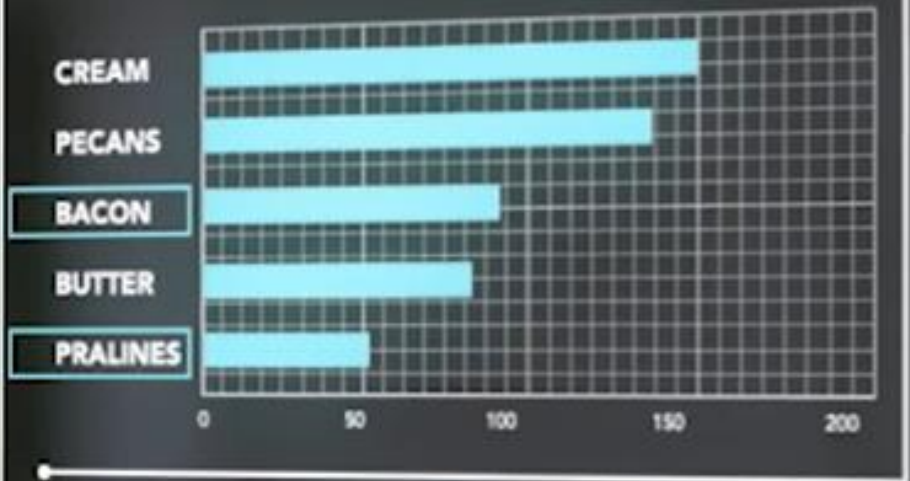


**BEST SELLER:  
PECANS & CREAM**

□ SOCIAL AFFINITY SEARCH



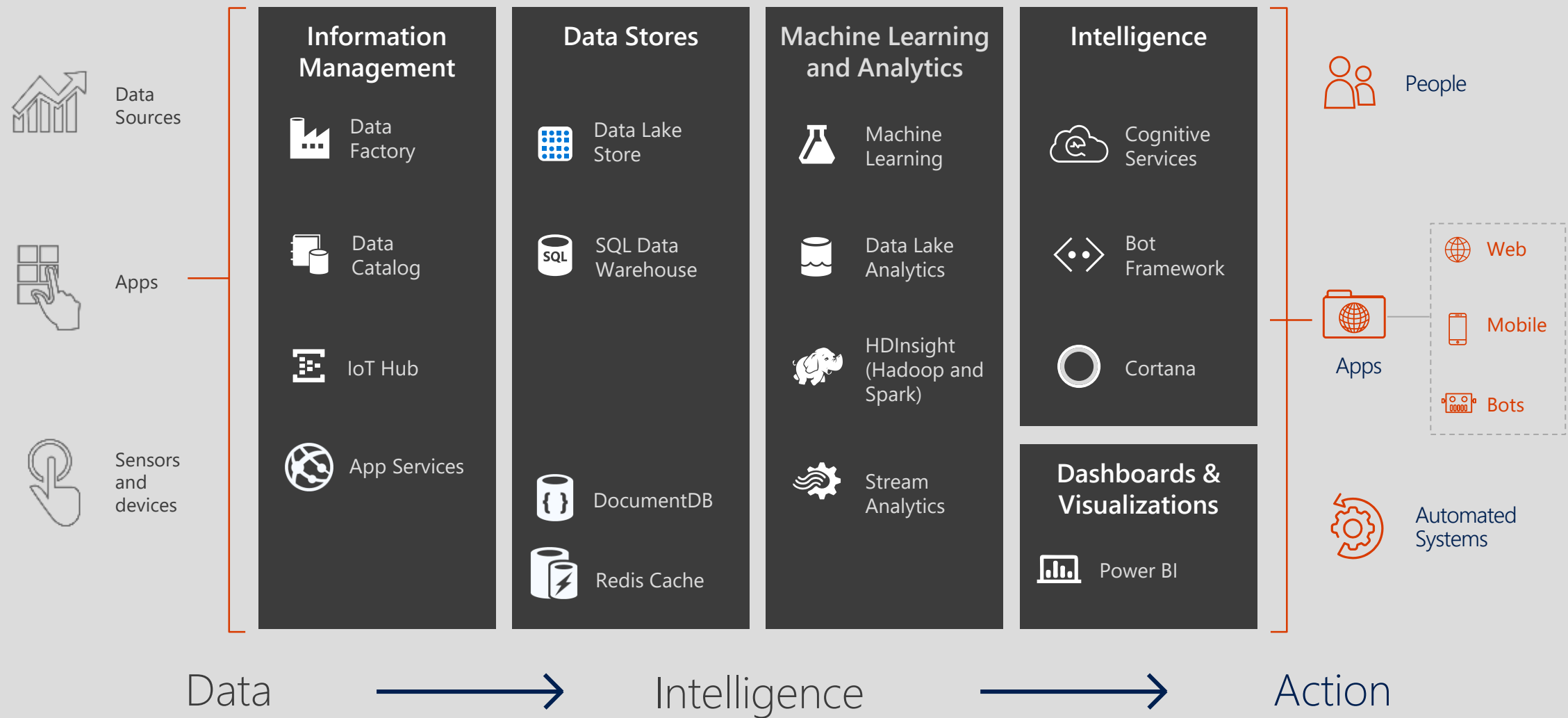
**SENTIMENT ANALYSIS: BACON + PRALINES**



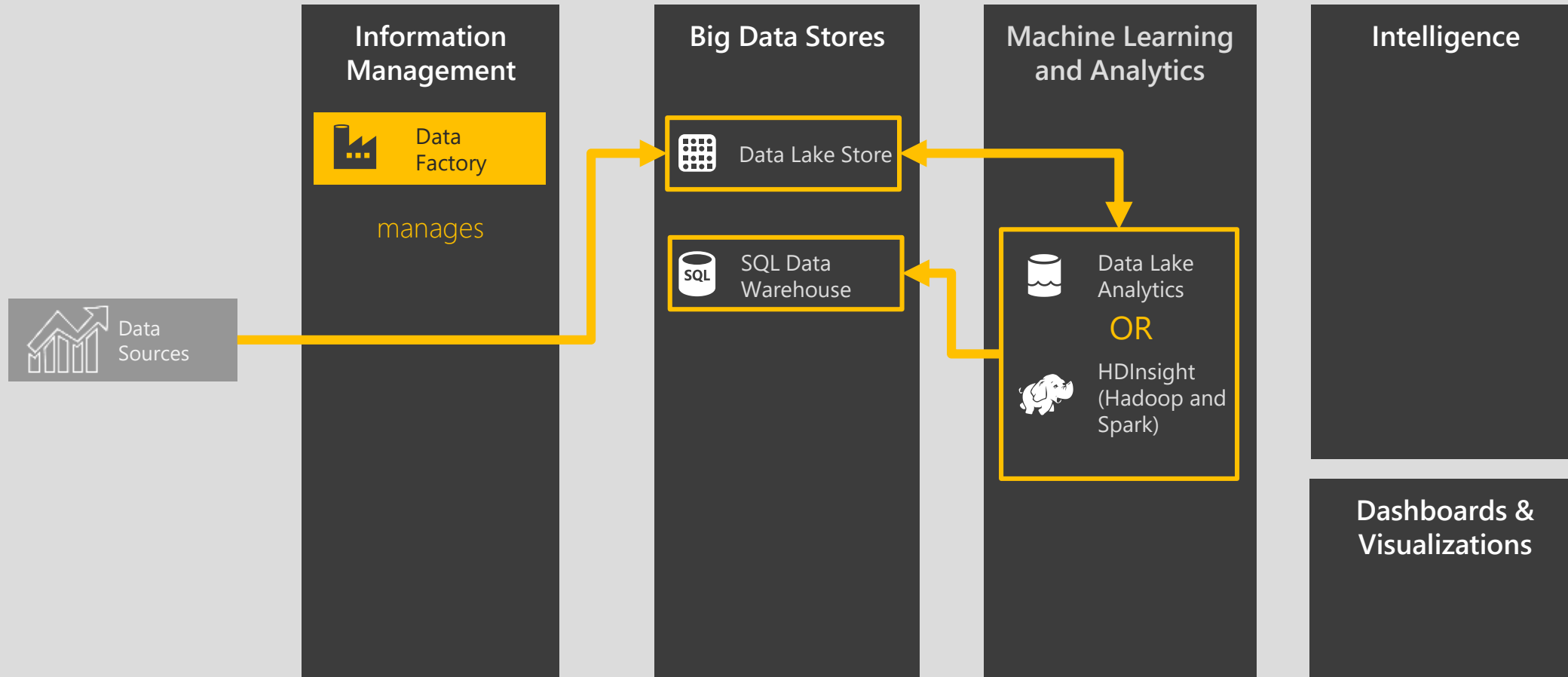
Transform data into  
intelligent action

David Gristwood

# Transform data into intelligent action

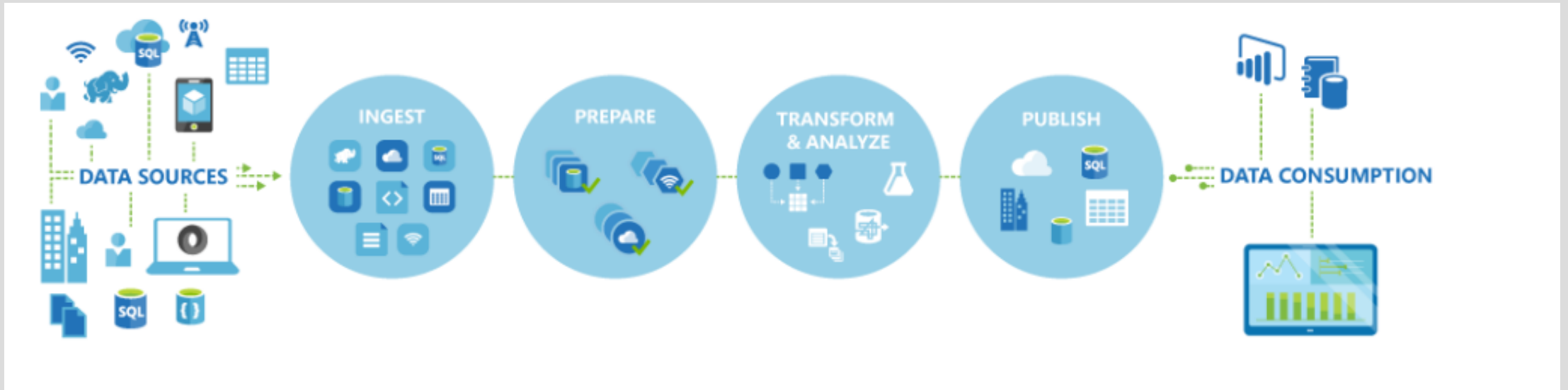


# Data at Scale



# Azure Data Factory

A managed cloud service for building and operating data pipelines

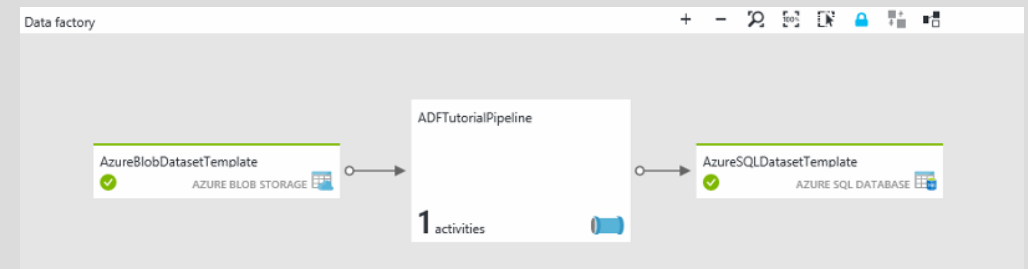


Create, schedule, orchestrate, manage data pipelines

Visualise data lineage

Monitor data pipeline health

Connect to on-premises and cloud data sources



# Data Factory Monitoring

Data factory

**RESOURCE EXPLORER**

- Data Factories
  - davidgriTutorialDataFactory
    - Pipelines
      - ADFTutorialPipeline
    - Datasets
      - AzureBlobDatasetTempl...
      - AzureSQLDatasetTempl...
    - Linked Services
      - AzureSqlLinkedService
      - AzureStorageLinkedSer...
    - Gateways

▶ || □

Start time (UTC): 03/10/2016 05:17 pm End time (UTC): 06/11/2016 04:17 pm Apply

Last edited by David Gristwood 11 m...

---

### ACTIVITY WINDOWS

↺ ↻ 🔍

No filter applied.

Last refreshed 2 minutes ago.

Pipeline ▼	Activity ▼	Window Start ↓ ...	Window End ▼	Status ▼	Type	Attempt Start	Attempt End	Duration	Retry Attempts
ADFTutorialPip...	CopyFromBlobT...	06/10/2016 2:00...	06/10/2016 3:00...	⌚ Waiting: Data	Copy	--	--	--	1
ADFTutorialPip...	CopyFromBlobT...	06/10/2016 1:00...	06/10/2016 2:00...	⌚ Waiting: Data	Copy	--	--	--	1
ADFTutorialPip...	CopyFromBlobT...	06/10/2016 12:0...	06/10/2016 1:00...	⌚ Waiting: Data	Copy	--	--	--	1
ADFTutorialPip...	CopyFromBlobT...	06/10/2016 11:0...	06/10/2016 12:0...	⌚ Waiting: Data	Copy	--	--	--	1
ADFTutorialPip...	CopyFromBlobT...	06/10/2016 10:0...	06/10/2016 11:0...	⌚ Waiting: Data	Copy	--	--	--	1

Data Factory Name  
**davidgriotutorialdatafactory**

Duration  
--

Input Dataset Ids 0  
2a2da873-bb4d-4b36-a0a9-9e1a7ab6...

Input Datasets 0  
AzureBlobDatasetTemplate

Output Dataset Ids 0  
6283d5c9-b69e-48cc-a41c-bef16f537...

Output Datasets 0  
AzureSQLDatasetTemplate

Percent Complete  
--

Pipeline Name  
**ADFTutorialPipeline**

Resource Group Name  
**adftutorialresourcegroup**

Run Attempts  
**1**

Attempt End  
--

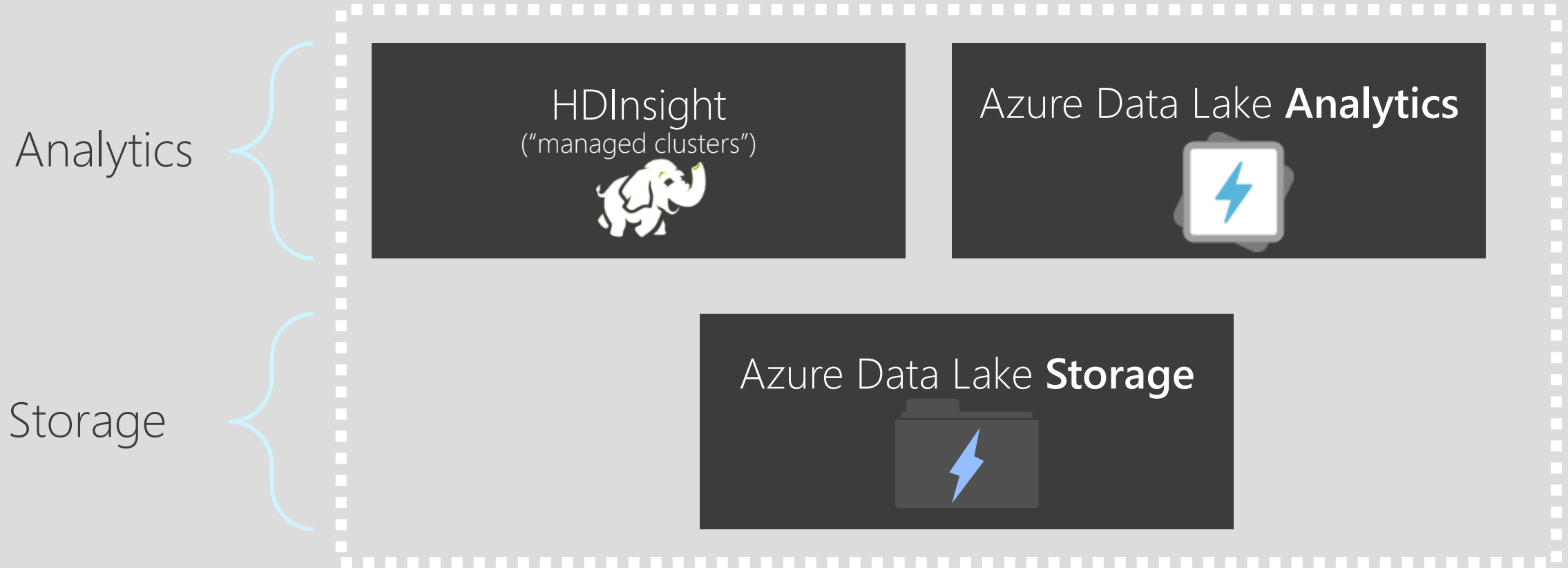
Attempt Start  
--

State Description  
The upstream dependencies are not r...

Window State  
**Waiting**

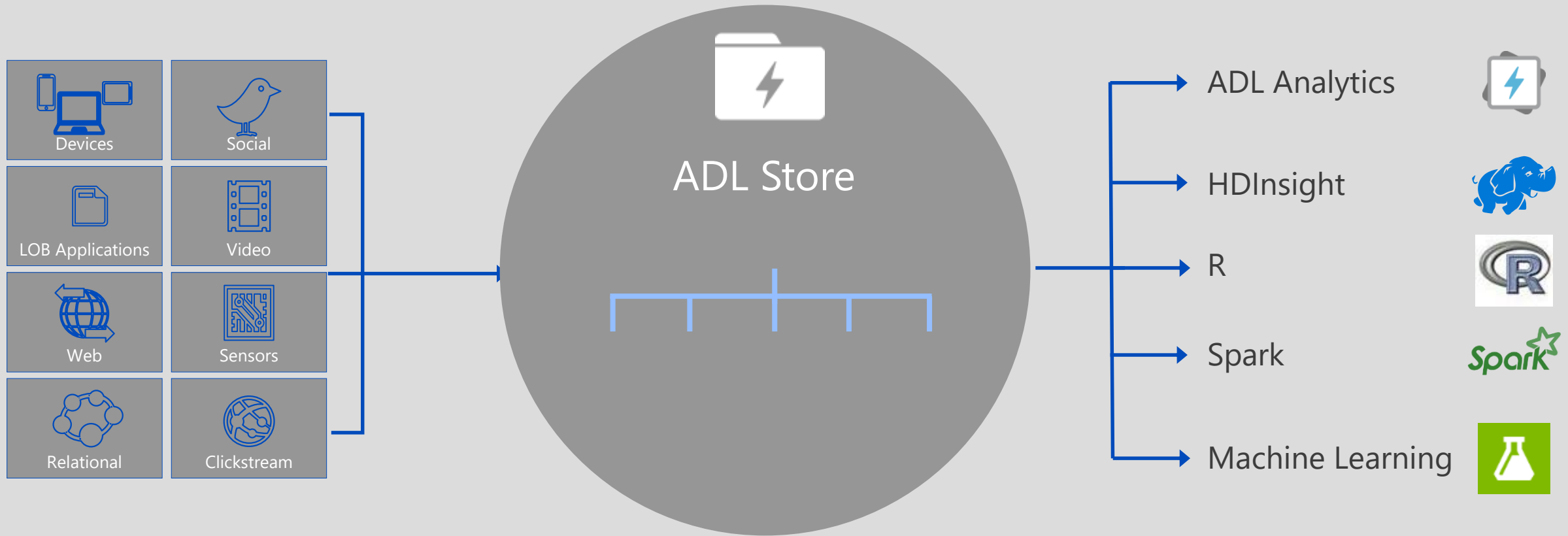
Window Substate  
Dataset dependencies

# Azure Data Lake



# Azure Data Lake Store

A highly scalable, distributed, parallel file system in the cloud  
Designed to work with multiple analytic frameworks



# Azure Data Lake Analytics Service

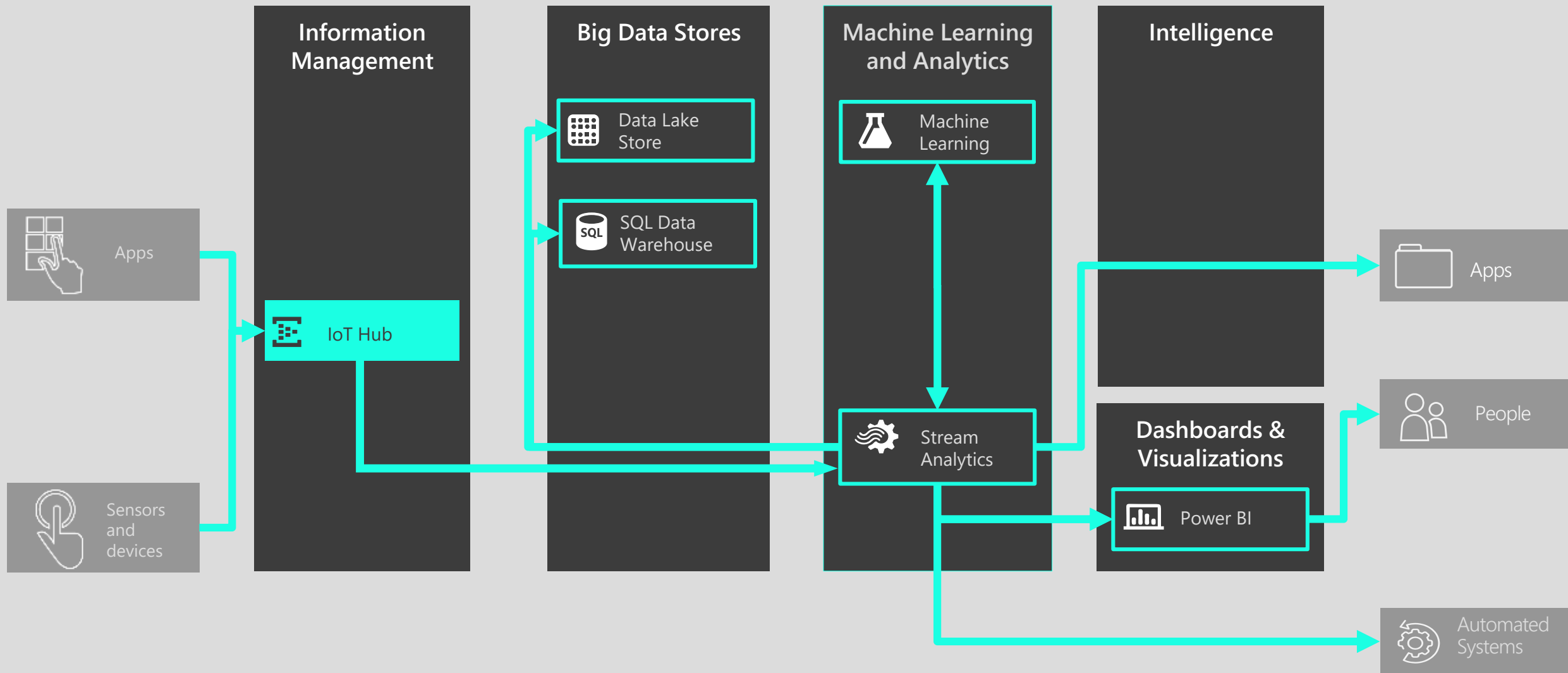
A new distributed  
analytics service



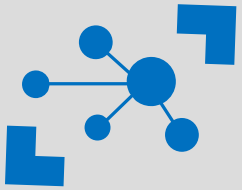
- Built on Apache YARN
- Scales dynamically with the turn of a dial
- Pay by the query
- Supports Azure AD for access control, roles, and integration with on-premise identity systems
- Built with U-SQL to unify the benefits of SQL with the power of C#



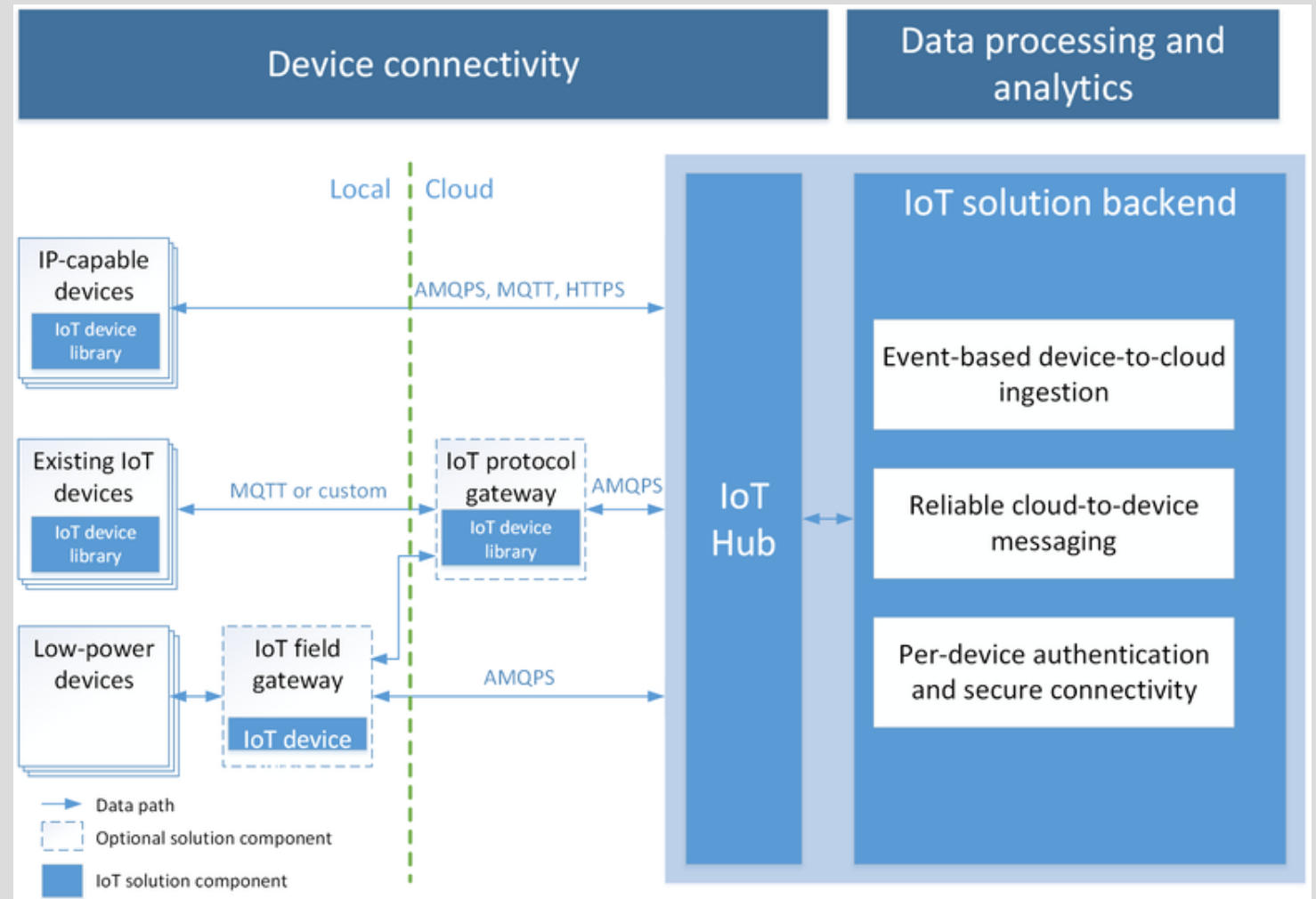
# Data at Speed



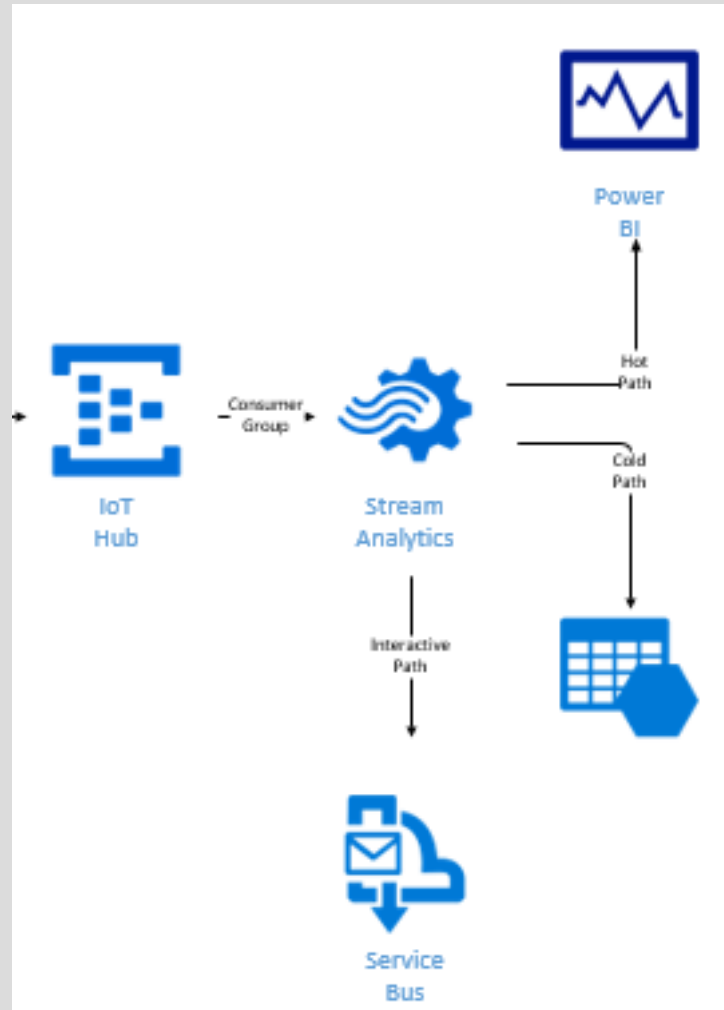
# Azure IoT Hub



- Telemetry ingestion
- Millions of devices
- Secure bi-directional communication
- AMQP, HTTP, MQTT, custom protocols
- Multiple platforms device SDKs



# Azure Stream Analytics



Stream millions events per second

Perform time based work

Mission critical reliability

SQL-like language

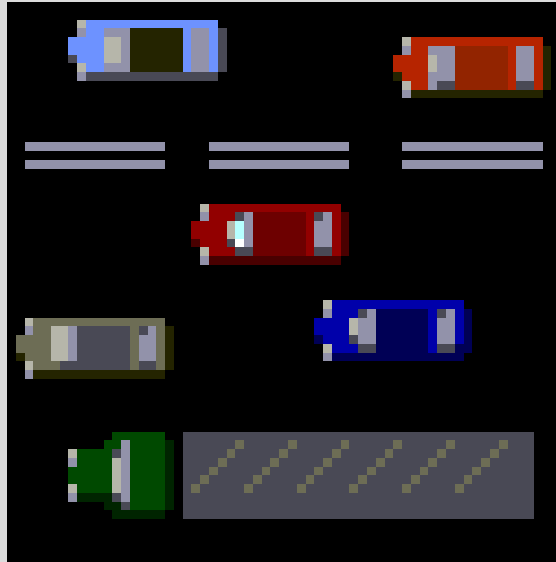
Create real time alerts

# Azure Streaming Analytics

Data at Rest



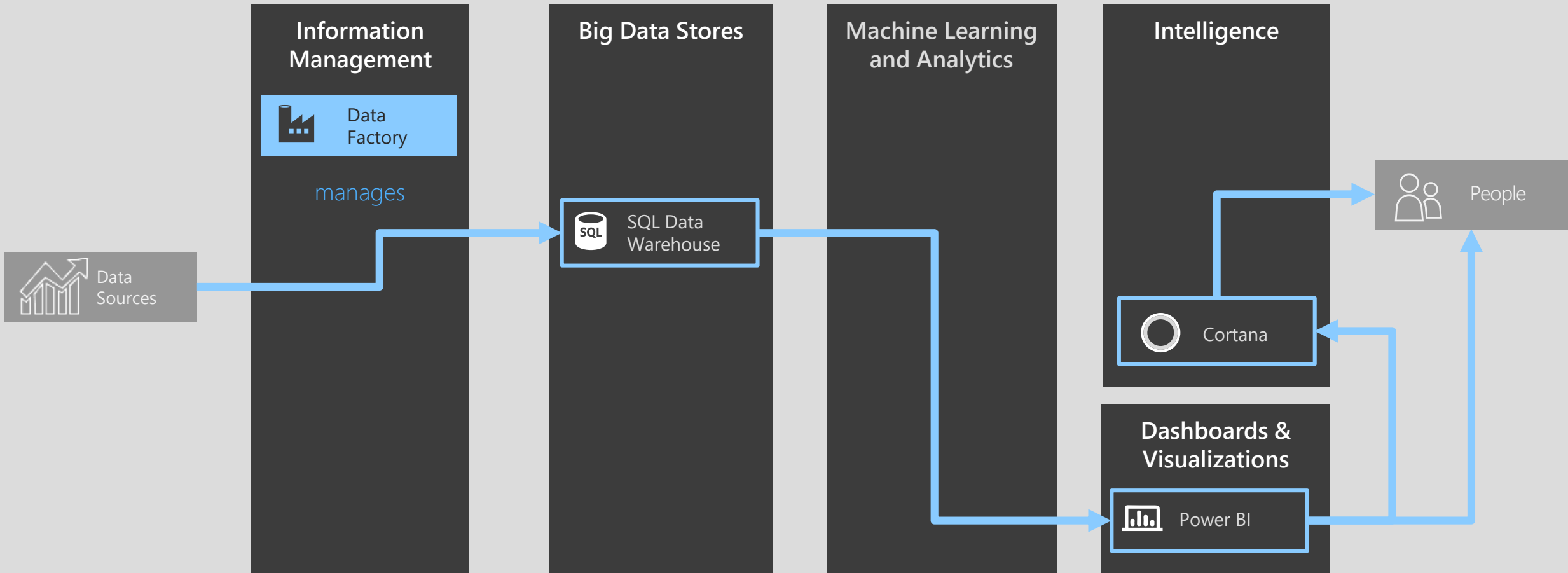
Data in Motion



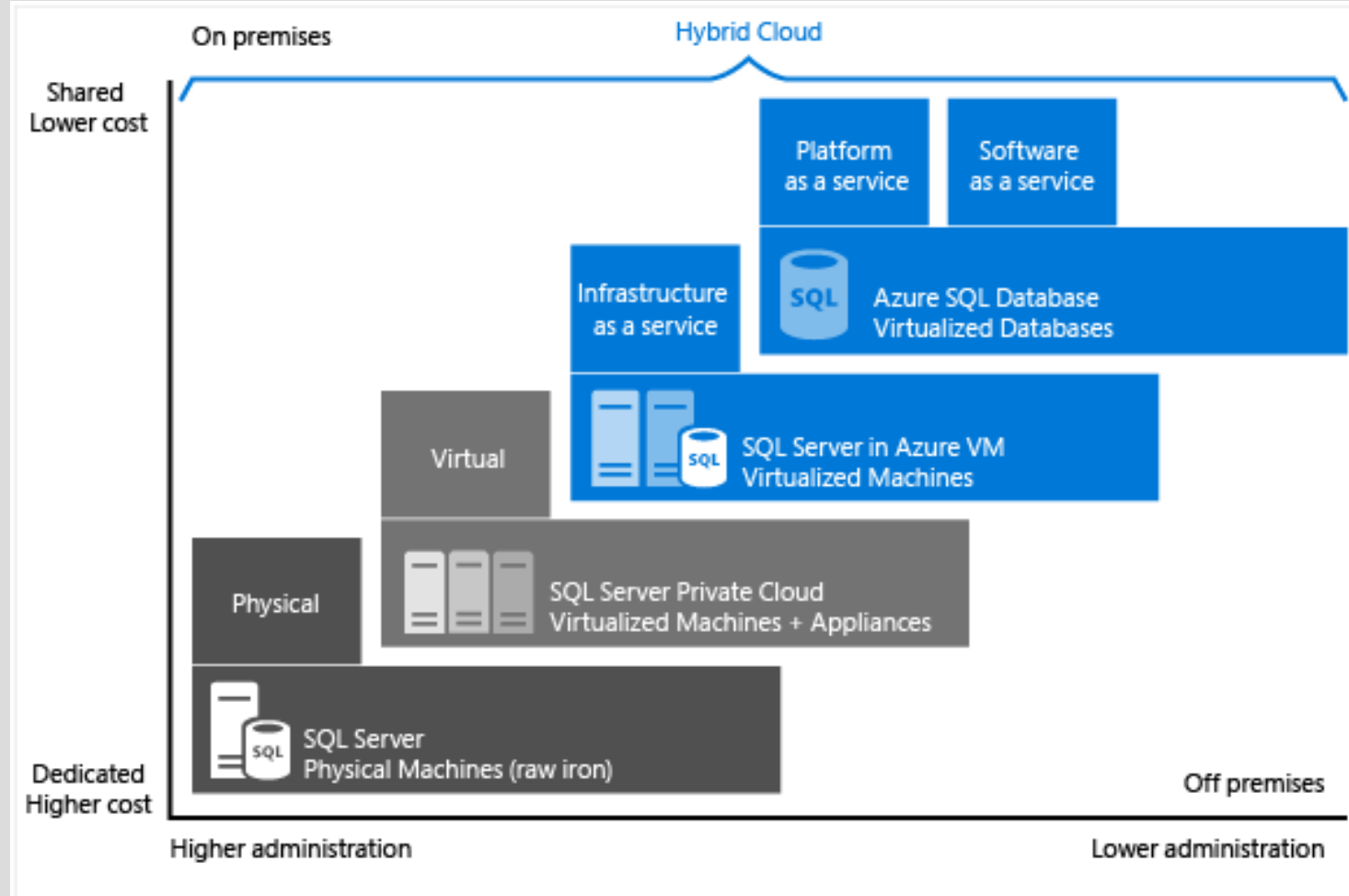
*"Every 5 seconds give me  
the count of tweets over  
the last 10 seconds"*

```
SELECT Topic, Count(*) AS  
TotalTweets  
FROM TwitterStream  
TIMESTAMP BY CreatedAt  
GROUP BY Topic,  
HoppingWindow(second, 10, 5)
```

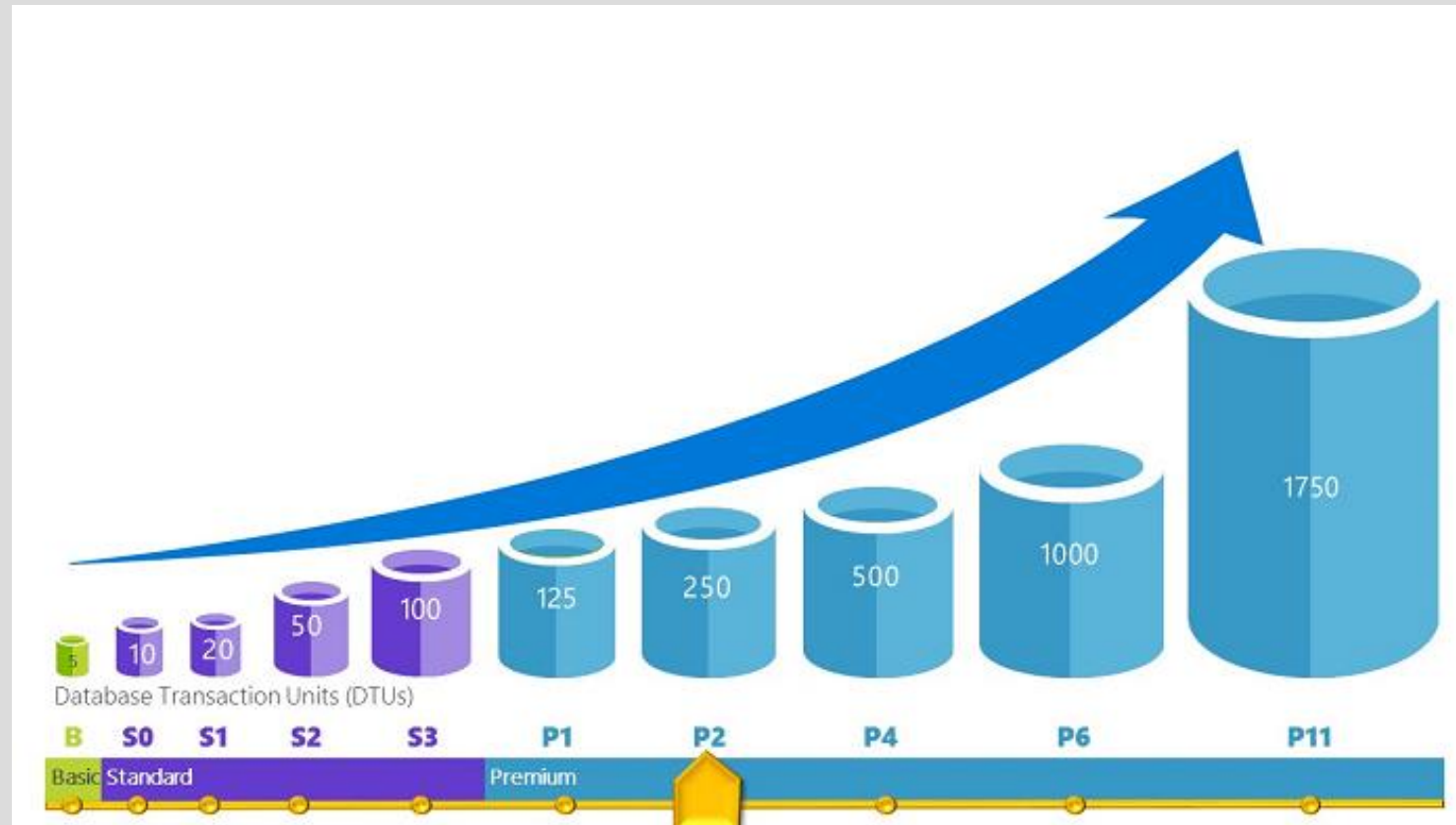
# Traditional BI



# SQL Database (aka 'SQL Azure')



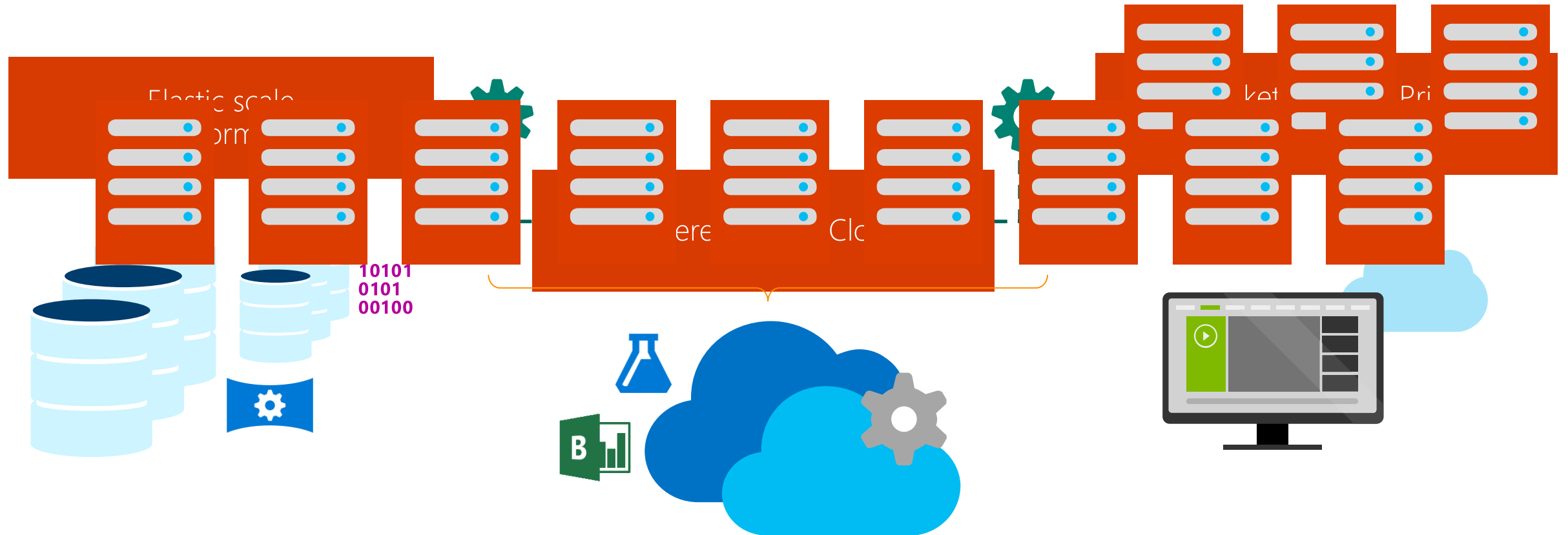
# SQL Database



- Database Transaction Unit (DTU) is unit of measure represents relative power databases
- Provides increasing set of resources ('power') to offer increasingly higher throughput
- Rating measured using the Azure SQL Database Benchmark

# Elastic Scale

- Spin up for heavy workloads, cycle down for daily activity
- Buy time to insight based on what you need, when you need it
- Choose the combo of compute and storage that meets your needs





# Microsoft Power BI

## Data sources

- SaaS solutions  
e.g. Marketo, Salesforce, GitHub, Google analytics
- On-premises data  
e.g. Analysis Services
- Organizational 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

Content packs

Live dashboards

Visualizations



Reports

01001  
10101

tassets

Data refresh

Natural language query

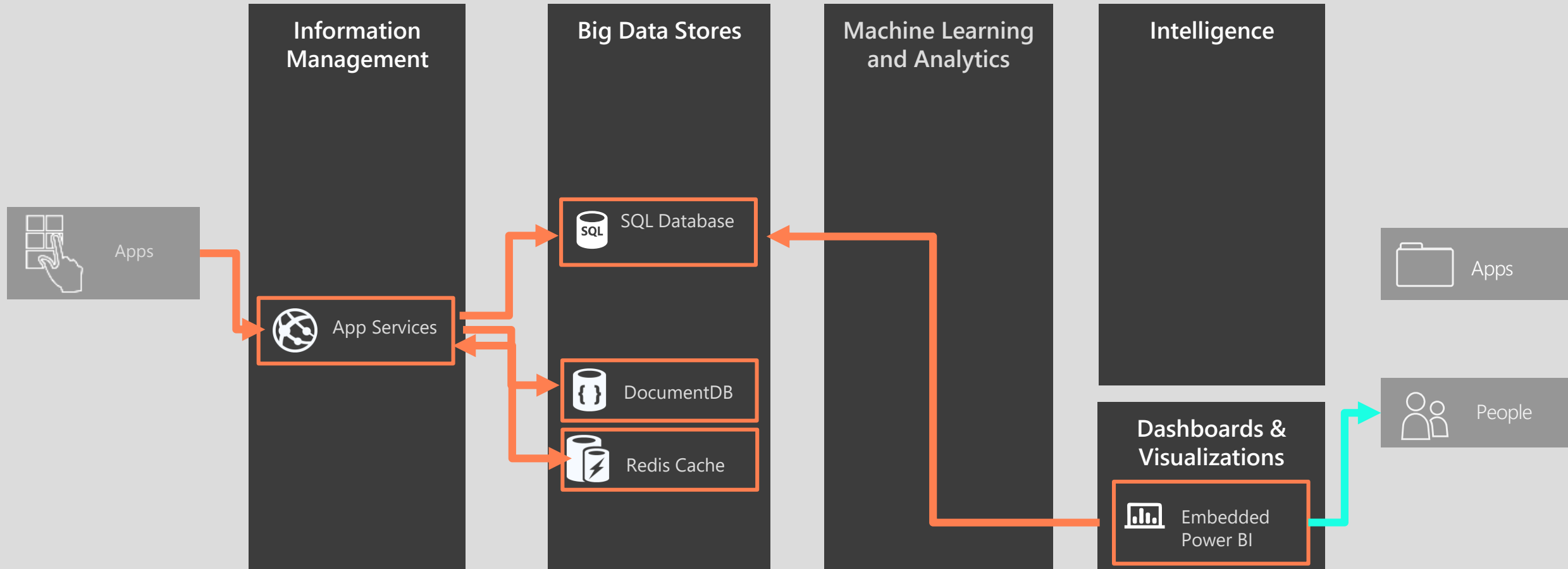
Sharing & collaboration



*Create beautiful visualizations to tell compelling data stories*

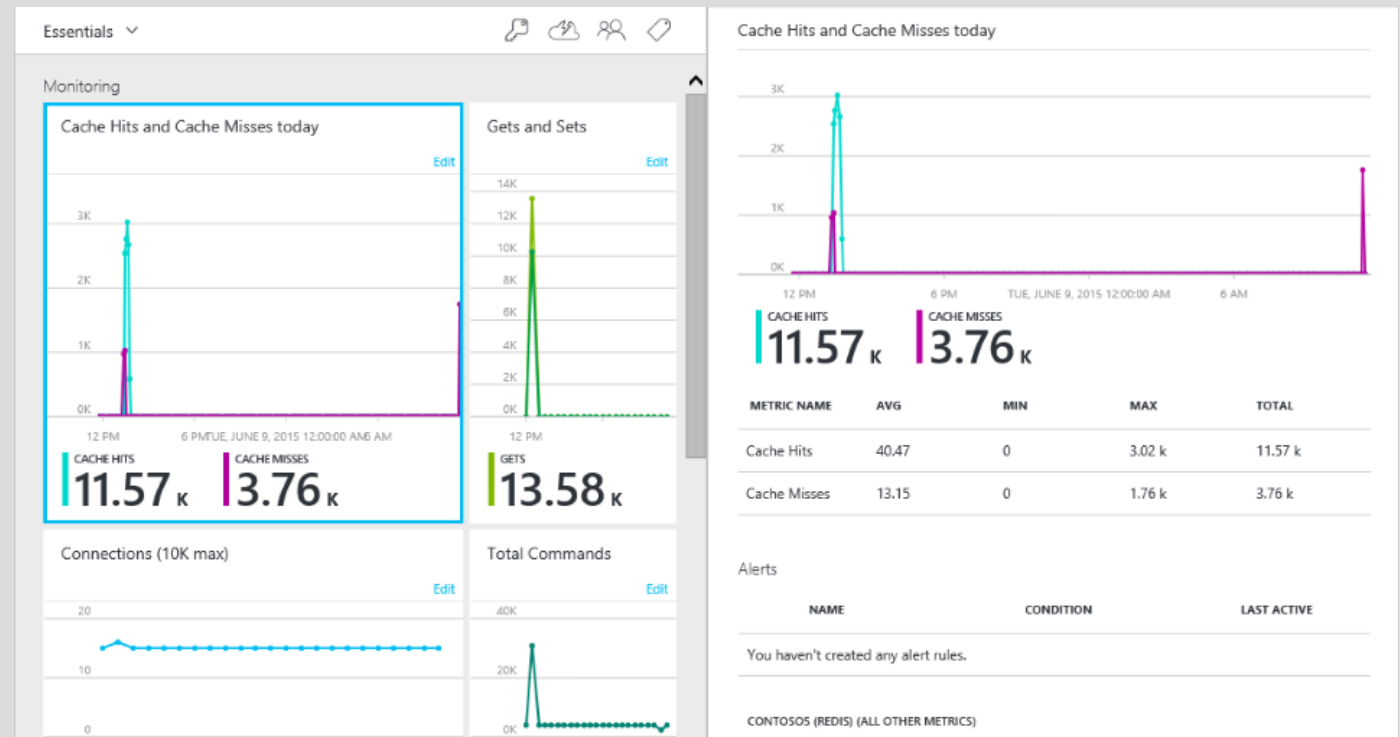
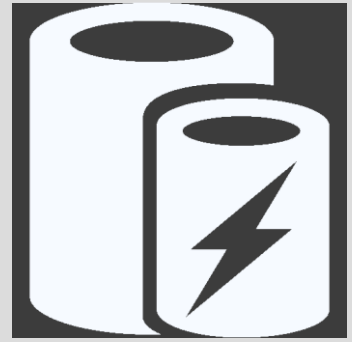
*Build rich, live dashboards that turn business intelligence into real insights*

# Business Applications



# Redis Cache

- Dedicated, managed Redis cache
- Advanced key-value store
- High-availability offering



# Azure DocumentDB



- NoSQL as a service
- Highly available, globally distributed
- Benefits of SQL and JavaScript, schema-free
- Guaranteed low latency write-optimised, latch-free database engine
- Protocol support for MongoDB

# Azure Search



- Fully managed service search service
- Based on Elasticsearch engine
- Avoid hassle of index corruption, service availability, scaling and service updates
- Load and update with integrated indexers
- Control over search rankings with facets, weightings
- Integration with Microsoft's Natural Language Processors