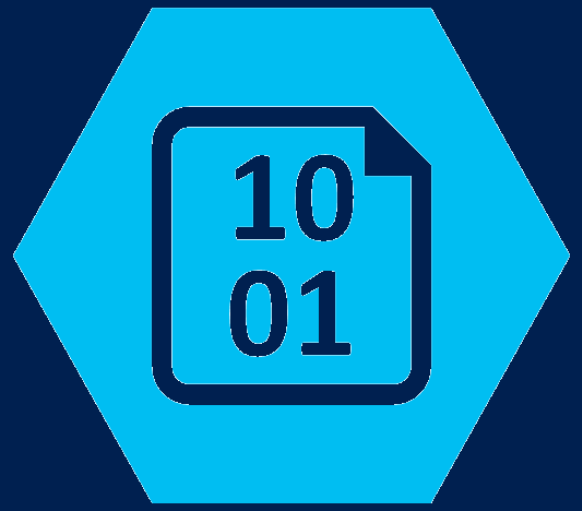
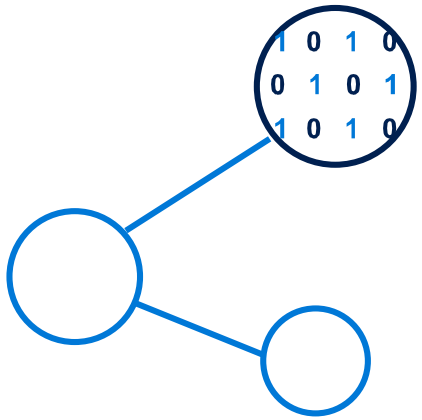
Data Science as a Service & Microsoft's Tools for Big Data

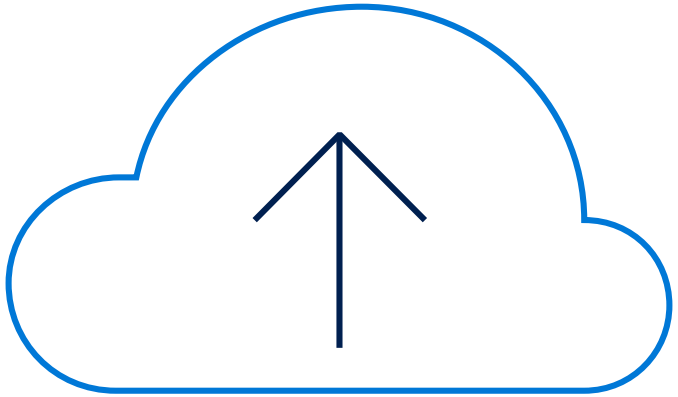


Alyssa Ong
Technical Evangelist
@alyssaong1337

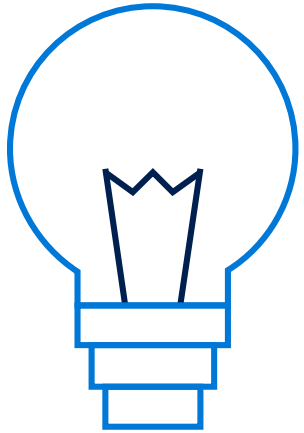
Businesses are being transformed by three trends



Big Data

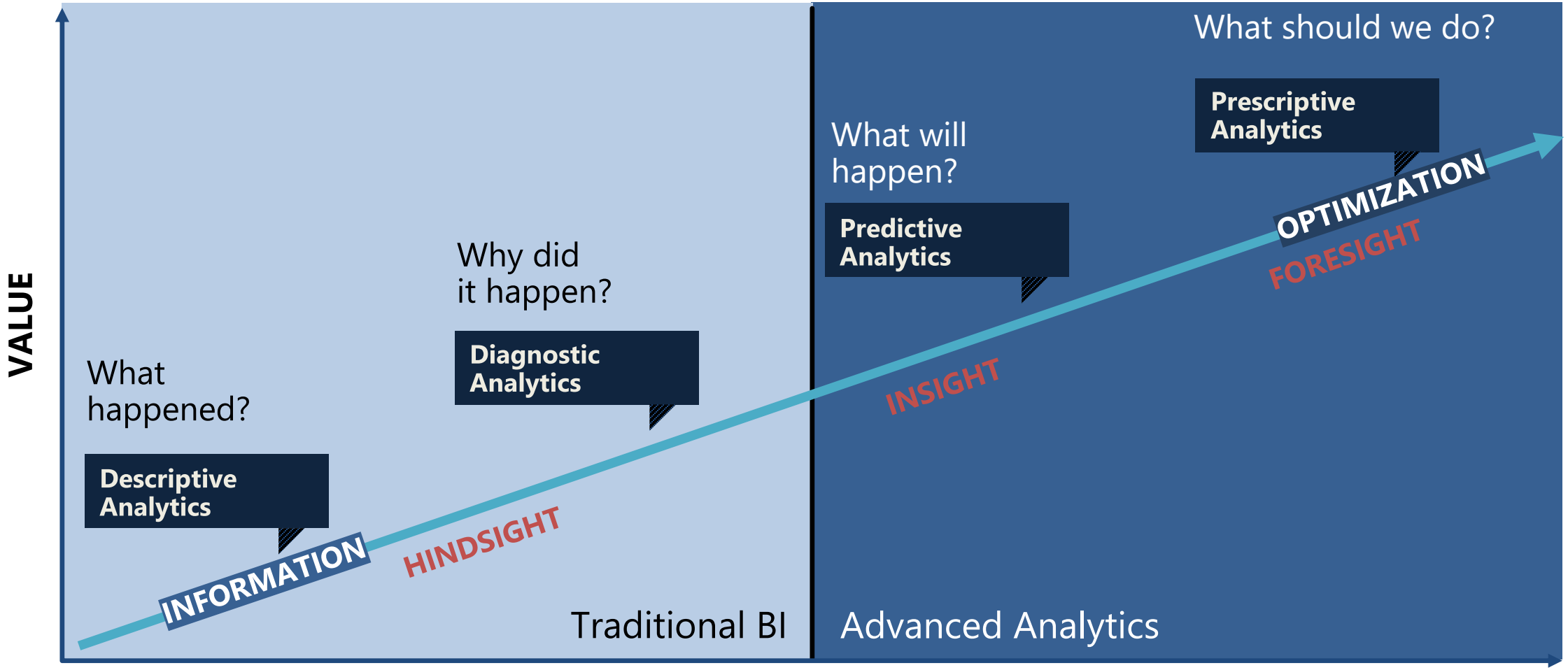


Cloud

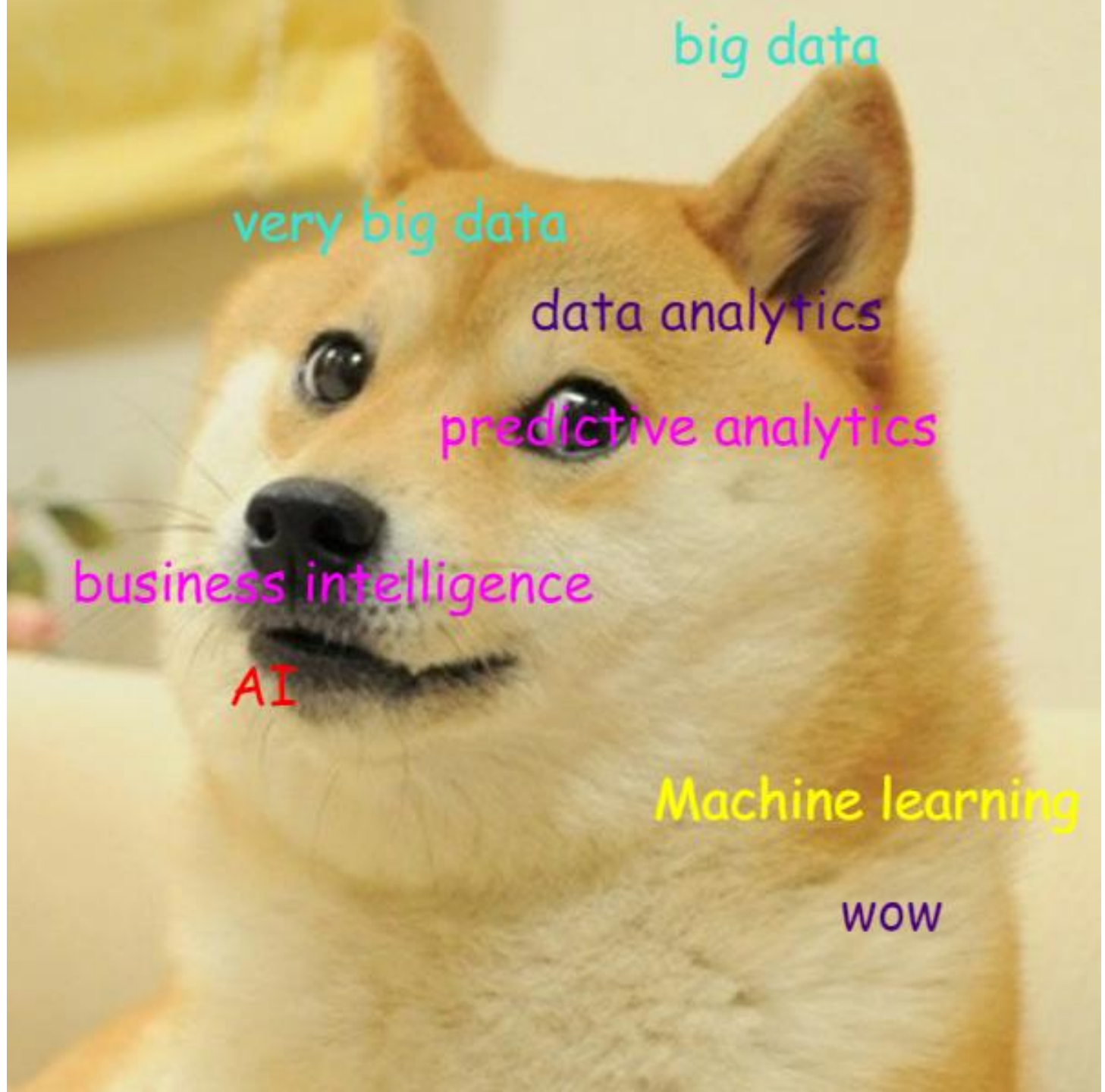


Intelligence

Going beyond traditional business intelligence



Source: Gartner

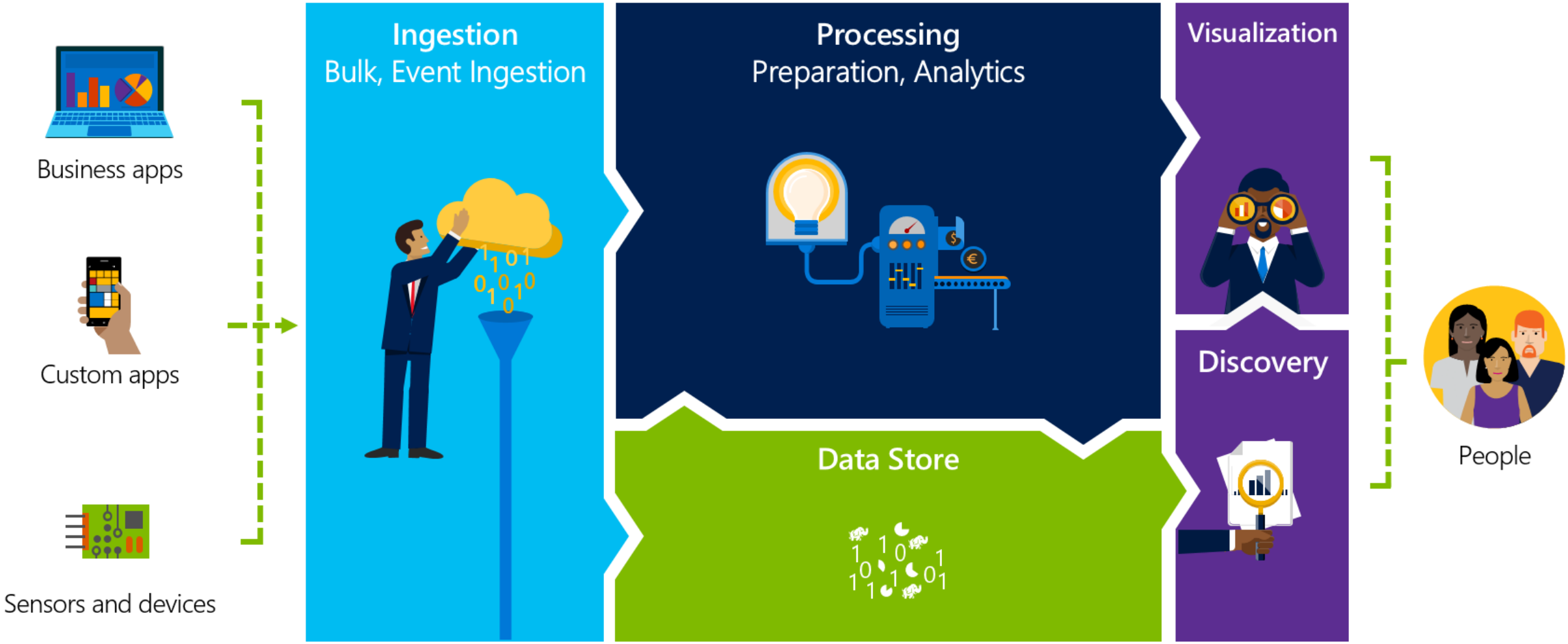

big data
very big data
data analytics
predictive analytics
business intelligence
AI
Machine learning
wow

What exactly is big data???

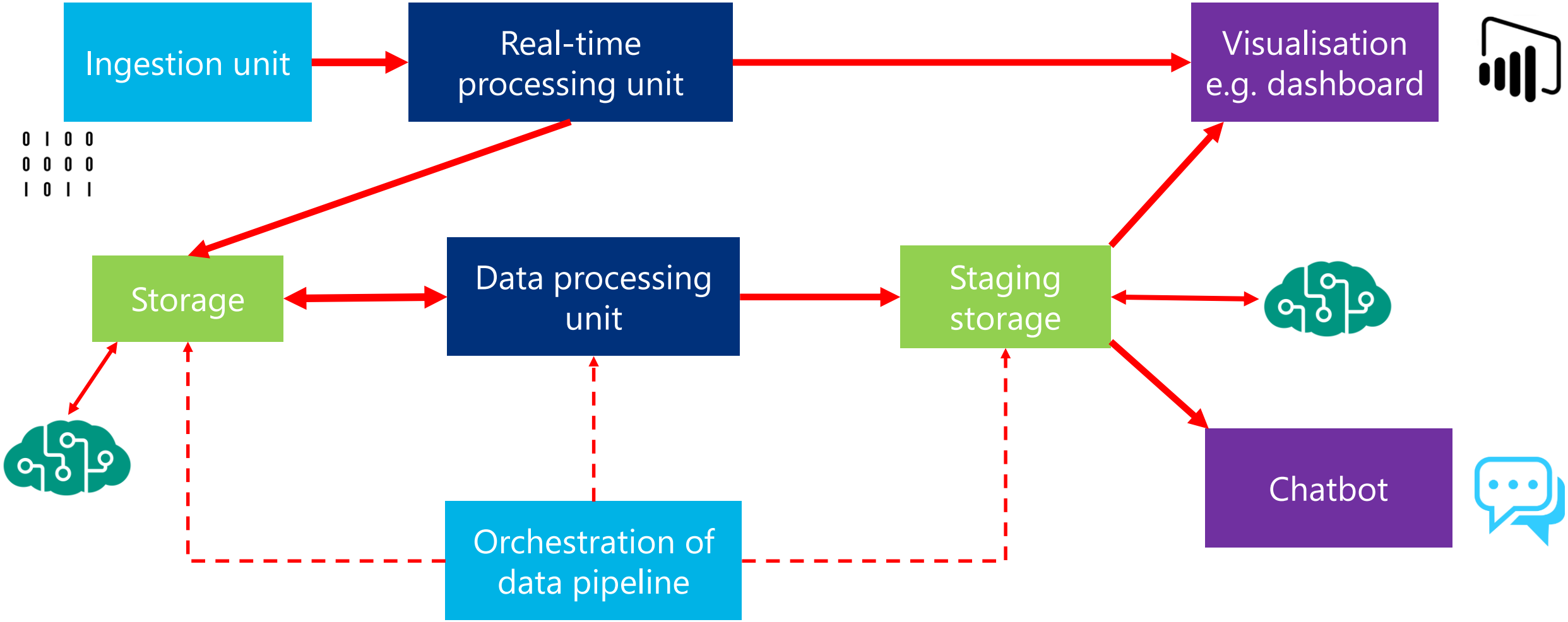
Let's break it down a little.

A TON of data you need to process, analyse and get intelligence from.

Big Data Analytics flow



An example of a big data analytics process…

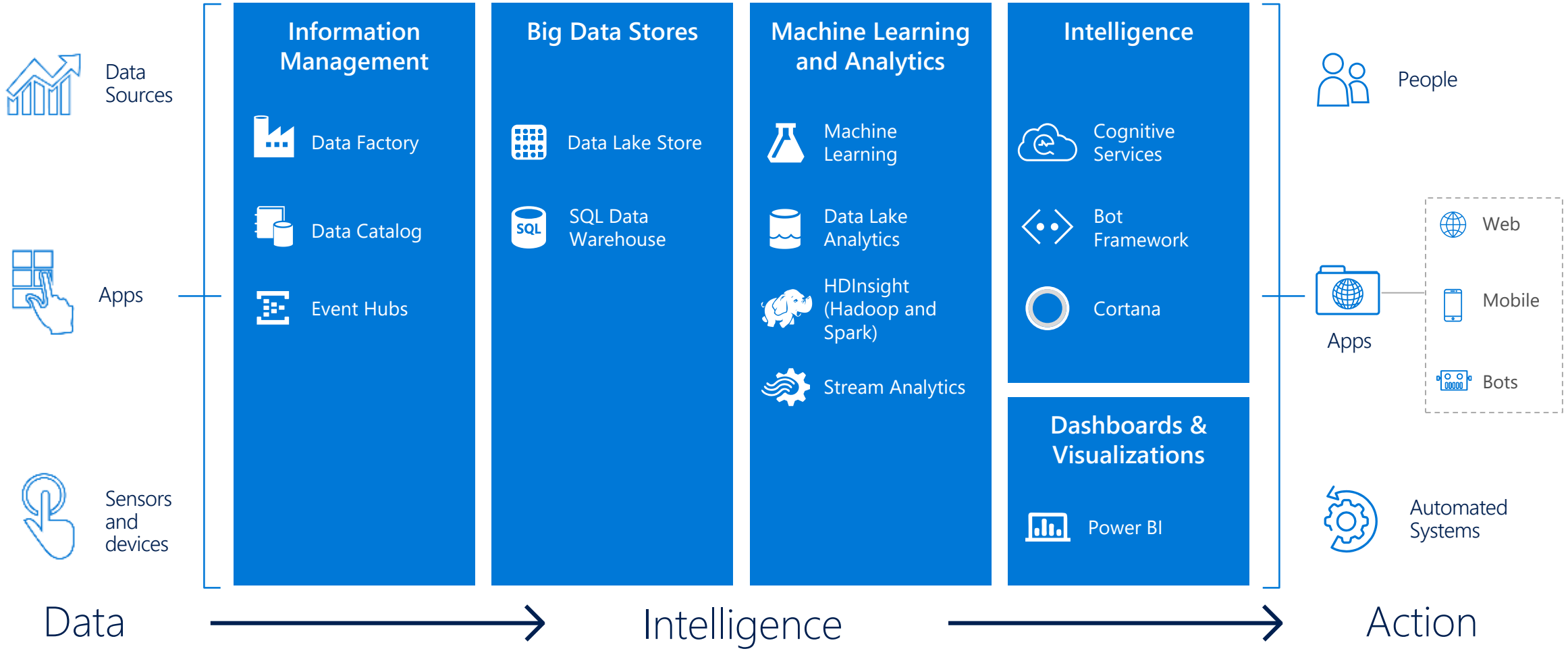


Cortana Intelligence Suite = Microsoft's Big Data toolkit

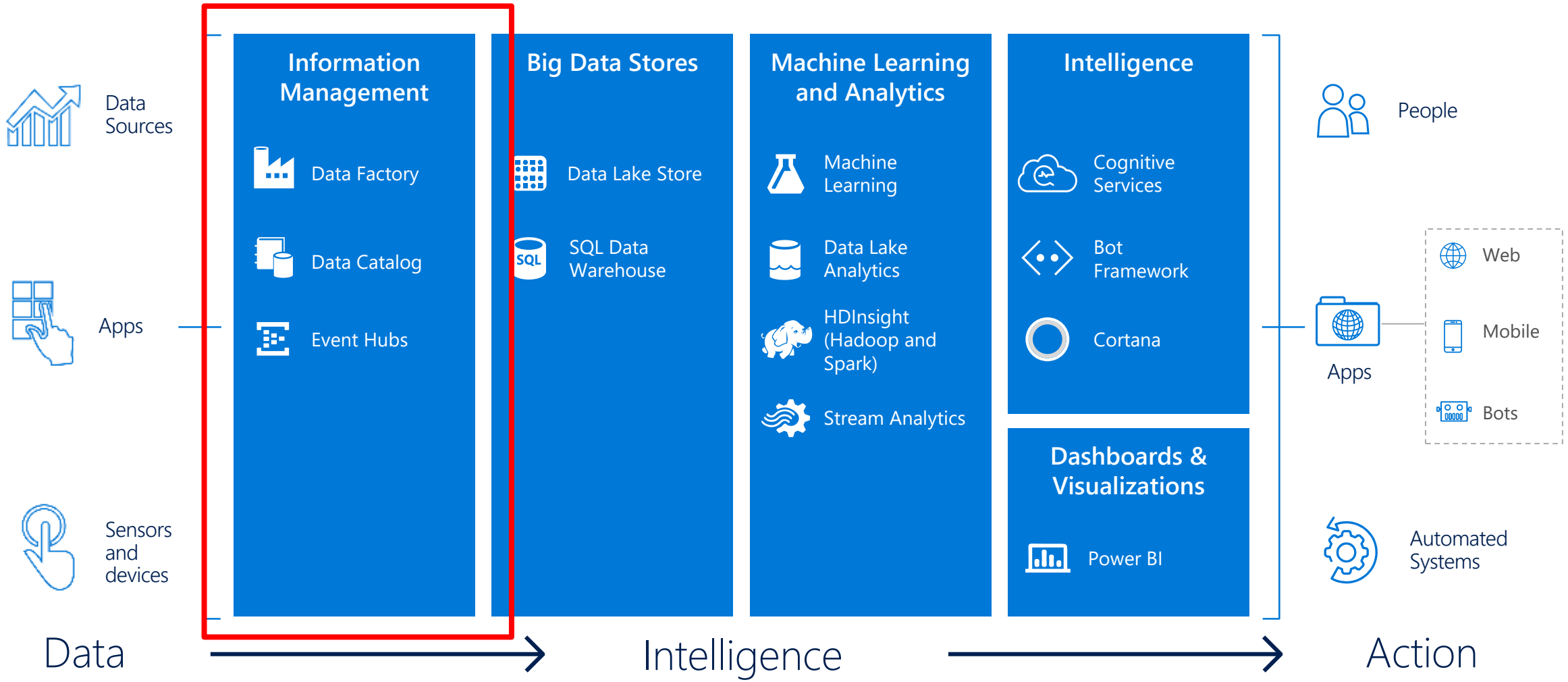
Stay ahead of the curve with Cortana Intelligence Suite

Data → Intelligence → Action

Easily turn data into intelligent action



Easily turn data into intelligent action



Ingestion and Information Management

Ingest

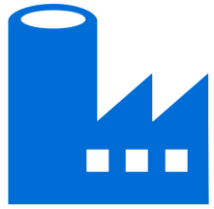
Event Hub

- High throughput data ingestion – up to 1 million/s
- Great for IoT and streaming

Manage

Data Catalog

- Extract metadata from data assets or add tags
- Makes data searchable and filterable

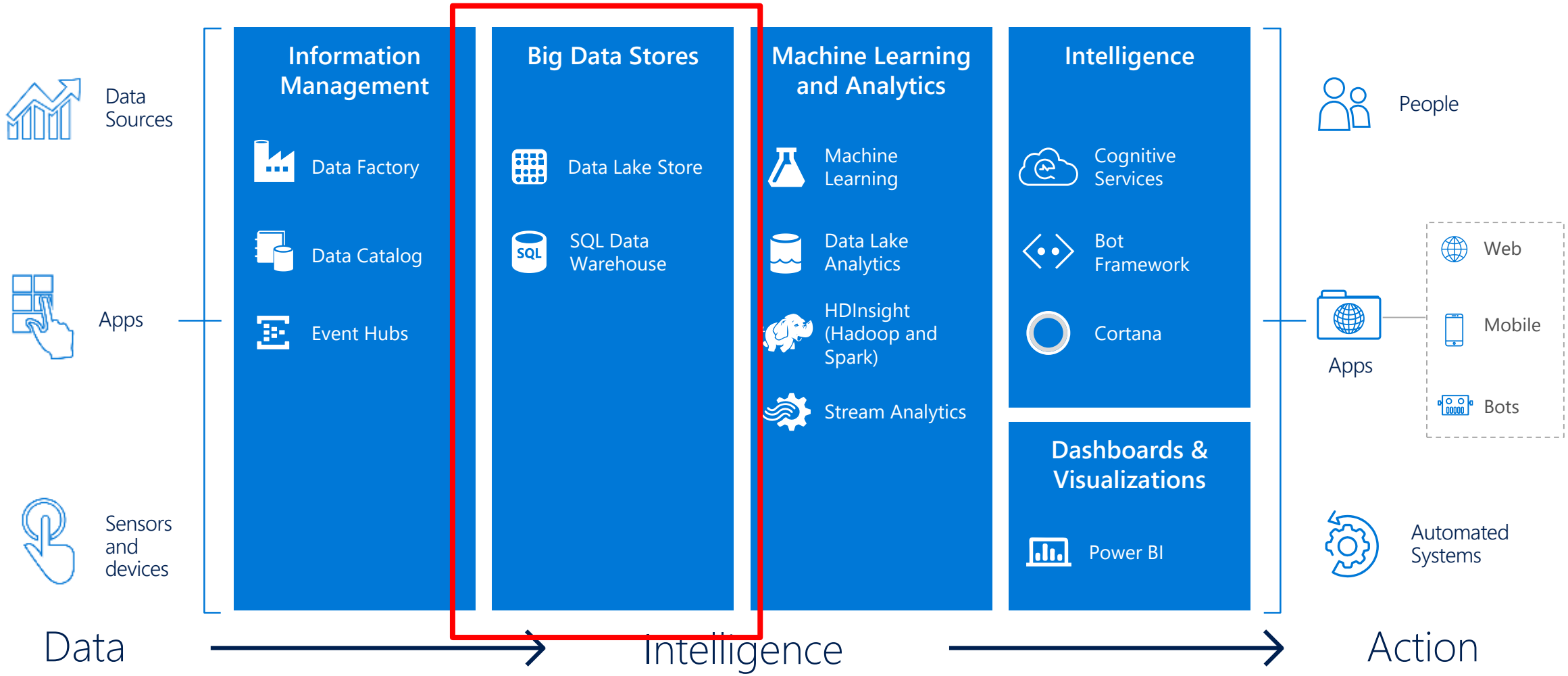


Orchestrate

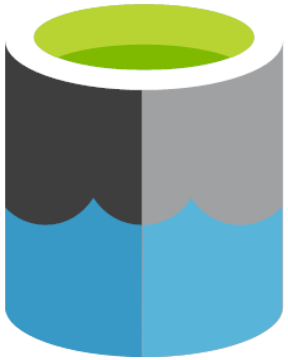
Data Factory

- Orchestrates data pipelines
- Does NOT do any processing
- Scheduling activities
- Pull data from on-premise systems

Easily turn data into intelligent action



Big Data Stores



Azure Data Lake

- NoSQL
- Optimized for big data e.g. log files, IoT data
- Works with Hadoop and U-SQL

Azure SQL Warehouse

- Relational
- Optimized for big data e.g. log files, IoT data
- Works with Hadoop and U-SQL

Normal Data Stores

Azure SQL database

- Relational db
- Good when you need integrity and consistency

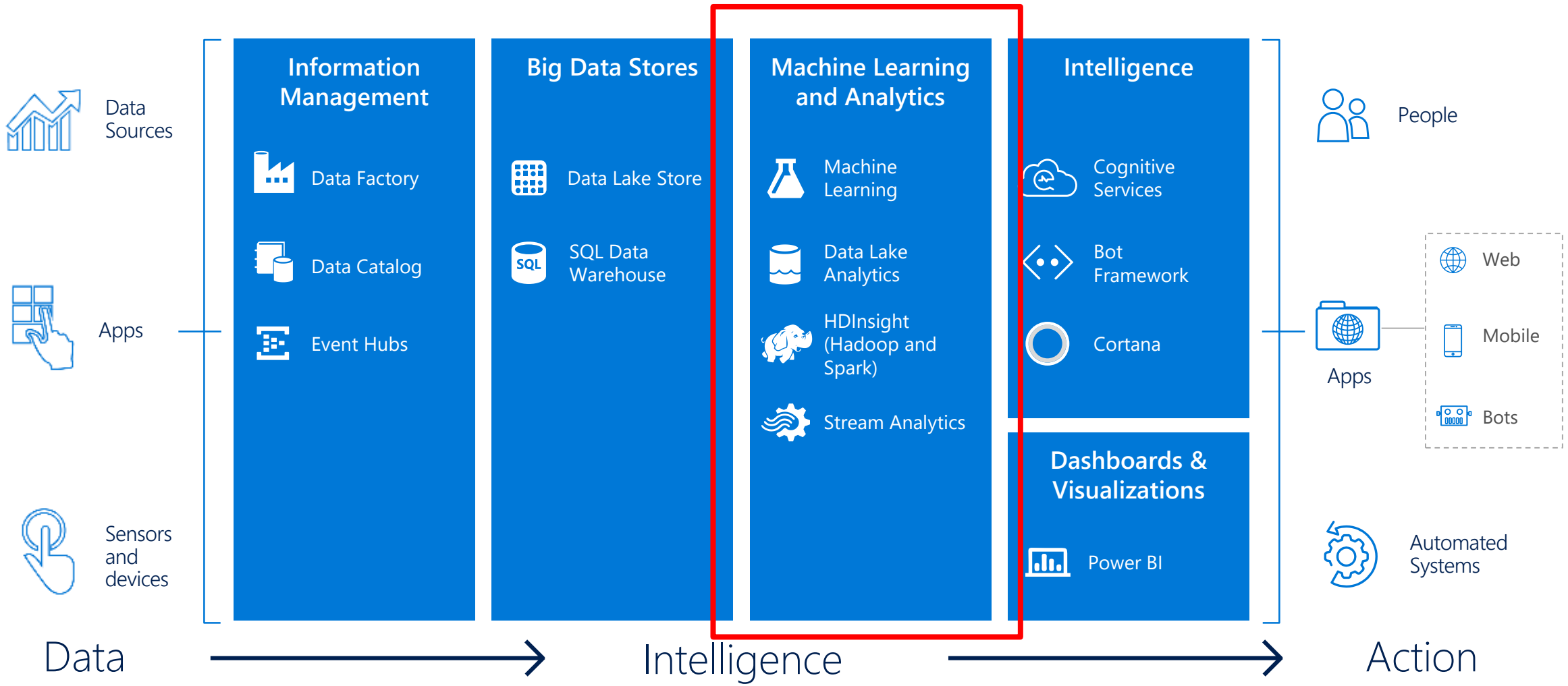
DocumentDb

- NoSQL db
- Great for ad-hoc queries

Storage Blob/Table

- NoSQL db
- No ad-hoc querying
- Cheaper than Documentdb

Easily turn data into intelligent action

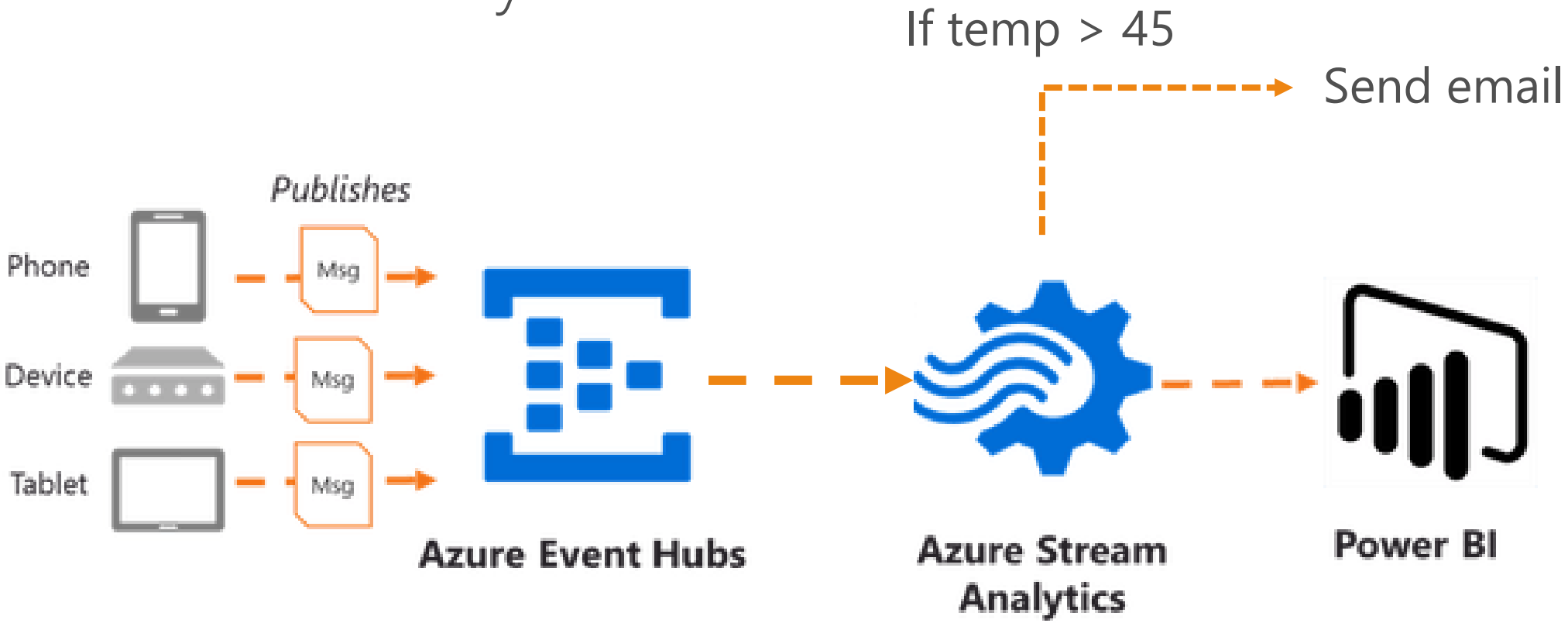


Real time data processing

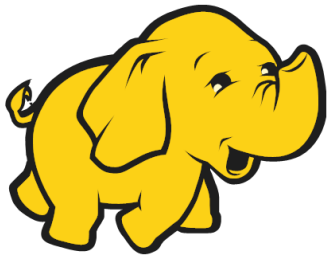
Azure Stream Analytics

- IoT data, web logs
- Used for notifications and alerting
- Real time scenarios

Visualise real-time analytics



Big Data Processing



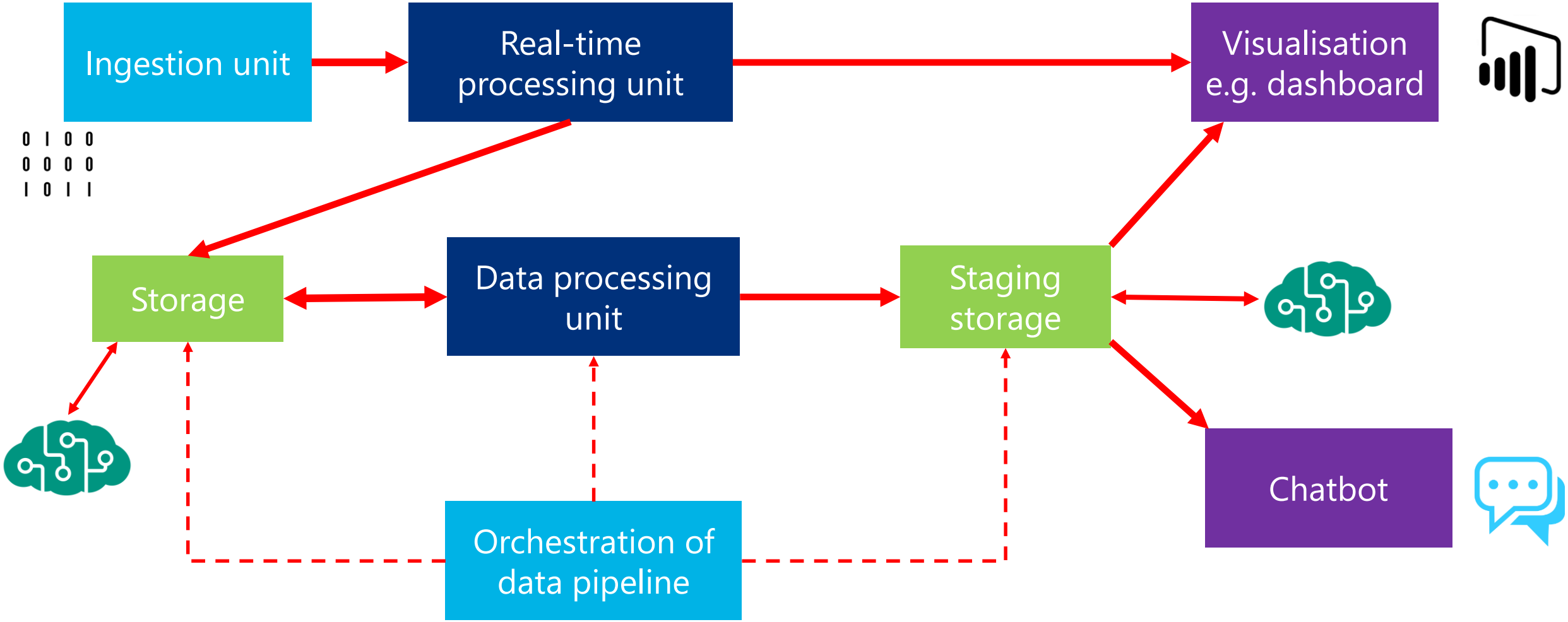
HDInsight

- Perform queries to process and transform your big data
- Supports open source services (Spark, Hadoop, etc.)

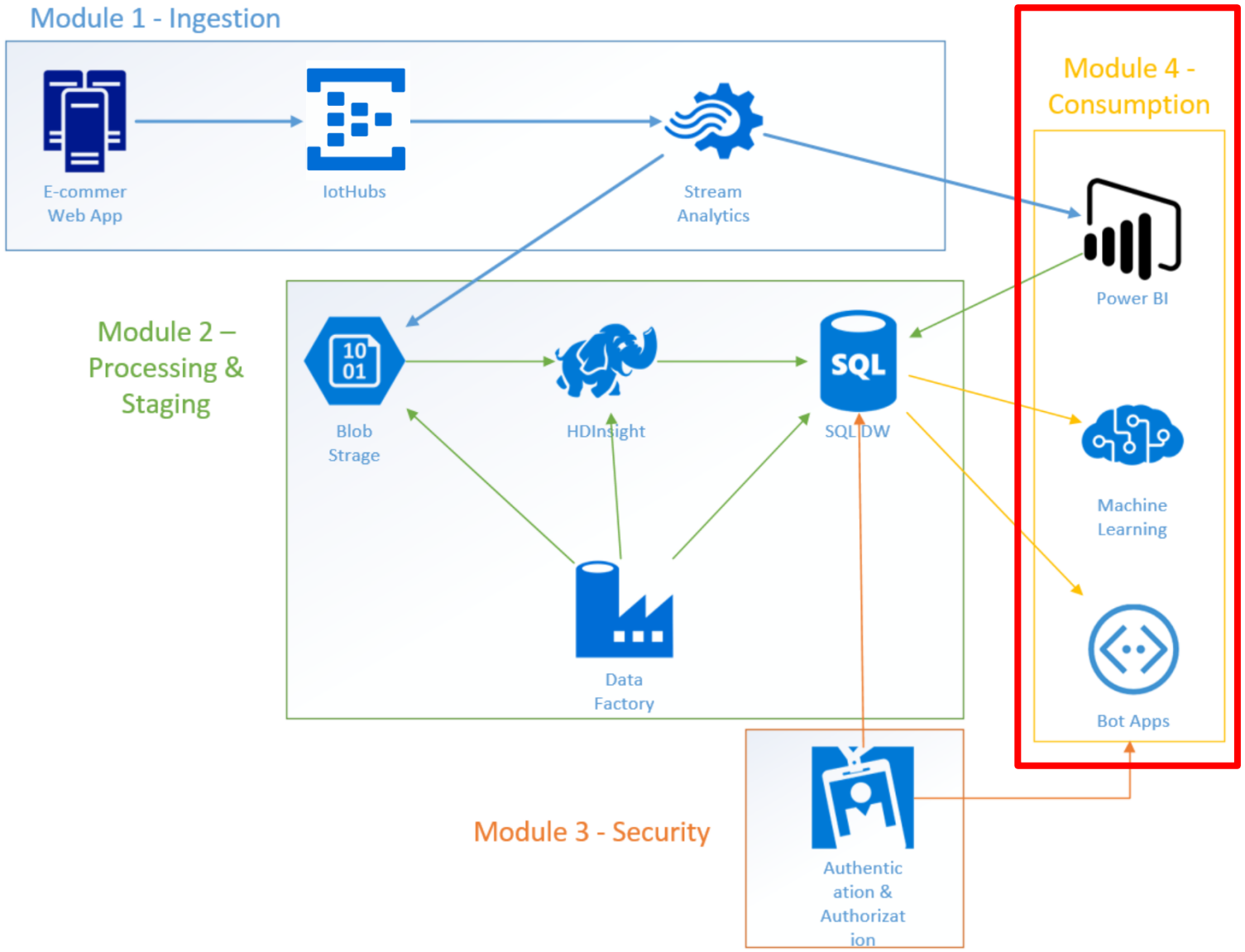
Data Lake Analytics

- Uses U-SQL (mix of C# and SQL)

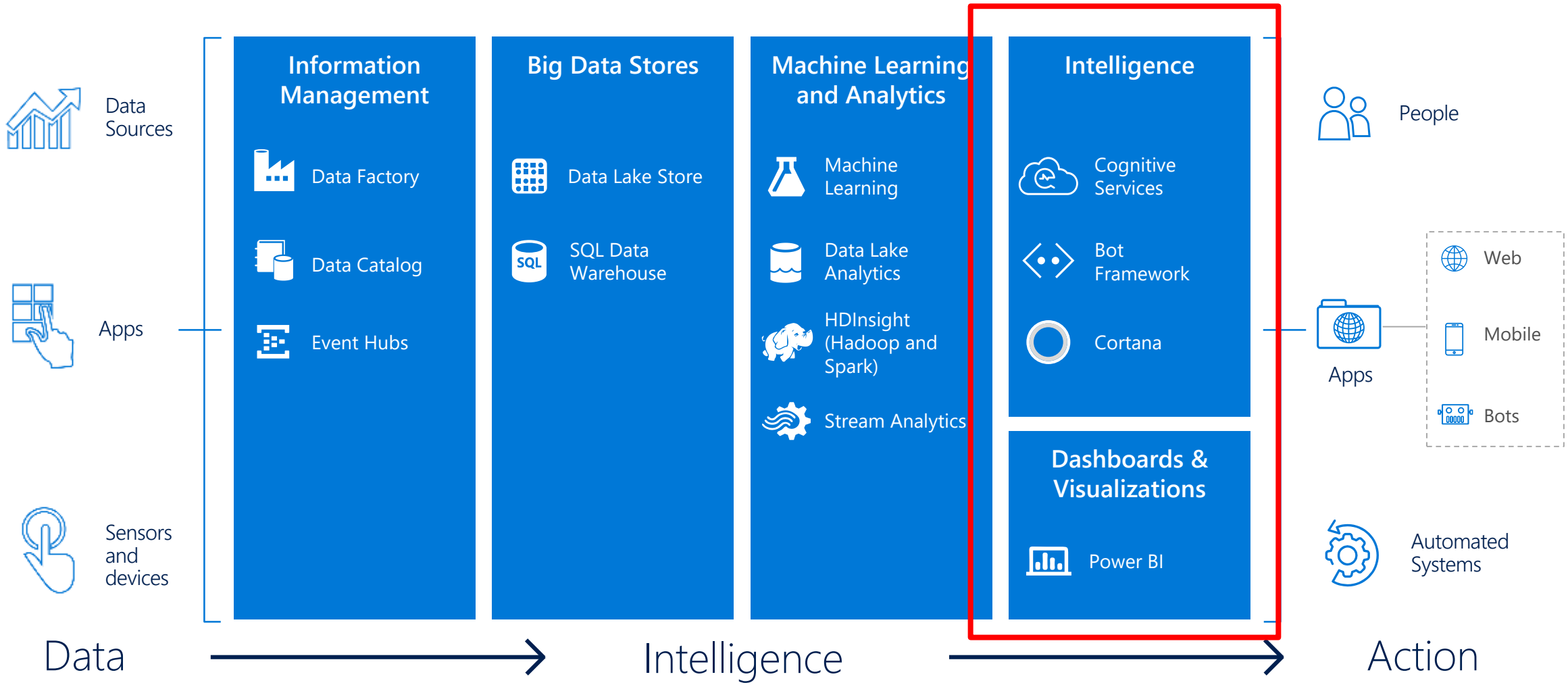
An example of a big data analytics process…



How does it all fit in so far?



Easily turn data into intelligent action



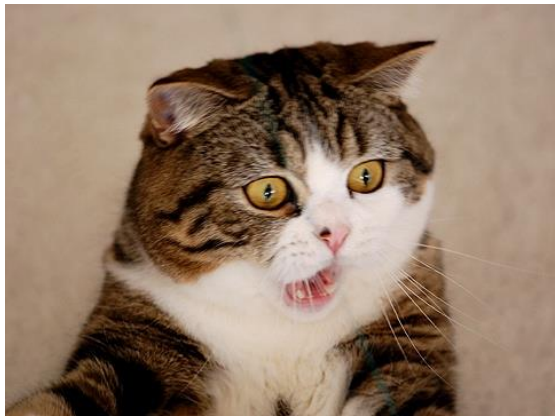
Machine Learning

ML can answer these main types of questions:

- **Classification** – is this group A, B, C or D?
- **Anomaly detection** – is this weird?
- **Regression** – how much? Or how many?
- **Clustering** – how is this organized?

Classification

- Will this tire fail in the next 1,000 miles: Yes or no?
- Which brings in more customers: a $5 coupon or a 25% discount?
- What breed of cat is this? A, B, C or D?



Anomaly detection

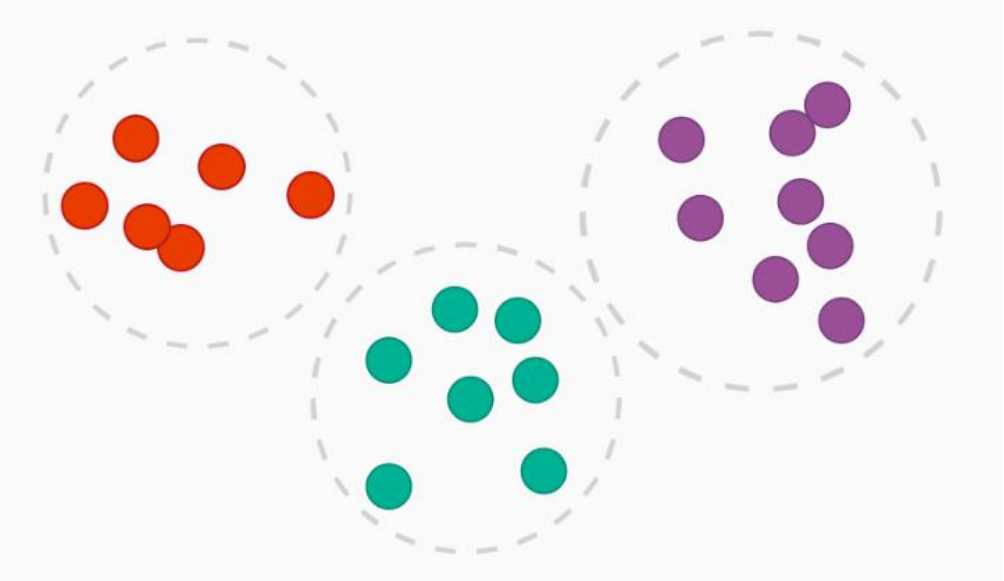
- Flags unusual events
- Is this credit card transaction normal?
- Is this pressure gauge reading normal?

Regression

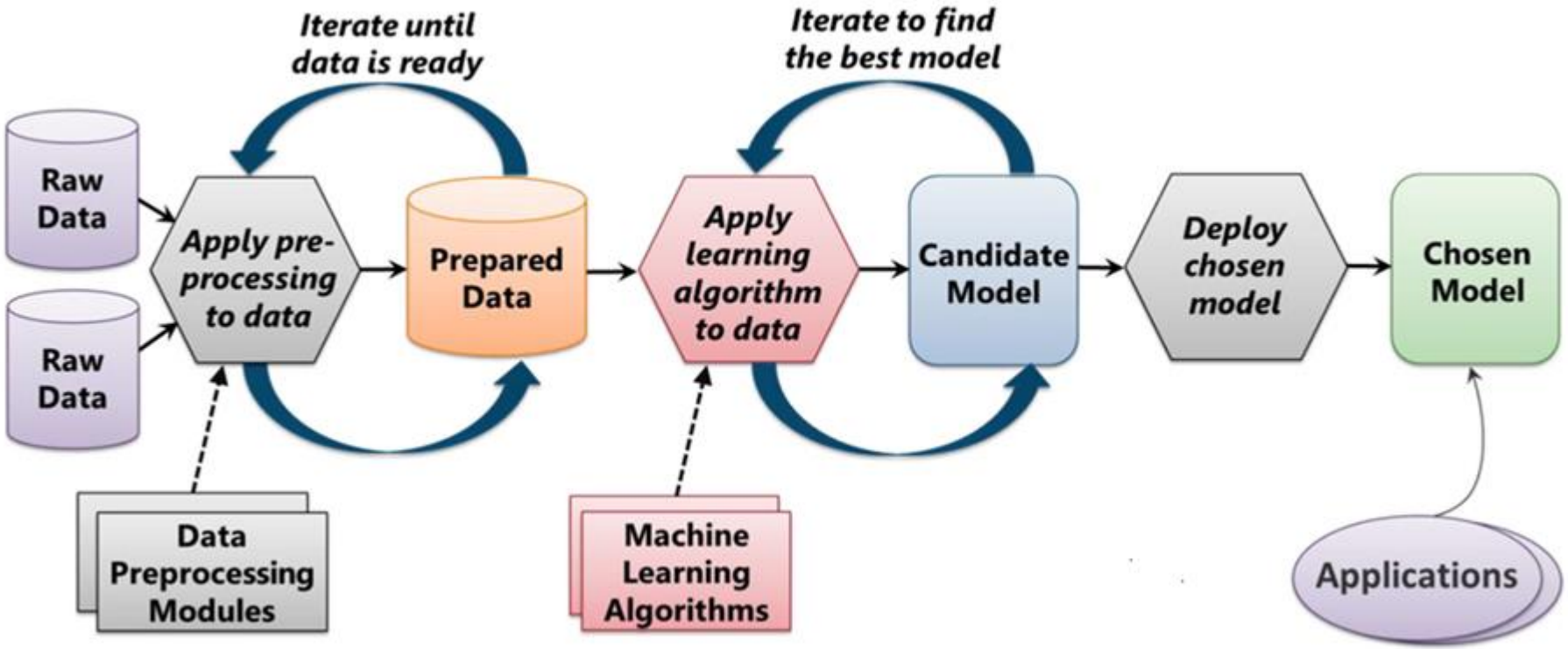
- Numerical predictions
- What will the temperature be next Tuesday?
- What will my fourth quarter sales be?

Clustering

- Understand how data is organized
- Which viewers like the same types of movies?
- Which printer models fail the same way?



The Machine Learning Process



From "Introduction to Microsoft Azure" by David Chappell

Azure Machine Learning

- Fully managed cloud service for building and operationalizing ML models

Fully managed

No software to install, no hardware to manage, and one portal to view and update.

Integrated

Simple drag, drop and connect interface for Data Science. No need for programming for common tasks.

Best in Class Algorithms + R

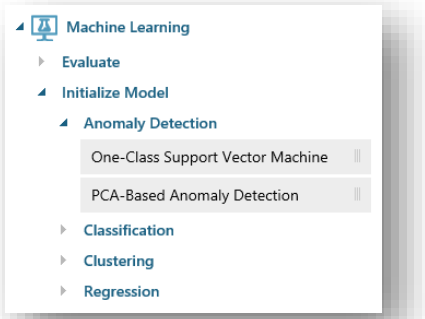
Built-in collection of best of breed algorithms. Support for R and popular CRAN packages.

Deploy in minutes

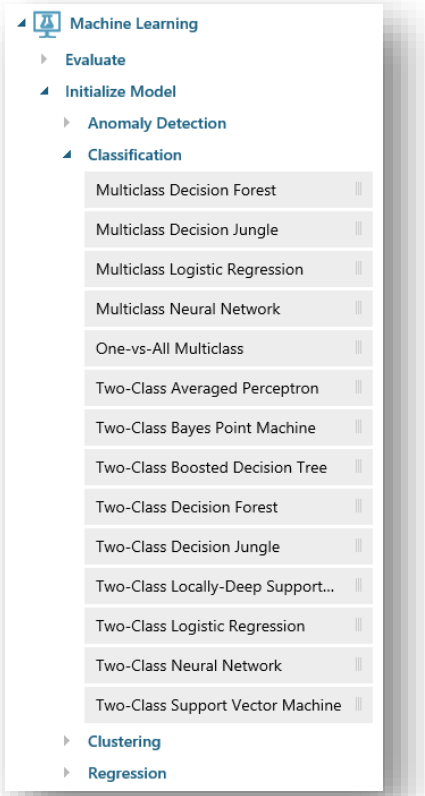
Operationalize models with a single click. Monetize in Machine Learning Marketplace.

Azure Machine Learning Algorithms

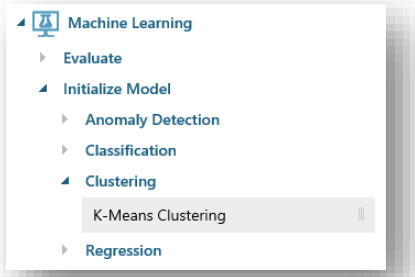
Anomaly



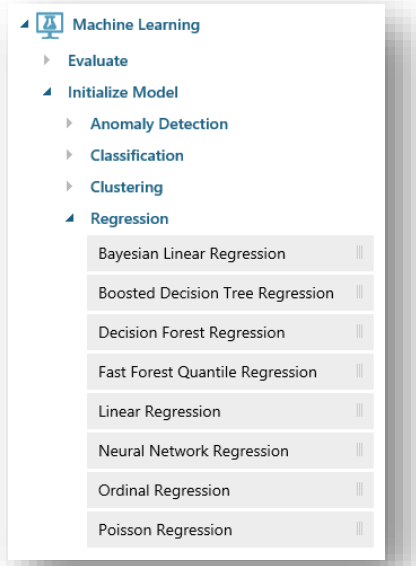
Classification



Clustering



Regression



Why Azure ML?

- Data Science is complex
  - Cost of accessing/using efficient ML algorithms is high
  - Comprehensive knowledge required on different tools/platforms to develop a complete ML project
  - Difficult to put the developed solution into a scalable production stage

- Azure Machine Learning provides an easier and faster solution

Azure Machine Learning Service

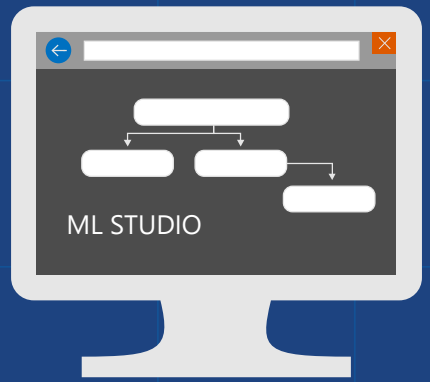
Data -> Predictive model -> Operational web API in minutes

Data

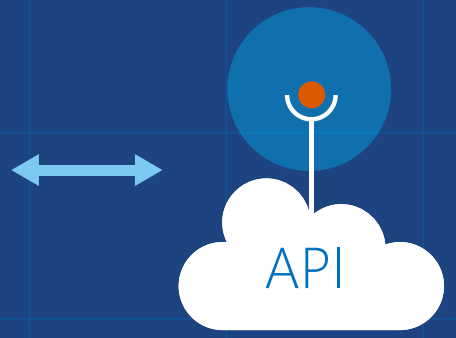
Blobs and Tables

Hadoop (HDInsight)

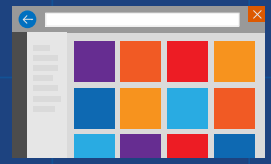
Relational DB (Azure SQL DB)



Integrated development environment for Machine Learning



Model is now a web service that is callable



Monetize the API through our marketplace

Clients

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

Demo – Azure ML

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.

START

Finding unusual data points

Discovering structure

Predicting categories

Predicting values

Three or more

Two

ANOMALY DETECTION

- **One-class SVM** — >100 features, aggressive boundary
- **PCA-based anomaly detection** — Fast training

CLUSTERING

- **K-means**

MULTI-CLASS CLASSIFICATION

- Fast training, linear model → **Multiclass logistic regression**
- Accuracy, long training times → **Multiclass neural network**
- Accuracy, fast training → **Multiclass decision forest**
- Accuracy, small memory footprint → **Multiclass decision jungle**
- Depends on the two-class classifier, see notes below → **One-v-all multiclass**

REGRESSION

- **Ordinal regression** — Data in rank ordered categories
- **Poisson regression** — Predicting event counts
- **Fast forest quantile regression** — Predicting a distribution
- **Linear regression** — Fast training, linear model
- **Bayesian linear regression** — Linear model, small data sets
- **Neural network regression** — Accuracy, long training time
- **Decision forest regression** — Accuracy, fast training
- **Boosted decision tree regression** — Accuracy, fast training, large memory footprint

TWO-CLASS CLASSIFICATION

- **Two-class SVM** — >100 features, linear model
- **Two-class averaged perceptron** — Fast training, linear model
- **Two-class logistic regression** — Fast training, linear model
- **Two-class Bayes point machine** — Fast training, linear model

- Accuracy, fast training → **Two-class decision forest**
- Accuracy, fast training, large memory footprint → **Two-class boosted decision tree**
- Accuracy, small memory footprint → **Two-class decision jungle**
- >100 features → **Two-class locally deep SVM**
- Accuracy, long training times → **Two-class neural network**

Created by the Azure Machine Learning Team

Email: AzurePoster@microsoft.com

Download this poster: http://aka.ms/MLCheatSheet

Microsoft

http://aka.ms/MLCheatSheet

Cognitive Services

Give your solutions a human side

Microsoft Cognitive Services preview

Vision

From faces to feelings, allow your apps to understand images and video

Speech

Hear and speak to your users by filtering noise, identifying speakers, and understanding intent

Language

Process text and learn how to recognize what users want

Knowledge

Tap into rich knowledge amassed from the web, academia, or your own data

Search

Access billions of web pages, images, videos, and news with the power of Bing APIs

Cognitive Services

Give your solutions a human side

Microsoft Cognitive Services preview

Vision

Computer Vision | Emotion | Face | Video | Content Moderator

Speech

Custom Recognition | Speaker Recognition | Speech

Language

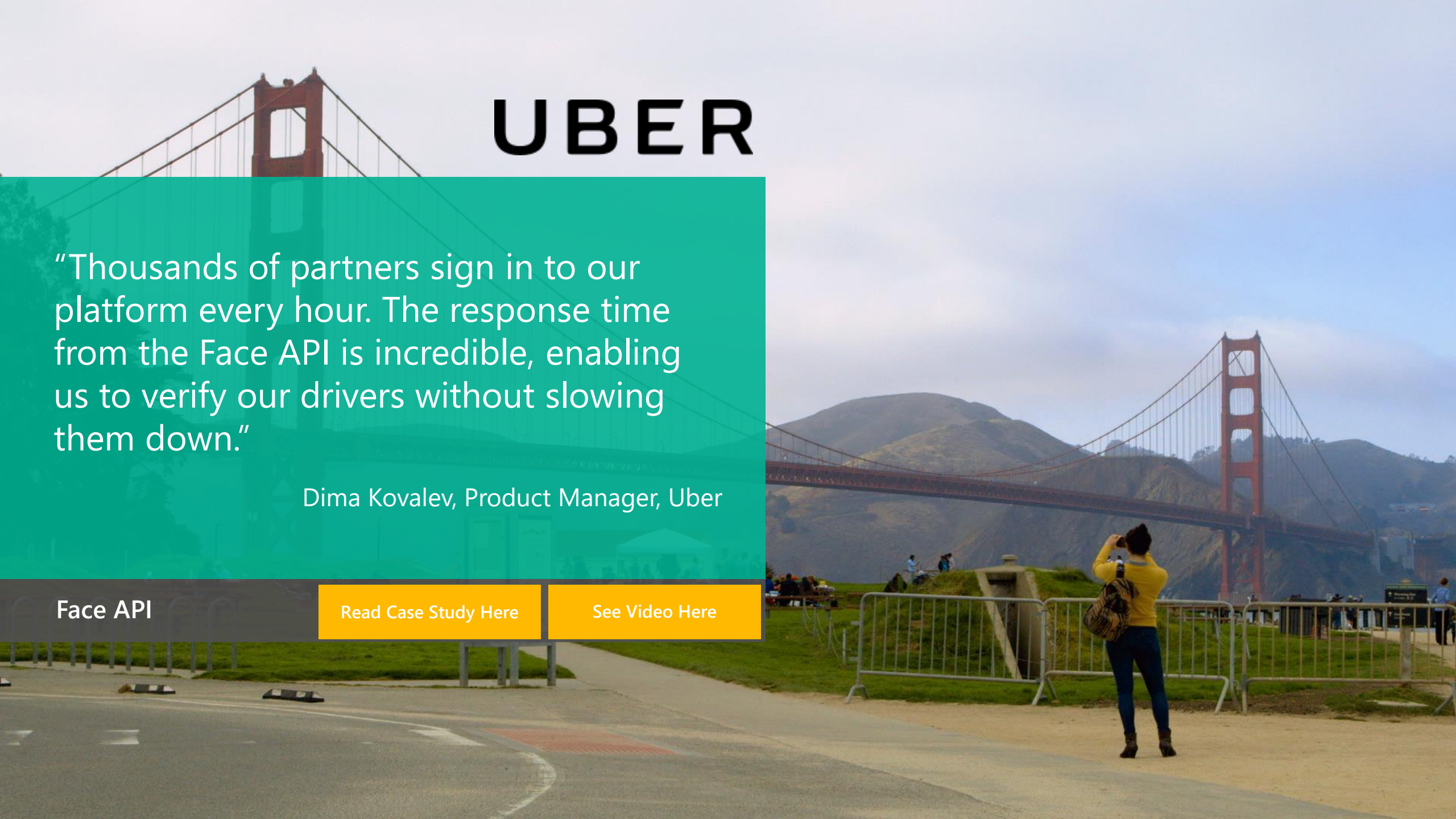
Bing Spell Check | Language Understanding | Linguistic Analysis | Text Analytics | Web Language Model | Translator Text and Speech

Knowledge

Academic Knowledge | Entity Linking | QnA Maker Knowledge Exploration | Recommendations

Search

Bing Auto Suggest | Bing Image Search | Bing News Search | Bing Video Search | Bing Web Search



UBER

"Thousands of partners sign in to our platform every hour. The response time from the Face API is incredible, enabling us to verify our drivers without slowing them down."

Dima Kovalev, Product Manager, Uber

Face API

Read Case Study Here

See Video Here

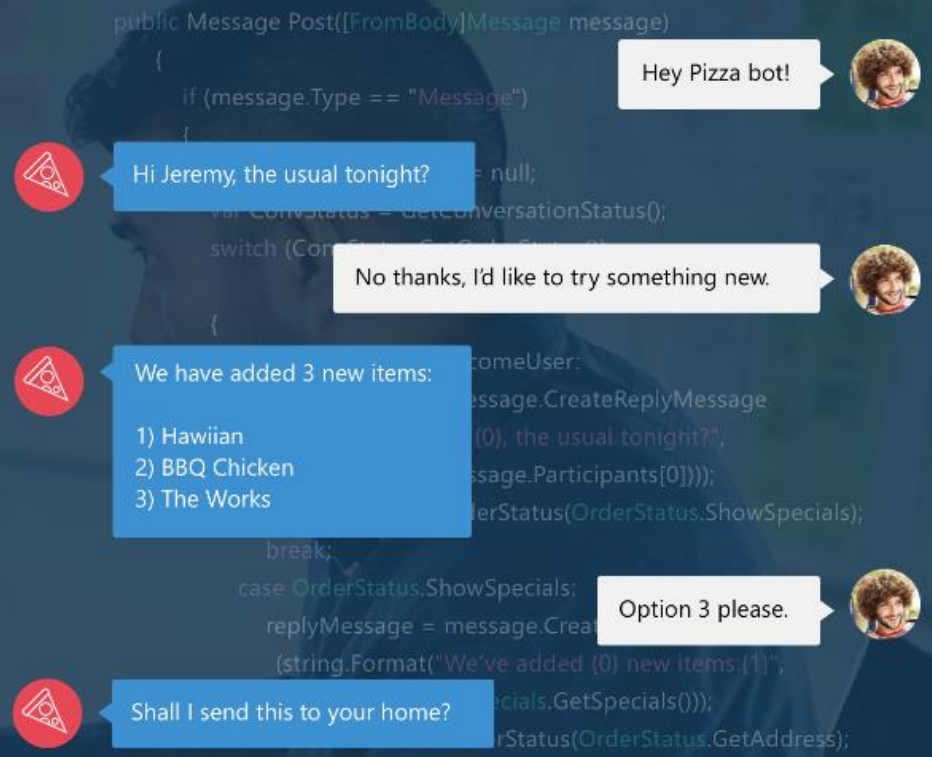
Microsoft

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



www.botframework.com

Bot Framework Components

Your Bot Framework Bot

Bot Builder SDKs

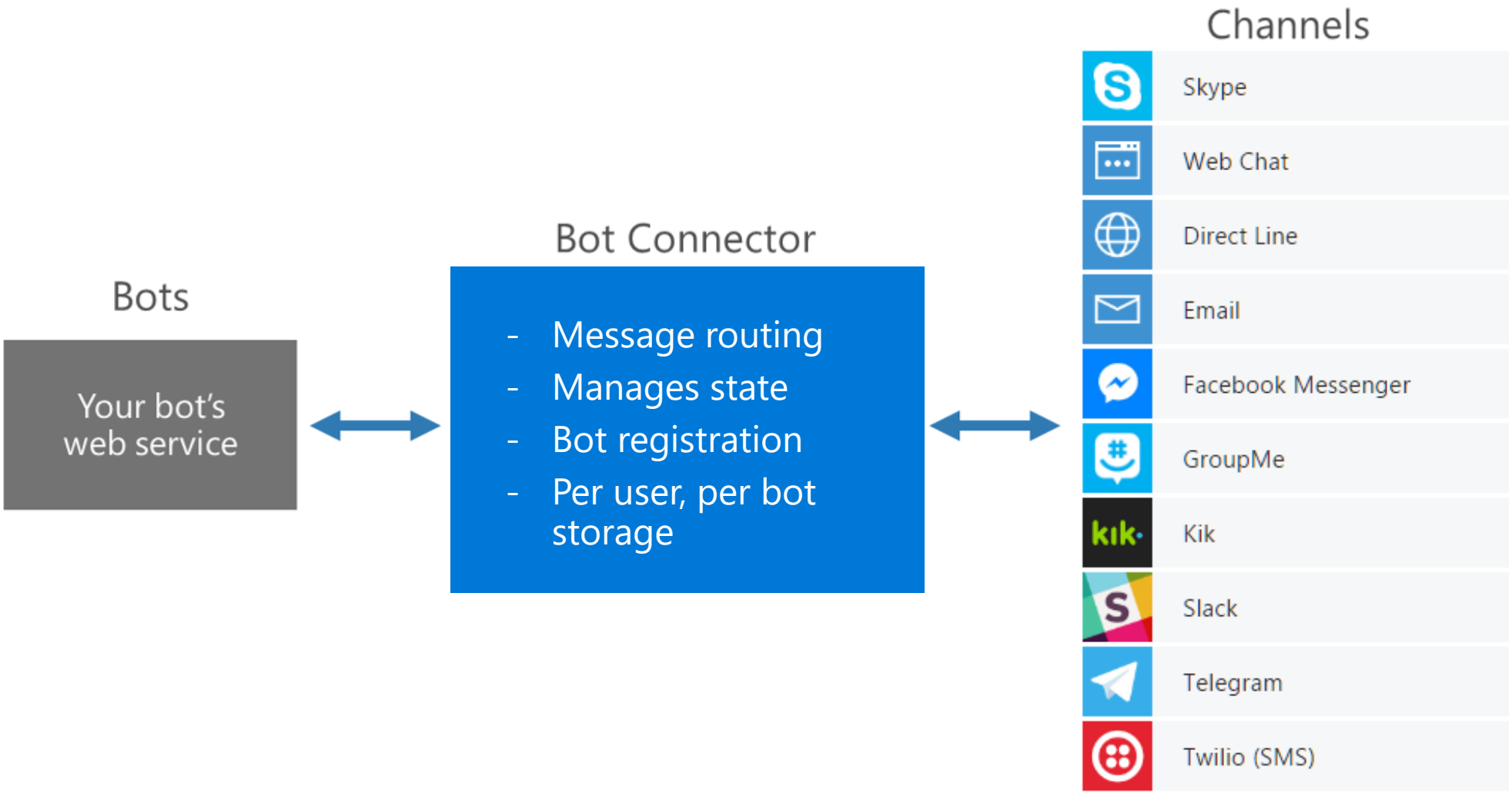
Build great dialogs within your Node.js- or C#-based bot

Bot Connector

Connect your bot to sms, Office 365 mail, Skype, Slack, etc.

Bot Directory

Publish bots and try others on the bot directory


Bots
Your bot's web service
Bot Connector
- Message routing
- Manages state
- Bot registration
- Per user, per bot storage
Channels
Skype
Web Chat
Direct Line
Email
Facebook Messenger
GroupMe
Kik
Slack
Telegram
Twilio (SMS)

What did we cover today?

- General architecture for Big Data
- Cortana Intelligence Suite – Microsoft's tools for Big Data
- Azure Machine Learning
- Cognitive Services
- Bot Framework

Thank you!

ongalyssa@outlook.com