Data Analytics
In the Era of Cloud Computing

Raghu Ramakrishnan
Technical Fellow
CTO for Data
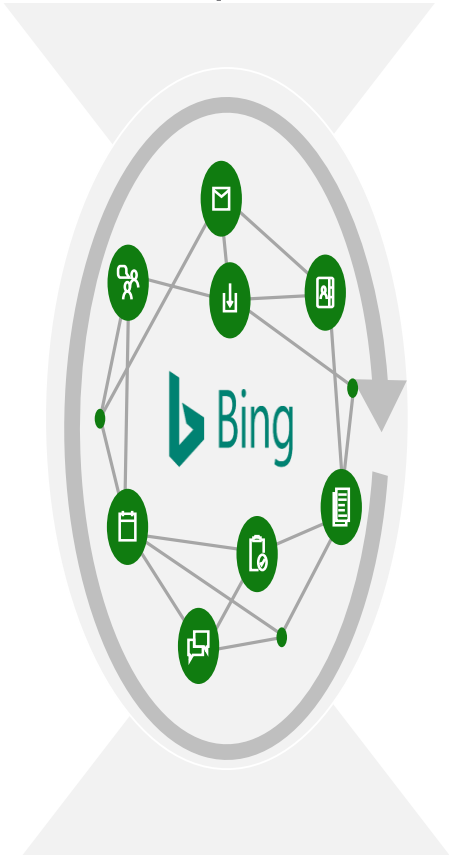
54

Azure regions

Microsoft Cloud ←→ Windows Linux Azure IoT Intelligent Edge

A Microsoft Example of Data and AI
Intelligent Local Queries

Query: Where's the nearest fruit smoothies
Location: Omaha, Nebraska

Search queries, views, click throughs, ...

World graph
- People
- Places
- Things
- Actions

2B+ entities
130B+ Web pages
50B+ facts

Web pages, Web documents, Images, ...

Tropical Smoothie Cafe — Share
Yelp (9) · $$$$
Sandwiches · Juices · Open · 15.6 mi

CALL   DIRECTIONS   WEBSITE

7902 Towne Center Pkwy Ste 109, Papillion, NE 68046

Open   10:00 AM - 9:00 PM

View menu

(402) 614-9050

Suggest an edit

Local results for Where's the nearest fruit smoothies near Omaha, Nebraska

1 Juice Stop   1.0 mi
   Yelp (15)
   $$$$ · Juices · Fast Food
   Open 8:00 AM - 9:00 PM
   510 S 12th St, Omaha

2 Juice Stop   6.3 mi
   Yelp (16)
   $$$$ · Juices · Fast Food
   Open 7:00 AM - 9:00 PM
   2119 S 67th St, Omaha

3 Smoothie King   10.0 mi
   Yelp (13)
   $$$$ · Juices · Fast Food
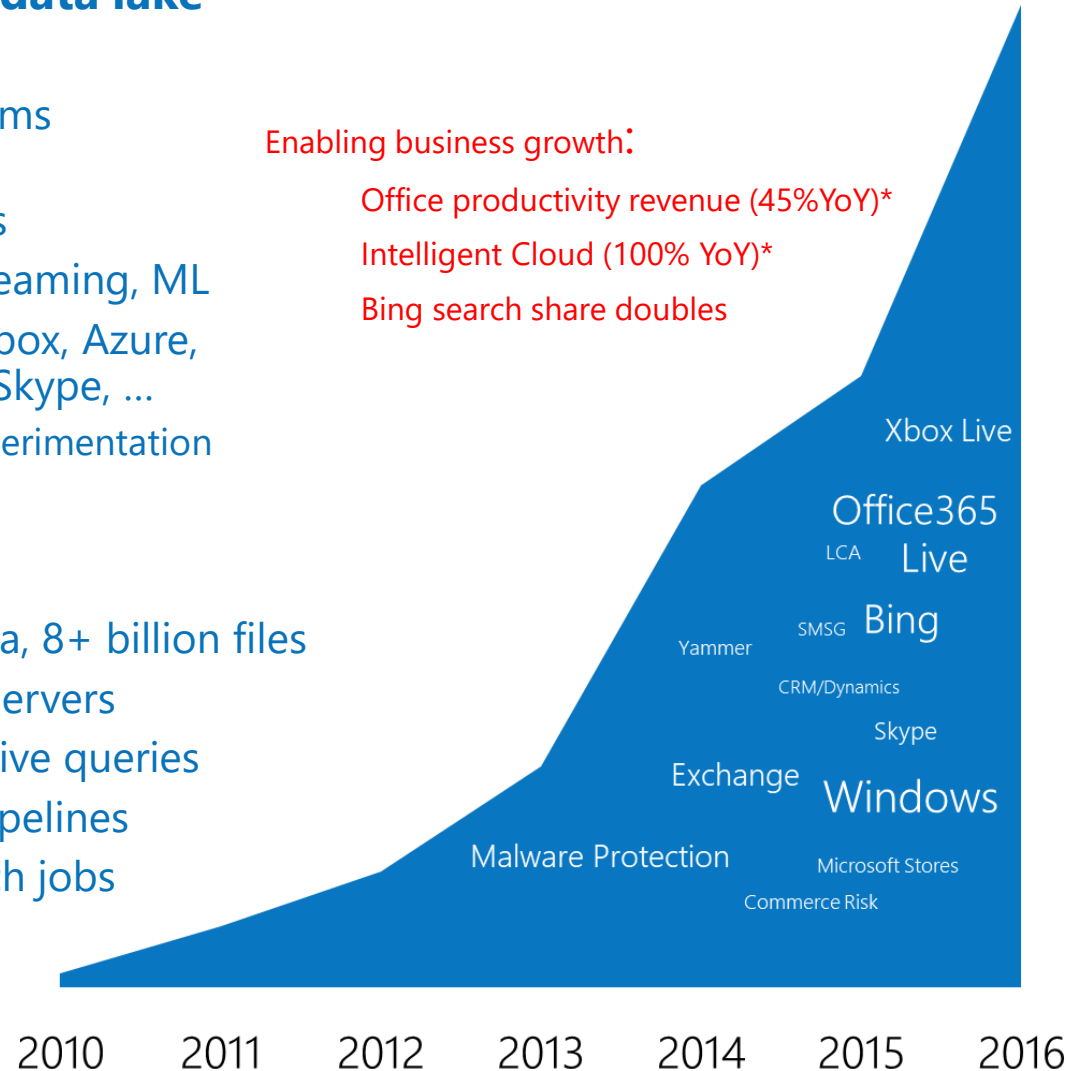   Open 7:00 AM - 9:00 PM
   8900 W Center Rd Ste 101, Omaha

Microsoft's Internal Big Data Service

Microsoft's internal data lake

- A data lake for all teams @Microsoft
- Good developer tools
- Batch, Interactive, Streaming, ML
- Used across Office, Xbox, Azure, Windows, Ads, Bing, Skype, ...
- Production jobs and experimentation

By the numbers
- 9+ Exabytes of data, 8+ billion files
- 100Ks of physical servers
- Millions of interactive queries
- Huge streaming pipelines
- 100Ks of daily batch jobs
- 15K+ developers
- 300+ teams

Enabling business growth:
- Office productivity revenue (45%YoY)*
- Intelligent Cloud (100% YoY)*
- Bing search share doubles

Xbox Live

Office365 Live

LCA

SMSG Bing

Yammer

CRM/Dynamics

Skype

Exchange Windows

Malware Protection

Microsoft Stores

Commerce Risk

2010 2011 2012 2013 2014 2015 2016

Azure Data Lake Store

A data lake for everyone

- Microsoft's serverless Big Data platform
- **Fully aligned with Hadoop ecosystem** and standards, with full support for Hadoop tools and engines as well as unique Microsoft capabilities
- Migrated to ADLS
- **1P = 3P**

Tesla's Digital Feedback Loop
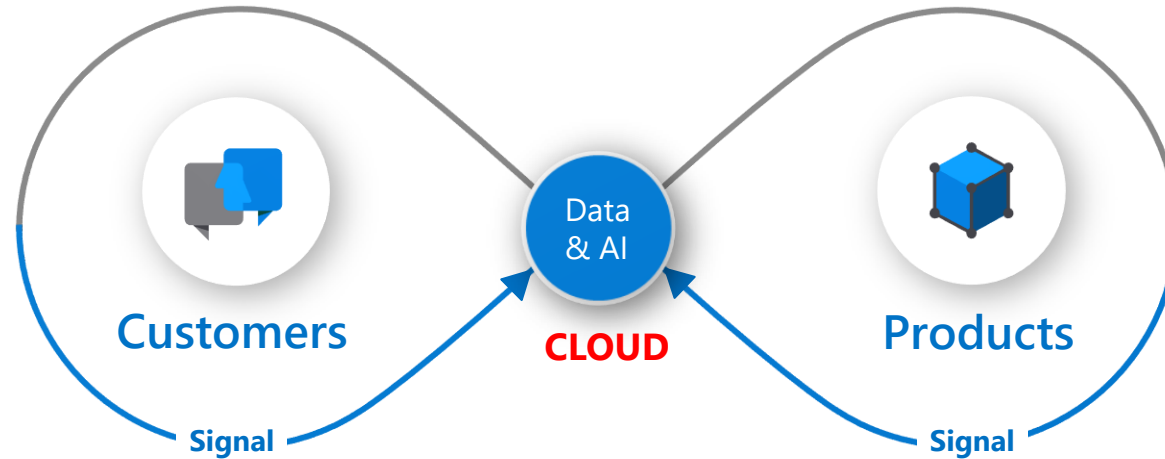
Customers

CLOUD

Products

Data & AI

Signal

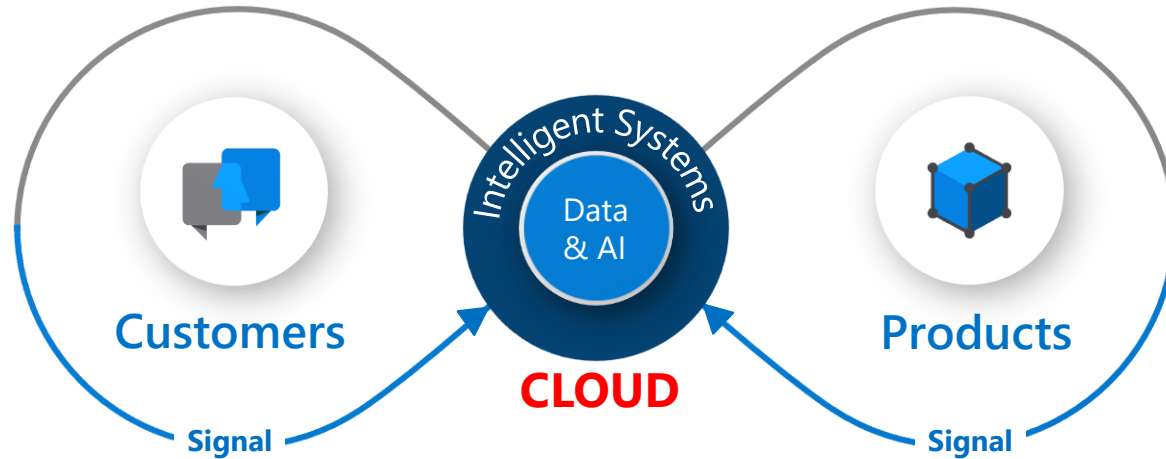Signal

Customer interaction data

TESLA Forums

My Model S takes the curve easily at 45, but I frequently come up on cars who take it at 30, and I have to disengage and hit the brake myself

Jnsparke

Driving telemetry captured through instrumented vehicles **(EDGE)**

Tesla's Digital Feedback Loop

TESLA

Intelligent Systems

Data & AI

Customers

CLOUD

Products

Signal

Signal

Customer interaction data

TESLA Forums

My Model S takes the curve easily at 45, but I frequently come up on cars who take it at 30, and I have to disengage and hit the brake myself

Jnsparke

Issues identified by combining customer and product signal

Over **90%**

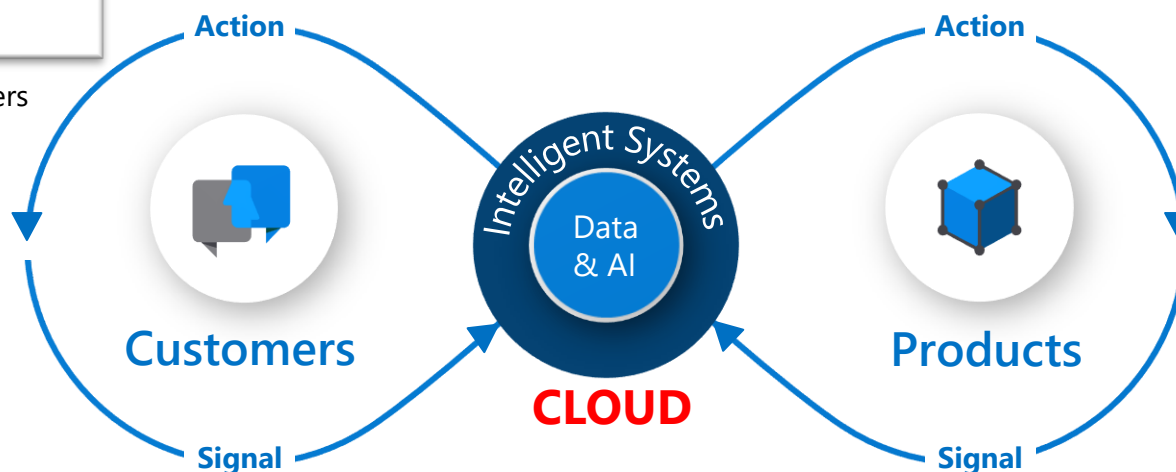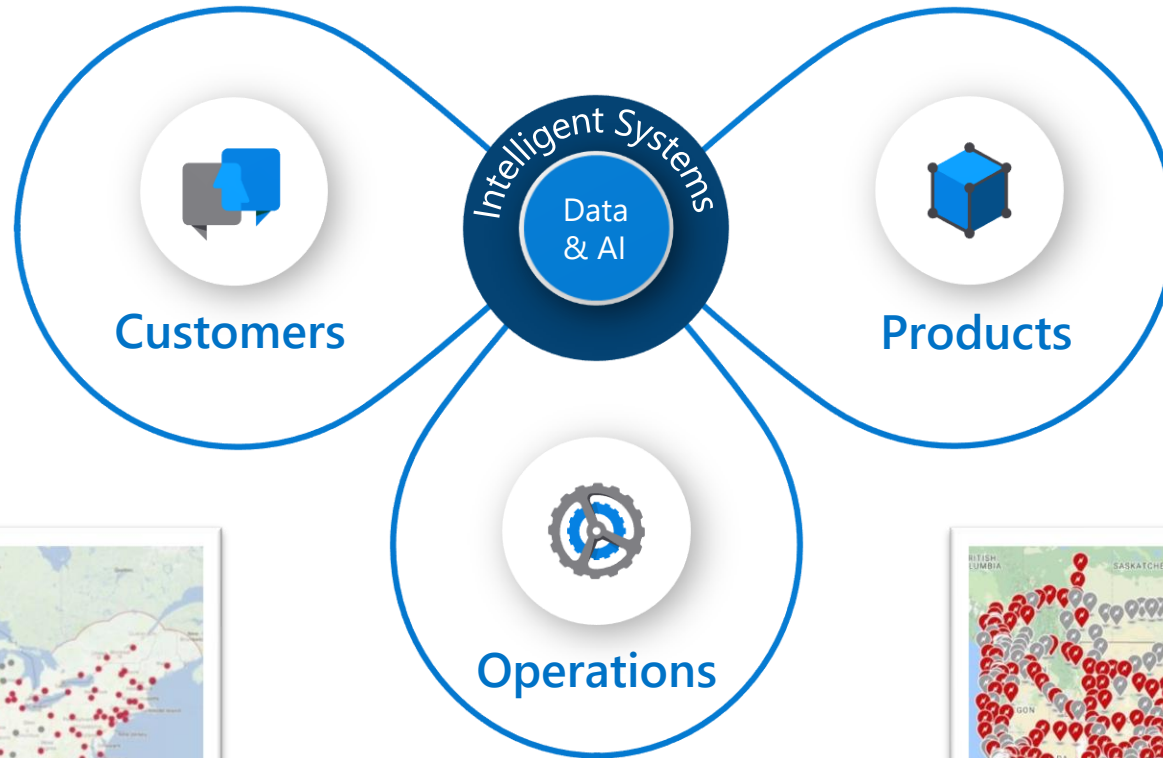of issues are identified remotely

Driving telemetry captured through instrumented vehicles **(EDGE)**

Tesla's Digital Feedback Loop

TESLA

> **Elon Musk** ✔
> @elonmusk
>
> Autopilot 1.01 coming soon: curve speed adaption, controller smoothness, better lane holding on poor roads, improved fleet learning!
>
> 🔔 Follow

Closing the loop with customers

Pushes updates for Autopilot over the air

Action

Intelligent Systems

Data & AI

Action

Customers

Products

CLOUD

Signal

Signal

Customer interaction data

Issues identified by combining customer and product signal

Driving telemetry captured through instrumented vehicles **(EDGE)**

TESLA Forums

My Model S takes the curve easily at 45, but I frequently come up on cars who take it at 30, and I have to disengage and hit the brake myself

Jnsparke

Over **90%**

of issues are identified remotely

Tesla's Digital Feedback Loop

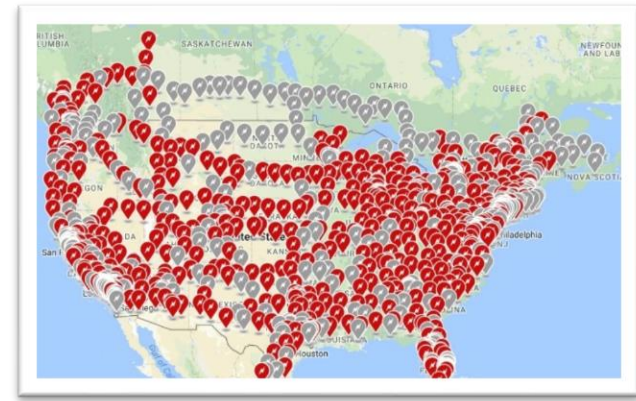Customers' driving data helps inform the expansion of Tesla's Supercharger network
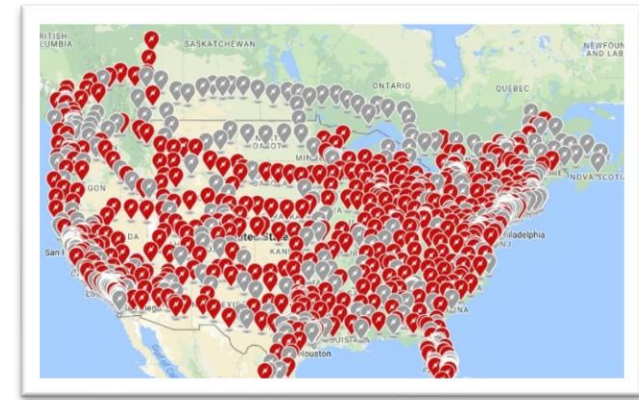
Tesla's Digital Feedback Loop

Technicians are empowered with historical car telemetry to aid in troubleshooting

TESLA

People

Intelligent Systems

Data & AI

Customers

Products

Operations

Customers' driving data helps inform the expansion of Tesla's Supercharger network

The Digital Feedback Loop

1 **Signal:** Capture digital signal from all touchpoints

2 **Intelligence:** Connect and synthesize data

3 **Action:** Improve key outcomes

▶ Improved decisions at the edge, guided by data analysis in the cloud

▶ Autonomous action with intermittent connectivity

People

Signal

Action

Action

Signal

Customers

Intelligent Systems

Data & AI

Products

Signal

Action

Action

Signal

Operations

USACE NavPortal and Real-time Dredging

- 25,000+ miles of navigable inland water in CONUS
- ESRI maps not showing real-time data
- Too much sensor data to parse out into actionable insights
  - Water Temp
  - Sediment analysis
  - Vessel traffic
  - Wave height
- Dredging on a pre-set schedule, not based on need

➢ **NavPortal** "The ~~Google~~ maps of vessel traffic and waterway management"

Bing

Data Analytics Trends

➢ Digital loops across Cloud and Edge

➢ Unified Insights and Governance

Challenges

1 Large portions of the enterprise are not digitized

34% of organizations indicate poor data reliability is an obstacle for growth, while 28% cite insufficient data systems

People

DATA

Customers

DATA

Products

DATA

2

DATA

Operations

3

Data is siloed across different parts of the organization

31% indicate data silos prevent their enterprise from maximizing value
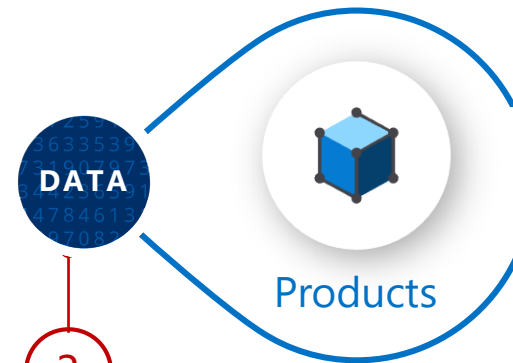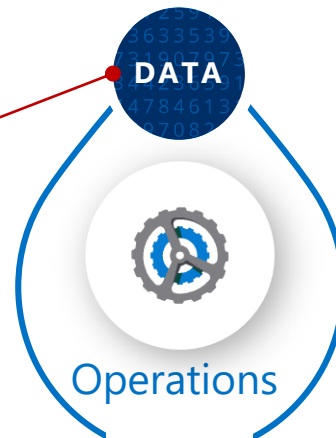
Generating insights requires a high degree of technical expertise

30% cite lack of analytical talent as a hindrance to generating insights

Organizational Silos

Many sources and types of data

Siloed Analytic Tools

Microsoft Azure Government (IL4)

Off-shore Sensors
- Pipeline
- Hopper
- SCOW

IoT SDK

SOAP

Message Routing by Plant Type

IoT Hub

HttpRequest

WCF App Service

AVRO blob (ROUTE)

CSV blob

Virtual machine (python)

AVRO blob (CAPTURE)

CSV blob

Blob Trigger

Event Hub

EH Trigger

SQL Database
Error Logging/Admin

Cosmos DB
Activity Tracker
Backup Message Repo

Logic App
Inactivity Alert Workflow

SQL Data Warehouse (polybase)

Power BI

On-Premises

Secure SSL Outbound-Only

EventProcessorHost Windows Service

Oracle Queue

Matlab Cycle-State Algorithms

NavPortal Architecture

Data Silos Optimized for Different Tasks

Big Data

Data Warehouse

OLTP

Global Apps

Product
silos

Spark, Hive, ML...

Data Lake

Meta data

Governance

Azure SQL DW

Analytics-optimized

Meta data

XACT_STATE

Governance

Azure SQL DB

Update-optimized

Meta data

XACT_STATE

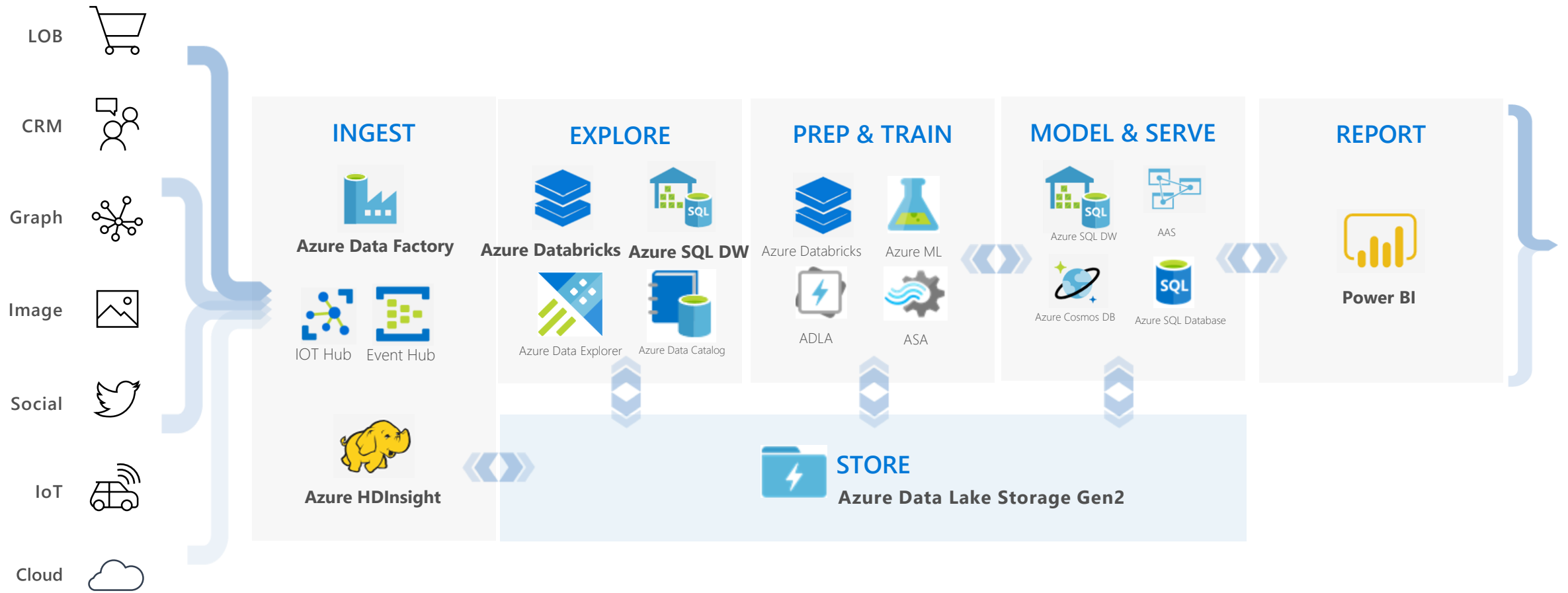Governance

Azure Cosmos DB

Document model

Meta data

XACT_STATE

Governance

Real Scenarios Involve Many Tasks!

Azure Synapse Analytics

LOB

CRM

Graph

Image

Social

IoT

Cloud

INGEST   **EXPLORE**   **PREP & TRAIN**   **MODEL & SERVE**   **REPORT**

Azure Synapse Analytics

Power BI

STORE
Azure Data Lake Storage Gen2

https://azure.microsoft.com/en-us/services/synapse-analytics/
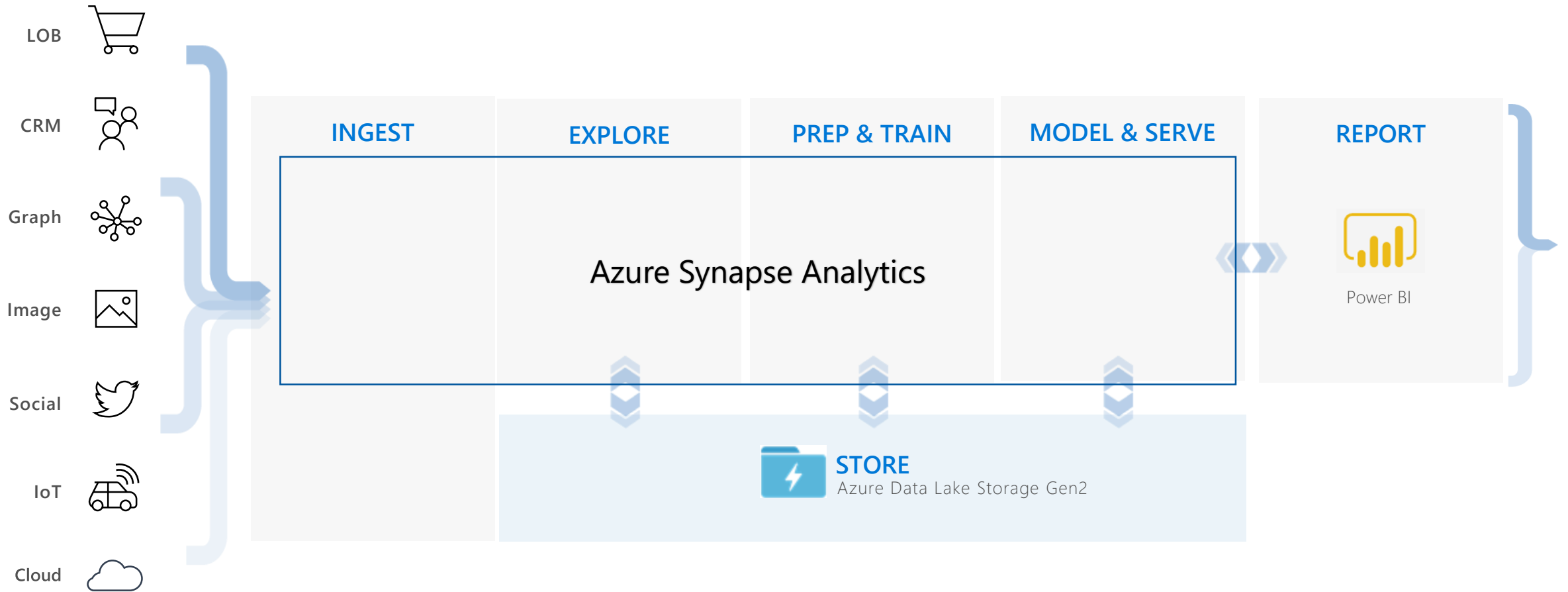
Scalability: All TPC-H Queries at 1PB Scale!

Elastic DQP – Unlimited Scale

TPCH 1PB

■Query time in minutes
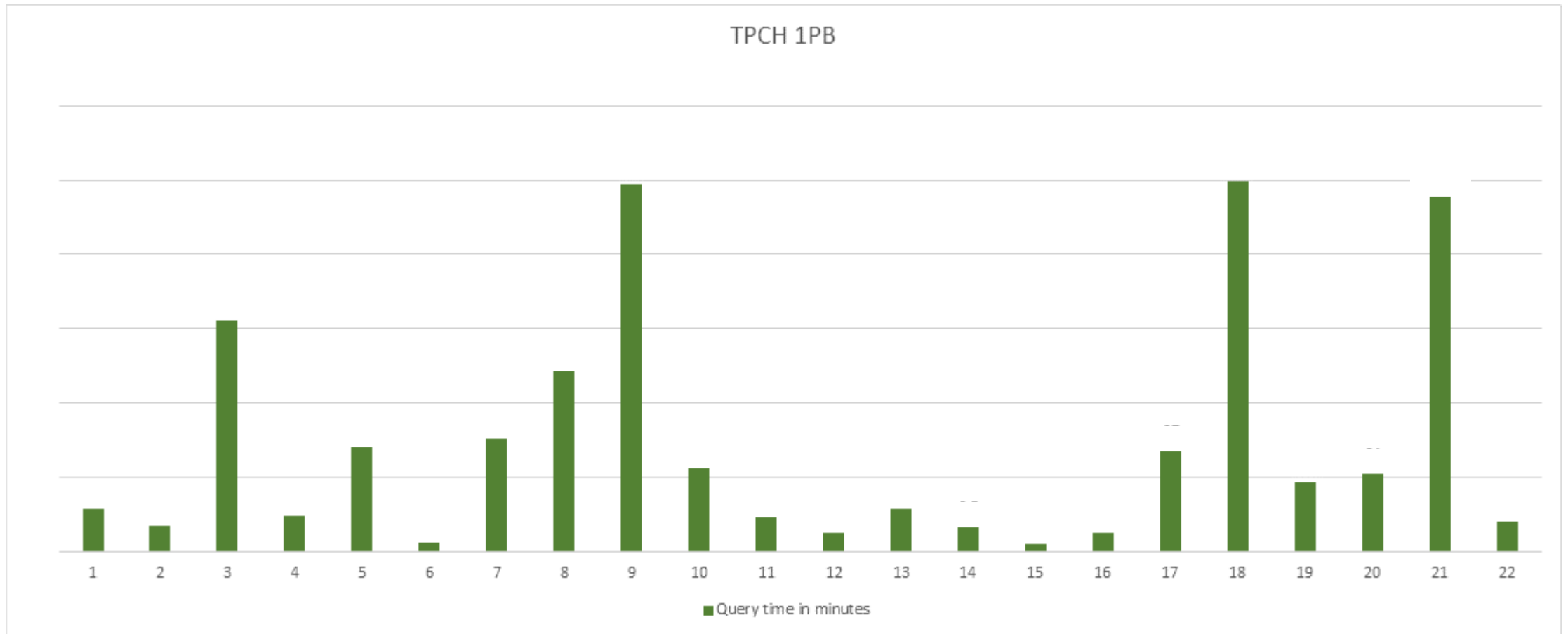
Data In the Cloud
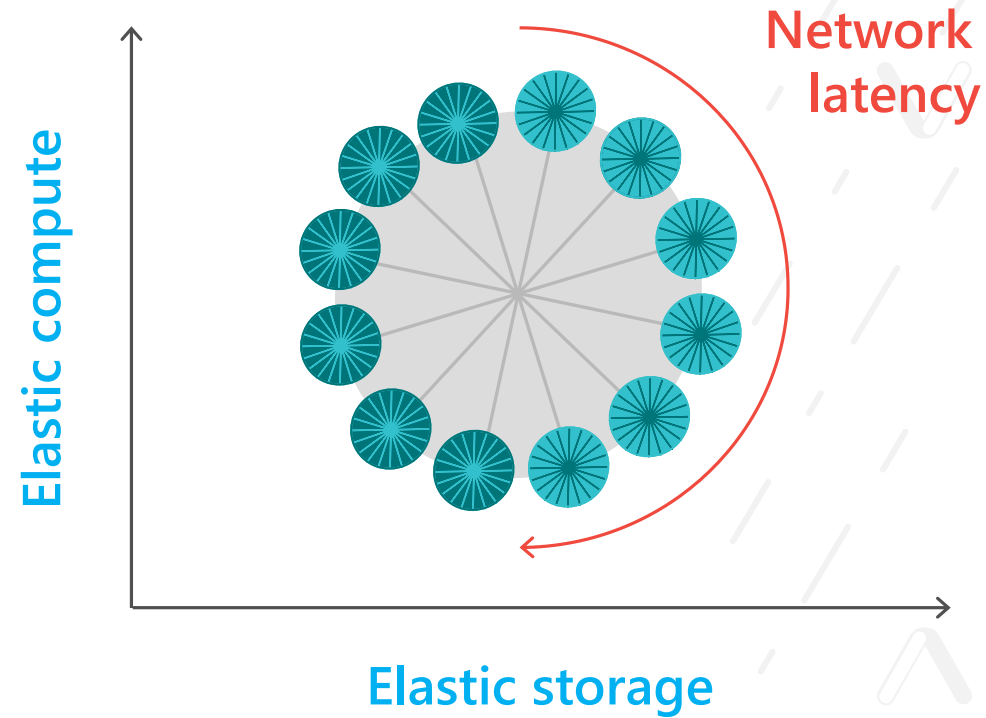
Unified Data Suite and Governance

One SQL Head

| Spark, Hive, ML... | SQL | Azure Cosmos DB |

Data Lake

Analytics-optimized

Update-optimized

Document storage

Meta data

Meta data

Meta data

XACT_STATE

XACT_STATE

XACT_STATE

Global apps

Complex tools must be simple to use

One pane of glass for Governance

🛡 Governance
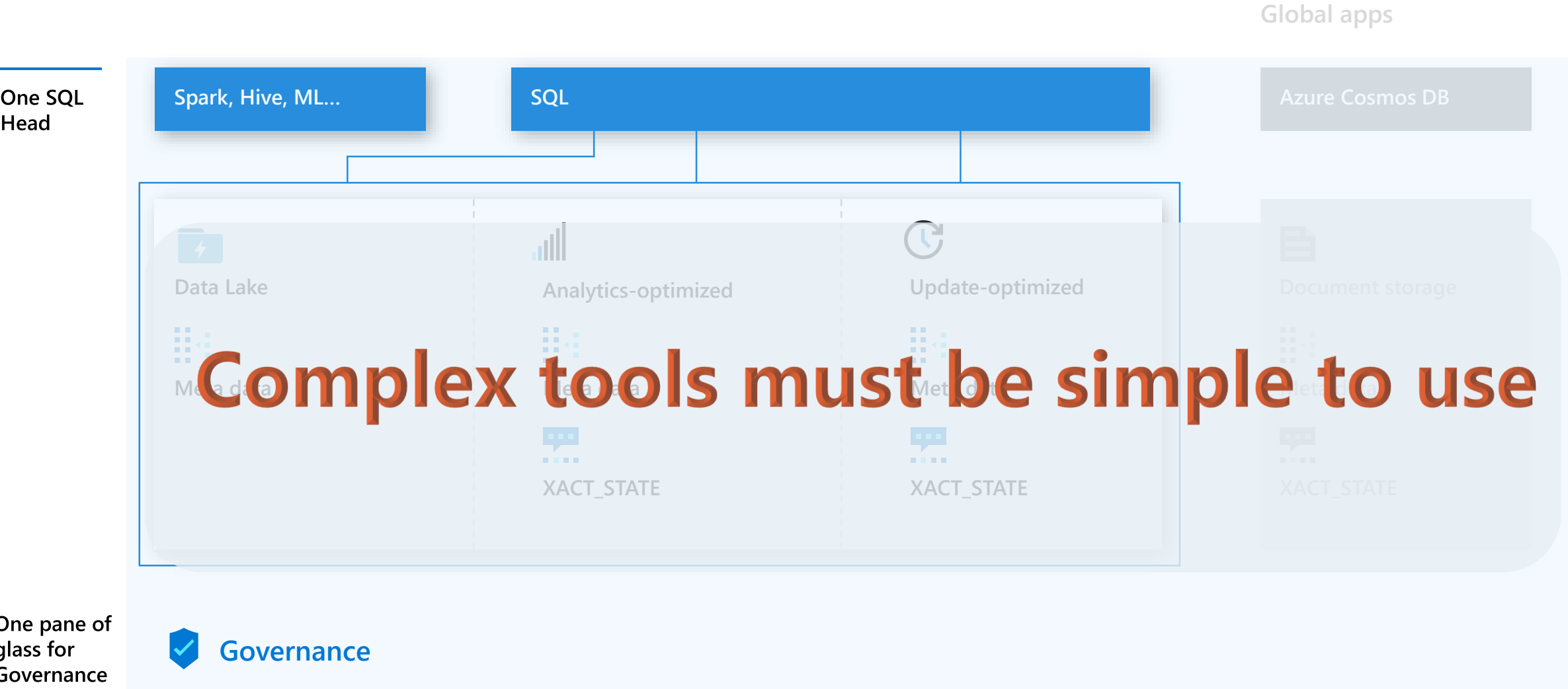
Azure Hybrid

Innovation anywhere with Azure

Microsoft Azure

Azure Stack
Integrated systems

Azure Arc
Any datacenter, any cloud

Azure IoT
Any edge device

Management | **Security + Identity** | **App + Data Services** | **Dev Tools + DevOps**

The Trusted Cloud

More certifications than any other cloud provider

GLOBAL

| ISO 27001 | ISO 27018 | ISO 27017 | ISO 22301 | SOC 1 Type 2 | SOC 2 Type 2 | SOC 3 | CSA STAR Self-Assessment | CSA STAR Certification | CSA STAR Attestation |

US GOV

| Moderate JAB P-ATO | High JAB P-ATO | DoD DISA SRG Level 2 | DoD DISA SRG Level 4 | DoD DISA SRG Level 5 | SP 800-171 | FIPS 140-2 | Section 508 VPAT | ITAR | CJIS | IRS 1075 |

INDUSTRY

| PCI DSS Level 1 | CDSA | MPAA | FACT UK | Shared Assessments | FISC Japan | HIPAA / HITECH Act | HITRUST | GxP 21 CFR Part 11 | MARS-E | IG Toolkit UK | FERPA | GLBA | FFIEC |

REGIONAL

| Argentina PDPA | EU Model Clauses | UK G-Cloud | China DJCP | China GB 18030 | China TRUCS | Singapore MTCS | Australia IRAP/CCSL | New Zealand GCIO | Japan My Number Act | ENISA IAF | Japan CS Mark Gold | Spain ENS | Spain DPA | India MeitY | Canada Privacy Laws | Privacy Shield | Germany IT Grundschutz workbook |