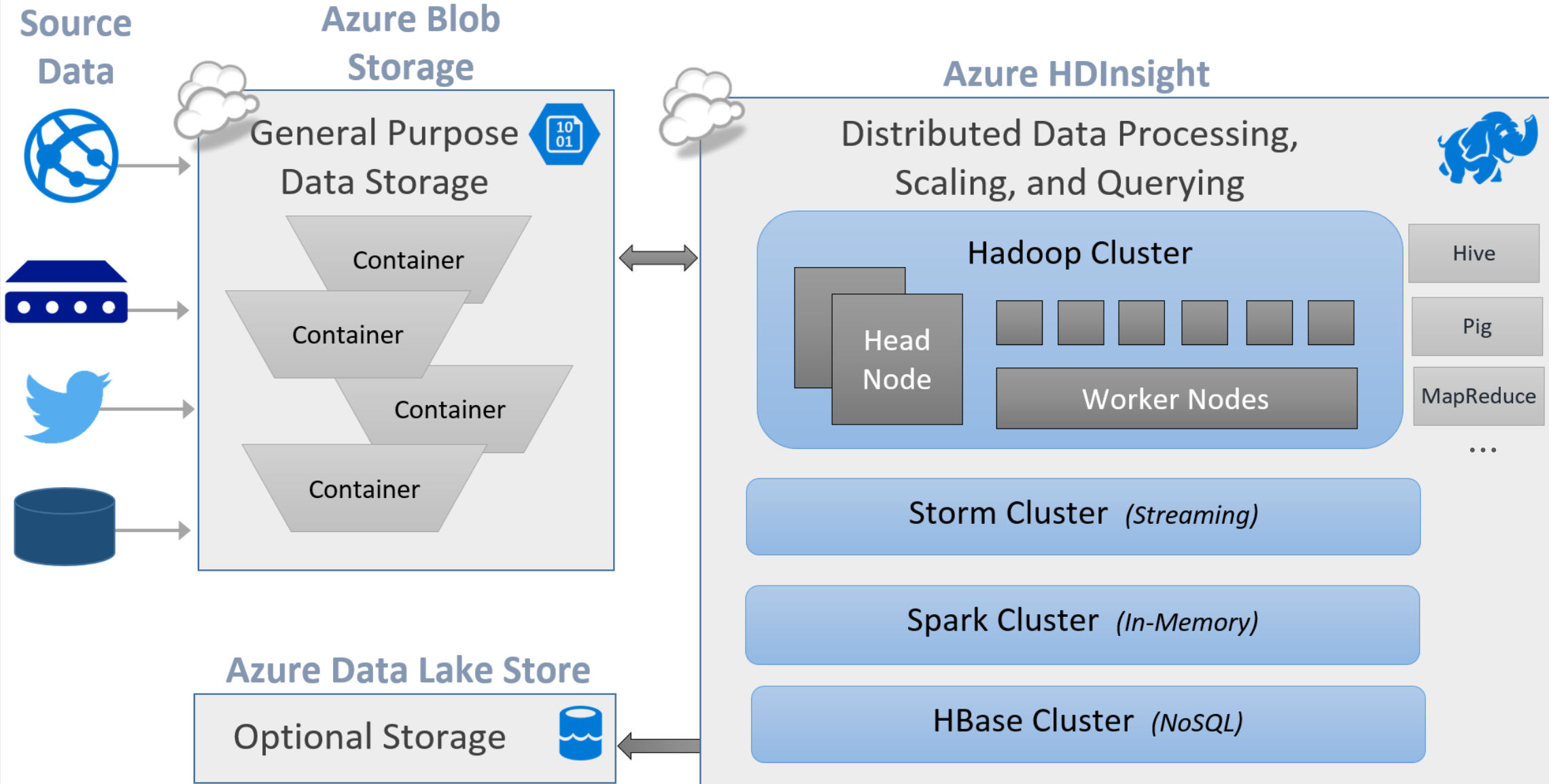
## Building Blocks

Batch Analysis of Data Stored in the Data Lake



- 1 U-SQL: execute federated query from SQL DW + ADL Store to analyze data
- 2 U-SQL: write results back to ADL Store

## Hadoop-based distribution for 'Big Data' solutions



## Purpose

### Big Data Processing

- ✓ A Big Data 'cluster-as-a-service' for distributed data processing, scaling, and querying capabilities
- ✓ Supports the Apache Hadoop open source ecosystem: Hive, Spark, R, Solr, Storm, etc
- ✓ Considered a 'compute' service
- ✓ Linux or Windows



### 3 Ways to Manage HDInsight

- ✓ As a service
- ✓ On-demand (ADF)
- ✓ Inside a virtual machine

### Hortonworks Partnership

- ✓ Based on Hortonworks Data Platform (HDP) distribution
- ✓ Microsoft + Hortonworks joint engineering team

## Common Use Cases

### Big Data Scenarios

Volume | Variety | Velocity

- ✓ Leveraging the Hadoop ecosystem
- ✓ You want to manage a cluster & go beyond what U-SQL can easily do with ADL Analytics
- ✓ Integration with other open source projects



### Development/POC

- ✓ Inexpensive way to test out proof of concept before investing in an on-prem big data cluster

### Data Processing Engine

- ✓ Computations, transformations and data movement for data sent to DW and analytics systems

### Data Exploration

- ✓ Part of data scientist's toolbox

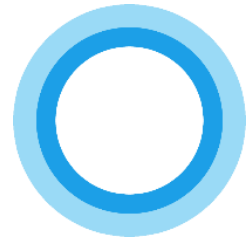
### One-Time Data Loads

- ✓ Large batch jobs (ex: Sqoop)



## Deciding on ADLA vs HDInsight

	Azure Data Lake Analytics (ADLA)	HDInsight
Data Storage	Azure Data Lake Store	Azure Data Lake Store or Azure Blob Storage
Analytical Capability	U-SQL batch processing jobs	Supports all open source projects (Hive, Pig, Spark, SQL-on-Hadoop, etc)
Open Source	No	Yes
Pricing	Pay-as-you-go Scale per U-SQL job	Big data cluster Permanent & on-demand



# Cortana Intelligence Suite

---

## Azure SQL Data Warehouse



Generally Available as of July 2016

# Relational Data Warehousing in Azure\*



## SQL Server in a Virtual Machine (IaaS)

---

Run SQL Server workloads (including SSAS, SSRS, SSIS, etc) in an Azure Virtual Machine.

Best for:

- ✓ Migrating/extending existing database
- ✓ Administer all aspects
- ✓ Bring your own license
- ✓ Dev/test scenarios



## Azure SQL Database (PaaS)

---

A relational database-as-a-service (DBaaS). Close feature parity with SQL Server.

Best for:

- ✓ < 1TB data volume (sharding across DBs is not suitable for DW workloads)
- ✓ OLTP with scaling & pooling needs (unpredictable workloads)
- ✓ Reduced administration of DB, O/S, and hardware



## Azure SQL Data Warehouse (PaaS)

---

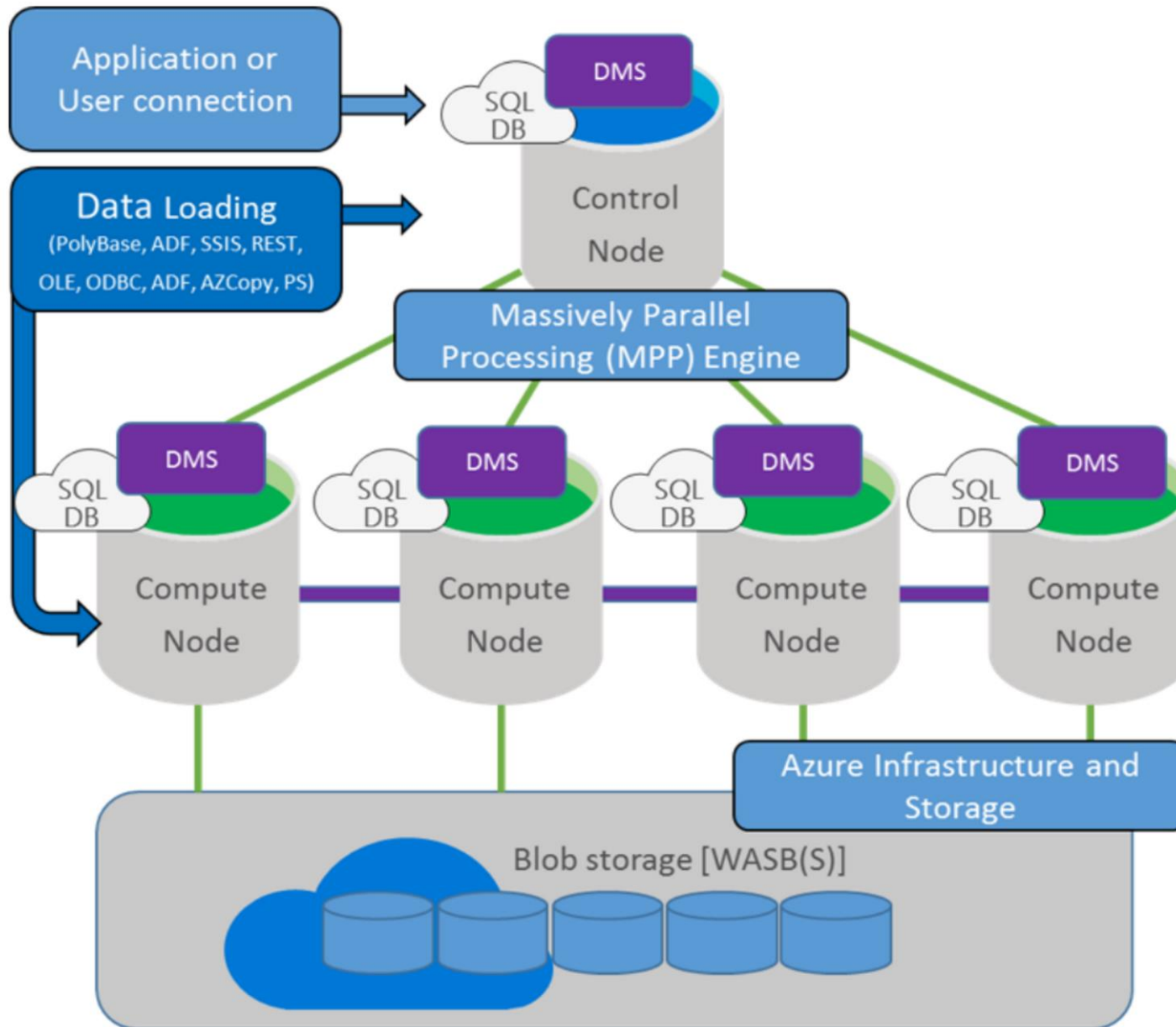
A DW-as-a-service (DWaaS) optimized for performance and large scale, distributed workloads.

Best for:

- ✓ Larger data volumes on MPP architecture
- ✓ Ability to scale up/down/pause on-demand
- ✓ Combining relational + nonrelational data

*\*Excluding: Other technologies in a VM such as Oracle, and Big Data technologies like Hive, etc*

# Azure SQL DW



- ✓ Large-scale DW workloads
- ✓ Massively parallel processing (MPP) architecture across distributions
- ✓ Part of the SQL Server family (with some key differences)

## Purpose

### MPP Scale-Out Query Engine

- ✓ Cloud-based, multi-tenant, platform-as-a-service (PaaS)
- ✓ Massively parallel processing (MPP)
- ✓ Built on SQL Server (with some differences & limits)
- ✓ Clustered columnstore indexes used by default



### Elastic Scale

- ✓ Scale up/down on-demand or on schedule

### PolyBase

- ✓ T-SQL for Hadoop queries & data loads

### Storage + Compute

- ✓ Storage and compute is decoupled
- ✓ Separate billing & scaling for storage vs. compute
- ✓ Data Warehouse Units (DWUs) controls compute billing
- ✓ Increase/decrease/pause compute ability independently of data storage

## Common Use Cases



### Analytical and Ad Hoc Workloads

- ✓ Batch inserts and updates
- ✓ OLTP workloads are \*not\* suitable for SQL DW

### Varying Workloads

- ✓ Large workloads which suit the ability to scale 'compute' up/down (ex: data loads or intensive analytical operations)

### Large Scale Cloud DW

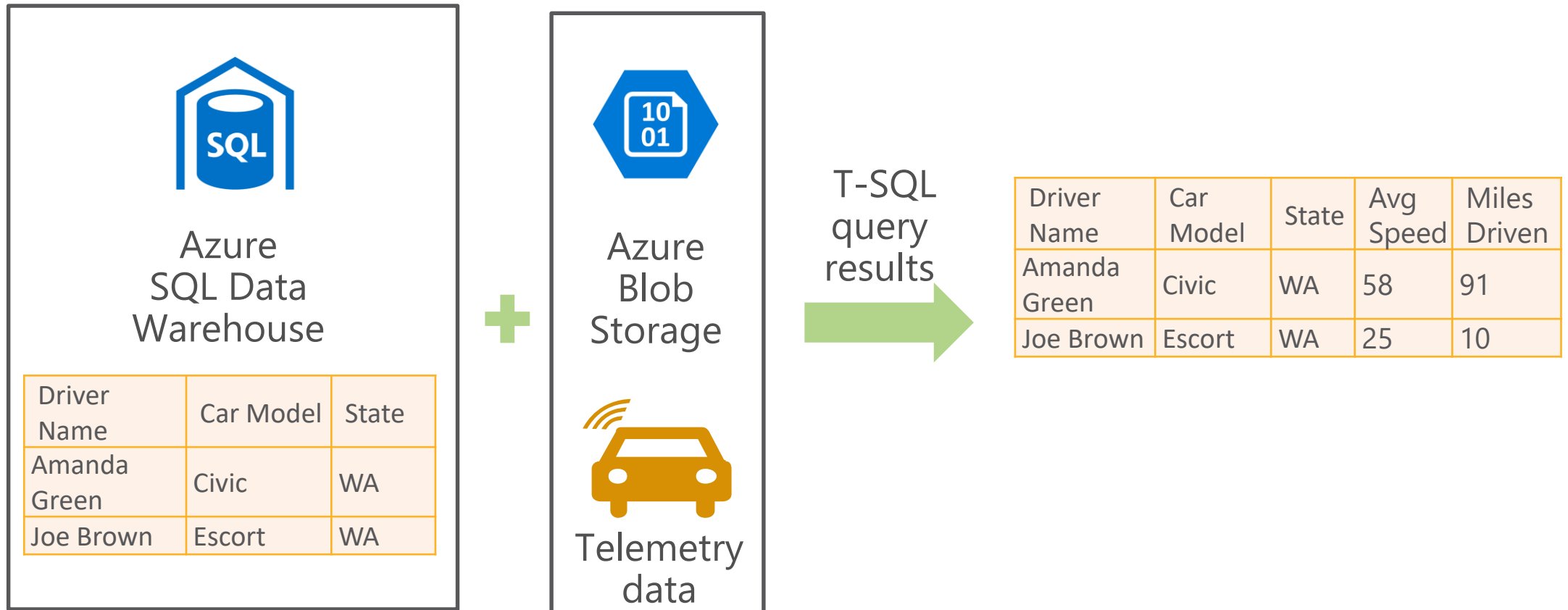
- ✓ Easier to provision large-scale environments in cloud than on-premises

### Data Variety

- ✓ Integration with various data source types and data structures (i.e., takes advantage of PolyBase)

## Two Ways of Using PolyBase

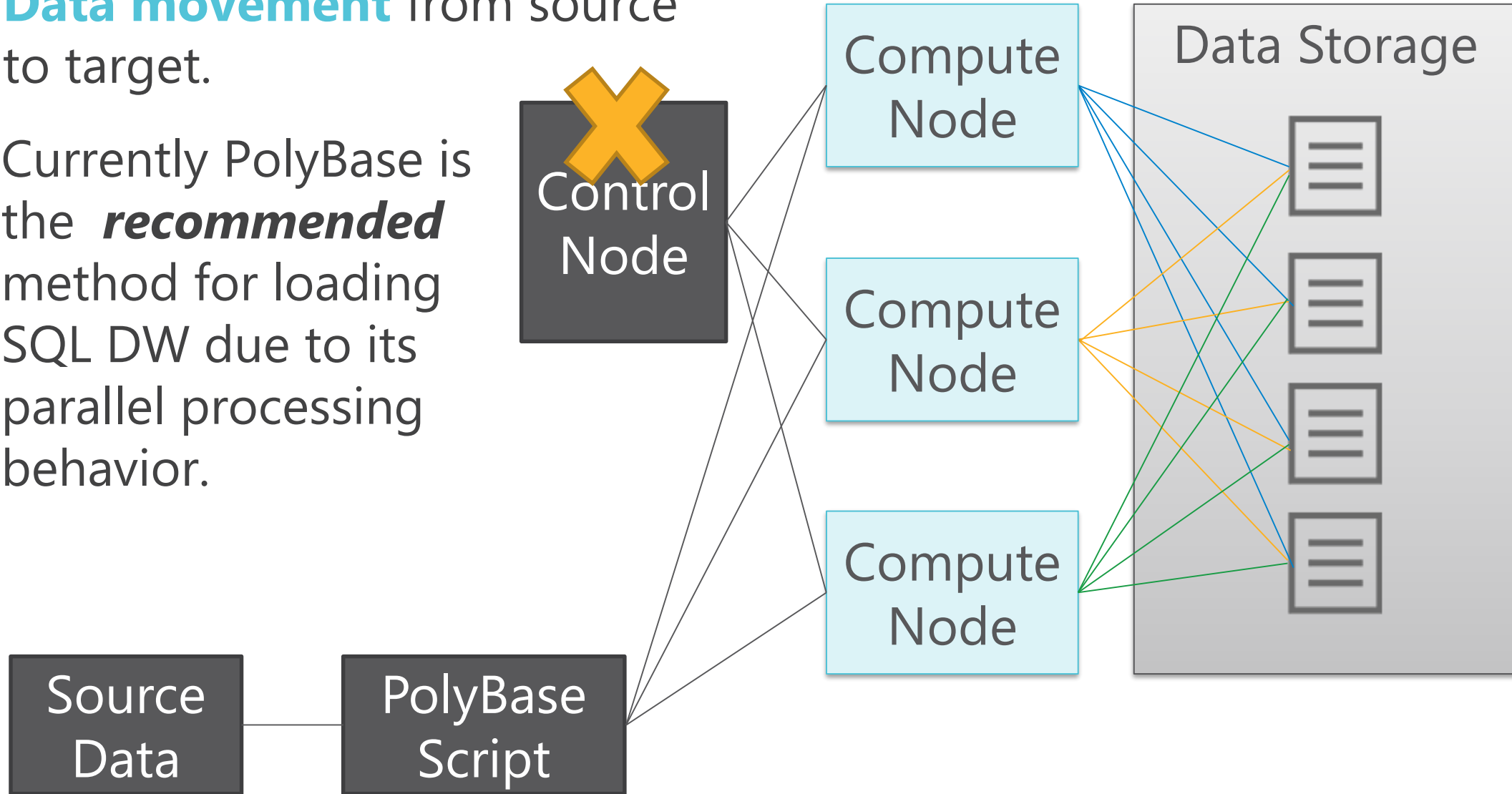
- 1 **Querying** of relational + semi/unstructured data in a single consolidated query. Objective: avoid data movement from where the data currently resides.



## Two Ways of Using PolyBase

- 2 **Data movement** from source to target.

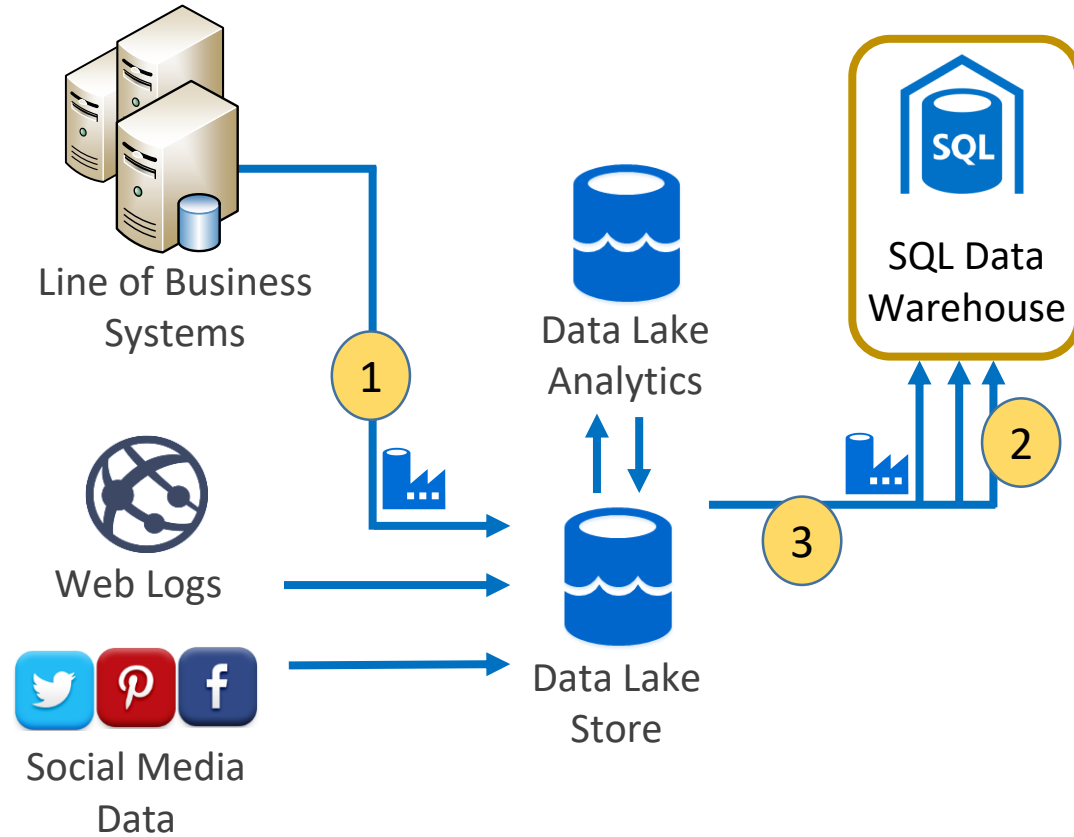
Currently PolyBase is the **recommended** method for loading SQL DW due to its parallel processing behavior.



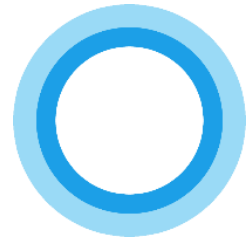


## Building Blocks

Modern Data Warehouse with a Data Lake



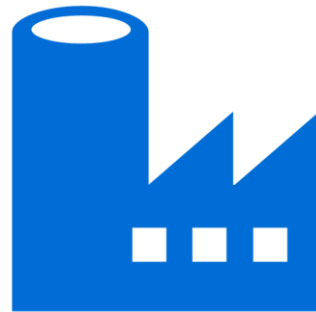
- 1 Load staging data from on-premises source systems to the data lake
- 2 PolyBase used for loading from the data lake to the data warehouse
- 3 PolyBase used for federated queries between the DW & ADL Store  
*\*Performance is not yet optimized for this use case, but it works\**



# Cortana Intelligence Suite

---

## Azure Data Factory

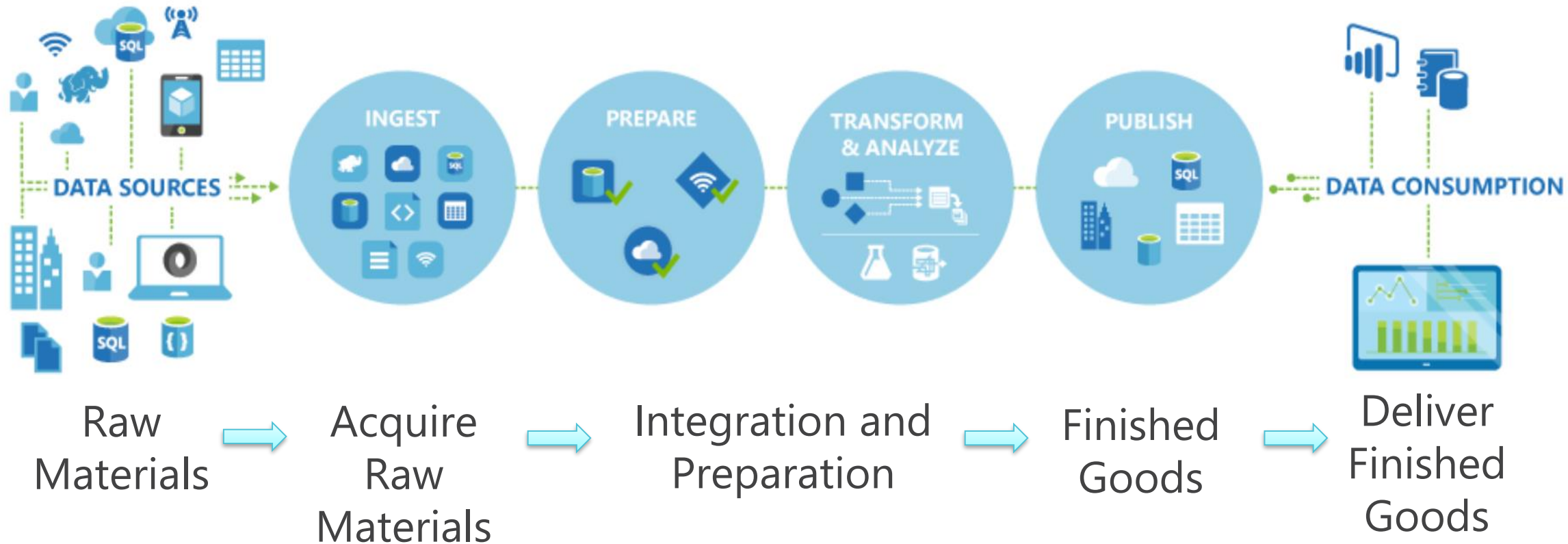


Generally Available as of August 2015

# Azure Data Factory

A service for building and operating data pipelines

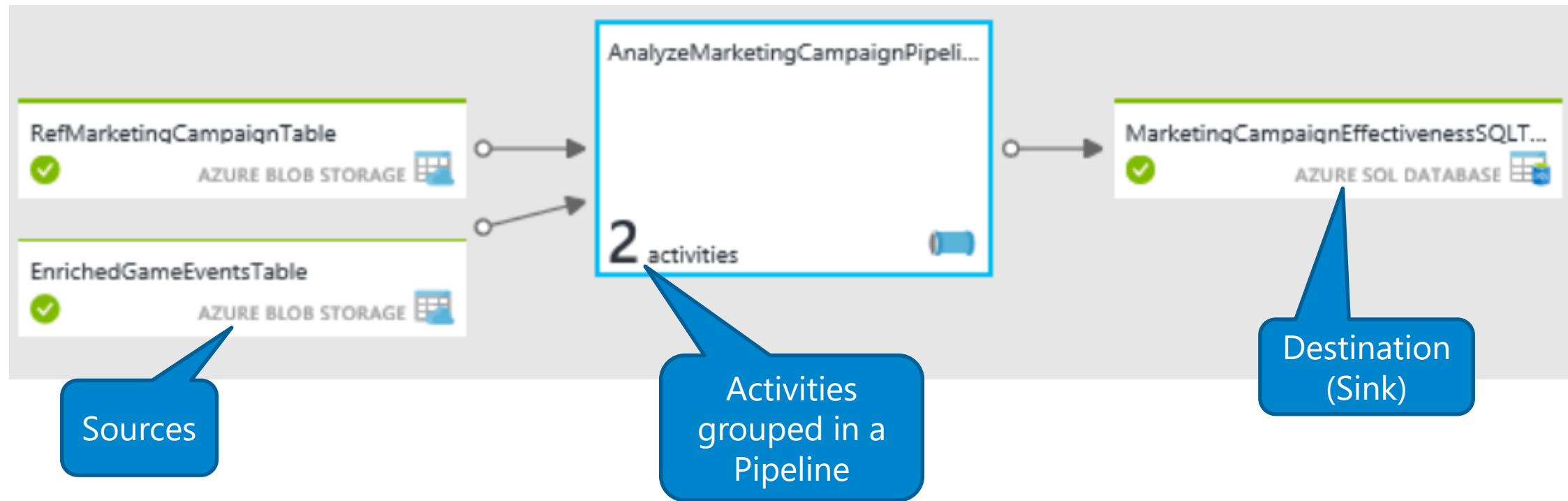
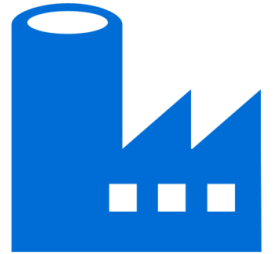
*Factory analogy:*



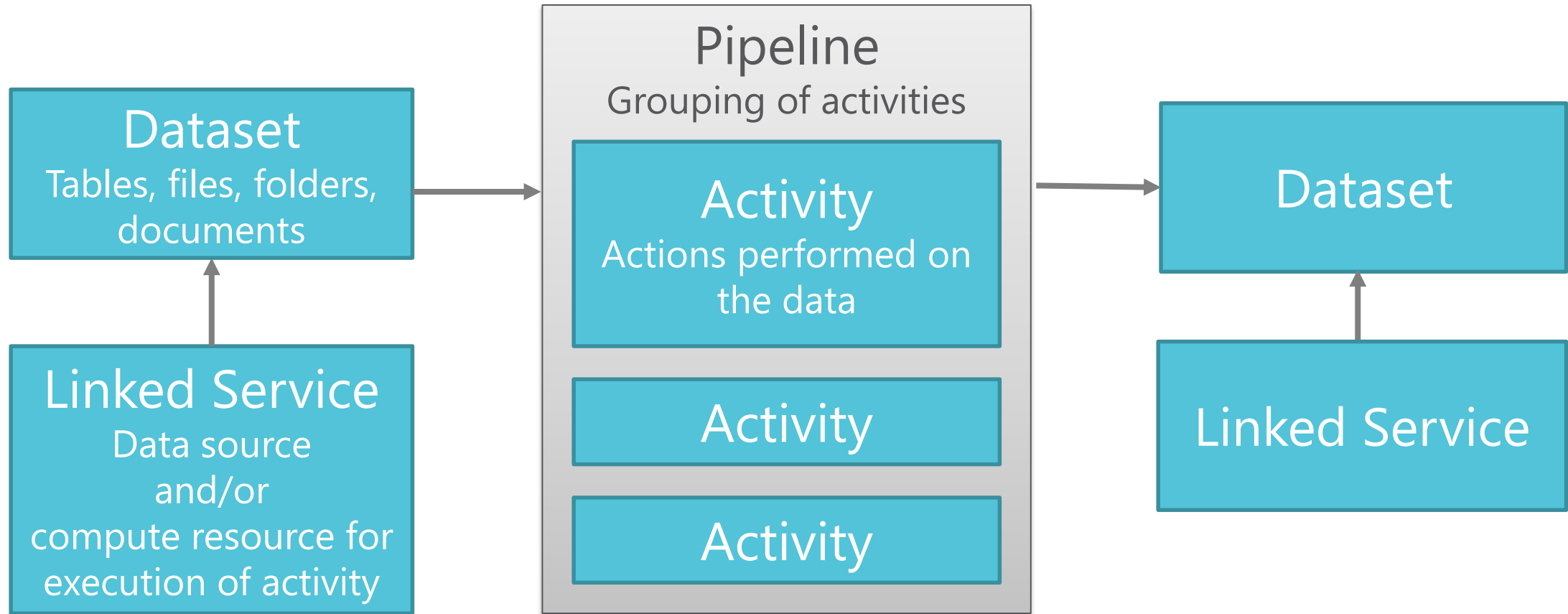
## Purpose

### Data Orchestration

- ✓ Automation of data ingestion, orchestration, and data processing
- ✓ Serves as the 'glue' for stitching together services



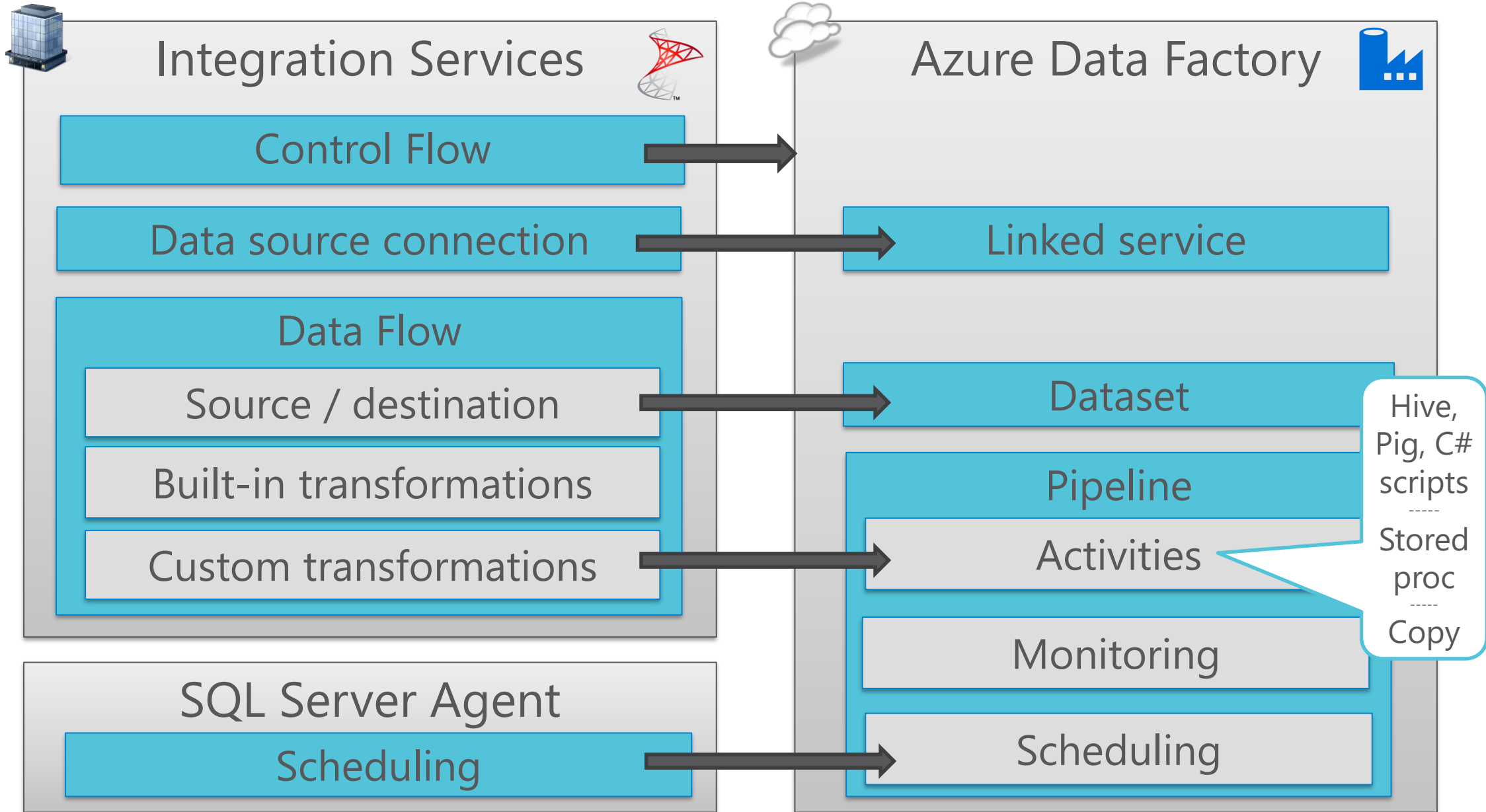
## Components of a Factory



	Source	Compute Environment	Activity	Sink
Ex1	Azure SQL Database	Azure SQL Database	Stored Procedure	Azure SQL Data Warehouse
Ex2	Azure Data Lake Store	Azure Data Lake Analytics	U-SQL Script	Azure Data Lake Store

# Azure Data Factory

## How Does ADF compare to Integration Services?

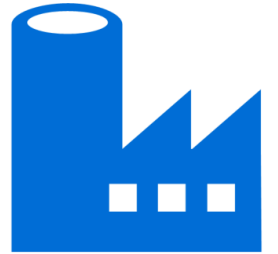


## Key Differences from SSIS

- ✓ Transformation capabilities: Hive, Pig, C#, or SQL DB stored procs, Copy activity
- ✓ Pipeline intent and scheduling are all combined together in ADF (not modular)

## JSON

- ✓ Most ADF elements are hand-written JSON scripts



## Planning for One Data Factory vs. Multiple Factories

- ✓ An ADF data management gateway can be used for only one ADF
- ✓ ADF diagram can get large
  - ✓ Grouping activities within pipelines can help with volume
  - ✓ Data lineage can be tracked across one ADF diagram
- ✓ Alignment of each factory with a Visual Studio project

# Azure Data Factory

BI-ADF-Dev

Pipelines/PL\_MoveFromDocDBtoADLS

New data store ... More

- Linked services
  - LS\_ADLS
  - LS\_DocDB
- Datasets
  - DS\_ADLS
  - DS\_DocDB
- Pipelines**
  - PL\_MoveFromDocDBtoADLS**
- Data Gateways
  - No data gateways in this data factory.
  - Drafts

Add activity Encrypt Clone Discard Deploy

```
{
  "name": "PL_MoveFromDocDBtoADLS",
  "properties": {
    "description": "Standardize the different JSON document structures, and relocate from DocumentDB to Azure Data Lake Store in CSV format.",
    "activities": [
      {
        "type": "Copy",
        "typeProperties": {
          "source": {
            "type": "DocumentDbCollectionSource",
            "query": "$Text.Format('SELECT TC.SessionID,TC.ClientID,TC.TimestampUtc,TC.EventData.EventClass,TC.EventData.FeaturePath ?? \\N/A\"
            "nestingSeparator": "."
          },
          "sink": {
            "type": "AzureDataLakeStoreSink",
            "writeBatchSize": 0,
            "writeBatchTimeout": "00:00:00"
          }
        },
        "inputs": [
          {
            "name": "DS_DocDB"
          }
        ],
        "outputs": [
          {
            "name": "DS_ADLS"
          }
        ],
        "policy": {
          "timeout": "1.00:00:00",
          "concurrency": 1,
          "executionPriorityOrder": "NewestFirst",
          "style": "StartOfInterval",
          "retry": 1,
          "longRetry": 0,
          "longRetryInterval": "00:00:00"
        },
        "scheduler": {
          "frequency": "Day",
          "interval": 1
        }
      }
    ],
    "name": "AC_MoveFromDocDBtoADLS"
  }
},
"start": "2016-12-22T00:00:00Z",
"end": "2099-12-31T05:00:00Z",
"isPaused": true,
"hubName": "bi-adf-dev_hub",
"pipelineMode": "Scheduled"
}
}
```

\$\$\$Text.Format('

```
SELECT
TC.SessionID,
TC.MachineID,
TC.TimestampUtc,
TC.EventData.ControlName ?? \\N/A\" AS ControlName
FROM TC
WHERE TC.TimestampUtc >= \\'{0:yyyy-MM-ddTHH:mm:ssZ}\\'
AND TC.TimestampUtc < \\'{1:yyyy-MM-ddTHH:mm:ssZ}\\',
WindowStart, WindowEnd)"
```



## ADF Monitoring & Management App

Data factory

**RESOURCE EXPLORER**

- ▲ Data Factories
  - ▲ adftesting
    - ▲ Pipelines
      - Copy\_BlobToSQLDB\_MortgageDefaults
      - Copy\_MortgageDefaults\_SourceToDest
    - ▲ Datasets
      - AzureBlobLoc\_enterprisesourcefiles\_fin...
      - AzureSqlTableLoc\_AdventureWorksLT...
      - AzureSqlTableLoc\_AdventureWorksLT...
      - AzureSqlTableLoc\_AdventureWorksLT...
    - ▲ Linked Services
      - AzureSqlLS\_AdventureWorksLT
      - AzureStorageLS\_enterprisesourcefiles
    - Gateways

**adftesting**

Start time (UTC): 04/10/2016 05:23 pm End time (UTC): 07/11/2016 05:23 pm Apply

**ACTIVITY WINDOWS**

No filter applied. Last refreshed a few seconds ago.

Pipeline	Activity	Window...	Window...	Status	Type	Attempt...	Attempt...	Duration	Retry Att...
Copy_Blo...	Copy_Mo...	07/09/20...	07/10/201...	⌚ Waiting	Copy	--	--	--	1
Copy_Mo...	Copy_Mo...	07/09/20...	07/10/201...	✅ Ready	Copy	07/10/201...	07/10/201...	00:01:31	1
Copy_Blo...	Copy_Mo...	07/08/20...	07/09/20...	⌚ Waiting	Copy	--	--	--	1
Copy_Mo...	Copy_Mo...	07/08/20...	07/09/20...	✅ Ready	Copy	07/09/20...	07/09/20...	00:02:02	1
Copy_Mo...	Copy_Mo...	07/07/20...	07/08/20...	✅ Ready	Copy	07/08/20...	07/08/20...	00:02:02	1
Copy_Blo...	Copy_Mo...	07/07/20...	07/08/20...	⌚ Waiting	Copy	--	--	--	1
Copy_Blo...	Copy_Mo...	07/06/20...	07/07/20...	⌚ Waiting	Copy	--	--	--	1
Copy_Mo...	Copy_Mo...	07/06/20...	07/07/20...	✅ Ready	Copy	07/07/20...	07/07/20...	00:01:01	1
Copy_Blo...	Copy_Mo...	07/05/20...	07/06/20...	⌚ Waiting	Copy	--	--	--	1
Copy_Mo...	Copy_Mo...	07/05/20...	07/06/20...	✅ Ready	Copy	07/06/20...	07/06/20...	00:01:31	1
Copy_Blo...	Copy_Mo...	07/04/20...	07/05/20...	⌚ Waiting	Copy	--	--	--	1

**Activity Window Explorer**

← JUL 2016 →

Su	Mo	Tu	We	Th	Fr	Sa
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

07/09/2016 12:00 AM UTC - 07/10/2016 12:00 A...

**Activity Window**

Start & End Time  
07/09/2016 12:00 AM UTC - 07/10/2016 12:00 A...

Status  
✅ Ready

Activity  
Copy\_MortgageDefaults\_SQLDBSourceToDest

Pipeline  
Copy\_MortgageDefaults\_SourceToDest

## Common Use Cases

### Big Data Processing

- ✓ HDInsight & big data stores are its strength
- ✓ Provide the script you want to run (Hive, Pig, etc) & it will spin up/tear down an HDInsight cluster on-demand

### Operationalize Solutions

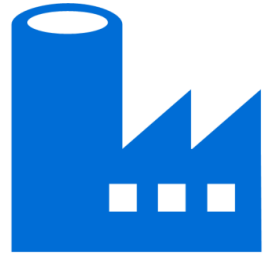
- ✓ Scheduling of data processing scripts

### Stage Into Supported Data Source

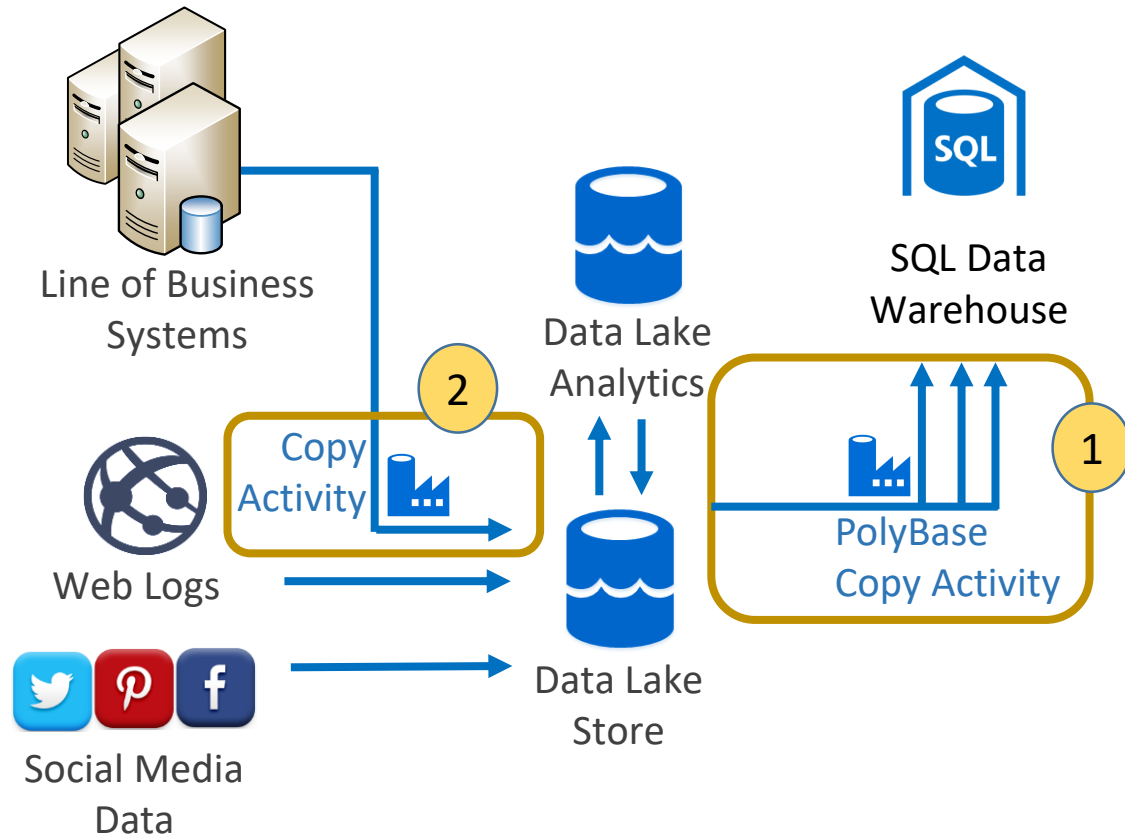
- ✓ Certain Azure services only support cloud data sources

### Data is in the Cloud

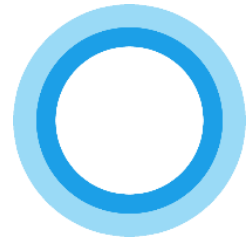
- ✓ Source and/or destination is in the cloud
- ✓ Comfortable with scripting



## Building Blocks Hybrid Data Integration



- 1 PolyBase is executed within a Data Factory copy activity for movement of data; Activity windows handle incremental data loads
- 2 Data Factory copy activity is used for movement of data; Activity windows handle incremental data loads



# Cortana Intelligence Suite

---

## Azure Machine Learning



Generally Available as of February 2015

## A service for building predictive analytics solutions

The screenshot displays the Microsoft Azure Machine Learning Studio interface for a project titled "Binary Classification: Direct marketing". The interface includes a left-hand navigation pane with categories such as "Saved Datasets", "Data Format Conversions", "Data Input and Output", "Data Transformation", "Feature Selection", "Machine Learning", "OpenCV Library Modules", "Python Language Modules", "R Language Modules", "Statistical Functions", "Text Analytics", "Web Service", and "Deprecated".

The main workspace shows a workflow diagram with the following modules: "Reader", "Metadata Editor", "Project Columns remove columns that are part of the label", "Split", "Two-Class Boosted Decision T...", "Two-Class Support Vector Ma...", "Split", "Sweep Parameters", "Score Model", and "Evaluate Model".

Three blue callout boxes highlight specific features:

- Data pre-processing modules:** Points to the "Data Transformation" category in the left navigation pane.
- Execution of algorithms:** Points to the "Two-Class Boosted Decision T..." module in the workflow.
- Custom R or Python:** Points to the "R Language Modules" and "Python Language Modules" categories in the left navigation pane.

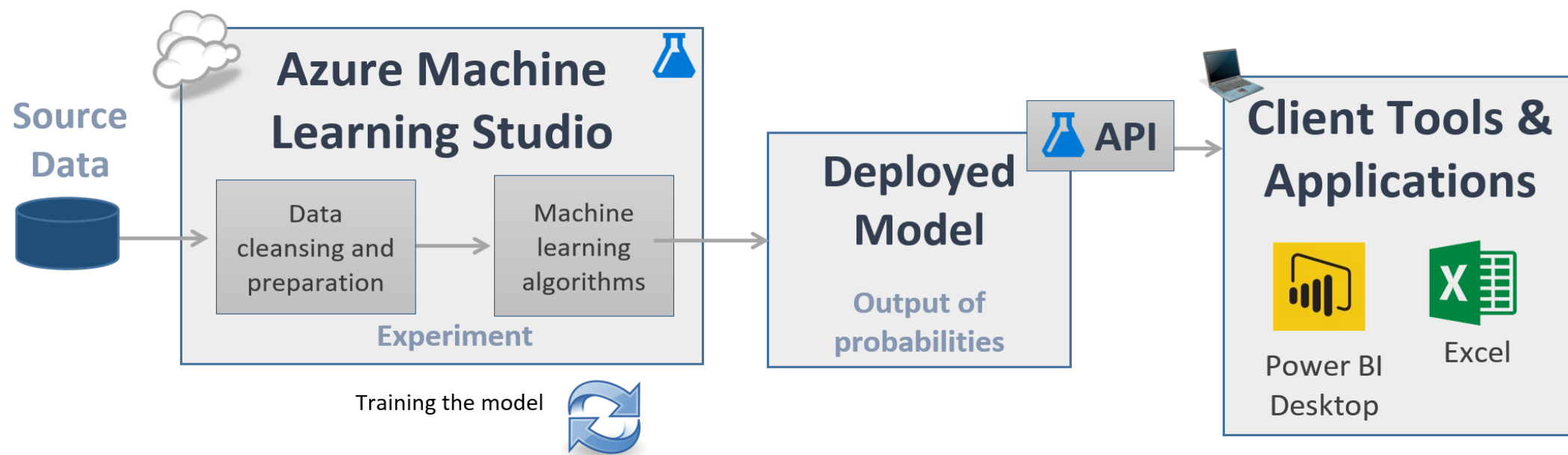
The right-hand pane shows the "Properties" for the "Two-Class Boosted Decision Tree" module, including settings for "Create trainer mode" (Single Parameter), "Maximum number of leav..." (20), "Minimum number of sam..." (10), "Learning rate" (0.2), "Number of trees construct..." (100), "Random number seed" (0), and a checked option for "Allow unknown categ...".

At the bottom right, there is a "Quick Help" section with the text: "Creates a binary classifier using a boosted decision tree algorithm (more help...)".

## Purpose

### Predictive Analytics

- ✓ Build predictive models using statistical techniques
- ✓ Learn from existing data to forecast future behaviors, outcomes & trends
- ✓ Minimize learning curve with predefined algorithms and drag & drop authoring environment
- ✓ Extensible with R and Python



## Common Use Cases

### Finding Anomalies

- ✓ Examining patterns for detection of fraud
- ✓ Locating unusual or abnormal equipment readings



### Descriptive Analytics

- ✓ Analysis of returns
- ✓ Customer segmentation (ex: by buying habits or age group) to improve customer service
- ✓ Personalized offer recommendations

### Predictive Analytics

- ✓ Credit risk
- ✓ Product demand & revenue predictions
- ✓ Customer retention
- ✓ Weather predictions
- ✓ Machine maintenance & smart buildings
- ✓ Hospital readmissions
- ✓ Student dropouts

## Common Use Cases

### Getting Started with Machine Learning

- ✓ Lowers the learning curve
- ✓ Extensible with custom R or Python



### Operationalizing a Solution

- ✓ Deploy as a web service

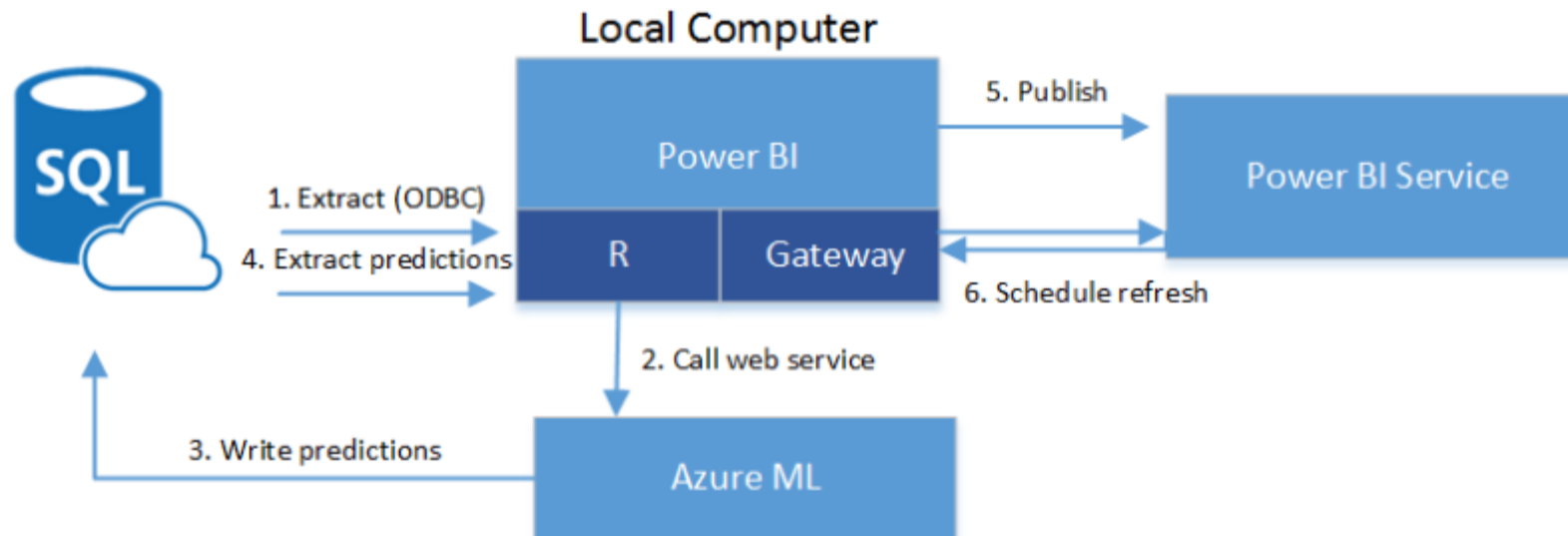
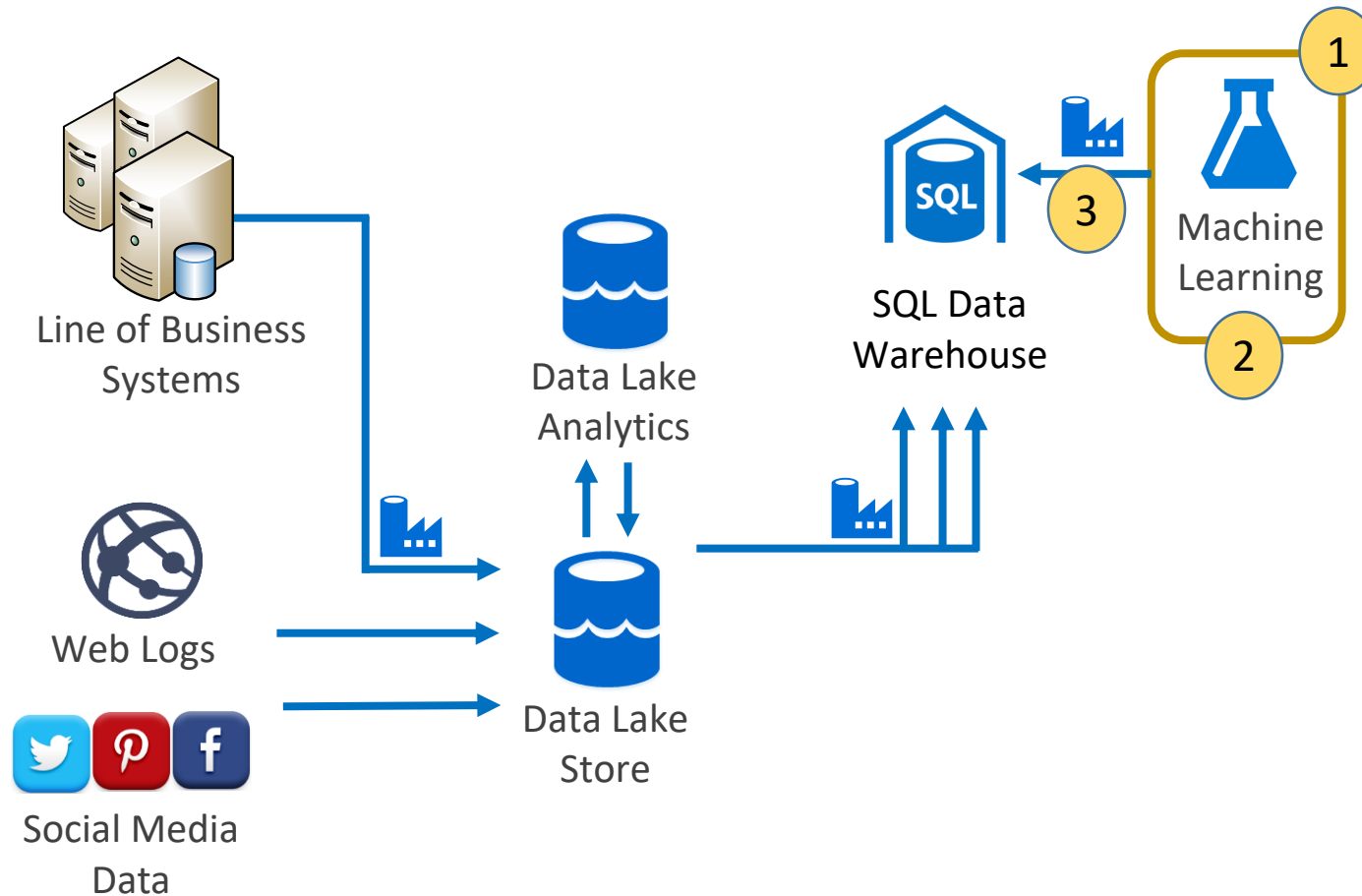


Image from: <https://powerbi.microsoft.com/en-us/blog/power-bi-azure-ml/>

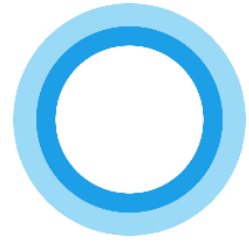


## Building Blocks

### Operationalizing an ML Model



- 1 Tested ML model is published as a web service
- 2 Execution of the ML model is invoked by calling the web service
- 3 Integrate scored results to the data warehouse for further analysis, using Data Factory



# Cortana Intelligence Suite

---

## Power BI

Generally Available as of July 2015

# Data analysis and visualization tools

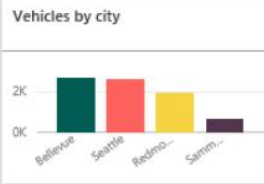
Ask a question about the data on this dashboard

[How to ask](#)



Vehicles in operation  
**8065**

Vehicles requiring maintenance  
**892**



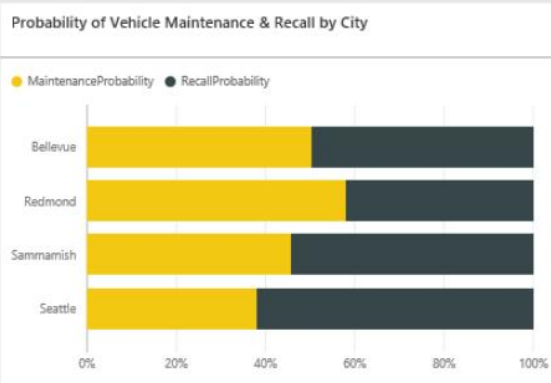
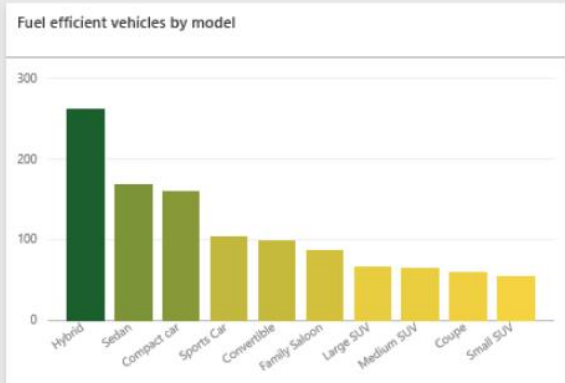
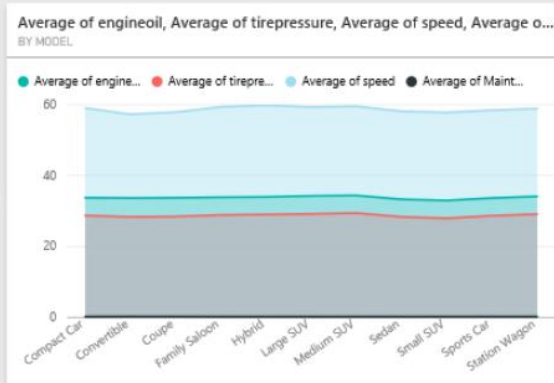
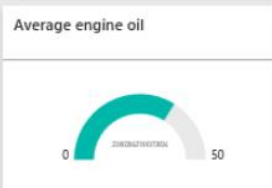
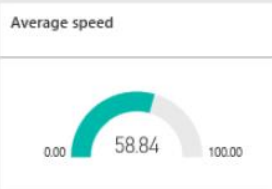
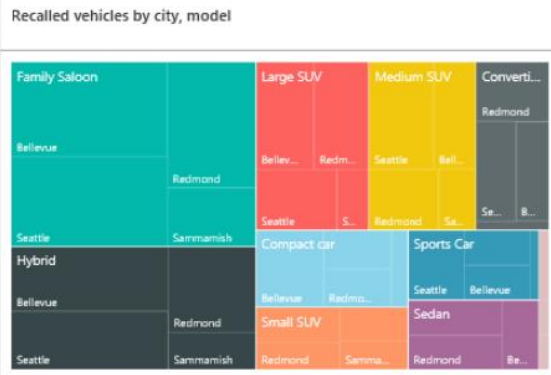
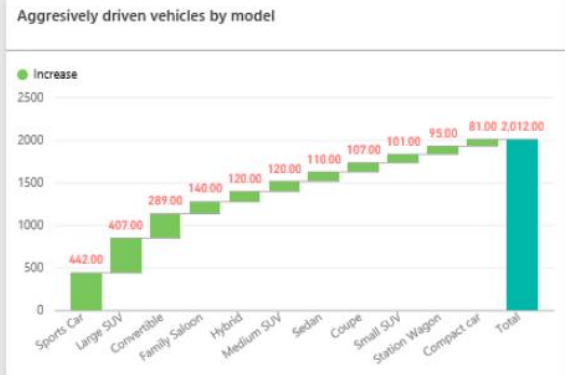
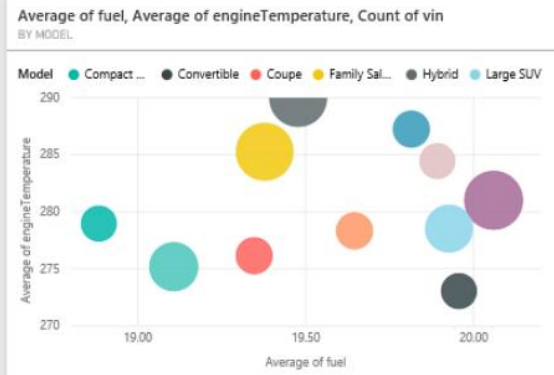
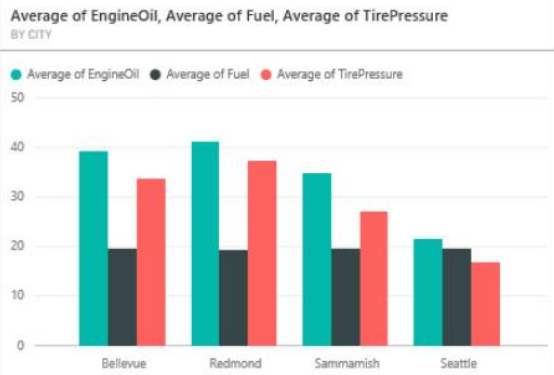
Aggressively driven vehicles  
**2012**

Fuel efficient vehicles  
**1133**

Recalled vehicles  
**505**

Recalled vehicles by city

city	Count of vin
Bellevue	159
Redmond	134
Sammamish	53
Seattle	159
<b>Total</b>	<b>505</b>

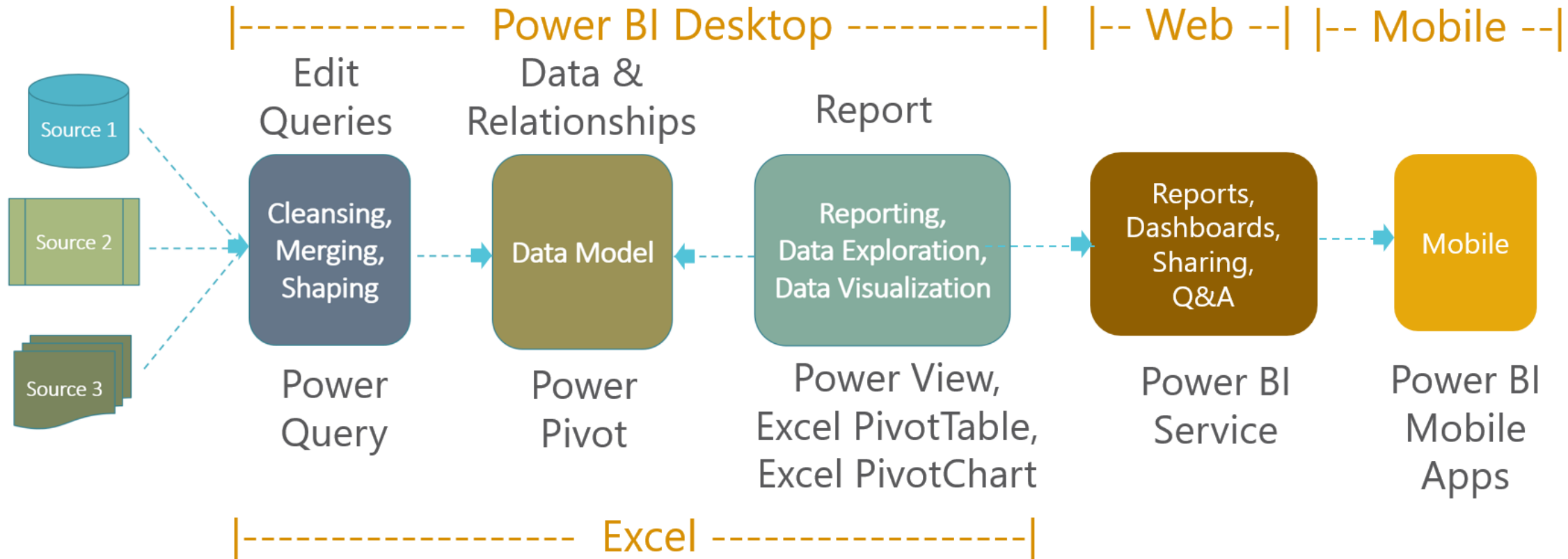


Power BI

## Purpose

Data Preparation, Data Modeling, Data Visualization

✓ Set of desktop, web, and mobile tools



## Common Use Cases

### Self-Service BI

- ✓ Mashup of data into a small data model which is imported & refreshed in Power BI Service
- ✓ Data preparation & cleansing
- ✓ Data visualizations



### Front-End Reporting

- ✓ Reports and dashboards from Corporate BI sources via live queries

### Prototyping

- ✓ Test out ideas for data structure, calculations, and reports

### Third Party Reporting

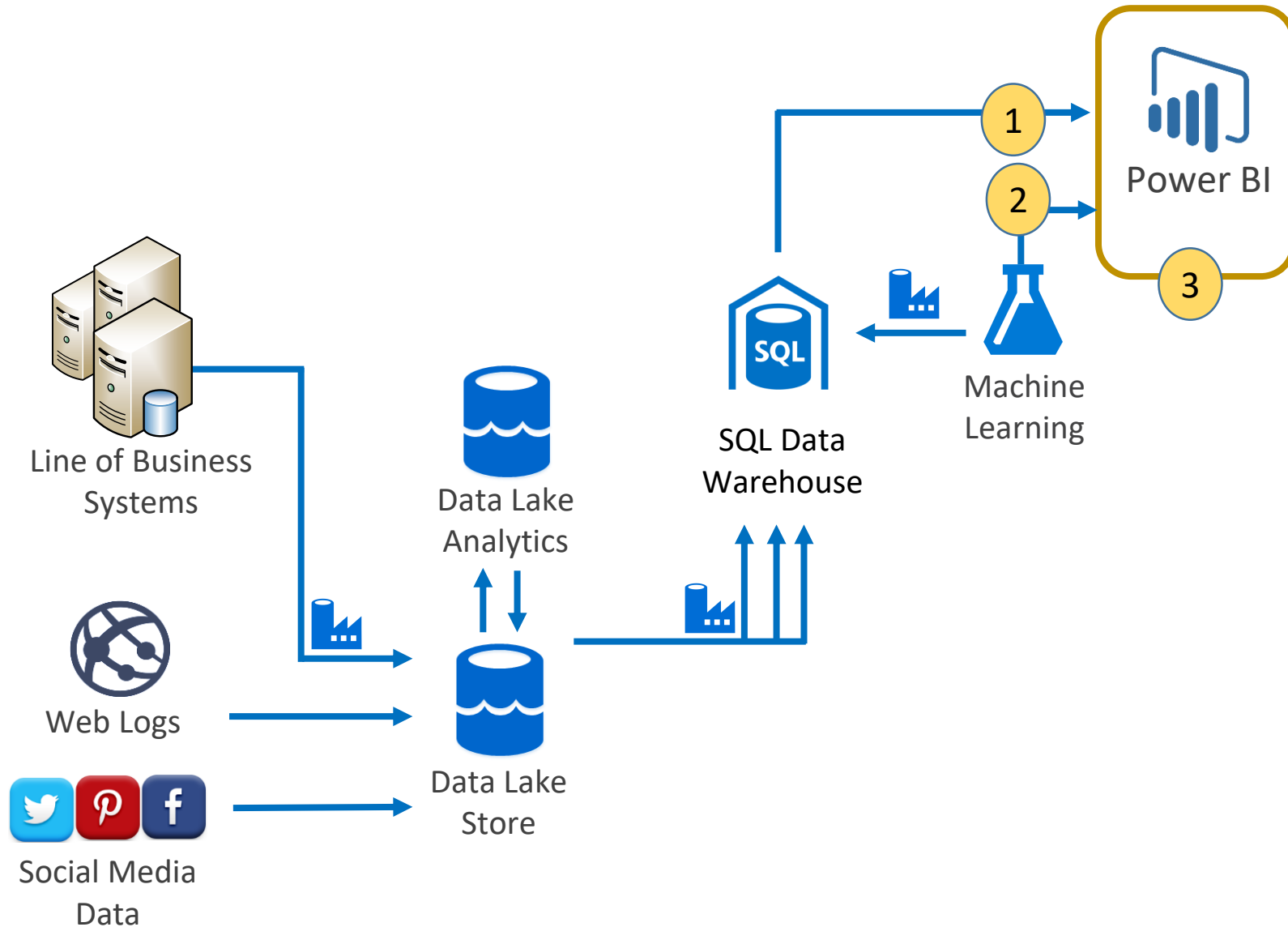
- ✓ 3<sup>rd</sup> party content packs quick start for isolated reporting scenarios

### Embed in Custom App

- ✓ Power BI Embedded (separate Azure Service)

## Building Blocks

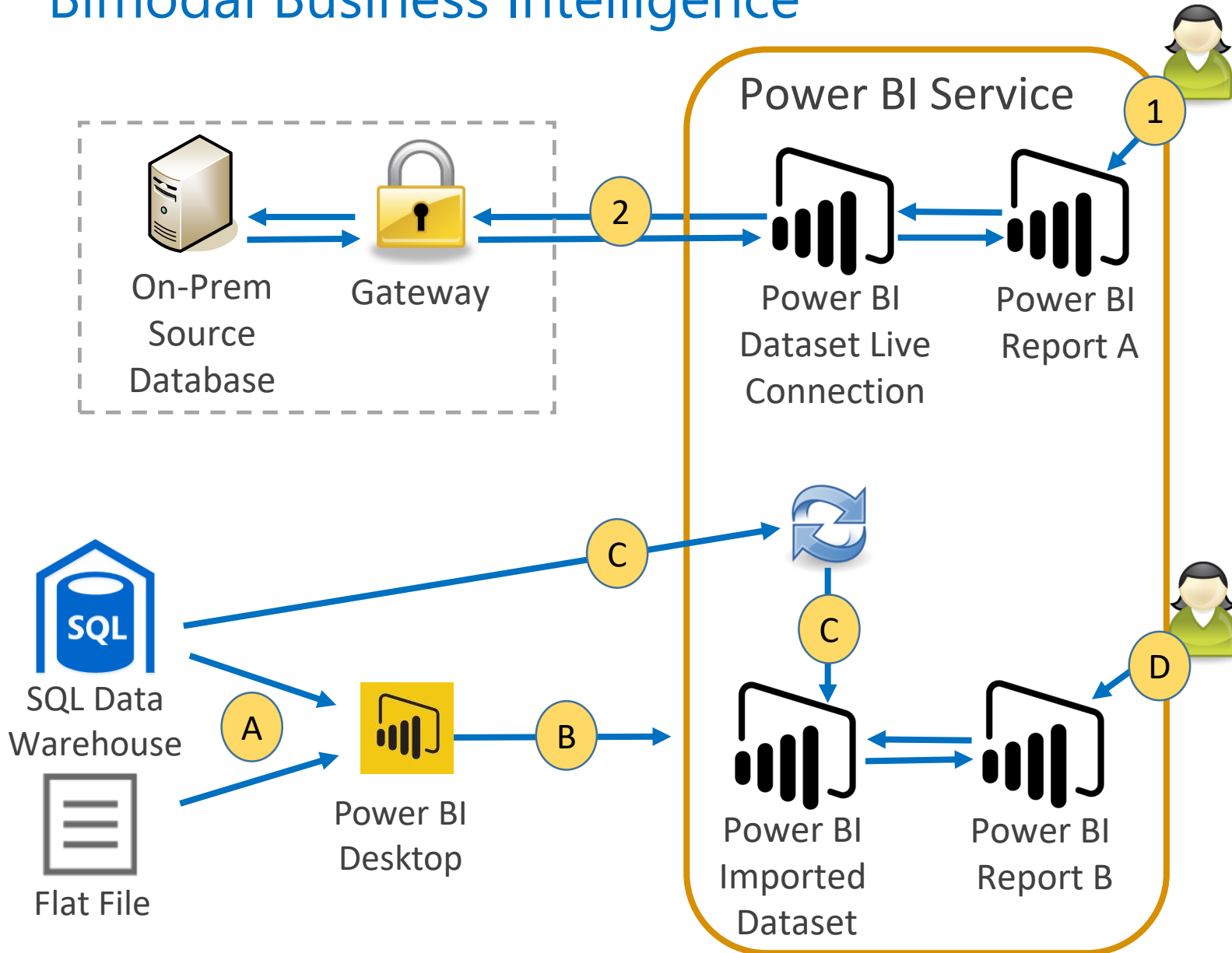
### Self-Service and Corporate BI Scenarios



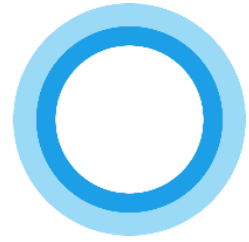
- 1 Connect Power BI to the data warehouse in DirectQuery mode to facilitate corporate reporting scenarios
- 2 ML scored results can be returned directly to Power BI
- 3 Utilize Power BI to facilitate self-service BI  
*See next page*

# Building Blocks

## Bimodal Business Intelligence



- 1 User runs Report A
  - 2 Data is requested via gateway to present in report
- 
- A Data mashup prepared
  - B Data model and reports published
  - C Data refresh schedule created for imported dataset
  - D User runs Report B & results are returned from the imported dataset



# Cortana Intelligence Suite

---

## Azure Data Catalog



Generally Available as of April 2016



# Azure Data Catalog

## Register, manage, search, and explore organizational data sources

The screenshot displays the Microsoft Azure Data Catalog interface. At the top, there is a navigation bar with icons for Home, Discover, Glossary, Publish, Settings, and User. Below this is a search bar and a 'Results Per Page' dropdown set to 10. The main content area shows search results for a 'Customer' table. A blue callout box points to a 'Data Preview (first x rows)' section. Another blue callout box points to the 'Data Profile' section, which includes a 'Table Profile' table and a 'Column Profile' table. A third blue callout box points to the search filters on the left side of the interface. The right side of the interface shows the 'Properties' panel for the selected table, including fields for Name, Friendly Name, Description, and Experts.

Microsoft Azure Data Catalog

Home Discover Glossary Publish Settings User

Search

Results Per Page: 10 Highlight Open In ... Delete

Searches

Current Search

Save Clear All

User Tags PII

Saved Searches (1)

Filters

User Tags

- Samples (14)
- Sales (2)
- Sensors (2)
- Weather (2)

see more

Object Type

- Table (1)

Source Type

- SQL Server (1)

Experts

- 

datacataloguser@sqlchi

Customer

click tile to add a description...

Experts:

Data Preview (first x rows)

Preview Columns **Data Profile** Documentation

Table Profile

Number Of Rows	Size	Last Data Update	Last Schema Update
847	856 KB		4/14/2016

Column Profile

Column Name	Data Type	Null Count	# Distinct Values	Minimum	Maximum
CustomerID	int	0	847	1	30118
		0	1		0
		7	4	Mr.	Sra.
		0			Yvonne

Search by: Tag, Object Type, Source Type, Expert Name

Data Profiling Info

Properties

Preview Columns Data Profile Docs

Name: Customer

Friendly Name: add friendly name...

Description: add your description...

Experts: DataCatalogUser@...com Add...

# Azure Data Catalog

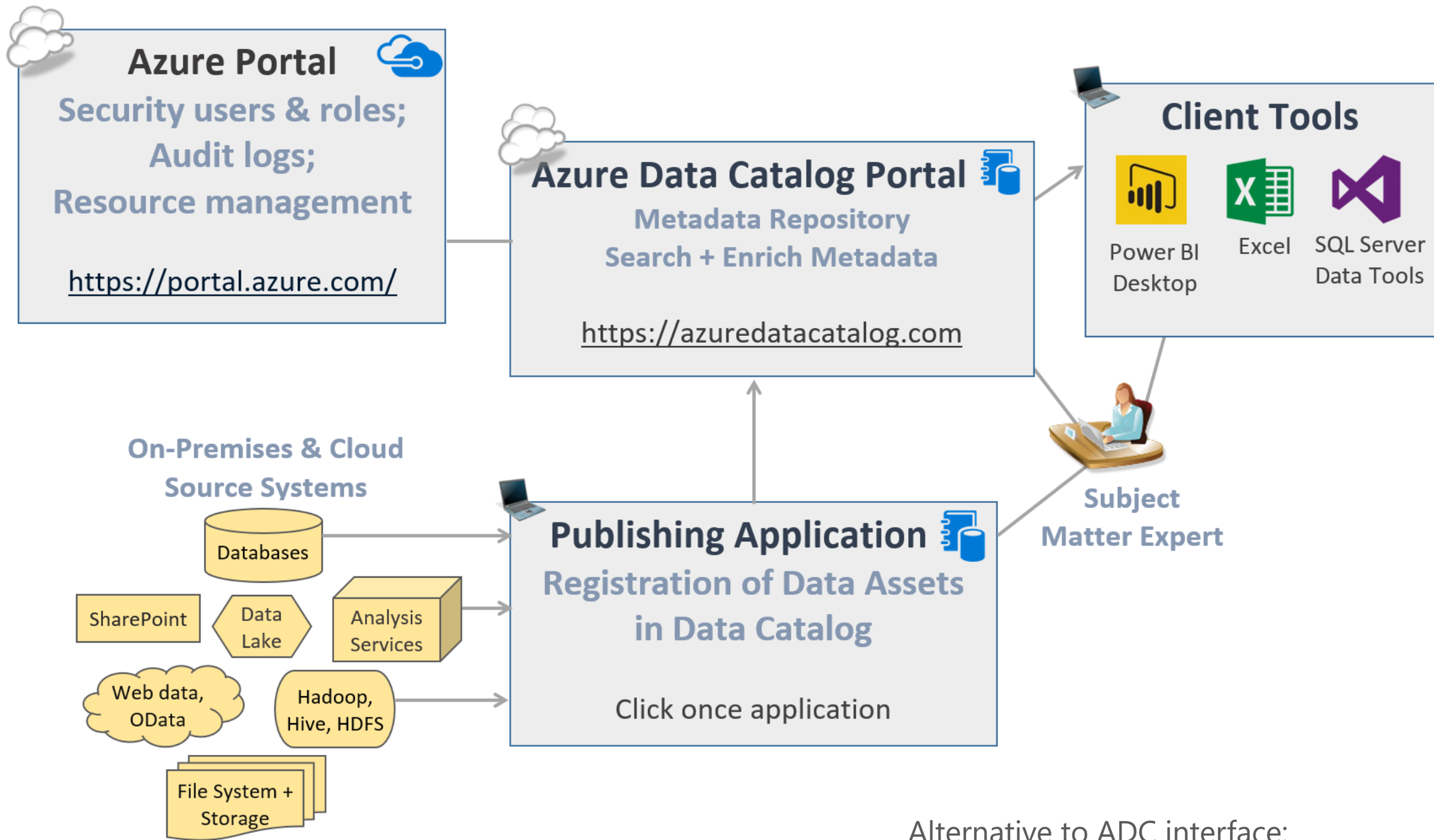
## Register, manage, search, and explore organizational data sources

The screenshot displays the Microsoft Azure Data Catalog interface. At the top, the navigation bar includes 'Home', 'Discover', 'Glossary', 'Publish', 'Settings', and 'User'. A search bar is present with a search icon, a grid icon, and a menu icon. Below the search bar, the interface shows '1 search results, 1 selected' and a 'Select All' checkbox. The main content area displays a search result for a 'Customer' table. A callout box points to the 'Description' column in the table, stating 'Descriptions for each column'. Another callout box points to the 'Tags' column, stating 'Tags for each column'. A third callout box points to the 'Connection Info' section on the right, stating 'How to connect & how to request access'. The 'Customer' table has the following columns:

Column Name	Data Type	Tags	Description
SalesPerson	nvarchar	<a href="#">Add...</a>	Current, primary, salesperson assigned to customer
EmailAddress	nvarchar	<a href="#">Add...</a> PII	Customer's preferred email - should be excluded from most data exports
Phone	nvarchar	<a href="#">Add...</a> PII	Customer's preferred # - should be excluded from most data exports
Password	nvarchar	<a href="#">Add...</a>	

The right-hand side of the interface shows the 'Properties' section for the 'Customer' table. It includes fields for 'Name' (Customer), 'Friendly Name' (add friendly name...), 'Description' (add your description...), 'Experts' (DataCatalogUser@...), 'Tags' (Samples), and 'Connection Info' (Server Name: sampledatabases.database.windows.net, Database Name: AdventureWorksLT, Schema Name: SalesLT, Object Name: Customer, Request Access: Submit form B-4 for access to Customer data. Contact John Brown at x-7482 with questions.).

# Azure Data Catalog



Alternative to ADC interface:  
APIs in conjunction with custom portal

## Purpose

### Data Documentation & Discovery

- ✓ Enterprise-wide metadata catalog for data assets
- ✓ Simplified data source discovery via search
- ✓ Enrich & understand assets with tags & annotations
- ✓ Collaboration between data producers & data consumers

### Important to Know

- ✓ One Data Catalog per organization  
*(not one per subscription)*
- ✓ Authentication only accepts an organizational account  
*(cannot use a Microsoft account)*



## Common Use Cases



### Documentation for Centralized Data Sources

- ✓ Line of business systems
- ✓ Data warehouse, marts
- ✓ Analytic systems
- ✓ Reporting Services
- ✓ File system
- ✓ Data lake

### Facilitate Self-Service BI

- ✓ Data dictionary
- ✓ Assist combining data from multiple sources
- ✓ Reduce duplication of effort

### Capture 'Tribal Knowledge'

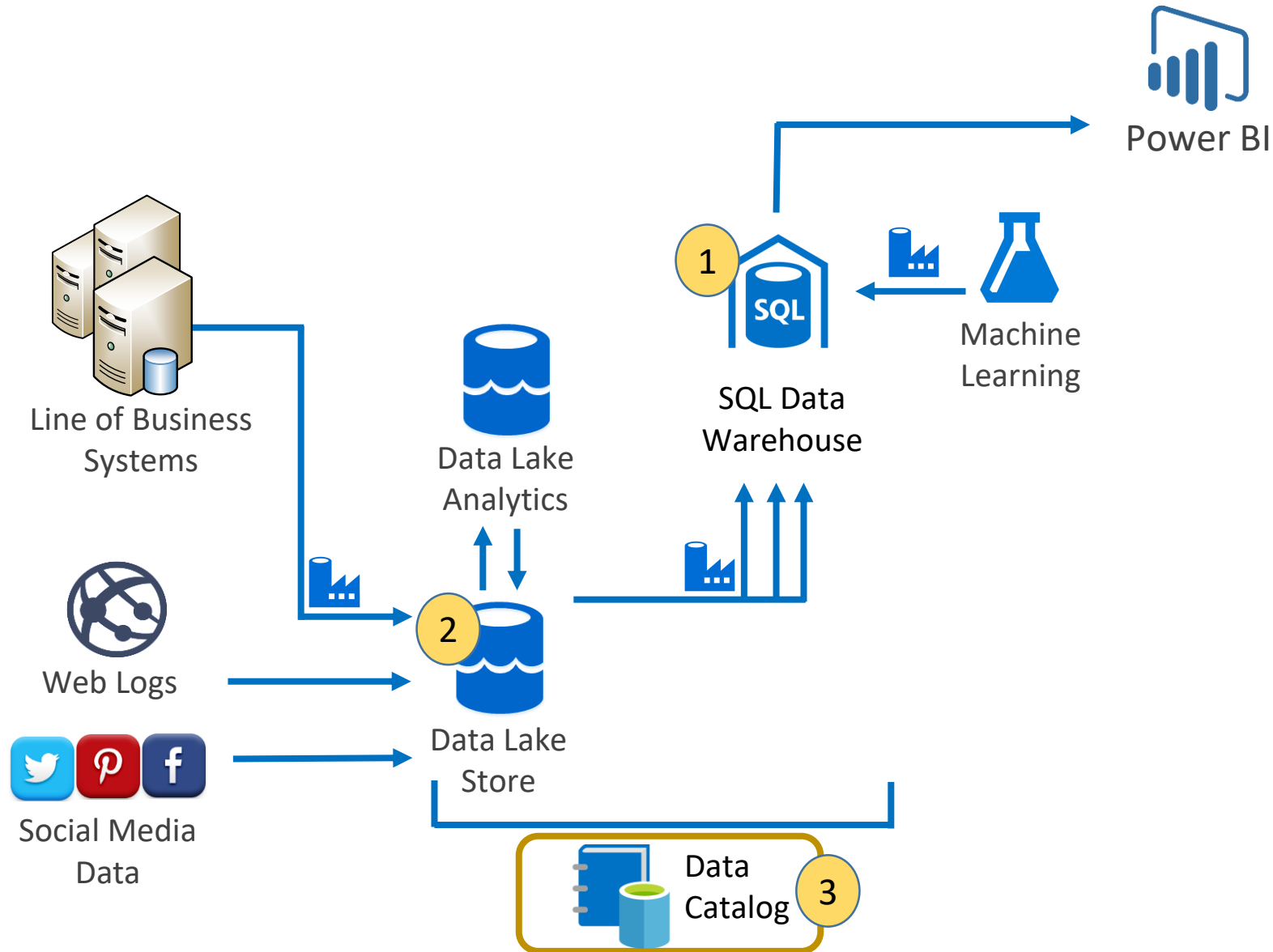
- ✓ Documentation about the data is maintained by subject matter experts
- ✓ Enhance understanding

### Data Discovery & Provisioning

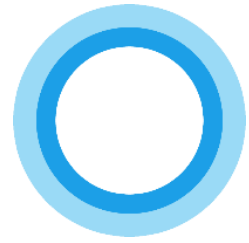
- ✓ Users search to discover data assets
- ✓ Who to contact to request access

## Building Blocks

Data discovery



- 1 Register data from the data warehouse
- 2 Selectively register curated data from the data lake store
- 3 View metadata; view data from a table or view using Power BI or Excel



# Cortana Intelligence Suite

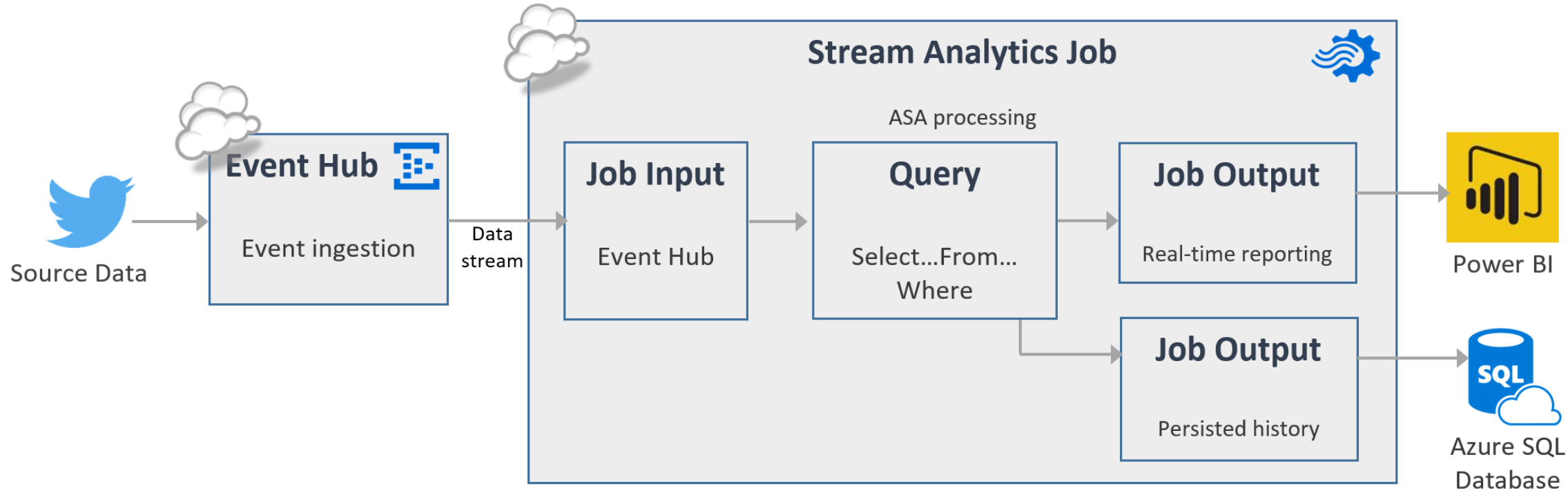
---

## Azure Stream Analytics + Event Hub



Generally Available as of April 2015

## Real-time analytics for high velocity streaming events





## Purpose

### Stream Analytics

- ✓ Analytic processing engine for streaming events
- ✓ Internet of Things (IoT) solutions for data in motion from devices, sensors, social media, etc

### Event Hub

- ✓ A publish-subscribe service that handles high volume & high velocity data streams
- ✓ Allows events to be ingested into Azure from many platforms & devices (another option: IoT Hub)
- ✓ The preferred method of event ingestion for Stream Analytics

### Simplification

- ✓ Alternative to batch loading processes
- ✓ Lower bar to entry for developers by using SQL-like language
- ✓ More straightforward than an HDInsight Storm cluster



## Common Use Cases

### Device Monitoring & Telemetry

- ✓ Level beyond acceptable threshold
- ✓ Business continuity

### Traffic

- ✓ Accident & traffic conditions

### Web Logs

- ✓ Clickstream analysis
- ✓ A/B testing
- ✓ Errors or degraded experience

### Demand-Based Pricing

- ✓ Bookings in past x minutes

### Identity Protection

- ✓ Real-time fraud alerts
- ✓ Identity theft scenarios

### Social Media

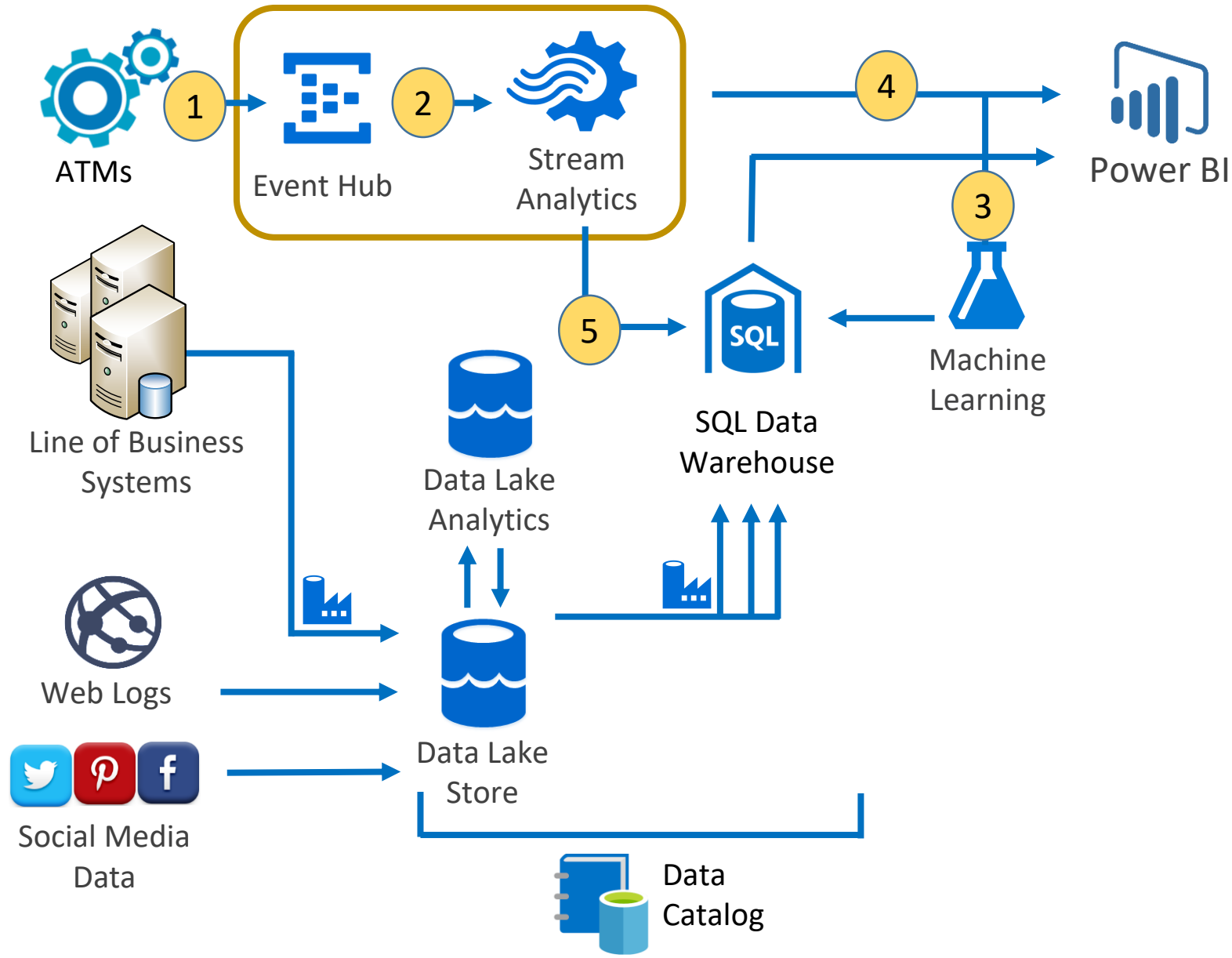
- ✓ Real-time sentiment analysis

### Inventory Levels

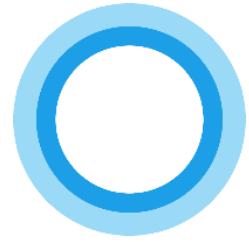
- ✓ Shelf volume vs. register checkouts



## Building Blocks Fraud Alerts for Unusual Activity



- 1 Events ingested to the raw data event queue
- 2 Consume & process data for a window of time
- 3 Invoke ML for fraud predictions
- 4 Real-time reporting and/or alerts
- 5 Persist data for historical reporting & analysis



# Cortana Intelligence Suite

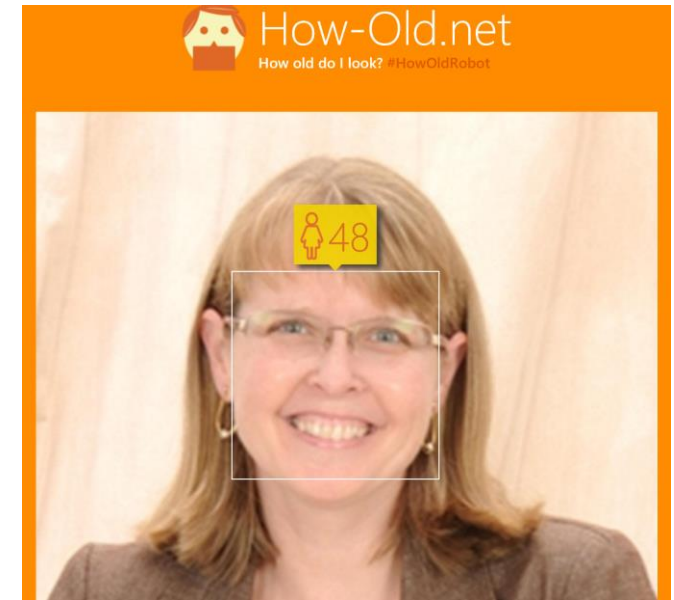
---






## Cognitive Services

General Availability Dates Vary Per API

# Cognitive Services

Set of APIs, SDKs, and cloud services to build intelligent systems



 Vision	 Speech	 Language	 Knowledge	 Search
Computer Vision	Speaker Recognition	Text Analytics	Academic Knowledge	Bing Search API
Face	Speech	Bing Speller	Entity Linking Service	Bing Image Search API
Emotion	CRIS	Web Language Model	Knowledge Exploration Service	Bing Video Search API
Video		Linguistic Analysis	Recommendations	Bing News Search API
		Language Understanding Intelligent Service		Bing Auto Suggest API

## Purpose

### Give Applications a "Human Side"

- ✓ Designed to make applications more personalized, intelligent, and engaging
- ✓ Set of APIs, SDKs, and cloud services to see, hear, and interpret
  - ✓ Vision
  - ✓ Speech
  - ✓ Language
  - ✓ Emotion
  - ✓ Face Recognition
  - ✓ etc...
- ✓ Some APIs supported by Azure Machine Learning or Bing services
- ✓ Integration with U-SQL in Azure Data Lake Analytics



## Common Use Cases

### Sentiment Analysis

- ✓ Detect key phrases & topics being discussed
- ✓ Identify feedback posted to website
- ✓ Convert speech to text
- ✓ Emotion recognition



### Security Systems

- ✓ Facial detection and recognition
- ✓ Video intelligence

### Search & Auto-Suggestions

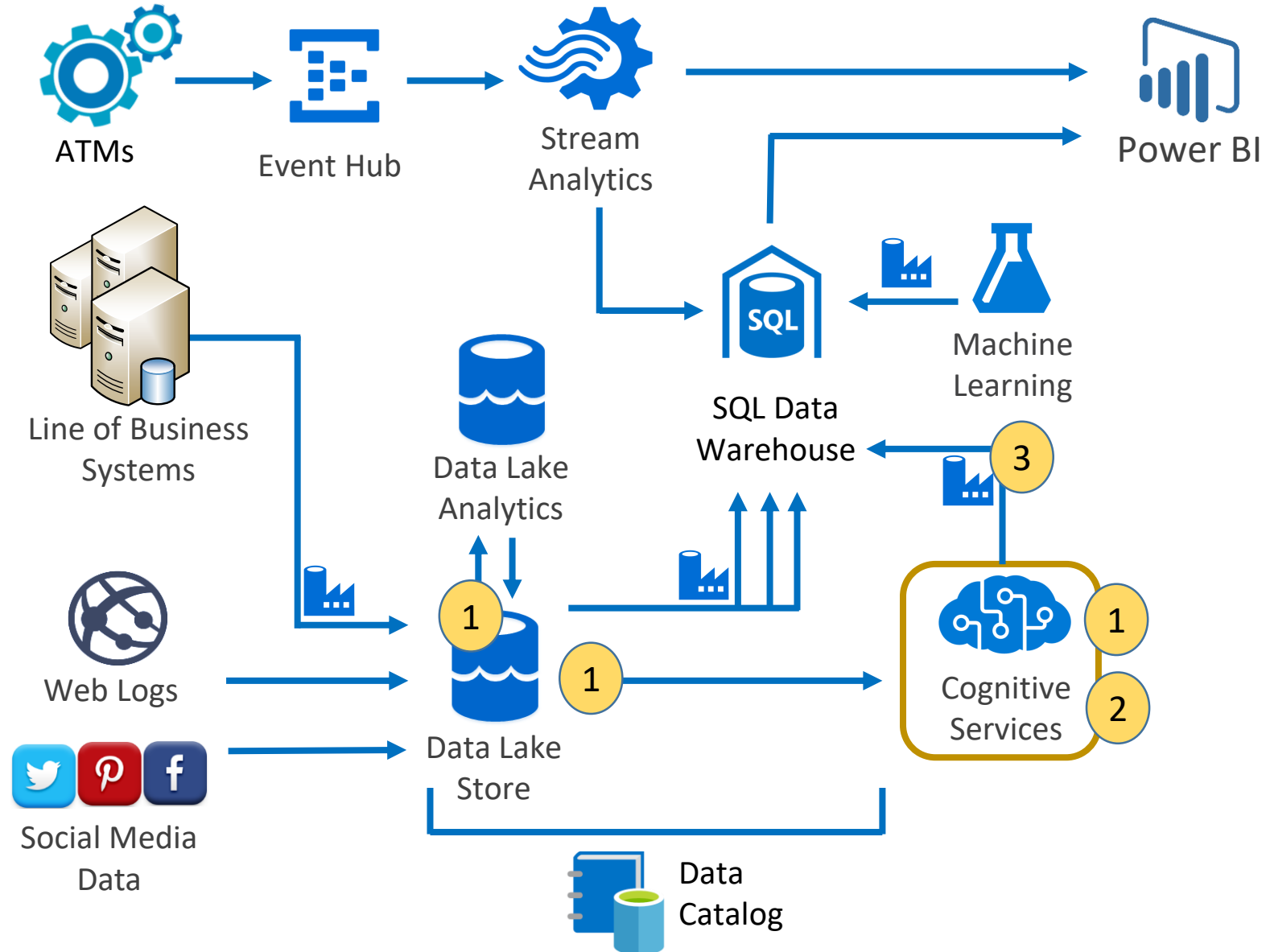
- ✓ Completion of partial search queries

### Personalized Shopping Experience

- ✓ Recommend items likely to be purchased together

## Building Blocks

### Text analytics & Sentiment analysis

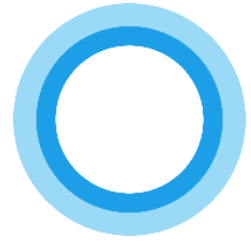


1 Perform text analytics on customer e-mail records using cognitive capabilities integrated in U-SQL

2 Perform sentiment analysis on customer phone records

3 Integrate results in the data warehouse for further analysis





# Cortana Intelligence Suite

---

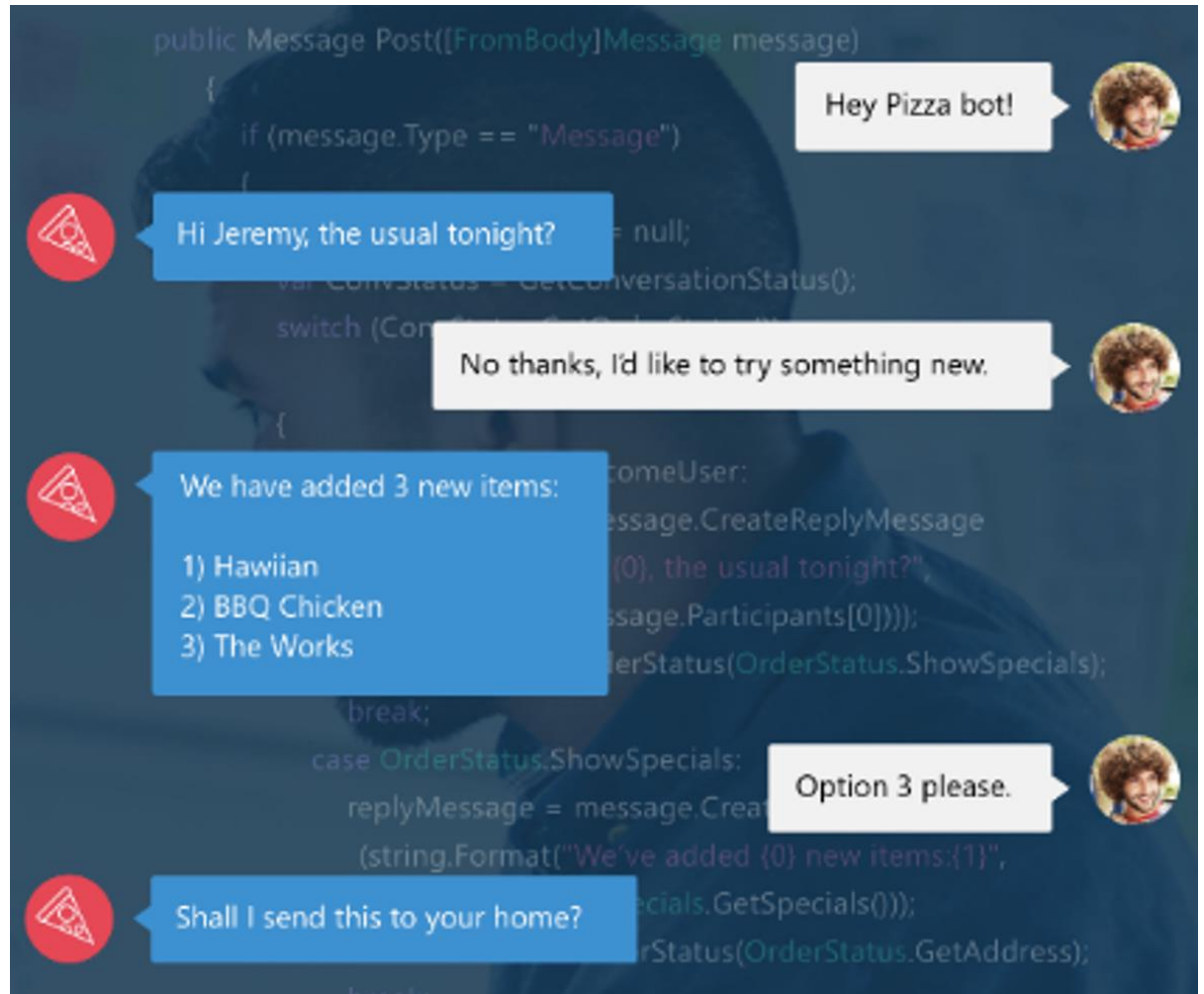
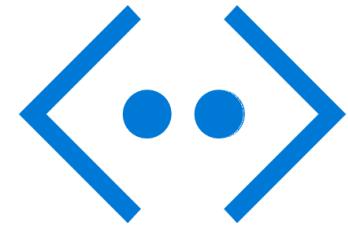
## Bot Framework

In Public Preview

## Purpose

### Conversation Agent

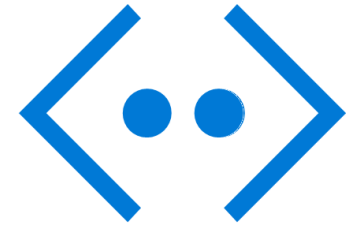
- ✓ Automated, yet contextual & natural, interactions with users
- ✓ Enable applications and services to have a conversational user interface (CUI)



## Common Use Cases

### Messaging Bots

- ✓ Web chats
- ✓ Text/SMS conversation



### Content Bots

- ✓ Share relevant content, such as news or weather, with you

### Watcher Bots for Alerting

- ✓ Flight is delayed
- ✓ Dental appointment reminder

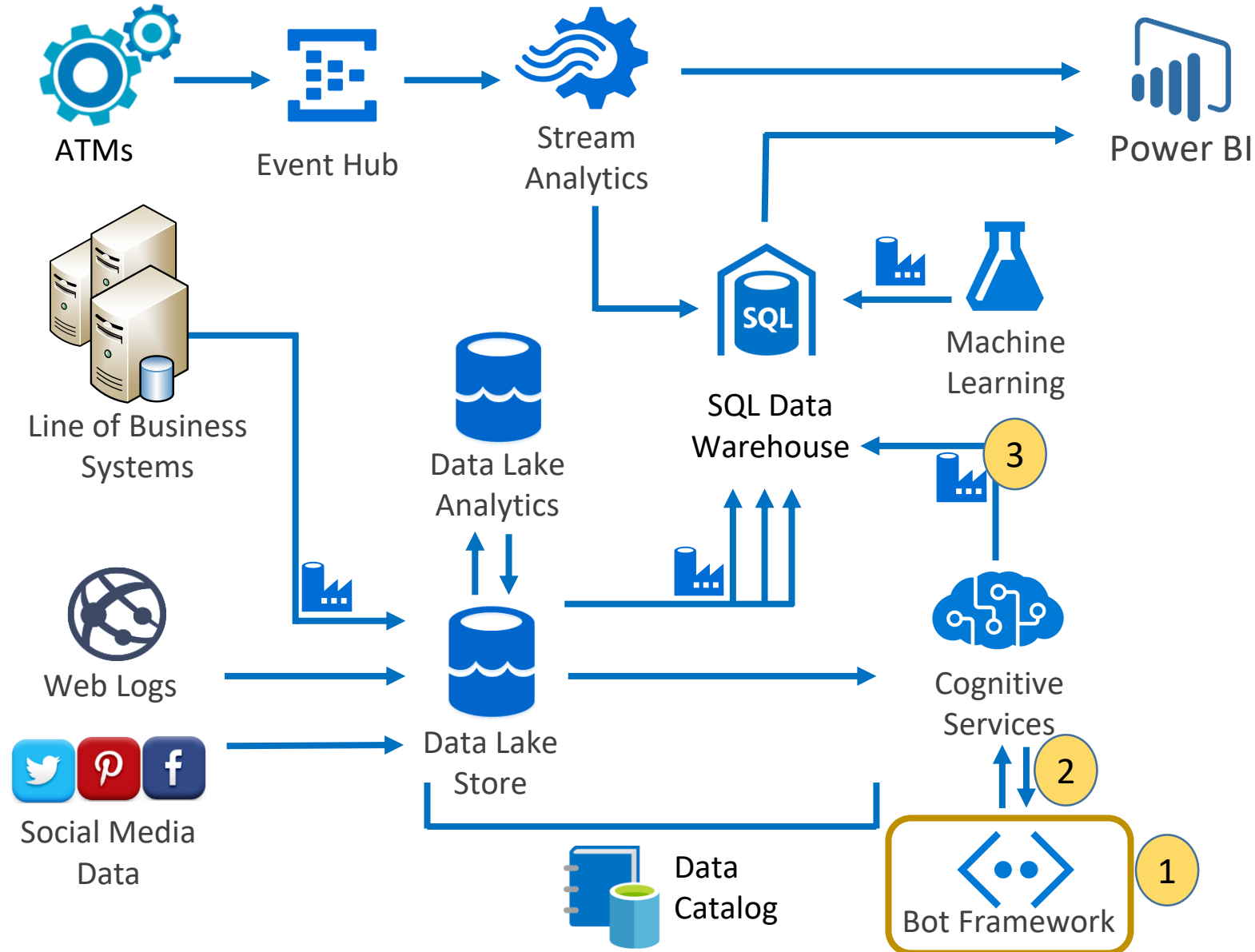
### E-Commerce Bots

- ✓ Order food
- ✓ Book a flight
- ✓ Check inventory for a product

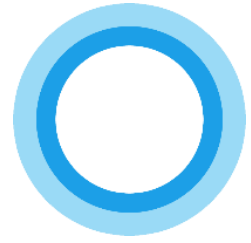
# Bot Framework

## Building Blocks

Web-based customer service



- 1 Conversation agent on website to handle common customer questions via chat
- 2 Integrate bot with Cognitive Services to enhance capabilities & perform text analytics
- 3 Integrate results in the data warehouse for further analysis



# Cortana Intelligence Suite

---

Cortana Assistant

# Why is it Called "Cortana Intelligence Suite" Anyway?

## Halo

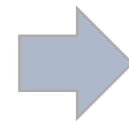
Cortana is a fictional synthetic intelligence character in the Halo novel & video game series. Cortana is "smart AI" which can learn and adapt.



## Windows Digital Assistant

Cortana was the inspiration for the Windows Digital Assistant.

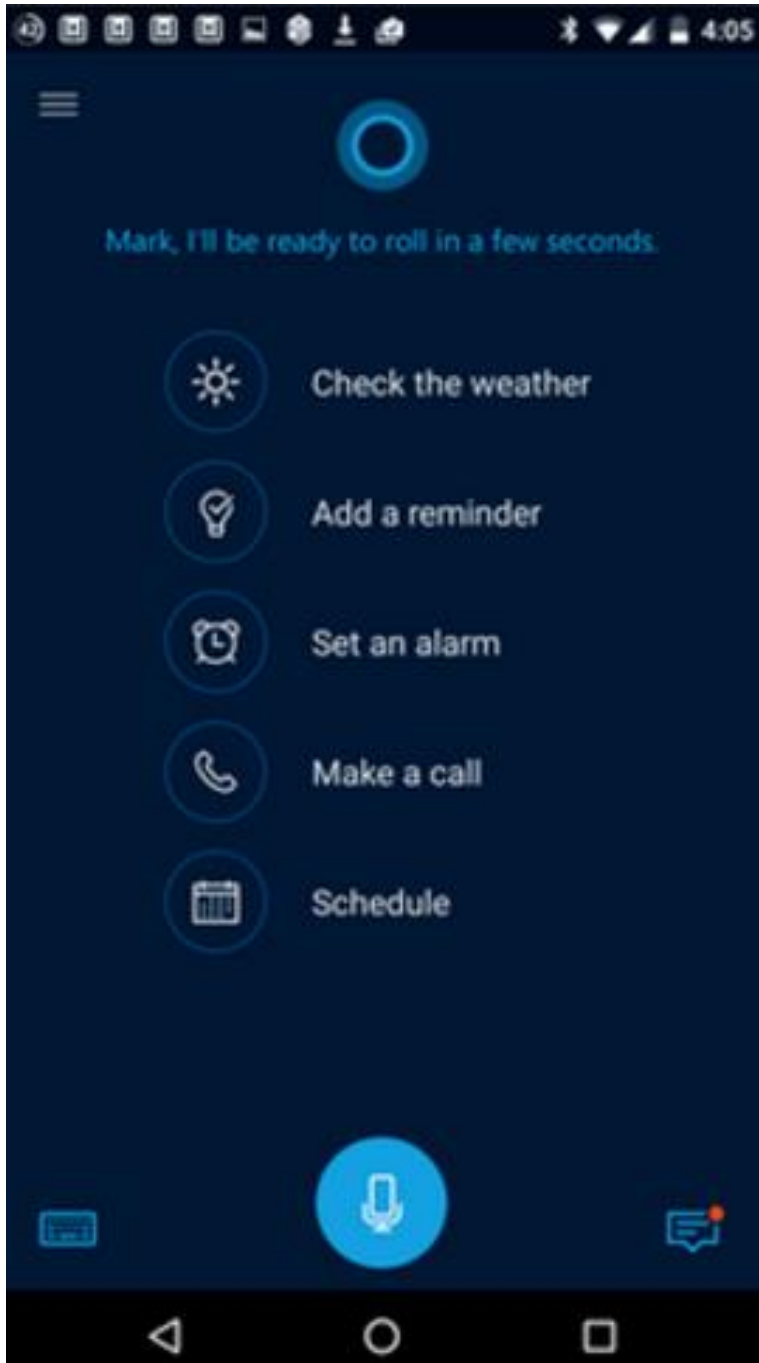
Initially Cortana was a codename, but got such a strong response that Microsoft kept the name.



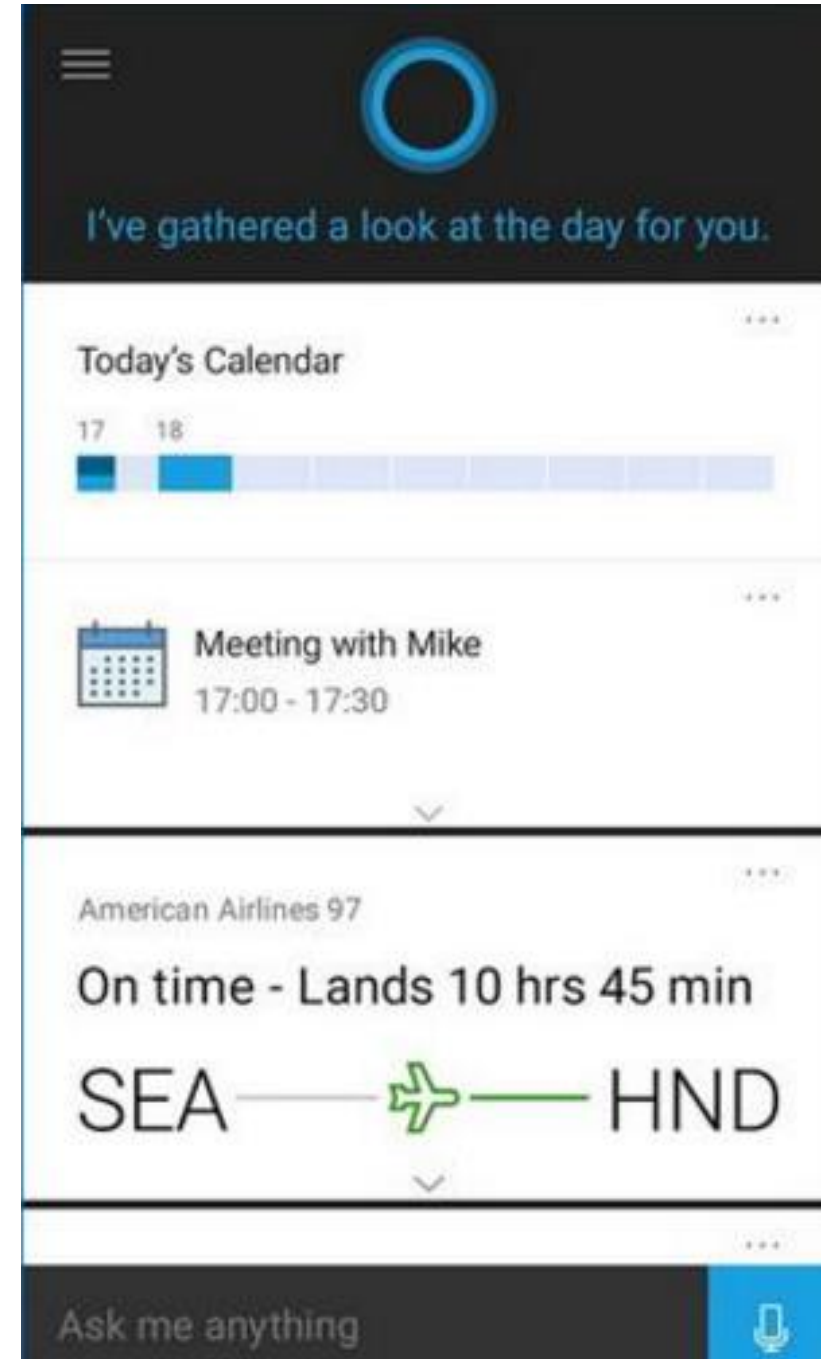
## Cortana Intelligence Suite

According to Joseph Sirosh, Cortana symbolizes the contextualized intelligence they intend to achieve with the suite of tools.

# Cortana Assistant



## Personal Digital Assistant





Power BI integration not currently available

The screenshot displays a Power BI dashboard titled "Executive Metrics Dashboard" with a search bar and a search history overlay. The dashboard includes various charts and KPIs:

- Expected Revenue THIS YEAR:** \$95.99M
- Total Invoice THIS MONTH:** \$3.99M
- Discount Savings THIS MONTH:** \$4.07M
- Total Invoice BY MONTH CATEGORY:** A line chart showing trends from January to June for Direct, Indirect, Logistics, and Other categories.
- Total Invoice BY SUB CATEGORY:** A treemap chart showing breakdowns for Hardware, Indirect Goods & Services, Raw Materials, Outsourced, and Other.
- AccountCount BY INDUSTRY:** A donut chart showing industry distribution.
- Variance to Budget BY MONTH:** A bar chart showing monthly variances from January to June, with a total bar.
- Sales Pipeline BY STAGENAME:** A funnel chart showing stages from Prospecting to Closed Lost.
- Customer Size vs Revenue:** A scatter plot showing the relationship between customer size and revenue.
- Total Invoice, Budget Amount BY MONTH CATEGORY:** A bar chart comparing monthly invoice and budget amounts.
- Number of Visits, Expenditures (\$M):** A line chart showing monthly trends for visits and expenditures.
- Marketing Campaign Stats BY DATE:** A bar chart showing daily metrics for Requests, Delivered, Opens, Clicks, and Blocked.
- Marketing Campaign Metrics:** Summary statistics for Requests (9515), Blocked (1343), Bounces (7112), and Delivered (747054).

The Cortana Assistant overlay, highlighted with a red border, shows a search result for "largest accounts in California map by city". The search text reads: "I found the california accounts and their billing city which they are in." Below the text is a map of California with various cities marked and color-coded. The map includes labels for cities such as Fairfield, Sacramento, Stockton, Vallejo, San Francisco, San Jose, Salinas, Fresno, Bakersfield, Palmdale, Thousand Oaks, San Bernardino, Los Angeles, Huntington Beach, Santa Ana, Escondido, Oceanside, San Diego, and Tijuana. The Cortana interface also shows a search bar, navigation icons, and a Windows taskbar at the bottom.



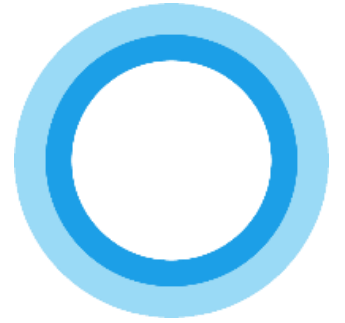
## Purpose

### Apply Language More Pervasively

- ✓ Virtual personal assistant for:
  - ✓ Asking questions
  - ✓ Finding things on PC
  - ✓ Managing calendar
  - ✓ Task completion
  - ✓ Monitoring & alerts
  - ✓ Tracking packages

### Integration

- ✓ Reminders are integrated between Windows devices
- ✓ Search within other apps, such as a calendar
- ✓ Interact with bots to make requests



## Common Use Cases

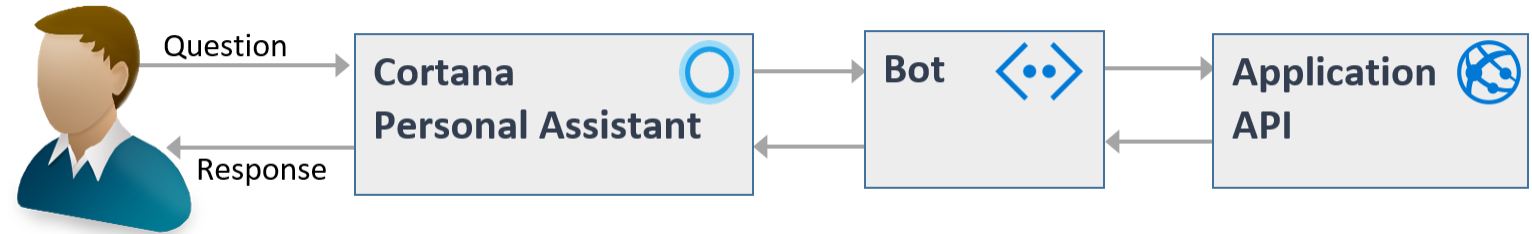
### Reminders

- ✓ Help with remembering appointments



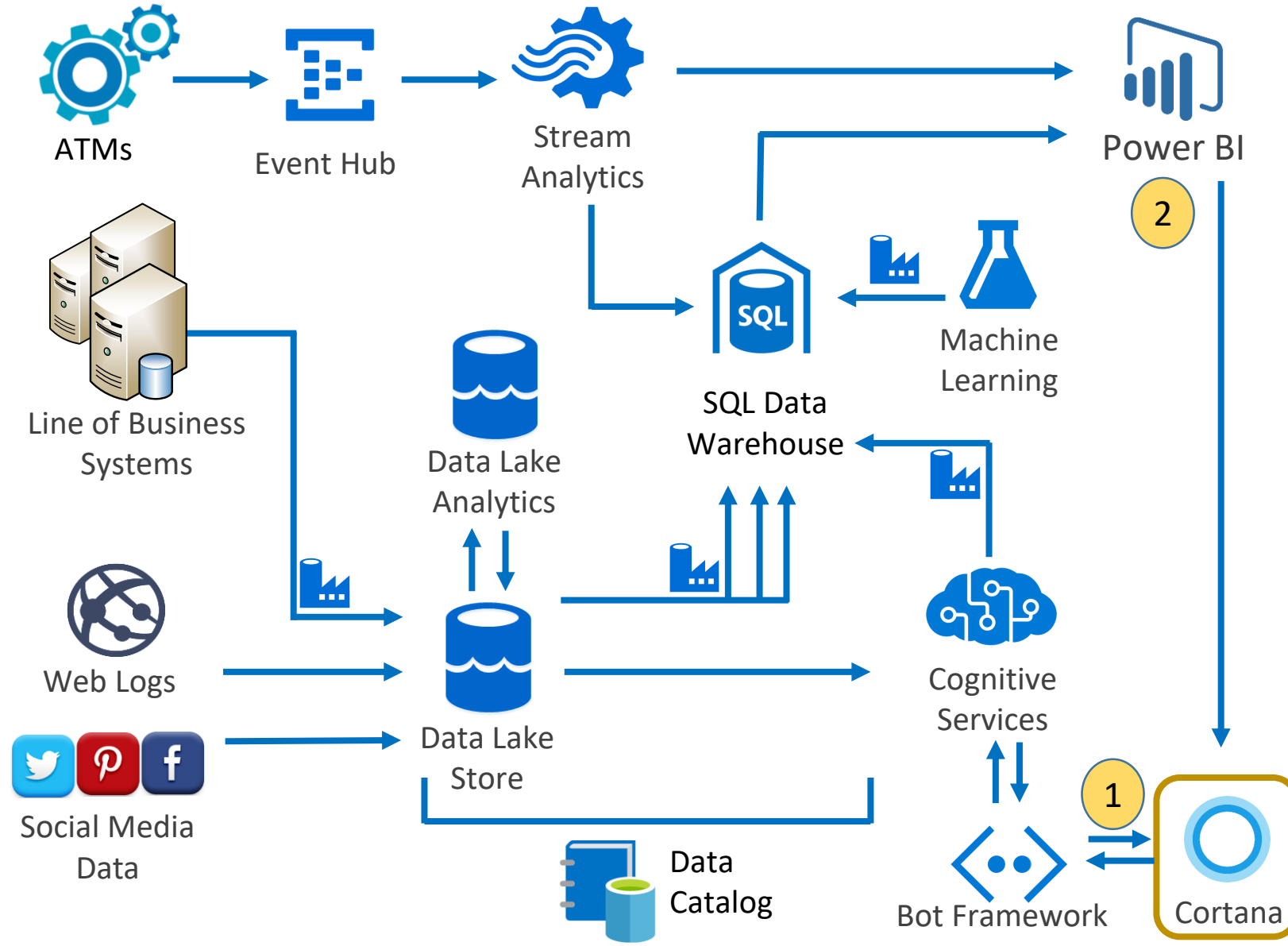
### Work With Bots

- ✓ Interact with a bot to place an order or schedule a meeting

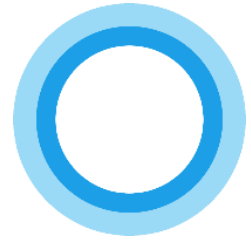


## Building Blocks

Voice-based customer assistance



- 1 Cortana for voice-based customer assistance, integrated with the Bot Framework
- 2 Render a Power BI report using Cortana

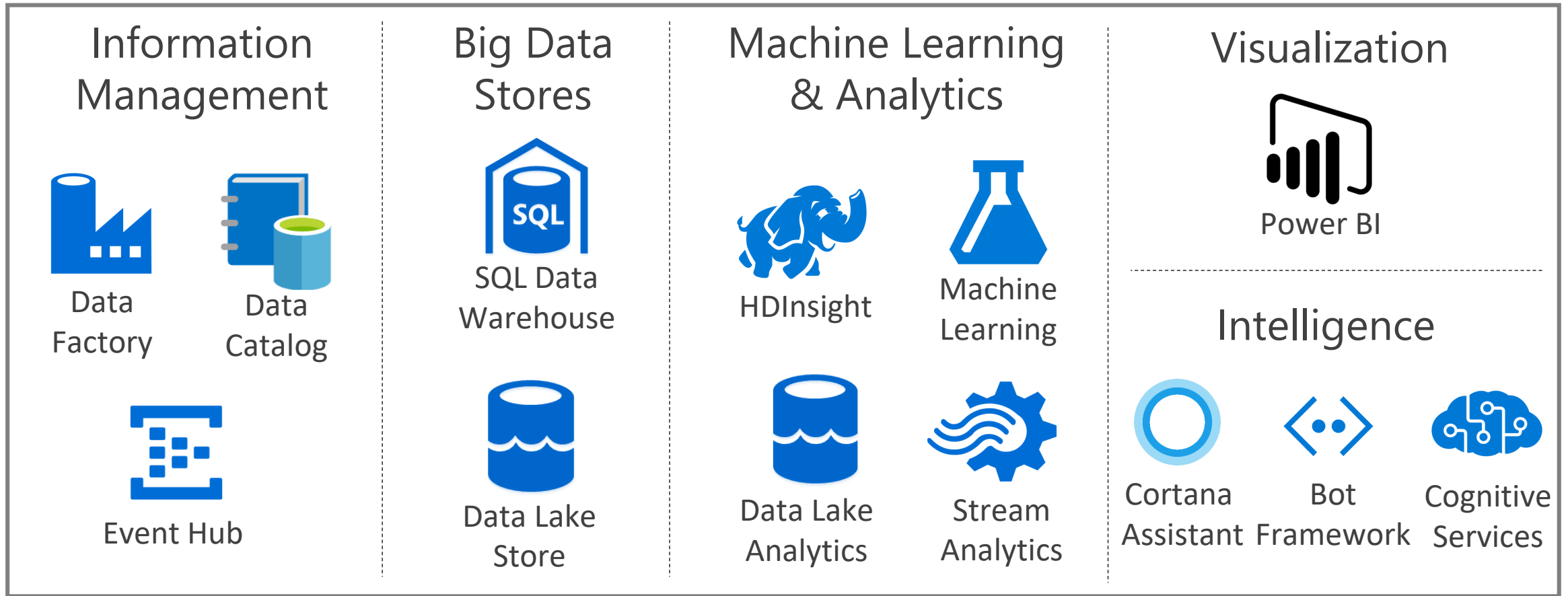


# Cortana Intelligence Suite

---

Wrap-Up

# Cortana Intelligence Components



Plus other solution components such as:



# Quick Reference of Cortana Intelligence Components



Data Catalog

Enterprise metadata catalog & data dictionary



Data Lake Store

Big data storage



Event Hub

Event ingestion for streaming & IoT



Power BI

Data visualization + self-service data prep & modeling



Data Factory

Data ingestion, orchestration, and batch processing



Data Lake Analytics

Query service for big data processing



Stream Analytics

Processing engine for streaming & IoT



Cortana Assistant

Personal digital assistant



Bot Framework

Agents which automate processes



SQL Data Warehouse

Relational data warehousing at scale integrated with big data



HDInsight

Big data clusters with support for Hadoop open source projects



Machine Learning

Predictive analytics service



Cognitive Services

Extends applications to see, hear & interpret

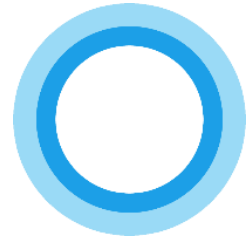
# Cortana Intelligence Suite – Current State

## Young Set of Services

- ✓ Many services have become generally available relatively recently
- ✓ Functionality is still maturing

## Integration Still Evolving

- ✓ The ultimate goal is deep integration between many Azure services



Cortana Intelligence Suite

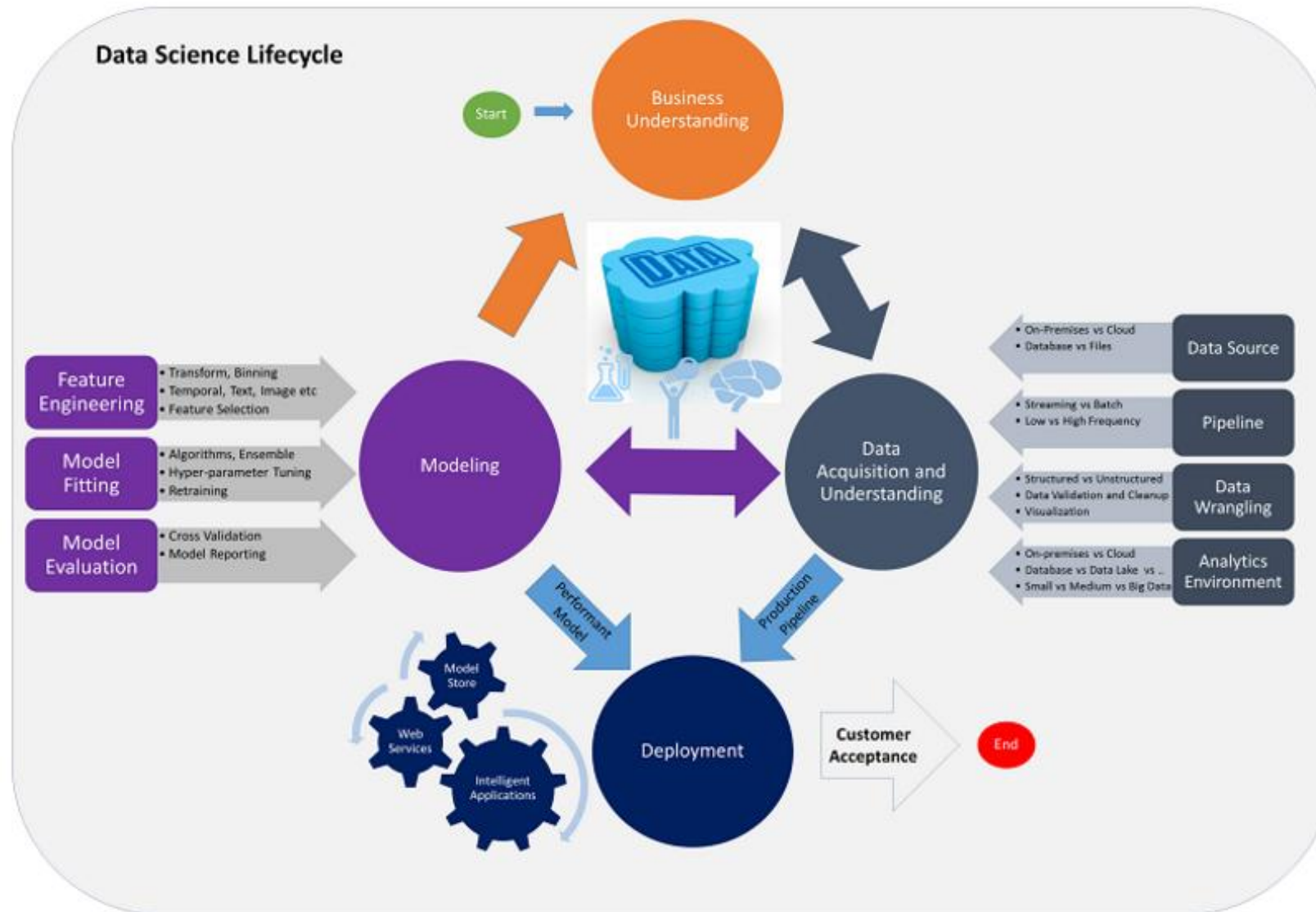
---

Resources for  
Samples & Tutorials



# Team Data Science Process

The 'Process' part  
of Cortana  
Intelligence Suite



The TDSP lifecycle is composed of five major stages that are executed iteratively. These include:

- **Business Understanding**
- **Data Acquisition and Understanding**
- **Modeling**
- **Deployment**
- **Customer Acceptance**

For each stage, we provide the following information:

- **Goals:** the specific objectives itemized.
- **How to do it:** the specific tasks outlined and guidance provided on completing them.
- **Artifacts:** the deliverables and the support for producing them.

# Cortana Intelligence Gallery

Refine by

CATEGORIES ▾

- Solution
- Experiment
- Machine Learning API
- Competition
- Tutorial
- Collection
- Notebook
- Custom Module
- Classroom Training
- Video Training

SHOW ▾

- Microsoft content only

TAGS ▾

- R
- Classification
- Microsoft R Server
- Python

## Results

Sort by: Popular ▾



 EXPERIMENT

### Sample 1: Download dataset from UCI: Adult 2 class dat...

This sample demonstrates how to download a dataset from a http location, add column names to the dataset and examine the dataset and compute some basic statistics.

👁 52141 ↓ 175929 one year ago



 EXPERIMENT

### Tutorial: Building a classification model in Azu...

This experiment serves as a tutorial on building a classification model using Azure ML. We will be using the Titanic passenger data ...

 Two-Class Decision Forest

👁 22551 ↓ 937010 months ago



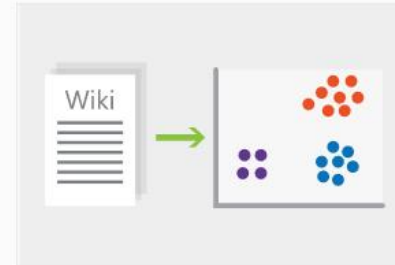
 EXPERIMENT

### Telco Customer Churn

Customer churn can take different forms, such as switching to a competitor's service, reducing the number of services used, or swit...

 Two-Class Decision Forest

👁 18049 ↓ 5503 6 months ago



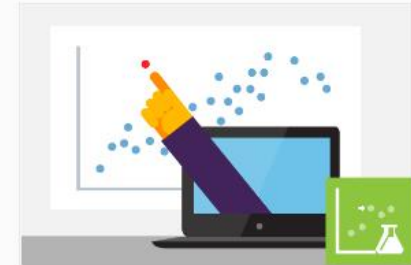
 EXPERIMENT

### Clustering: Find similar companies

This experiment clusters similar companies into same group given their Wikipedia articles and can be used to assign cluster to new ...

 K-Means Clustering

👁 16393 ↓ 101191 months ago



 MACHINE LEARNING API

### Anomaly Detection

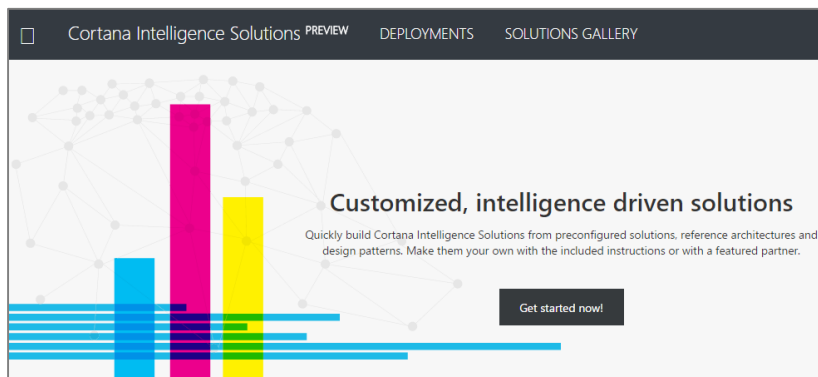
Detect different anomalous patterns in your time series data using machine learning algorithms. Level changes, trend changes, spikes are supported on seasonal and non-seasonal time series.

👁 14314 ↓ 3 months ago



# Cortana Intelligence Gallery – Solutions

(1/3)



## TUTORIAL

### Data Warehousing and Modern BI on Azure

By AzureML Team for Microsoft • April 17, 2017



#### Summary

This one-click deploy tutorial creates a fact and dimension generation pipeline fronted with tabular models to showcase data warehousing on the cloud using the AdventureWorks dataset.

#### Description

#### Overview

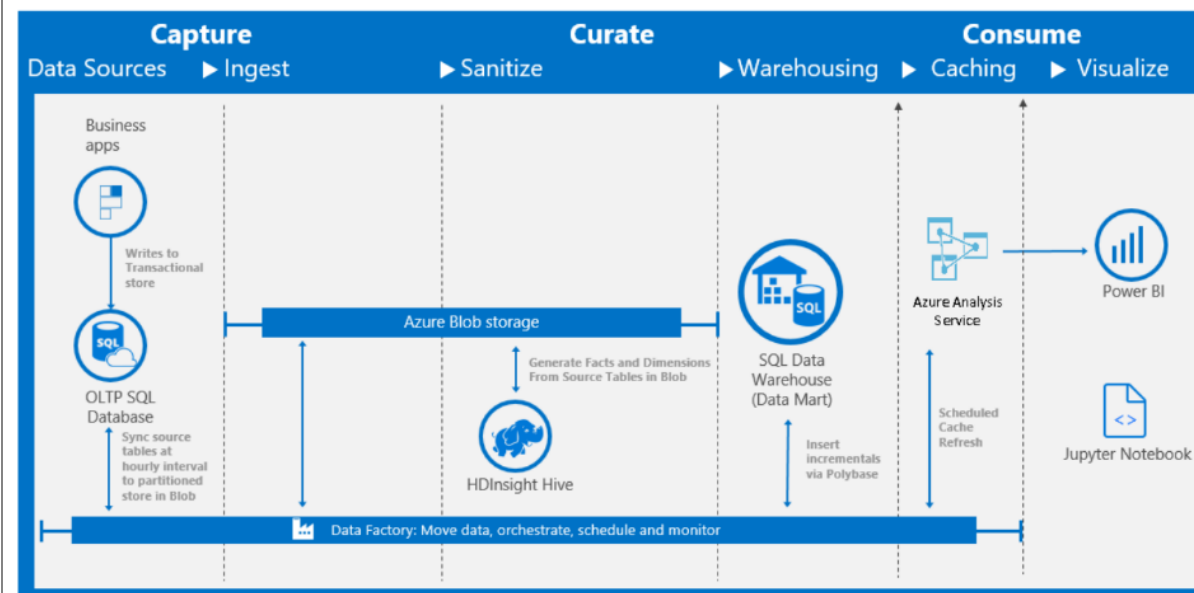
With the advent of Big Data infrastructure and Cloud, enterprises have started to collect, persist, process and analyze large amount of structured and unstructured data from various internal and external data sources. Furthermore, traditional appliance-based on-premises Enterprise Data Warehouse (EDW) systems in whole or parts onto

<https://caqs.azure.net/>

## Pricing

This solution will cost **\$131.49/day** to run. [Refer here for details.](#)

## Architecture



In this solution, we demonstrate how a hybrid EDW scenario can be implemented on Azure using:

1. **Azure SQL Data Warehouse** as a Data mart to vend business-line specific data.
2. **Azure Analysis Services** as an analytical engine to drive reporting.
3. **Azure Blob Storage** as a Data Lake to store raw data in their native format until needed in a flat architecture.
4. **Azure HDInsight** as a processing layer to transform, sanitize and load raw data into a de-normalized format suitable for analytics.
5. **Azure Data Factory** as our orchestration engine to move, transform, monitor and process data in a scheduled time-slice based manner.

# Cortana Intelligence Gallery - Solutions

(2/3)



Deploy

Cortana Intelligence Solutions PREVIEW DEPLOYMENTS SOLUTIONS GALLERY

1 | **Create new deployment**

2 | Provide configuration parameters

3 | Resource provisioning (automated)

4 | Done

### Data Warehousing and Modern BI on Azure

**Prerequisites**

- This pattern requires **12** Azure Data Factory 'reserve cores' to operate. Ensure adequate cores are available before provisioning.
- At the time of deployment, please select a region that supports creation of V12 SQL Server instance for Azure SQL Data Warehouse.

Estimated Provisioning Time: **40 Minutes**

[Not ready to deploy or need more information on Cortana Intelligence Solutions? Contact us.](#)

**Deployment name**

(Deployment name must be between 3 and 9 characters, start with a lowercase letter, and contain only lowercase letters and numbers.)

**Subscription**

(2aa1-38e7d)

**Locations**

**Description (optional)**

[License](#)

# Cortana Intelligence Gallery – Solutions

(3/3)

Cortana Intelligence Solutions PREVIEW **DEPLOYMENTS** SOLUTIONS GALLERY

All Deployments ▾ + New ▾ □ Clean up Deployments

1 deployment(s)

Deployment name ↑	Type	Subscription	Date created	Status	Created by
<a href="#">dwbitest</a>	Data Warehousing and Modern BI on Azure	Microsoft Azure Sponsorship	04-22-2017 07:57:04	Ready	

Cortana Intelligence Solutions PREVIEW **DEPLOYMENTS** SOLUTIONS GALLERY

- 1 | Create new deployment
- 2 | Provide configuration parameters
- 3 | Resource provisioning (automated)
- 4 | **Done**

**dwbitest**

Solution: Data Warehousing and Modern BI on Azure [Links](#)  
Resource group: [dwbitest](#) [Azure activity log](#)  
Status: **Ready** [Metrics](#)

### Next Steps

#### What's deployed into my subscription ?

You now have a data pipeline that

- Takes transactional data from your OLTP Store (Azure SQL Database)
- Creates a data lake off this data.
- Runs transforms on this raw data to create fact and dimension tables.
- Creates a tabular model off these fact/dimension tables.

This pipeline also refreshes at a 15 minute interval and accordingly updates the data lake/warehouse/tabular model with new data.

#### Overview

The following are now available to you in your subscription:

- The Adventure Works OLTP data source on Azure SQL Database.
- Raw OLTP source data cloned to Azure Blob.
- Star-schema based Fact & Dimension Hive tables backed by Azure Blob persistence built from the OLTP sources.
- Data Warehouse hosting the generated Fact & Dimension tables in Azure SQL Data warehouse.
- Azure Analysis services based tabular model sourcing data from the data mart with scheduled model refreshes handled by Azure Automation and Azure Data Factory.
- Batch (initial one-time load) & Incremental (change data capture for time slices) pipelines to orchestrate the ETL process.

Info on what has been provisioned & what steps need to be taken next



# GitHub

The screenshot shows the GitHub interface for the repository 'Azure / Cortana-Intelligence-Gallery-Content'. The top navigation bar includes 'This repository', a search bar, and links for 'Pull requests', 'Issues', and 'Gist'. On the right, there are icons for notifications, a plus sign, and a user profile. Below the repository name, there are statistics for 'Watch' (118), 'Star' (55), and 'Fork' (36). A secondary navigation bar shows 'Code', 'Issues' (1), 'Pull requests' (0), 'Projects' (0), 'Wiki', 'Pulse', and 'Graphs'. The main content area shows the current branch as 'master' and the path 'Cortana-Intelligence-Gallery-Content / Tutorials /'. There are buttons for 'Create new file', 'Upload files', 'Find file', and 'History'. A commit by 'miguelgferro' is highlighted, with the message 'changes urls' and the latest commit hash '71cdf7e' on 'Jan 25'. Below this, a list of folders and their commit messages and dates is shown:

..		
Data-Lake	update	10 months ago
Deep-Learning-for-Text-Classification-in-Azure	changes urls	2 months ago
MRS-for-SAS-Users	Fixed some errors in the course readme files.	11 months ago
MXNet-Azure-GPU	small typo	6 months ago
R-for-SAS-Users	Fixed some errors in the course readme files.	11 months ago
SQL-Data-Warehouse	New screenshots - the portal GUI changed	6 months ago
Step-by-Step Guide for End-to-End Pipeline Depl...	update links	7 months ago
Training-ResNet-on-ImageNet-with-MRS-and-GP...	updated readme	5 months ago

<https://github.com/Azure/Cortana-Intelligence-Gallery-Content/tree/master/Tutorials>

# LearnAnalytics@MS

LearnAnalytics@MS   Analytics Training   Course Catalog   Data Science Certifications   Analytics Training Partners

## Start Learning Today

Dive into Webinars, On-Demand Videos, and Classroom Training to quickly master big data and advanced analytics techniques with Microsoft.



### Data Science Virtual Machine – A Walkthrough of end-to-end Analytics Scenarios

View this webinar to learn how the Data Science Virtual Machine (DSVM) in Microsoft Azure conveniently enables key end-to-end data analytics scenarios. See a variety of popular scenarios which will review test and development, training and scoring for deep-learning on GPU based instances of the DSVM.

[Learn more](#)







### Tutorial: Using R for Scalable Data Analytics: From single machines to Hadoop Spark clusters

Learn about Spark data-science and ML workflows using SparkR, sparklyr, RevoScaleR in the hands-on sections, and play with Microsoft R Server operationalization, parallelization using RxExec and other CRAN packages. (These materials were presented at Strata, San Jose - Mar 14, 2017.)

[Learn more](#)

- Refine by
- AVAILABILITY
- Instructor Led
  - Self Paced
- TRAINING TOPICS
- Bot Framework
  - Cognitive Services
  - Cortana Intelligence
  - Data Catalog
  - Data Factory
  - Data Lake Analytics
  - Data Lake Store / Blob Store
  - Event Hub
  - HDInsight (Hadoop & Spark)
  - Machine Learning
  - Microsoft R Server
  - Power BI
  - Python
  - R Language
  - SQL Data Warehouse
  - Stream Analytics

### Instructor Led Courses

<p>INTERMEDIATE INSTRUCTOR LED</p>  <p><b>Cortana Intelligence Suite - Foundations</b></p> <p>Apr 7 / Telford, UK</p> <p>In this workshop you'll cover a series of modules that guide you from understanding an analytics workload, the Cortana Intelligence Suite Process, the foundations of data transfer and storage, data</p>	<p>INTERMEDIATE INSTRUCTOR LED</p>  <p><b>Data Science and Data Engineering Bootcamp</b></p> <p>Apr 10-14 / Chicago, IL</p> <p>Comprehensive, hands-on, in-person bootcamp designed to enable professionals to learn applied data science and data engineering in just 5 days.</p>	<p>INTERMEDIATE INSTRUCTOR LED</p>  <p><b>Cortana Intelligence Suite Workshop – Foundations And Microsoft R for the Architect</b></p> <p>Apr 10-12 / Sydney, Australia</p> <p>Welcome to the Cortana Intelligence Suite workshop delivered by your Microsoft Data Science team. In this</p>
<p>INTERMEDIATE INSTRUCTOR LED</p>  <p><b>Developing and Deploying Intelligent Chat Bots</b></p> <p>Apr 13 / Sydney, Australia</p> <p>This 1-day course, designed for</p>	<p>INSTRUCTOR LED</p>  <p><b>Learn Analytics Series - Scalable Data Science with Microsoft R Server + Spark and HDInsight (3 days)</b></p>	<p>INTERMEDIATE INSTRUCTOR LED</p>  <p><b>MicrosoftML 1.3.0: What's new for machine learning in Microsoft R Server</b></p> <p>Apr 18 / Pacific Time</p>

# Azure Portal

The screenshot shows the Azure Portal interface for Data Lake Analytics. The top navigation bar includes 'Data Lake Analytics (Default Directory)', 'dev', and 'Data Lake Analytics'. The main menu has 'Add', 'Columns', and 'Refresh'. The left sidebar shows 'Subscriptions: 2 of 3 selected' and a list of items. The central area has a search bar and a list of actions: 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', and 'Diagnose and solve problems'. The right sidebar shows 'New Job', 'Sample Scripts' (circled in red), 'Data Explorer', 'Delete', and 'View All Jobs'. Below this, the 'Essentials' section displays resource details: Resource group (change) BI..., DevRG; Status Running; Location East US 2; Subscription name (change) ..., Development; and Subscription ID.

The screenshot shows the Azure Portal interface for Data Factory 'Sample pipelines'. The top navigation bar includes 'Data Factory', 'Sample pipelines', and 'Data Factory'. The main menu has 'Delete' and 'Move'. The left sidebar shows 'Essentials' with resource details: Resource group BI..., DevRG; Location EastUS; Subscription name ..., Development; and Subscription id. Below this, the 'Actions' section has buttons for 'Author and deploy', 'Copy data (PREVIEW)', 'Monitor & Manage', 'Sample pipelines' (circled in red), 'Diagram', and 'Metrics and operations'. The 'Contents' section shows a summary of Datasets (2), Pipelines (1), and Linked services (2). The right sidebar shows 'Customer profiling' and 'Suggest a sample' cards.

The screenshot shows the Azure Portal configuration window for SQL Data Warehouse. The title bar is 'SQL Data Warehouse'. The configuration fields include: Database name (AdventureWorksDW), Subscription (y-Development), Resource group (BI..., DevRG) with 'Use existing' selected, 'Select source' (Sample), 'Select sample' (AdventureWorksDW), Server (sandboxsampledb (East US 2)), Server admin login, Password, Collation (Specified by sample), and Performance (100). At the bottom, there is a 'Pin to dashboard' checkbox, a 'Create' button, and a link for 'Automation options'.



# Azure Virtual Machines

The image shows two overlapping screenshots of the Azure portal's Virtual Machines section. The top screenshot shows a search for 'data science' with three results, the first of which is circled in red. The bottom screenshot shows a search for 'SQL Server 2016' with seven results.

**Virtual machines (Default Directory)**

**Compute**

**Subscriptions:** 1 of 3 selected – Don't see a subscription? [Switch directories](#)

Filter by name...

-Development

2 items

NAME
BI...MDev1
BI...MDev2

**Filter**

data science

**Results**

NAME	PUBLISHER	CATEGORY
Data Science Virtual Machine	Microsoft	Virtual Machine Images
Linux Data Science Virtual Machine	Microsoft	Virtual Machine Images
Linux Data Science Virtual Machine (BYOL/CSP)	Microsoft	Virtual Machine Images

**Subscriptions:** 2 of 3 selected – Don't see a subscription? [Switch directories](#)

Filter by name...

2 subscriptions

4 items

NAME
BI...Dev1
BI...Dev2
BI...Test1
BI...Test2

**Filter**

SQL Server 2016

**Results**

NAME	PUBLISHER	CATEGORY
Free License: SQL Server 2016 SP1 Developer on Windows Server 2016	Microsoft	Recommended
R Server Only SQL Server 2016 Enterprise	Microsoft	Recommended
SQL Server 2016 SP1 Standard on Windows Server 2016	Microsoft	Recommended
SQL Server 2016 SP1 Web on Windows Server 2016	Microsoft	Recommended
(BYOL) SQL Server 2016 RTM Enterprise on Windows Server 2012 R2	Microsoft	Recommended
SQL Server 2016 SP1 Enterprise on Windows Server 2016	Microsoft	Recommended

# Azure Documentation

Azure / Machine Learning

Filter

- > Overview
- ▼ Get Started
  - Create your first experiment
  - ▼ Example walkthrough
    - Create a predictive solution
    - 1: Create a workspace
    - 2: Upload data
    - 3: Create experiment
    - 4: Train and evaluate
    - 5: Deploy web service
    - 6: Access web service
  - > Data Science for Beginners
    - R quick start
- > How To

Azure / Data Factory

Filter

- ▼ Overview
  - Introduction to Azure Data Factory**
  - > Concepts
- ▼ Get Started
  - > Tutorial: Create a pipeline to copy data
  - > Tutorial: Create a pipeline to transform data
- FAQ
- > How To
- ▼ Reference
  - PowerShell
  - .NET
  - REST
- ▼ Resources
  - Release notes for Data Management Gateway
  - Learning path
  - > Case Studies
  - Service updates

Azure / Data Lake Analytics

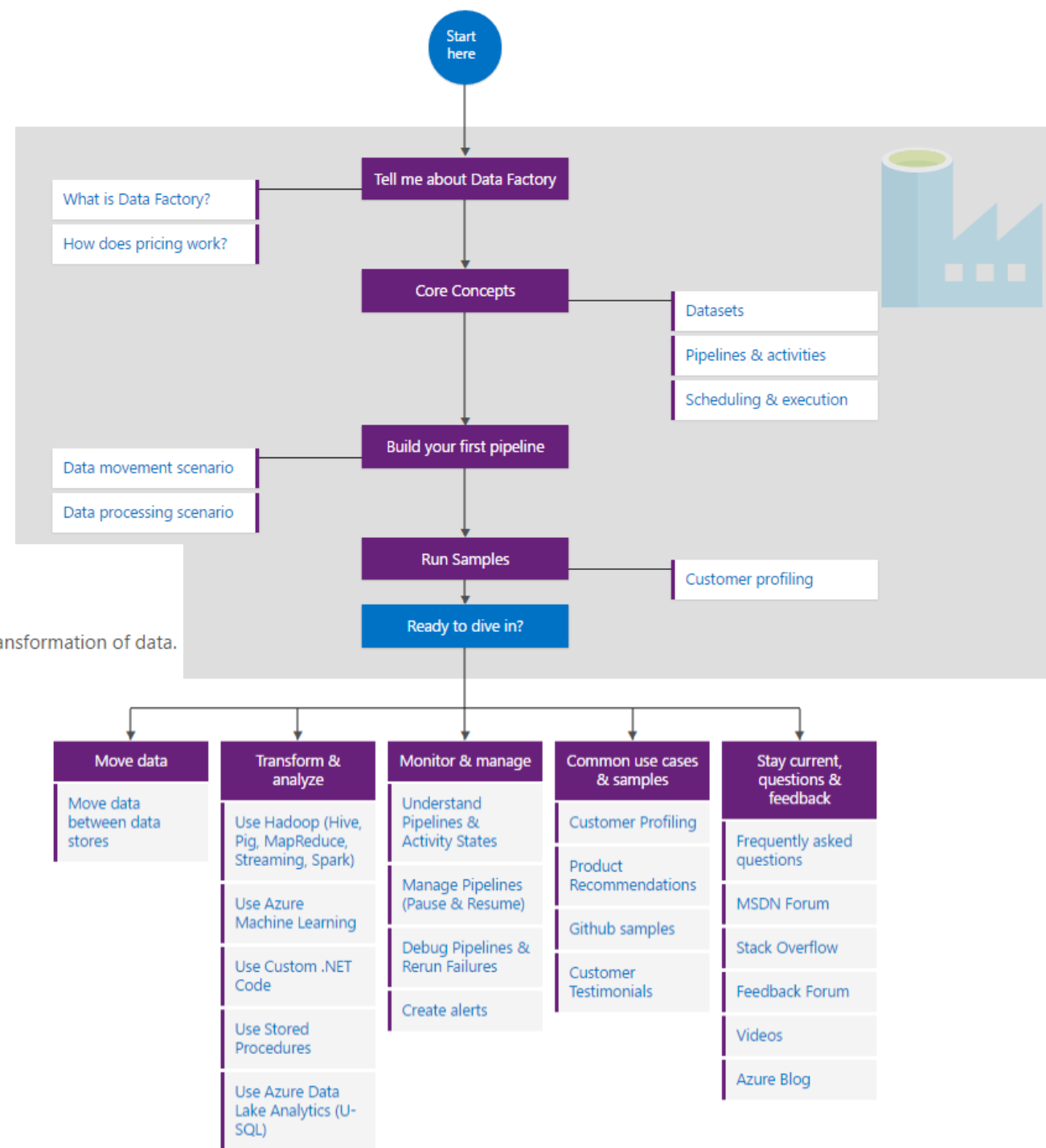
Filter

- ▼ Overview
  - What is Data Lake Analytics?
- ▼ Get started
  - Interactive tutorial: Analyze website logs**
  - Azure portal
  - PowerShell
  - .NET SDK
  - Java SDK
  - Visual Studio
  - Azure CLI
  - Azure CLI 2.0
  - REST API
  - Python
  - > How to
  - > Reference
  - > Resources

# Azure Learning Paths

## Learning path for Azure Data Factory

Data Factory is a cloud-based data integration service that orchestrates and automates the movement and transformation of data. Follow the guidance on this page for an effective learning path through Data Factory content.



# Visual Studio Projects

The screenshot shows the 'New Project' dialog in Visual Studio. The left sidebar is expanded to 'Azure Data Lake' > 'U-SQL (ADLA)'. The main list shows several project templates, with 'USQL Sample Application' highlighted in blue. The right pane shows the details for this project.

Project Name	Framework	Type	Description
U-SQL Project	.NET Framework 4.5.2	Azure Data Lake	
Class Library (For U-SQL Application)	.NET Framework 4.5.2	Azure Data Lake	
U-SQL Unit Test Project	.NET Framework 4.5.2	Azure Data Lake	
<b>USQL Sample Application</b>	.NET Framework 4.5.2	Azure Data Lake	<b>Type:</b> Azure Data Lake A project for creating a U-SQL Sample Project.
U-SQL Unit Test Sample	.NET Framework 4.5.2	Azure Data Lake	

The screenshot shows the 'New Project' dialog in Visual Studio. The left sidebar is expanded to 'Azure Data Lake' > 'DataFactory'. The main list shows two project templates, with 'Data Factory Templates' highlighted in blue. The right pane shows the details for this project.

Project Name	Framework	Type	Description
Data Factory Templates	.NET Framework 4.5.2	DataFactory	<b>Type:</b> DataFactory A project for creating a data integration solution based on a template.
Empty Data Factory Project	.NET Framework 4.5.2	DataFactory	

# Channel 9 Videos

filters Recent<sup>x</sup>



Azure SQL DW

## Part 2 – working with tables in Azure SQL Data Warehouse

Sep 30, 2016 at 3:22PM  
by sushmav, ktolle

In this video, you will be able to create a database and tables within an Azure SQL Data Warehouse. The content will also explore the key considerations such as partitions, indexes and statistics. You...

★★★★★ 7 ratings 0 comments

[view episode](#)



Azure SQL DW  
Part 3–Loading Data into Azure SQL Data Warehouse



Azure SQL DW  
Part1– Azure SQL Data Warehouse Overview



Azure SQL DW  
BltoDCTeaser

<https://channel9.msdn.com/Series/Azure-SQL-DW>

# AppSource

AppSource Apps Partners List on AppSource Blog How it works

## Find the right app for your business needs

Get solutions tailored to your industry that work with the products you already use.

Search Microsoft AppSource

Refine by category

- Analytics 121
- Artificial intelligence 11
- Collaboration 46
- Customer service 87
- Finance 116
- Human resources 19
- IT + administration 44
- Marketing 116
- Operations + supply chain 125
- Productivity 142
- Sales 235

Refine by industry

- Agriculture
- Distribution
- Financial services
- Government
- Healthcare + life sciences
- Manufacturing
- Professional services
- Retail + consumer goods
- Education

Microsoft Cloud Mobility Productivity

AppSource Apps Partners List on AppSource Blog How it works

Search Microsoft AppSource

Products

App results (18) View partner results (9)

Web apps

**Cortana Intelligence**

Hitachi Solutions  
**IoT Predictive Service Hub**  
By Hitachi Solutions  
Cortana Intelligence  
Create a Data-Driven Service Business to Improve Operational Efficiencies  
Contact me

NS  
**Predict Product Quality**  
By New Signature  
Cortana Intelligence  
Predict product quality and reduce waste  
Test Drive

brillio  
**Brillio Smart Assets Management**  
By Brillio  
Web apps  
Smart Asset Management helps industries in Operations domain to monitor assets and regulate failures  
Test Drive

osisoft  
**PI Integrator for Micro...**  
By osisoft  
Web apps  
Driving Digital Transformation and /  
Free trial

Microsoft  
**Quality Assurance**  
By Microsoft  
Cortana Intelligence  
Identify errors and potential failures before they occur rather than after they are detected.  
Test Drive

CloudRecon  
By CloudAtlas  
Web apps  
Make your Cloud migration successful - for migrating your on-premise workloads to the Cloud.  
Free trial

Hitachi Solutions  
**IoT Predictive Service Hub**  
By Hitachi Solutions  
Cortana Intelligence  
Create a Data-Driven Service Business to Improve Operational Efficiencies  
Contact me

NS  
**Predict Product Quality**  
By New Signature  
Cortana Intelligence  
Predict product quality and reduce waste  
Test Drive

Categories

- Analytics
- Artificial intelligence
- Collaboration
- Customer service
- Finance
- Human resources
- IT + administration
- Marketing
- Operations + supply chain
- Productivity
- Sales

Personalized offers  
By Microsoft  
personalized offers to each customer.  
Test Drive

Demand Forecasting and Price Optimization

Saviant  
**Saviant Energy Analytics Platform (SEAP)**

VeriPark  
**VeriPark Next Best Action**

KenSci  
**KenSci Clinical Analytics**  
By KenSci

<https://appsource.microsoft.com/en-us/>



# Additional Reading

Setting up a PC for Azure Cortana Intelligence Suite Development

<http://www.sqlchick.com/entries/2016/6/17/setting-up-a-pc-for-azure-cortana-intelligence-suite-development>

What is the Cortana Intelligence Suite?

<http://www.sqlchick.com/entries/2015/8/22/what-is-the-cortana-analytics-suite>

Should You Use a SQL Server Marketplace Image for an Azure Virtual Machine?

<http://www.sqlchick.com/entries/2016/4/2/should-you-use-a-sql-server-marketplace-image-for-an-azure-virtual-machine>

How to Build a Demo/Test Environment for Azure Data Catalog

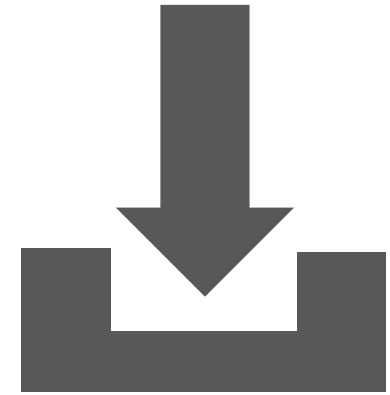
<http://www.sqlchick.com/entries/2016/4/20/how-to-create-a-demo-test-environment-for-azure-data-catalog>

Overview of Azure Data Catalog in the Cortana Intelligence Suite

<http://www.sqlchick.com/entries/2015/9/15/overview-of-azure-data-catalog-in-the-cortana-analytics-suite>

# Thank You for Attending!

To download a copy of this presentation:  
SQLChick.com "Presentations & Downloads" page



Melissa Coates  
Analytics Architect  
SentryOne

Blog: [sqlchick.com](http://sqlchick.com)  
Twitter: [@sqlchick](https://twitter.com/sqlchick)

*Creative Commons License:  
Attribution-NonCommercial-NoDerivative Works 3.0*

