

Azure Machine Learning Analytics

Marcin Szeliga

Agenda

- Azure ML Services
- Azure Databricks
- Azure Data Factory

Azure ML Services



Azure Machine Learning service



Democratize AI

Enable Domain Experts & Developers to get rapidly build AI solutions



Accelerate AI

Improve productivity for Data scientists

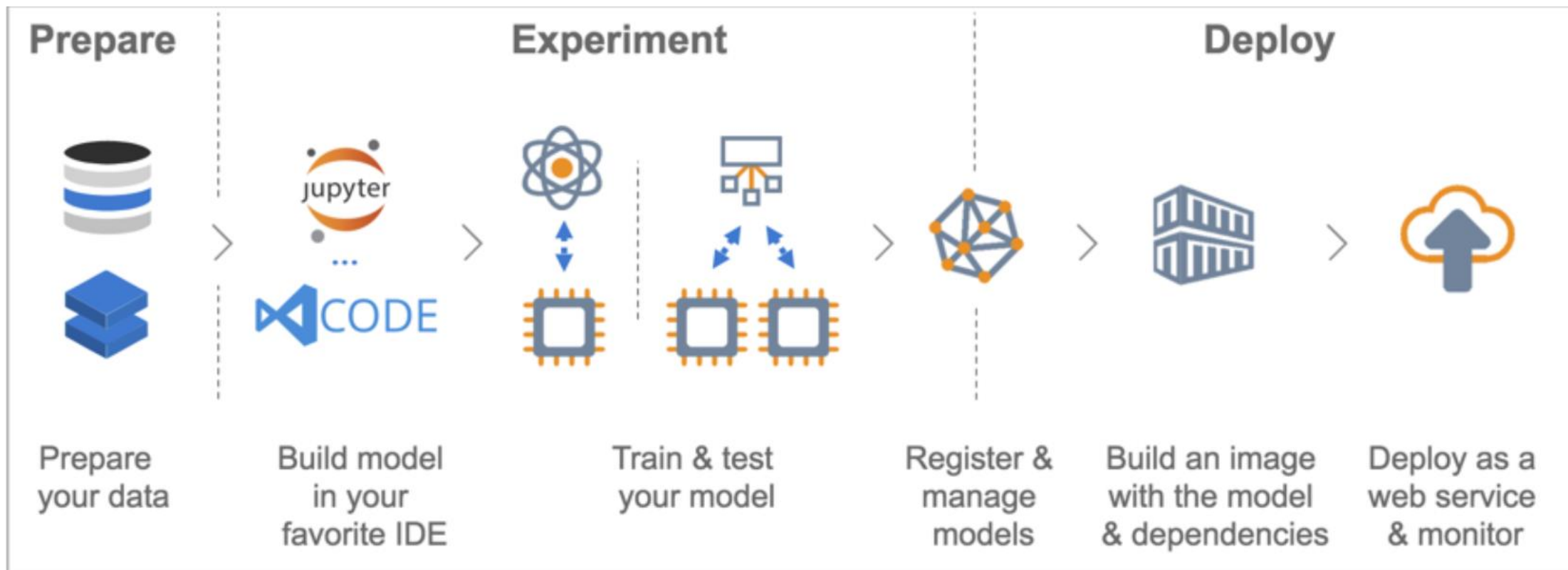


Scale AI

Build AI solutions at scale in an automated fashion

What is Azure Machine Learning service?

Cloud-based environment you can use to prep data, train, test, deploy, manage, and track machine learning models



How to use Azure Machine Learning service

Create a workspace

[Experiments](#) [Pipelines](#) [Compute](#) [Models](#) [Images](#) [Deployments](#) [Activities](#)

Welcome to your new Workspace

1. Getting started

Create your first experiment in Azure Notebooks to be able to view and track metrics.

[Open Azure Notebooks](#)
[View More Sample Notebooks](#)


2. Done getting started?

Once you run the Azure Notebook, you will be able to view the data from the experiment in the Experiments page.

[View Experiments](#)


What's possible with AML?

Using Azure Machine Learning service, you can track your models as you build, train, deploy, and manage them at cloud scale.




Run & Monitor Experiments

Submit Experiments for training and automatically track their progress and view logs.




Register Models

Save scoring logic operations into models to create Docker Images and Deployments.




Build Images

Quickly create Docker images that encapsulate models, scripts, and any associated files.



Deploy Models

Send scoring requests to web services in Azure Container Instances, Azure Kubernetes Service, or field programmable gate arrays (FPGA).



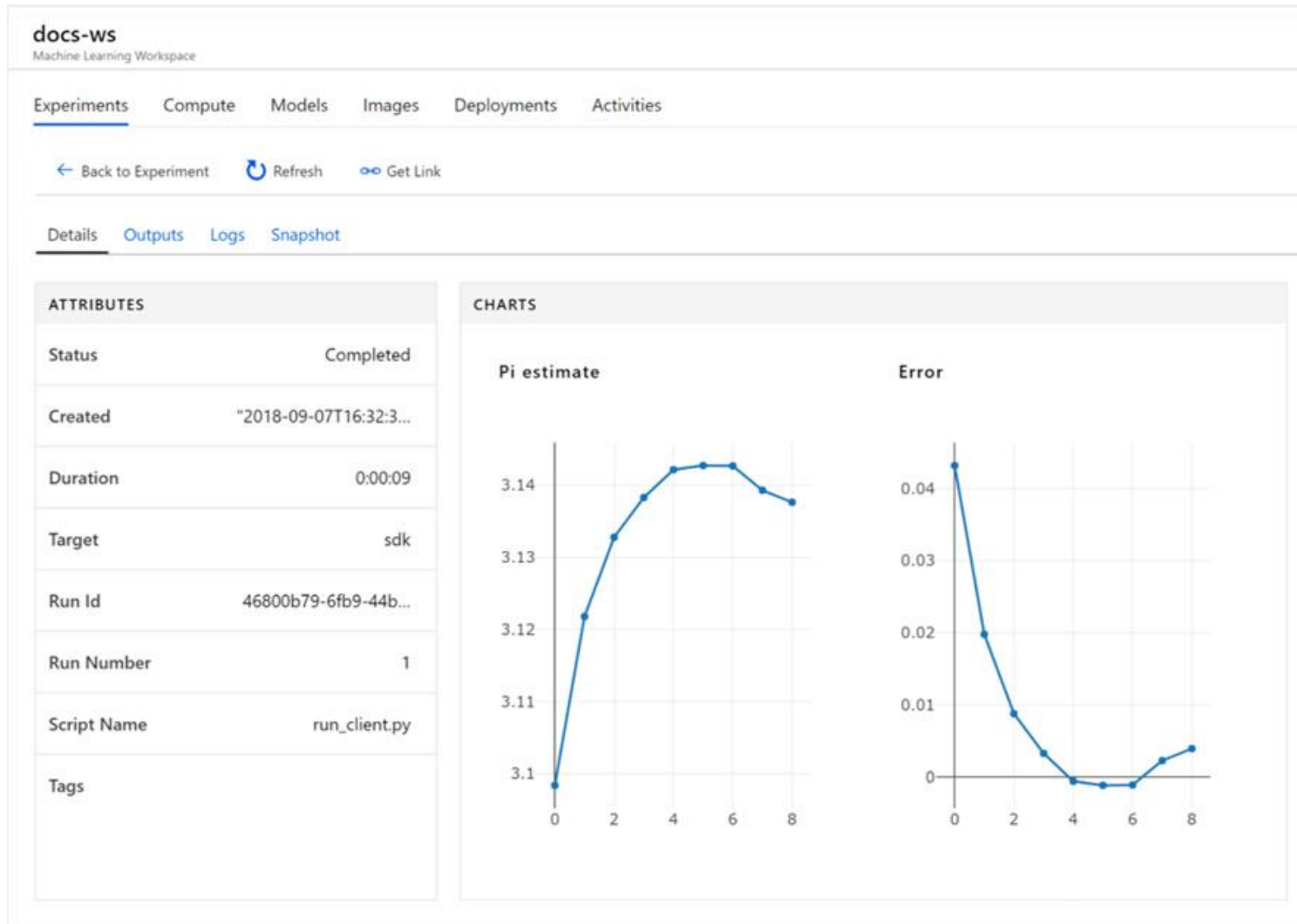
Create Pipeline

Pipelines are used to build, optimize, and manage machine learning workflows.

[Learn more about our features and capabilities here](#)

How to use Azure Machine Learning service

Build and train



How to use Azure Machine Learning service

Deploy and manage


[Experiments](#) [Pipelines](#) [Compute](#) [Models](#) [Images](#) [Deployments](#) [Activities](#)

aci-dog-breeds

[← Back to Deployments](#) [✎ Edit](#) [🗑 Delete](#)

[Details](#) [Models](#) [Images](#)

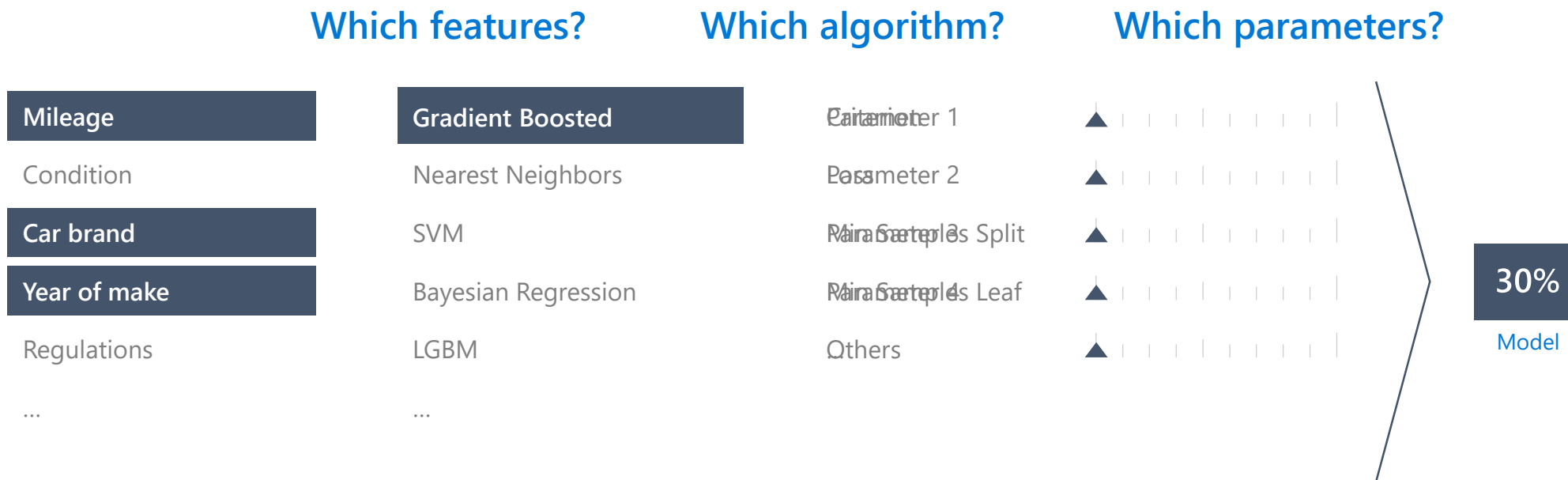
ATTRIBUTES

Description	Classify dog breeds using transfer learning with PyTorch
State	Healthy
Compute Type	ACI
Service ID	aci-dog-breeds
Tags	data : dog_breeds, method : transfer learning, framework : pytorch
Creation date	11/07/2018, 3:22:23 PM UTC
Last updated	11/07/2018, 3:22:29 PM UTC
Image ID	aci-dog-breeds:1
Scoring URI	http://137.117.58.22:80/sc 
CPU	1
Memory	1 GB

What is automated machine learning?

- Automated machine learning (automated ML) picks an algorithm and hyperparameters for you and generates a model ready for deployment
- The model can be downloaded to be further customized as well
- Automated ML Capabilities
 - Based on Microsoft Research
 - Brain trained with several million experiments
 - Collaborative filtering and Bayesian optimization
 - Privacy preserving: No need to “see” the data

Model Creation Is Typically Time-Consuming



Model Creation Is Typically Time-Consuming

Which features?

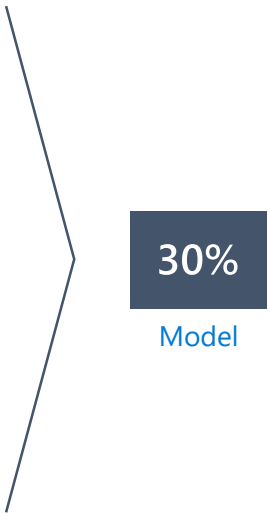
- Mileage
- Condition
- Car brand
- Year of make
- Regulations
- ...

Which algorithm?

- Gradient Boosted
- Nearest Neighbors
- SVM
- Bayesian Regression
- LGBM
- ...

Which parameters?

- Critereion
- Neighbors
- Weights
- Min Samples Split
- Min Samples Leaf
- Others



Model Creation Is Typically Time-Consuming

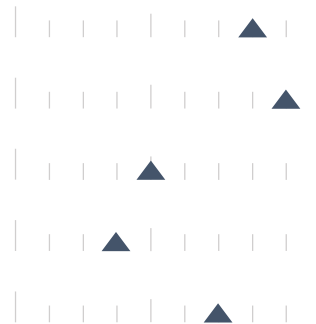
Which features?



Which algorithm?



Which parameters?

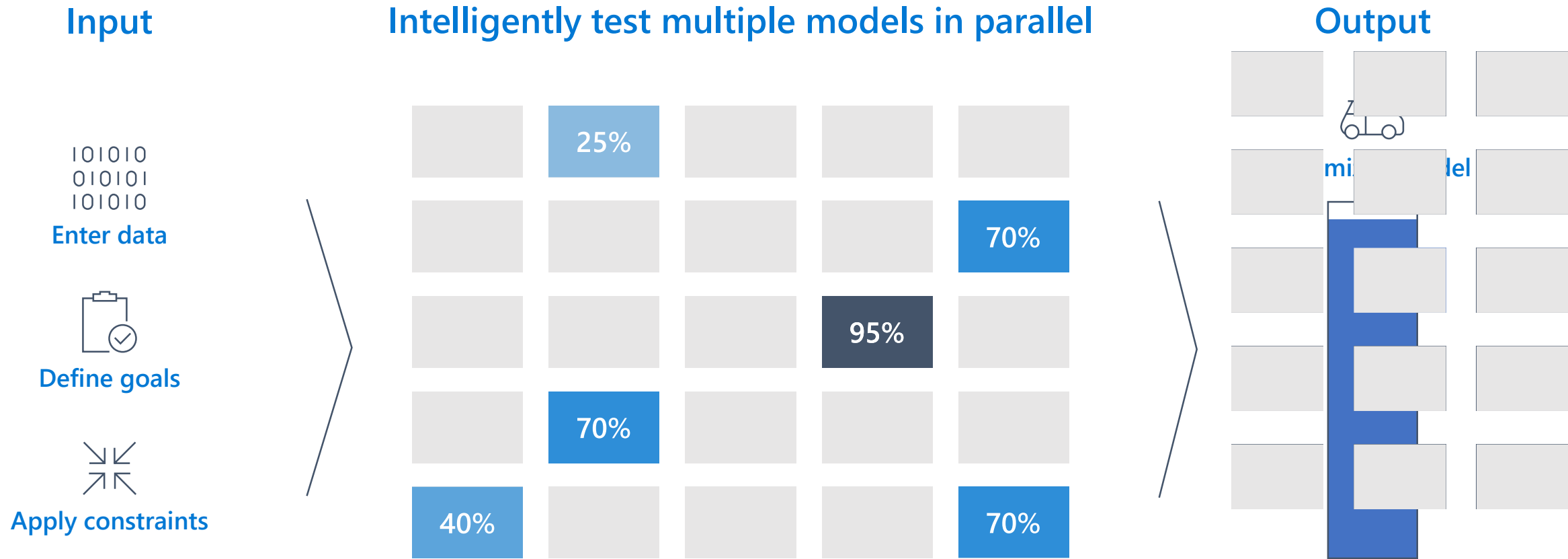


30%

15%

Iterate

Automated ML Accelerates Model Development



Automated ML

Data



Data cleaning support

Automated ML currently supports automated data cleaning

Feature



Feature engineering

Most time consuming part when done manually can now be done within minutes.

Algorithm



Pick and play

Testing many different algorithms at once.

Tuning



What to leave out

Hyperparameter tuning: what to include what to leave out

Ranking



Ranking

Having an overview of the best performing models based on accuracy & speed.

Explaining



Justification

Being able to explain what created an outcome and what features had the most significant impact

Why Use Azure ML Services?

Broadly useful

Is a part of Azure Cloud

Integration with data platforms, ex: PowerBI, SQL, CosmosDB

Efficient

Automated ML accelerates model development

Faster model training using multiple cores and parallel experiments

Integration

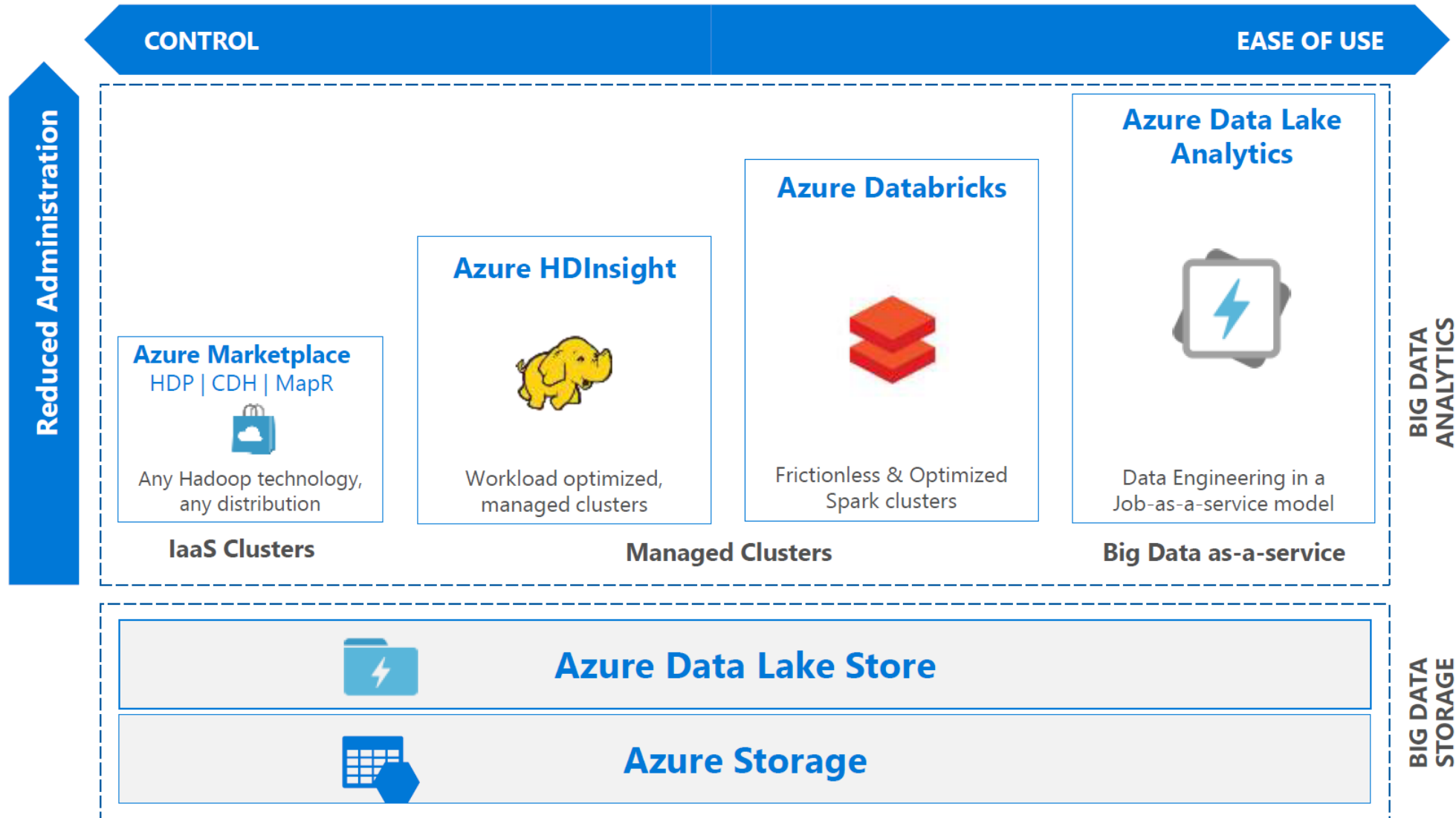
Azure Notebooks, Jupyter Notebooks

Python SDK for deployment and hosting for inference

Azure
Databricks

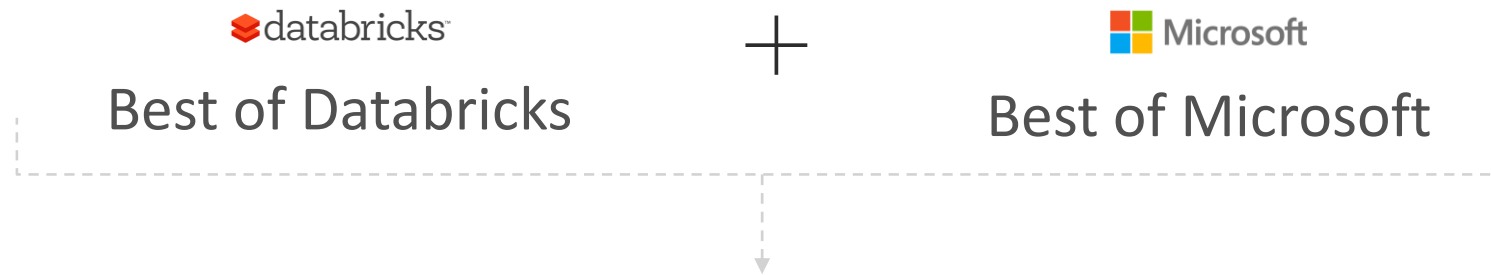


Knowing the various big data solutions





What is Azure Databricks ?


A fast, easy and collaborative Apache® Spark™ based analytics platform optimized for Azure



 Designed in collaboration with the founders of Apache Spark

 One-click set up; streamlined workflows

 Interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.

 Native integration with Azure services (Power BI, SQL DW, Cosmos DB, Blob Storage, ADF, SQL DB, AAD)

 Enterprise-grade Azure security (Active Directory integration, compliance, enterprise-grade SLAs – 99.95%)

What is Apache Spark

Apache Spark emerged to provide a parallel processing framework that supports in-memory processing to boost the performance of big-data analytical applications on massive volumes of data.

Interactive Data Analysis

Used by business analysts or data engineers to analyze and prepare data.

Streaming Analytics

Ingest data from technologies such as Kafka and Flume to ingest data in real-time.

Machine Learning

Contains a number of libraries that enables a Data Scientist to perform Machine Learning.

Why use Azure Databricks

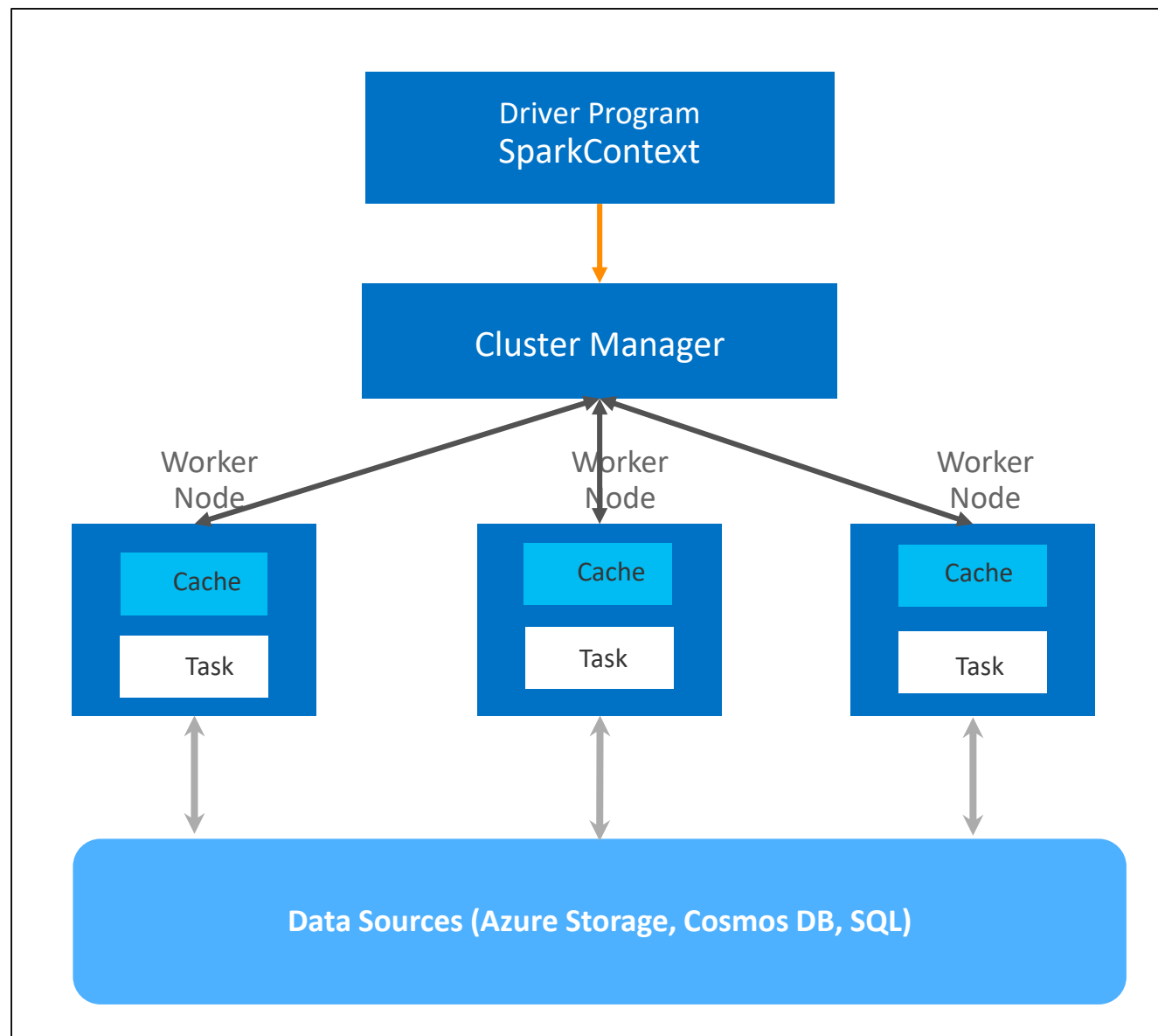
Azure Databricks is a wrapper around Apache Spark that simplifies the provisioning and configuration of a Spark cluster in a GUI interface

Azure Databricks components.

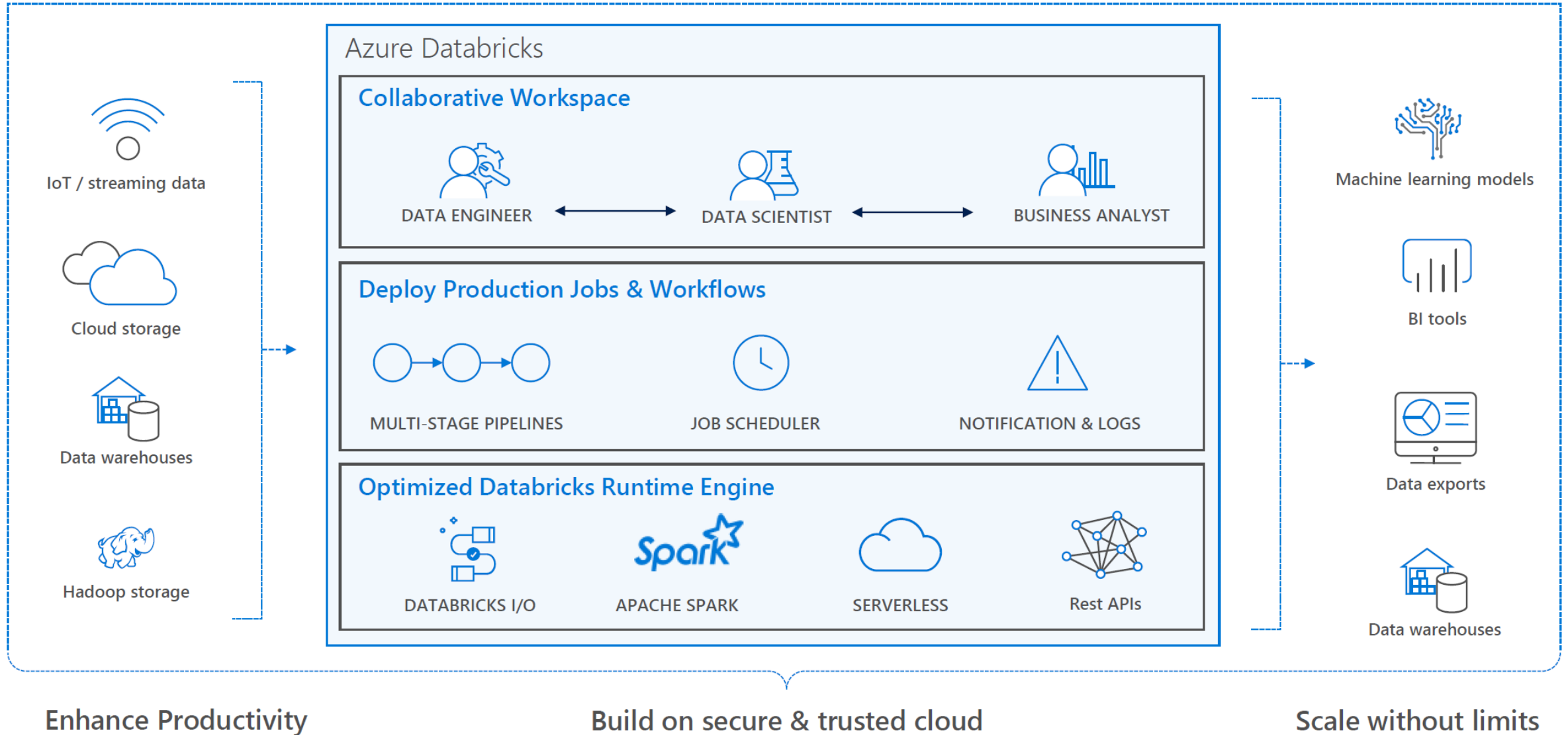
- Spark SQL and DataFrames
- Streaming
- Mlib
- GraphX
- Spark Core API

Spark Architecture & Dataframes

- 'Driver' runs the user's 'main' function and executes the various parallel operations on the worker nodes
- To take advantage of Spark – you use Dataframes as the data structure
- Once your Data is in the DataFrame – Spark can parallelize operations
- The Dataframes support both batch and streaming data
- The results of the operations are collected by the driver



Azure Databricks



Performing ETL to populate a data model

The goal of transformation in Extract Transform Load (ETL) is to transform raw data to populate a data model.

Extraction	Data Validation	Transformation	Corrupt Record Handling	Loading Data
<p>Connect to many data stores:</p> <ul style="list-style-type: none">• Postgres• SQL Server• Cassandra• Cosmos DB• CSV, Parquet• Many more..	<p>Validate that the data is what you expect.</p>	<p>Applying structure and schema to your data to transform it into the desired format.</p>	<p>Built-in functions of Databricks allow you to handle corrupt data such as missing and incomplete information.</p>	<p>Highly effective design pattern involves loading structured data back to DBFS as a parquet file.</p>

Why Use Azure Databricks ?

Enhance productivity

Get started quickly by launching your new Spark environment with one click

Share your insights in powerful ways through rich integration with Power BI

Build on the most compliant cloud

Simplify security and identity control with built-in integration with Active Directory

Regulate access with fine-grained user permissions to Azure Databricks' notebooks, clusters, jobs and data

Scale without limits

Operate at massive scale without limits globally

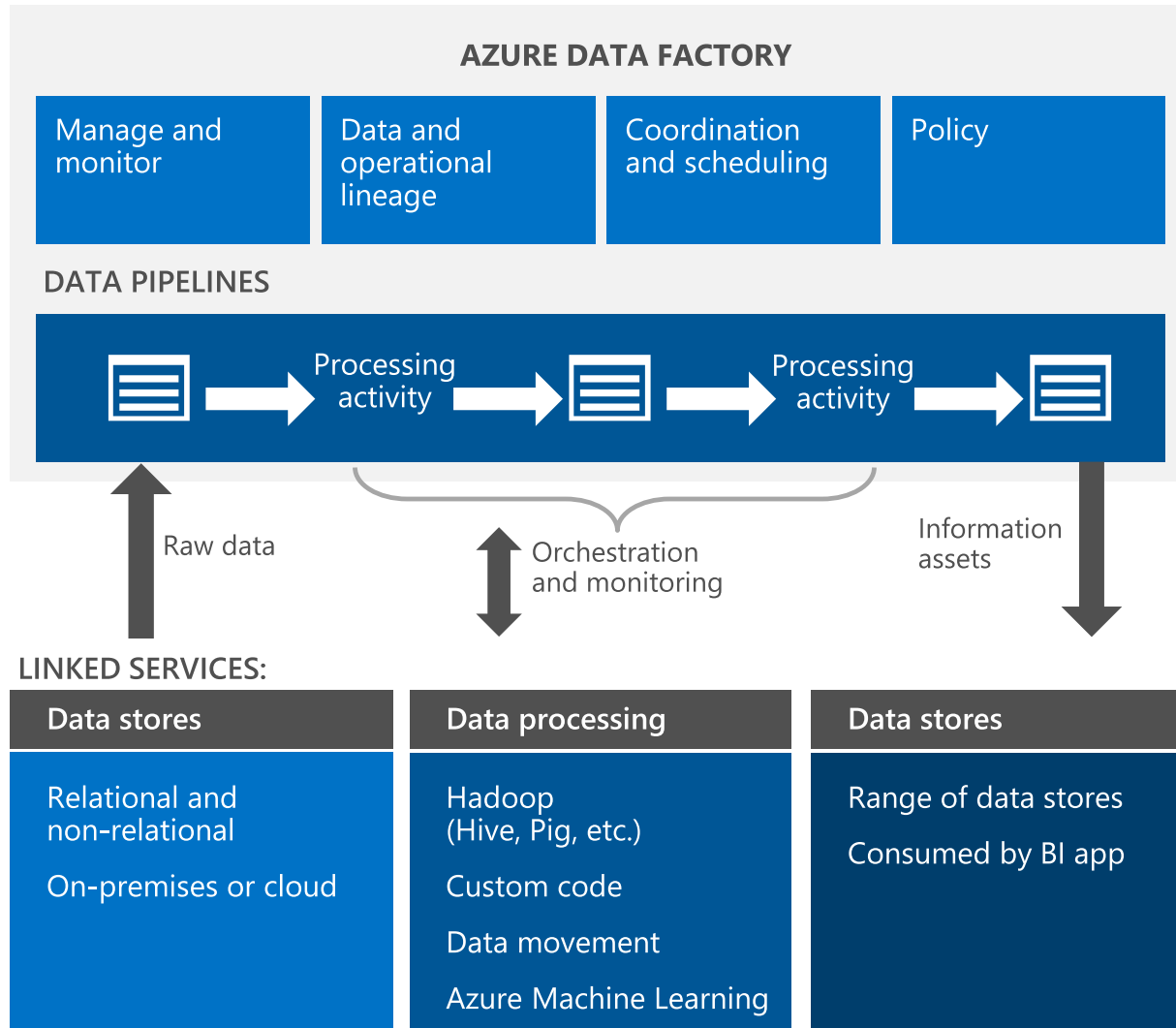
Accelerate data processing with the fastest Spark engine

Azure
Data
Factory



Azure Data Factory

Compose services to transform data into actionable intelligence



Activities

- Actions you perform on your data
- Inputs turned into outputs

Pipelines

- Logical grouping of activities for group operations

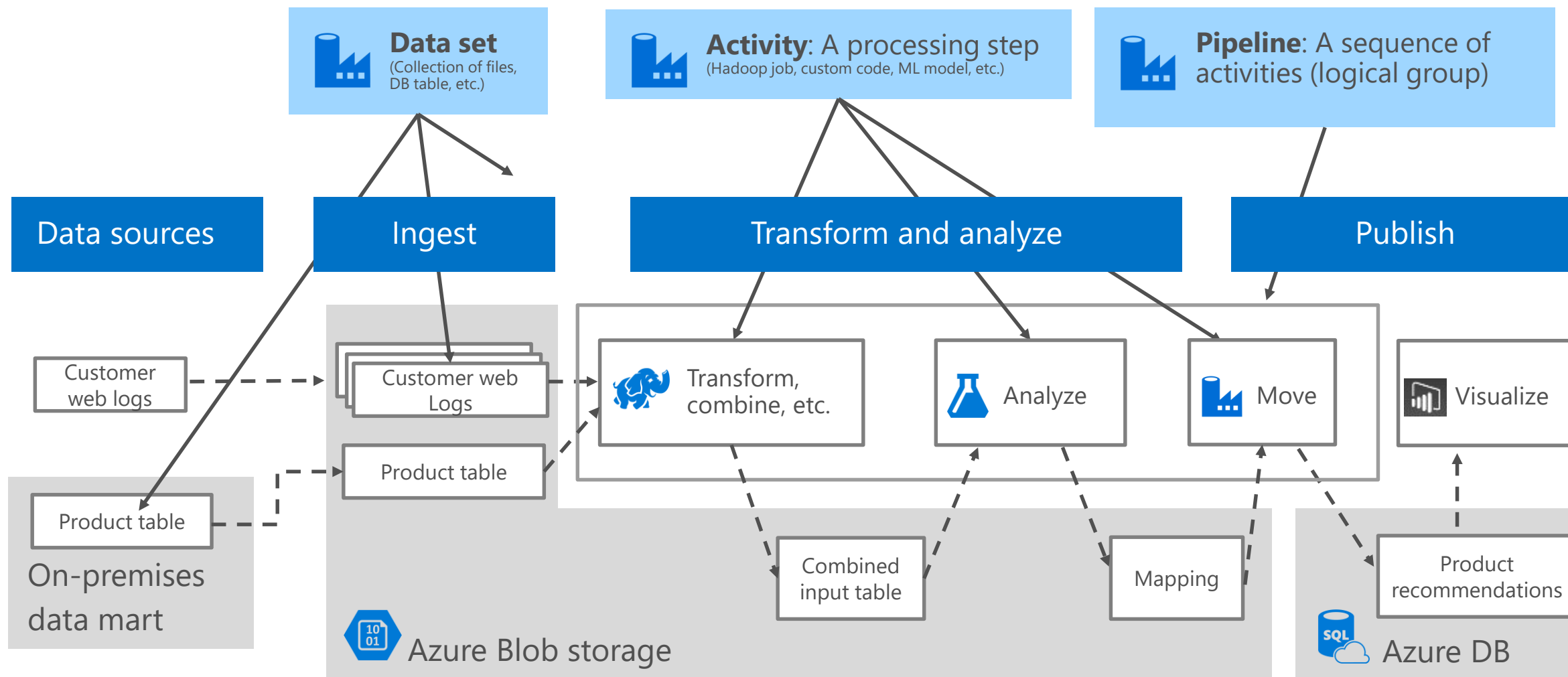
Data sets

- A named reference/pointer to data you want to use as an input or output of an activity

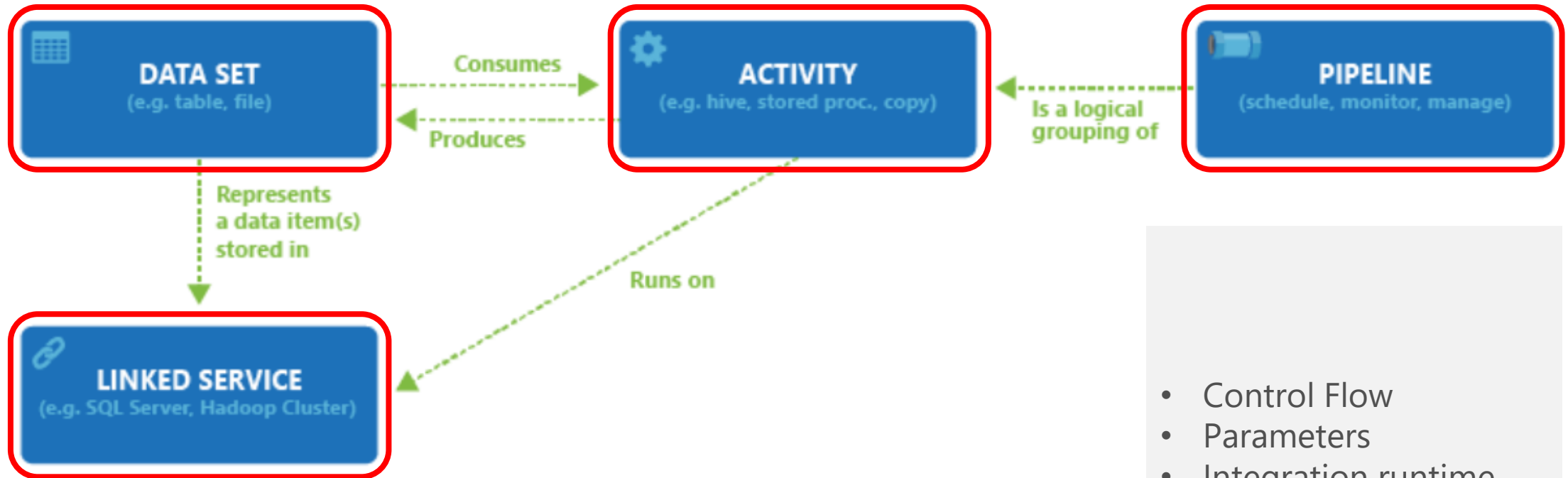
Linked services

- Connection of data factories to the resources and services you want to use
- Connection of data stores like Azure storage and on-premises SQL Server
- Connection of compute services like Azure Machine Learning, Azure HDInsight, and Azure Batch

Data Factory concepts

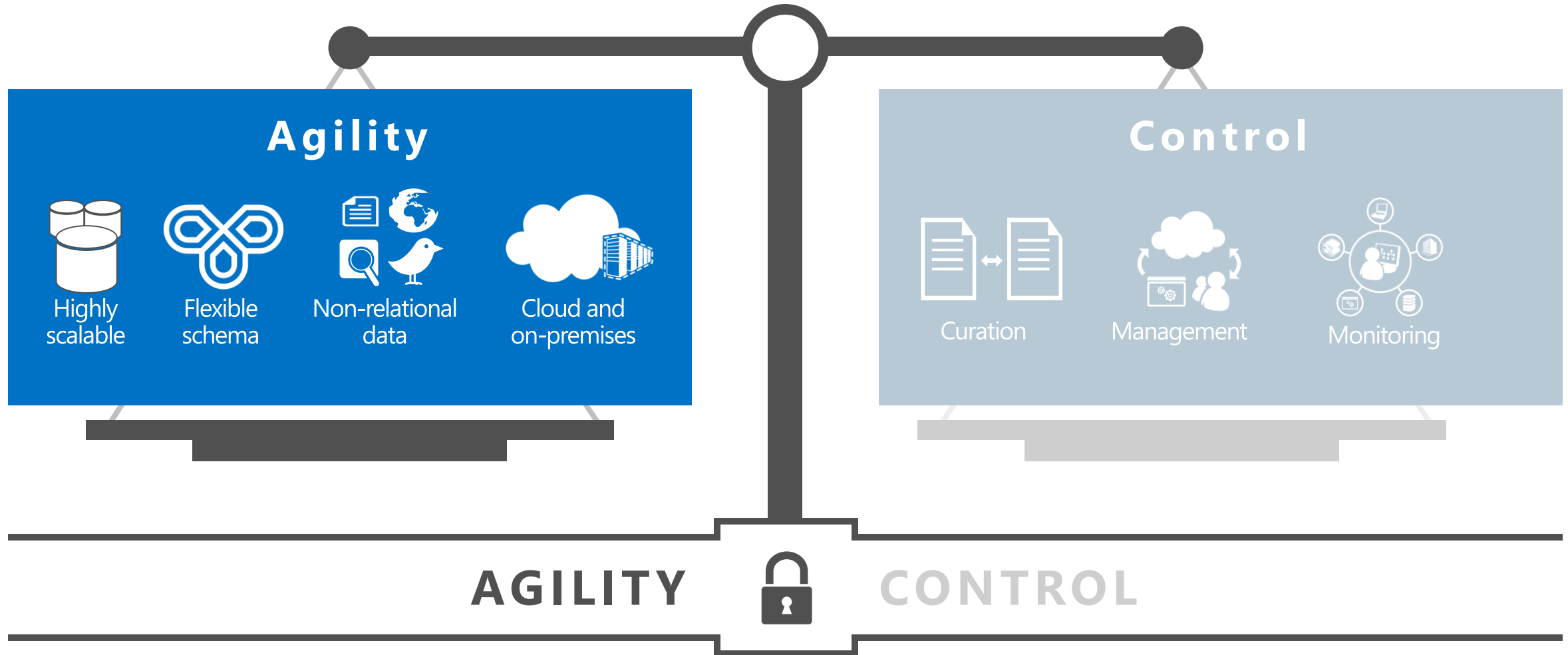


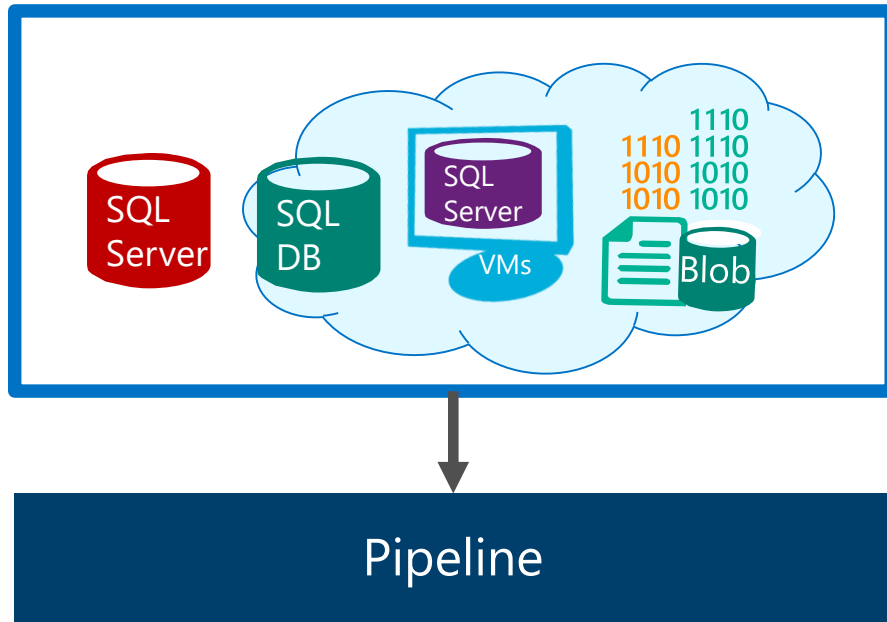
Azure Data Factory Components



Azure Data Factory

Produce trusted information from raw data





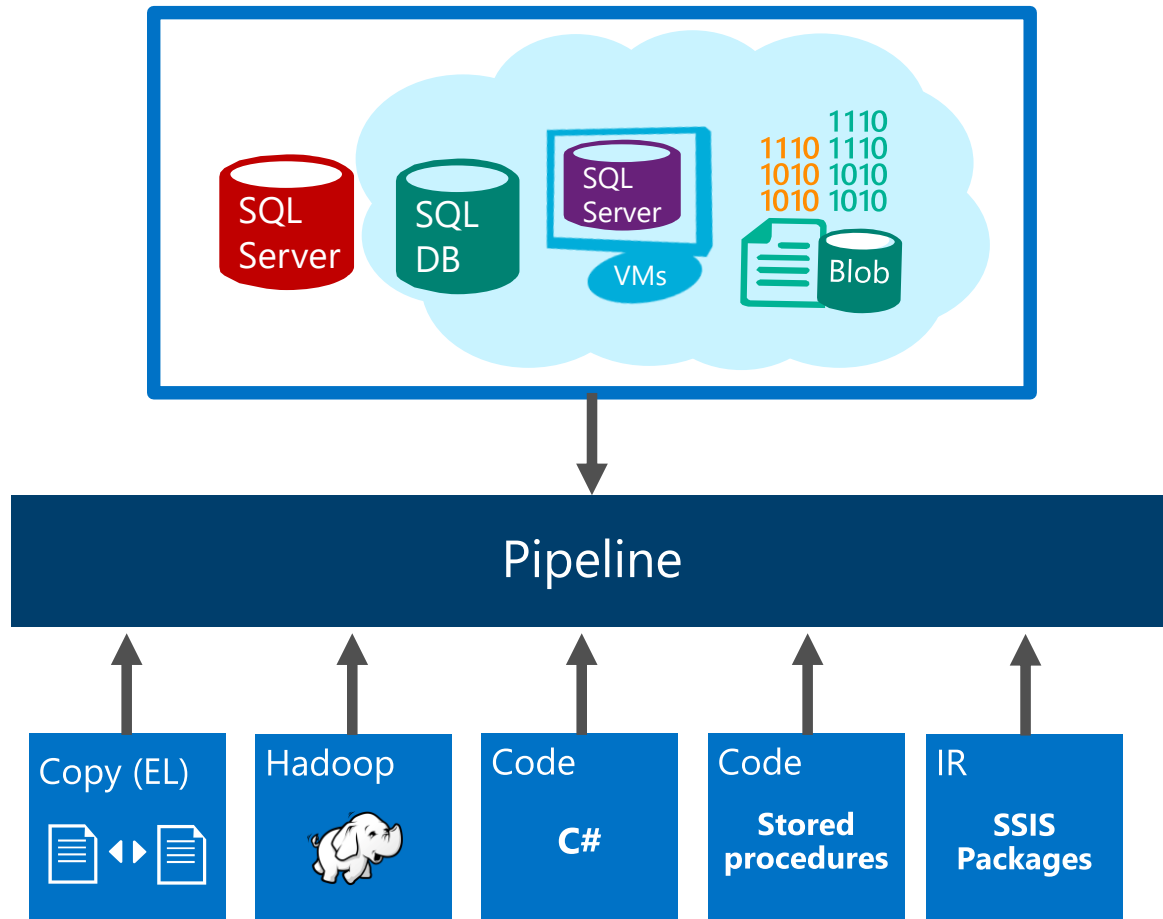
Produce trusted information from raw data

Connect to your data

- SQL Server on-premises
- SQL Server VMs, SQL Database, Azure blobs, tables
- Data movement (v2)

Gain agility on shapes and location

- Leverage highly scalable data stores with flexible schemas (Hadoop)
- Connect on-premises and cloud data
- Leverage relational and non-relational data



Compose processing to curate data

Processing on data

- Hadoop (Map/Reduce, Hive, Pig)
- C#
- Stored procedures
- Copy
- Activity dispatch

Azure-SSIS Integration Runtime (IR)

Data curation

- Join, aggregate, cleanse, enrich



Output trusted insights

Book1 - Excel

POWER QUERY SEARCH POWERPIVOT Michael Tejedor

S&P 500 Component Stocks - List of S&P 500 companies

List of S&P 500 companies - Wikipedia, the free encyclopedia

Ticker symbol	Company	SEC filings	GICS Sector	GICS Sub Industry
MMM	3M Co.	reports	Industrials	Industrial Conglomerate
ABT	Abbott Laboratories	reports	Health Care	Health Care Equipment
ABBV	AbbVie	reports	Health Care	Pharmaceuticals
ANF	Abercrombie & Fitch Company A	reports	Consumer Discretionary	Apparel, Accessories & L
ACE	ACE Limited	reports	Financials	Property & Casualty Insu
ACN	Accenture plc	reports	Information Technology	IT Consulting & Services
ACT	Actavis Inc	reports	Health Care	Pharmaceuticals
ADBE	Adobe Systems Inc	reports	Information Technology	Application Software
ADT	ADT Corp	reports	Industrials	Diversified Commercial S
AMD	Advanced Micro Devices	reports	Information Technology	Semiconductors
KSS	Kohl's Corp.	reports	Consumer Discretionary	General Merchandise St
KRFT	Kraft Foods Group	reports	Consumer Staples	Packaged Foods & Meat
LOW	Lowe's Cos.	reports	Consumer Discretionary	Home Improvement Ret
ISI	ISI Corporation	reports	Information Technology	Semiconductors

Columns [8]
Ticker symbol, Company, SEC filings, GICS Sector, GICS Sub Industry, Address of Headquarters, Date first added, Key
Last modified
Wednesday, August 28, 2013
From
Wikipedia
Data Source
http://en.wikipedia.org/wiki/List_of_S%26P_500_companies

ADD TO WORKSHEET FILTER & SHAPE DISCLAIMER

Online Search

S&P 500 Public

408 results

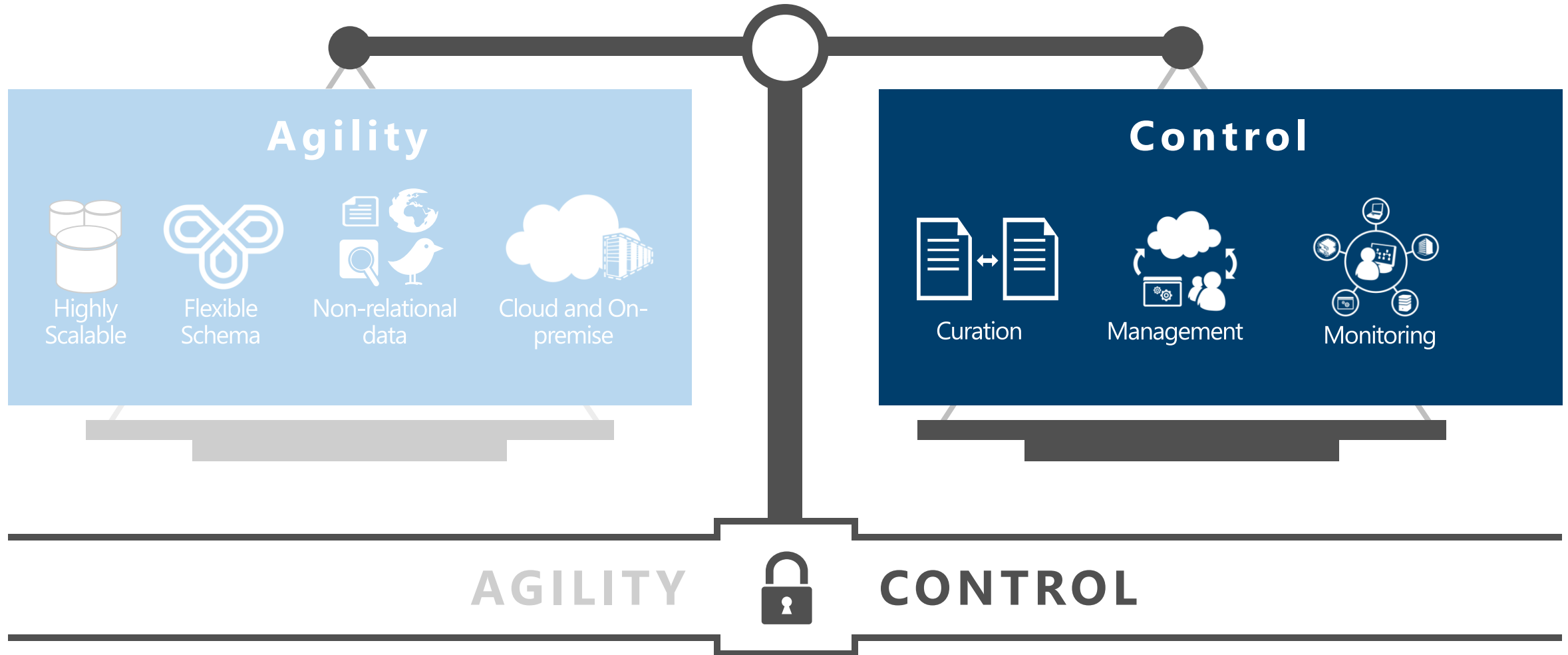
- Recent changes to the list of ...
From Wikipedia on Wednesday, August 28, 2013
List of S&P 500 companies - Wikipedia, the free encyclopedia
- S&P 500 Component Stocks ...
From Wikipedia on Wednesday, August 28, 2013
List of S&P 500 companies - Wikipedia, the free encyclopedia
- Investing - S&P 500
From Wikipedia on Wednesday, August 28, 2013
S&P 500 - Wikipedia, the free encyclopedia
- Index maintenance - S&P 500
From Wikipedia on Wednesday, August 28, 2013
S&P 500 - Wikipedia, the free encyclopedia
- Total annual returns - Market ...
From Wikipedia on Wednesday, August 28, 2013
S&P 500 - Wikipedia, the free encyclopedia
- Largest percentage changes ...

Consume data assets by BI tool or application

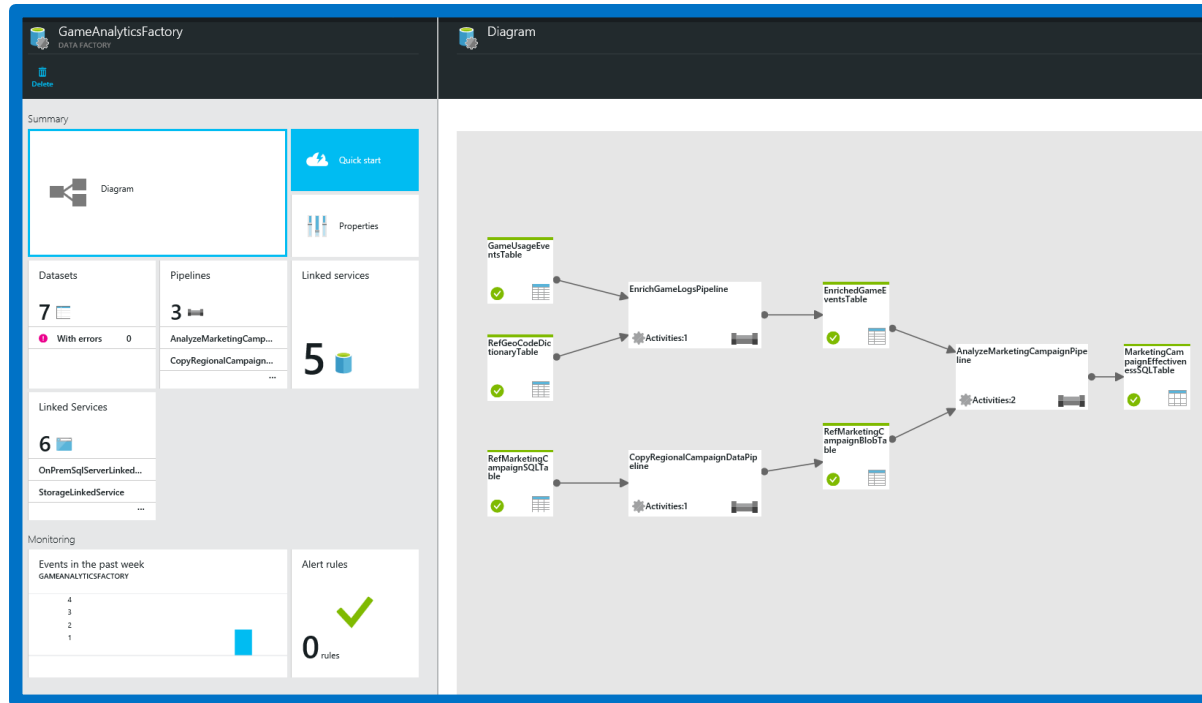
REST APIs for easy application integration

Azure Data Factory

Produce trusted information from raw data



Manage and monitor



Manage from single pane of glass

- No custom monitoring code required

Operations Management Suite

Set data production policy

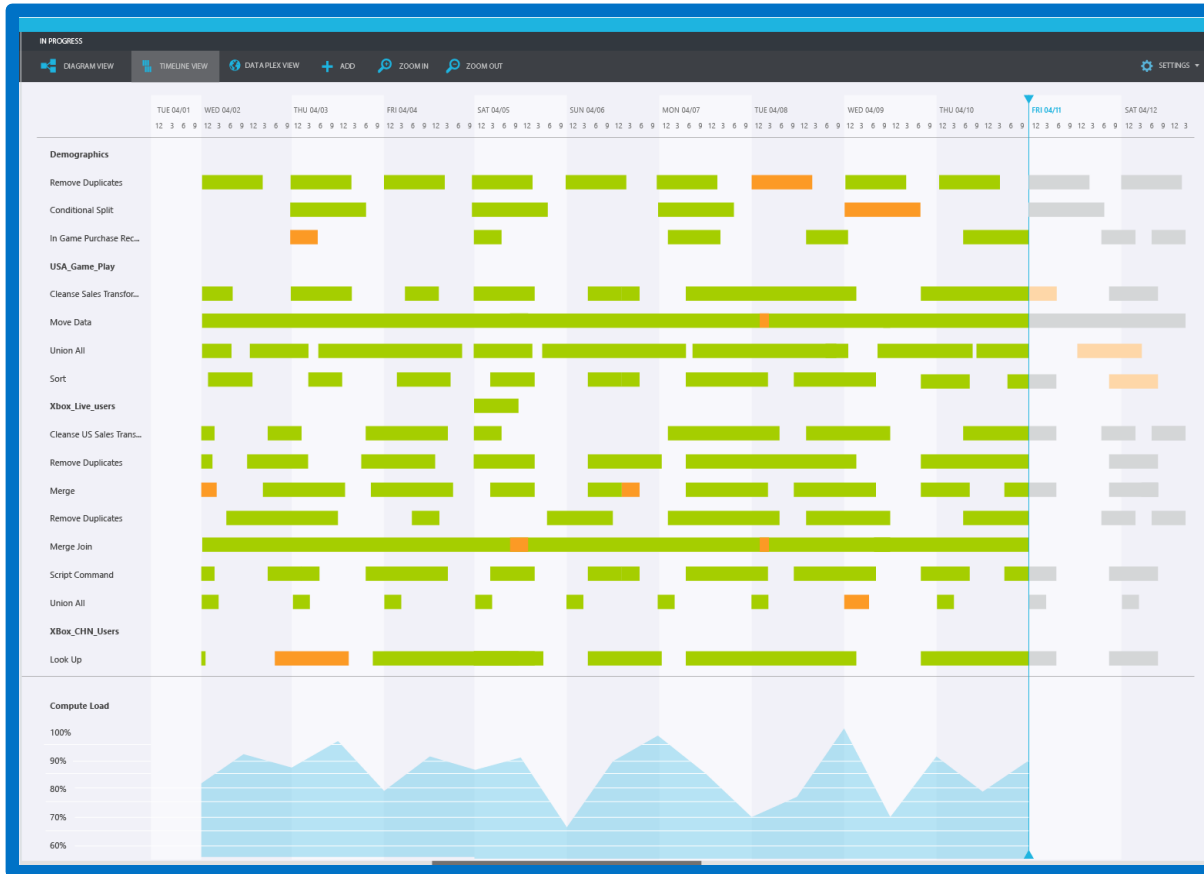
- Retry, concurrency, late data handling

Identify and debug errors

- Automatic data set health alerts
- Troubleshooting of complex pipelines

Fully scriptable

Rich scheduling



Rich scheduling controls

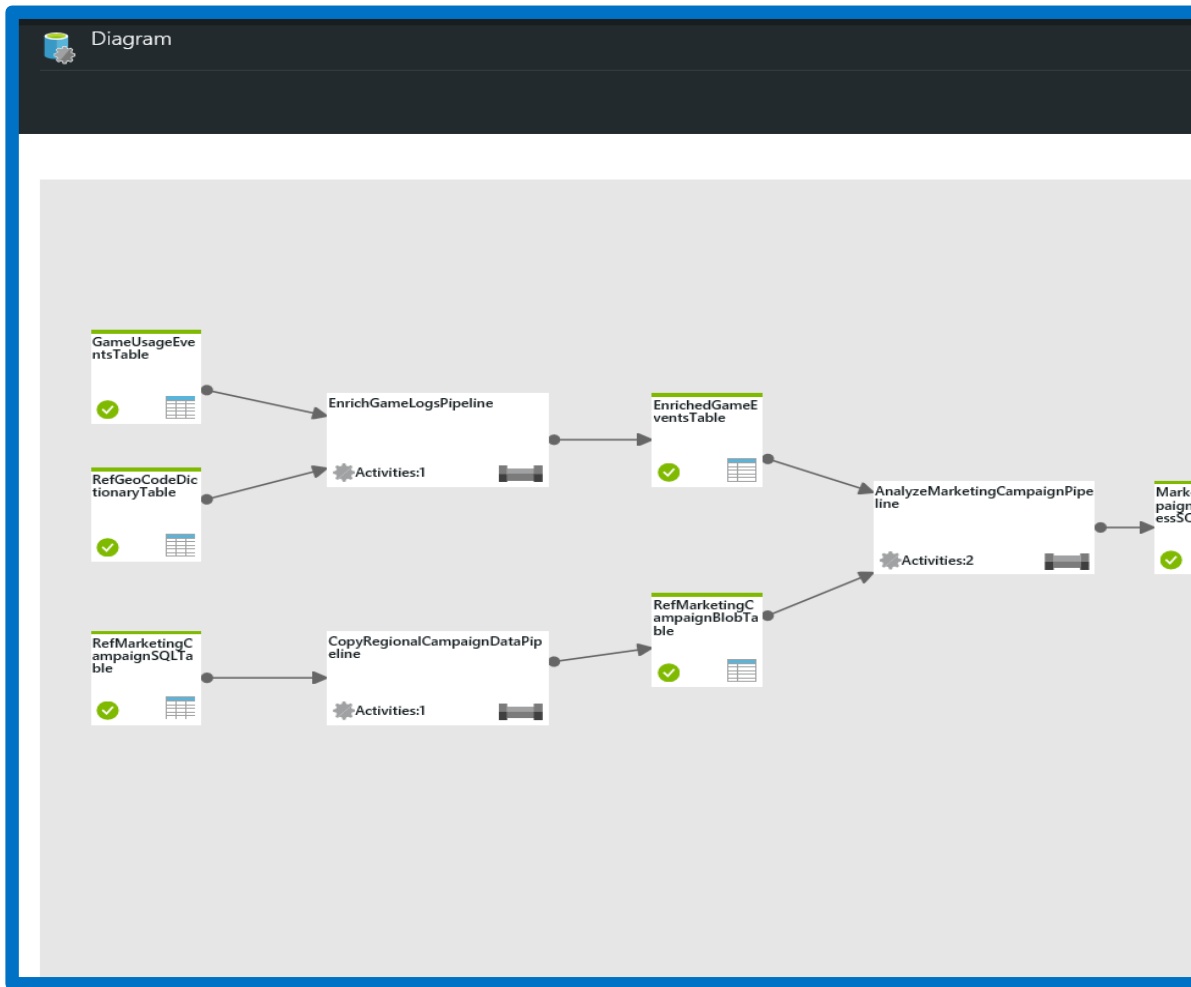
- Frequency: Minutes, hours, days, etc.
- Execute if precondition passes
- Easily chain pipelines running on differing schedules

Triggers

Easily restate previously produced data sets

- Rerun a single activity invocation
- Scheduled restatement with concurrency control (recalculate all data sets at once, etc.)

Lineage



Ability to trace data origins

- Understand who is consuming my data
- Determine how a field in my report was derived

Impact analysis

- Understand impact of changing data or processing

Why use Azure Data Factory?

Key tool to have a hybrid solution

Developers can use Data Factory to transform semi-structured, unstructured and structured data from on-premises and cloud sources into trusted information

Globally deployed data movement

Data Factory can access data stores and compute services in all Azure regions to move data between data stores or process data using compute services

Compatibility with SSIS

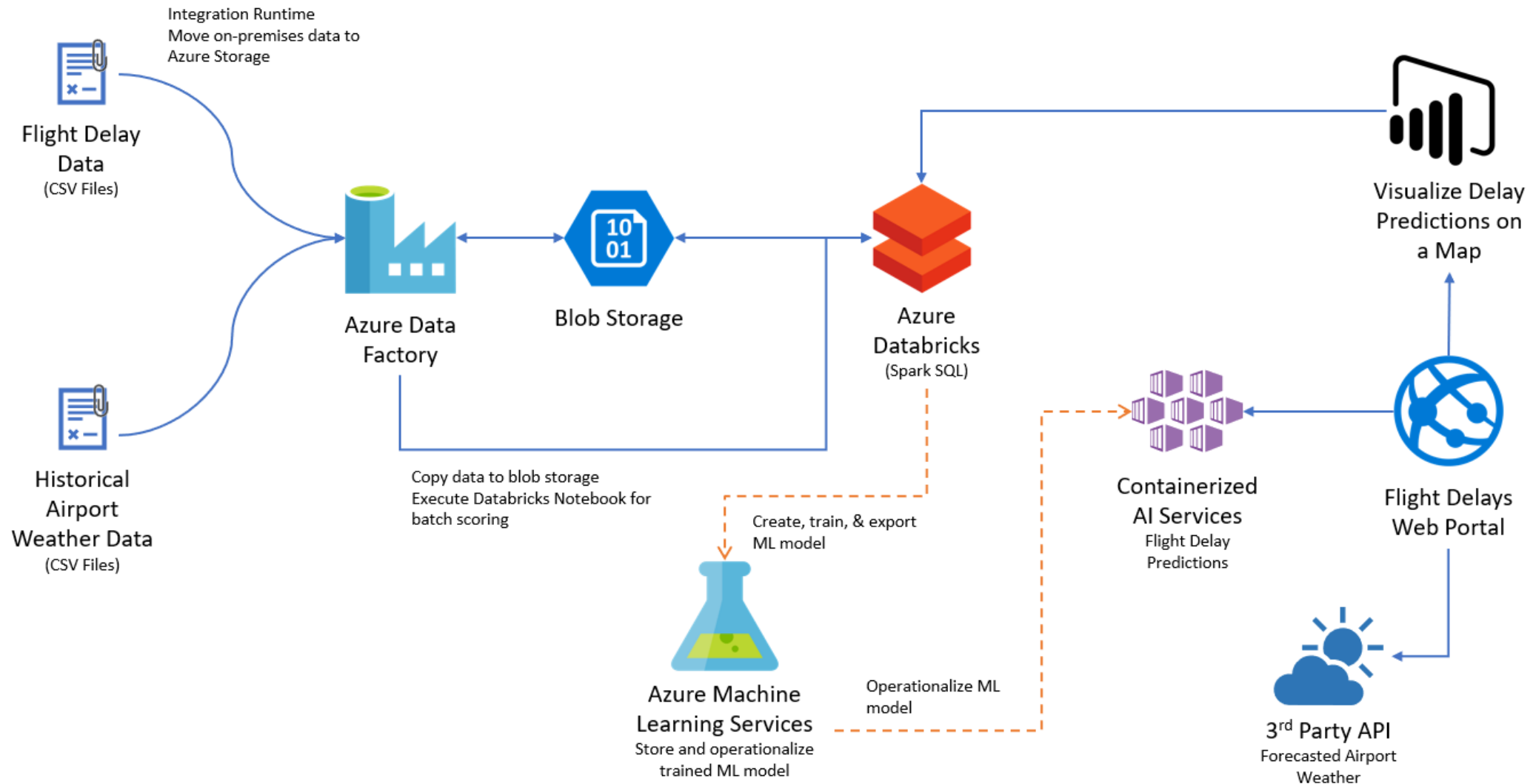
Three Integration Runtime types:

- Azure: Public Azure (PaaS)
- Self-hosted: On-premises machine or Azure VM (IaaS)
- Azure-SSIS: Can be provisioned in both public Azure network or private Azure VNet network

Lab Predictive modeling

- In this workshop, you will deploy a web app using Machine Learning Services to predict travel delays given flight delay data and weather conditions
- At the end of this workshop, you will be better able to build a complete machine learning model in Azure Databricks for predicting if an upcoming flight will experience delays
- In addition, you will learn to store the trained model in Azure Machine Learning Model Management, then deploy to Docker containers for scalable on-demand predictions, use Azure Data Factory (ADF) for data movement and operationalizing ML scoring, summarize data with Azure Databricks and Spark SQL, and visualize batch predictions on a map using Power BI

Solution architecture



Lab Predictive modeling

- Set up your environment
- Load Sample Data and Databricks Notebooks
- Setup Azure Data Factory
- Develop a data factory pipeline for data movement
- Operationalize ML scoring with Azure Databricks and Data Factory
- Summarize data using Azure Databricks

Questions

- Machine Learning practitioner
- Over 25 years of professional experience
- Artificial Intelligence MVP & MCT
- Microsoft Certified Solutions Expert
 - Data Management and Analytics
 - Cloud Platform and Infrastructure
 - Business Intelligence
- Microsoft Certified Solutions Developer
 - Azure Solution Architect

