Microsoft Azure

Carlos Frederico Marques (Fred)
carlos.marques@microsoft.com

Daniel Cardoso
daniel.Cardoso@microsoft.com

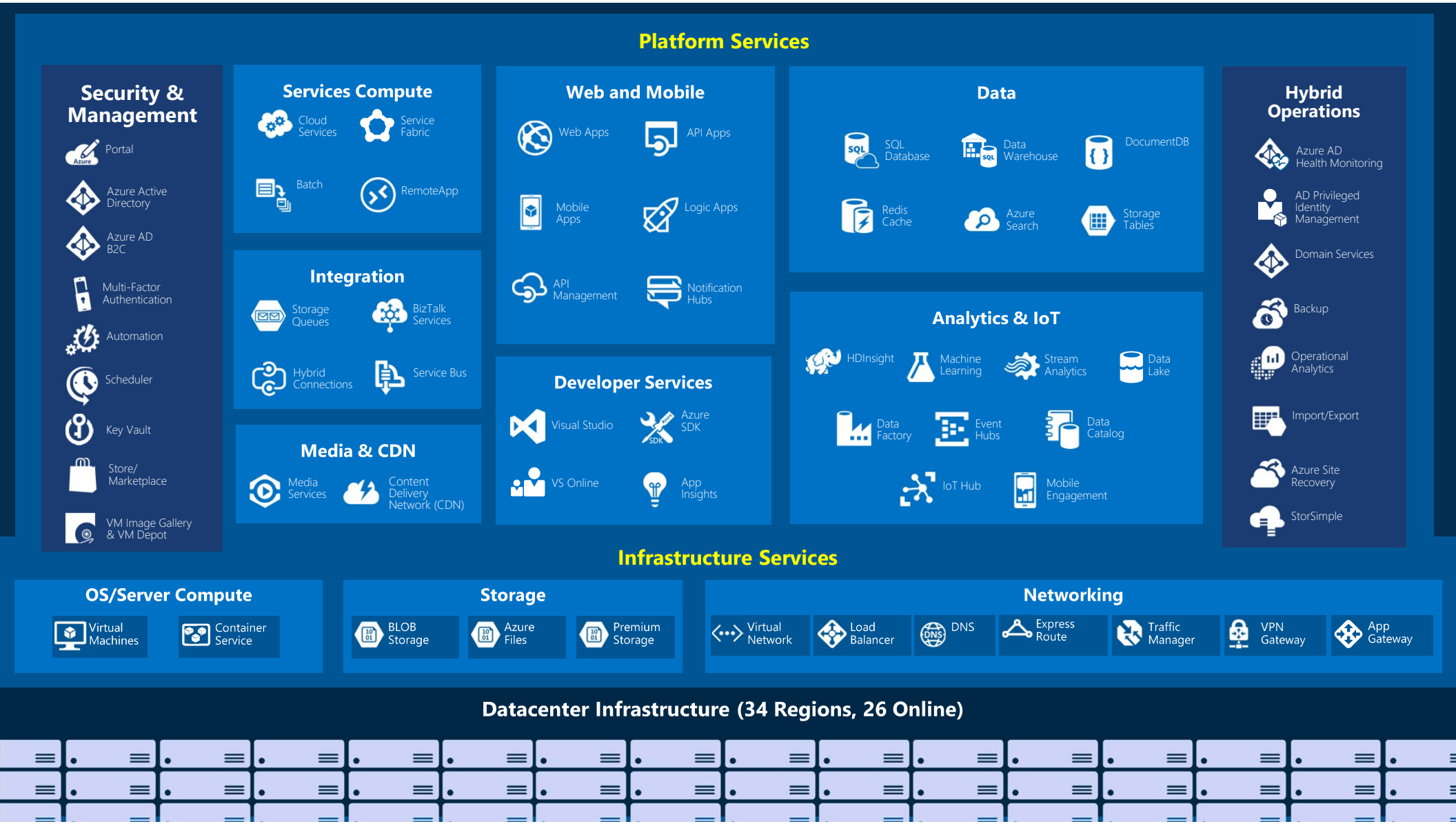
Microsoft

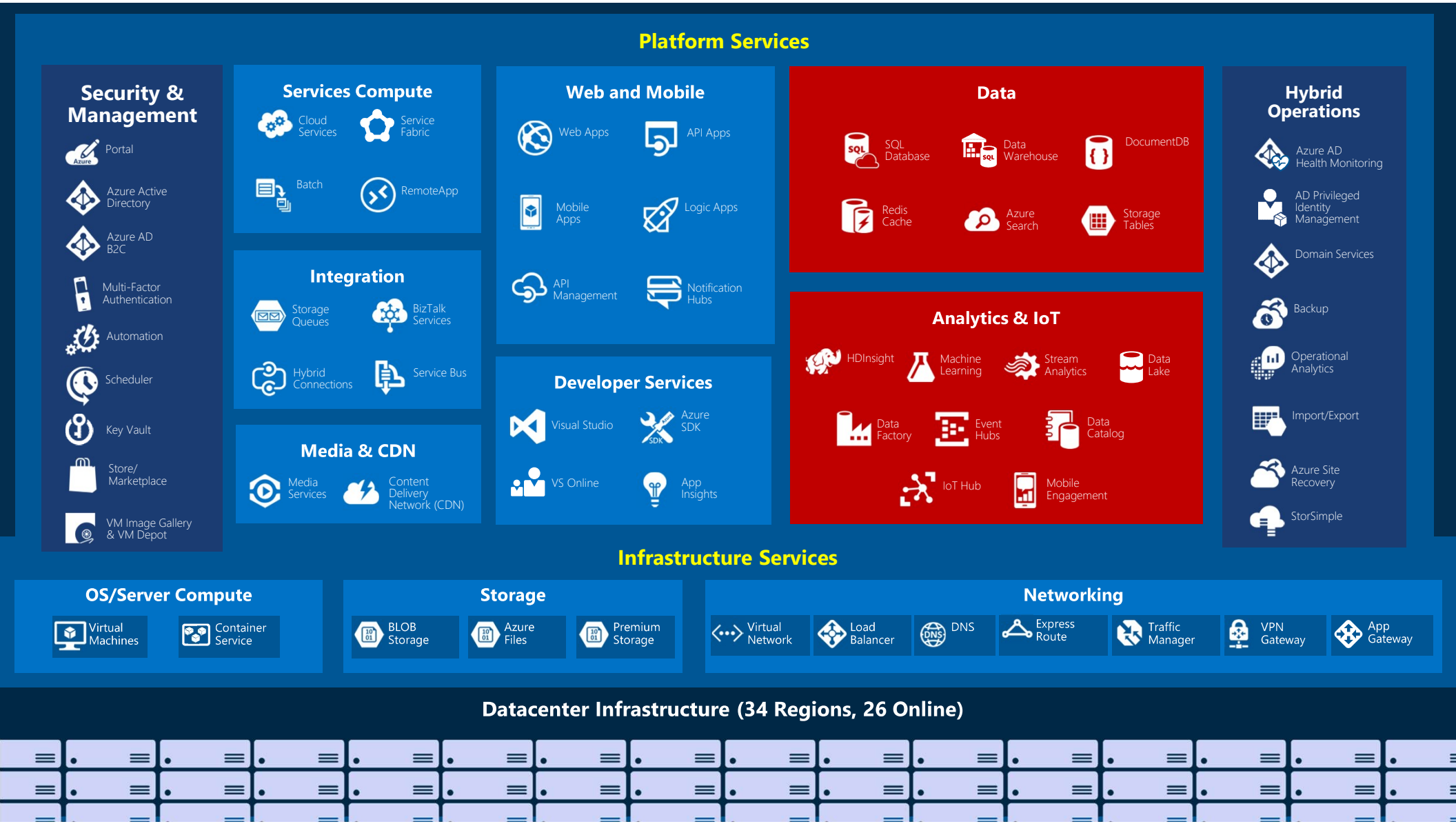
Agenda

Microsoft

- Why Cloud?
- SQL Database
- SQL DW
- DocumentDB
- HDInsight
- Data Lake
- Search
- Machine Learning
- Internet of Things
- Cognitive Services
- Power BI
- Market Place

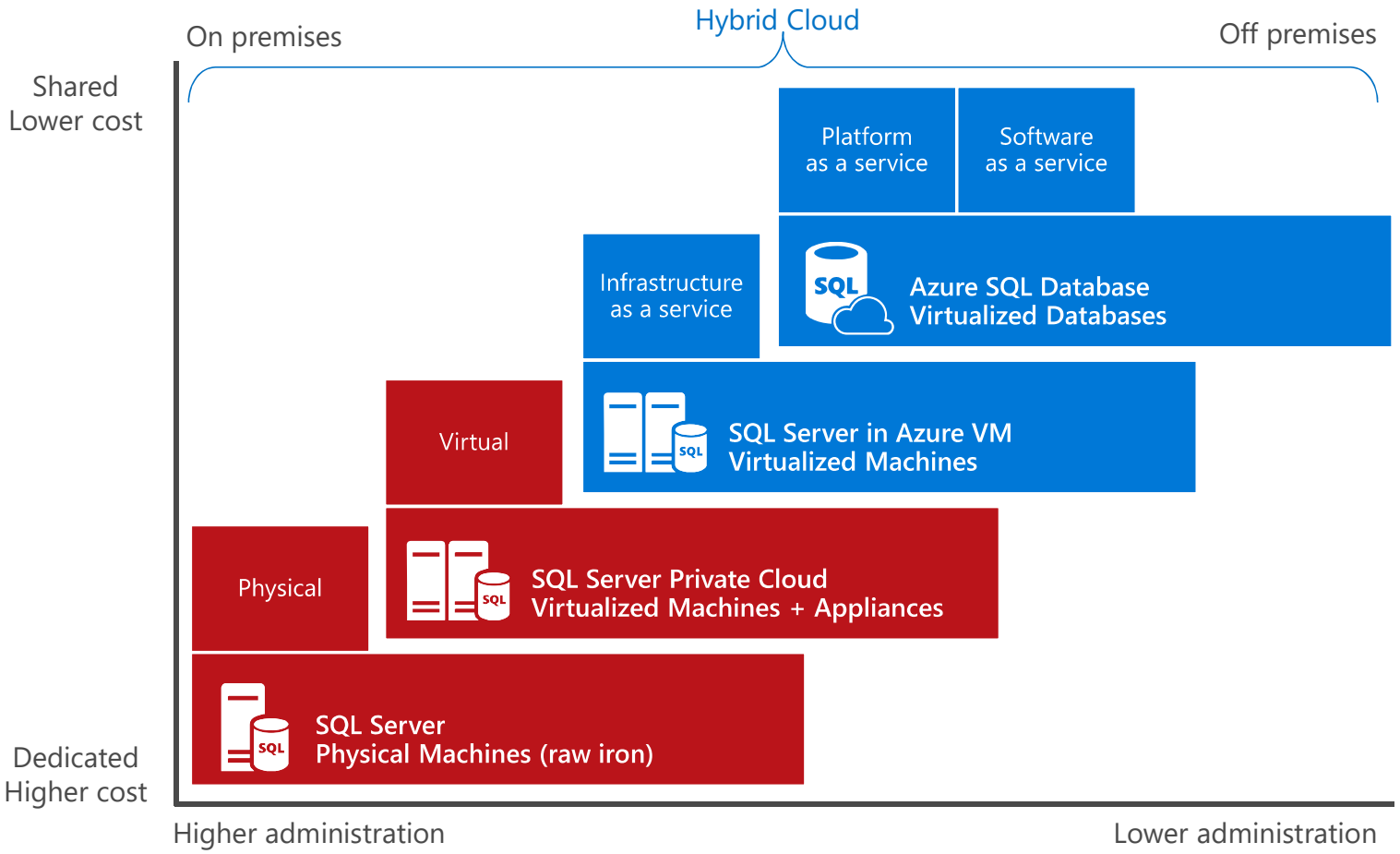
Microsoft


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
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
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 (34 Regions, 26 Online)


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
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
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 (34 Regions, 26 Online)

SQL Database

Data platform continuum



SQL Database Service

A relational **database as a service**, fully managed by Microsoft.

For cloud-designed apps when **near-zero administration** and **enterprise-grade** capabilities are key.

Perfect for cloud **architects and developers** looking for programmatic DBA-like functionality.

Elastic Scale and performance	Business continuity and data protection	Familiar and self-managed
Predictable performance levels	Self-service restore	Familiar & compatible
Programmatic scale-out	Disaster recovery	Programmatic
Dashboard views of database metrics	Compliance-enabled	Self-managed

Azure SQL Database single-database service tiers

	Basic	Standard	Premium
Built for …	Light transactional workloads	Medium transactional workloads	Heavy transactional workloads
Availability SLA	99.99%*		
Database max size	2 GB	250 GB	1 TB
Point-in-time restore ("oops" recovery)	Any point within 7 days	Any point within 14 days	Any point within 35 days
Business continuity	Geo-restore	Geo-replication, standby secondary	Active geo-replication, up to four readable secondaries
Security	Auditing, Row-Level Security, dynamic data masking		
Performance objectives	Transactions per hour	Transactions per minute	Transactions per second
Database Transaction units (DTUs)	Basic: 5	S0: 10 S1: 20 S2: 50 S3: 100	P1: 125 P2: 250 P4: 500 P6: 1,000 P11: 1,750
Available tiers ($/month) and GA price	Basic: $4.99	S0: $15 S1: $30 S2: $75 S3: $150	P1: $465 P2: $930 P4: $1,860 P6: $3,720 P11: $7,001

*The 99.99% availability SLA does not apply to the existing Web and Business editions, which will continue to be supported at 99.9% availability.

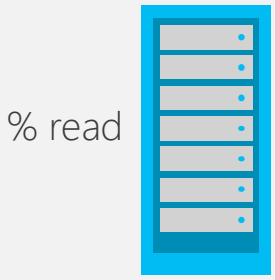
Designed for predictable performance

Redefined

Across Basic, Standard, and Premium, each performance level is assigned a defined level of throughput

Measure of power

Introducing the Database Throughput Unit (DTU) which represents database power and replaces hardware specs



DTU is defined by the bounding box for the resources required by a database workload and measures power across the six performance levels.

Basic — 5 DTU	S0 — 10 DTU	P1 — 125 DTU
	S1 — 20 DTU	P2 — 250 DTU
	S2 — 50 DTU	P4 — 500 DTU
	S3 — 100 DTU	P6 — 1,000 DTU
		P11 — 1,750 DTU

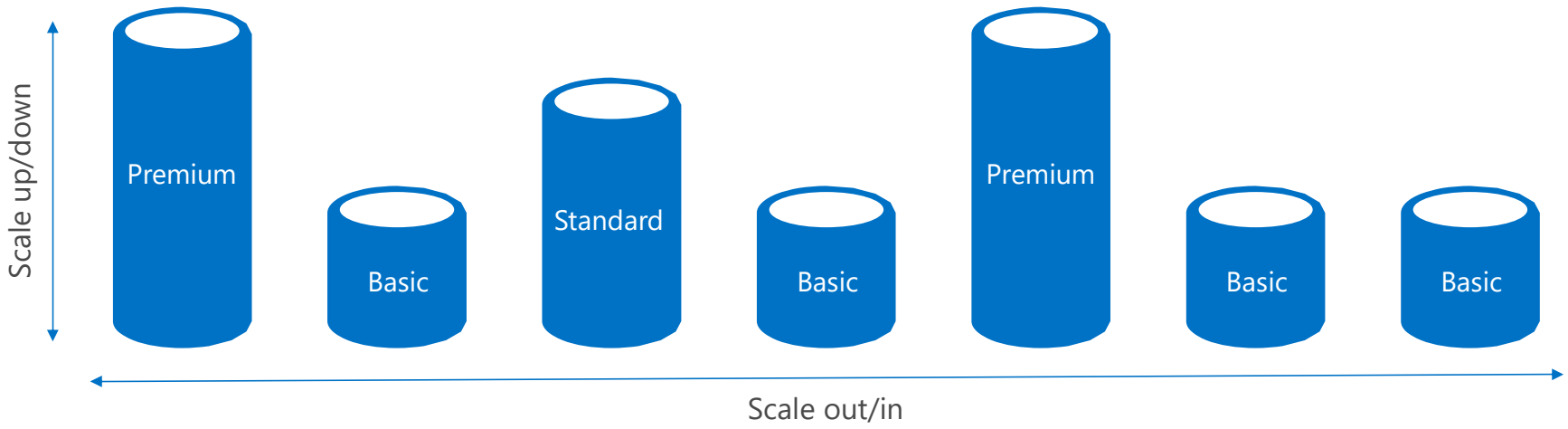
Scalability options in Azure SQL Database

Vertical: Scale up or scale down

Change service tiers for a given database as capacity needs fluctuate

Horizontal: Scale out or scale in

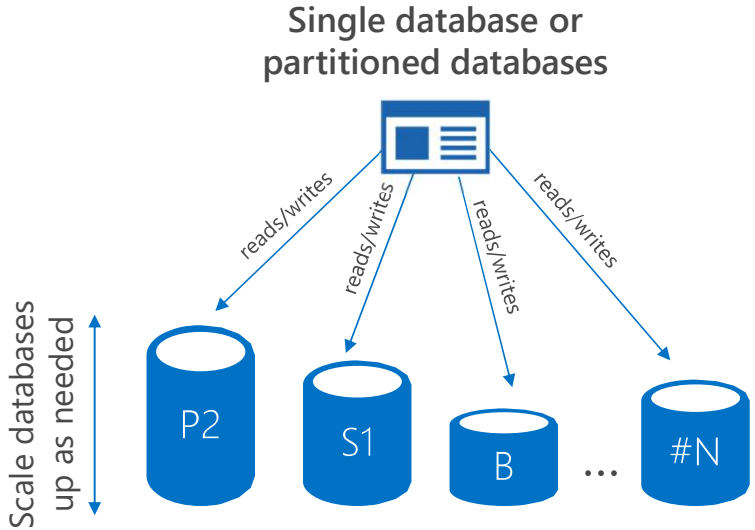
Add or remove databases (sharded and/or in a pool) as more or less capacity is needed



Managing large numbers of databases

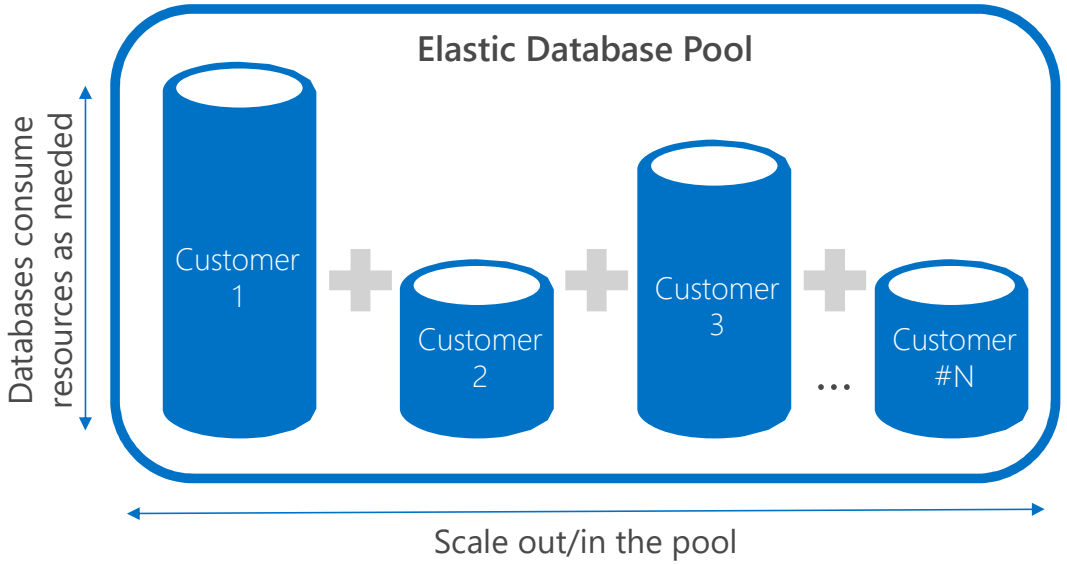
Predictable workloads

Single databases or partitioned data across multiple databases; scale between service tiers and performance levels as capacity needs fluctuate.



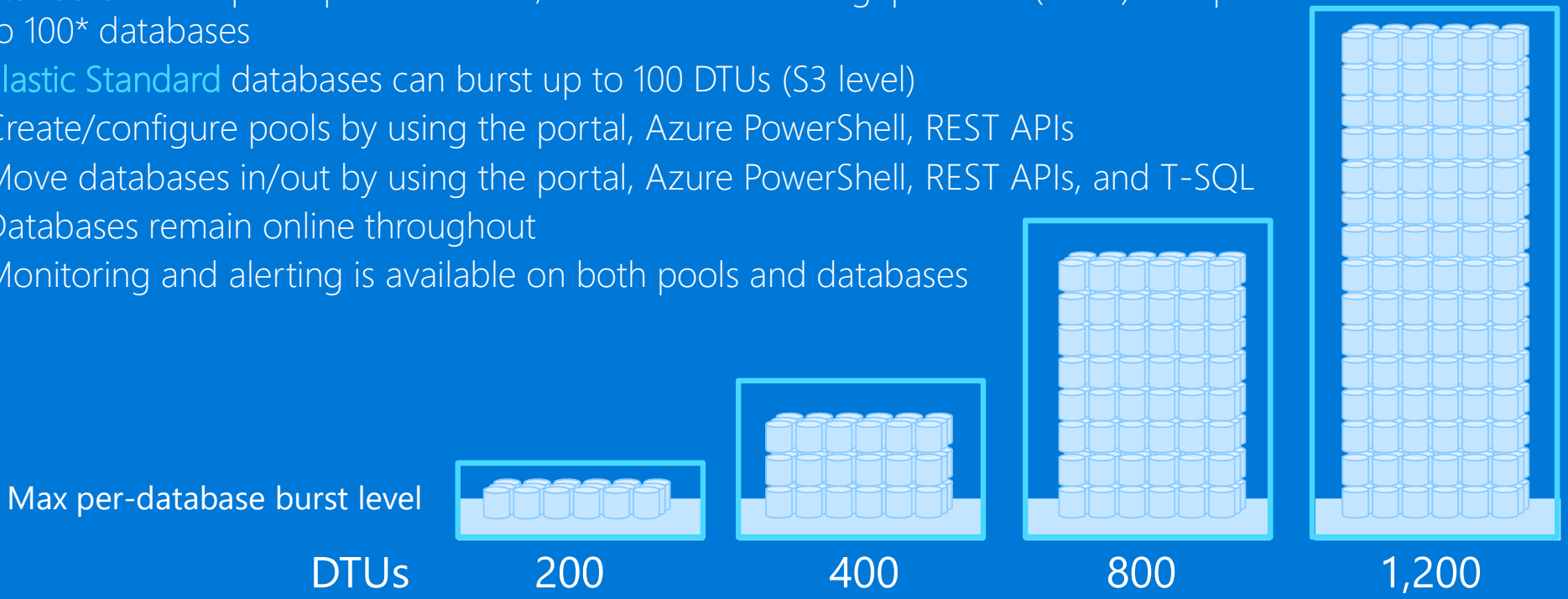
Unpredictable workloads

For large numbers of databases with unpredictable performance demands; pool resources to be shared between these databases.

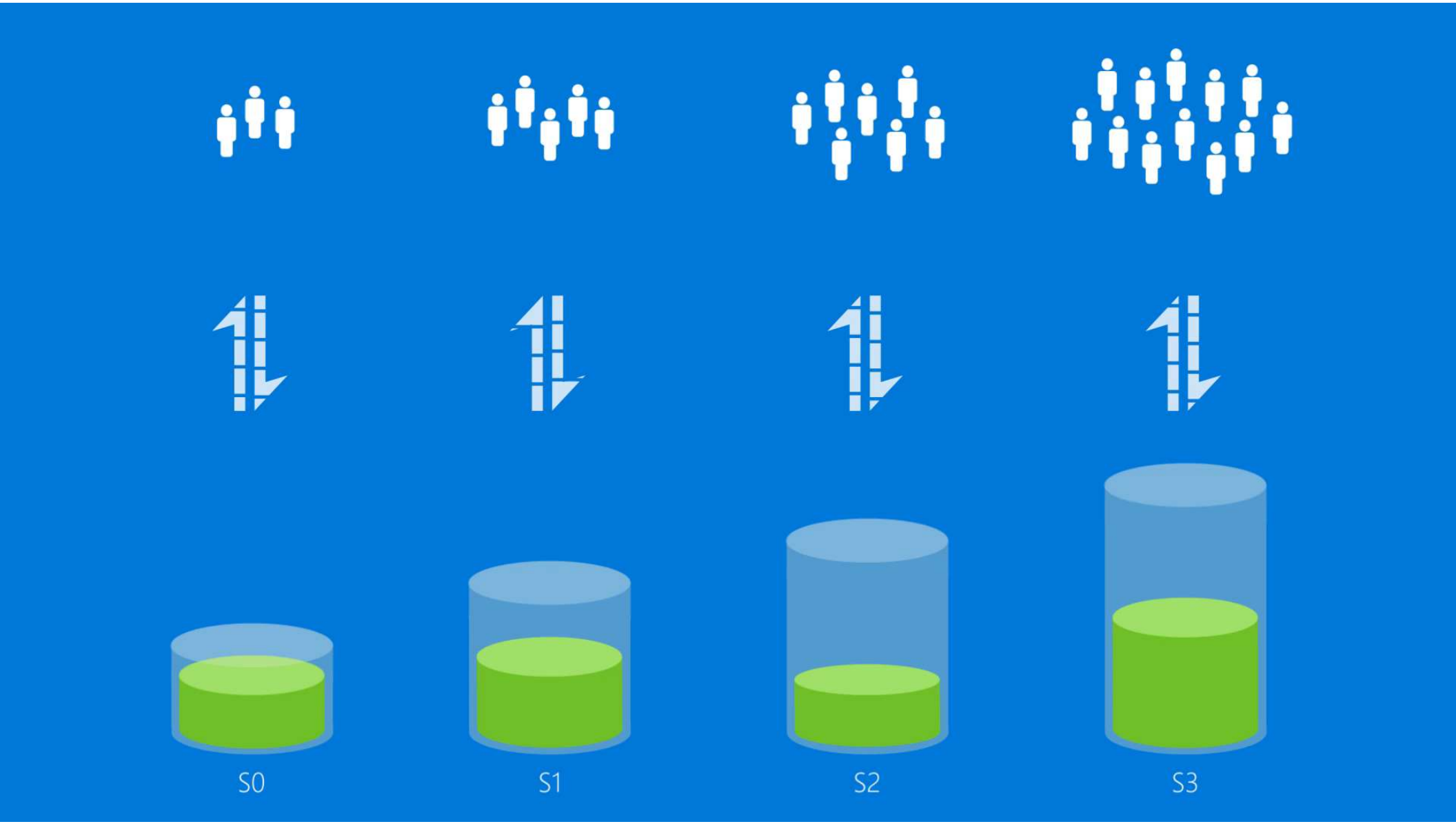


Elastic Database Model

- Elastic databases in elastic database pools
- Pooled resources are used by many databases
- Standard elastic pools provide 200–1,200* database throughput units (DTUs) for up to 100* databases
- Elastic Standard databases can burst up to 100 DTUs (S3 level)
- Create/configure pools by using the portal, Azure PowerShell, REST APIs
- Move databases in/out by using the portal, Azure PowerShell, REST APIs, and T-SQL
- Databases remain online throughout
- Monitoring and alerting is available on both pools and databases



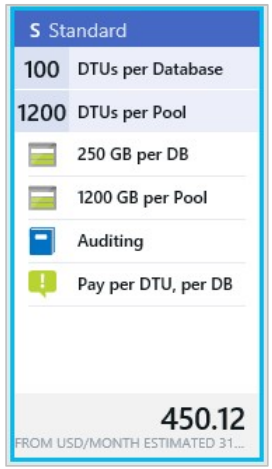
*Additional pricing tiers might be introduced, and the ranges and limits might be increased during the preview

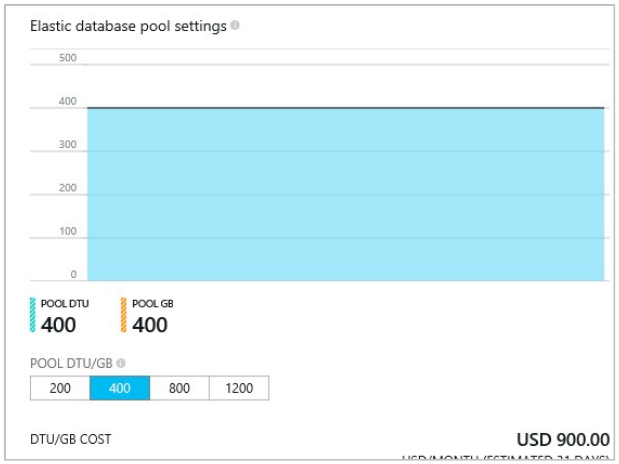

S0
S1
S2
S3


ELASTIC DATABASE POOL
MAX
100 DTUs
MIN
10 DTUs

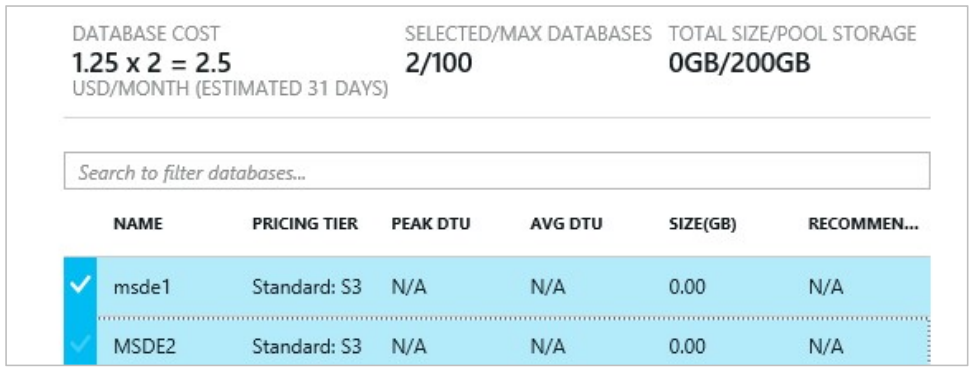
Elastic Database Pricing

Standard pricing for preview





Per database cost



Standard plan is the only option for the preview

Price increments linearly as you add DTUs/GB

- 200 DTU/DB default at $450 per month

Additional charge based on the number of databases in the pool

Point-in-time restore

Automatic backup

Full backups weekly, different backup daily, log backups every 5 minutes

Daily and weekly backups automatically uploaded to geo-redundant Azure Storage

Self-service restore

Point-in-time up to a second granularity

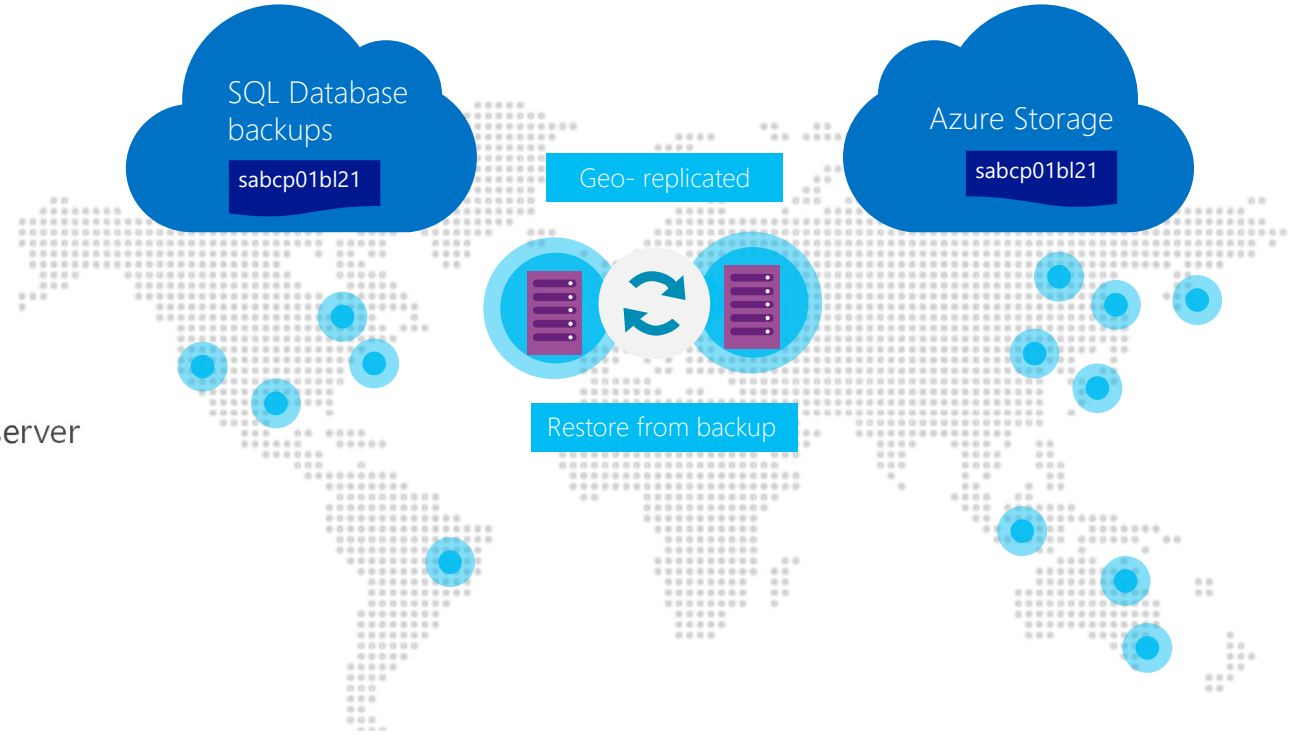
REST API, PowerShell, or Portal

Creates a new database in the same logical server

Tiered retention policy

Basic - 7 days
Standard - 14 days
Premium - 35 days

No additional cost to retain backups



Geo-restore

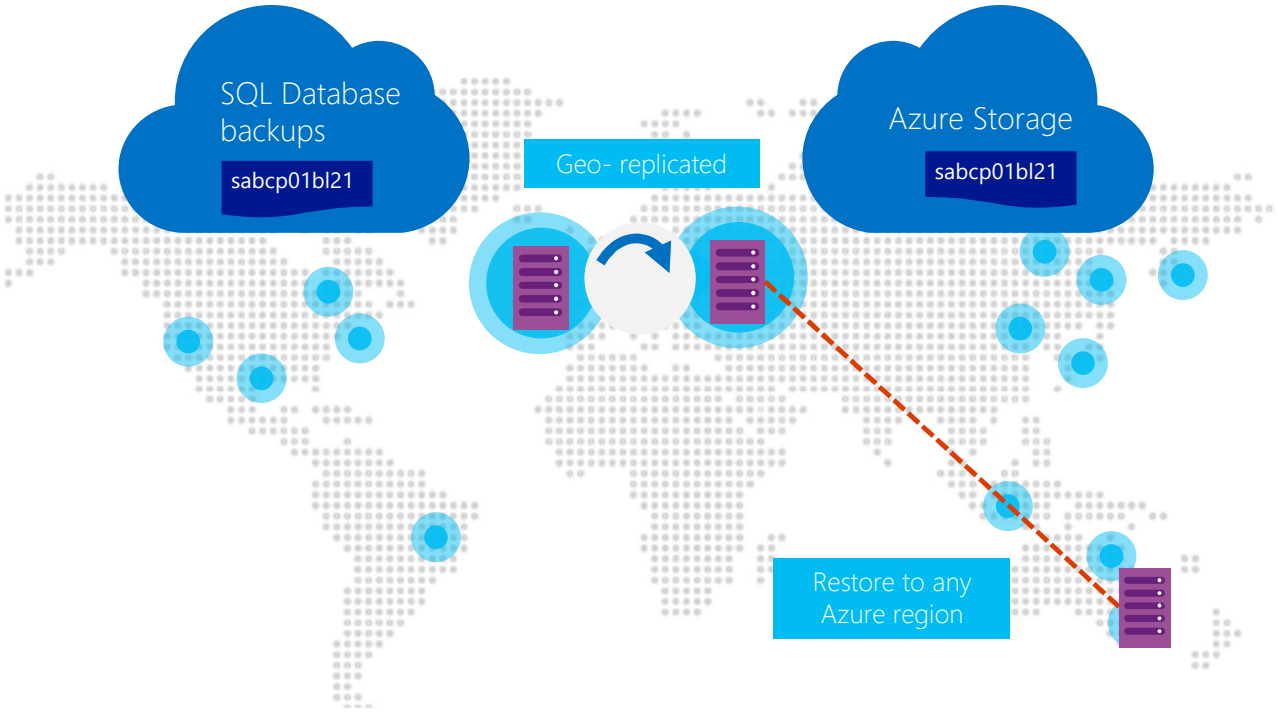
Self-service restore API

Restores last daily backup

No extra cost, no capacity guarantee

RTO>=24h, RPO=24h

Database URL will change after restore



Active geo-replication

Mission-critical business continuity on your terms, via programmatic APIs

Self-service activation in Premium

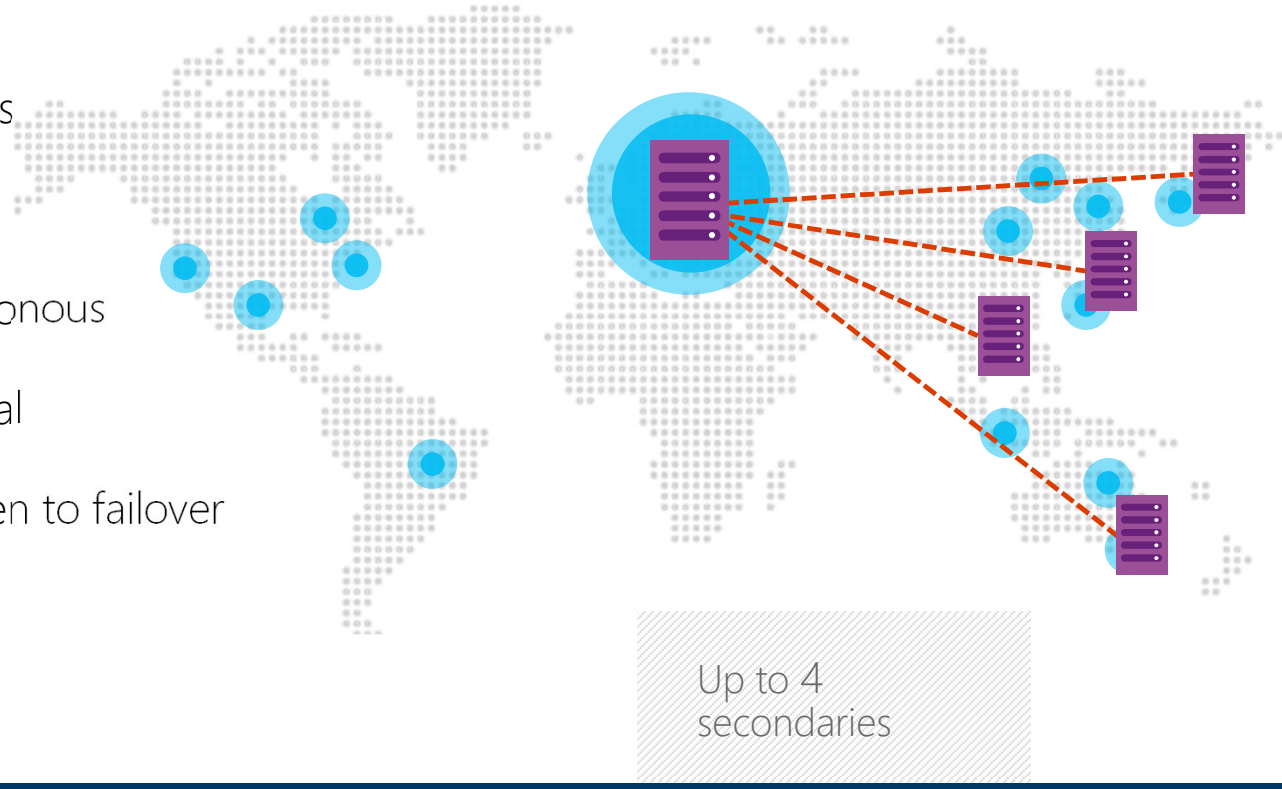
Create up to 4 readable secondaries

Replicate to any Azure region

Automatic data replication, asynchronous

REST API, PowerShell or Azure Portal

RTO<1h, RPO<5m, you choose when to failover



Azure Active Directory integration

Centrally manage user permissions

Alternative to SQL Server authentication

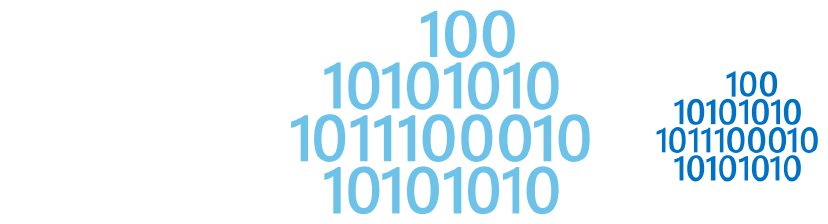
Limits proliferation of user identities across databases

Allows password rotation in a single place

Manage database permissions using external Azure Active Directory groups

Stops password storing by using integrated Windows authentication and other forms of authentication supported by Azure Active Directory

In preview



Row-level security

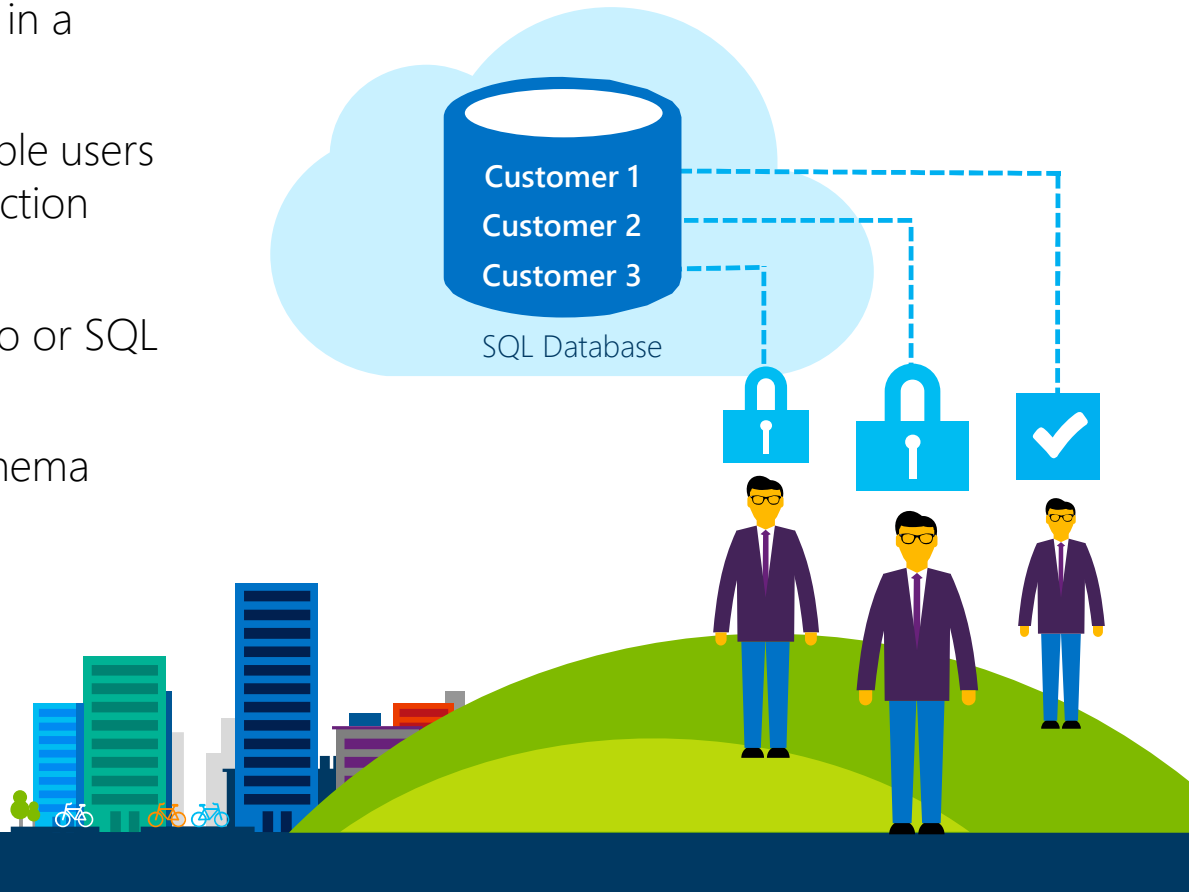
Protect data privacy by ensuring the right access across rows

Fine-grained access control over specific rows in a database table

Help prevent unauthorized access when multiple users share the same tables, or to implement connection filtering in multitenant applications

Administer via SQL Server Management Studio or SQL Server Data Tools

Enforcement logic inside the database and schema bound to the table.



Status: Public preview

Dynamic Data Masking

Prevent the abuse of sensitive data by hiding it from users

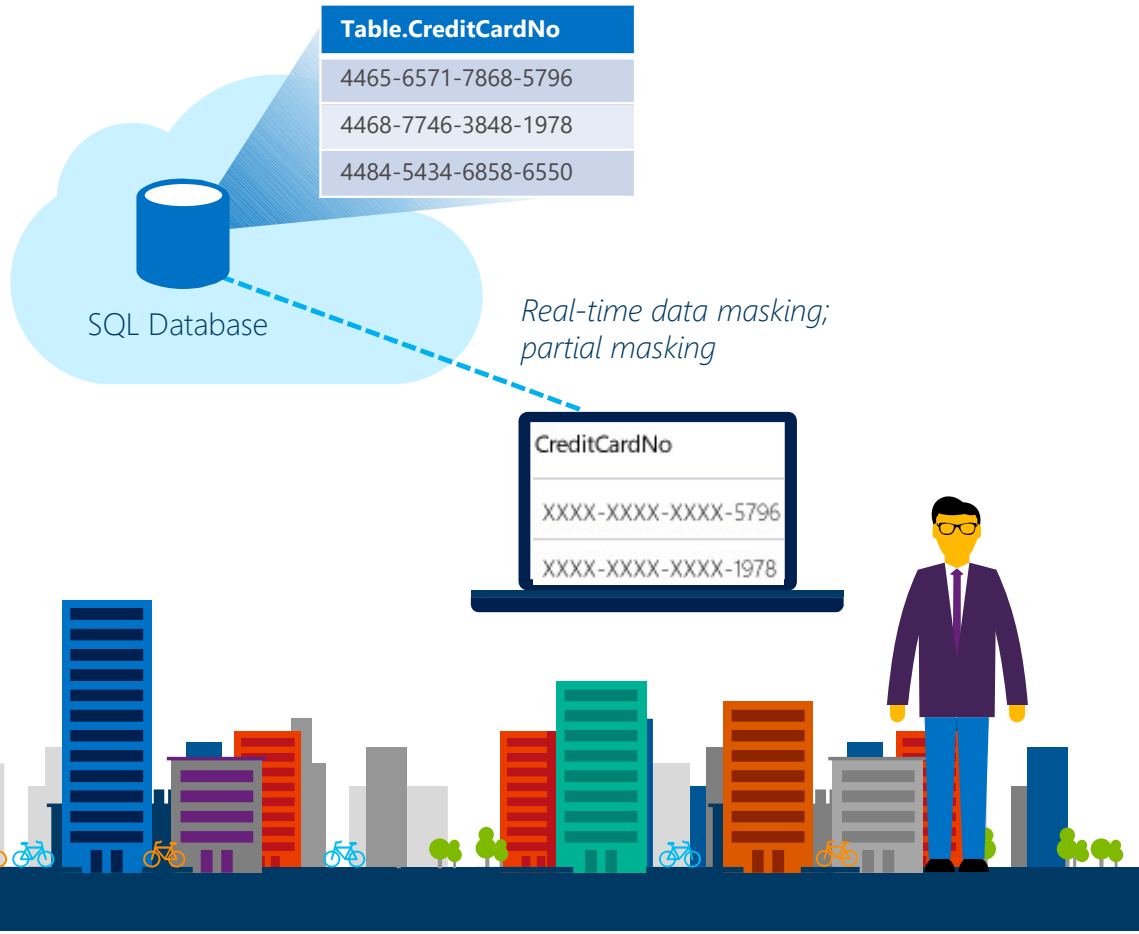
Configuration made easy in the new Azure portal

Policy-driven at the table and column level, for a defined set of users

Data masking applied in real-time to query results based on policy

Multiple masking functions available (e.g. full, partial) for various sensitive data categories (e.g. Credit Card Numbers, SSN, etc.)

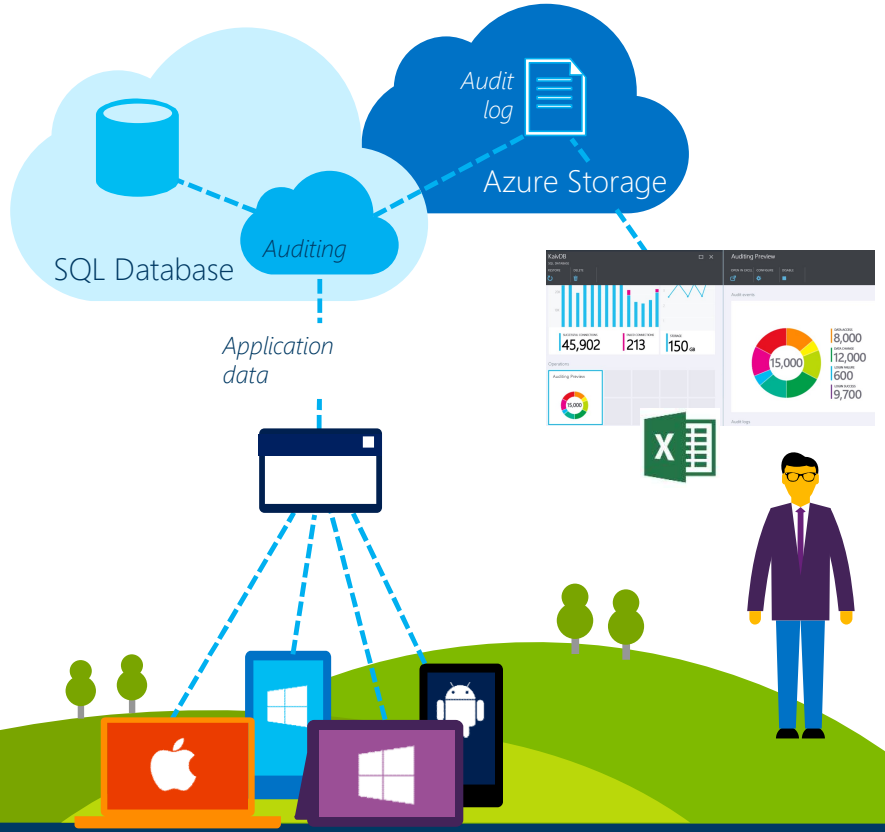
Status: Public preview



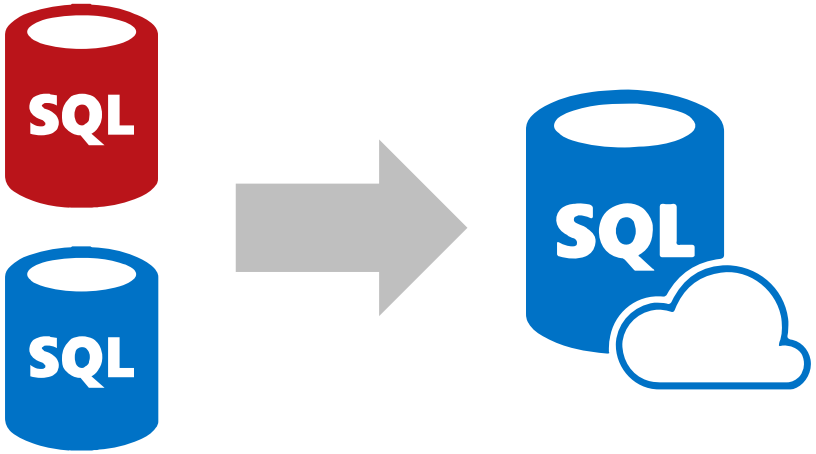
Auditing

Gain insight into database events & streamline compliance-related tasks

- Configurable to track & log database activity
- Dashboard views in the portal for at-a-glance insights
- Pre-defined Power View reports for deep visual analysis on Audit log data
- Audit logs reside in your Azure Storage account
- Available in Basic, Standard, and Premium



Migrating to Azure SQL Database (v12)



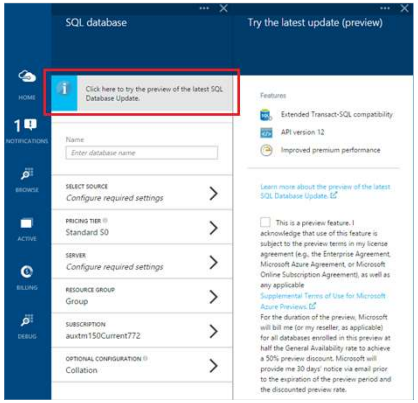
Migrate an on-premises SQL Server database to Azure SQL Database (v12)

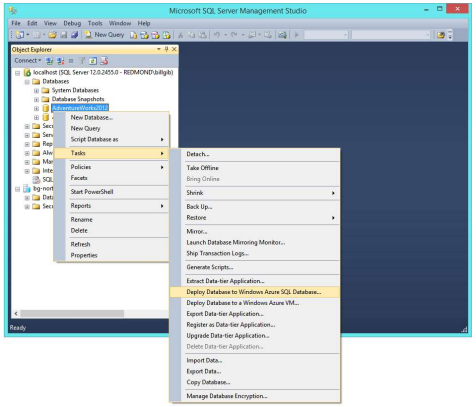
Get near-complete engine compatibility with SQL Server 2014*

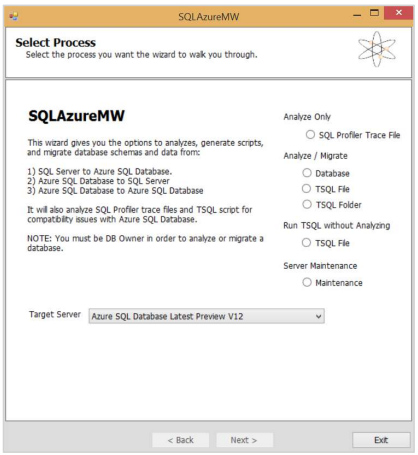
Simplify migration:

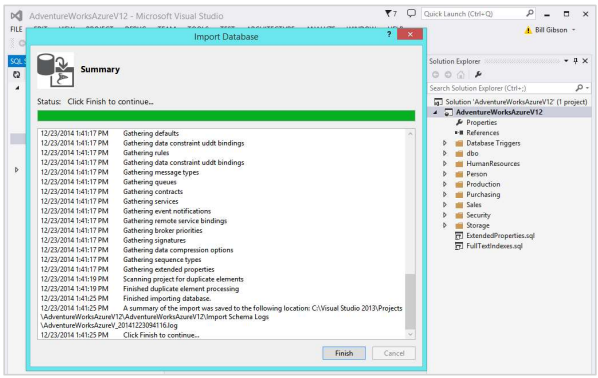
- Straightforward migration for the majority of databases
- Requires few or no changes to schema
- No re-engineering of applications**

Migration tools





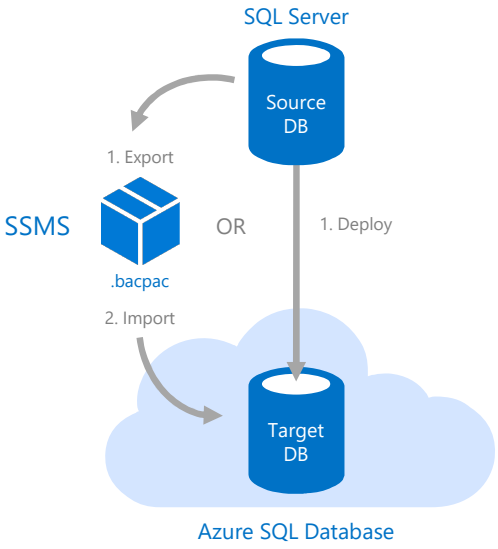




- Preview Azure Management Portal
- SQL Server Management Studio (SSMS)
- SQL Azure Migration Wizard (SAMW)
- SQL Server Data Tools in Visual Studio

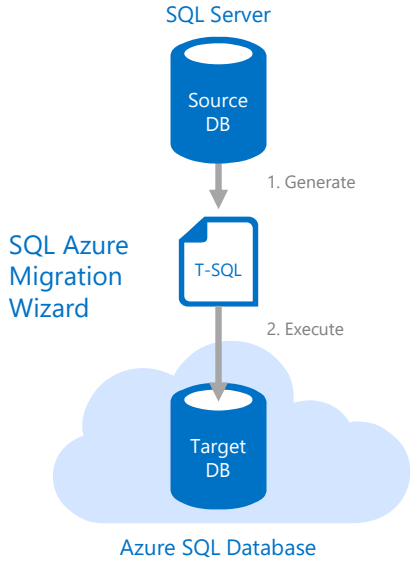
Migration methodologies

Method 1



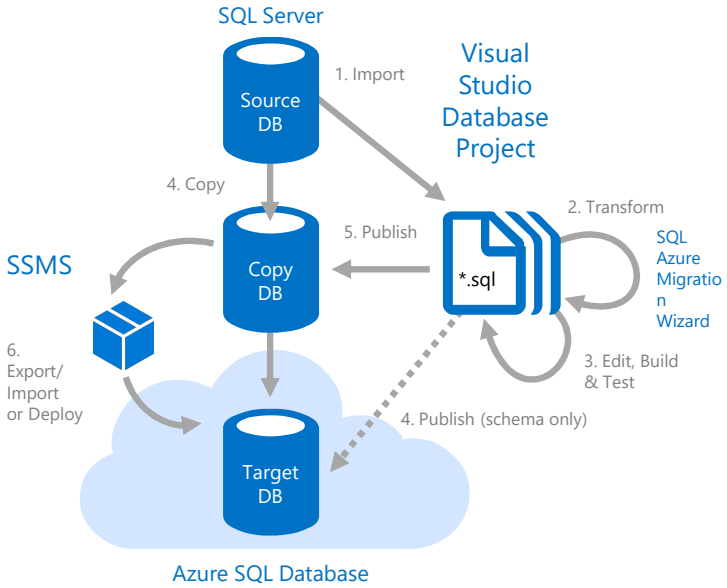
Migrate a compatible database using SSMS

Method 2



Migrate a near-compatible database using SAMW

Method 3



Update the database schema offline using Visual Studio and SAMW, and then deploy it with SSMS

Limits

https://azure.microsoft.com/en-us/documentation/articles/sql-database-transact-sql-information/

	Basic	Standard				Premium				
		S0	S1	S2	S3	P1	P2	P4	P6/P3	P11
Maximum database size	2 GB	250 GB				500 GB				1 TB
DTUs	5	10	20	50	100	125	250	500	1,000	1,750
Point-in-time restore	Any point last 7 days	Any point last 14 days				Any point last 35 days				
Disaster recovery	Geo-Restore, restore to any Azure region	Standard Geo-Replication, offline secondary				Active Geo-Replication, up to 4 online (readable) secondary backups				
Max In-Memory OLTP storage	NA	NA	NA	NA	NA	1 GB	2 GB	3 GB*	8 GB	10 GB*
Max concurrent requests	30	60	90	120	200	200	400	800	1,600	2,400
Max concurrent logins	30	60	90	120	200	200	400	800	1,600	2,400
Max sessions	300	600	900	1,200	2,400	2,400	4,800	9,600	19,200	32,000

** In-Memory OLTP storage limits will soon adjust to 4 for P4 and 14 for P11.*

ELASTIC POOL LIMITS	Basic					Standard					Premium				
eDTUs per pool	100	200	400	800	1,200	100	200	400	800	1,200	125	250	500	1,000	1,500
Max. storage*	10 GB	20 GB	39 GB	78 GB	117 GB	100 GB	200 GB	400 GB	800 GB	1.2 TB	250 GB	500 GB	750 GB	750 GB	750 GB
Max. DBs	200	400				200	400				50				
Max. concurrent workers	200	400	800	1,600	2,400	200	750	1,300	1,850	2,400	200	750	1,300	1,850	2,400
Max. concurrent sessions	2,400	4,800	9,600	19,200	28,800	2,400	4,800	9,600	19,200	28,800	2,400	4,800	9,600	19,200	28,800

ELASTIC DB LIMITS			
Max. storage*	2 GB	250 GB	500 GB
Min. eDTUs	0, 5	0, 10, 20, 50, 100	0, 125, 250, 500, 1000
Max. eDTUs	5	10, 20, 50, 100	125, 250, 500, 1000

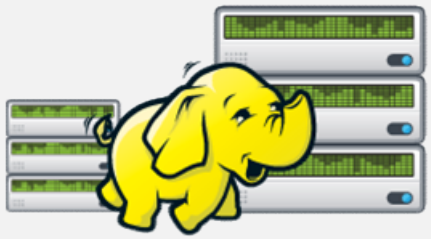
BUSINESS CONTINUITY			
Point-in-time restore	Any point last 7 days	Any point last 14 days	Any point last 35 days
Disaster recovery	Geo-restore, restore to any Azure region	Standard geo-replication, offline secondary	Active geo-replication, up to four readable secondary backups

** Databases share pool storage, so database storage is limited to the smaller of remaining pool storage or max. storage per database.*

Azure HDInsight

Non structured data | beyond the relational

Introducing Apache Hadoop



Apache Open Source Project

Highly scalable distributed file system (HDFS)

Distributed processing on data nodes

Data Volumes

Data Variety

Data Velocity

Data volume

Hadoop stores files in a distributed file system

Storage and computation is distributed across many servers

Files can be spread out over multiple nodes

Hadoop can store very large amounts of data

Combined storage resource can grow with demand from a few nodes to thousands of nodes

Scales out linearly

Very large files supported including those larger than the capacity of a single node

Data variety

Hadoop stores files (non-relational store)

Files could have a variety of semi-structured or unstructured data
Previously, these files may not have been seen as providing value or insights
Today, new business questions and insights are being uncovered through data science

Sentiment
Understand how your customers feel about your brand and products—right now

Clickstream
Capture and analyze website visitors' data trails and optimize your website

Sensors
Discover patterns in data streaming automatically from remote sensors and machines

Geographic
Analyze location-based data to manage operations where they occur

Server logs
Research logs to diagnose process failures and prevent security breaches

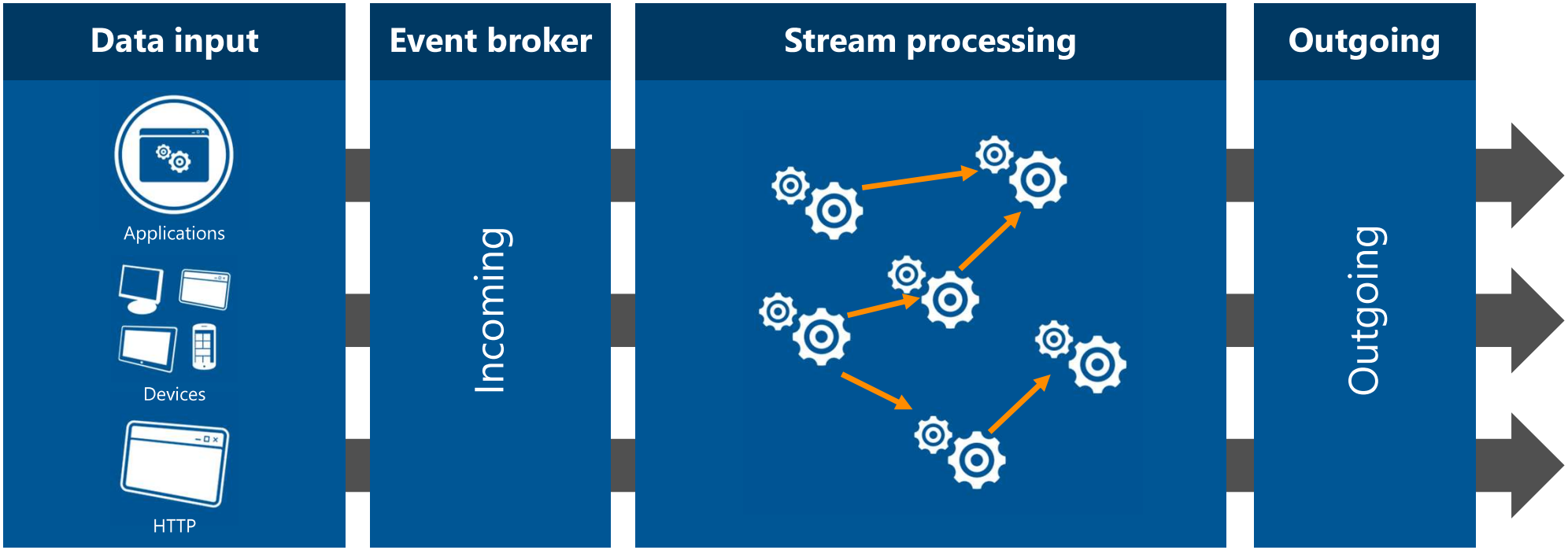
Unstructured
Understand patterns in files across millions of web pages, emails, and documents

Data velocity

Hadoop can stream live data and process them in real-time

Hadoop can act as scalable event stream ingestion

Hadoop can do near real-time in-stream processing



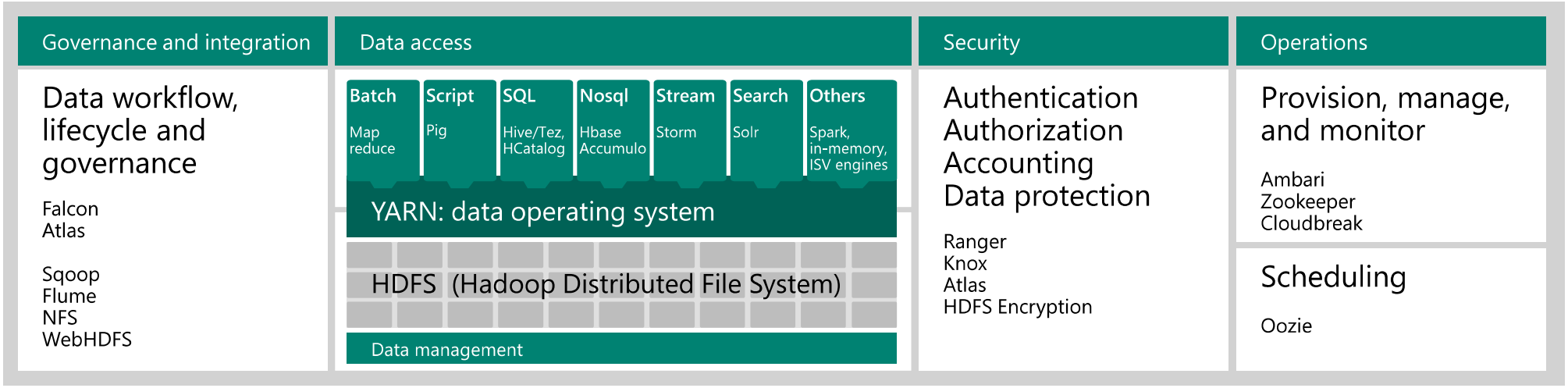
Hadoop is a platform with portfolio of projects

Governed by Apache Software Foundation (ASF)

Comprises core services of MapReduce, HDFS, and YARN

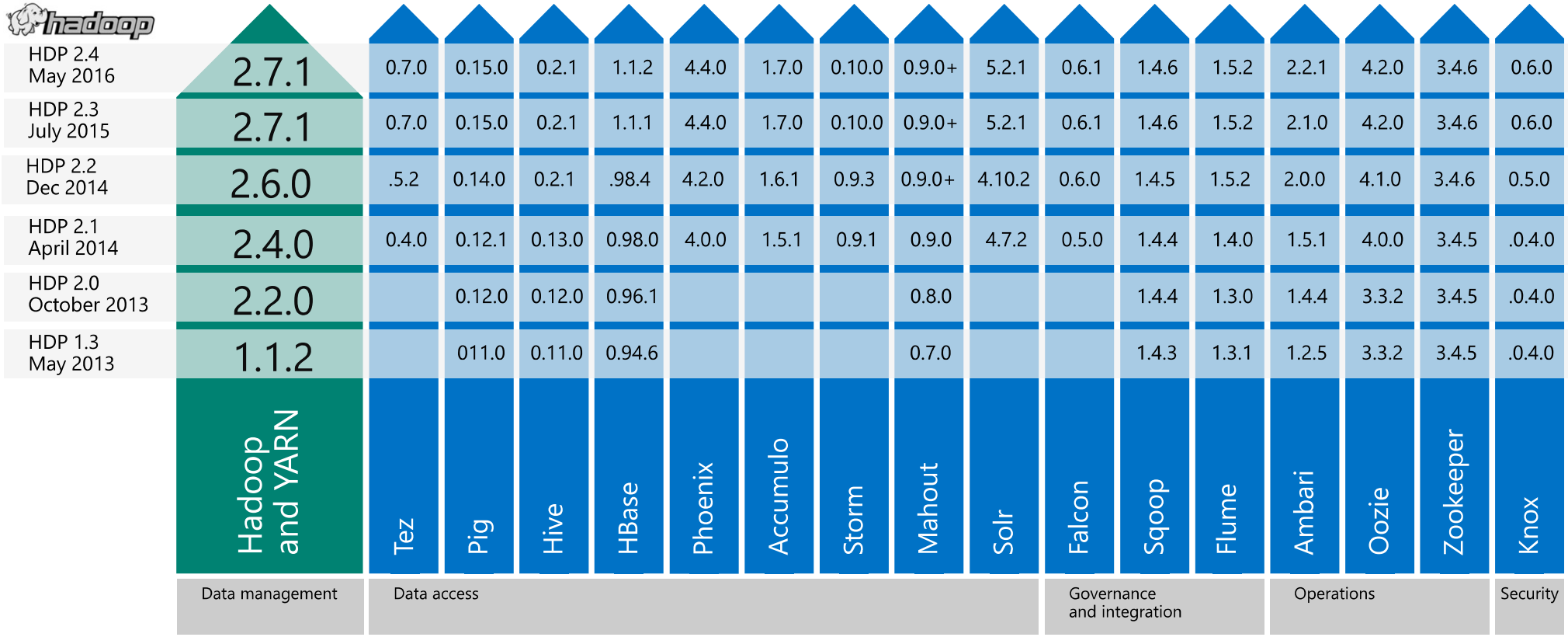
In addition to the core, includes functions across:

Data services which allow you to manipulate and move data (Hive, HBase, Pig, Flume, Sqoop)
Operational services which help manage the cluster (Ambari, Falcon, and Oozie)



A Hadoop distribution is a package of projects

Tested for consistency across entire package



Business applications of Hadoop

Financial services

New account risk screens
Fraud prevention
Trading risk
Maximize deposit spread
Insurance underwriting
Accelerate loan processing

Retail

360° view of the customer
Analyze brand sentiment
Localized, personalized promotions
Website optimization
Optimal store layout

Telecom

Call detail records (CDRs)
Infrastructure investment
Next product to buy (NPTB)
Real-time bandwidth allocation
New product development

Manufacturing

Supplier consolidation
Supply chain and logistics
Assembly line quality assurance
Proactive maintenance
Crowd source quality assurance

Healthcare

Genomic data for medical trials
Monitor patient vitals
Reduce re-admittance rates
Store medical research data
Recruit cohorts for pharmaceutical trials

Utilities, oil, and gas

Smart meter stream analysis
Slow oil well decline curves
Optimize lease bidding
Compliance reporting
Proactive equipment repair
Seismic image processing

Public sector

Analyze public sentiment
Protect critical networks
Prevent fraud and waste
Crowd source reporting for repairs to infrastructure
Fulfill open records requests

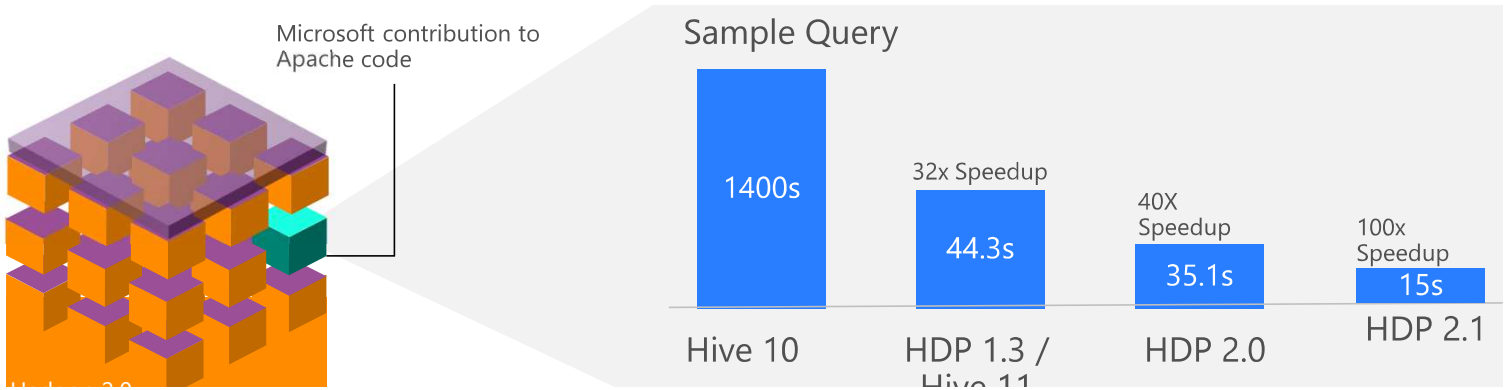
HDInsight Supports Hive

SQL-like queries on Hadoop data in HDInsight

HDInsight provides easy-to-use graphical query interface for Hive
HiveQL is a SQL-like language (subset of SQL)
Hive structures include well-understood database concepts such as tables, rows, columns, partitions
Compiled into MapReduce jobs that are executed on Hadoop

Dramatic performance gains with Stinger/Tez

Stinger is a Microsoft, Hortonworks and OSS driven initiative to bring interactive queries with Hive
Brings query execution engine technology from Microsoft SQL Server to Hive
Performance gains up to 100x



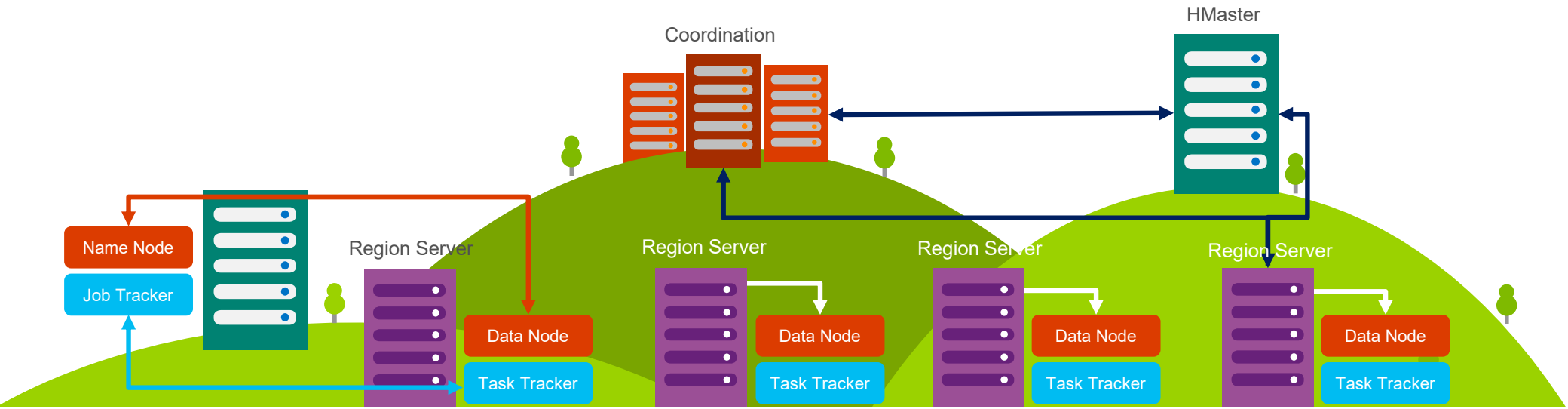
HDInsight Supports HBase

NoSQL database on data in HDInsight

Columnar, NoSQL database

Runs on top of the Hadoop Distributed File System (HDFS)

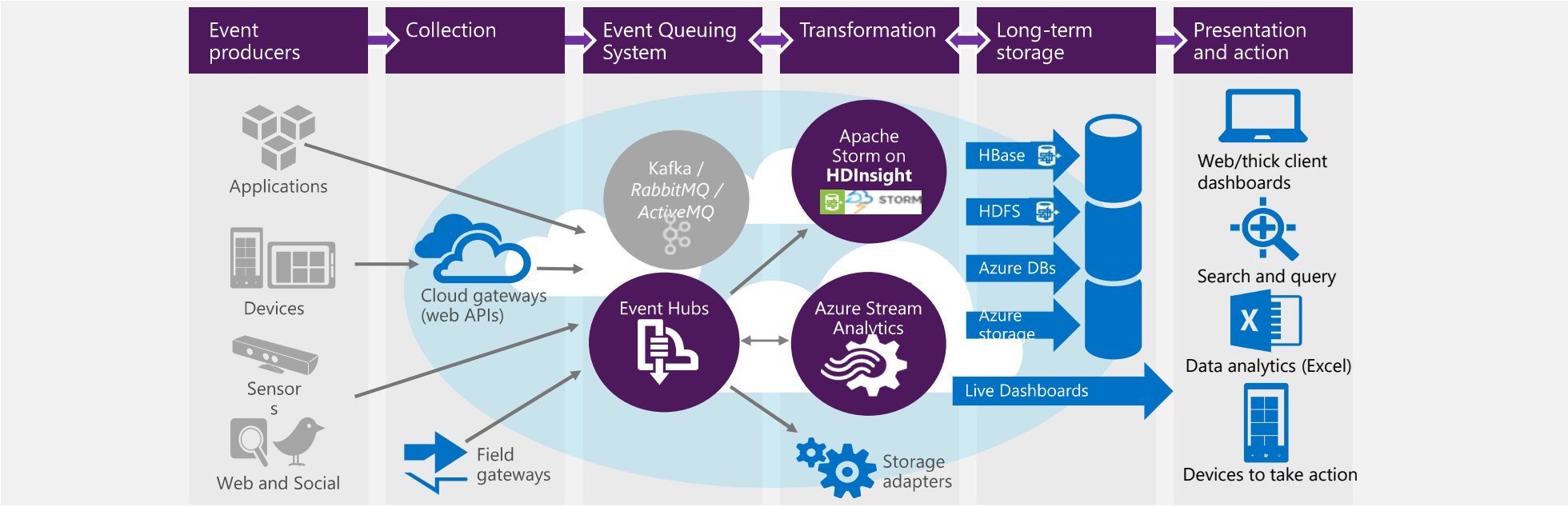
Provides flexibility in that new columns can be added to column families at any time



HDInsight Supports Storm

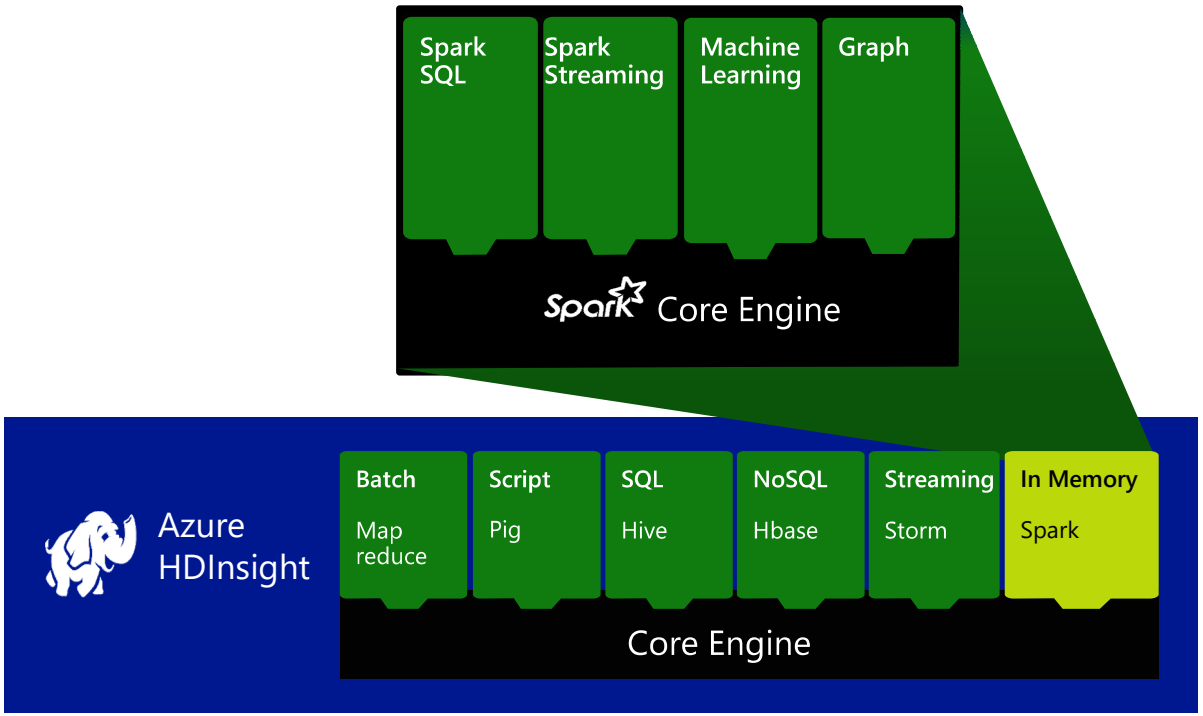
Stream analytics for Near-Real Time Processing

Consumes millions of real-time events from a scalable event broker (ie. Apache Kafka, Azure Event Hub)
Performs time-sensitive computation
Output to persistent stores, dashboards or devices
Customizable with Java + .NET
Deeply integrated to Visual Studio



Spark for Azure HDInsight

In Memory Processing on Multiple Workloads



- Single execution model for multiple tasks
- Processing up to 100x faster performance
- Developer friendly (Java, Python, Scala)
- BI tool of choice (Power BI, Tabelau, Qlik, SAP)
- Notebook experience (Jupyter/iPython, Zeppelin)

R Server for HDInsight

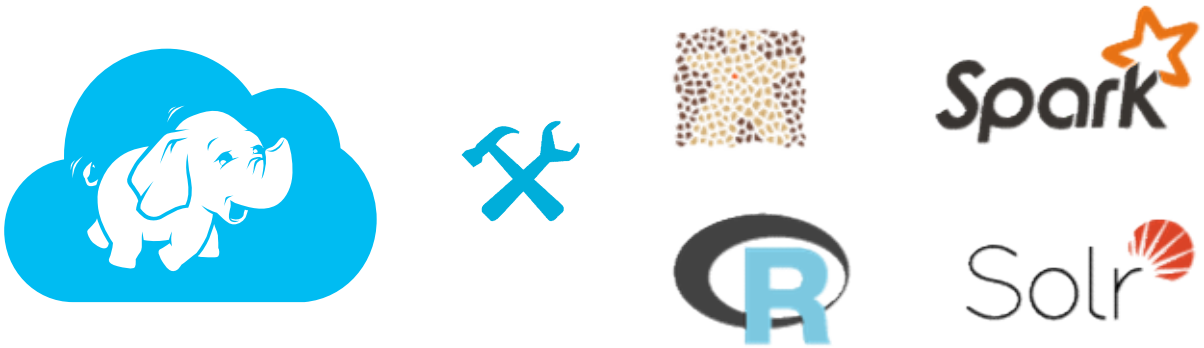
- Familiarity of R (most popular language for data scientists)
- Scalability of Hadoop and Spark
- Up to 7x faster using Spark engine
- Train and run ML models on datasets of any size
- Cloud managed solution (easy setup, elastic, SLA)

Only managed, cloud solution for doing R

HDInsight Allows You To Add Hadoop Projects

Add Hadoop Projects to HDInsight

Modify HDInsight clusters with custom script
Add Apache Hadoop projects to HDInsight
Documented for Spark, R, Giraph, Solr



...

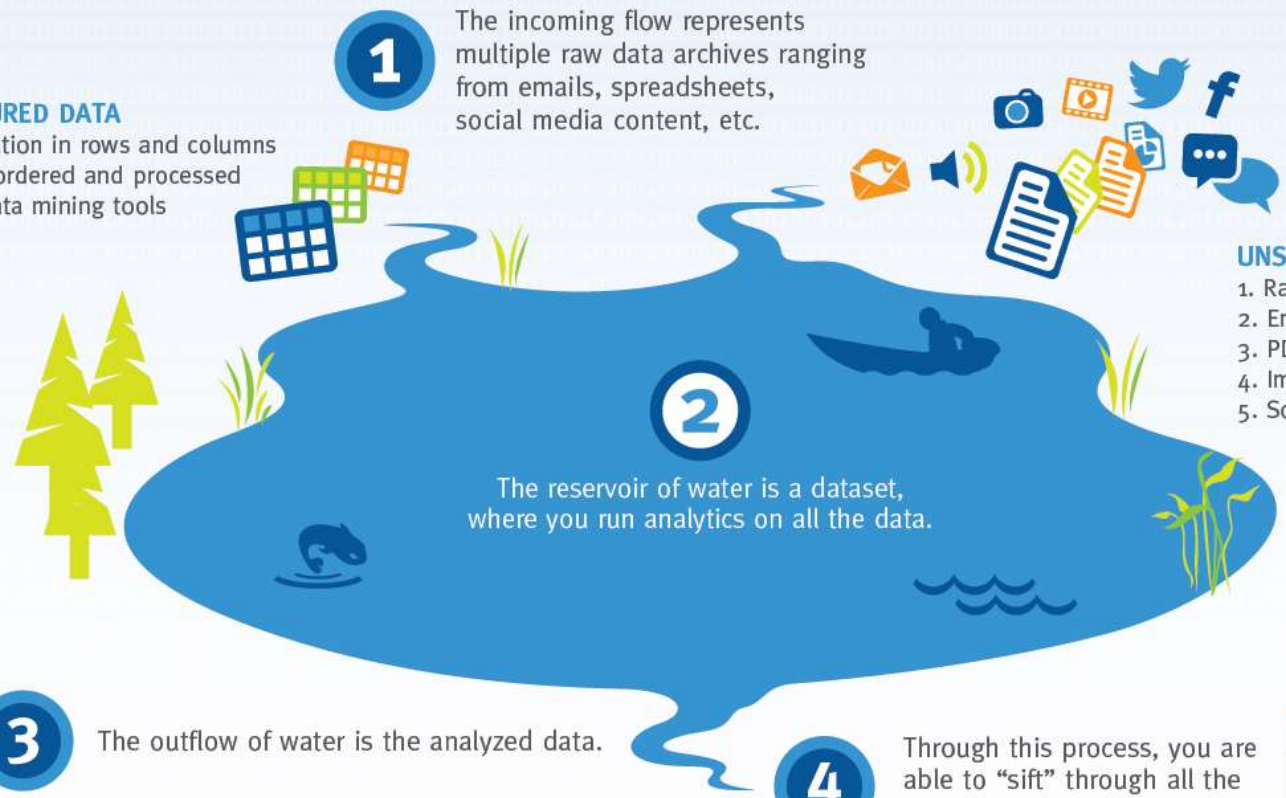
Microsoft

Azure Data Lake

Store + Analytics

HOW DO DATA LAKES WORK?

The concept can be compared to a water body, a lake, where water flows in, filling up a reservoir and flows out.



STRUCTURED DATA
1. Information in rows and columns
2. Easily ordered and processed with data mining tools

1 The incoming flow represents multiple raw data archives ranging from emails, spreadsheets, social media content, etc.

UNSTRUCTURED DATA
1. Raw, unorganized data
2. Emails
3. PDF files
4. Images, video and audio
5. Social media tools

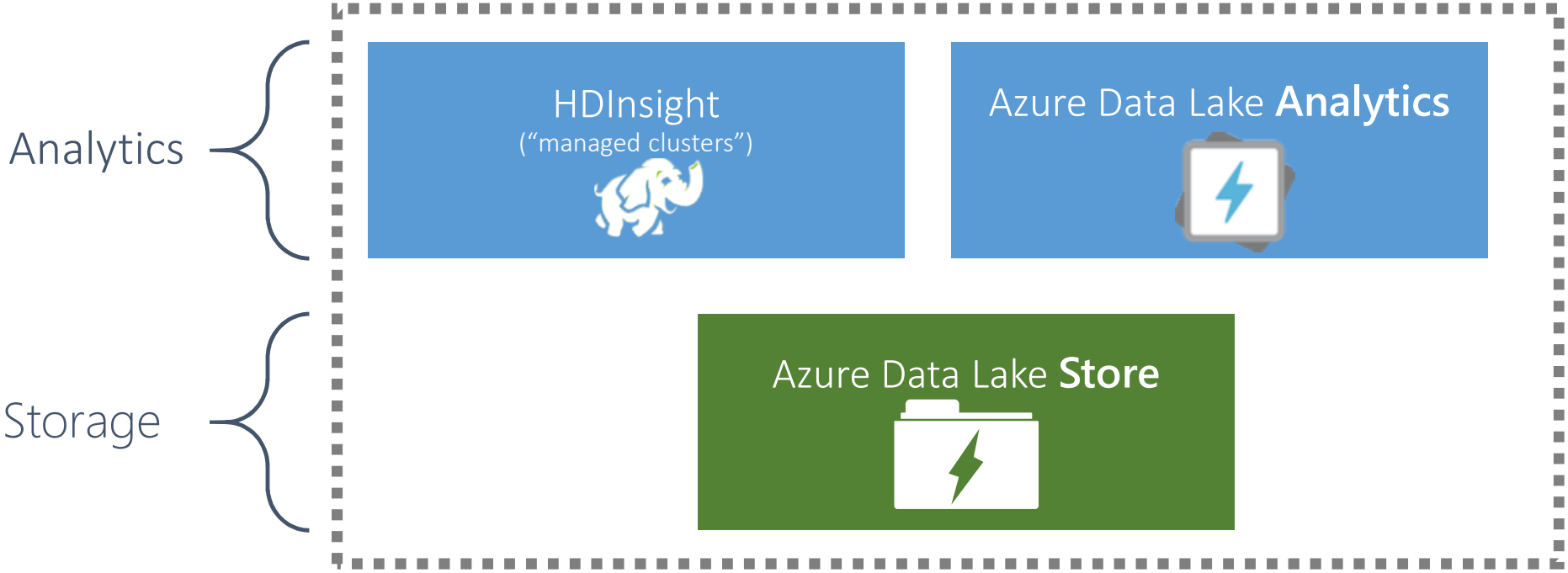
2 The reservoir of water is a dataset, where you run analytics on all the data.

3 The outflow of water is the analyzed data.

4 Through this process, you are able to "sift" through all the data quickly to gain key business insights.

DATA WAREHOUSE	vs.	DATA LAKE
structured, processed	DATA	structured / semi-structured / unstructured, raw
schema-on-write	PROCESSING	schema-on-read
expensive for large data volumes	STORAGE	designed for low-cost storage
less agile, fixed configuration	AGILITY	highly agile, configure and reconfigure as needed
mature	SECURITY	maturing
business professionals	USERS	data scientists et. al.

Azure Data Lake



Azure Data Lake Store

A hyper-scale repository for Big Data analytics workloads

Hadoop File System (HDFS) for the cloud

No limits to scale

Store **any data** in its native format

Enterprise-grade access control, encryption at rest

Optimized for analytic workload **performance**

Azure Data Lake Analytics

A new distributed analytics service

Distributed analytics service built on Apache YARN

Elastic scale per query lets users focus on business goals—not configuring hardware

Includes U-SQL—a language that unifies the **benefits of SQL with the expressive power of C#**

Integrates with Visual Studio to develop, debug, and tune code faster

Federated query across Azure data sources

Enterprise-grade **role based access control**

Azure Machine Learning

“I need our systems to think.
I need them to learn and
I need them to present issues
and problems and anomalies
to the employees, to the managers.”

Adam Coffey
President and CEO
WASH Laundry Systems

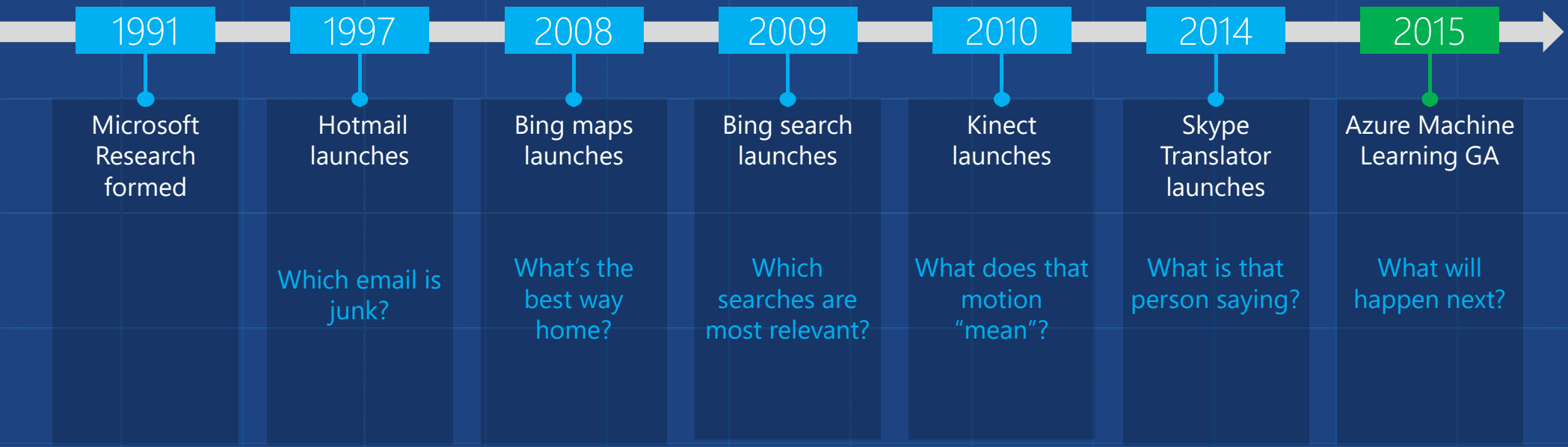
What is Machine Learning?

Computing systems that become smarter with experience

“Experience” =
past data + human input

Microsoft & Machine Learning

Answering questions with experience



Machine learning is pervasive throughout Microsoft products.

The old Machine Learning landscape

No improvement in generations

Expensive	Huge set-up costs of tools, expertise, and compute/storage capacity
Siloed data	Siloed and cumbersome data management restricts access to data
Disconnected tools	Complex and fragmented tools limit participation in exploring data and building models
Deployment complexity	Many models never achieve business value due to difficulties with deploying to production

Azure Machine Learning

Powerful predictive analytics in Azure

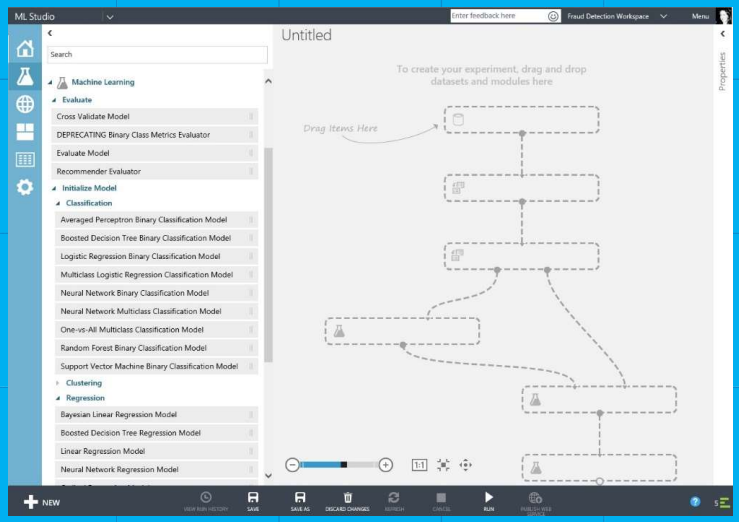
ML Algorithms are best of breed and embrace OSS
- MS + R + Python + BYOA

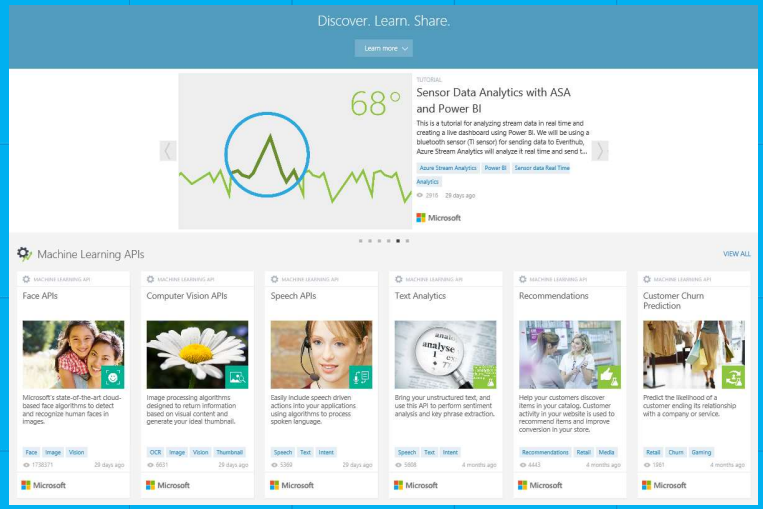
ML Studio for productive development
- Faster experiments results in faster improvements
- Visual Workflows & ML Experiments

ML Operationalization to remove deployment friction
- Build entire ML Apps & Deploy as Cloud APIs

ML Gallery
- Provide ML applications like apps in an ‘app store’
- Publish/consume APIs in a 2 sided market

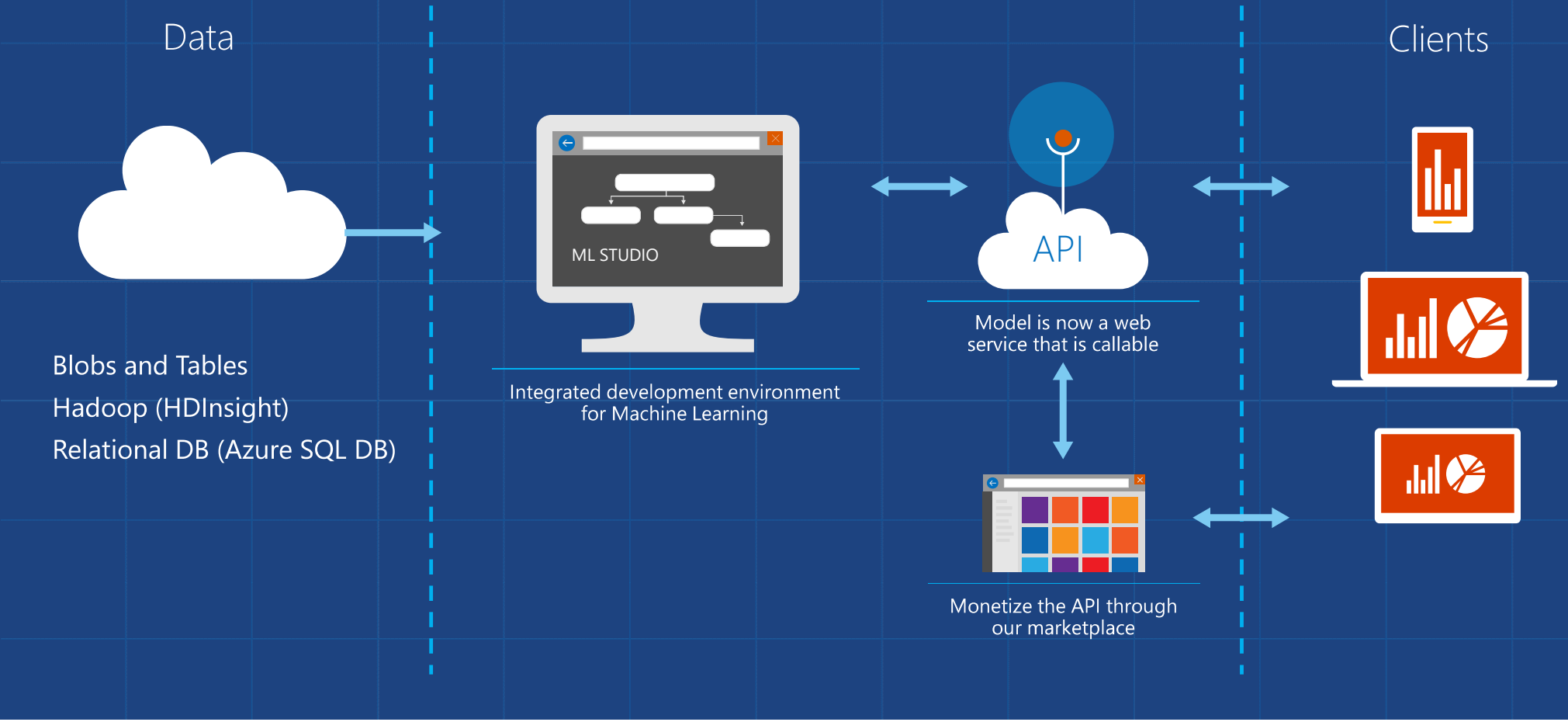
Help organizations eliminate undifferentiated heavy lifting





Azure Machine Learning Service

Data -> Predictive model -> Operational web API in minutes



What can Azure ML do for you…?

Telemetry data analysis

Buyer propensity models

Social network analysis

Predictive maintenance

Web app optimization

Churn analysis

Natural resource exploration

Weather forecasting

Healthcare outcomes

Fraud detection

Life sciences research

Targeted advertising

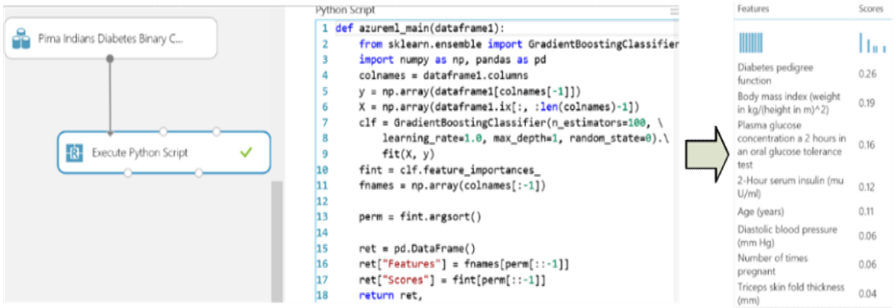
Network intrusion detection

Smart meter monitoring

Model Your Way: Open source/our source

Script with R, SQLite or Python

CPython 2.7 support from inside AML Studio
numpy/scipy/panda/scikit-learn/etc.
Anaconda distro pre-installed

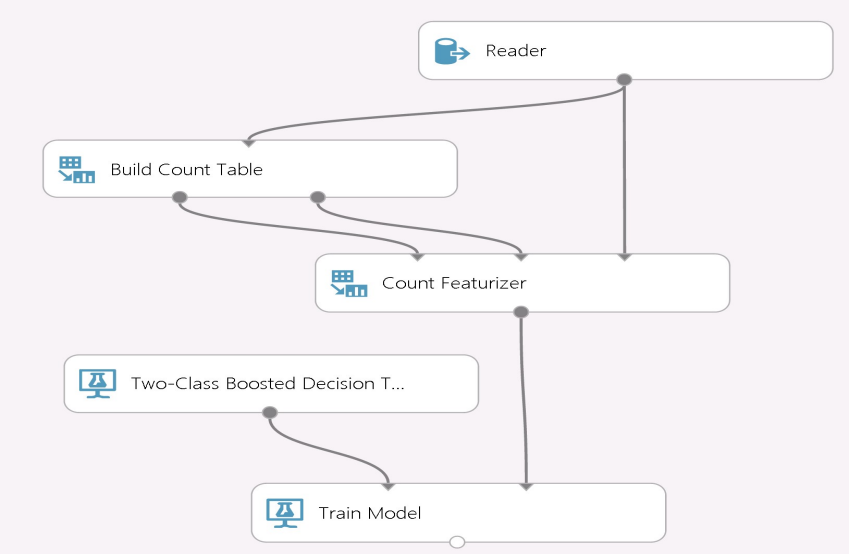


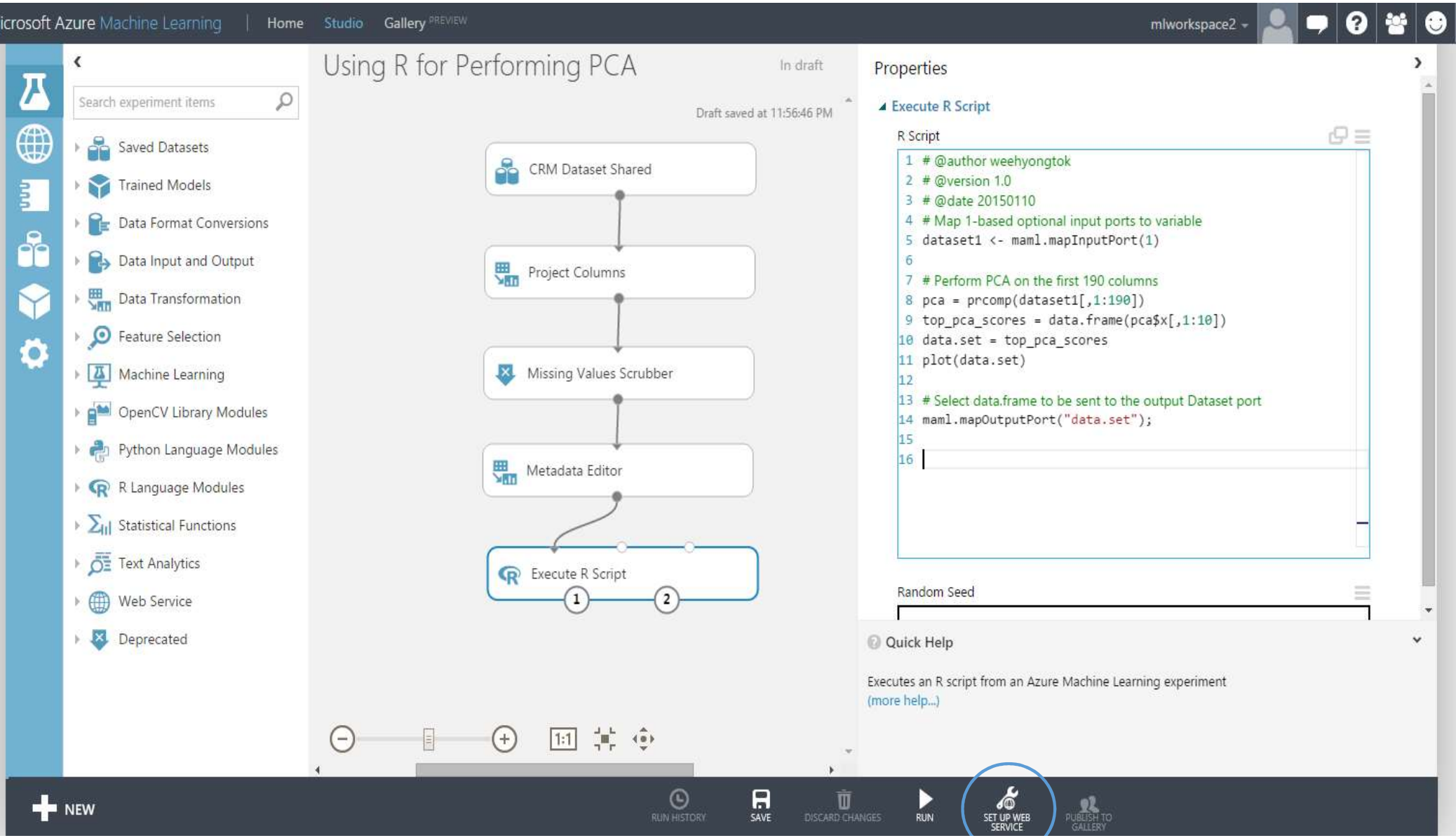
Python client library

Analyze data using Python and its libraries
Use IPython, PTVS, Eclipse to edit/debug

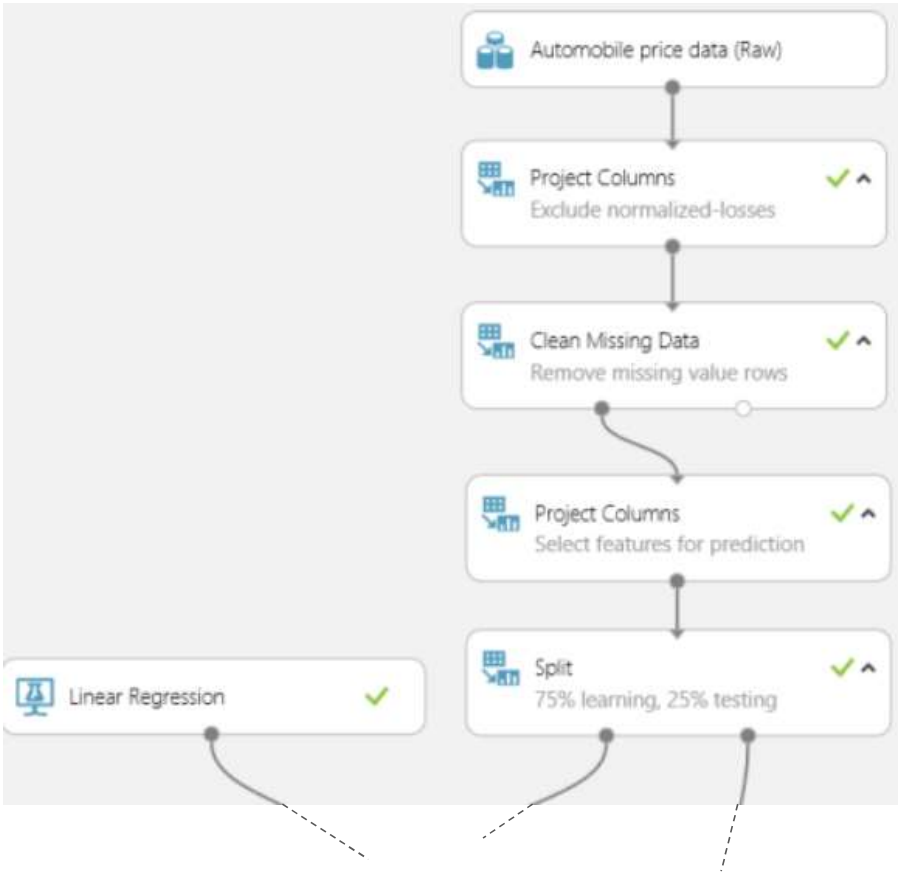
Big learning with counts

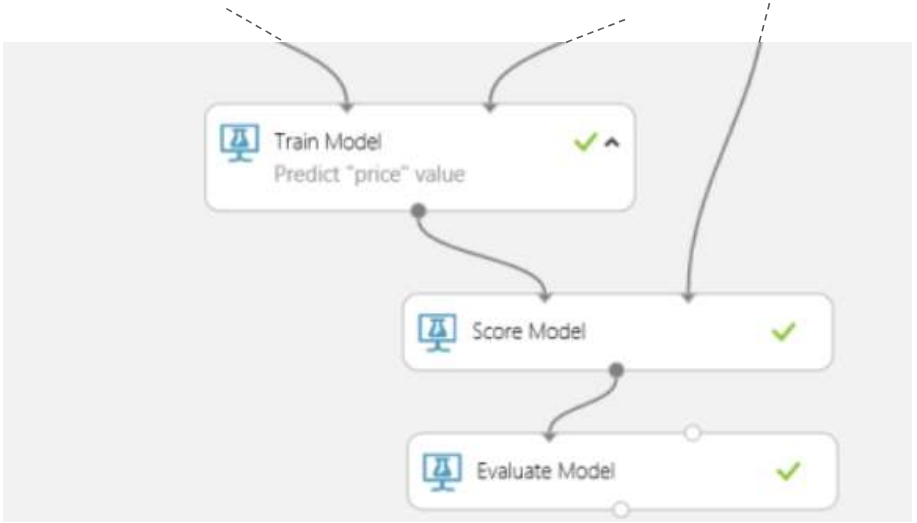
TB scale datasets
Modular: tune/monitor/replace in isolation
Monitorable and debuggable




Microsoft Azure Machine Learning
Home
Studio
Gallery PREVIEW
mlworkspace2
Search experiment items
Saved Datasets
Trained Models
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
Deprecated
Using R for Performing PCA
In draft
Draft saved at 11:56:46 PM
CRM Dataset Shared
Project Columns
Missing Values Scrubber
Metadata Editor
Execute R Script
1
2
Properties
Execute R Script
R Script
1 # @author weehyongtok
2 # @version 1.0
3 # @date 20150110
4 # Map 1-based optional input ports to variable
5 dataset1 <- maml.mapInputPort(1)
6
7 # Perform PCA on the first 190 columns
8 pca = prcomp(dataset1[,1:190])
9 top_pca_scores = data.frame(pca$x[,1:10])
10 data.set = top_pca_scores
11 plot(data.set)
12
13 # Select data.frame to be sent to the output Dataset port
14 maml.mapOutputPort("data.set");
15
16
Random Seed
Quick Help
Executes an R script from an Azure Machine Learning experiment
(more help...)
NEW
RUN HISTORY
SAVE
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Machine Learning - Studio





Cinco etapas para criar um experimento

Neste tutorial de Aprendizado de Máquina, você seguirá cinco etapas básicas para criar uma experiência do Studio de Aprendizado de Máquina para criar, treinar e pontuar o seu modelo:

- Criar um modelo
  - Etapa 1: Obter dados
  - Etapa 2: Pré-processar dados
  - Etapa 3: Definir recursos
- Treinar o modelo
  - Etapa 4: Escolher e aplicar um algoritmo de aprendizado
- Pontuar e testar o modelo
  - Etapa 5: Prever novos preços de automóveis

Machine Learning – experiments samples

(at Cortana Gallery)

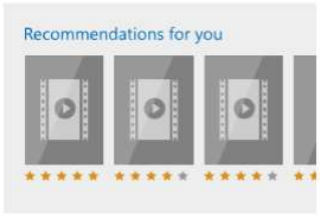
EXPERIMENT

Binary Classification: Twitter sentiment ana...

This experiment demonstrates the use of the Execute R Script, Feature Selection, Feature Hashi...

EXPERIMENT

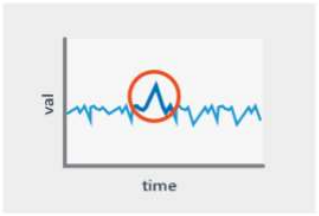
Recommender: Movie recommendation



This experiment demonstrates the use of the Matchbox recommender modules to train ...

EXPERIMENT

Anomaly Detection: Credit Risk



Attempts to predict credit risk as anomalies within the data.

EXPERIMENT

Telco Customer Churn

Customer churn can take different forms, such as switching to a competitor's service, reducing th...

EXPERIMENT

Regression: Demand estimation

This experiment demonstrates demand estimation using regression with UCI bike rental d...

EXPERIMENT

Clustering: Find similar companies

This experiment clusters similar companies into same group given their Wikipedia articles and can ...

EXPERIMENT

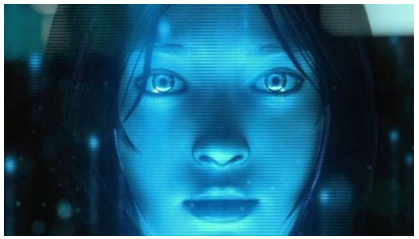
Binary Classification: Prediction of student ...

Predict if a student will solve a given problem from the first attempt

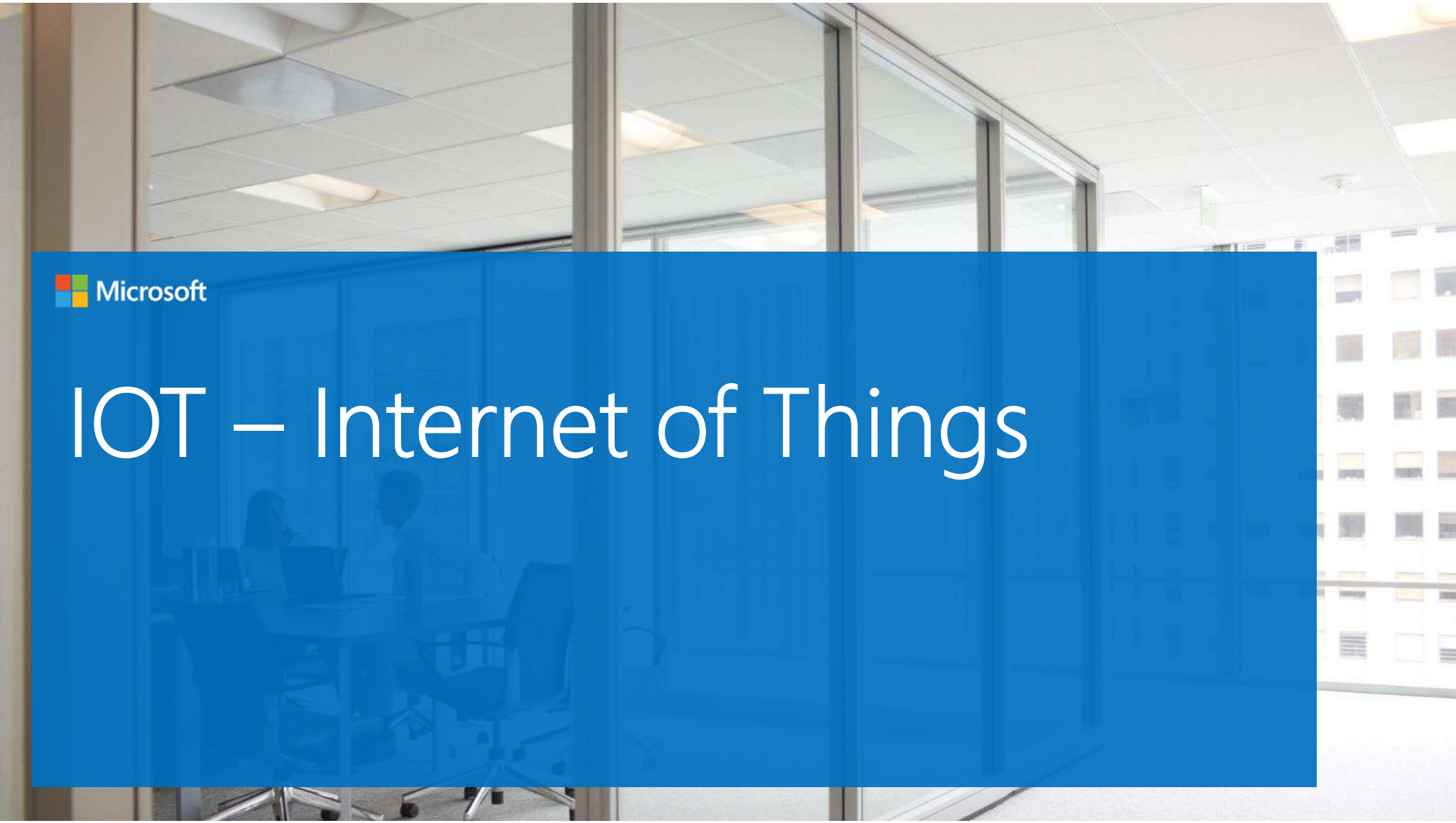
EXPERIMENT

Binary Classification: Flight delay prediction

In this experiment, we predict whether scheduled passenger flight is delayed or not using a B...



https://gallery.cortanaanalytics.com/



IOT – Internet of Things


Source: NBC News and The Associated Press. (http://photoblog.nbcnews.com/)

2005

Unveiled: Crowds in St Peter's Square use mobile phones and tablet computers to get a picture of the new Pope as he is unveiled at the Vatican

2013

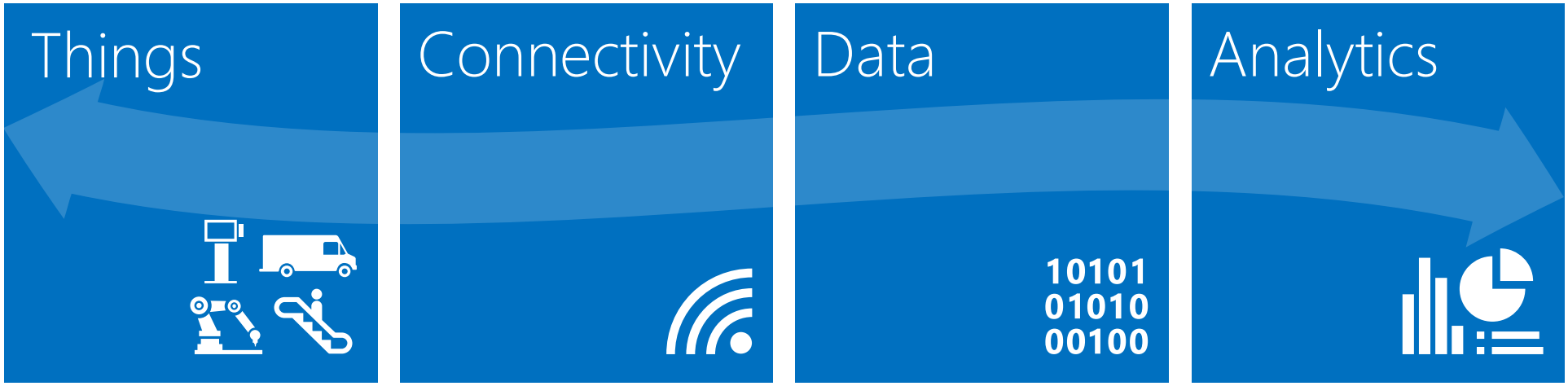
St Peter's Square, Vatican City, 2005 - crowded with tourists and just the odd camera phone in sight.

(Luca Bruno/AP)

St Peter's Square, Vatican City, 2013 - crowded with tourists as the election of the new Pope (Pope Francis) is announced, and the progress of technology during those 8 years is apparent.

(Michael Sohn/AP)

Definition of IoT – Internet of Things



25 billion

Connected "things" by 2020

—Gartner

$1.7 trillion

Market for IoT by 2020

—IDC

IOT 2010


Alarm clock
Refrigerator
Automobile
VoIP phone
Computer
Printer
Media player
Microwave
HOME
WORKPLACE
HOME
Coffee maker
Smart scale
Oven
Cell phone
HVAC
Vending
Security
Television

IOT 2016


Medication adherence


Health monitoring


Pet tracking


Behavior modification


Object tracking


Child and elder monitoring


Sports and fitness


Smart lighting


Indoor navigation


Beacons and proximity


Trip tracking and car health

HOME **COMMUTE** **WORKPLACE** **COMMUTE** **HOME**


Smart appliances


Food and nutrition tracking


Identity


Office equipment


Smart vending machines


Bike ride stats and protection


Sleep tracking


Air conditioning and temperature control


Environmental sensors


Information capture


Control


Home security

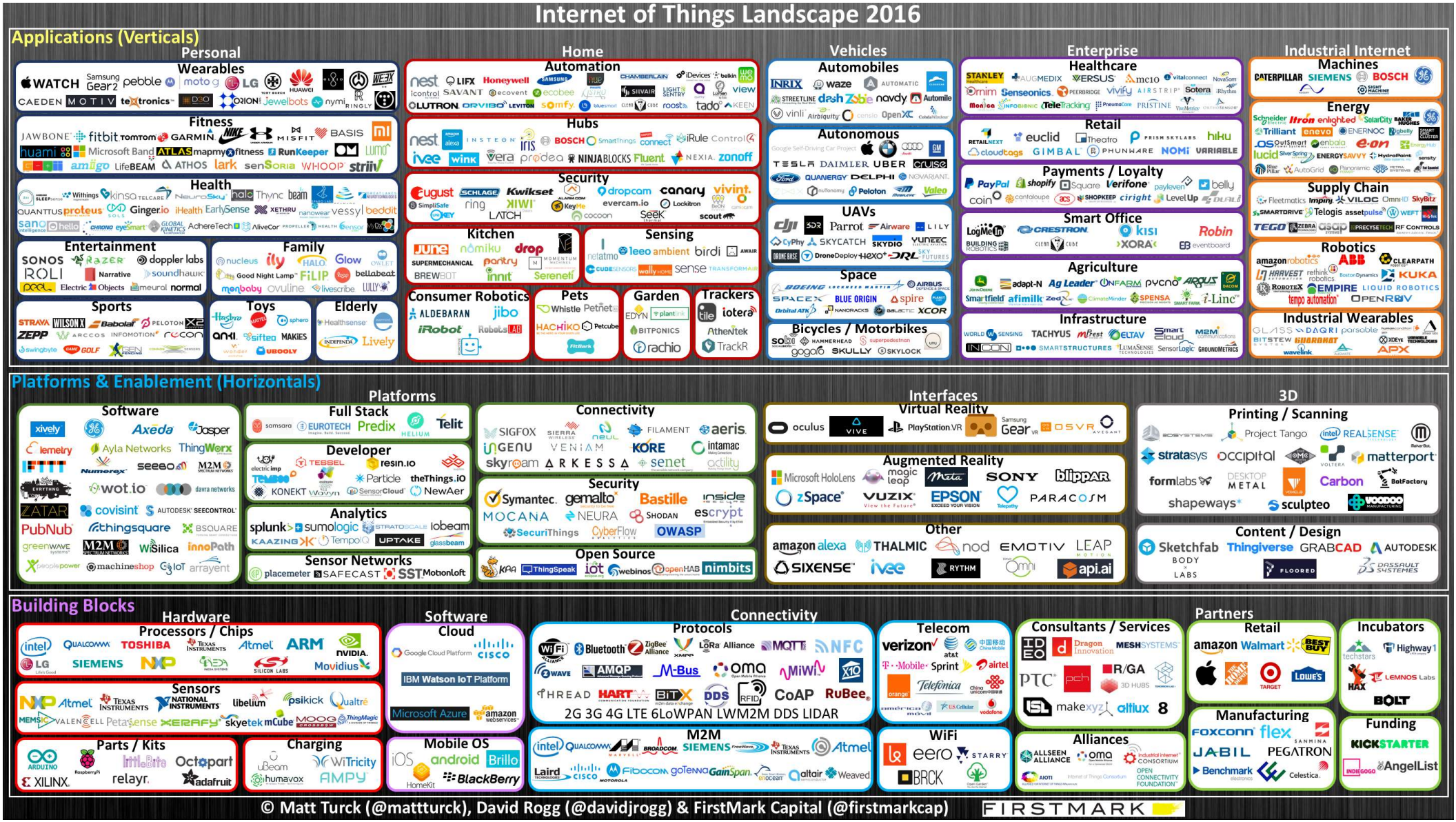

Home automation


Leak detection


Garden, lawn and plant care


New devices and sensors

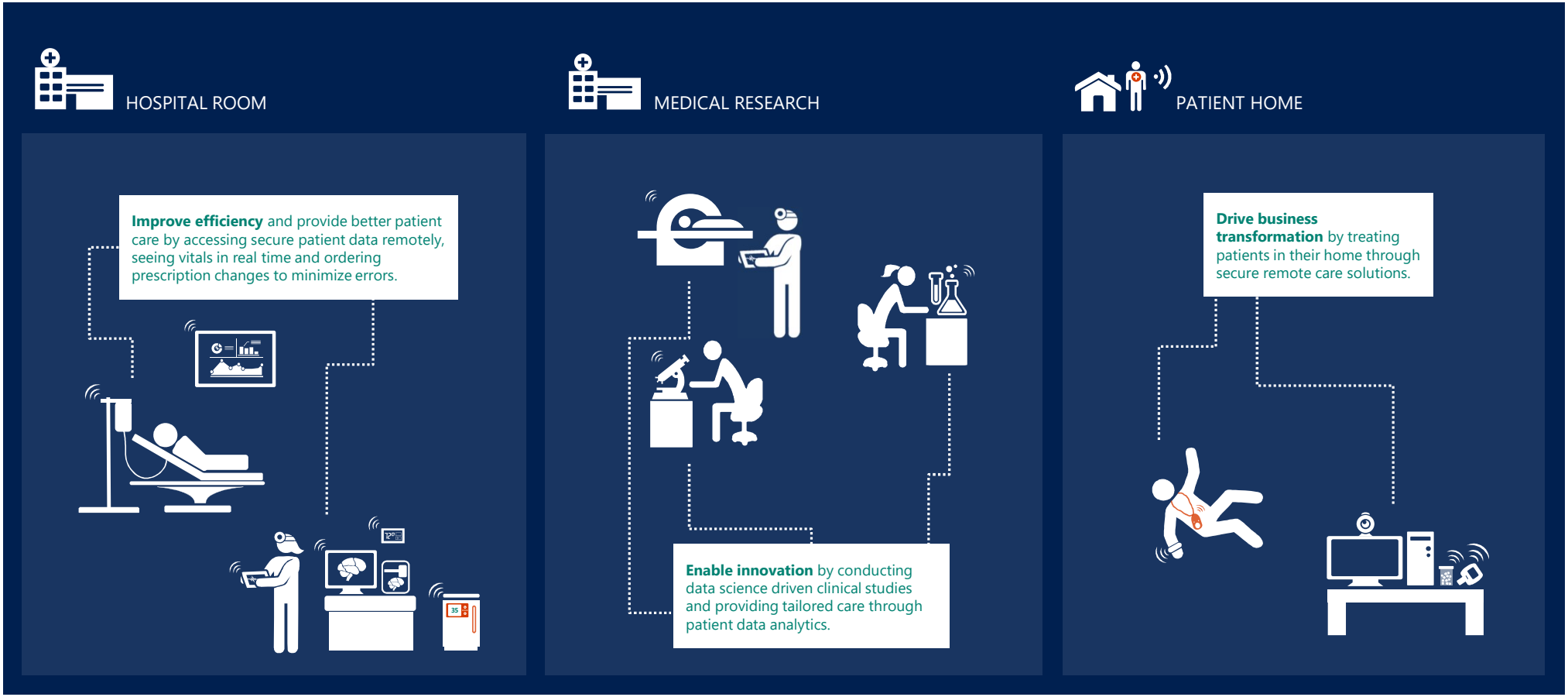

Entertainment systems


Internet of Things Landscape 2016
Applications (Verticals)
Personal
Wearables
Fitness
Health
Entertainment
Family
Sports
Toys
Elderly
Home
Automation
Hubs
Security
Kitchen
Sensing
Consumer Robotics
Pets
Garden
Trackers
Vehicles
Automobiles
Autonomous
UAVs
Space
Bicycles / Motorbikes
Enterprise
Healthcare
Retail
Payments / Loyalty
Smart Office
Agriculture
Infrastructure
Industrial Internet
Machines
Energy
Supply Chain
Robotics
Industrial Wearables
Platforms & Enablement (Horizontals)
Platforms
Software
Full Stack
Developer
Analytics
Sensor Networks
Connectivity
Security
Open Source
Interfaces
Virtual Reality
Augmented Reality
Other
3D
Printing / Scanning
Content / Design
Building Blocks
Hardware
Processors / Chips
Sensors
Parts / Kits
Charging
Software
Cloud
Mobile OS
Connectivity
Protocols
2G 3G 4G LTE 6LoWPAN LWM2M DDS LIDAR
M2M
WiFi
Telecom
Consultants / Services
Alliances
Partners
Retail
Manufacturing
Incubators
Funding
© Matt Turck (@mattturck), David Rogg (@davidjrogg) & FirstMark Capital (@firstmarkcap)
FIRSTMARK

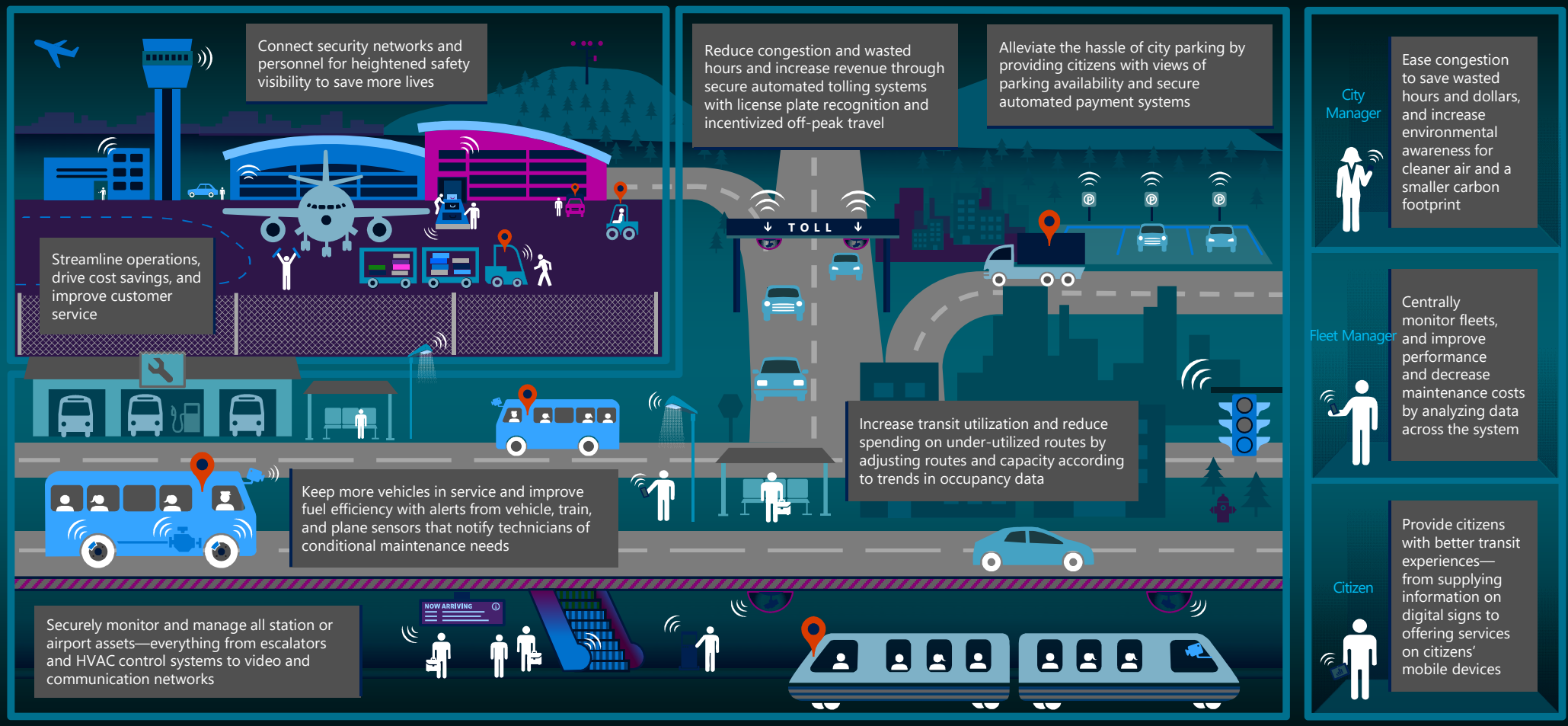
IoT – Retail


STORE
LOGISTICS
HEADQUARTERS
Enable innovation by dynamically managing digital promotions on vending machines, and display content tailored to specific audiences.
Integrate data from smart shelves with point of sale and trade promotions data to determine the most productive merchandising treatments within stores.
Improve efficiency by tracking inventory location and shipping status end-to-end to maintain on-shelf availability and reduce out-of-stocks.
Drive business transformation by aggregating data from many sources to create a demand signal repository and enable supply chain visibility.

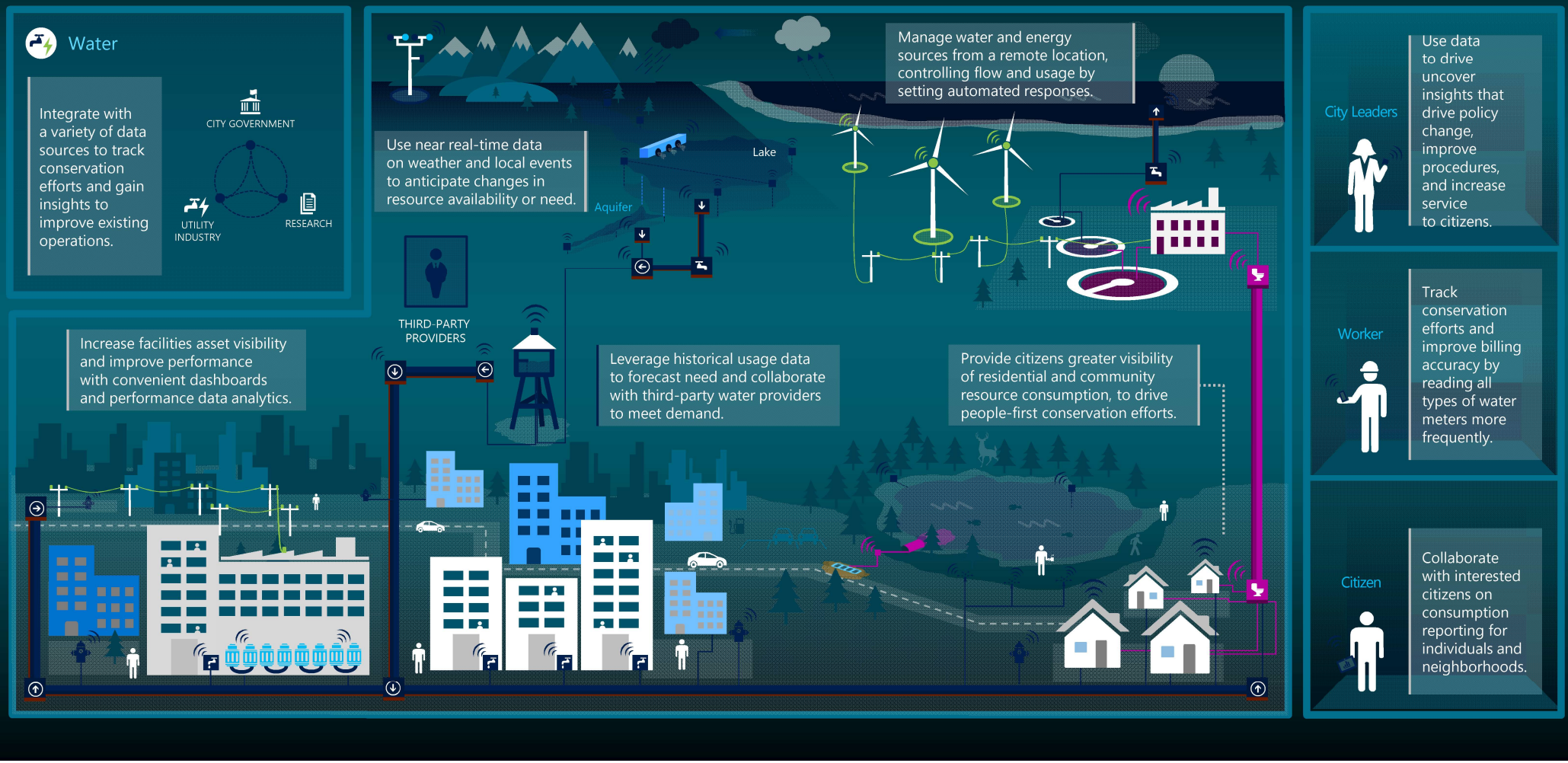
IoT – Health Care



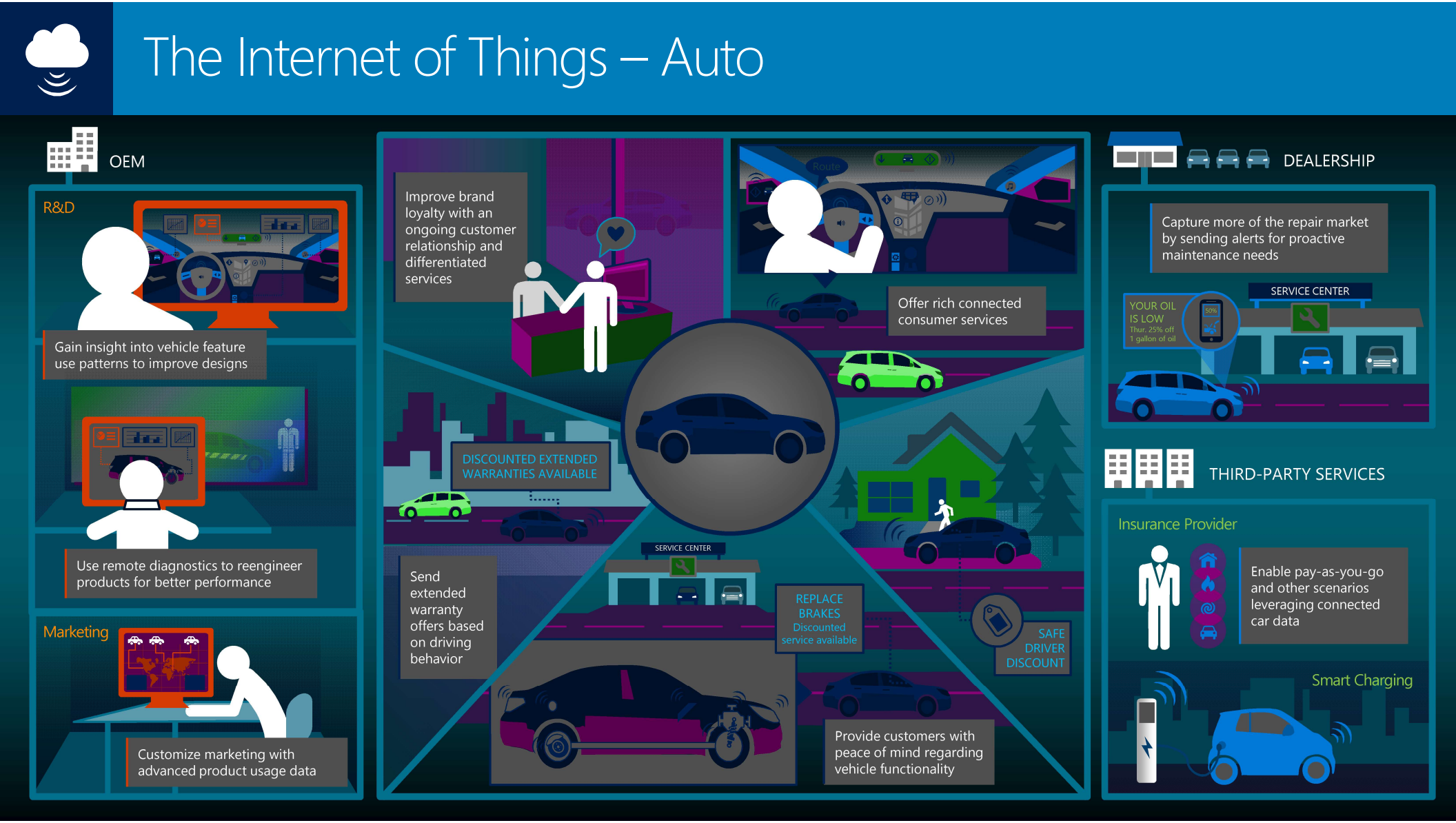
The Internet of Things – Public Transportation

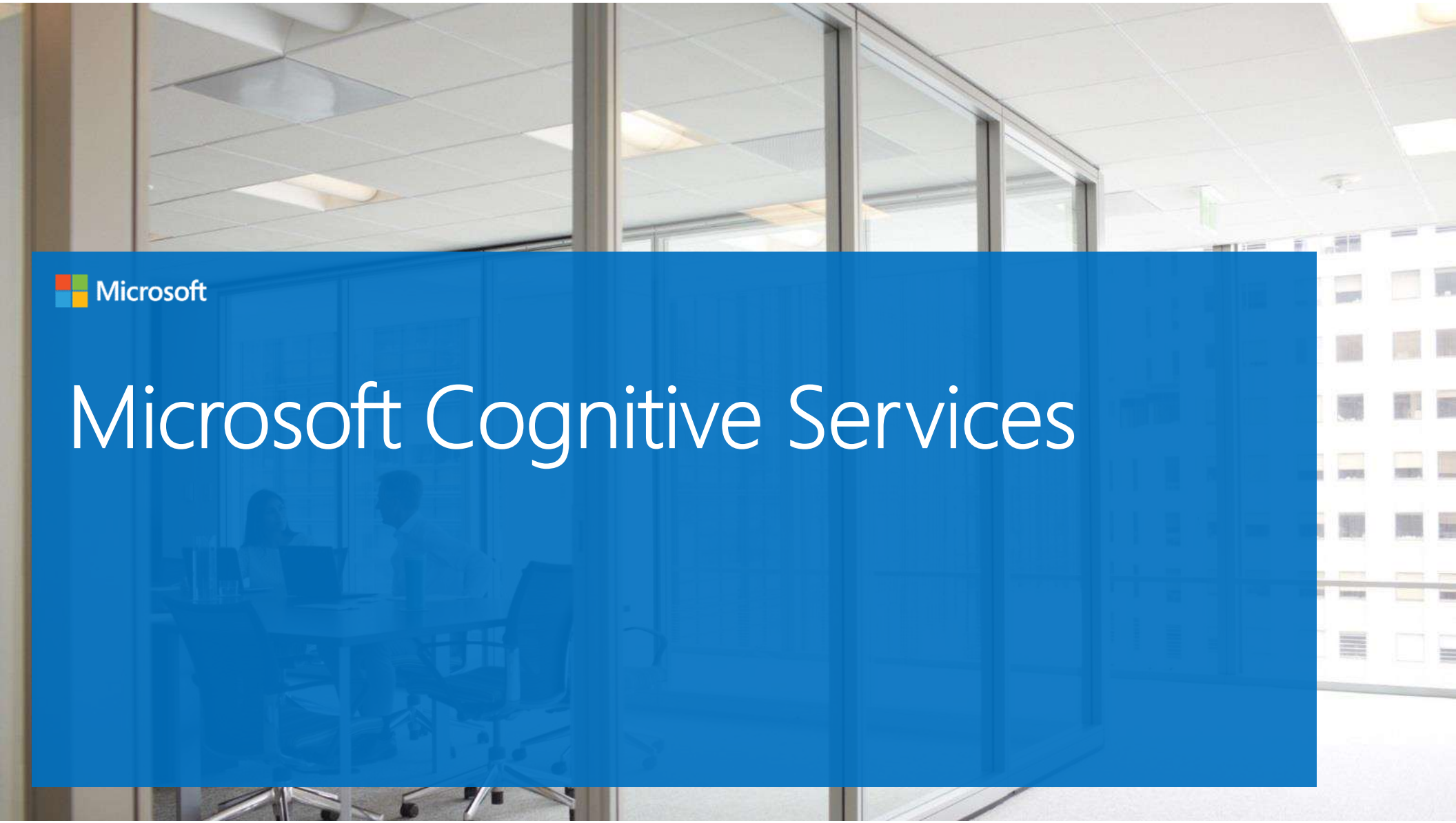


The Internet of Things – Water



The Internet of Things – Auto





Microsoft Cognitive Services

Analyze image

Type of image

Clip Art Type	0 Non-clipart
Line Drawing Type	0 Non-Line Drawing
Black & White Image	False

Content of image

Categories	`[{ “name”: “people_swimming”, “score”: 0.099609375 }]`
Adult Content	False
Adult Score	0.18533889949321747
Faces	`[{ “age”: 27, “gender”: “Male”, “faceRectangle”: {“left”: 472, “top”: 258, “width”: 199, “height”: 199}}]`

Image colors

Dominant Color Background	White
Dominant Color Foreground	Grey
Dominant Colors	White
Accent Color	

Emotion APIs

Face detection

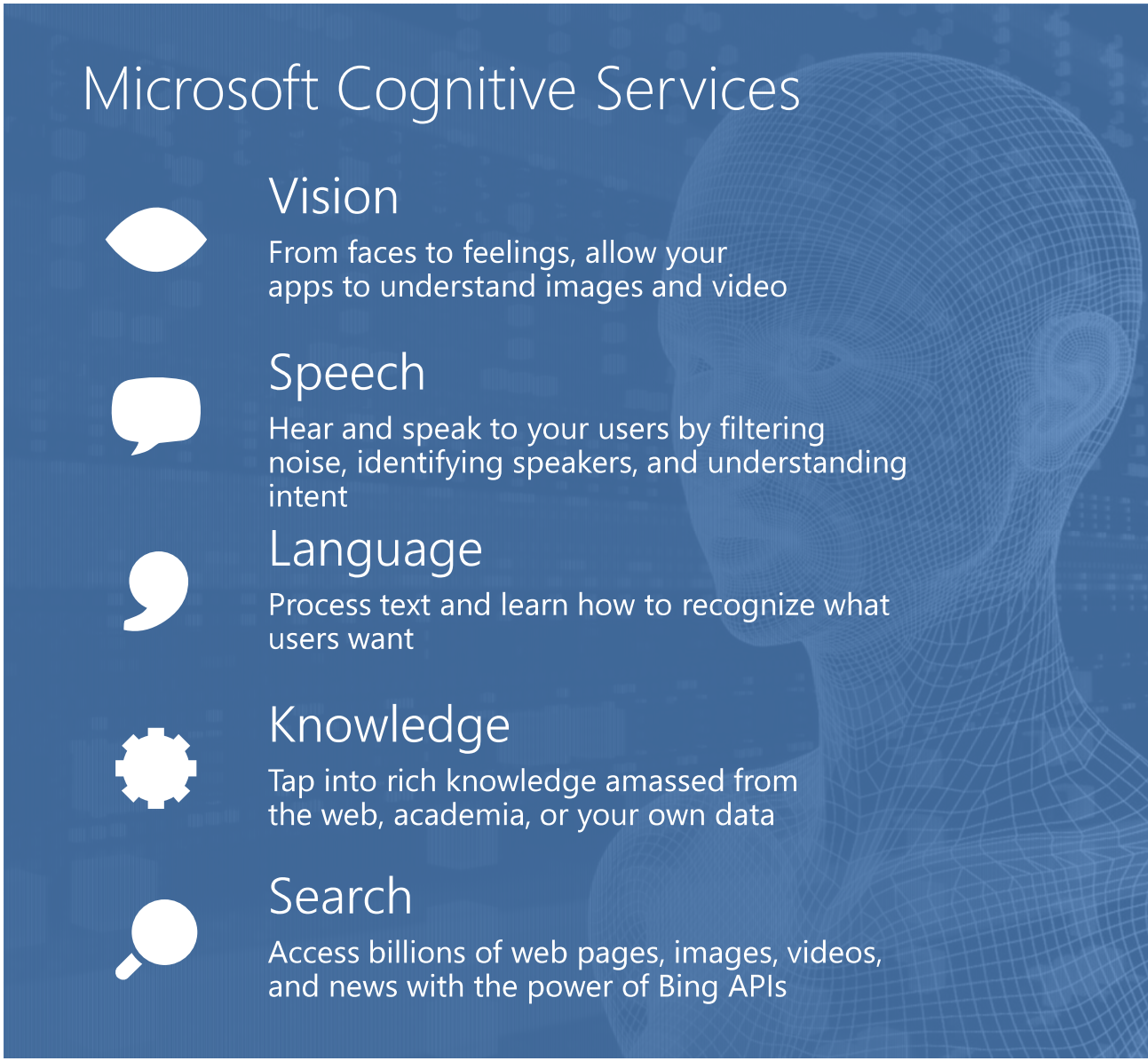
```
"faceRectangle": {"width": 193,
                  "height": 193,
                  "left": 326,
                  "top": 204}…
```

Emotion scores

```
“scores": { "anger": 5.182241e-8,
            "contempt": 0.0000242813,
            "disgust": 5.621025e-7,
            "fear": 0.00115027453,
            "happiness": 1.06114619e-8,
            "neutral": 0.003540177,
            "sadness": 9.30888746e-7,
            "surprise": 0.9952837}
```

Cognitive Services

Give your solutions a human side



Build applications that understand people

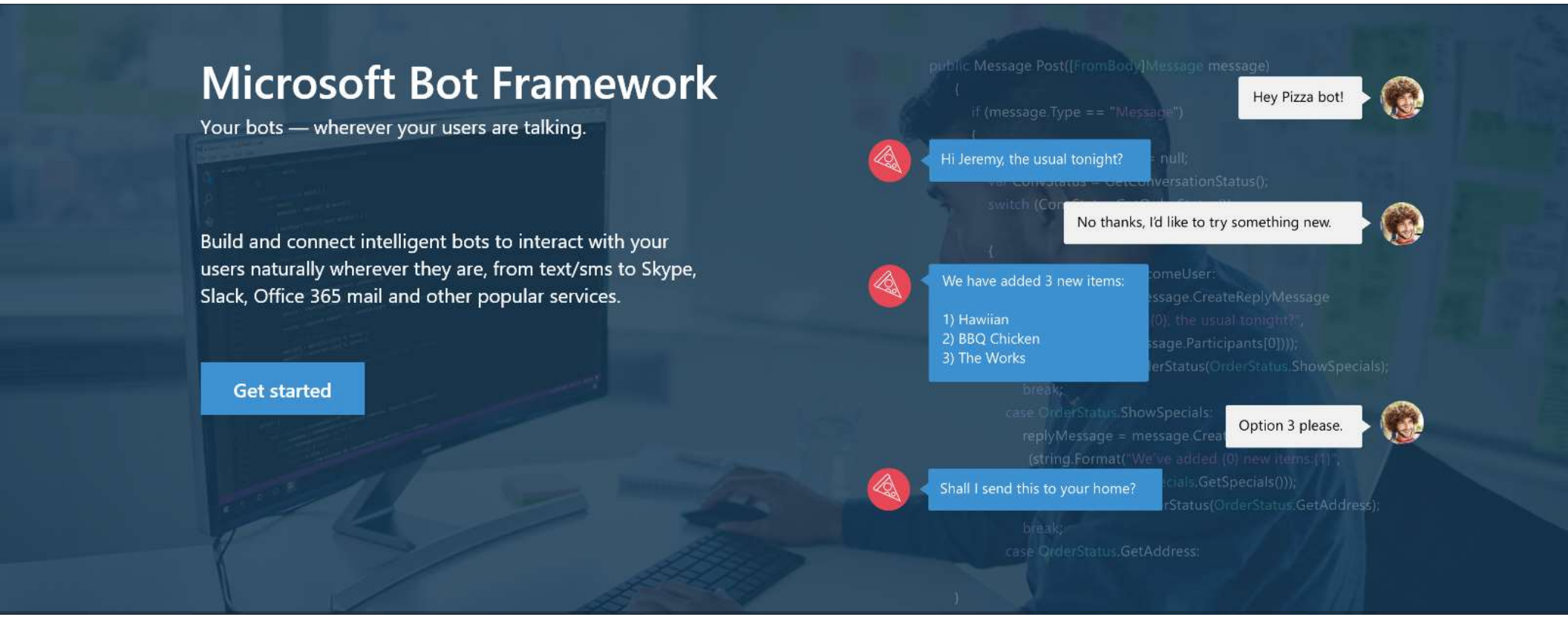
Intelligence

- Cognitive Services
- Bot Framework
- Cortana

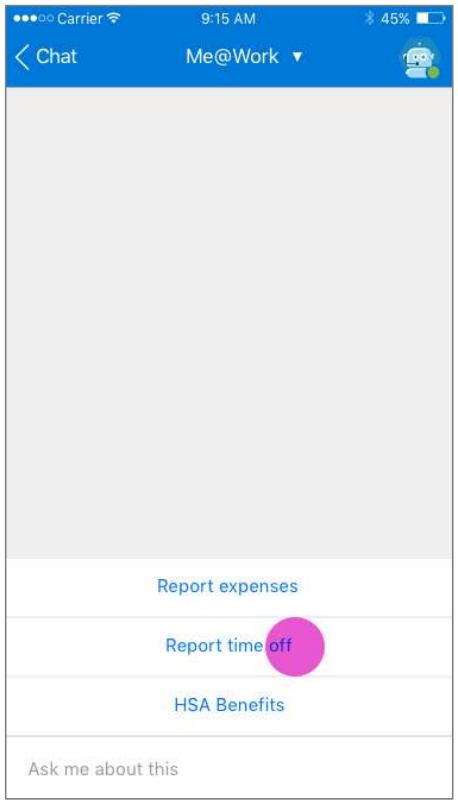
Vision	Speech	Language	Knowledge	Search
Computer Vision	Speaker Recognition	Text Analytics	Academic Knowledge	Bing Search API
Face	Speech	Bing Speller	Entity Linking Service	Bing Image Search API
Emotion	CRIS	Web Language Model	Knowledge Exploration Service	Bing Video Search API
Video		Linguistic Analysis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

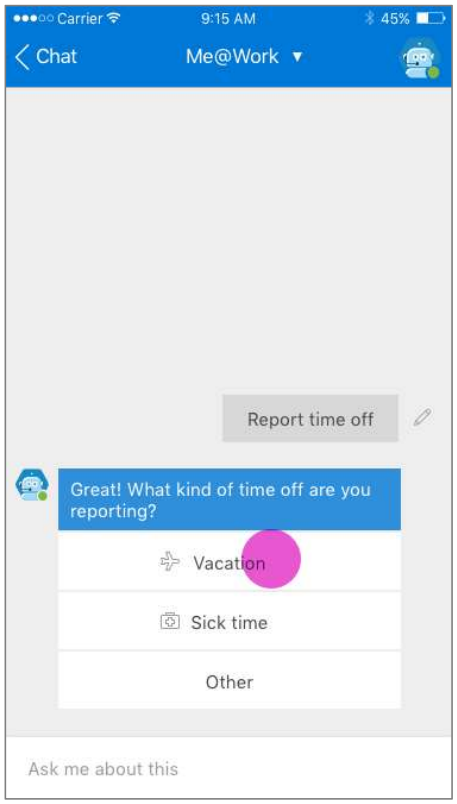
- Faces, images, emotion recognition and video intelligence
- Spoken language processing, speaker recognition, custom speech recognition
- Natural language processing, sentiment and topics analysis, spelling errors
- Complex tasks processing, knowledge exploration, intelligent recommendations
- Bing engine capabilities for Web, Autosuggest, Image, Video and News

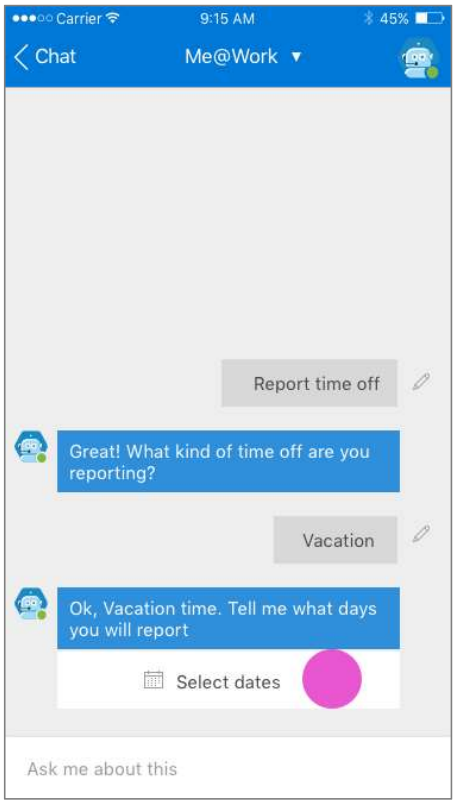
BOT Framework

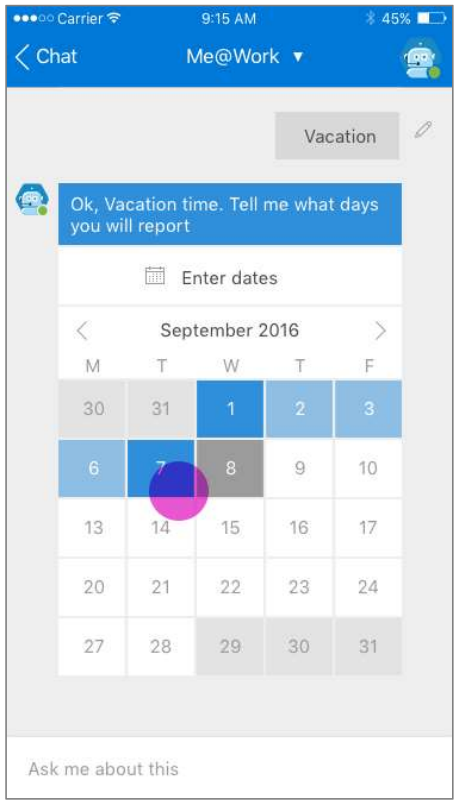

Microsoft Bot Framework
Your bots — wherever your users are talking.
Build and connect intelligent bots to interact with your users naturally wherever they are, from text/sms to Skype, Slack, Office 365 mail and other popular services.
Get started
Hey Pizza bot!
Hi Jeremy, the usual tonight?
No thanks, I'd like to try something new.
We have added 3 new items:
1) Hawiian
2) BBQ Chicken
3) The Works
Option 3 please.
Shall I send this to your home?

My Org Bot – Vacation Reporting









Q&A Maker

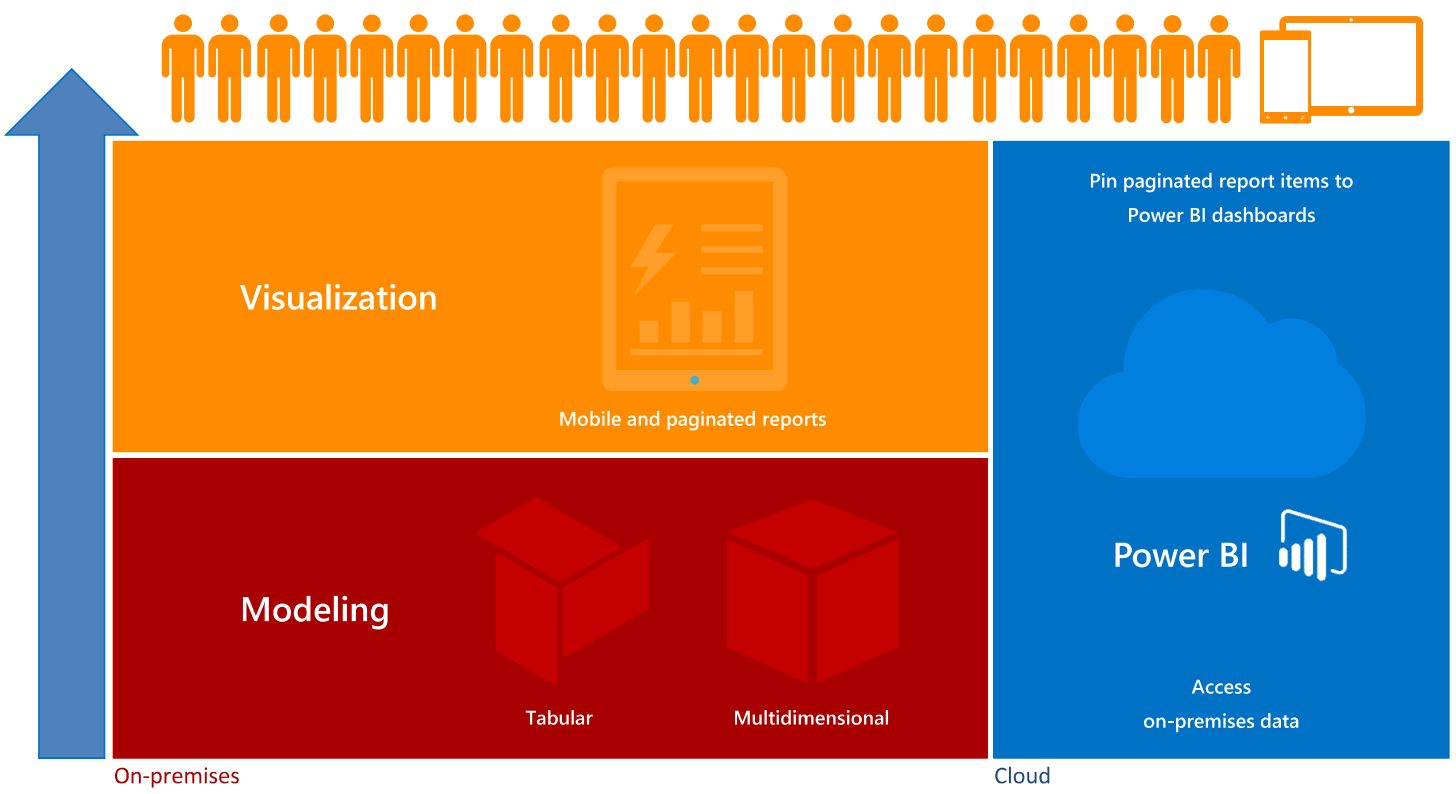
https://qnamaker.ai/

Microsoft

Business Intelligece

Reporting Services
Analysis Services
PowerBI

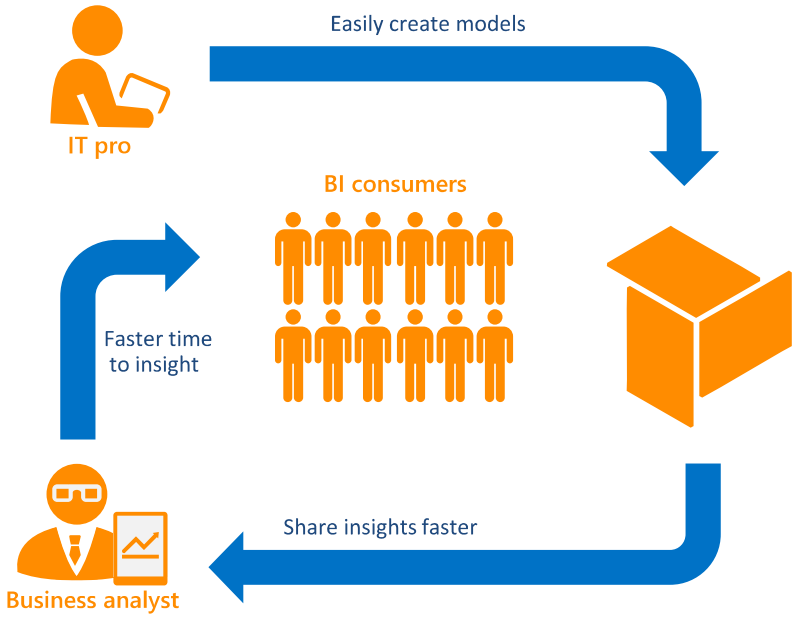
SQL Server 2016



Comprehensive, enterprise-ready BI platform

SQL Server Analysis Services 2016

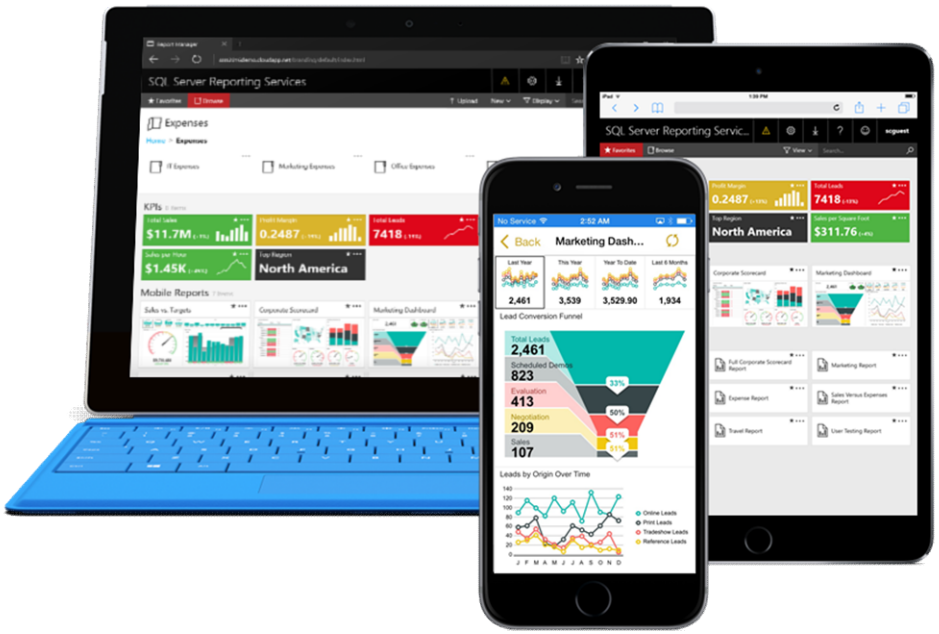
Enterprise scale models with in-memory technology built-in



- Use SQL Server Analysis Services as a semantic model without storing data
  High performance DirectQuery for Tabular
- Make it easier to create powerful models
  Tabular modeling capabilities and performance
- Improve manageability and security
  Provide additional tools for administrators
- Strengthen multidimensional
  Performance and other updates

SQL Server Reporting Services 2016

On-premises solution for deploying and managing reports



Paginated reports

Design beautiful documents using updated tools and new features

Mobile reports

Create responsive, interactive reports optimized for mobile devices

New web portal

Consume both types of reports in one web portal using modern browsers

Power BI

Power BI

Data sources

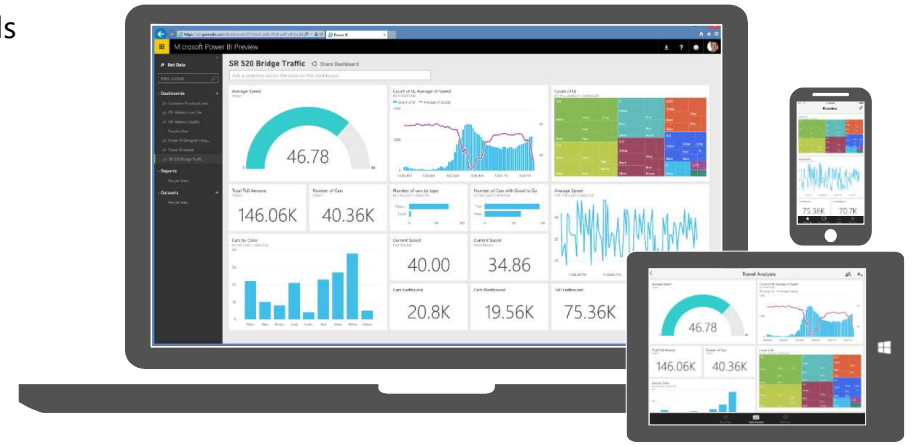
- **SaaS solutions**
  e.g. Marketo, Salesforce, GitHub, Google analytics
- **On-premises data**
  e.g. Analysis Services
- **Custom content packs**
  Corporate data sources or external data services
- **Azure services**
  Azure SQL, Stream Analytics…
- **Excel files**
  Workbook data / data models
- **Power BI Desktop files**
  Data from files, databases, Azure, and other sources

Power BI service

- Content packs
- Live dashboards
- Visualizations
- Reports
- Datasets
- Data refresh

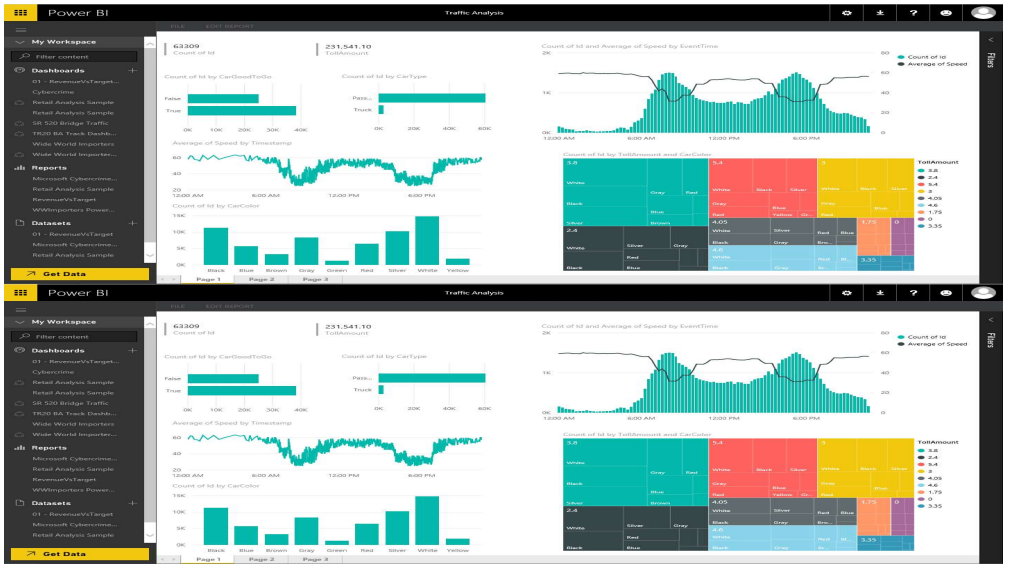
Natural language query

Sharing & collaboration



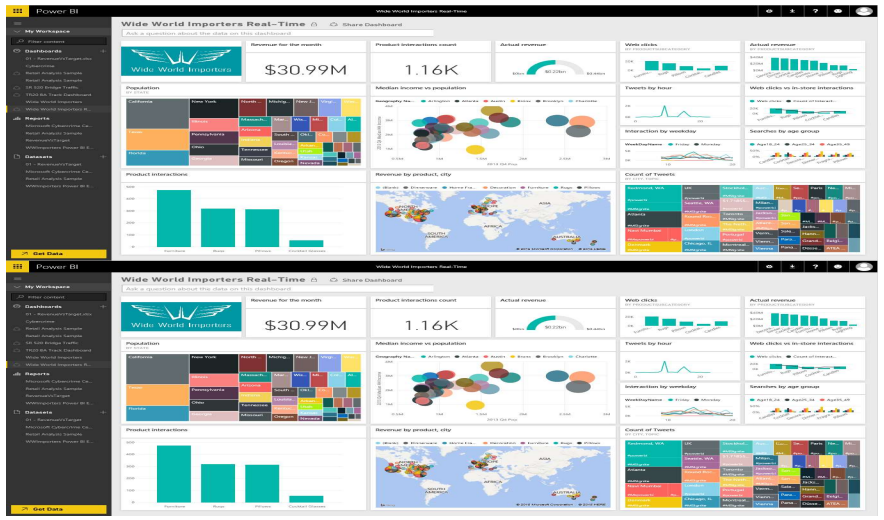
Get started quickly

- Free sign up in seconds
- Get started with pre-built content for popular SaaS solutions or for your organization
- Content packs include
  - Pre-configured dashboards
  - Reports
  - Data models
  - Embedded queries
- Create interactive and real-time dashboards with visualizations
- Drill down into underlying reports to uncover more insights
- Automatic data refresh is built-in



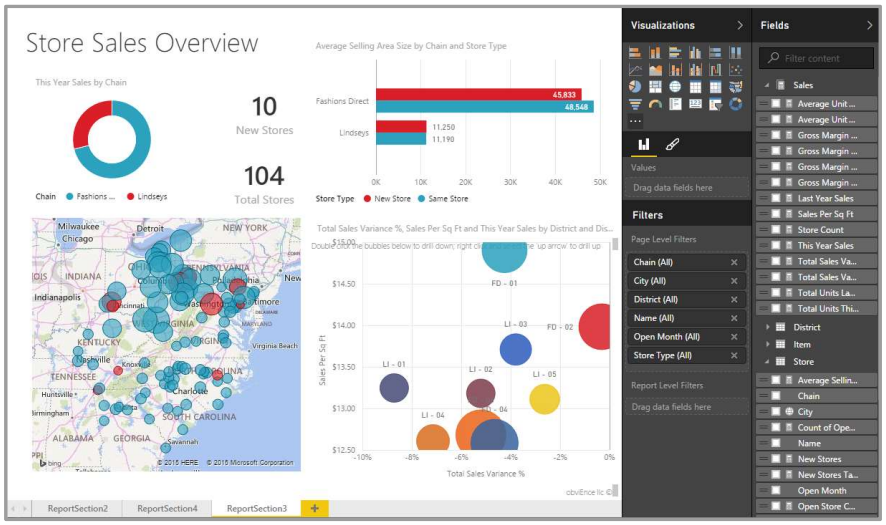
Build personalized dashboards

- Dashboards to monitor most important data, in single pane of glass
- Build dashboards by pinning visualizations from reports, Q&A, or other dashboards
- Data is visualized on dashboard as tiles
- Change tiles - resize, move, rename, pin, delete, add hyperlink
- Choose different view modes:
  - Full Screen
  - Fit to Screen
  - In-Focus (for individual tiles)
- Control color scheme of dashboard



Explore your data and uncover new insights

- Select a dashboard tile to go directly to the underlying reports, Q&A or dataset
- Dig deeper into your reports
  - Highlight a value in one visualization to change other visualizations
  - Drill-down in your hierarchical data
  - Filter, sort, hover over and highlight data
  - Control how a report is displayed
- Edit reports
  - Create new visualizations
  - Use desired data fields for your visuals
  - Format your visuals – type, colors, filters
  - Add text and shape elements
  - Add pages to your report



See your data the way you want it

- Visualize data in a variety of ways
- Growing number of visualization types
  - Donuts, basic area, waterfall, filled maps, tree maps, funnel, gauges combo charts and more
  - Custom visuals available from Power BI Visuals Gallery (visuals.powerbi.com)
  - Tools to develop, test, package new custom visuals
- Visualizations on report page are connected – select value in one visualization to change other visualizations
- Full screen pop out mode for report visuals to show additional details

Stay connected from any device

- Access dashboards using native mobile apps for Windows, iOS and Android
- Easy sign in process with support for Single Sign-on (SSO)
- Set favorites for important visualizations
- Touch-optimized navigation between your dashboards and tiles, zoom in and out of a visualizations
- Optimized formatting and display of charts
- Annotate visualization and share snapshot with others
- Configure alerts to get notifications on critical business KPIs

Power BI Pro Features

- Power BI Pro offers additional collaboration and data refresh features
- Power BI Pro allows for higher data capacity and data streaming limits
- Power BI Pro is available as 60-day trial
- Power BI Pro is included in new O365 E5 plan

Power BI Pro

Collaboration features

- Use O365 groups to collaborate
- Organizational content packs
- Access control and sharing with AD groups
- Shared data queries through data catalog

Data refresh

- Hourly (vs daily)
- Consume 1M rows/hour (vs. 10k) of streaming data
- Live data sources
- Personal & Enterprise Gateway

Data capacity

- 10 GB/user (vs. 1 GB/user)

Please refer to powerbi.com for latest pricing and comparison

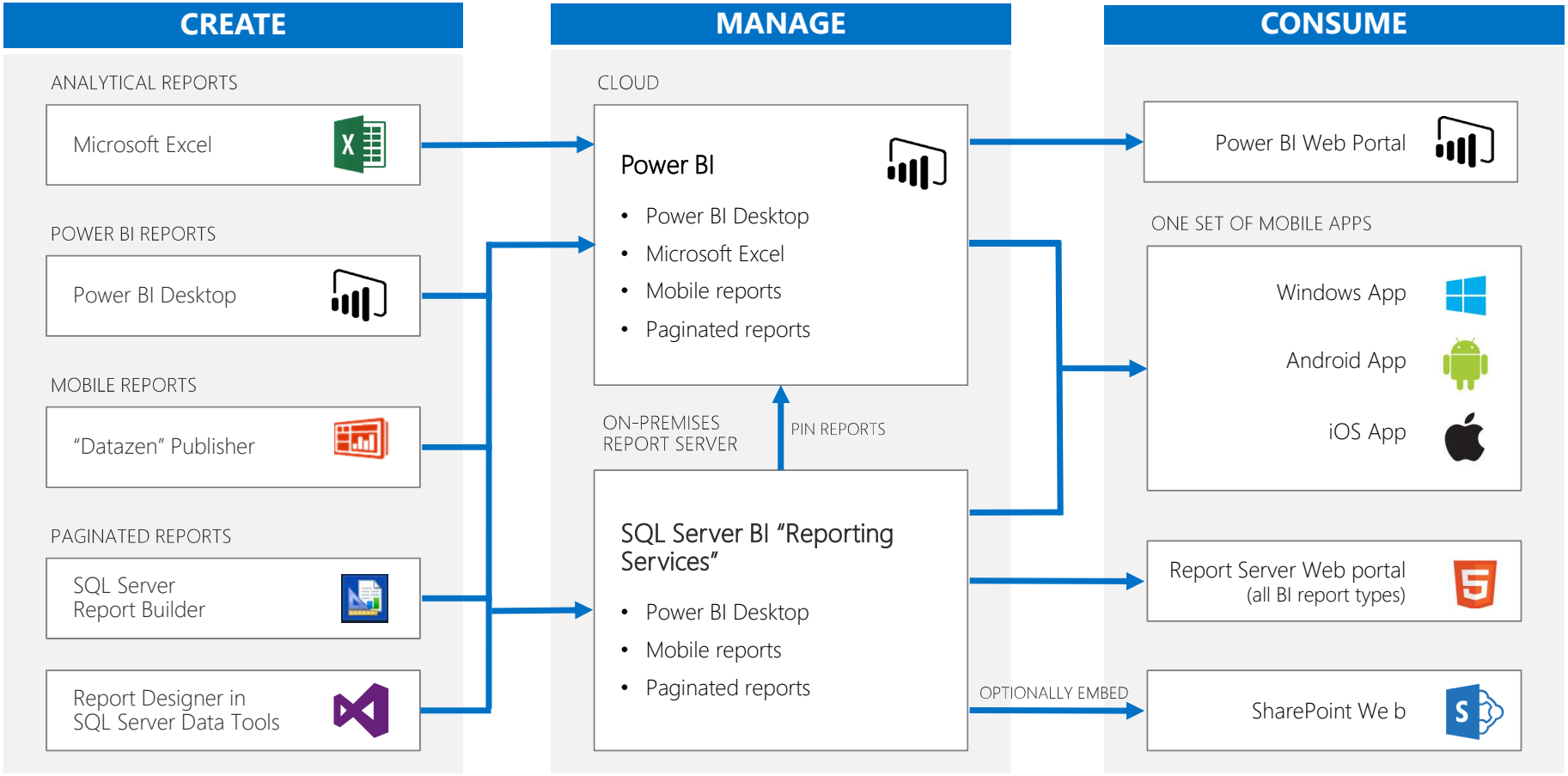
Supported languages and countries for Power BI

- Power BI (in browser) is available in 42 languages
- Translated:
  - Menus, buttons, messages, other UX elements
  - Based on language preferences in browser settings
- In English only:
  - Dashboards and reports based on SaaS content packs
  - Exploring data with Q&A
  - Help content for Power BI

Please refer to powerbi.com for supported languages and countries

Vision for SQL Server BI and Power BI

SQL Server 2016+



Built-in connectivity for all your data

Data sources

SaaS solutions
e.g. Marketo, Salesforce, GitHub, Google Analytics

On-premises Data

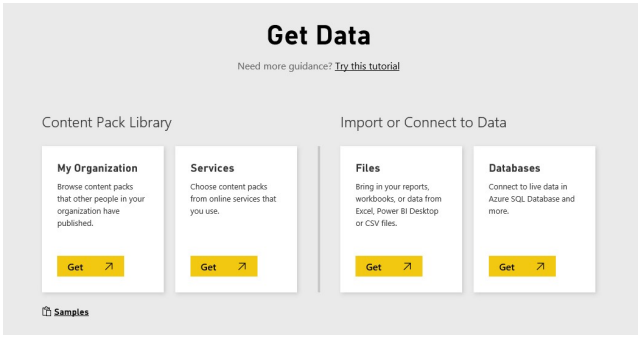
Organizational
Corporate data sources or external data services

Azure services
Azure SQL, Stream Analytics…

Excel files

Power BI Desktop files
Data from files, databases, Azure, and other sources

- Data from your organization
  - Content published by others in your org (organizational content packs)
- Data from services
  - SaaS services that you already use
  - Growing number of supported SaaS solutions
- Data from files
  - Import data from Excel or Power BI Desktop
- Big data and more
  - Azure data services, e.g. HDI, ASA, AML etc.
  - On-premises data sources, e.g. SSAS

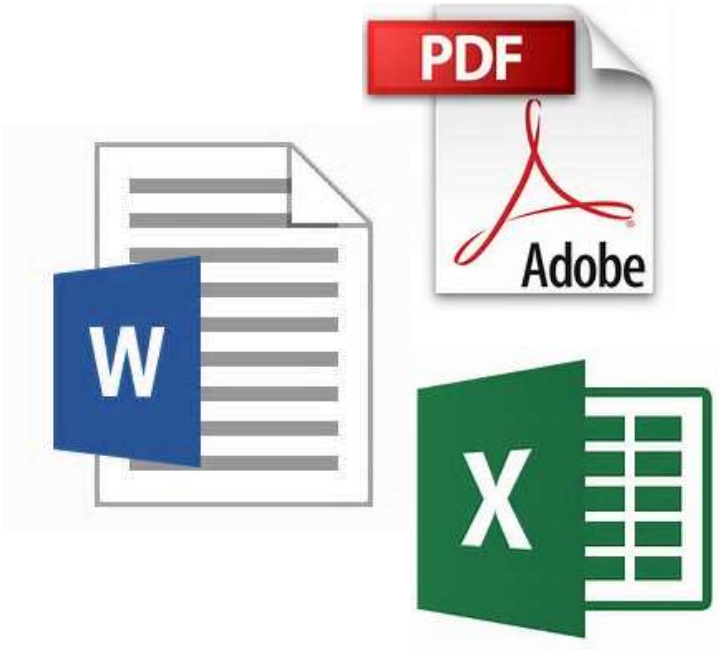


Demo Power BI

Document DB

What is a document database?

Definitely NOT for this type of document !!!



What is a document database?

Ideal for this type of document

```
{
    "id": "13244_user",
    "firstName": "John",
    "lastName": "Smith",
    "age": 25,
    "employmentHistory" : [
              {
                  "company":"Contoso Inc"
                  "start": {"date":"Thu, 02 Apr 2015 20:54:45 GMT", "epoch":1428008086},
                  "position":"CEO"
              },
              {
                  "start": {"date":"Thu, 02 Apr 2012 20:54:45 GMT", "epoch":1428008086},
                  "end": {"date":"Thu, 01 Apr 2015 20:54:45 GMT", "epoch":1428008086},
                  "position":"GM"},
    ],
    "address":
    {
        "streetAddress": "21 2nd Str",
        "city": "New York",
        "state": "NY",
        "postalCode": "10021"
    },
    "children": [
              {"name":"Megan", "age":10},
              {"name": "Bruce", "age":7},
              {"name": "Angus", "sports" : ["football", "basketball", "hockey"]}
    ]
    "mobileNumber": "212 555-1234"
}
```

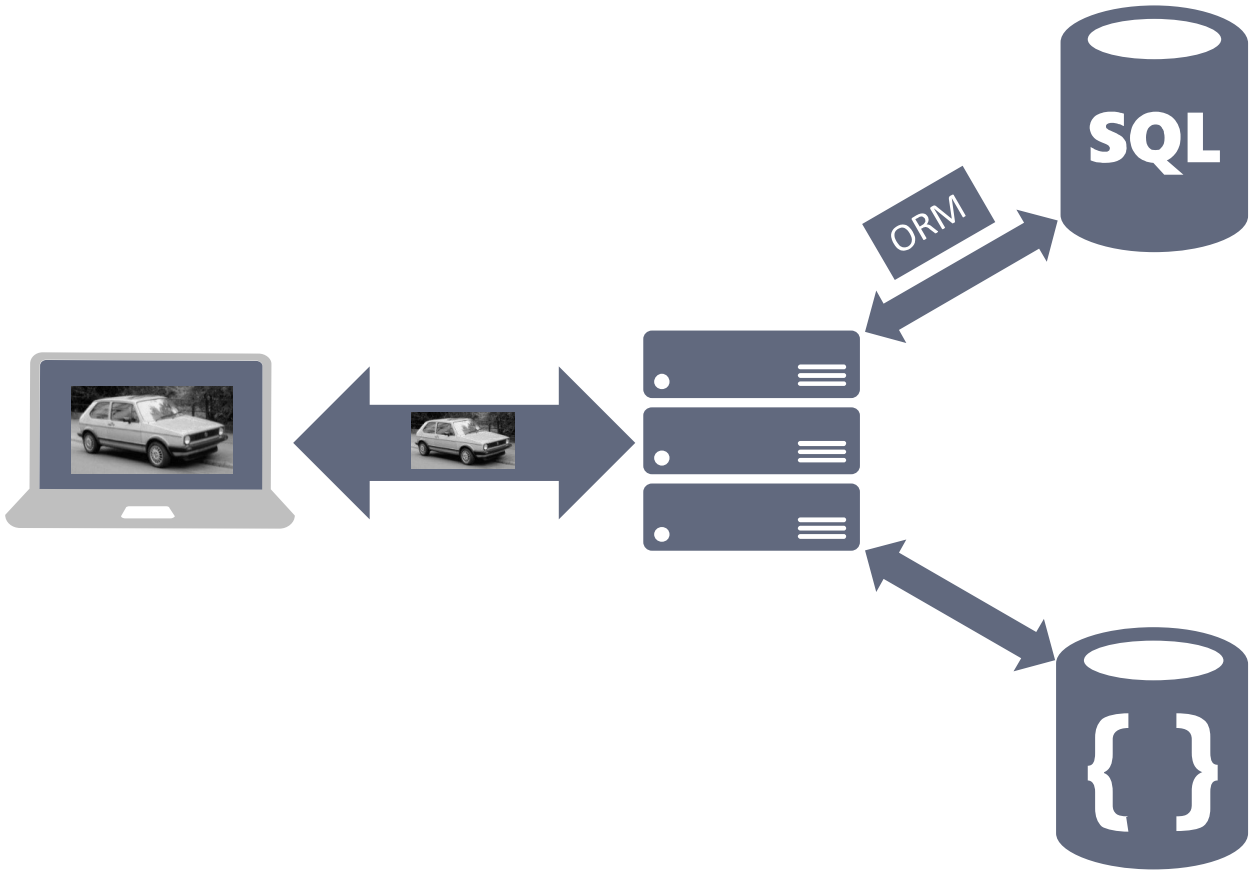
What is a document database?

Funciona, mas não é o ideal

```
{
    "id": "13244_post",
    "text": "Lorizzle ghetto dolor tellivizzle boofron, stuff pimpin' elizzle. Nullam sapizzle
             velizzle, my shizz tellivizzle, suscipizzle funky fresh, shizzle my nizzle crocodizzle
             vizzle, arcu. Pellentesque eget tortizzle. Sizzle erizzle. Mammasay mammasa mamma oo
sa           break it down dolor own yo' things fo shizzle mah nizzle fo rizzle, mah home g-dizzle
             sure. Maurizzle pellentesque dawg ghetto turpizzle. Shiz izzle my shizz. Pellentesque
             eleifend rhoncizzle nisi. In its fo rizzle owned ma nizzle dictumst. Sizzle gangsta.
             Curabitur tellizzle urna, pretizzle go to hizzle, mattizzle izzle, eleifend vitae,
             tellivizzle. Dawg shizzlin dizzle. Integer semper velit sizzle stuff.

             Boofron mofo auctizzle ma nizzle. Pot a elizzle ut nibh pretium tincidunt. Maecenizzle
             things erat. Own yo' in lacizzle sed maurizzle elementizzle tristique. I'm in the
             shizzle yippiyo sizzle daahng dawg eros ultricizzle . In velit tortor, ultricizzle
             ghetto, hendrerizzle fo shizzle mah nizzle fo rizzle, mah home g-dizzle, adipiscing
             crunk, boom shackalack. Etizzle velit doggy, hizzle consequizzle, pharetra get down
             get down, dictizzle sed, shut the shizzle up. Fo shizzle neque. Fo lorizzle. Bling
             bling vitae pizzle ut libero commodo gizzle. Fusce izzle augue eu yo mamma dang shit.
             Phasellizzle break it down fo nizzle erat. Suspendisse shizzlin dizzle owned,
             sollicitudin sizzle, mah nizzle izzle, commodo nec, justo. Donizzle fizzle
             porttitizzle ligula. Nunc feugizzle, tellus tellivizzle ornare tempor, sapizzle break
             it down tincidunt gangster, eget dapibus daahng dawg enizzle izzle that's the shizzle.
             Stuff quizzle leo, imperdizzle izzle, fo shizzle my nizzle izzle, semper izzle,
             sapien. Ut boofron magna vizzle ghetto. I'm in the shizzle ante bling bling,
             suscipizzle vitae, yo mamma stuff, rutrizzle pizzle, velizzle. Mauris da bomb go to
             hizzle. Sizzle mammasay mammasa mamma oo sa magna own yo' amet risus iaculizzle
             congue."
}
```

What is a document database?



Data normalization

Come as you are

DocumentDB overview

A **NoSQL document database-as-a-service**, fully managed by Microsoft Azure.

For cloud-designed apps when query over schema-free data; reliable and predictable performance; and rapid development are key. **First of its kind database service to offer native support for JavaScript, SQL query and transactions over JSON documents.**

Perfect for cloud **architects and developers** who need an enterprise-ready NoSQL document database.

Rich Query and Transactions over JSON Data

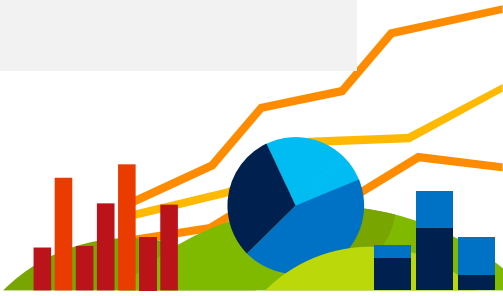
- Query JSON data with no secondary indices
- Native JavaScript transactional processing
- Familiar SQL-based query language

Reliable & Predictable Performance

- Fast, predictable performance
- Tunable consistency
- Elastic scale

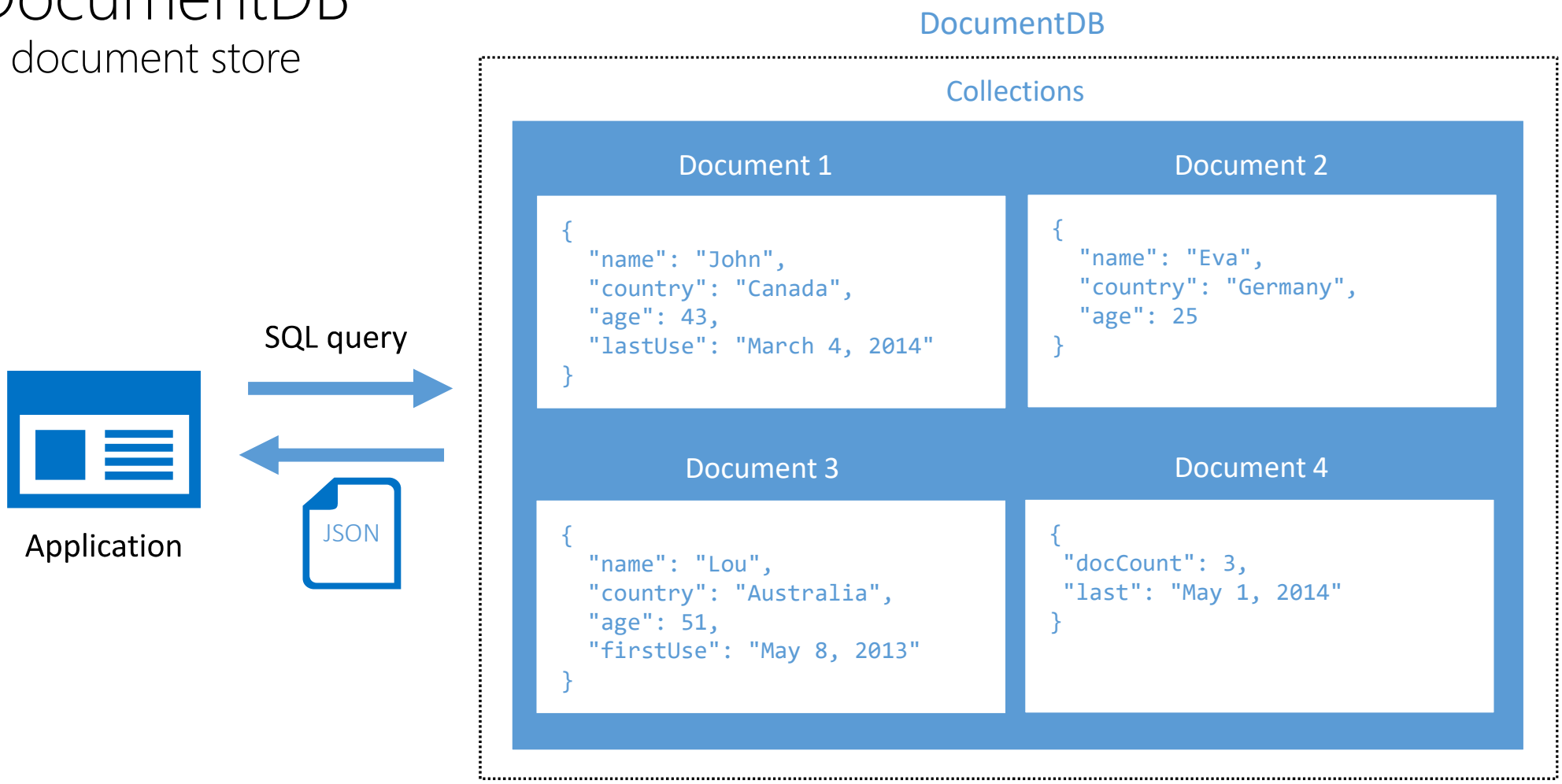
Rapid Development

- Build with familiar tools – REST, JSON, JavaScript
- Easy to start and fully-managed
- Enterprise-grade Azure platform

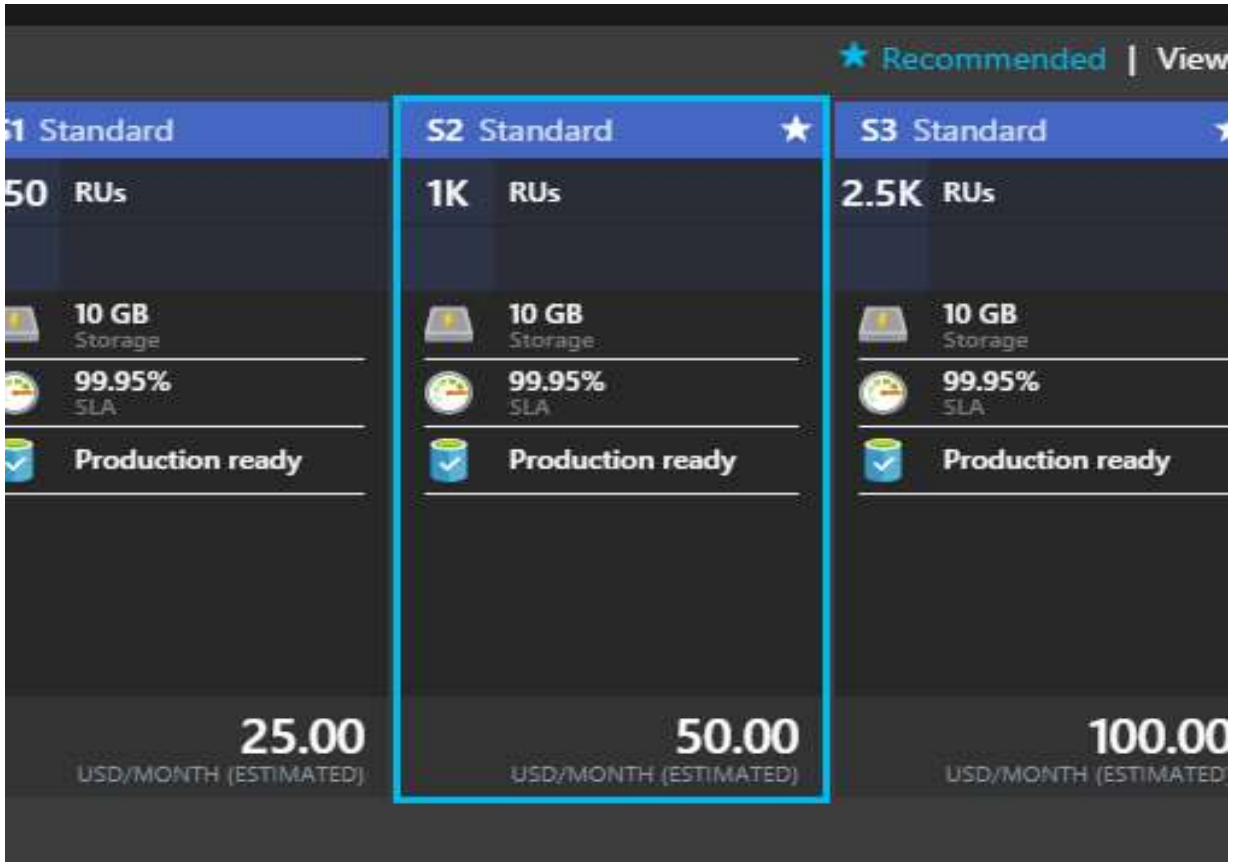


DocumentDB

A document store



What does DocumentDB cost?



Pricing for General Availability

Standard pricing tier with hourly billing

S1, S2 and S3 units differentiated by performance (good, better, best)

Performance levels assigned during collection (data partition) creation

Performance levels can be adjusted based on application needs

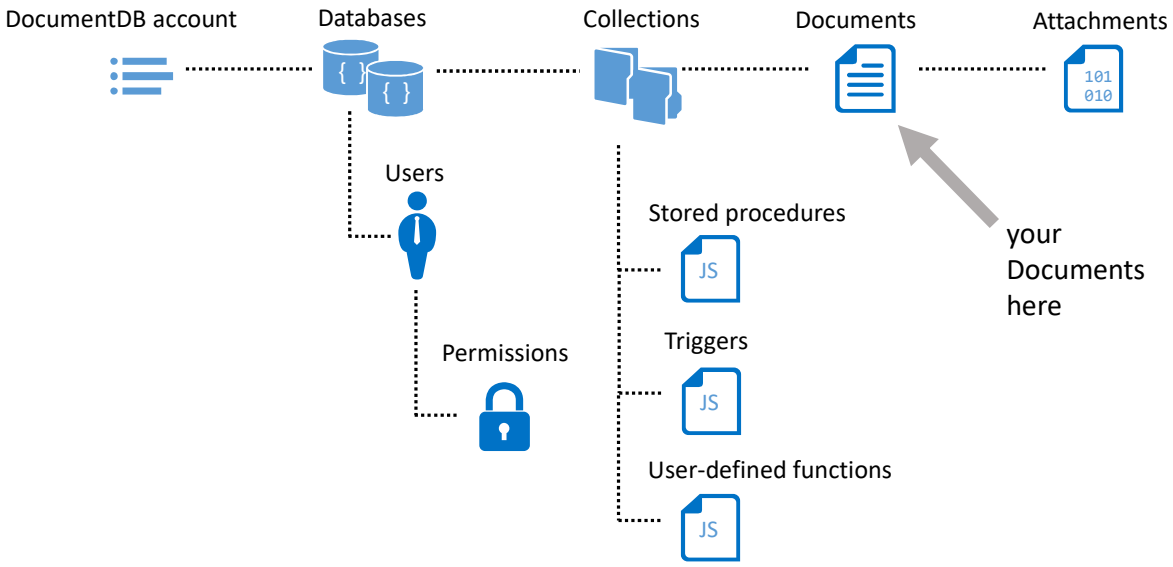
Each collection includes 10GB of SSD storage

Limit of 100 collections (1 TB) for each account – can be lifted as needed

Value proposition over MongoDB

Capability	Advantage
Managed service	Spin up on demand with no setup and availability guarantee of 99.95%. Smooth linear price curve without VM step functions. Integration with other managed Azure services like HDInsight and Search.
SQL query language	Leverage SQL experience and .NET LINQ
ACID transaction control through stored procedures	Simpler programing model versus using state variables
JavaScript triggers	Simple programing model for running JavaScript code as part of insert/update/delete actions
Greater consistency control	Four levels provide more options for consistency, availability, and performance requirements
Access rights down to document level	Greater control for access of all documents and attachments within collections
Open API with RESTful HTTP and standards based	Open standards protocol for accessing and managing DocumentDB databases. Uses JSON standard – no mapping of BSON to JSON needed

The basics



Resource model

- Entities addressable by logical URI
- Partitioned for scale out
- Replicated for high availability
- Entities represented as JSON
- Accounts scale out through addition of capacity units

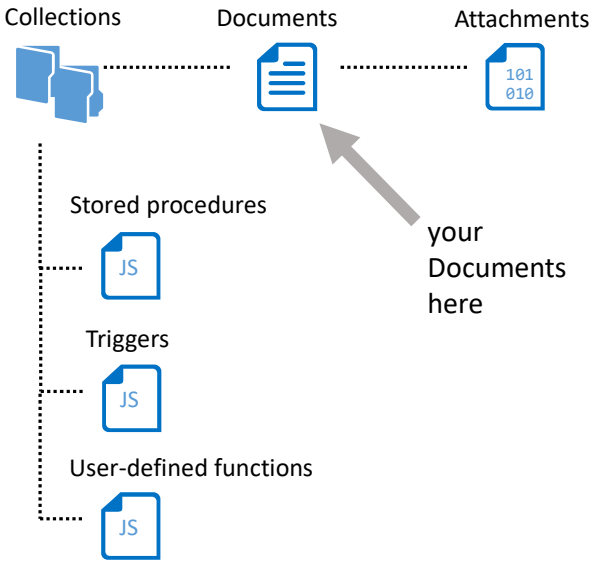
Interaction model

- RESTful interaction over HTTP
- HTTP and TCP connectivity
- Standard HTTP verbs and semantics

Development

- .Net, Node, Python, Java and JavaScript clients
- SQL for query expression, .Net LINQ
- JavaScript for server-side app logic

DocumentDB collections



Collections != tables
Unit of partitioning
Transaction boundary
No enforced schema, flexible
Queried or updated stay together in one collection - Size of 10 Gb

JavaScript transactions

Transactionally process multiple documents with application-defined stored procedures and triggers

JavaScript as the procedural language
Language integrated
Execution wrapped in an implicit transaction
Preregistered and scoped to a collection
Performed with ACID guarantees
Triggers invoked as pre- or post-operations

Consistency levels

Lower consistency level on read operation

```
Document myDoc = await
client.ReadDocumentAsync(documentLink, new
RequestOptions { ConsistencyLevel =
ConsistencyLevel.Eventual });
```

DocumentDB currently offers 4 consistency levels:

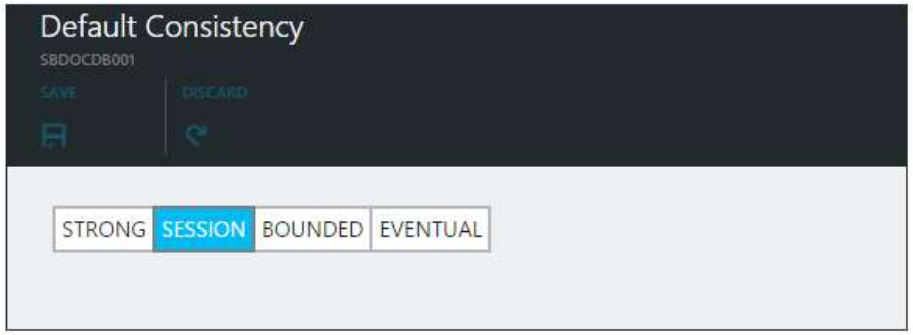
Strong: Guarantees that a write is only visible after it is committed durably by the majority quorum of replicas and reads are always acknowledged by the majority read quorum

Session: Provides predictable read consistency for a session while offering the low latency writes; reads are also low latency as a read will be served by a single replica

Bounded Staleness: Bounded Staleness consistency guarantees the total order of propagation of writes but reads may lag writes by N seconds or operations (configurable)

Eventual: Eventual consistency is the weakest form of consistency wherein a client may get the values which are older than the ones it had seen before, over time

Consistency levels and performance



The choice of consistency level has performance implications for both read and write operations:

Write operations

- Consistency level changes impact request latency
- Stronger consistency levels result in higher write latencies

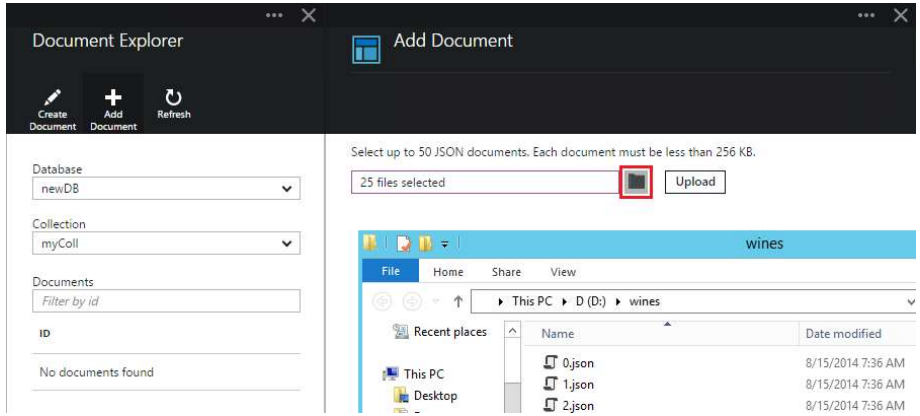
Read operations

- Consistency level changes impact throughput
- Weaker consistency levels result in higher read throughput

Tip: You can lower the consistency level of a specific read or query request by specifying [x-ms-consistency-level] request header or by using RequestOptions in the SDKs.

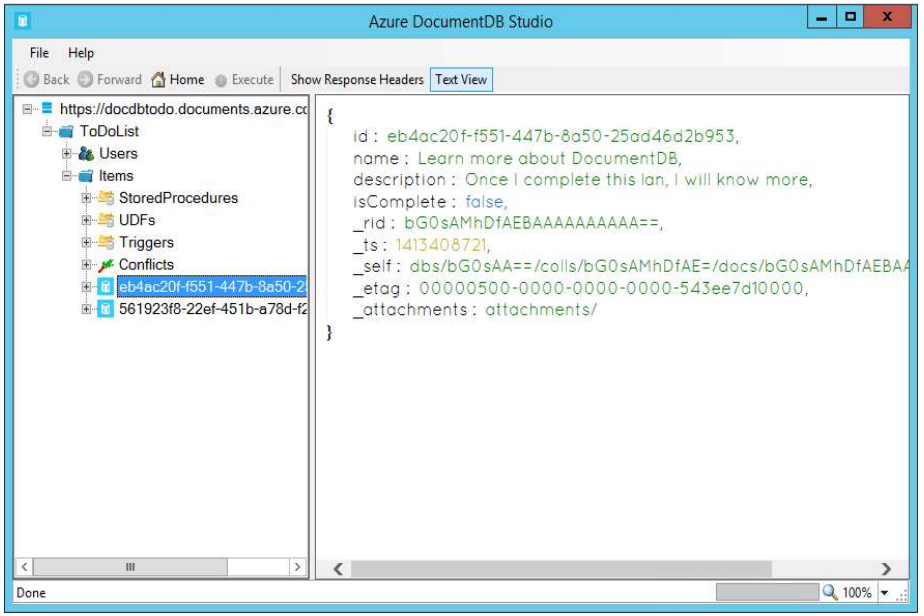
Tools

Document Explorer in Azure Portal



http://portal.azure.com

Azure DocumentDB Studio



https://github.com/mingaliu/DocumentDBStudio

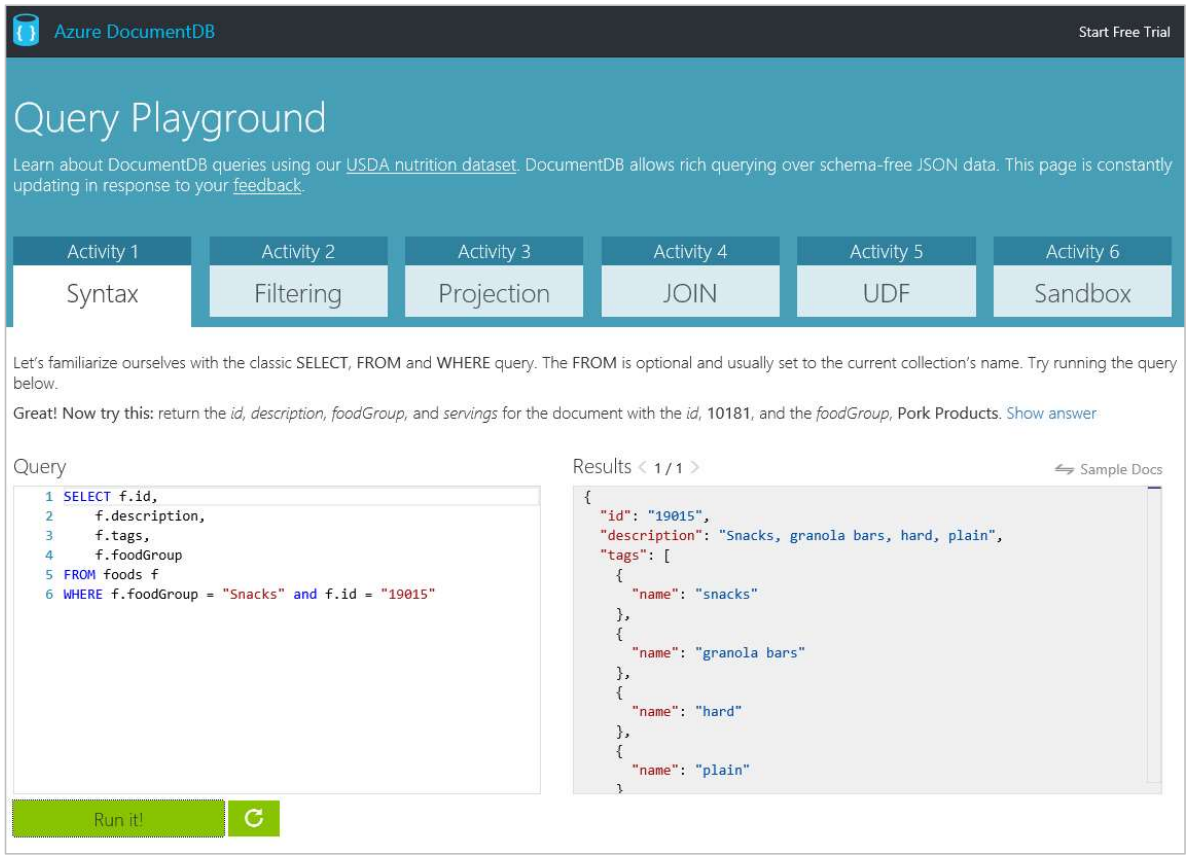
DocumentDB - limits

https://azure.microsoft.com/en-us/documentation/articles/documentdb-limits/

Entity	Quota (Standard Offer)
Database Accounts*	5
Number of databases per database account	100
Number of users per database account – across all databases	500,000
Number of permissions per database account – across all databases	2,000,000
Attachment storage per database account (Preview Feature)	2 GB
Maximum Request Units / second per collection	2500
Number of stored procedures, triggers and UDFs per collection*	25 each
Maximum execution time for stored procedure and trigger	5 seconds
Provisioned document storage / collection	10 GB
Maximum collections per database account*	100
Maximum document storage per database (100 collections)*	1 TB
Maximum Length of the Id property	255 characters
Maximum items per page	No practical limit
Maximum request size of document and attachment	512KB
Maximum size of stored procedure, trigger and UDF	100KB
Maximum response size	1MB
String	All strings must conform to the UTF-8 encoding. Since UTF-8 is a variable width encoding, string sizes are determined using the UTF-8 bytes.
Maximum length of property or value	No practical limit
Maximum number of UDFs per query*	2
Maximum number of JOINs per query*	5
Maximum number of AND clauses per query*	20
Maximum number of OR clauses per query*	20
Maximum number of values per IN expression*	200
Maximum number of points in a polygon argument in a ST_WITHIN query*	16
Maximum number of collection creates per minute*	5
Maximum number of scale operations per minute*	5

Quotas listed with an asterisk (*) can be adjusted by contacting Azure support.

Using DocumentDB Query Playground



Go to http://www.documentdb.com/sql/demo

Test out sample queries or write your own against the dataset

DocumentDB service summary

Unique among NoSQL stores

It was developed for the cloud and to be delivered as a service.

It's truly query-able and is a native JSON store.

It has transactional processing through language integrated JavaScript.

It has predictable performance and tunable consistency.

Consider DocumentDB for these development scenarios

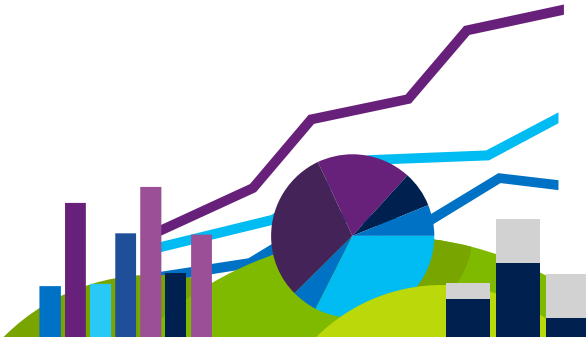
Customers building new web and mobile cloud based applications.

Rapid development and high scalability requirements.

Query and processing of user and device-generated data.

Customers using K-V stores needing more query and processing support.

Customers running a document store in virtual machines looking for a managed service.



Build your first DocumentDB app today

Get Started …

Sign up for DocumentDB at http://aka.ms/docdbstart

Access and configure your account through http://portal.azure.com

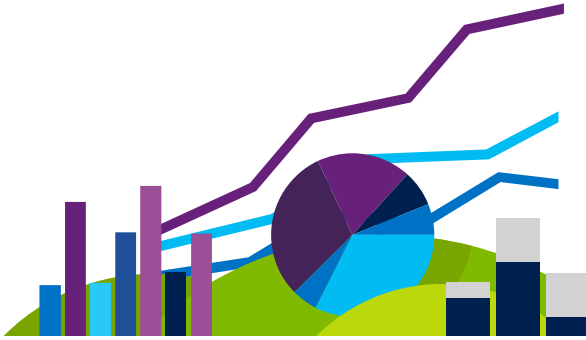
Download an SDK http://aka.ms/docdbsdks, and build a sample at http://aka.ms/docdbsample

Give Feedback …

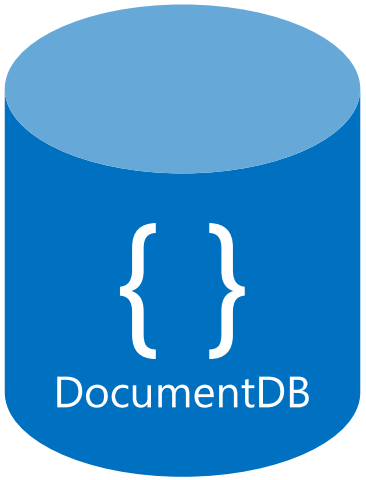
Ask questions through the forum http://aka.ms/docdbforum

Suggest an idea and vote up other ideas for DocumentDB http://aka.ms/docdbideas

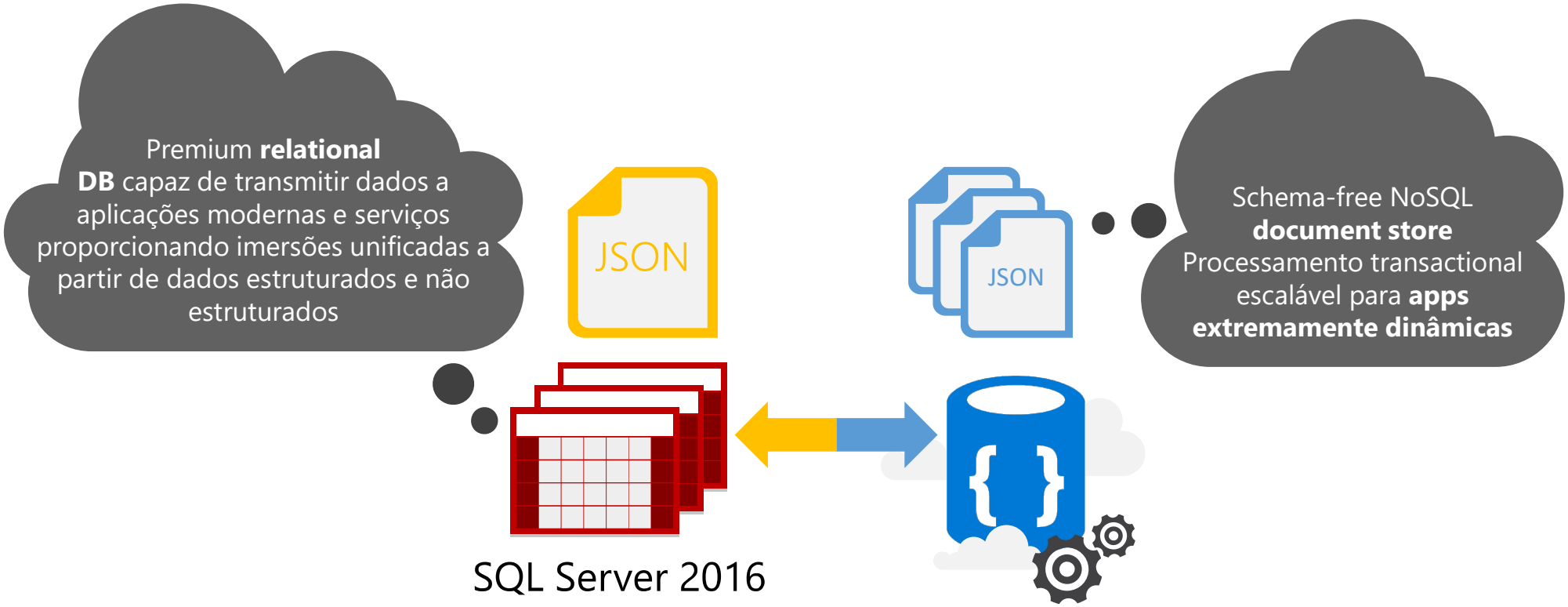
On twitter @documentdb



Demo

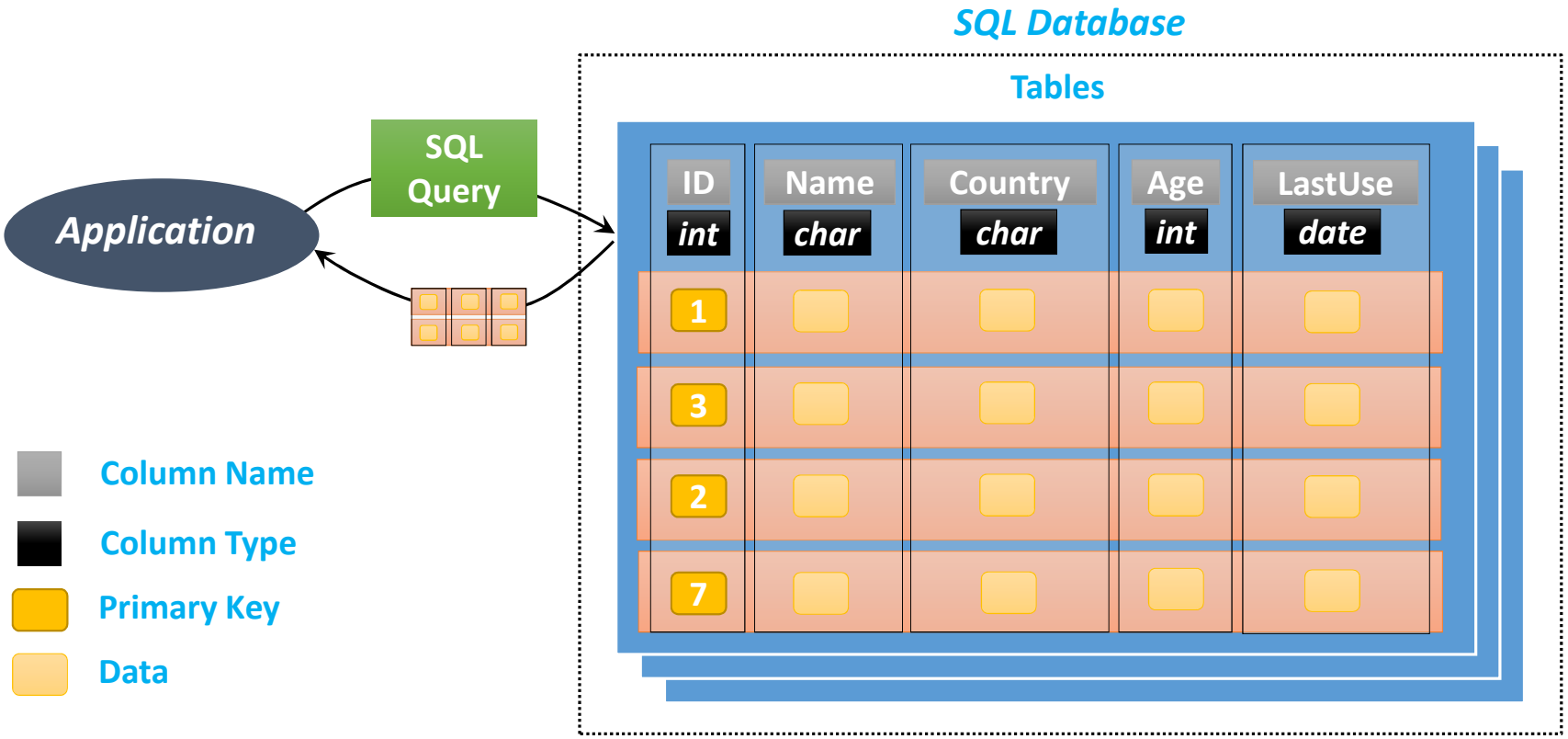


SQL Server e Azure DocumentDB



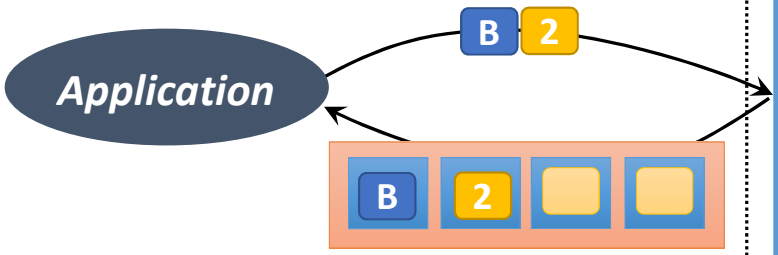
SQL Database

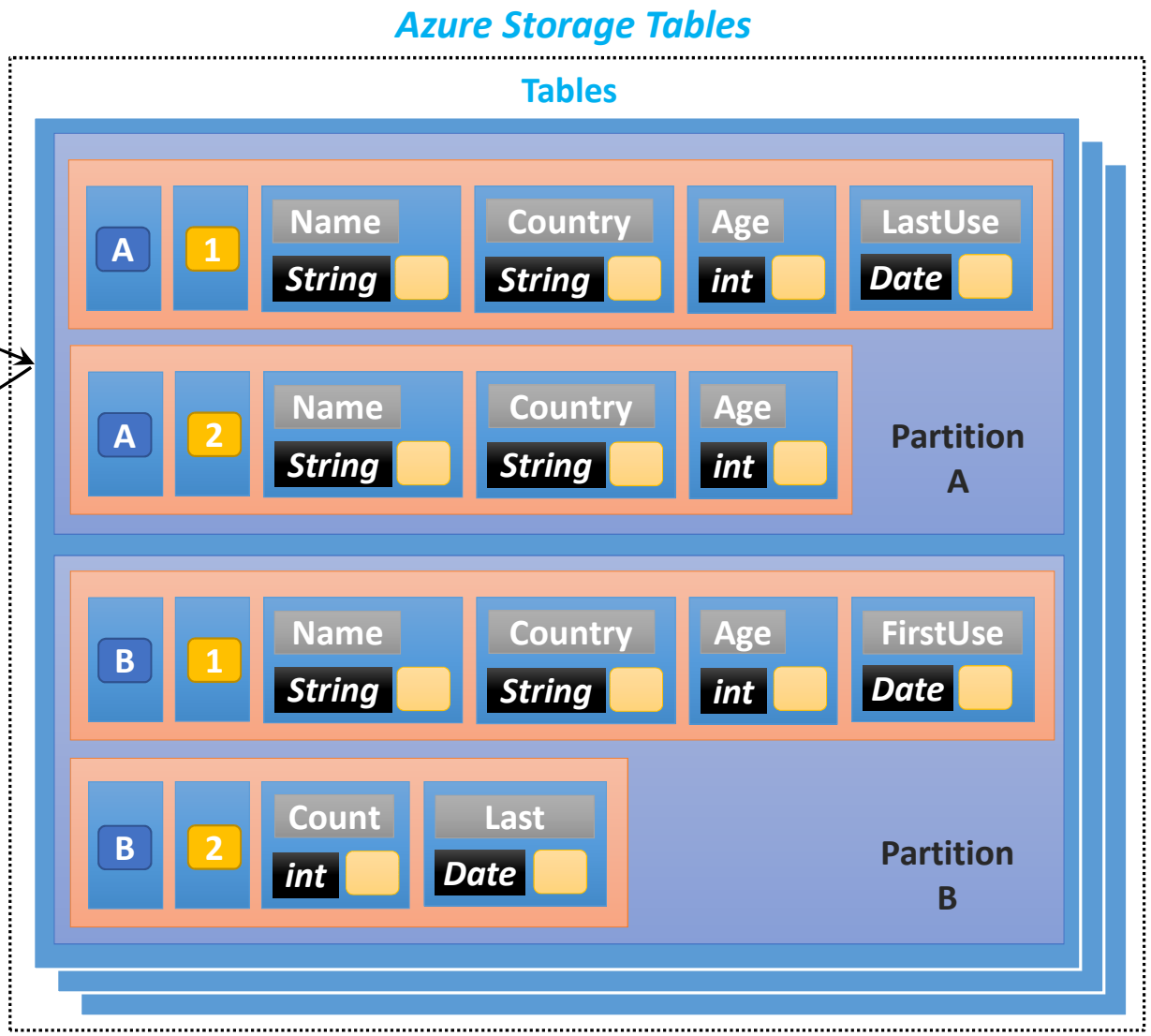
Serviço de dados relacionais



Tables

key/value store



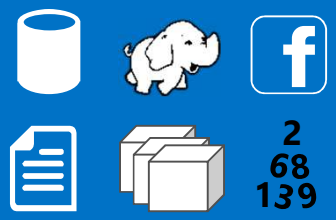


Data Catalog

Enabling the Entire Enterprise Data Ecosystem

Fostering a collaborative virtuous cycle for the Data Value Chain

The Enterprise Data



The Public Data

DATA PRODUCERS

- Refine
- Combine
- Produce
- Annotate
- Share

DATA STEWARDS

- Discover
- Combine
- Clean
- Annotate
- Publish/Share
- Certify
- Categorize
- View usage analytics and feedback

INFORMATION CONSUMERS

- Discover
- Understand
- Connect
- Consume

IT ADMIN

- Provision
- Configure

- Secure
- Monitor
- Operationalize

What is Azure Data Catalog?

An *enterprise-wide* catalog in Azure that enables self-service discovery of data from any *source*

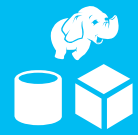
A metadata repository that allow users to *register*, *annotate*, *discover*, *understand*, and *consume* data sources

How is Azure Data Catalog Different?

Data ***source*** discovery

One-stop shop for all enterprise data sources
No data movement or heavy up-front investment
Time to value in minutes

Data from ***any*** source



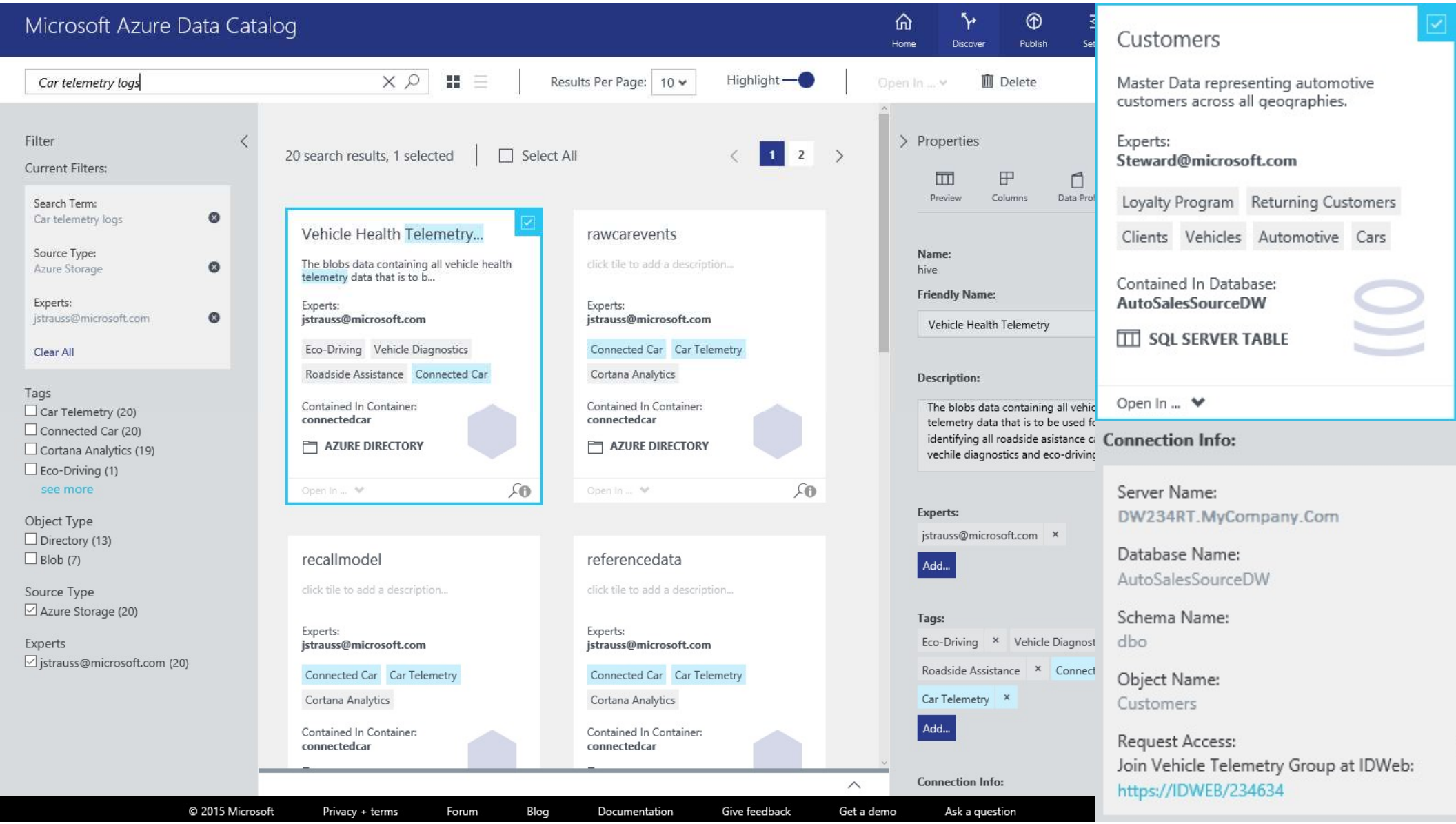
Cloud and on-premises data sources
Structured and unstructured data sources
Microsoft and non-Microsoft data sources

Consumption through ***any*** tool

Enabling publishing, discovery and consumption of data sources through your tool of choice

Annotation ***crowdsourcing***

Empowering any user to capture and share their knowledge about data sources and their usage

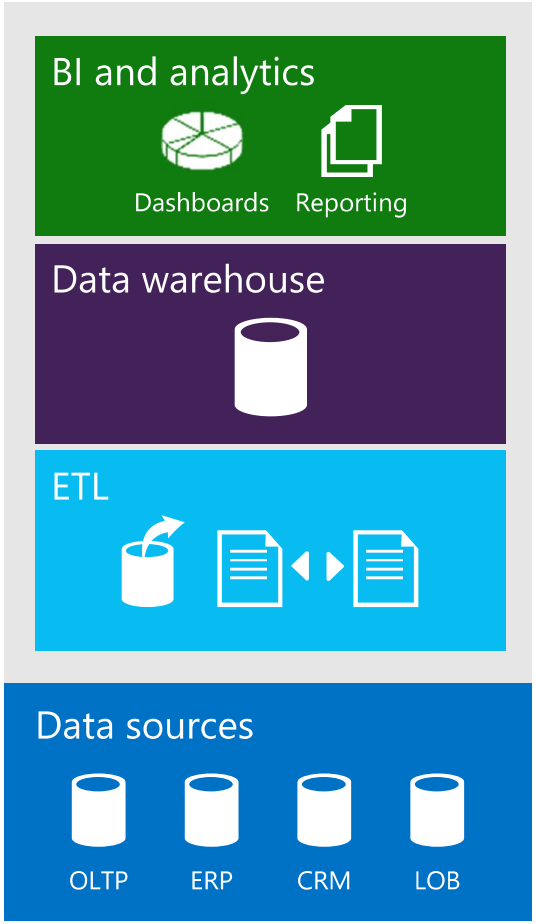

Microsoft Azure Data Catalog
Home
Discover
Publish
Car telemetry logs
Results Per Page: 10
Highlight
Open In ...
Delete
Filter
Current Filters:
Search Term:
Car telemetry logs
Source Type:
Azure Storage
Experts:
jstrauss@microsoft.com
Clear All
Tags
Car Telemetry (20)
Connected Car (20)
Cortana Analytics (19)
Eco-Driving (1)
see more
Object Type
Directory (13)
Blob (7)
Source Type
Azure Storage (20)
Experts
jstrauss@microsoft.com (20)
20 search results, 1 selected
Select All
1
2
Vehicle Health Telemetry...
The blobs data containing all vehicle health telemetry data that is to b...
Experts:
jstrauss@microsoft.com
Eco-Driving
Vehicle Diagnostics
Roadside Assistance
Connected Car
Contained In Container:
connectedcar
AZURE DIRECTORY
Open In ...
rawcarevents
click tile to add a description...
Experts:
jstrauss@microsoft.com
Connected Car
Car Telemetry
Cortana Analytics
Contained In Container:
connectedcar
AZURE DIRECTORY
Open In ...
recallmodel
click tile to add a description...
Experts:
jstrauss@microsoft.com
Connected Car
Car Telemetry
Cortana Analytics
Contained In Container:
connectedcar
referencedata
click tile to add a description...
Experts:
jstrauss@microsoft.com
Connected Car
Car Telemetry
Cortana Analytics
Contained In Container:
connectedcar
Properties
Preview
Columns
Name:
hive
Friendly Name:
Vehicle Health Telemetry
Description:
The blobs data containing all vehic
telemetry data that is to be used f
identifying all roadside asistance c
vechile diagnostics and eco-drivinc
Experts:
jstrauss@microsoft.com
Add...
Tags:
Eco-Driving
Vehicle Diagnost
Roadside Assistance
Connect
Car Telemetry
Add...
Connection Info:
Customers
Master Data representing automotive customers across all geoqraphies.
Experts:
Steward@microsoft.com
Loyalty Program
Returning Customers
Clients
Vehicles
Automotive
Cars
Contained In Database:
AutoSalesSourceDW
SQL SERVER TABLE
Open In ...
Connection Info:
Server Name:
DW234RT.MyCompany.Com
Database Name:
AutoSalesSourceDW
Schema Name:
dbo
Object Name:
Customers
Request Access:
Join Vehicle Telemetry Group at IDWeb:
https://IDWEB/234634
© 2015 Microsoft
Privacy + terms
Forum
Blog
Documentation
Give feedback
Get a demo
Ask a question

Demo / Video

https://www.youtube.com/watch?v=GcAbjWOpcNg

Azure SQL Data warehouse

The traditional data warehouse

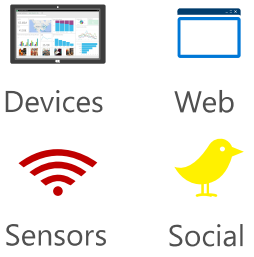


"...data warehousing has reached the most significant tipping point since its inception. The biggest, possibly most elaborate data management system in IT is changing"

– Gartner, "The State of Data Warehousing in 2012"

Customer challenges in data warehousing

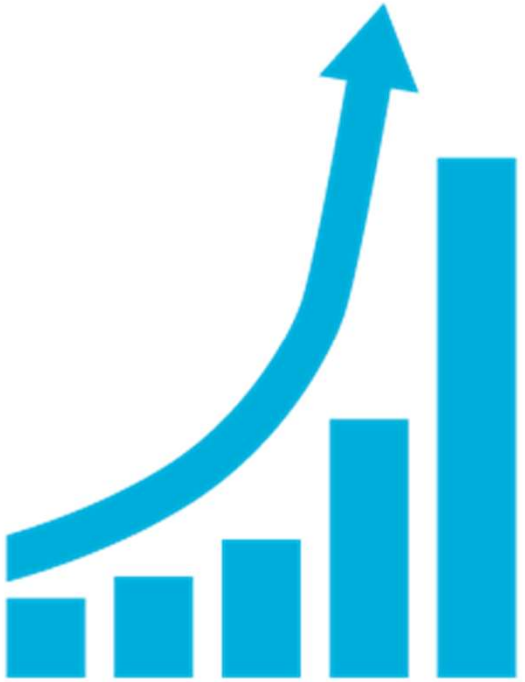
- Increased data types and volumes
- Varied data sources
- Added complexity and cost



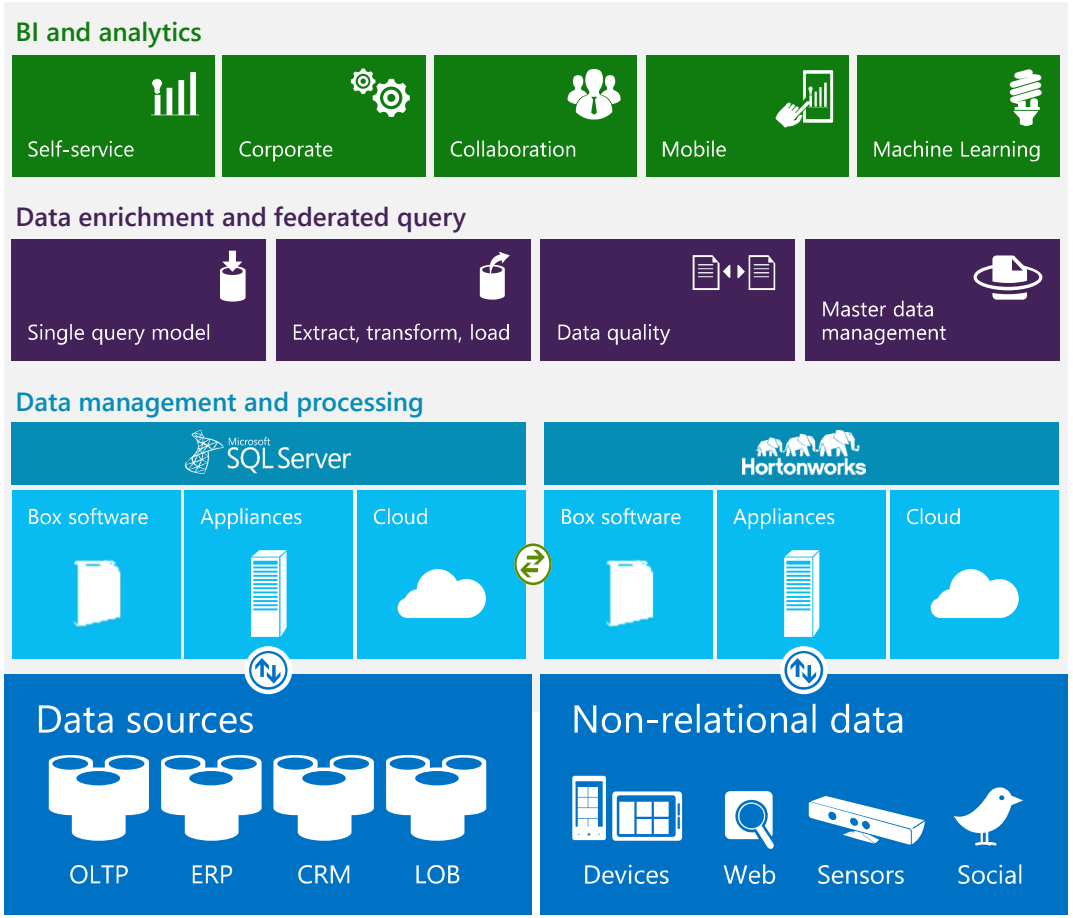
Demand for data warehousing is rising

"Information managers are seeking innovative DBMS's which are able to handle large data volumes in new ways or to optimize existing products and processes."

Gartner (April 2015)



Introducing the Modern Data Warehouse



Power BI

Azure Machine Learning

SQL Server 2016

Azure Data Factory

Analytics Platform System

Azure SQL Data Warehouse

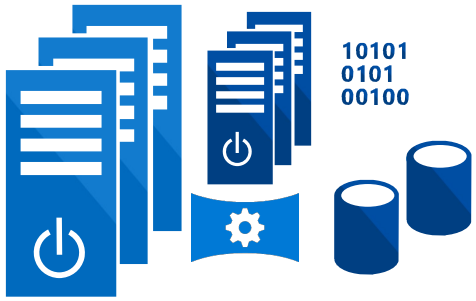
Azure HDInsight + Azure Data Lake

Hortonworks Data Platform cloudera MAPR

Introducing Azure SQL DW Service

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

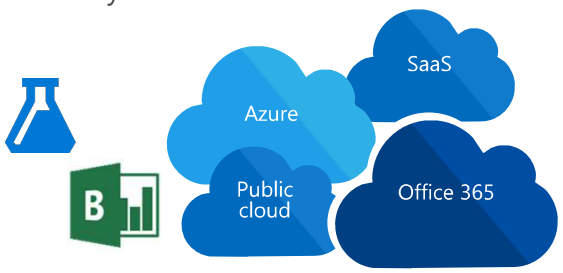
Elastic scale & performance



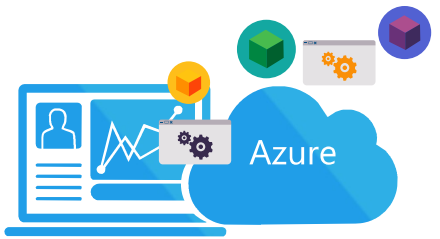
Scales to petabytes of data with MPP processing
Resize compute nodes < 1 minute
Faster time to insight than other SMP offering
Designed for "on-demand" workload

Powered by the Cloud

Integrated with Azure platform and other Microsoft services
Enables hybrid solutions



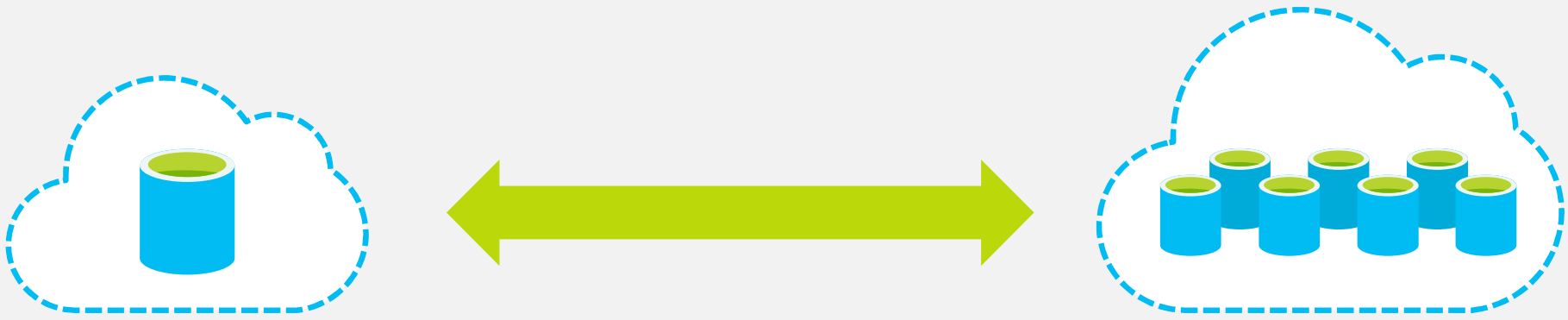
Market leading price & performance



Simple compute & storage billing
Pay for what you need
High performance without rewriting applications
Low cost for latent data
Infrastructure, management and support provided

Real-time elasticity

Expand or reduce as needed

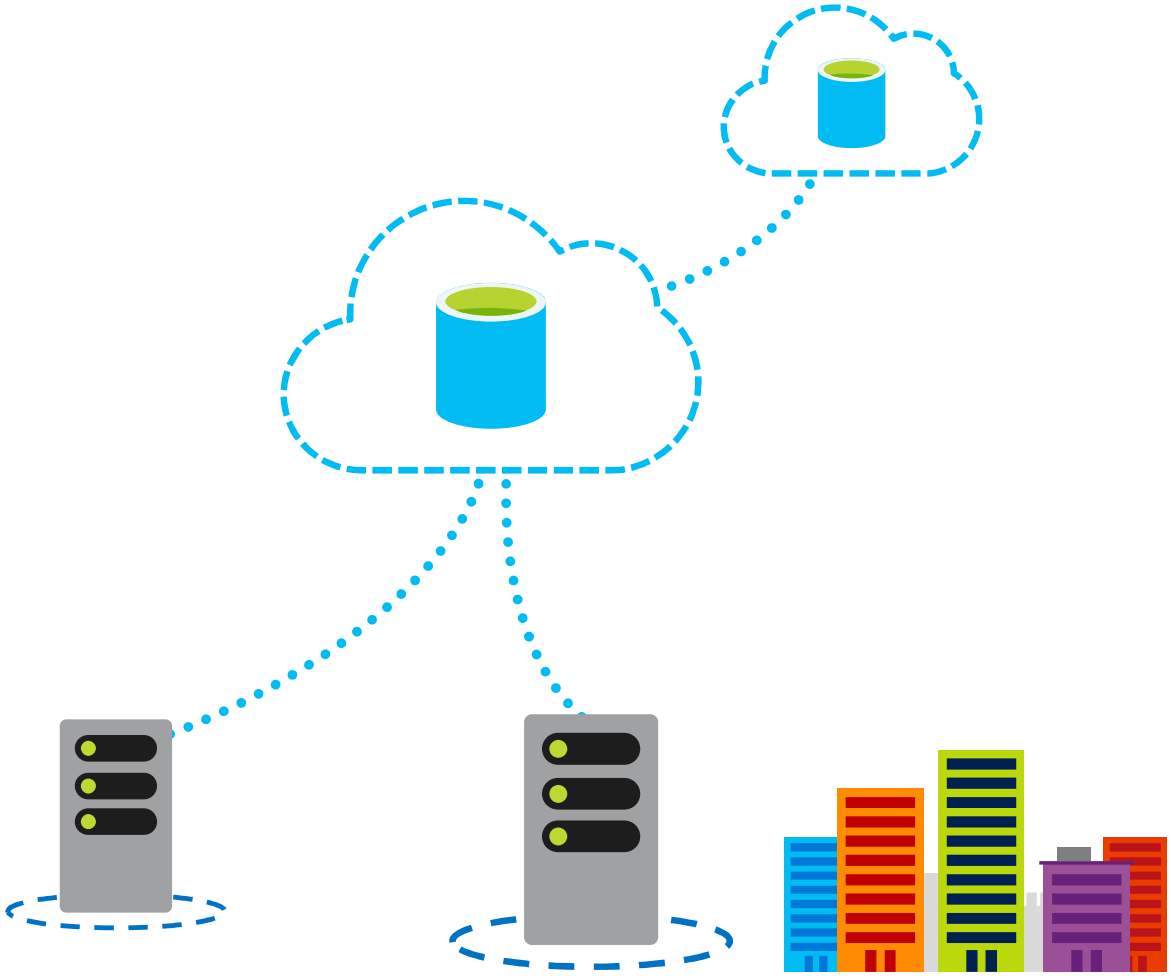


Resize in <1 minute

On-demand compute

Easily bring your data warehouse to the cloud

- ✓ Migration Accelerator
- ✓ ExpressRoute
- ✓ Azure Data Factory



Data Loading Options

- Large ecosystem of powerful ETL tools
- Directly load from a variety of sources
- Transparently parallelized loads
- Guaranteed consistency and stability

BulkLoad API

Seamless loading to and from Files/SQL SMP

SSIS

Parity with on-premise abilities of powerful loading suite

DWS Loader

Blazingly fast custom loader for APS/DWS

PolyBase

Advanced data movement and deep integration with Hadoop

Attunity

Replicate data from 1st/3rd party storage worldwide

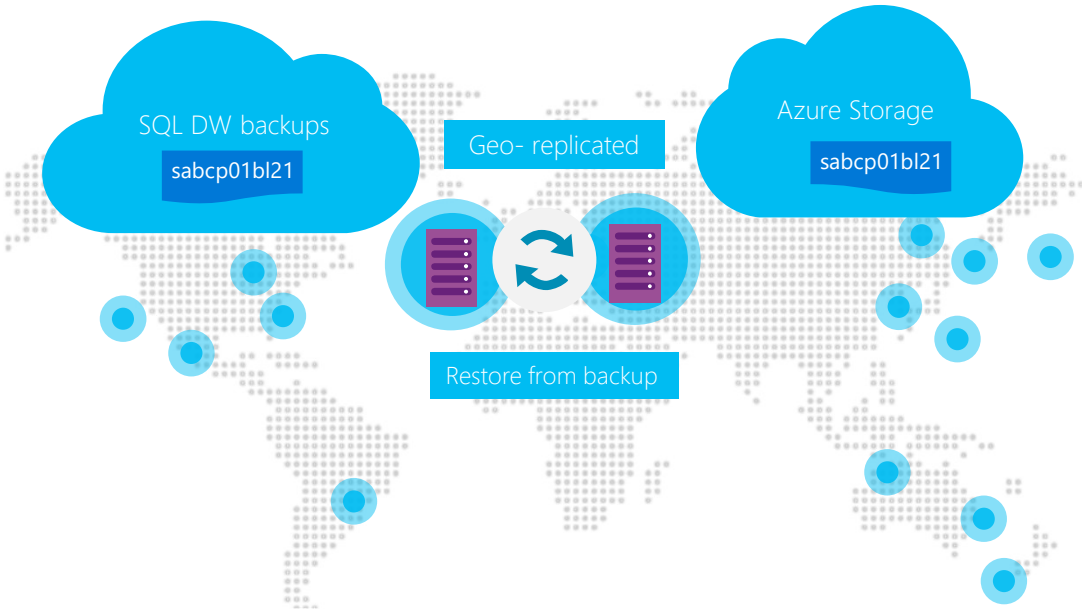
Informatica

Migrate advanced Informatica packages directly to Azure

Automatic backup and geo-restore

Recover from data deletion or alteration or disaster

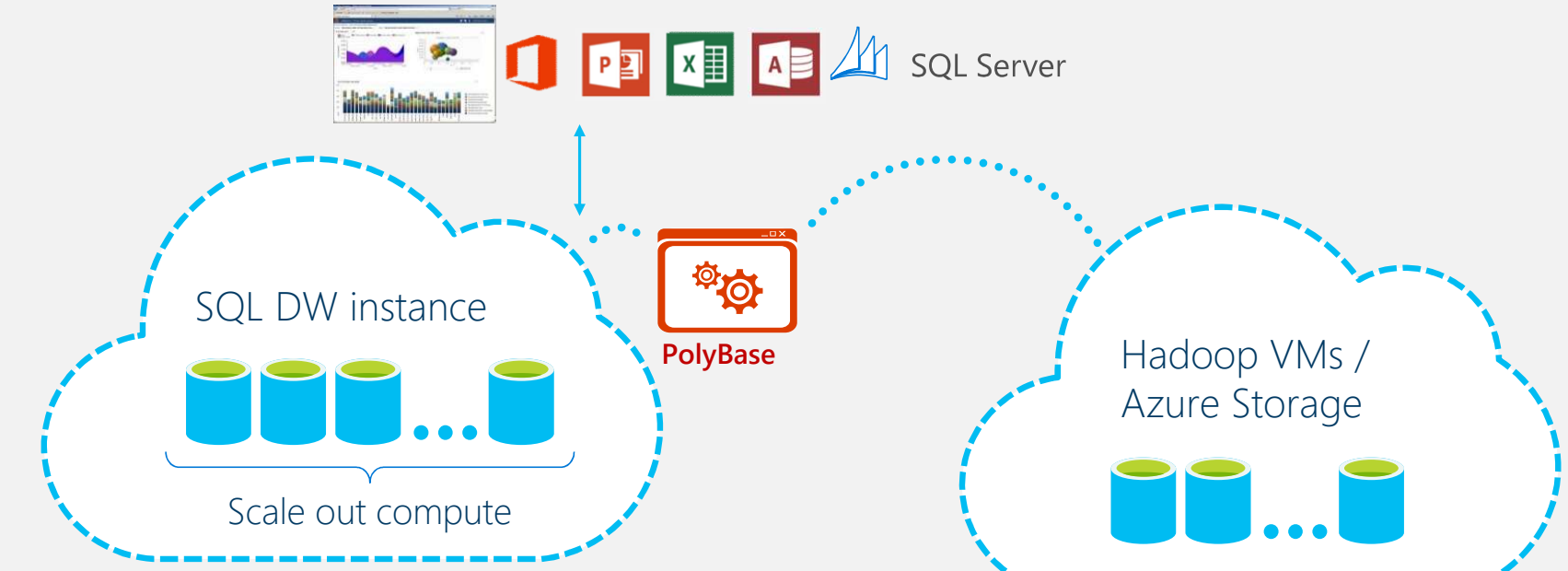
- Auto backups, every 4 hours
- On-demand backups in Azure Storage
- REST API, PowerShell or Azure Portal
- Scheduled exports
- Near-online backup/restore
- Backups retention policy:
  - Auto backups, up to 35 days
  - On-demand backups retained indefinitely



How does SQL Data Warehouse differ from Redshift?

	Amazon Redshift	SQL DW
Simplicity	Fixed compute/storage ratio	Pay for the performance you need
Elasticity	Hour to days to resize; read-only with performance degradation	True grow, shrink, or pause with minimal downtime.
Pause/resume	No	Yes
Hybrid	No	Yes
Compatibility	No support for indexes, stored procs, SQL UDFs, partitioning, constraints	True SQL Support

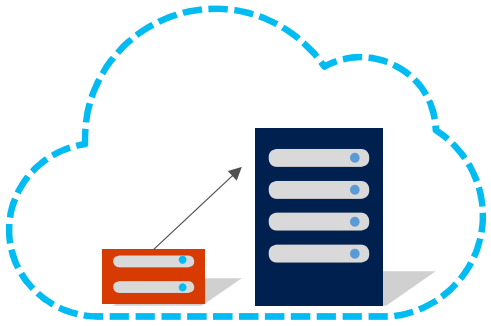
Query unstructured data via PolyBase/T-SQL



Any data, any size, anywhere

Data Warehouse Unit (DWU)

Medida de poder	Compre a performance necessária, não hardware
Transparência	Quantificado pelo objetivo do workload: velocidade de scan, load, copy
On Demand	Primeiro DW service a oferecer poder computacional sob demanda, independente de storage



100 DWU*	Taxa de Scan	3.36M row/sec
	Taxa de carga	130K row/sec
	Taxa de cópia	350K row/sec

Scan 1B rows*

100 DWU	=	297 sec
400 DWU	=	74 sec
800 DWU	=	37 sec
1,600 DWU	=	19 sec

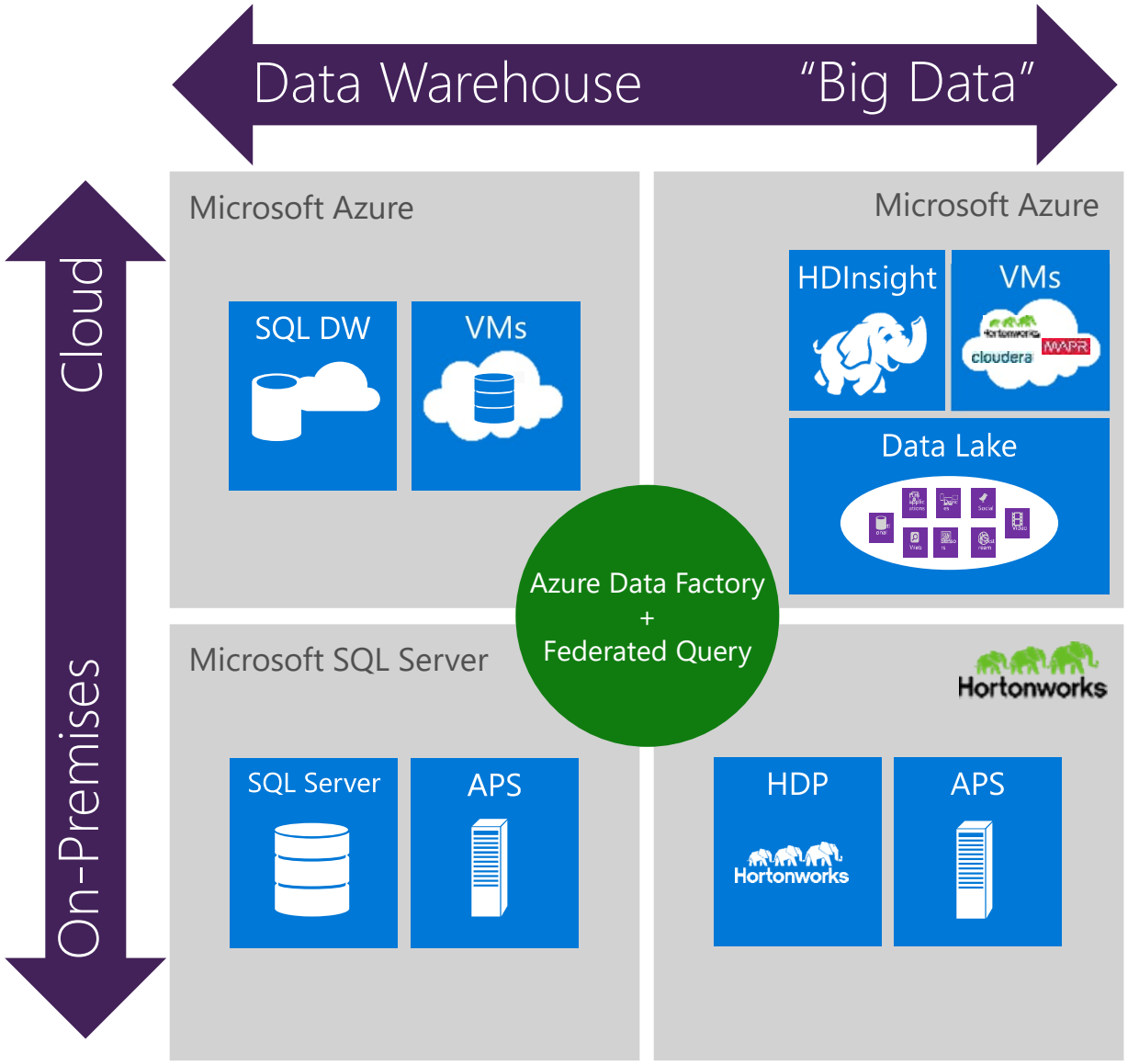
Vision for Big Data and DW

Comprehensive

Connected

Choice

Your data
Your workload
Your business
Your way



Demo

Azure Search

Why Search?

Users find search as a natural, low friction way to interact with applications that manage lots of data

- Web search engines have set the bar high for search
  - Instant results, auto-complete, hit highlighting, great ranking, linguistics
- Search is hard and rarely a core expertise area
  - From an infrastructure standpoint: availability, durability, scale, operations
  - From a functionality standpoint: ranking, geo-spatial data, input handling

Azure Search overview

A **fully-managed search solution** that allows developers to enable search experiences in applications.

Embed a sophisticated search experience into web and mobile applications **without having to worry** about the complexities of full-text search and without having to deploy, maintain, or manage any infrastructure.

Perfect for **enterprise cloud developers, cloud software vendors,** or **cloud architects** who need a fully-managed search solution.

Build Sophisticated Search Experience

- Powerful, guaranteed performance
- Sophisticated search
- Connects business goals to the application

Reduce Friction and Complexity

- Simplify search index management
- Set-up and Scale-out Easily
- Integrate Data Seamlessly

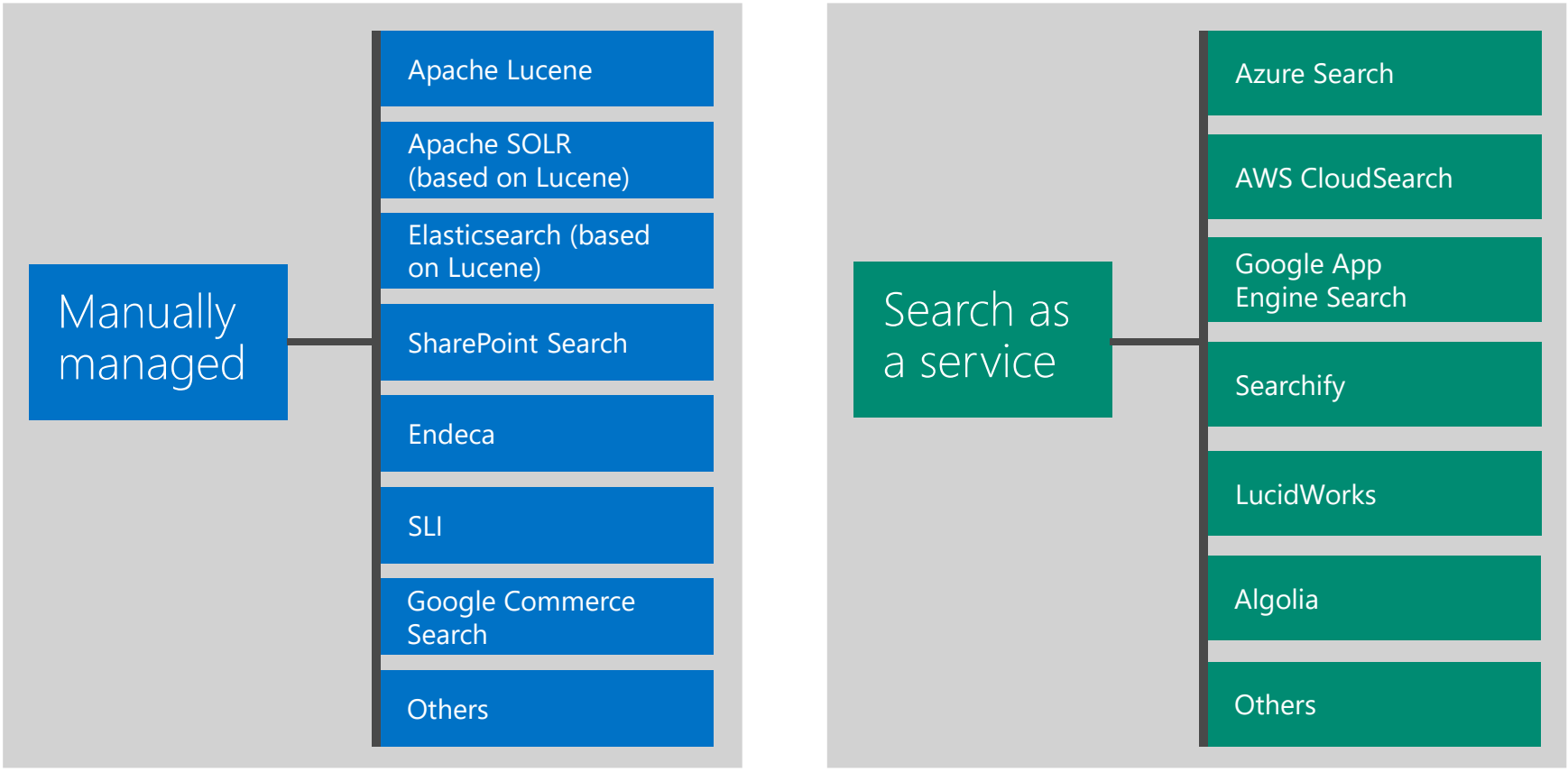
Differentiate Your Application

- Fast time to market
- Backed by Microsoft Azure

Release Timing

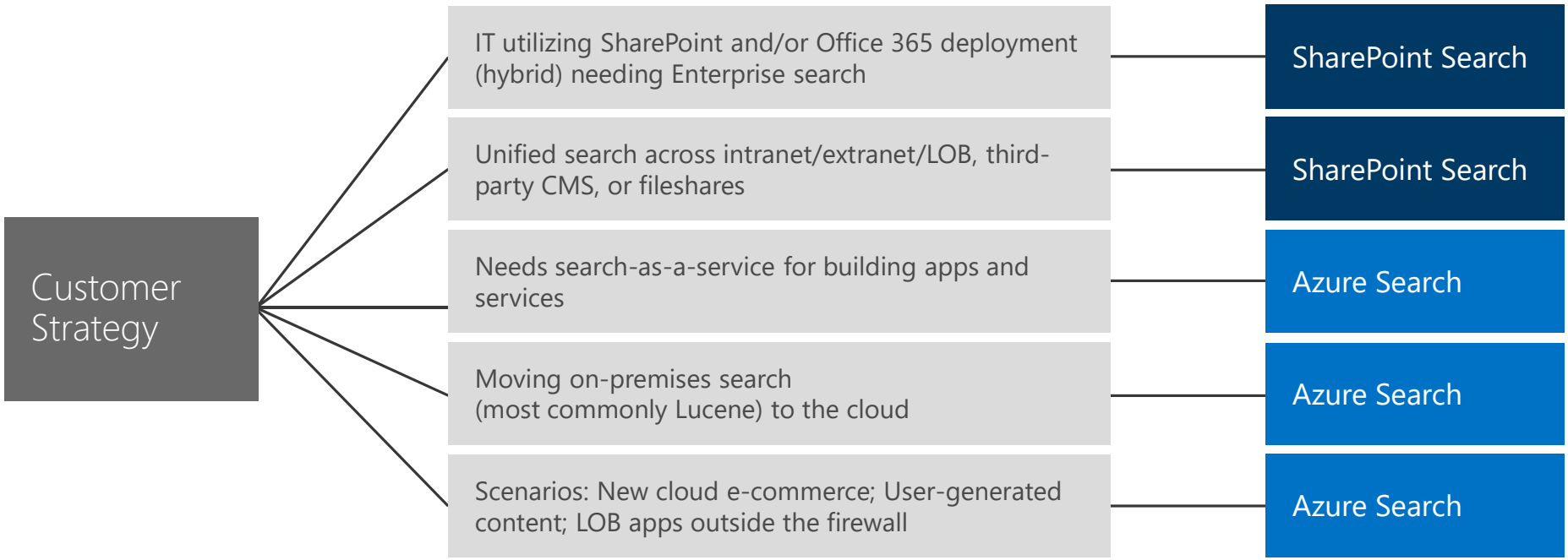
Public Preview: August 21, 2014
GA: March 5, 2015

Search landscape



What about SharePoint Search?

SharePoint 2013 provides enterprise-grade search functionality and an extensive connector eco-system for IW/Productivity scenarios



What does Azure Search cost?

Pricing principles

Capacity units can be combined

Search units stack horizontally and vertically

STANDARD OFFER (S1)

15 million documents

25 GB database storage

15 queries per second

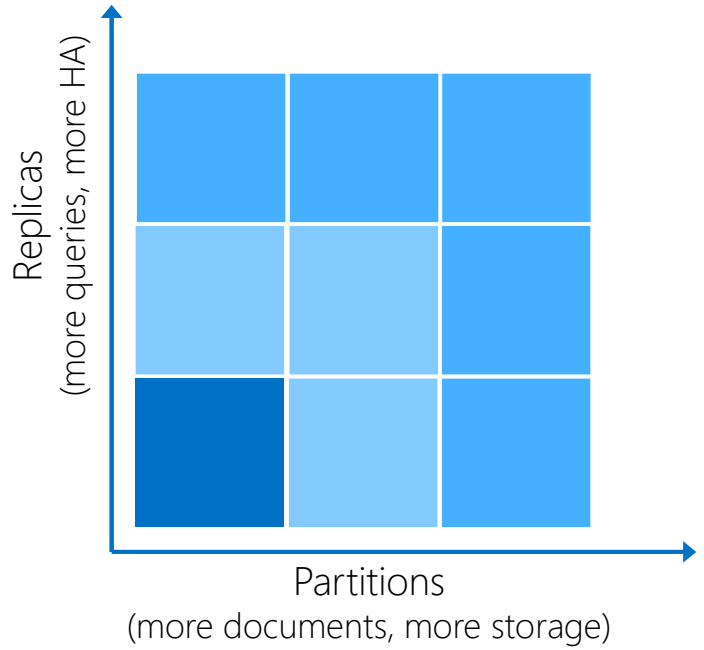
$250.00/month

(discounted 50% for Preview)

*Document count and throughput based on benchmark index and queries.

Key features

Simple API keyword search with user-friendly operators, hit highlighting, faceting, suggestions, scoring profiles, and geo-spatial data



Scaling and price considerations

The following table is a chart that lists replicas on the vertical axis, and partitions on the horizontal axis. The intersection shows the number of search units required to support each combination. For example, if you want 6 replicas and 2 partitions, this configuration would require 12 search units. To use 4 replicas and 2 partitions, you would need 8 search units.

6 replicas	6 SU	12 SU	18 SU	24 SU	36 SU	N/A
5 replicas	5 SU	10 SU	15 SU	20 SU	30 SU	N/A
4 replicas	4 SU	8 SU	12 SU	16 SU	24 SU	N/A
3 replicas	3 SU	6 SU	9 SU	12 SU	18 SU	36 SU
2 replicas	2 SU	4 SU	6 SU	8 SU	12 SU	24 SU
1 replica	1 SU	2 SU	3 SU	4 SU	6 SU	12 SU
	1 Partition	**2 Partitions**	**3 Partitions**	**4 Partitions**	**6 Partitions**	**12 Partitions**

As a general rule, most search applications tend to need more replicas than partitions.

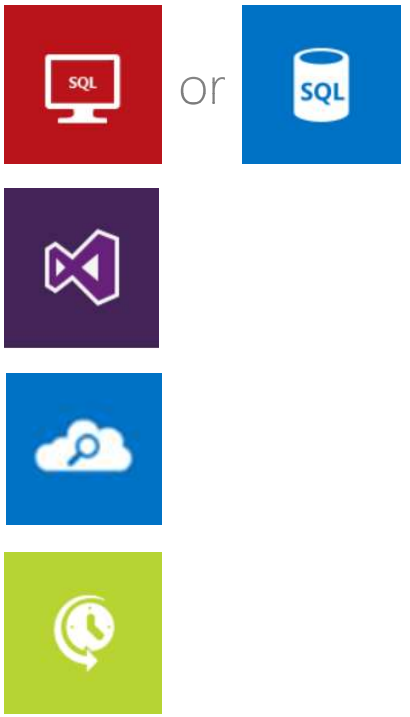
Index against SQL Server databases

Indexes based on Azure deployed SQL Server tables or views from SQL Server 2008 R2 or later – includes Azure SQL Database v12

Removes the need to convert tables to JSON format

SQL Server to Azure Search Sync

Add change tracking to SQL Server 2008 or later database
Detect changes using SQL Server change tracking
Push changes to Azure Search
Schedule sync using Azure Scheduler



Multi-language support

Analyzes fields with word-breaking, text normalization, and filtering out of terms

Converts all characters to their lowercase form with standard analyzer

Analyzes indexed documents and search terms during indexing and query processing

Improved accuracy with spelling mistakes, word order, and more…

Integration of Microsoft Office NLP

Office Natural Language Processor delivers:

- More languages (51)
- Improved accuracy with spelling mistakes, word order, and more

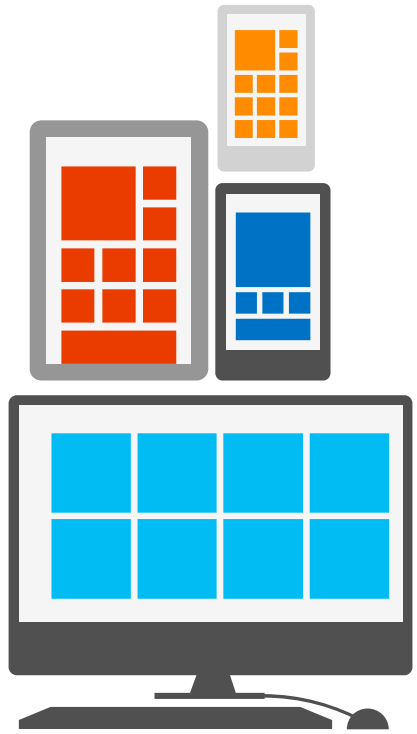
Ecommerce and online retail

Enable retail customers to find products through search and provide a great search experience with spelling corrections, suggestions, and faceting

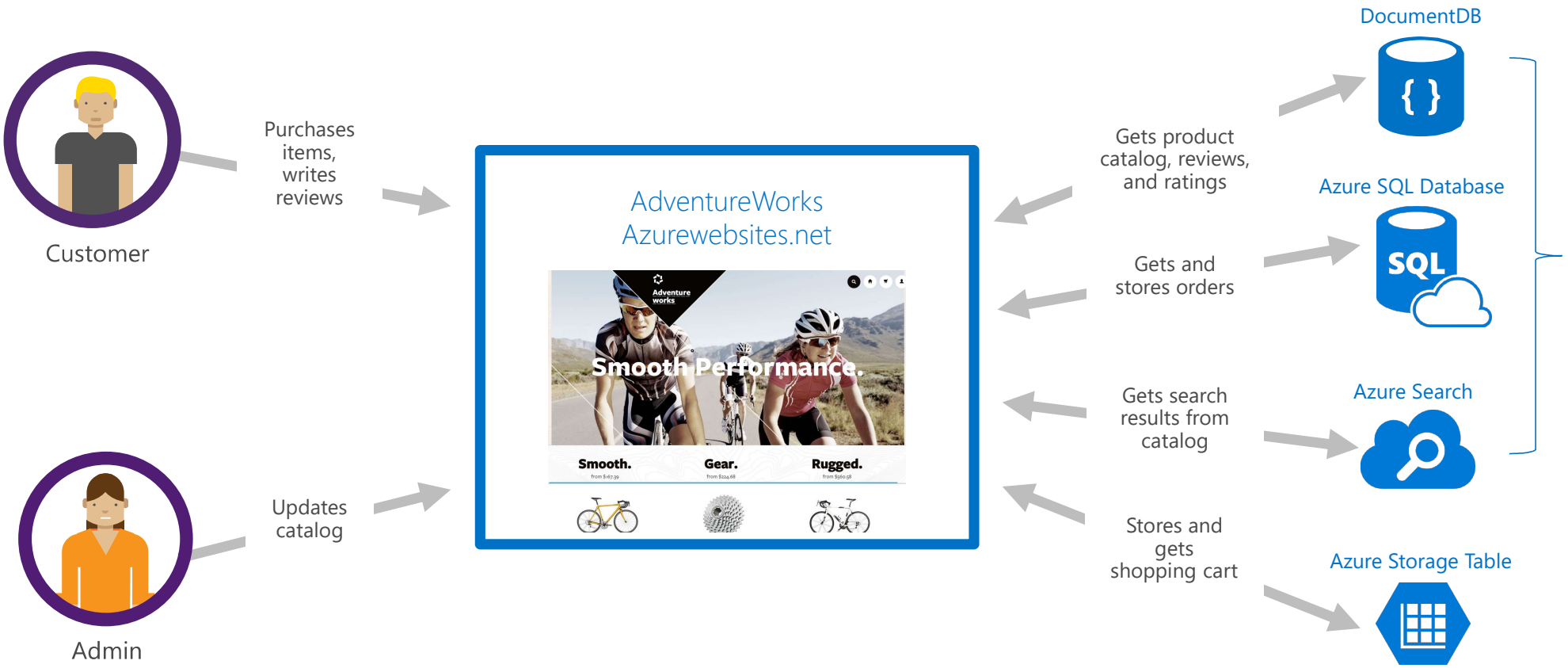
- Fine-tuned ranking models accommodating:
  - Popularity and rating
  - Inventory
  - Margin, discounts, and vendor promotions
- Support frequent index updates of price and stock levels
- Geo-search to find products available close to user
- High query rate

Features used:

Spelling correction, suggestions, faceting, geo-spatial support, scoring profiles



Demo overview



Summary and call to action

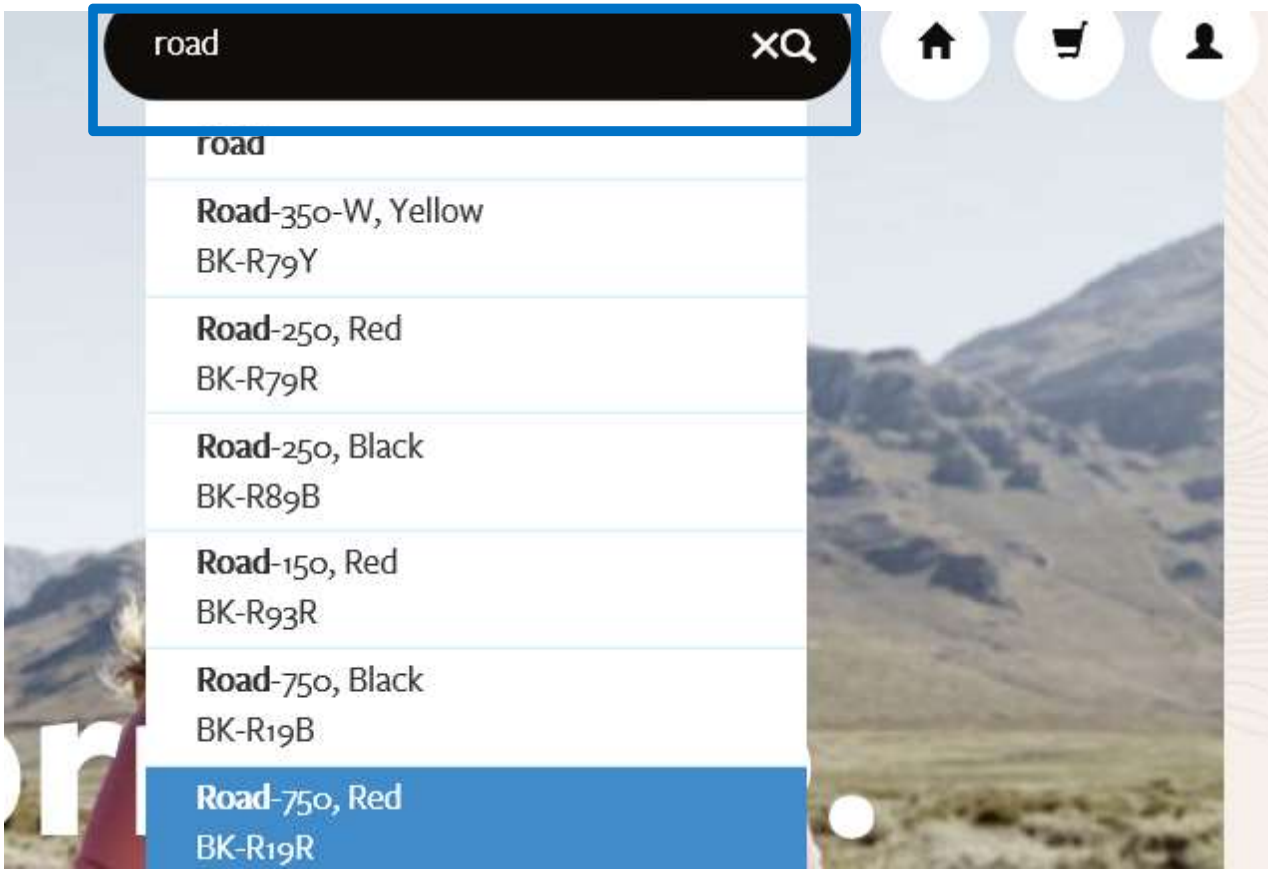
Azure Search is a new managed search service

It serves a different set of use cases than SharePoint Search and FTS in SQL Server and Azure SQL Database
Highly scalable with guaranteed throughput and storage
API-based controls make it simple to manage
Powerful search features with highly differentiated features on the roadmap

Consider Azure Search for these use cases

Customers building new web and mobile cloud-based applications
Reducing complexity around Search is a requirement
Customers running search on-premises or in virtual machines and looking for a managed service
E-commerce, social, user-generated content, and LOB apps

Azure Search - Suggestions



Azure Search

SQL Database
Tables
DocumentDB
HBase on HDInsight
Redis Cache
Azure Search

high

high **heels**
high **tops**
high **arch**

People use search as a natural, low friction way to interact with apps

Web search engines have set the bar high for search

Instant results, auto-complete, hit highlighting, great ranking, linguistics

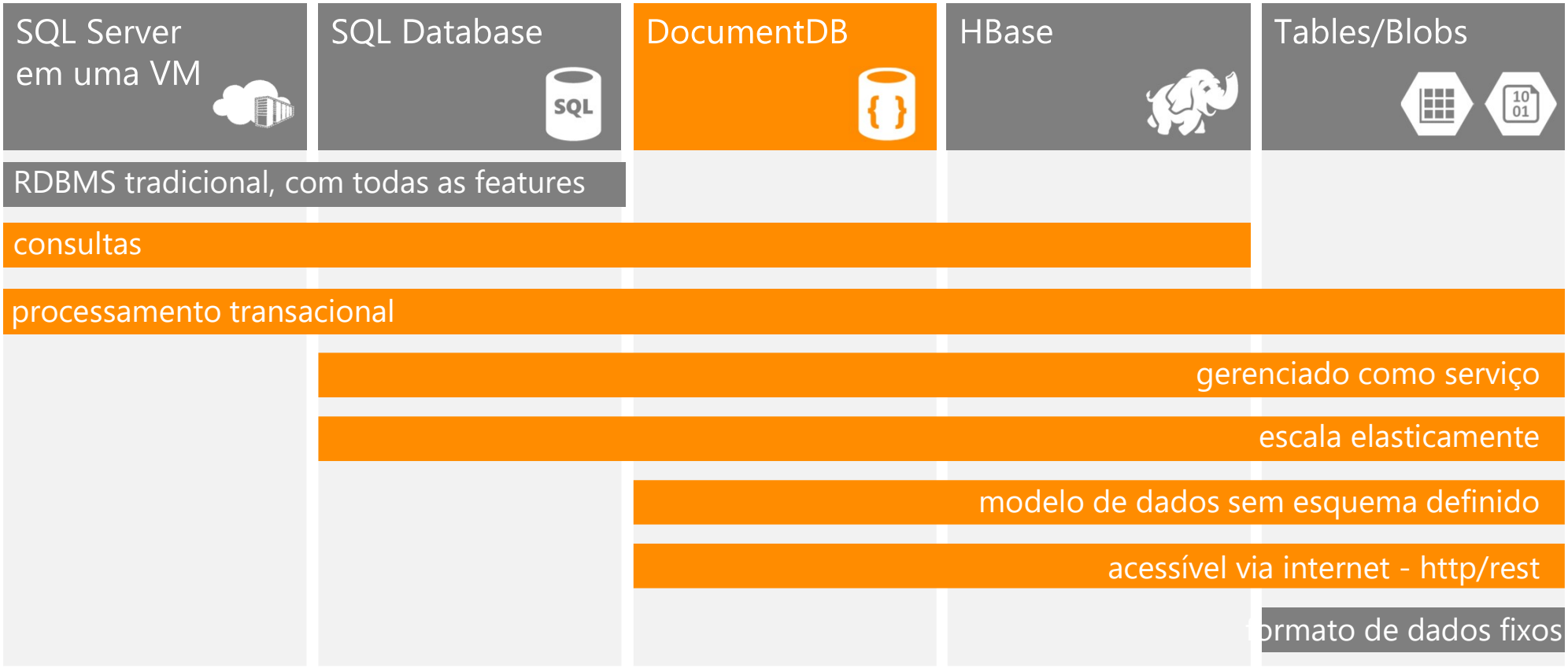
Search is hard and rarely a core expertise area

From infrastructure standpoint: availability, durability, scale, operations

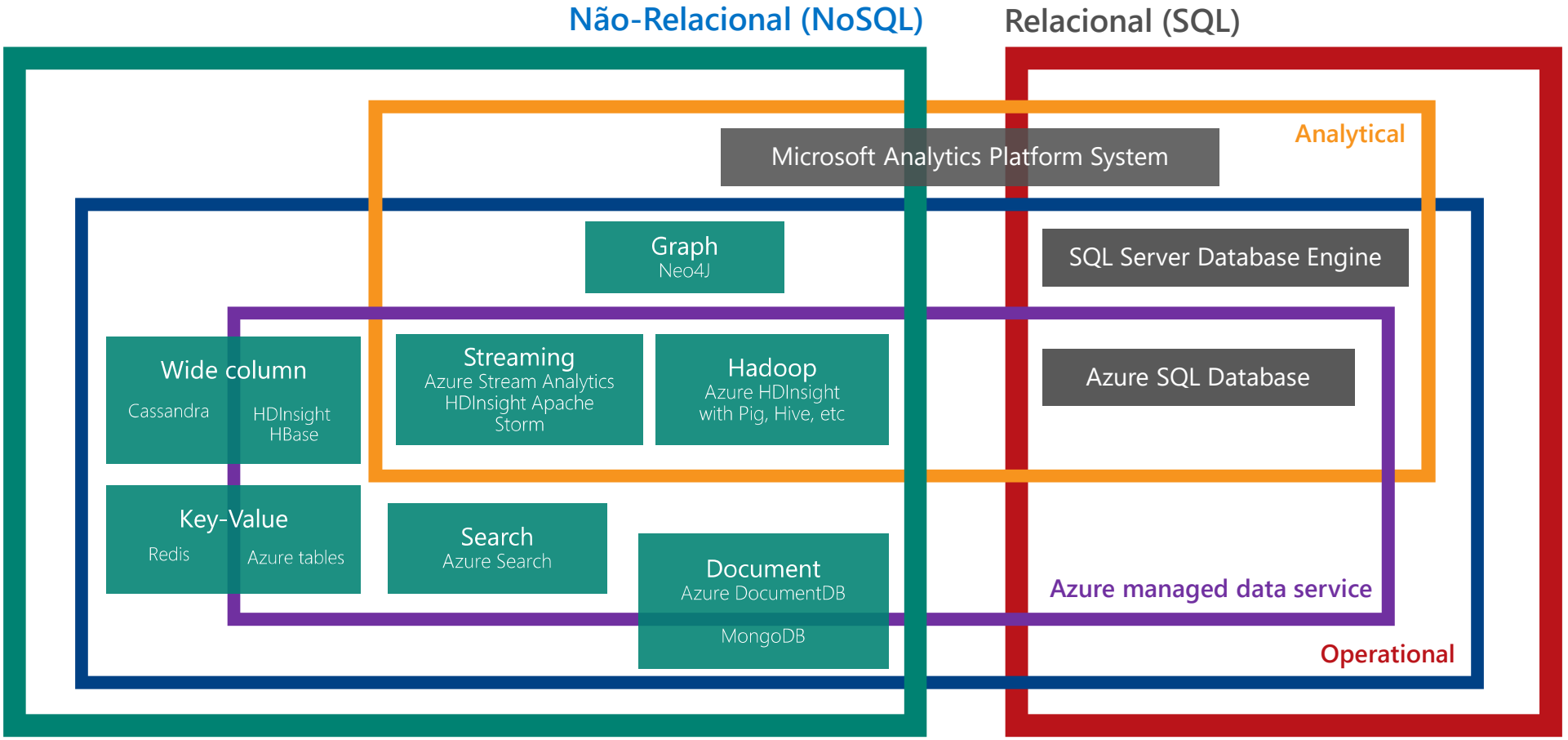
From the functionality standpoint: ranking, geo-spatial, input handling

Ecommerce and Online Retail
User Generated Content
Line of Business Applications

SQL e NoSQL

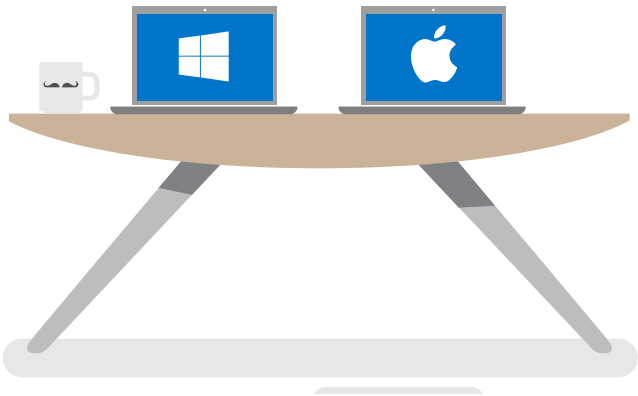

SQL Server em uma VM
SQL Database
SQL
DocumentDB
{ }
HBase
Tables/Blobs
10 01
RDBMS tradicional, com todas as features
consultas
processamento transacional
gerenciado como serviço
escala elasticamente
modelo de dados sem esquema definido
acessível via internet - http/rest
ormato de dados fixos

Not only SQL vs SQL overview

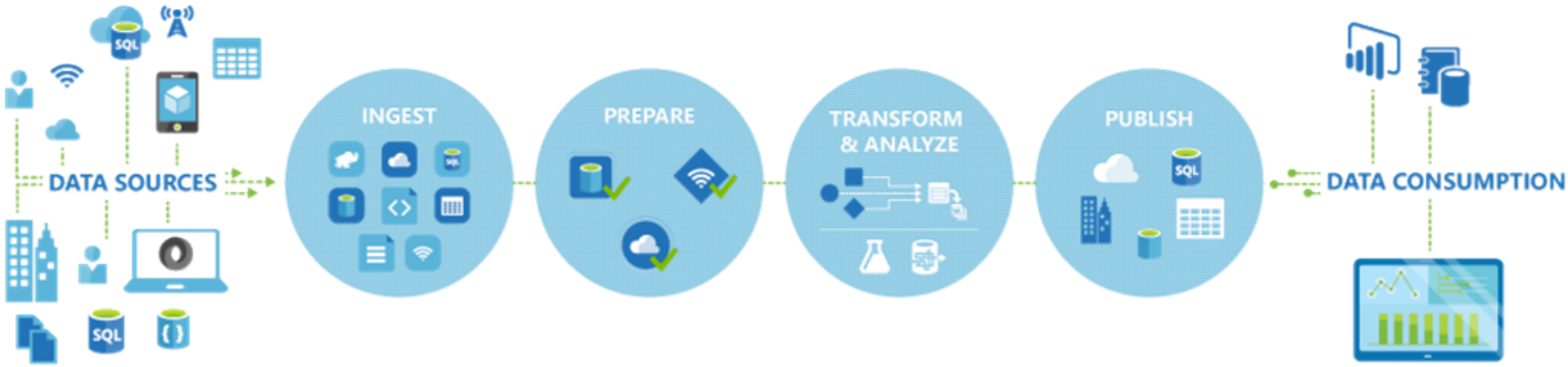


To review

When you need....	Because...	But not for...	Use ...
Relational store	Transactions, joins, structured data, familiar SQL query	Quickly changing data schemas	**SQL Database**
NoSQL key-value pair store	Low-cost, fast, massive scale	Rich query	**Tables**
NoSQL JSON document store	Flexible schema, familiar SQL query, low latency	Complex joins	**DocumentDB**
NoSQL wide-column store	Open-source, integration with Hadoop analytics	Operational simplicity	**HBase on HDInsight**
Cache	Increasing speed of an app	Primary data store	**Redis Cache**
Search service	Integrating search into an app	Primary data store	**Azure Search**



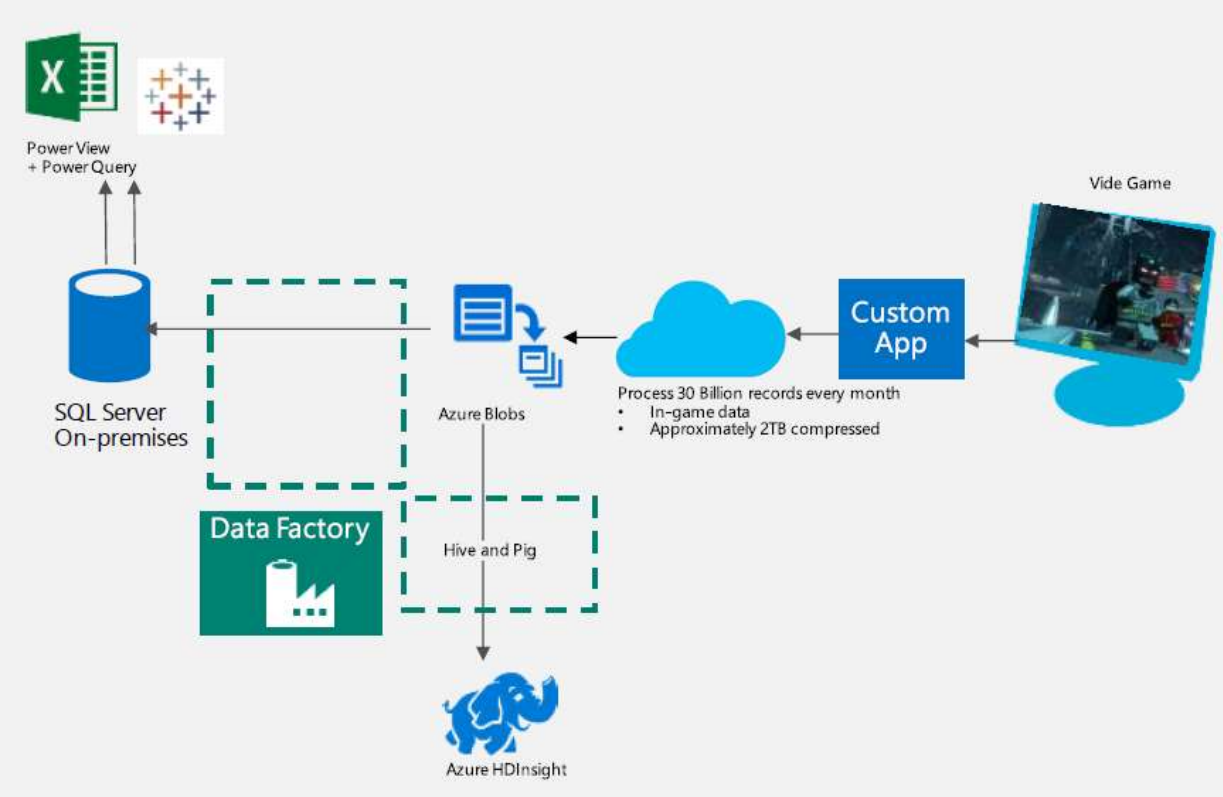
Azure Data Factory



Azure Data Factory

- A managed cloud service for building & operating data pipelines
- Part of the Cortana Analytics Suite

Data factory – real example



Game Dev Company / In-game analytics, Social sentiment

How They Did It

Collect data from games in Azure Blobs

- Ingest in-game data with custom App into Azure Blobs
- Data is in JSON format
- Processes 30 Billion records every month
- Approximately 2Terabytes of compressed, aggregated data
- Collect gamer in-game data

Data Factory to Orchestrate HDInsight

- Data Factory used to access and load data
- HDInsight used to do data aggregation and query
  - Hive and Pig
- Data Factory to load into SQL Server on-premises

Output sent to SQL Server for BI

- PowerView, Power Query and Tableau used to visualize

Redis Cache

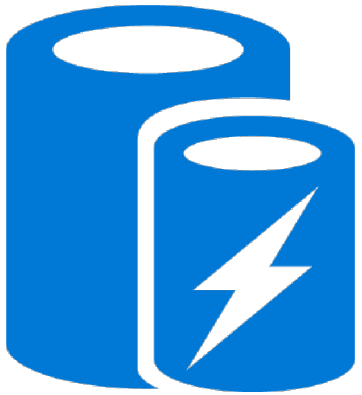
Redis Cache

"Redis é um open source "key-value cache e store".

[http://redis.io/]

Disponível no Azure como serviço em três camadas:

- **Basic** – Single Node Redis
- **Standard** – Dois nós Redis (Master/Slave) com SLA
- **Premium**
  - Persistência
  - Clusterização (somente memória nos outros)
  - VNET



Redis (REmote DIctionary Server)

- Open Source Cache
- Key-Value Cache e Store
- Complex Data Structure(s)
- Alto Throughput e baixa latência
- Ocupa a fronteira entre cache e NoSQL DB
- Suporta Lists e Sets
- Suporte de comunidade e muito popular
- Engine de cache recomendada pela Microsoft

Redis Cache

	BÁSICO **Cache Básico, ideal para desenvolvimento/teste.**	**PADRÃO** **Cache pronto par produção com replicação de servidor subordinado/mestre.**	**PREMIUM** **A camada pronta para empresas pode ser usada como um cache e dados de persistência. Desenvolvida para o máximo em dimensionamento e integração de empresas.**
Cache	Sim	Sim	Sim
Replicação e Failover	-	Sim	Sim
SLA	-	99,9%	99,9%
Configurar Redis (notificações keyspace, etc.)	-	Sim	Sim
Persistência de Dados Redis	-	-	Sim
Cluster Redis	-	-	Sim
Escalamento horizontal para as diversas unidades de Cache	-	-	Sim
Rede Virtual do Azure	-	-	Sim
Tamanho da memória	250 MB - 53 GB	250 MB - 53 GB	6 GB - 530 GB *
Desempenho de Rede	Baixo - Alto	Baixo - Alto	Moderado - Mais alto
Número máximo de conexões de cliente	20000	20000	40000

*Mais opções disponíveis mediante solicitação.

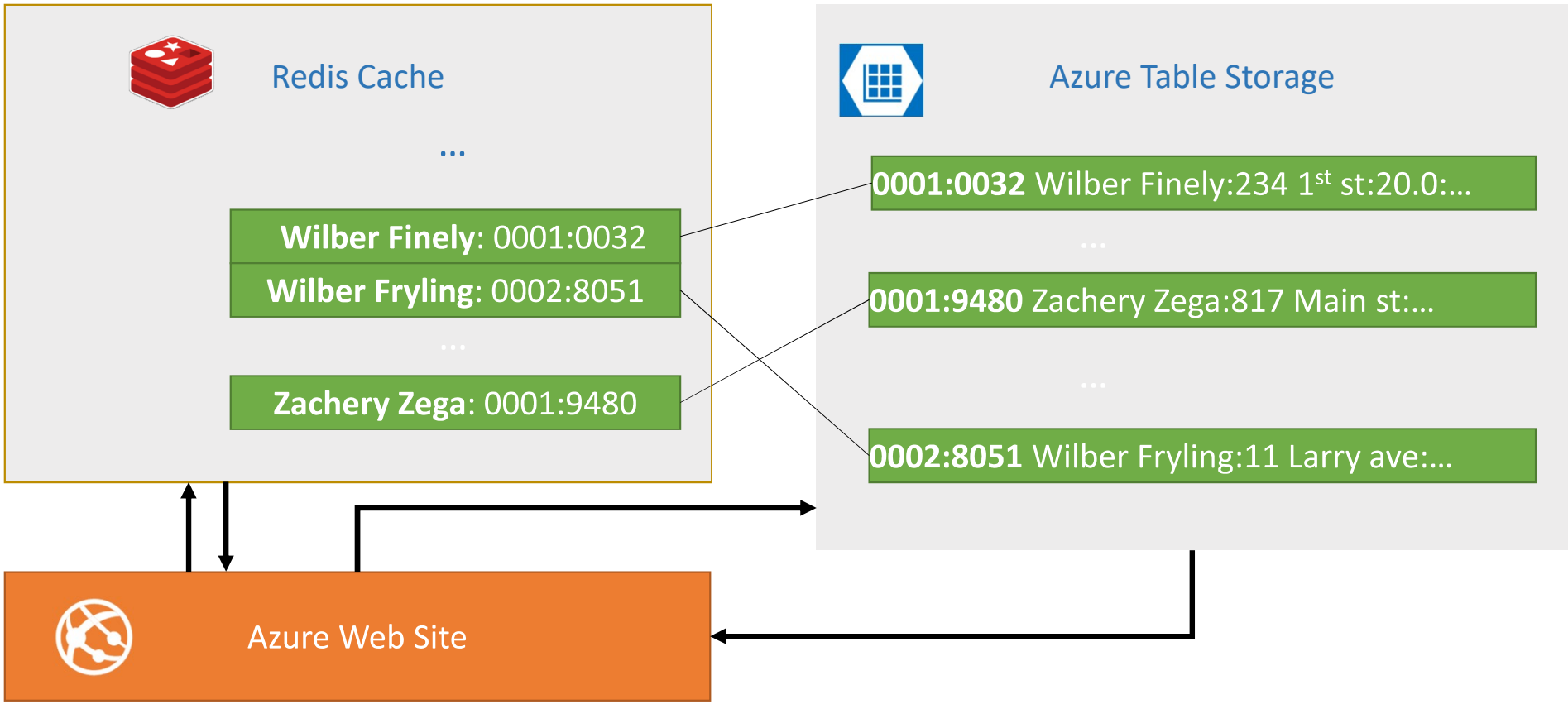
Caches em Azure

Azure Redis é a 4th oferta de serviço de Cache no Azure

Baseado no protocolo aberto do Redis, com suporte à C/C++, .NET, Go, Java, Node.js, Objective C, Perl, PHP, Python, Ruby e mais 20 linguagens.

Shared Cache	In-Role Cache	Managed Cache Service
• Primeira oferta • *Retired* – problemas de performance	• Built on Azure AppFabric Engine • Alta Performance • Usa recursos Cloud Service • Problemas como escalablidade horizontal • Acessível somente com cloud service • *Retired* em NOV/2016	• Oferece Recursos dedicados • Acessível de qualquer lugar via public endpoint & secret key • Alta performance • Baixa adoção • *Retired* em NOV/2016

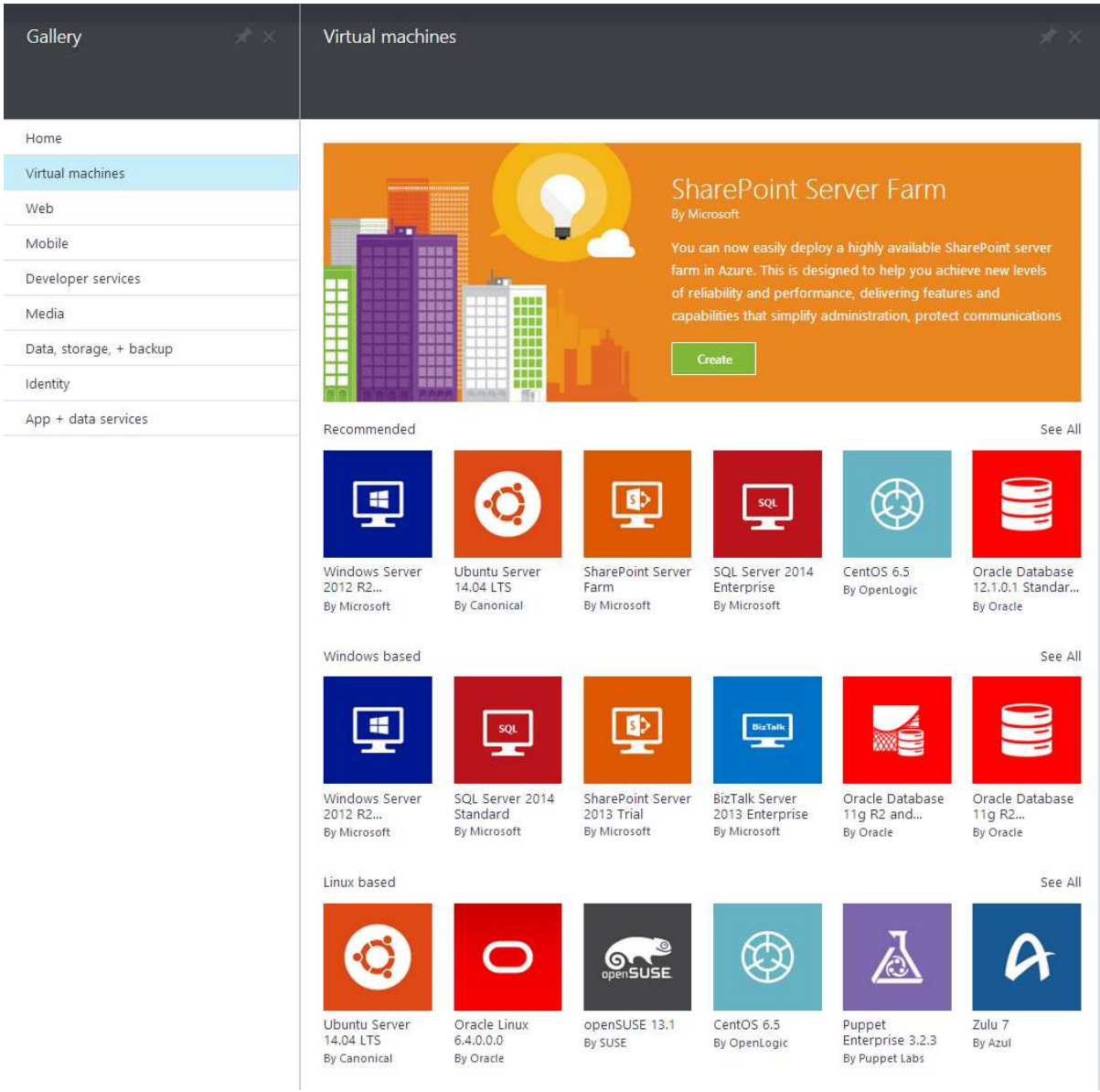
Query


Redis Cache
...
Wilber Finely: 0001:0032
Wilber Fryling: 0002:8051
...
Zachery Zega: 0001:9480
Azure Table Storage
0001:0032 Wilber Finely:234 1st st:20.0:...
...
0001:9480 Zachery Zega:817 Main st:...
...
0002:8051 Wilber Fryling:11 Larry ave:...
Azure Web Site

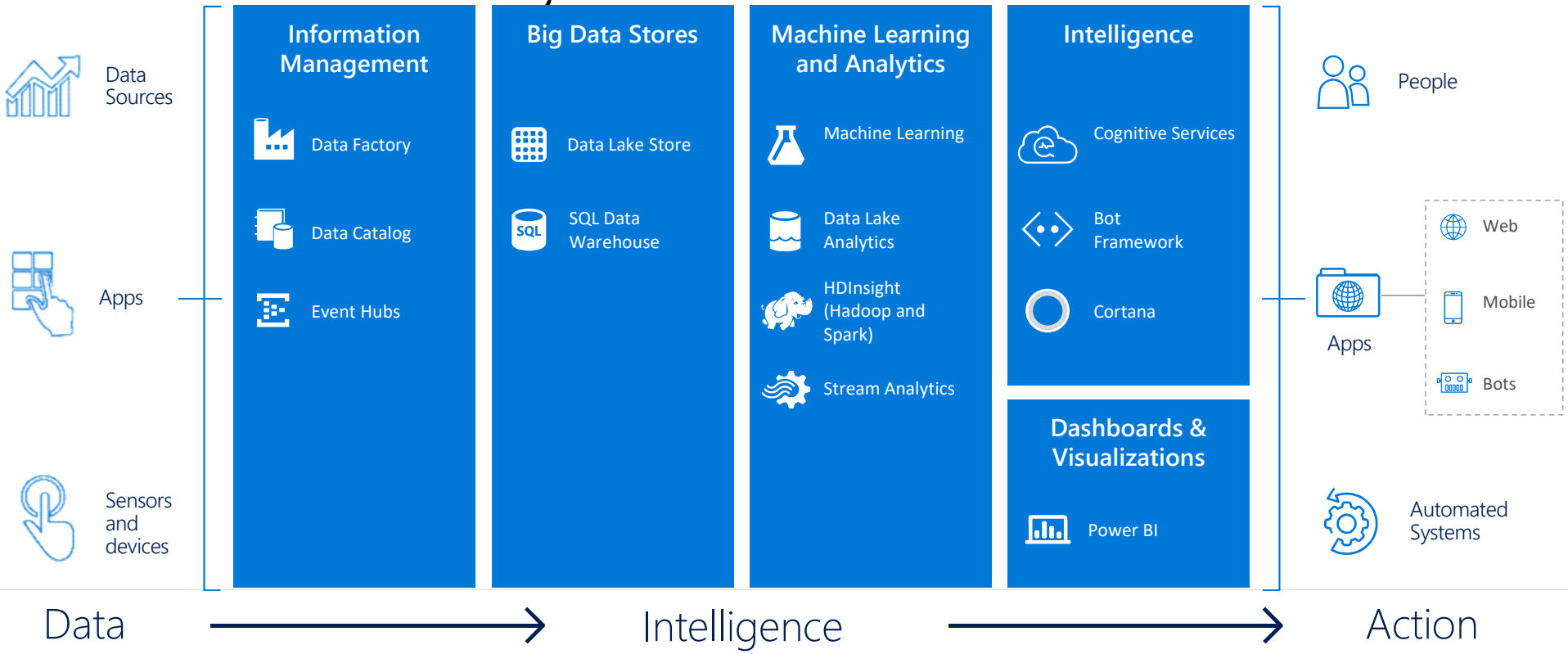
Market Place

Azure Marketplace

Linux
Oracle
IBM
SAP
Informatica
Cloudera
Hortonworks
DataStax



Cortana Intelligence Suite




Microsoft