

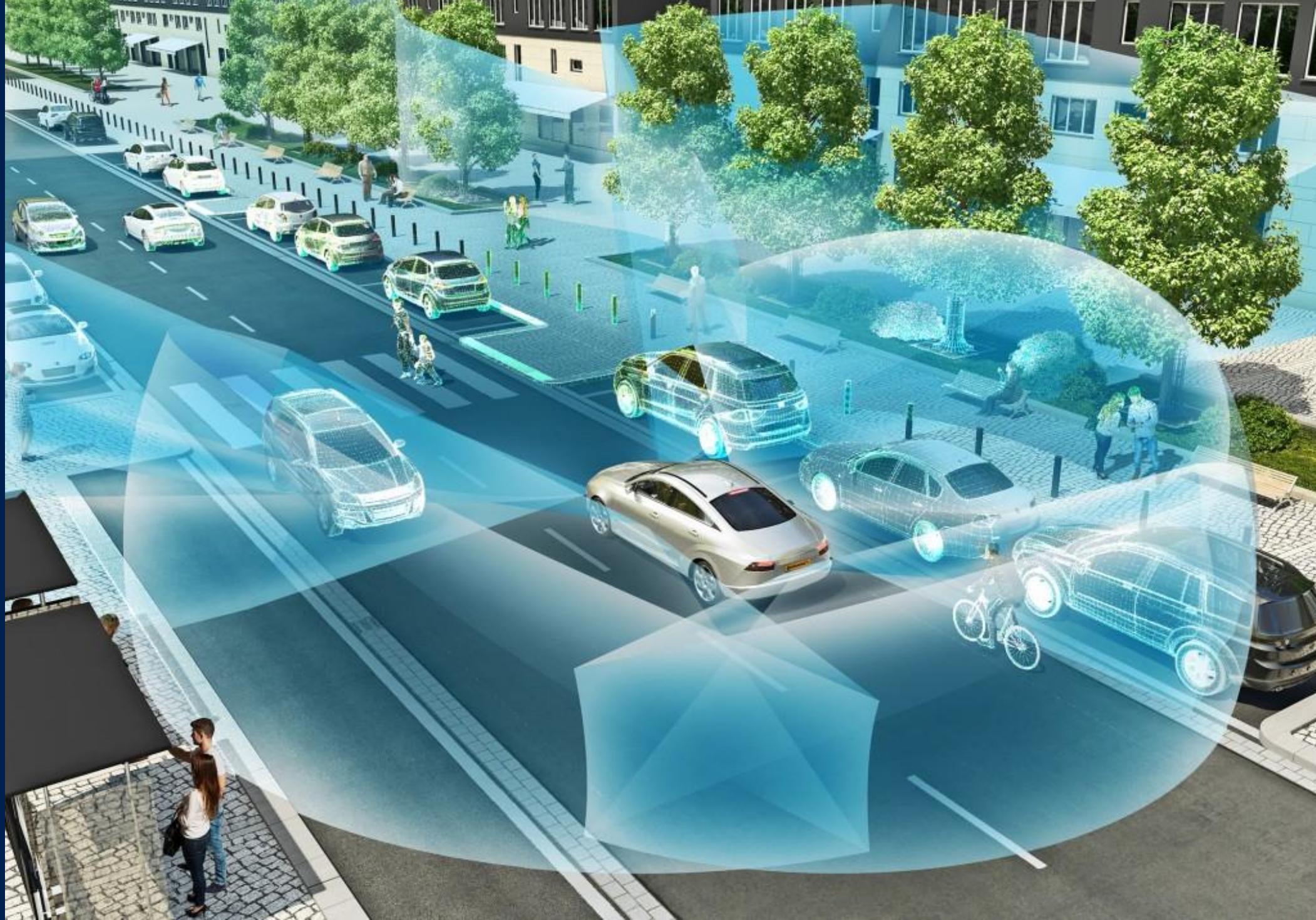


Met Azure naar Advanced Analytics

De weg naar digitale transformatie

RDW Techday, 14 Maart 2017

Ruben Pool, Data Solution Architect





Azure Introduction

Azure = datacenters



Happening in our backyard (Middenmeer)





Achieve global scale, in local regions





Any tool, application, framework

Build and run open source solutions

Applications



WORDPRESS



Drupal



OPENSHIFT
ENTERPRISE
by Red Hat®



CLOUD
FOUNDRY™



Joomla!

DevOps



CHEF™



A NSIBLE



Frameworks



.NET



node.js™



python



Java



php



Databases



cassandra

Ask Bigger Questions



Hortonworks



MariaDB



cassandra

Containers



mesos



docker



DC/OS



Core OS

Infrastructure



DC/OS



redhat



consensys



blockstack

ubuntu



docker



FreeBSD

Trustworthy



ISO 27001



SOC 1 Type 2



SOC 2 Type 2



PCI DSS Level 1



Cloud Controls
Matrix



ISO 27018



Content Delivery and
Security Association



Shared
Assessments



FedRAMP
JAB P-ATO



HIPAA/
HITECH



FIPS 140-2



21 CFR
Part 11



FERPA



DISA Level 2



CJIS



IRS 1075
ITAR-ready



Section
508 VPAT



European
Union Model
Clauses



EU Safe
Harbor



United Kingdom
G-Cloud



China Multi
Layer Protection
Scheme



China
GB 18030



China
CCCPPF



Singapore
MTCS Level 3



Australia
Signals
Directorate



New
Zealand
GCIO



Japan
Financial
Services



ENISA
IAF

Platform Services

Security & Management

Portal

Azure Active Directory

Azure AD B2C

Multi-Factor Authentication

Automation

Scheduler

Key Vault

Store/ Marketplace

VM Image Gallery & VM Depot

Services Compute

Cloud Services

Service Fabric

Batch

RemoteApp

Integration

Storage Queues

BizTalk Services

Hybrid Connections

Service Bus

Media & CDN

Media Services

Content Delivery Network (CDN)

Web and Mobile

Web Apps

API Apps

Mobile Apps

Logic Apps

API Management

Notification Hubs

Developer Services

Visual Studio

Azure SDK

VS Online

App Insights

Data

SQL Database

Data Warehouse

DocumentDB

Redis Cache

Azure Search

Storage Tables

Analytics & IoT

HDIInsight

Machine Learning

Stream Analytics

Data Lake

Data Factory

Event Hubs

Data Catalog

IoT Hub

Mobile Engagement

Hybrid Operations

Azure AD Health Monitoring

AD Privileged Identity Management

Domain Services

Backup

Operational Analytics

Import/Export

Azure Site Recovery

StorSimple

Infrastructure Services

OS/Server Compute

Virtual Machines

Container Service

Storage

BLOB Storage

Azure Files

Premium Storage

Networking

Virtual Network

Load Balancer

DNS

Express Route

Traffic Manager

VPN Gateway

App Gateway

Datacenter Infrastructure (30 Regions, 22 Online)





Microsoft

THE OCEAN
CLEANUP

www.theoceancleanup.com

How Microsoft Azure Helps Cleaning the Ocean

29 November 2016

Bruno Sainte-Rose, Lead Computational Modeler

Hendrik Wrenger, Computational Modeler

[Refresh](#)[+ Add job](#)[Scale](#)[Delete](#)

Resizing

Current nodes

0

Target nodes

100

VM size

standard_a1

Allocation state

Resizing

Summary

No items in heat map

IDLE	0
RUNNING	0
CREATING	0
STARTING	0
REBOOTING	0
WAITING FOR S...	0
START TASK FAI...	0
LEAVING POOL	0
REIMAGING	0
OFFLINE	0
UNKNOWN	0

[Refresh](#)[Add job](#)[Scale](#)[Delete](#)

Current nodes

100

Target nodes

100

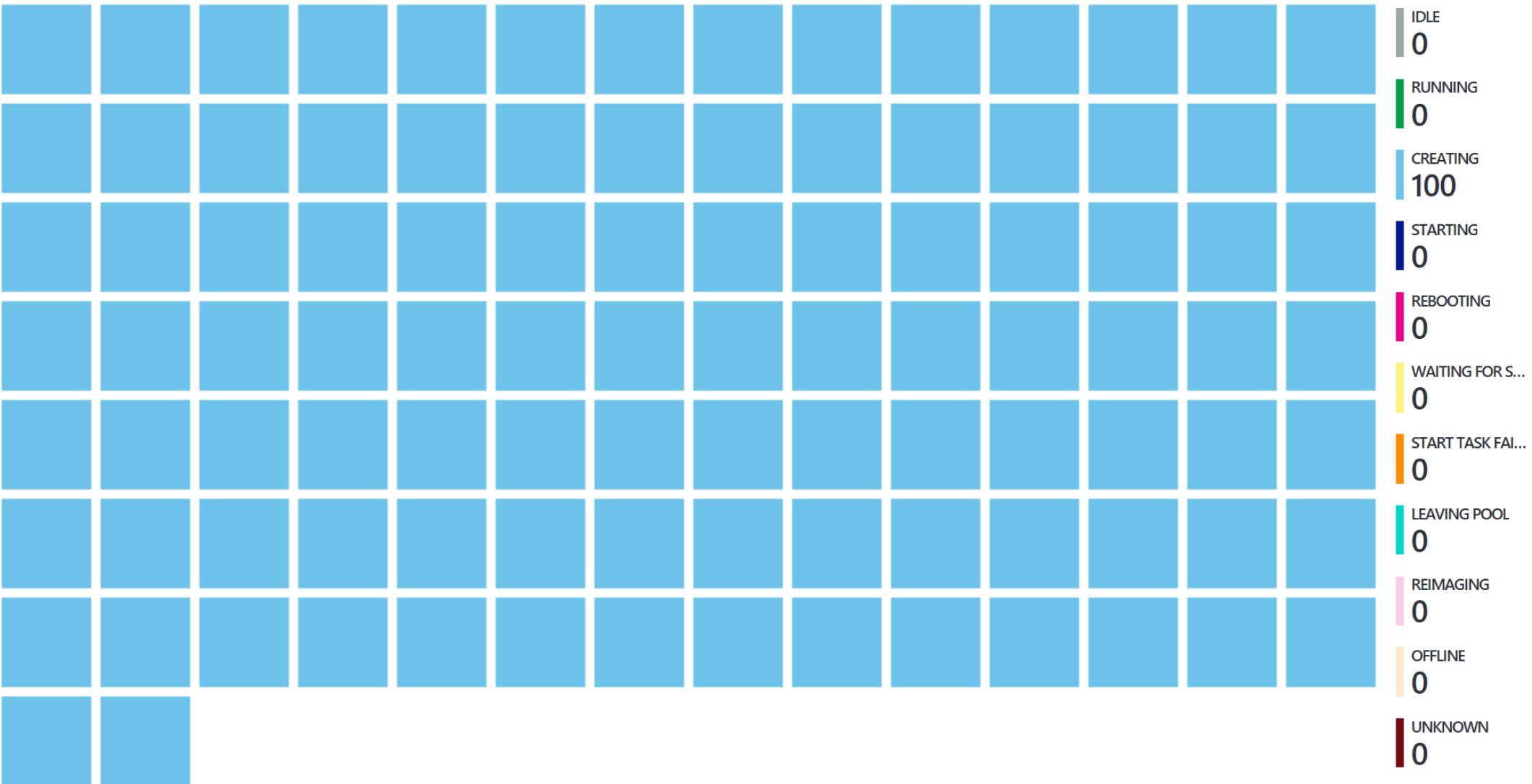
VM size

standard_a1

Allocation state

Steady

Summary



[Refresh](#)[Add job](#)[Scale](#)[Delete](#)

Essentials ^

Current cores

100

Current nodes

100

Target nodes

100

Operating System

Windows Server 2012 R2 *

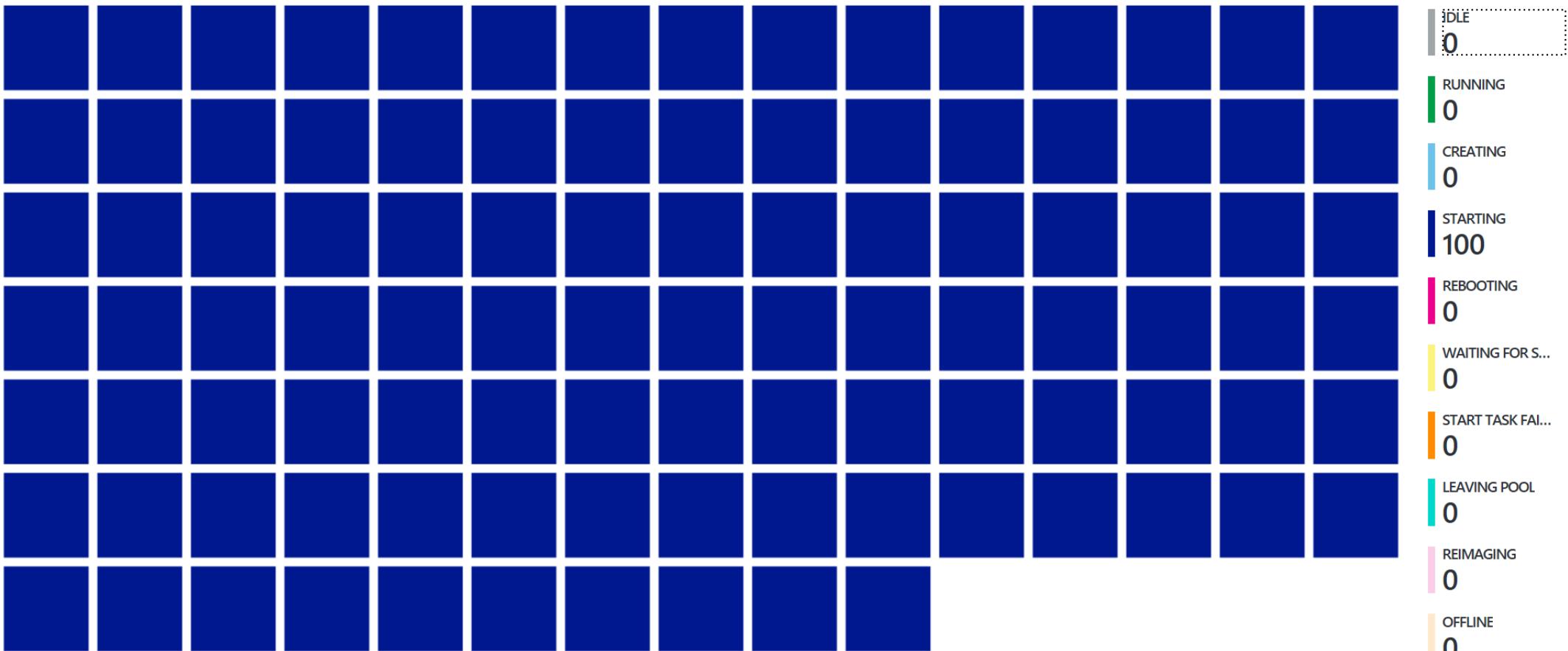
VM size

standard_a1

Allocation state

Steady

Summary



[Refresh](#)[Add job](#)[Scale](#)[Delete](#)

Essentials ^

Current cores

100

Current nodes

100

Target nodes

100

Operating System

Windows Server 2012 R2 *

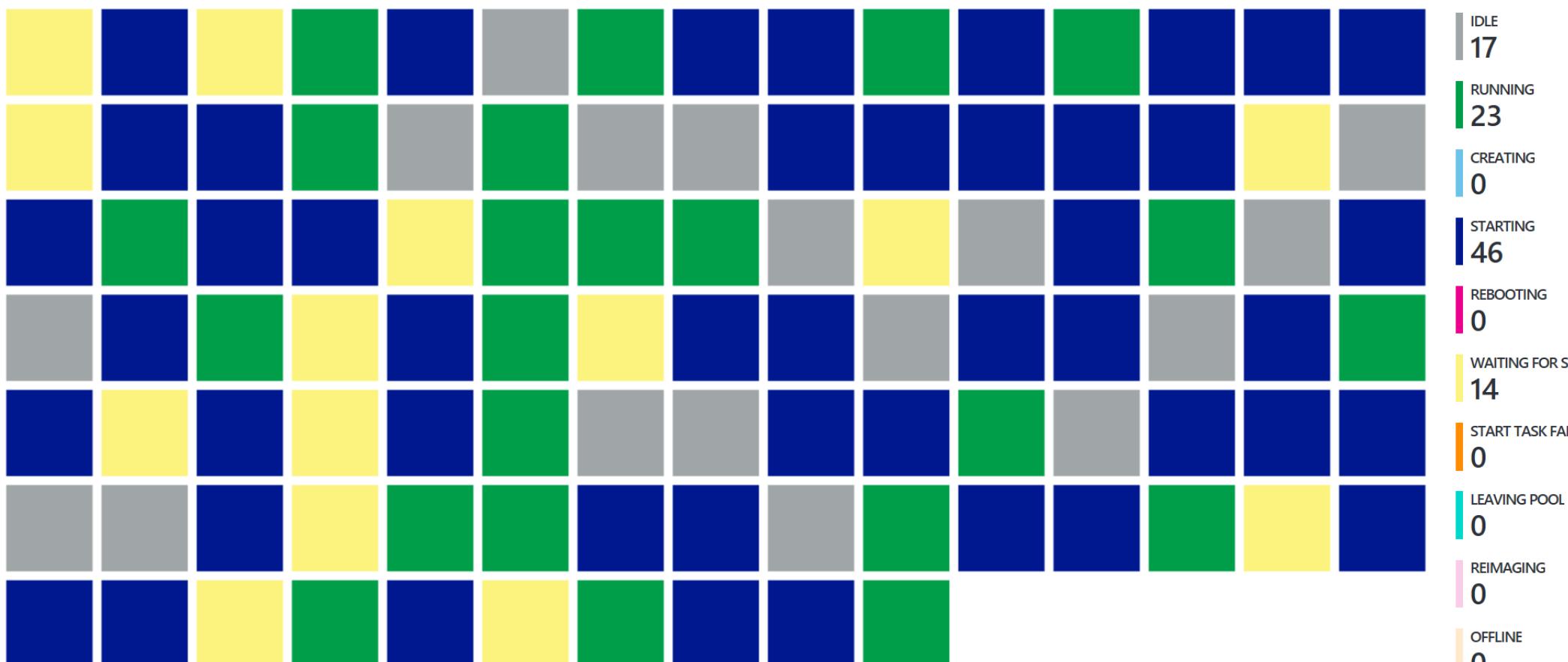
VM size

standard_a1

Allocation state

Steady

Summary



[Refresh](#)[+ Add job](#)[Scale](#)[Delete](#)

Essentials ^

Current cores

100

Current nodes

100

Target nodes

100

Operating System

Windows Server 2012 R2 *

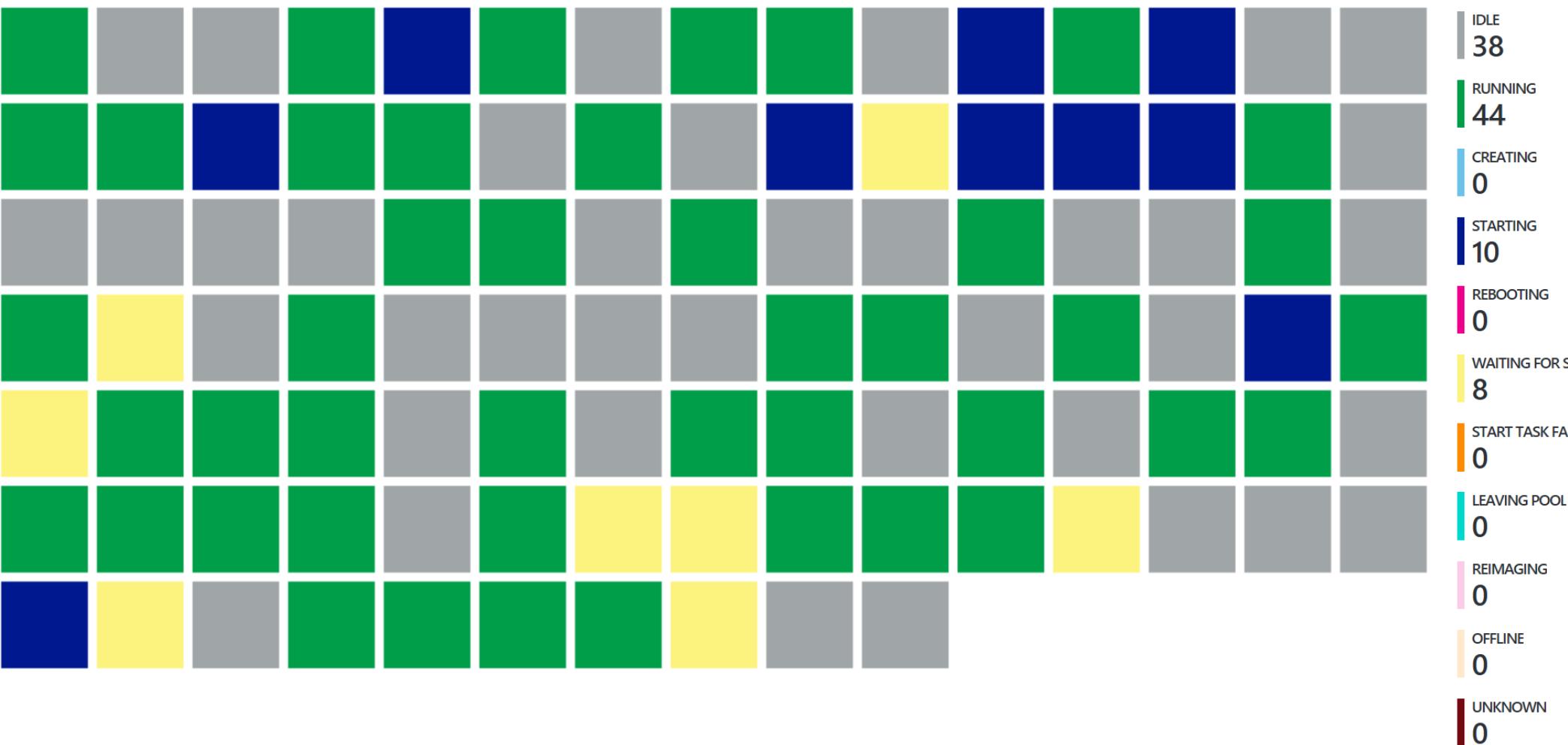
VM size

standard_a1

Allocation state

Steady

Summary





Building solutions
in Azure

Platform Services

Security & Management



Portal



Azure Active Directory



Azure AD B2C



Multi-Factor Authentication



Automation



Scheduler



Key Vault



Store/ Marketplace



VM Image Gallery & VM Depot

Services Compute



Cloud Services



Service Fabric



Batch



RemoteApp

Web and Mobile



Web Apps



API Apps



Mobile Apps



Logic Apps



API Management



Notification Hubs

Data



SQL Database



Data Warehouse



DocumentDB



Redis Cache



Azure Search



Storage Tables

Analytics & IoT



HDInsight



Machine Learning



Stream Analytics



Data Lake



Data Factory



Event Hubs



Data Catalog



IoT Hub

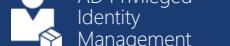


Mobile Engagement

Hybrid Operations



Azure AD Health Monitoring



AD Privileged Identity Management



Domain Services



Backup



Operational Analytics



Import/Export



Azure Site Recovery



StorSimple

OS/Server Compute



Virtual Machines



Container Service

Storage



BLOB Storage



Azure Files



Premium Storage

Infrastructure Services

Networking



Virtual Network



Load Balancer



DNS



Express Route



Traffic Manager



VPN Gateway



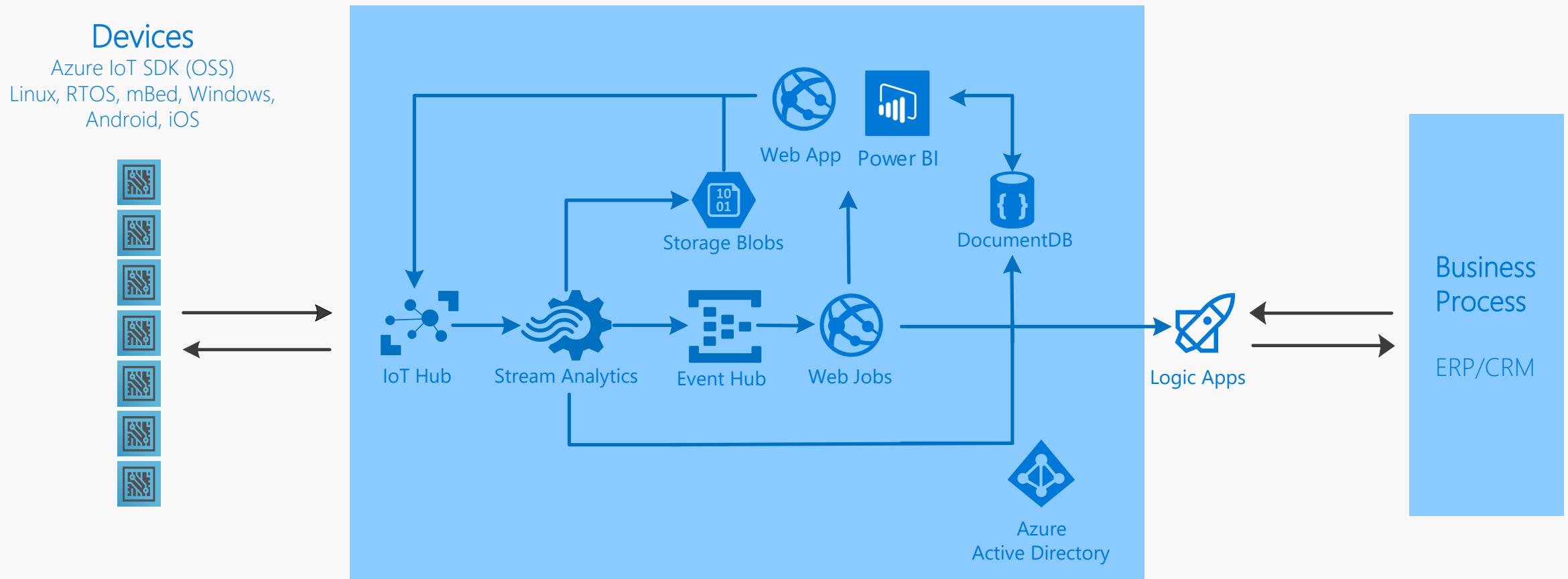
App Gateway

Datacenter Infrastructure (30 Regions, 22 Online)





Example: Azure IoT Suite.

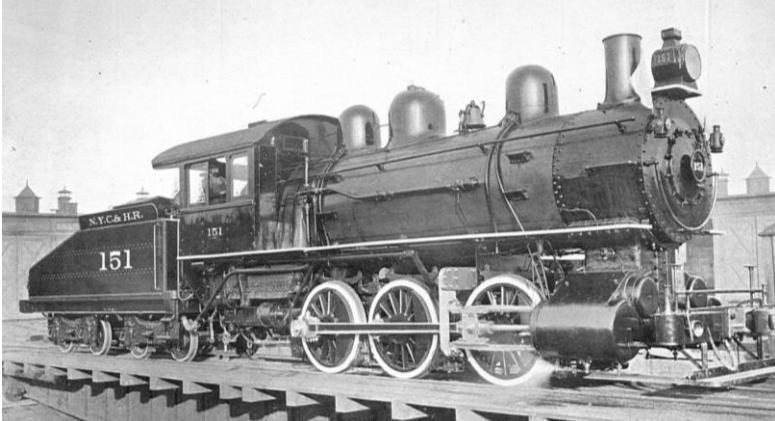


Introducing Cortana Intelligence



Some innovations fundamentally disrupt the economic landscape

1st Industrial Revolution



2nd Industrial Revolution



3rd Industrial Revolution

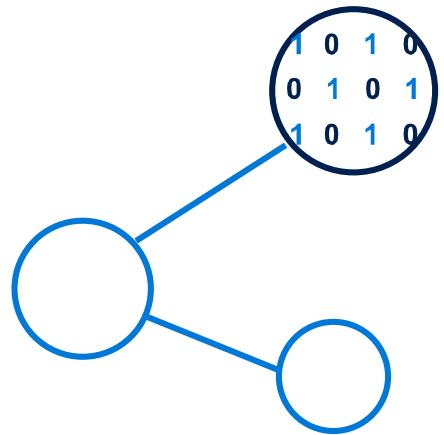


1760

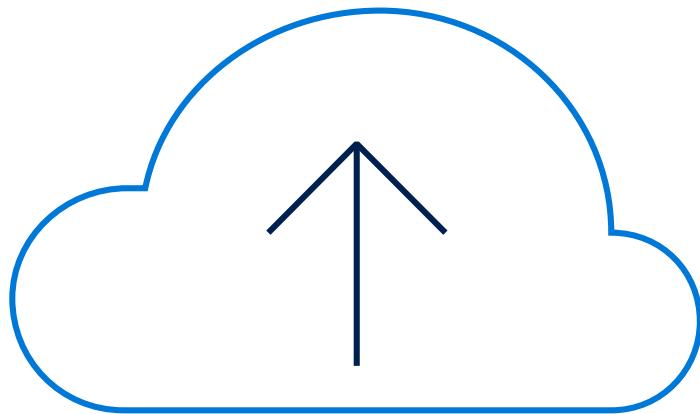
1870

1950

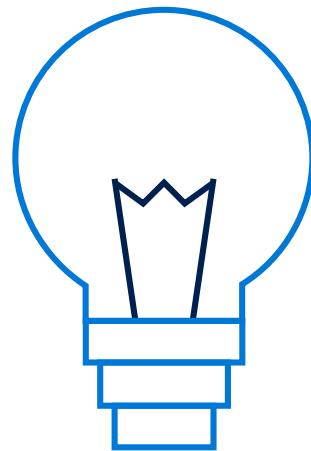
We now face the next disruption –
the fourth industrial revolution



Big Data

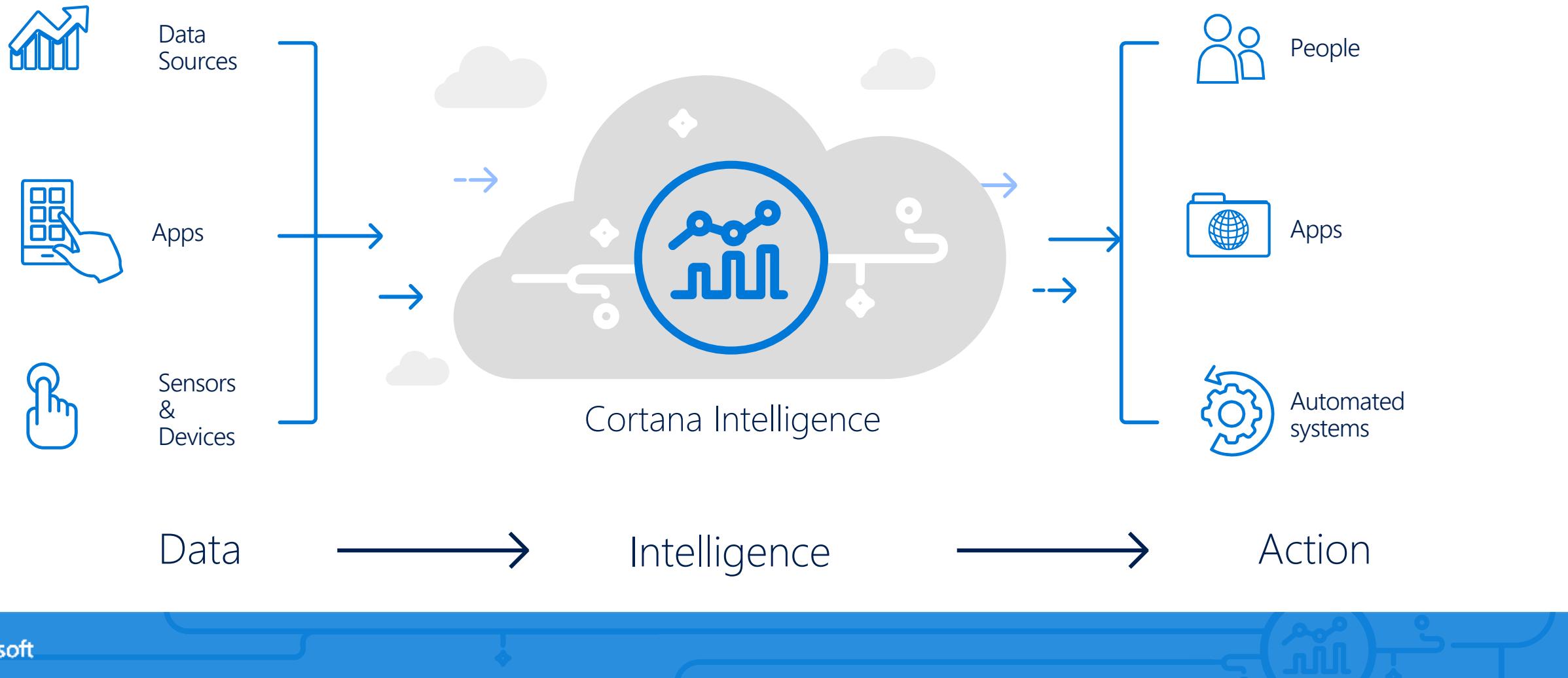


Cloud

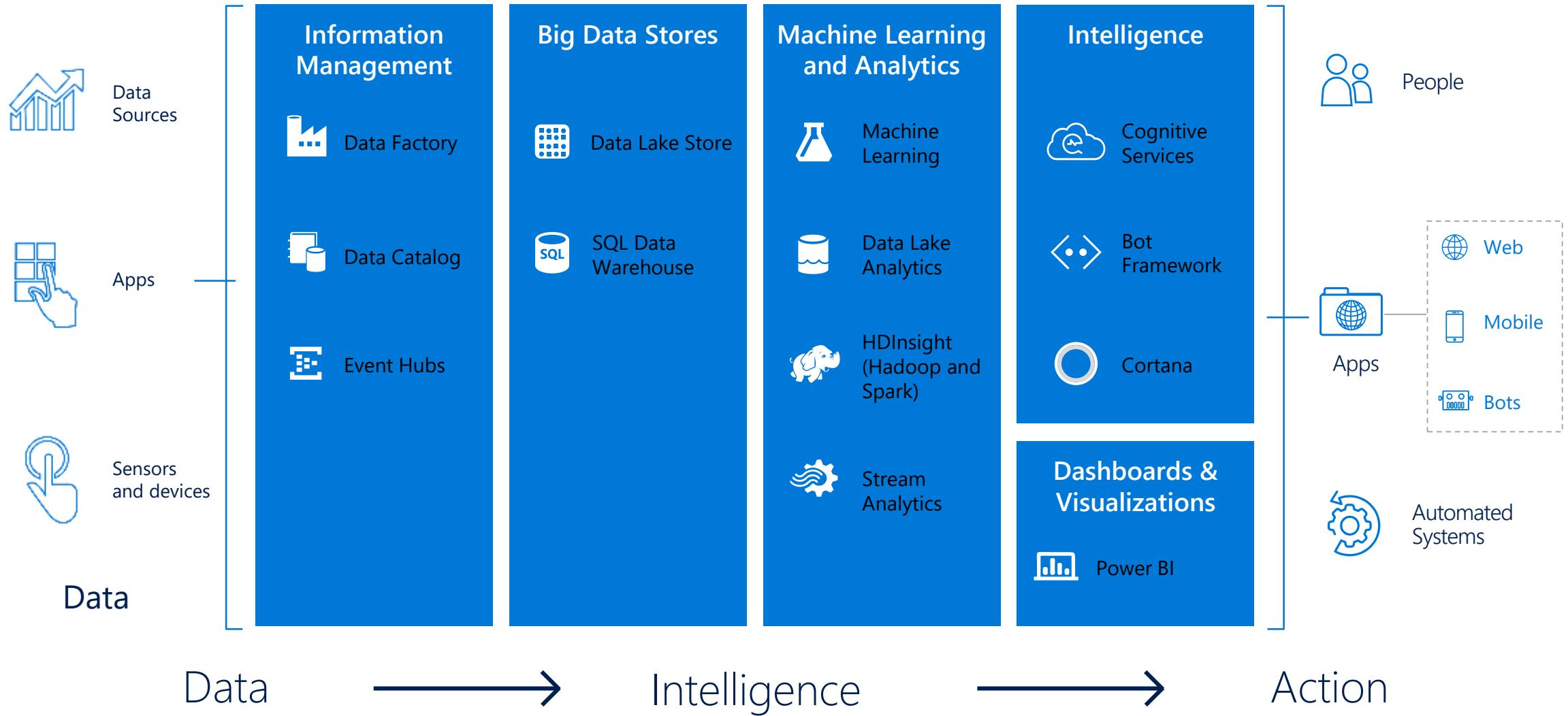


Intelligence

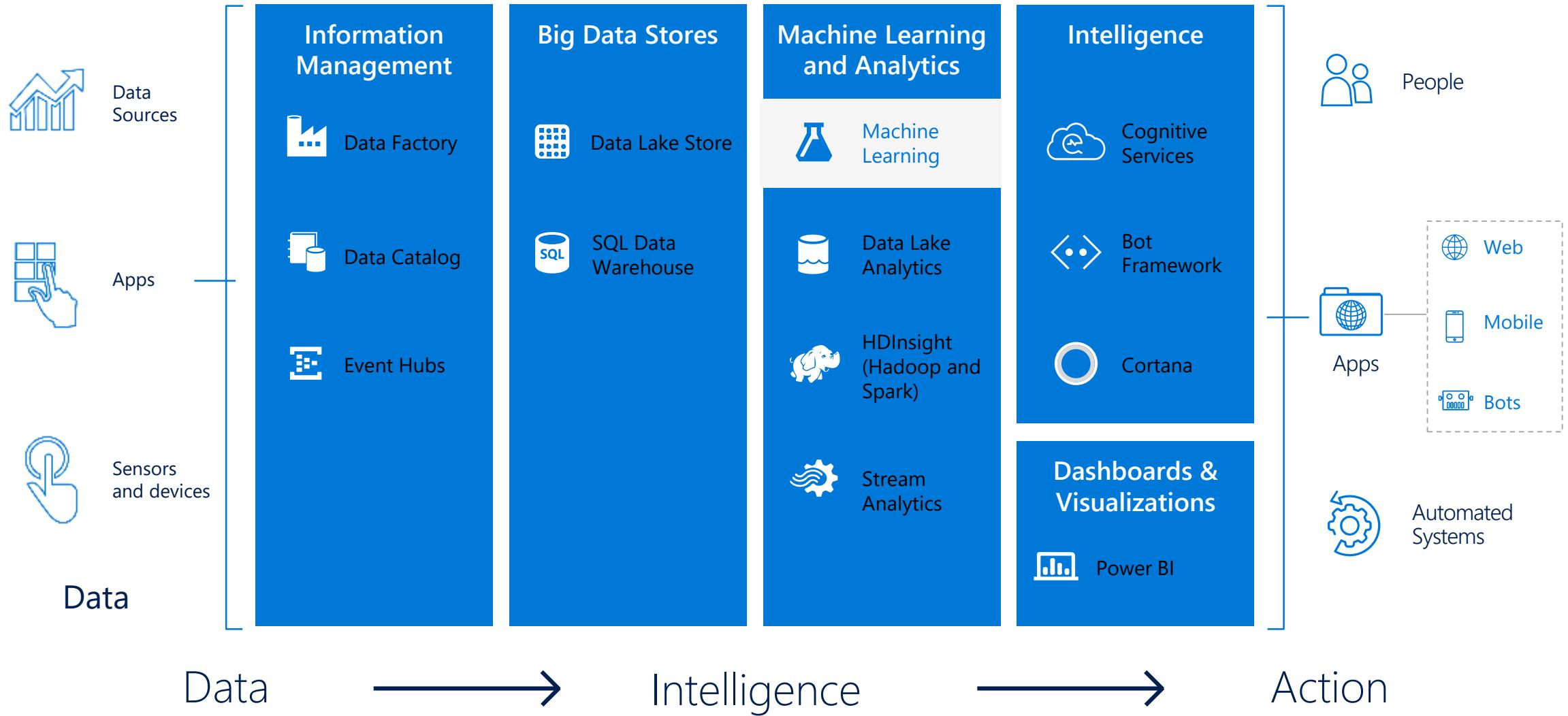
Transform data into intelligent action with Cortana Intelligence



Transform data into intelligent action



Transform data into intelligent action



Easily build, deploy, and share predictive analytics solutions

Machine Learning and Analytics



Machine
Learning



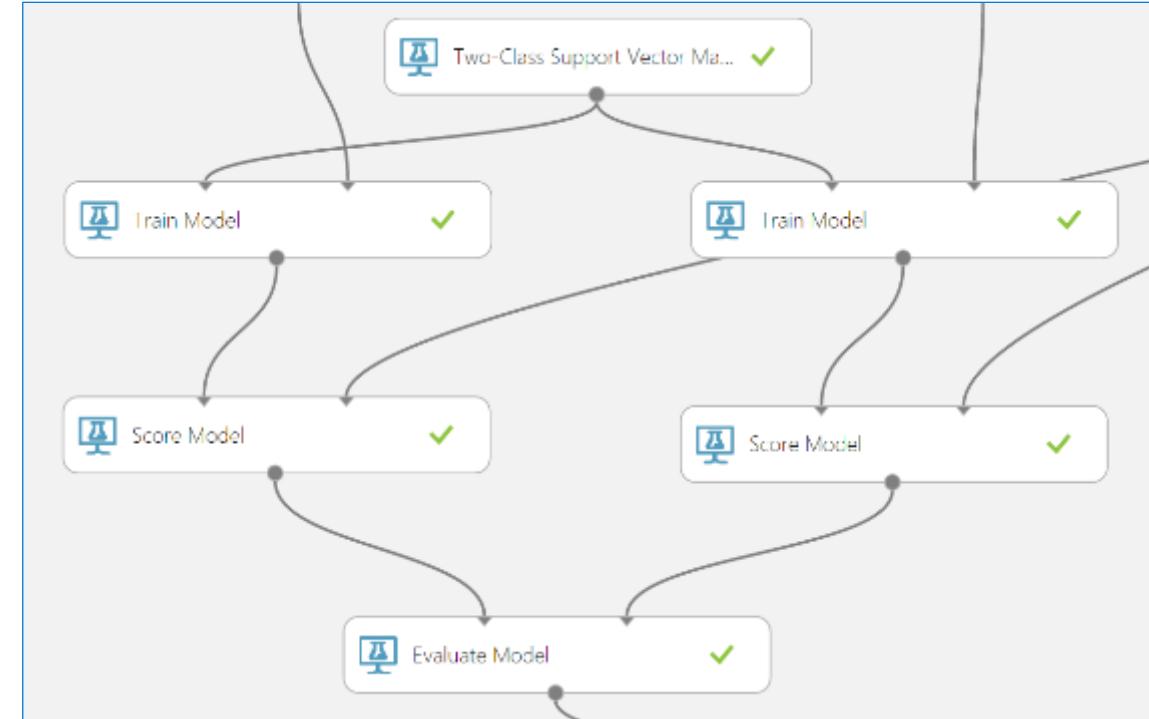
Data Lake
Analytics

HDInsight (Hadoop and Spark)



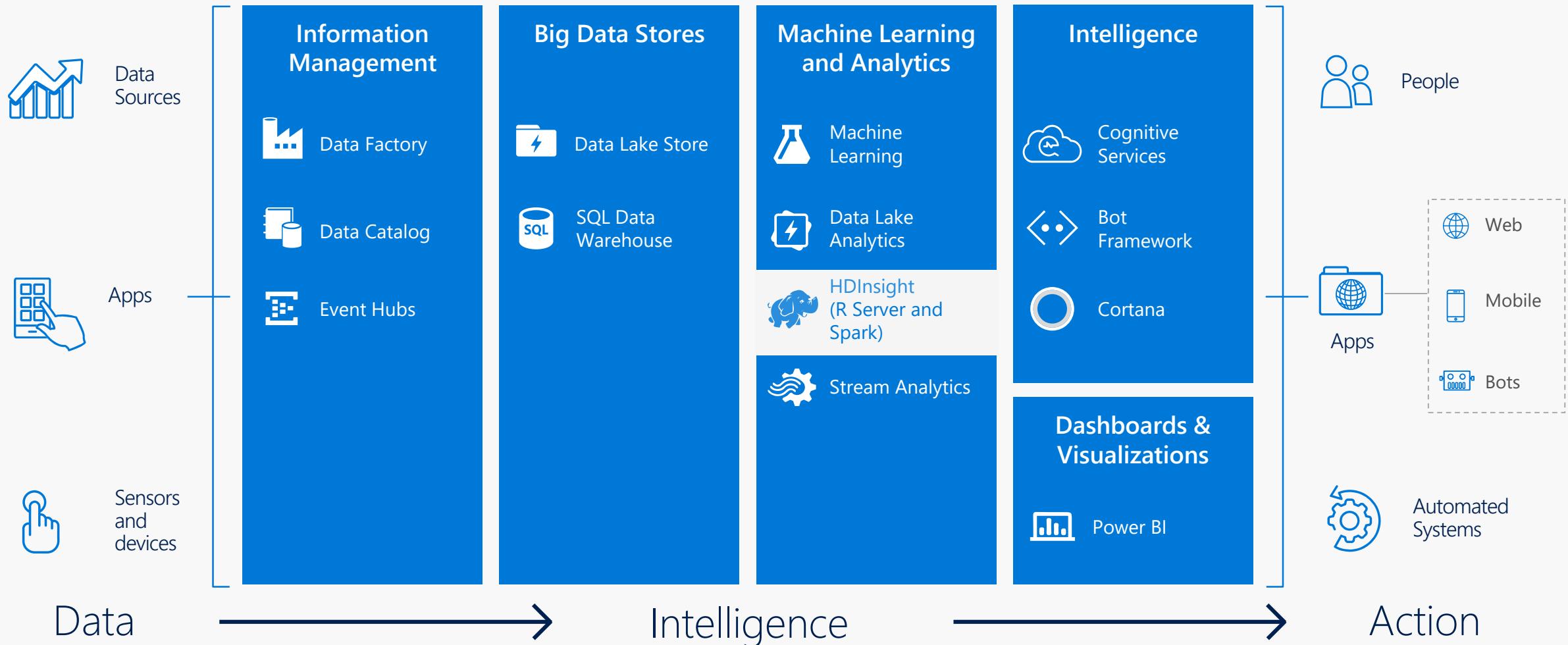
Stream
Analytics

The screenshot shows the Cortana Analytics Gallery interface. At the top, there are navigation links: 'Browse all' (highlighted in green), 'Solution Templates', and 'Experiments'. Below this, there's a 'Refine by' section with dropdown menus for 'CATEGORIES' and 'SHOW'. Under 'CATEGORIES', options like 'Solution Template', 'Experiment', and 'Machine Learning API' are listed. Under 'SHOW', there's a checkbox for 'Microsoft content only'. Below these are 'TAGS' filters for 'R', 'Classification', 'DA1203x', and 'classification'. The main area is titled 'Results' and displays two items: 'Face APIs' and 'Text Analytics'. Each item has a thumbnail image, a title, a brief description, and a timestamp ('1071687 · 7 months ago').

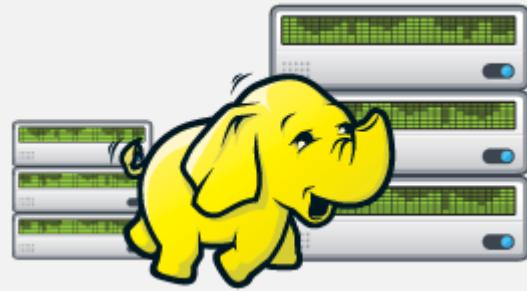


Scenario's: education,

Cortana Intelligence Suite



Let's Talk About Hadoop



Apache Open Source Project
Highly scalable distributed file system (HDFS)
Distributed processing on data nodes

Data Volumes



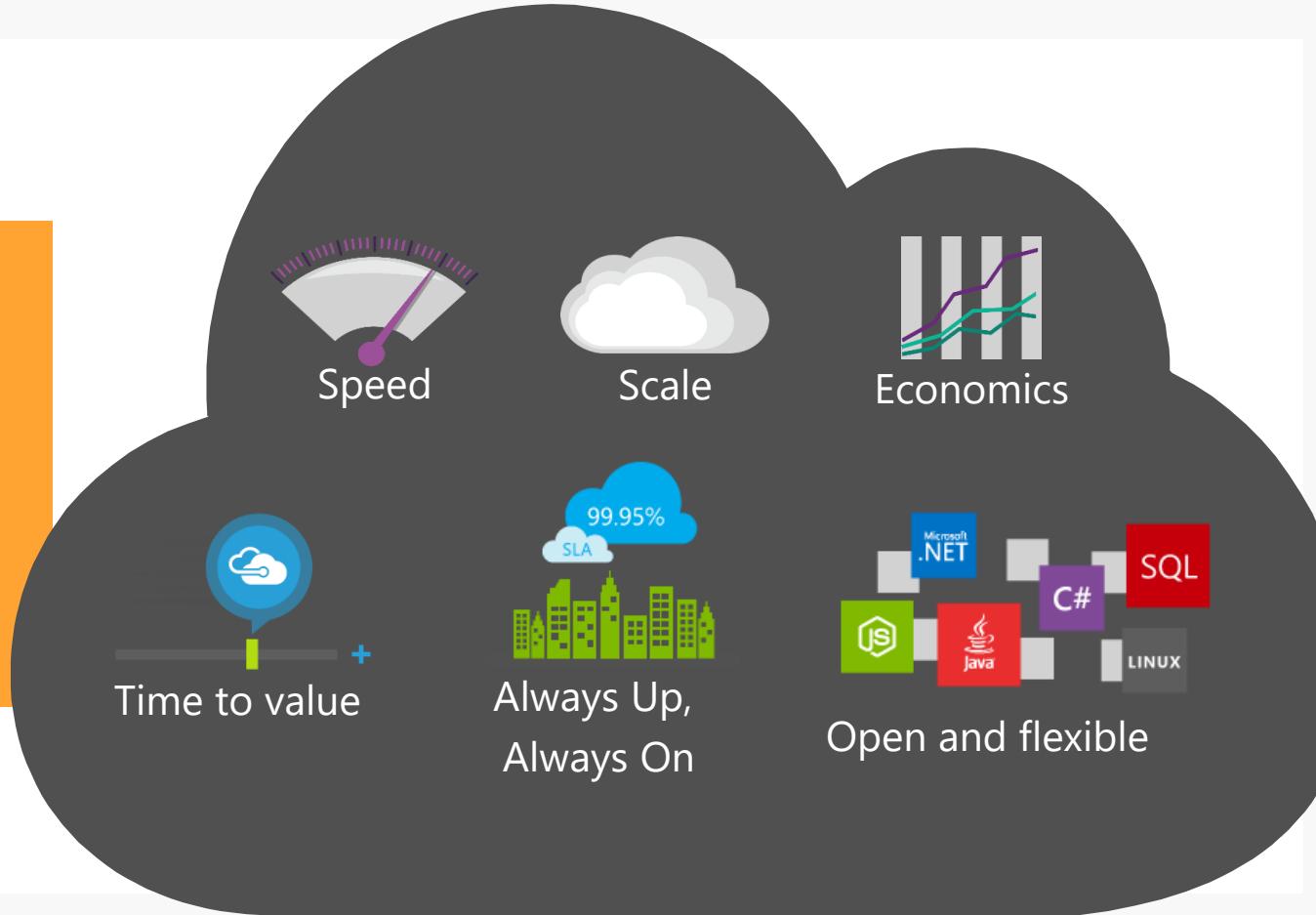
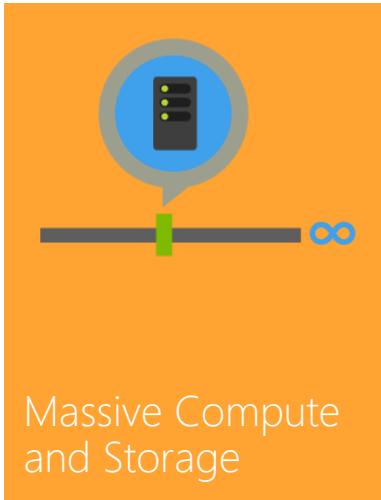
Data Variety



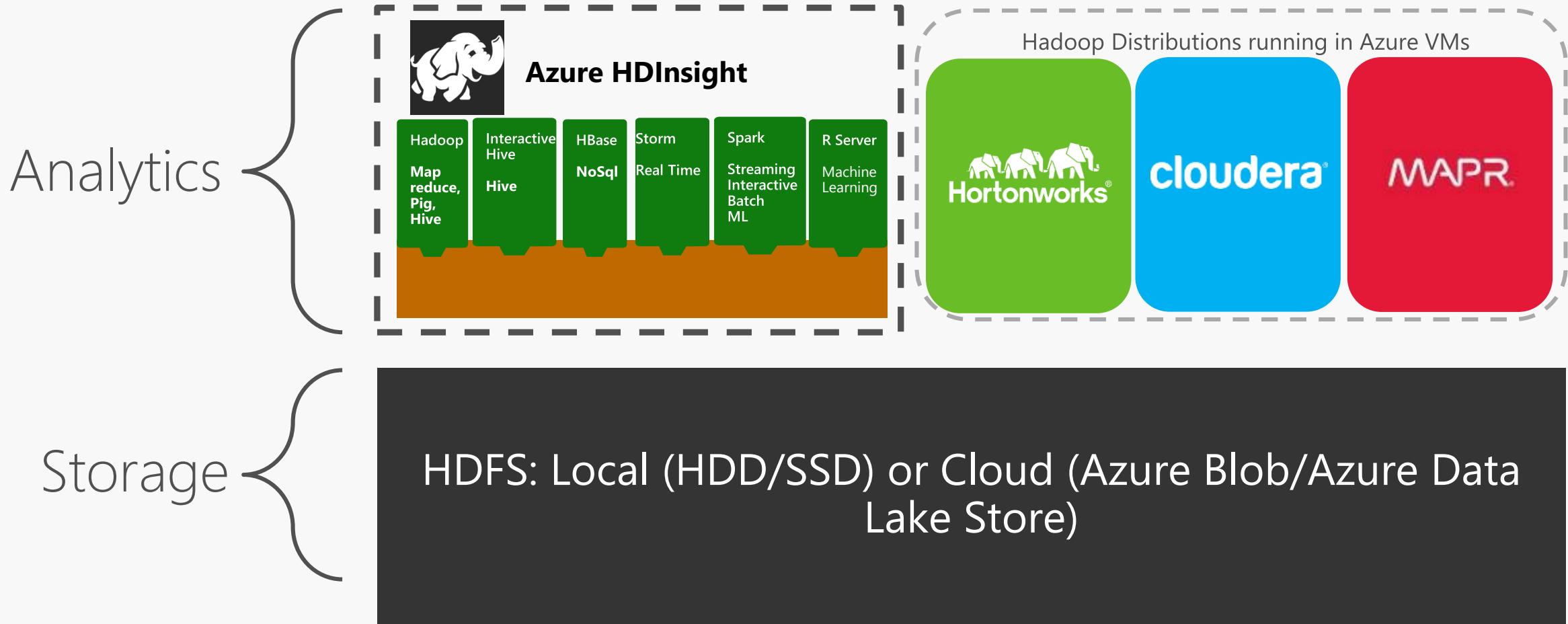
Data Velocity



Why Cloud + Hadoop?

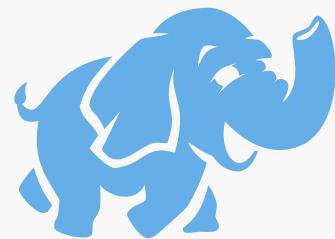


Big Data as part of Cortana Analytics



Azure HDInsight

Hadoop as a Service
on Azure



Fully-managed Hadoop for the cloud

100% Open Source Hortonworks data platform

Clusters up and **running in minutes**

Supported by Microsoft with **industry's best SLA**

Familiar **BI analysis tools**, Visual Studio integration

Open source notebooks for **interactive data science**

63% lower TCO than deploying Hadoop on-premise*

*IDC study "The Business Value and TCO Advantage of Apache Hadoop in the Cloud with Microsoft Azure HDInsight"

Interactive Data analysis through Zeppelin

Zeppelin Notebook Interpreter Connected

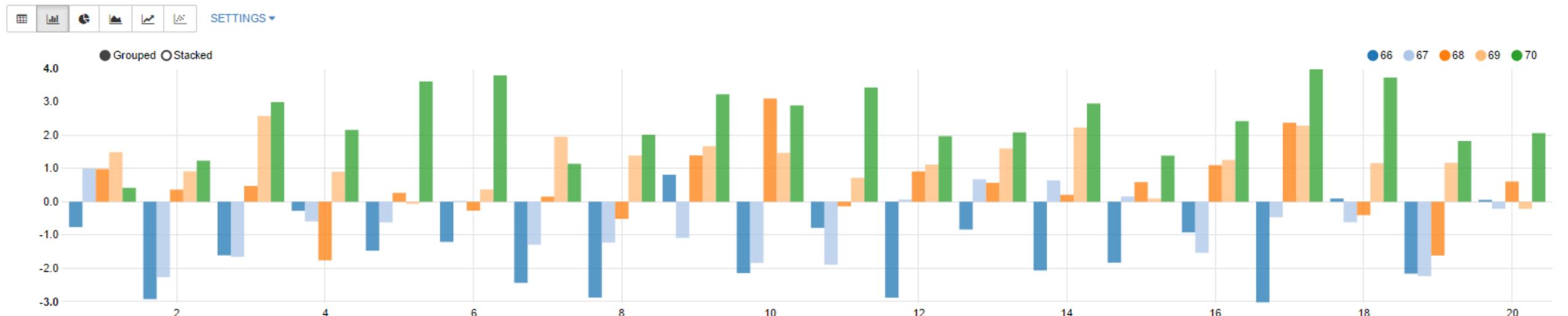
```
case class Hvac(date: String, time: String, targettemp: Integer, actualtemp: Integer, buildingID: String)
val hvac = hvacText.map(s => s.split(",")).filter(s => s(0) != "Date").map(
  s => Hvac(s(0),
    s(1),
    s(2).toInt,
    s(3).toInt,
    s(6)
  )
).toDF()
hvac.registerTempTable("hvac")
```

hvacText: org.apache.spark.rdd.RDD[String] = wasb://crimes@asadkstorage.blob.core.windows.net/HdiSamples/SensorSampleData/hvac/HVAC.csv MapPartitionsRDD[678905] at textFile at <console>:32
defined class Hvac
hvac: org.apache.spark.sql.DataFrame = [date: string, time: string, targettemp: int, actualtemp: int, buildingID: string]
Took 2 seconds

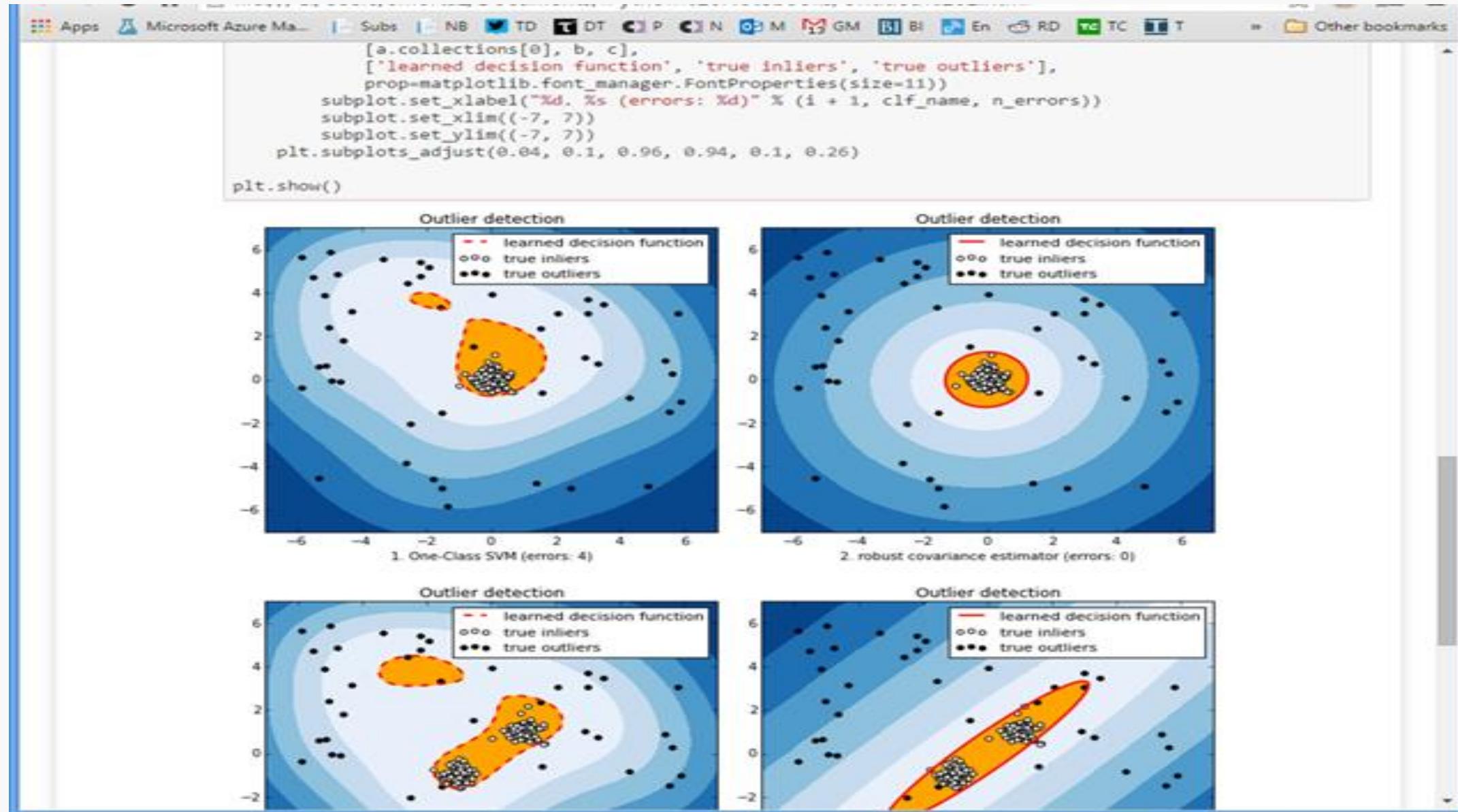
```
|%sql select buildingID, date, targettemp, (targettemp - actualtemp) as temp_diff
from hvac
where targettemp > "${Temp = 65,65|75|85}"
```

Temp 75 ▾

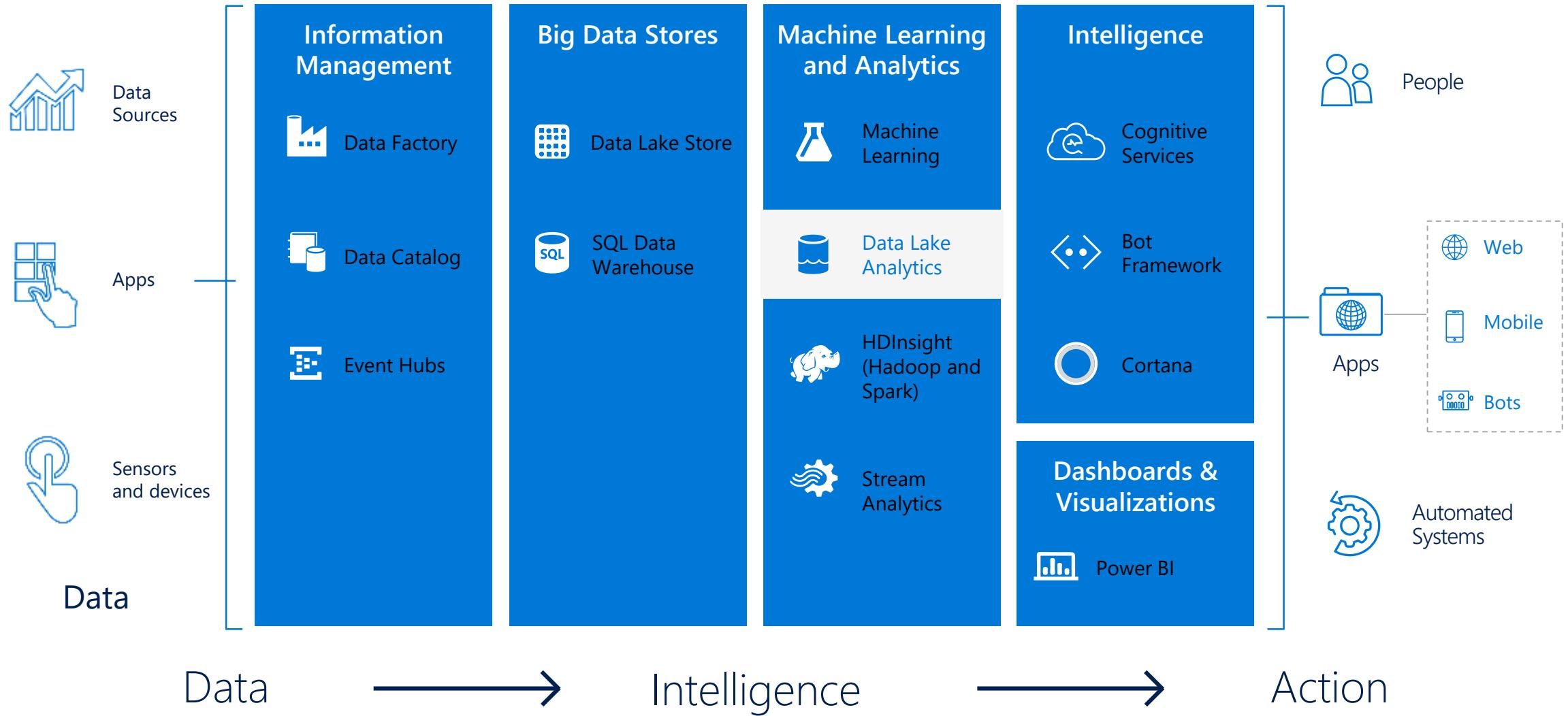
FINISHED ▶ ✎ 📈



Machine learning though Jupyter



Transform data into intelligent action





New U-SQL Job

Data Explorer

datalakeruben

Submit Job Data Explorer Open File Save As

* Job Name

DemoJob

Priority

1

AUs

37

Estimated Cost

1.23 USD/minute

```
1 REFERENCE ASSEMBLY ImageCommon;
2 REFERENCE ASSEMBLY FaceSdk;
3 REFERENCE ASSEMBLY ImageEmotion;
4 REFERENCE ASSEMBLY ImageTagging;
5 REFERENCE ASSEMBLY ImageOcr;
6
7 @imgs =
8     EXTRACT FileName string, ImgData byte[]
9     FROM @"images/{FileName:*.jpg"
10    USING new Cognition.Vision.ImageExtractor();
11
12 // Extract the number of objects on each image and tag them
13 @objects =
14     PROCESS @imgs
15     PRODUCE FileName,
16             NumObjects int,
17             Tags string
18     READONLY FileName
19     USING new Cognition.Vision.ImageTagger();
20
21 OUTPUT @objects
22     TO "/output/output2.tsv"
23     USING Outputters.Tsv();
24
25 @emotions =
26     PROCESS @imgs
27     PRODUCE FileName string,
28             NumFaces int,
29             Emotion string
30     READONLY FileName
```

Data Explorer

datalakeruben

Storage accounts

datalakeruben (default)

Catalog

datalakeruben

master

Filter New Folder Upload

datalakeruben

NAME

catalog

images

output

samples

system

usqlext



DemoJob

Job Details

► Resubmit
⟳ Refresh
 Duplicate Script
🚫 Cancel Job

Job Summary

Preparing
Queued
Running
Finalizing
35s
21s
55s

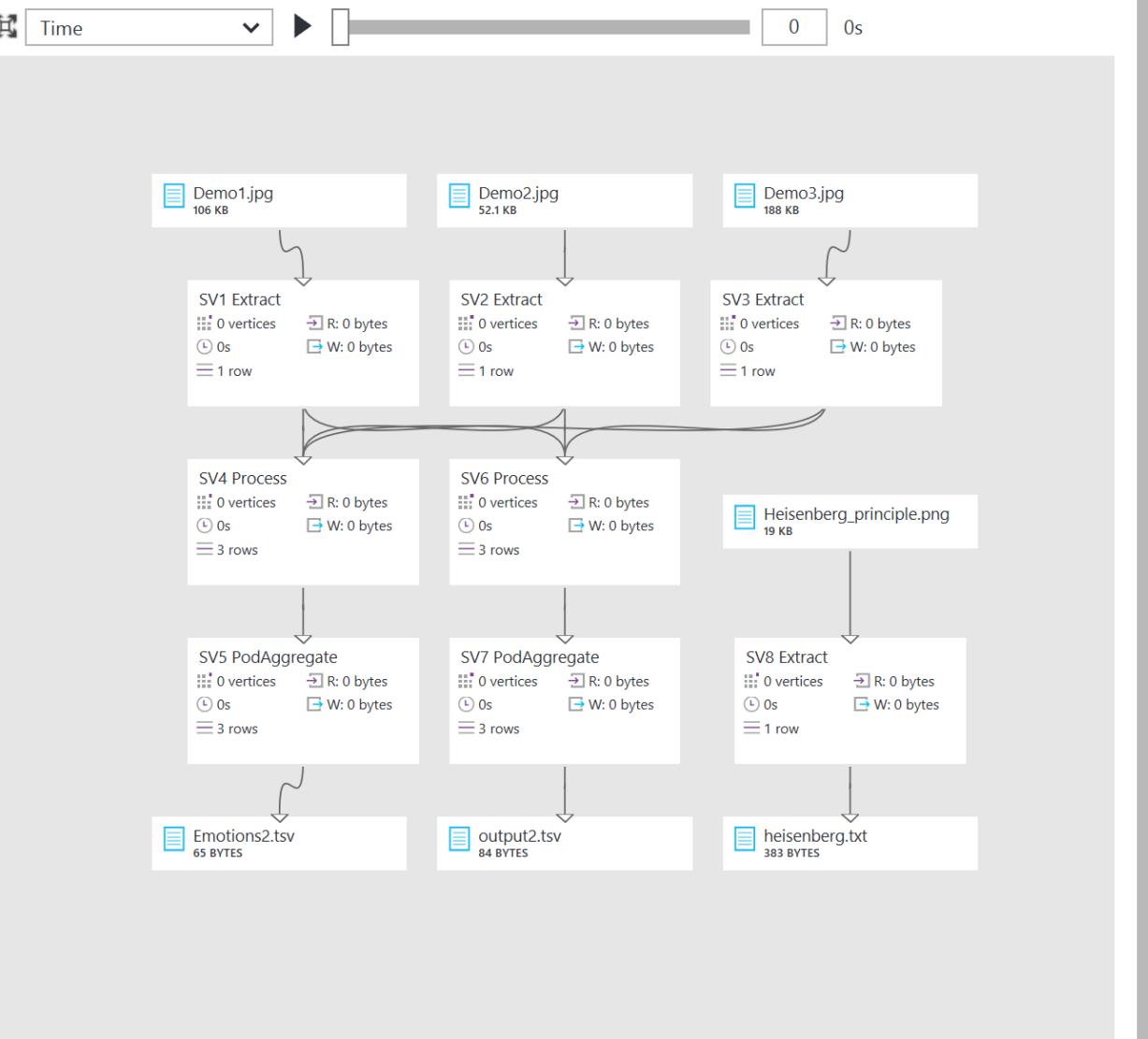
State: Succeeded
 Duration: 1min 51s
 Author: rupool@microsoft.com
 Submitted: 3/7/2017 3:58:18 PM

Show more...

Input
Output

NAME

Demo1.jpg	106 KB
Demo2.jpg	52.1 KB
Demo3.jpg	188 KB
Heisenberg_principle.png	19 KB



New U-SQL Job

► Submit Job
>Data Explorer
↑ Open

* Job Name

DemoJob

```

1 REFERENCE ASSEMBLY ImageC...
2 REFERENCE ASSEMBLY FaceSd...
3 REFERENCE ASSEMBLY ImageEr...
4 REFERENCE ASSEMBLY ImageTa...
5 REFERENCE ASSEMBLY ImageO...
6
7 @imgs =
8   EXTRACT FileName string
9   FROM @"~/images/{FileName}
10  USING new Cognition.V...
11
12 // Extract the number of o...
13 @objects =
14   PROCESS @imgs
15   PRODUCE FileName,
16     NumObjects int,
17     Tags string
18   READONLY FileName
19   USING new Cognition.V...
20
21 OUTPUT @objects
22 TO "/output/output2.ts...
23 USING Outputters.Tsv()
24
25 @emotions =
26   PROCESS @imgs
27   PRODUCE FileName string,
28     NumFaces int,
29     Emotion string
30   READONLY FileName
  
```

Visual exploration through BI

Power BI

Power BI

FILE SAVE READING VIEW A Text Box

My Workspace

Filter content

Dashboards

HDIBusiness Weekly Rep...

How Old Dashboard: ho...

Retail Analysis Sample

Spark

Spark

Reports

test1

Datasets

HDIBusiness Weekly Rep...

SalesMarketing

SparkDataset

SparkDataset *

Count of id by crimetype

Crimetype	Count (M)
ARSON	~0.01
ASSAULT	~0.35
BATTERY	~1.10
BURGLARY	~0.35
CONCEALED C...	~0.02
CRIM SEXUAL ...	~0.02
CRIMINAL DA...	~0.60
CRIMINAL TRE...	~0.18
DECEPTIVE PR...	~0.18
DOMESTIC VI...	~0.01
GAMBLING	~0.01
HOMICIDE	~0.01
HUMAN TRAF...	~0.01
INTIMIDATION	~0.01
KIDNAPPING	~0.01
LICOR LAW ...	~0.01
MOTOR VEHIC...	~0.28
NARCOTICS	~0.65
NON - CRIMIN...	~0.01
NON-CRIMIN...	~0.01
OBSCENITY	~0.01
OFFENSE INV...	~0.01
OTHER MARC...	~0.01
OTHER OFFEN...	~0.35
Primary Type	~0.01
PROSTITUTION	~0.05

Visualizations >

Fields >

crime

arrest

beat

block

casenumber

communityarea

Count of Rows i...

crimetype

date

description

district

domestic

fbicode

icur

id

latitude

location

locationdescrip...

longitude

updatedon

ward

xcoordinate

ycoordinate

year

crimeraw

crimetest

hivepeople1

hivesamplable

test

Get Data

Page 1 +



*“Does your car have any idea why
my car pulled it over?”*

