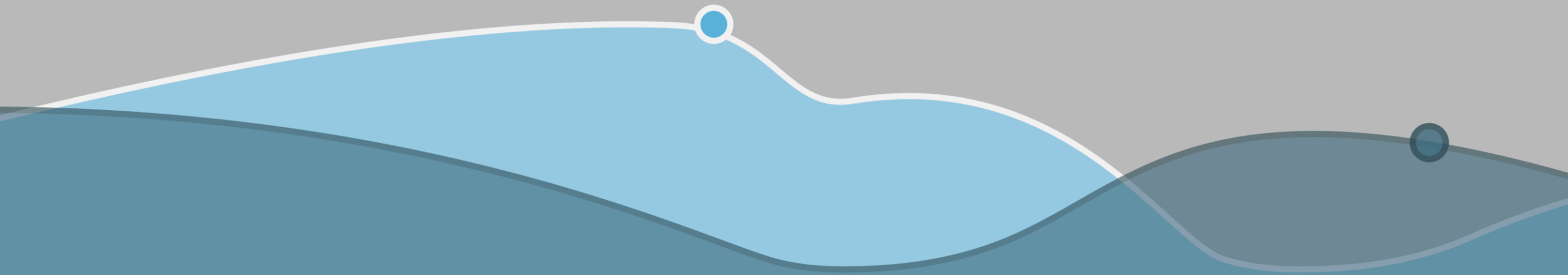
Microsoft

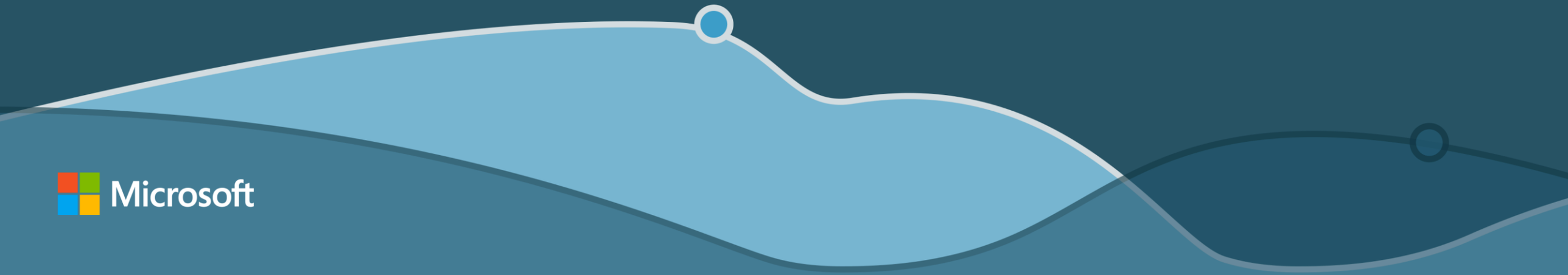
Cortana Analytics Workshop

Sept 10 – 11, 2015 • MSCC

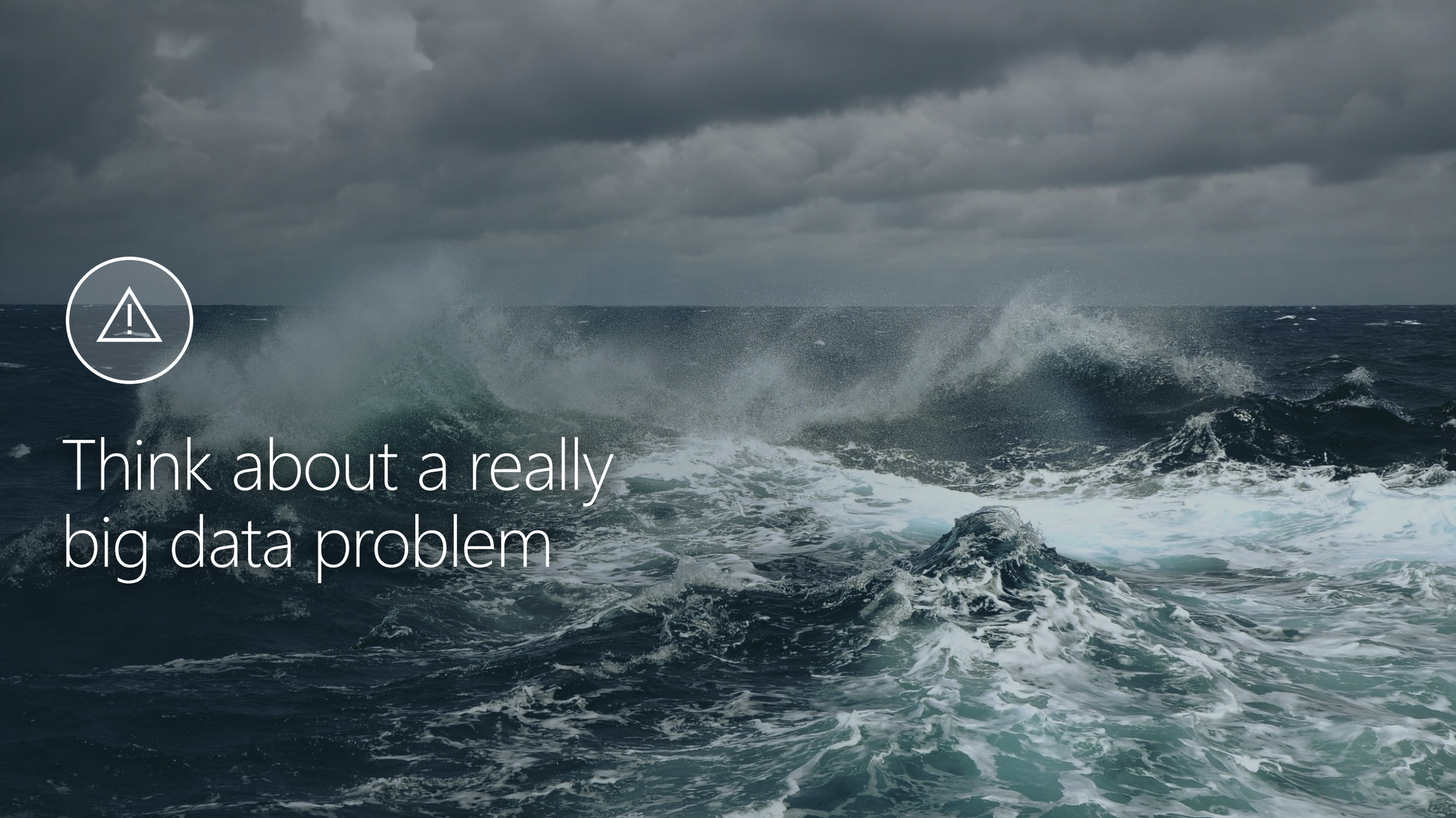


Demystifying Cortana Analytics

Jason Wilcox
Director of Engineering



Microsoft



Think about a really big data problem

Tough Love for Microsoft Search

Danny Sullivan, Search Engine Land
December 30, 2008

Processing **hundreds of billions** of documents

Understanding **billions** of entities

Reasoning **trillions** of entity relationships

Seeing **hundreds of billions** of queries and clicks

Serving and adapting for **billions** of users

Search is a
big data play

We tried to muscle our way to success

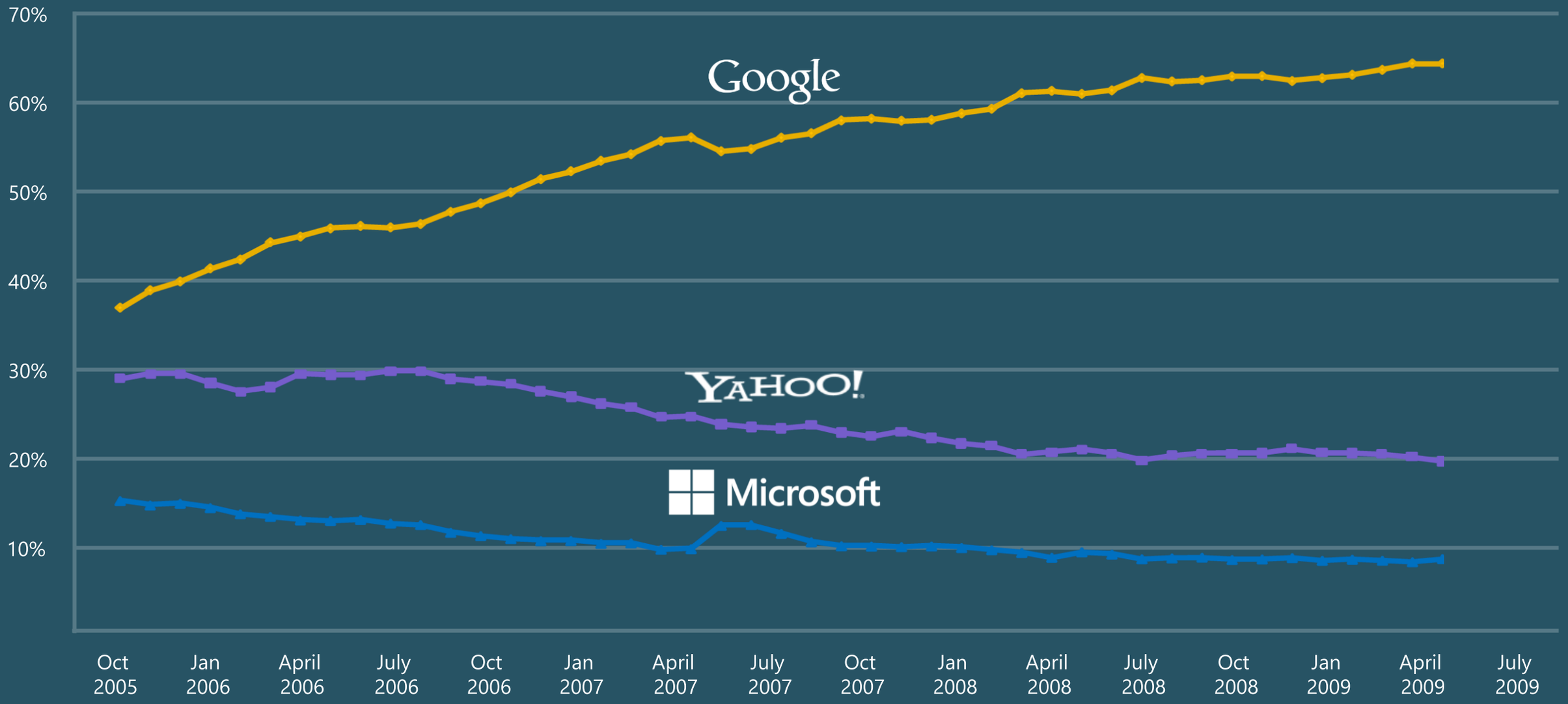
We hired the **best** and **brightest**

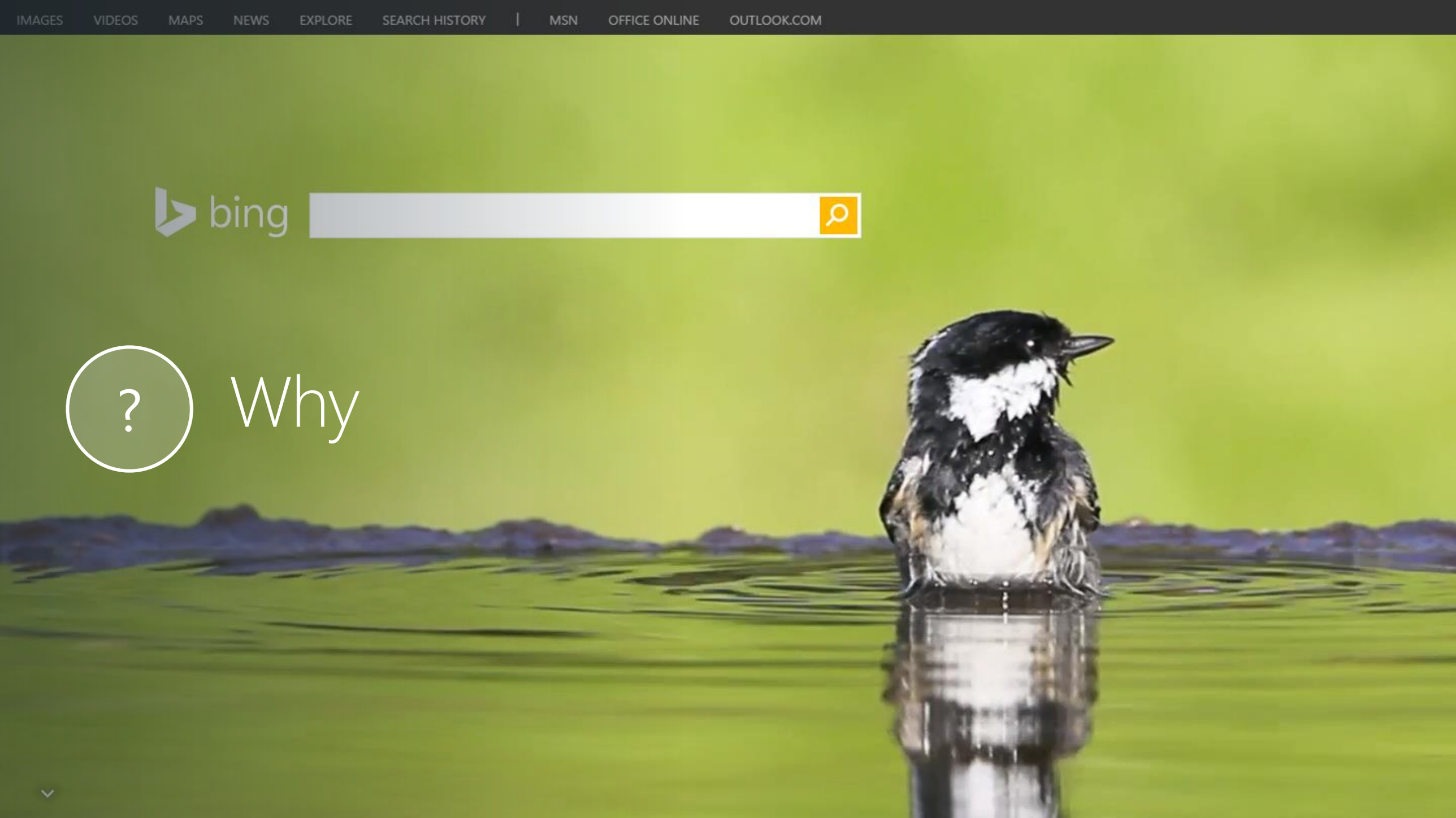
We threw **technology** at the problem

We threw **people** at the problem

And, it **didn’t work**

Microsoft had a sliding US search share 2005-2009



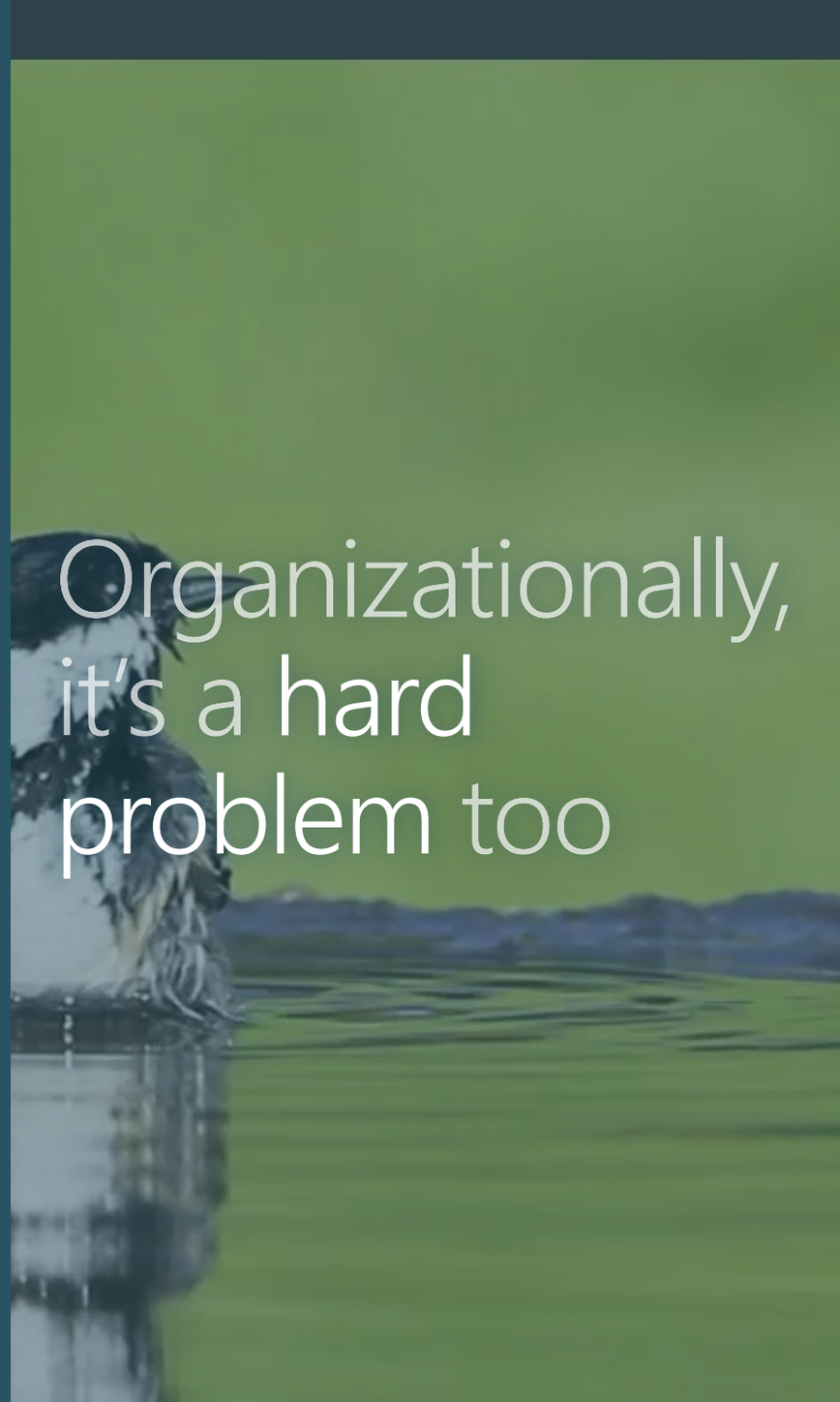

IMAGES
VIDEOS
MAPS
NEWS
EXPLORE
SEARCH HISTORY
MSN
OFFICE ONLINE
OUTLOOK.COM
bing
?
Why

Hard to hire **data scientists**

Lots of people involved **need access** to data

Huge technical integration **challenges**

Massive **volume**, **velocity**, and **variety** of data

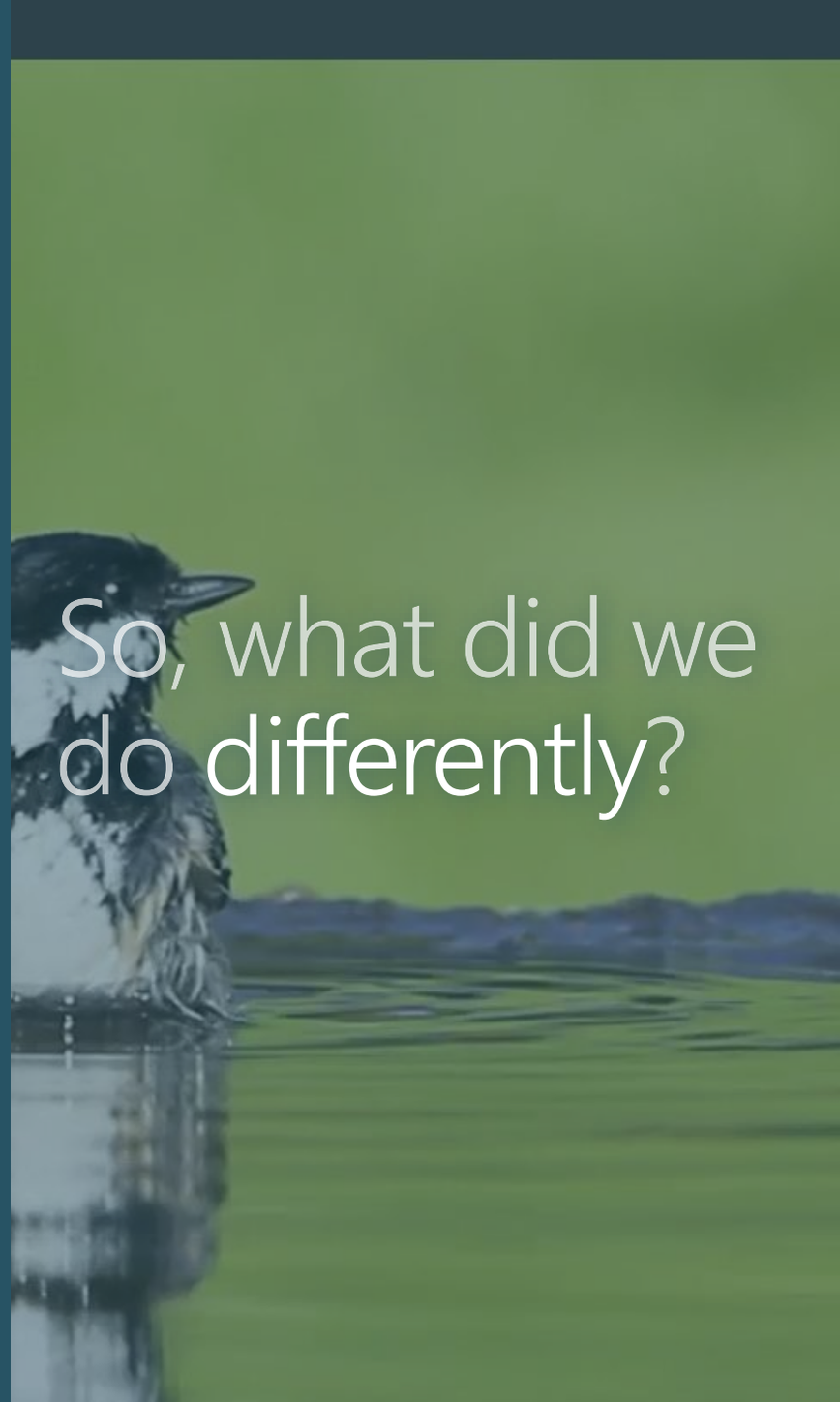


Qi Lu was brought on board—a new perspective

Laser focused on measuring relevance (NDCG)

Measure the rate of exploration (count of experiments)

Empower everyone to explore (100x people)



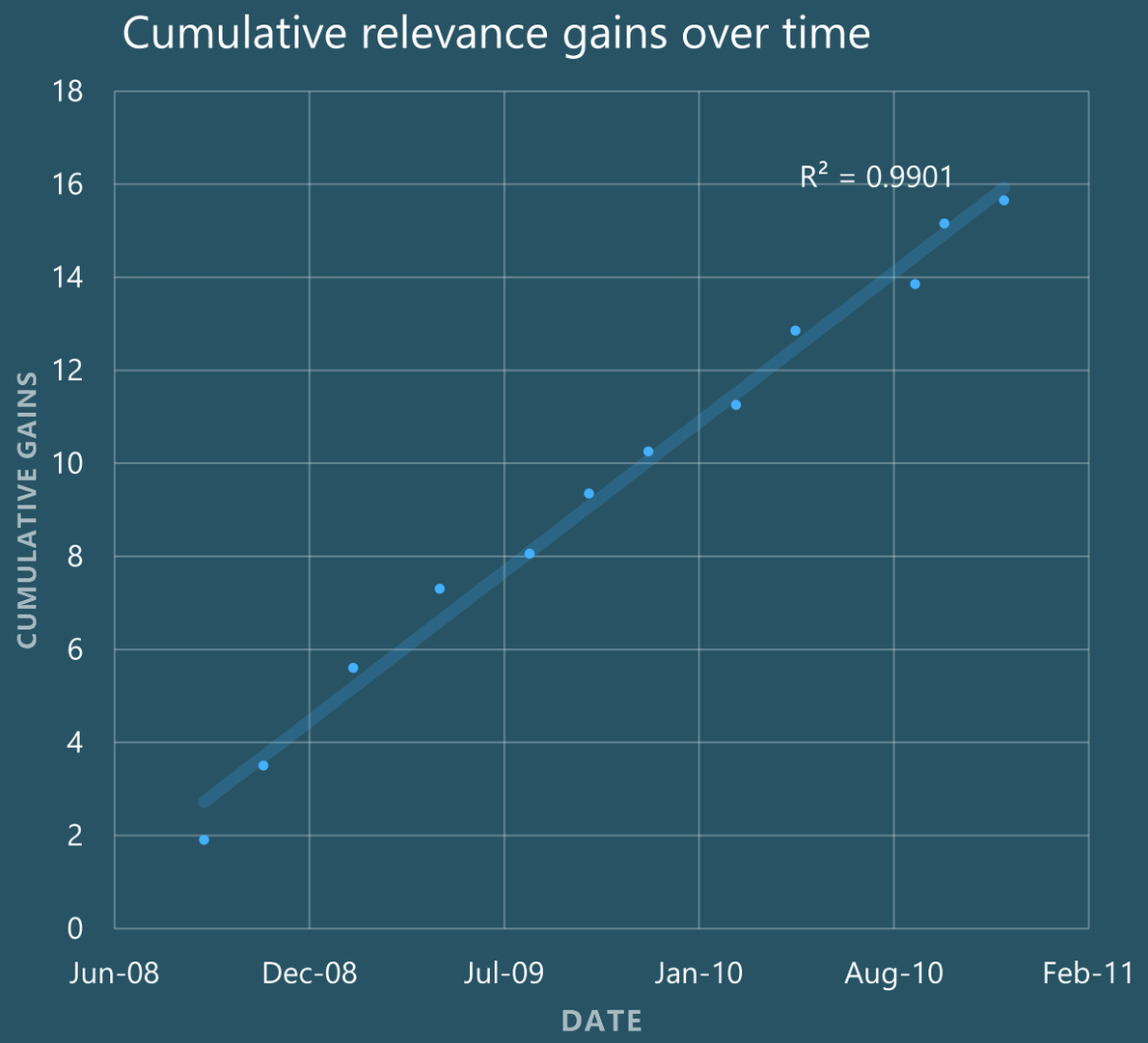
Infrastructure investments are risky

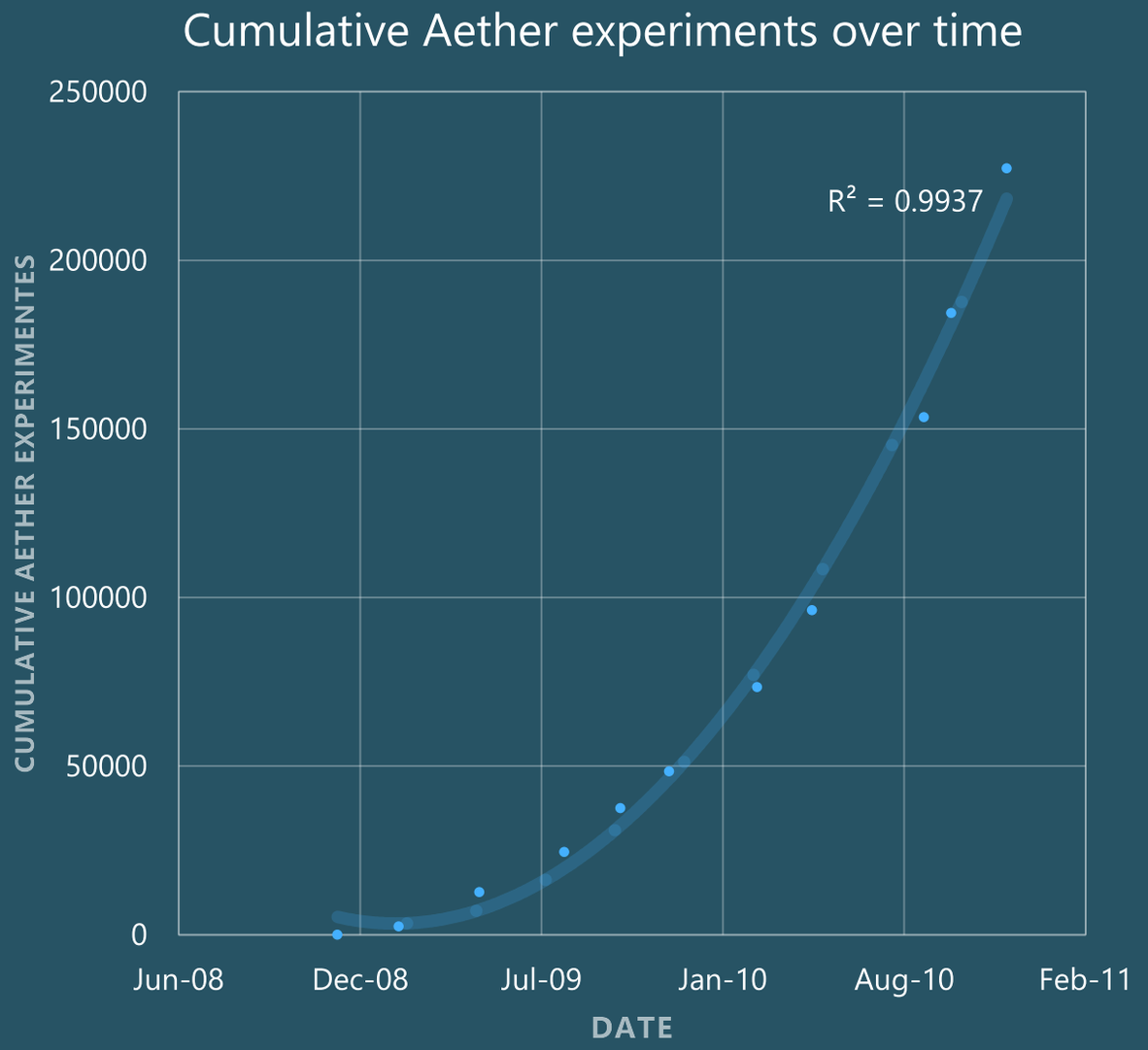
Built an exabyte-scale data lake for everyone to put their data of all types (structured and unstructured)

Built tools approachable by any developer

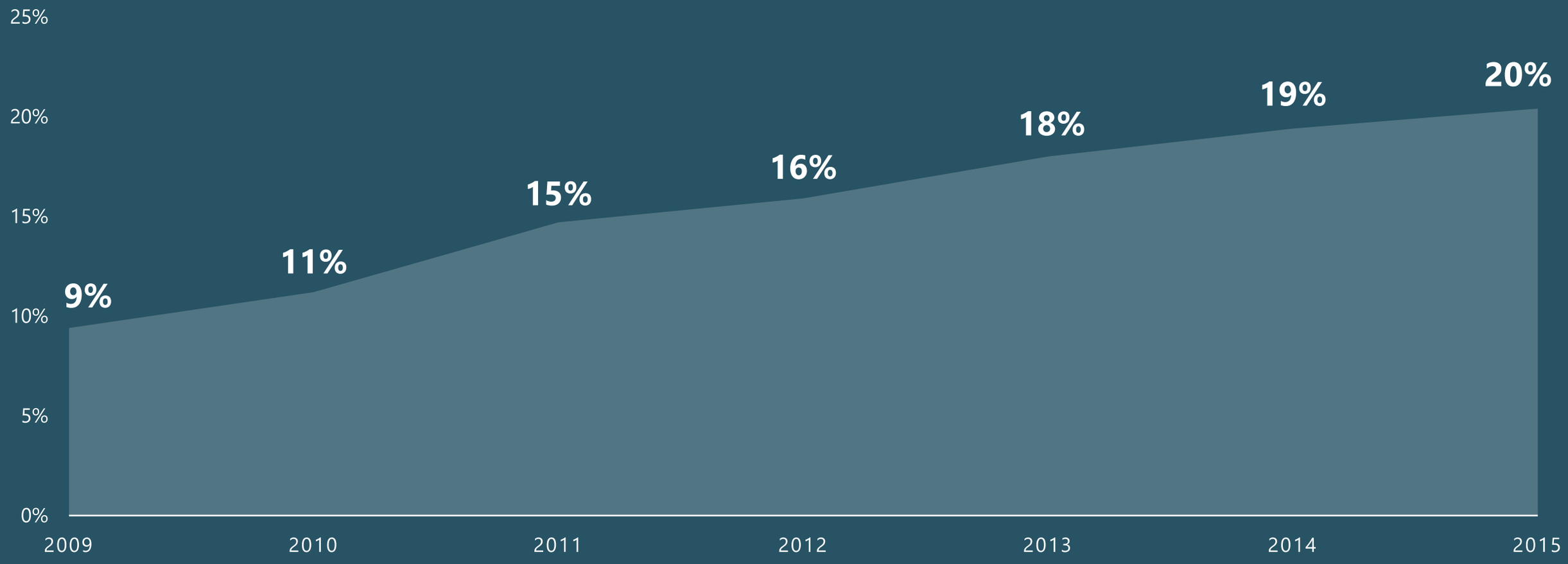
Built machine learning tools for collaborating across large experiment models

...and it helped





Microsoft doubles search share



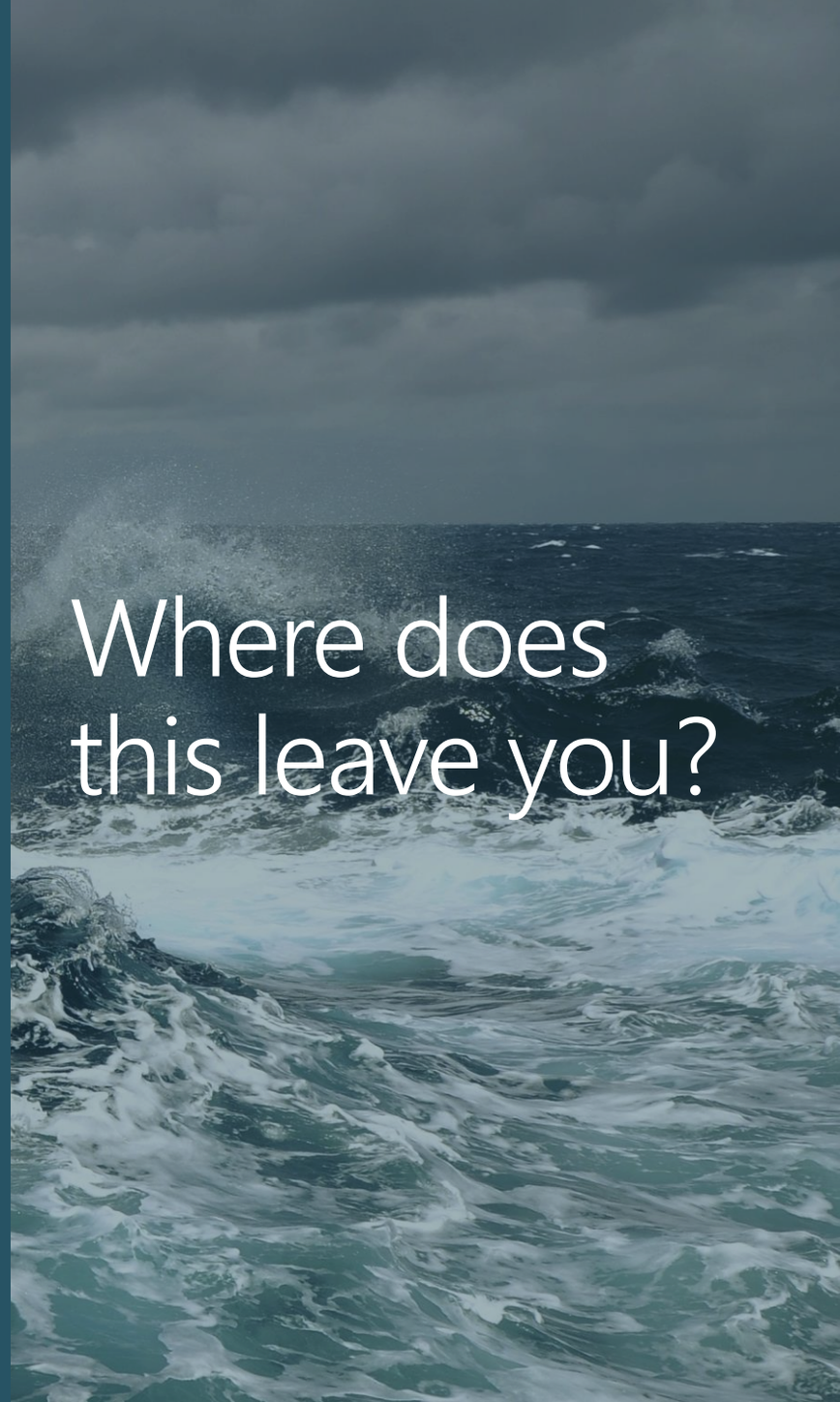
Source: ComScore 2009-2015 Search Report US

We went through these learnings

It is a hard problem

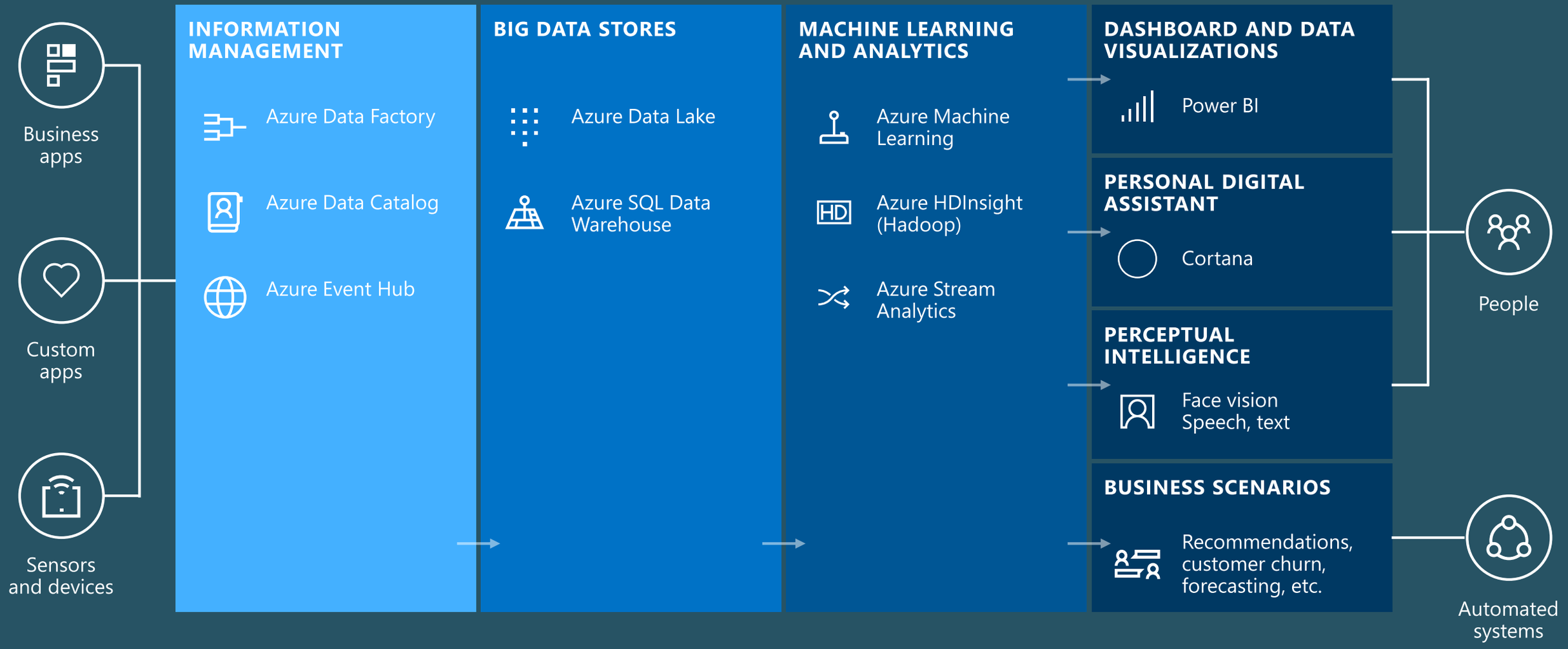
It requires a lot of investment and know-how to get it right

With Cortana Analytics, you don't have to be the size of Bing to solve the problem



Cortana Analytics Suite

TRANSFORM DATA INTO INTELLIGENT ACTION



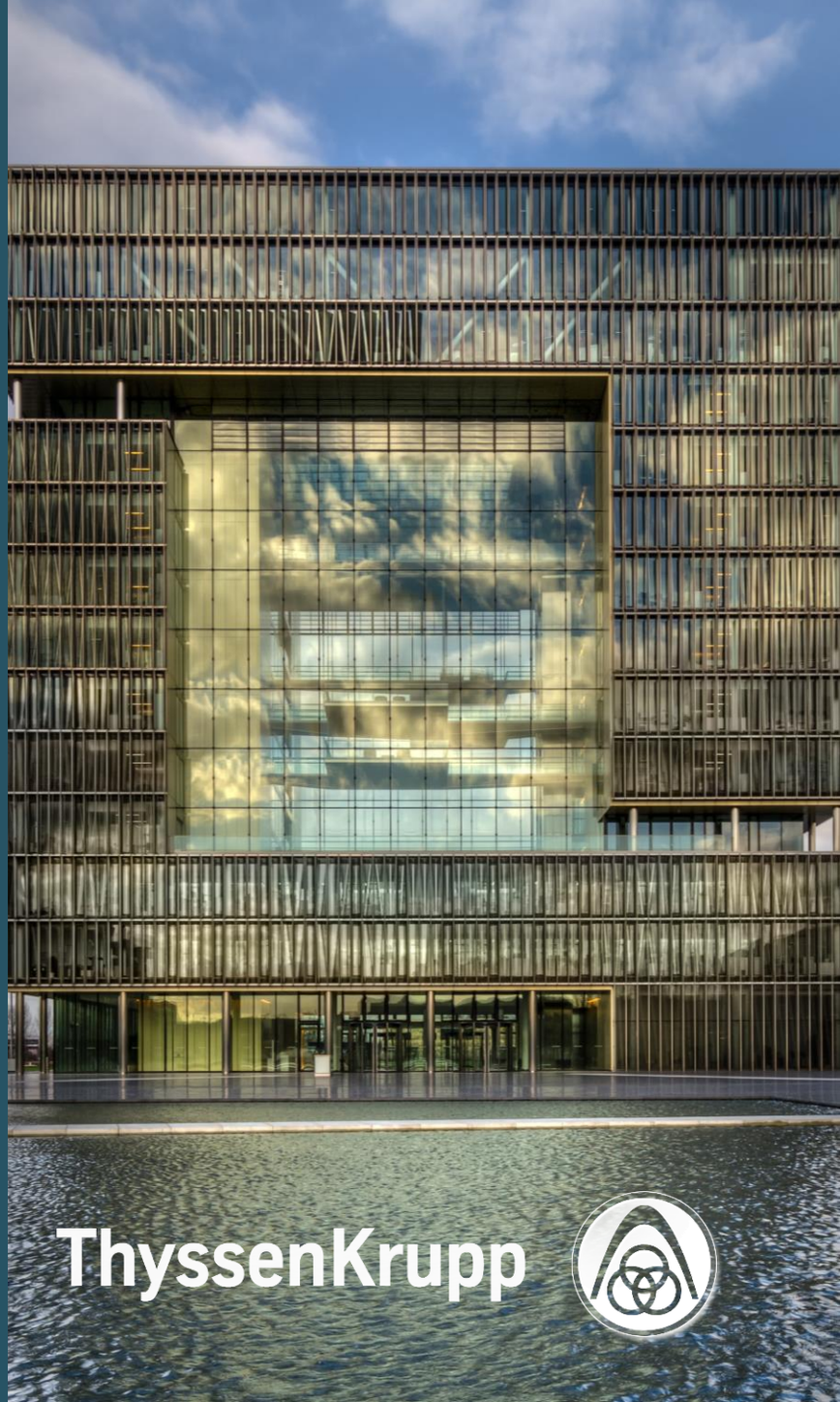

What are the questions people ask of their data?

Predictive maintenance

THYSSENKRUPP WANTED TO BOTH OFFER BETTER CUSTOMER SERVICE AND LOWER THEIR COSTS

> We wanted to go beyond the industry standard of preventative maintenance, to offer predictive and even preemptive maintenance.

ANDREAS SCHIERENBECK
CEO, ThyssenKrupp Elevator



Personalized offers

PIER 1 IMPORTS WANTED TO OFFER A CONNECTED, PERSONAL EXPERIENCE, BOTH ONLINE AND IN STORE

> At Pier 1 Imports, we've embraced the cloud. It helps us operationalize technology quickly and react to our ever-evolving business needs.

ANDREW LAUDATO
Pier 1 Imports

Churn prediction

TACOMA PUBLIC SCHOOL WANTED TO LEVERAGE DATA TO PREDICT STUDENT DROPOUT RISKS TO INCREASE GRADUATION RATES

> “With the addition of the Azure recommendations, user activity on the site has increased tremendously. And we see an increase in users who are coming to the site through those recommendations.”

SHAUN TAYLOR
Chief Information Officer

Smart buildings

CMU WANTED TO USE SENSOR DATA FOR MORE THAN REACTIVE REPAIR AND DIAGNOSIS

> We see Azure ushering in an era of self-service predictive analytics for the masses. We can only imagine the possibilities.

BERTRAND LASTERNAS
Carnegie Mellon

You get a question from a stakeholder…

TO ANSWER THE QUESTION, TEAMS MUST:

1. Find out what data sets are available
2. Gain access to the data
3. Shape the data
4. Run first experiment
5. Repeat steps 1, 2, 3, and 4 until you get it right
6. Find the insight
7. Operationalize/action

How can Cortana Analytics help?

FULLY MANAGED

ALL TYPES

OPERATIONALIZE

SHARE AND COLLABORATE

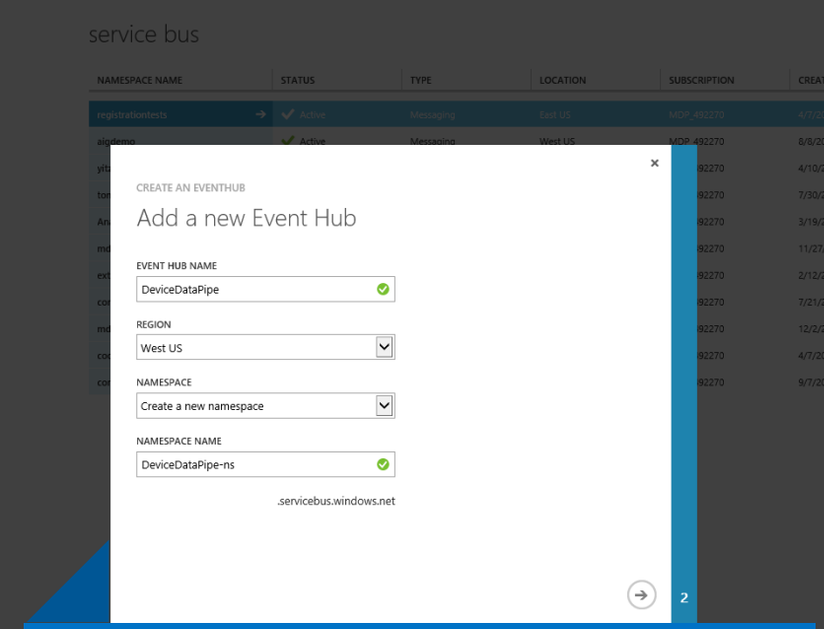
OPENNESS

END-TO-END

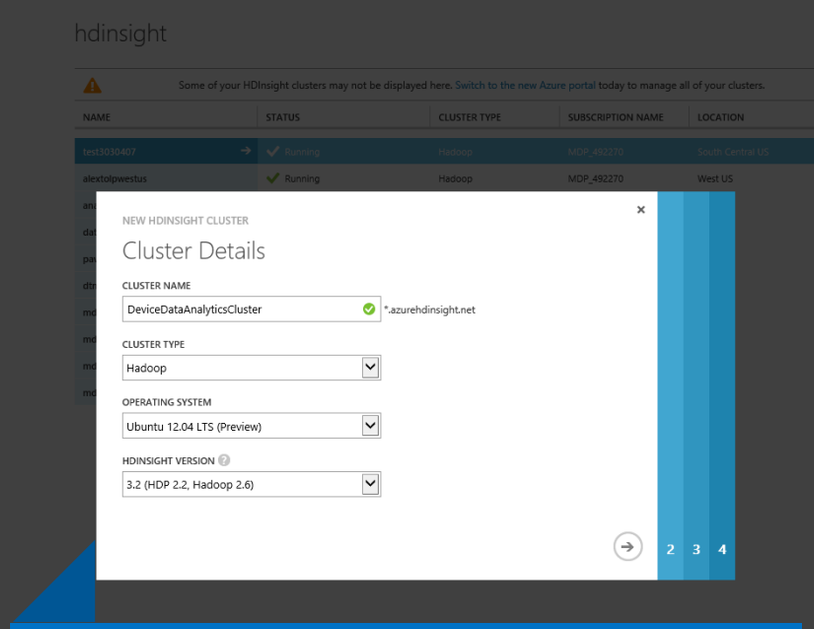
1

Fully managed set of services

Focus on the data problem, not the infrastructure

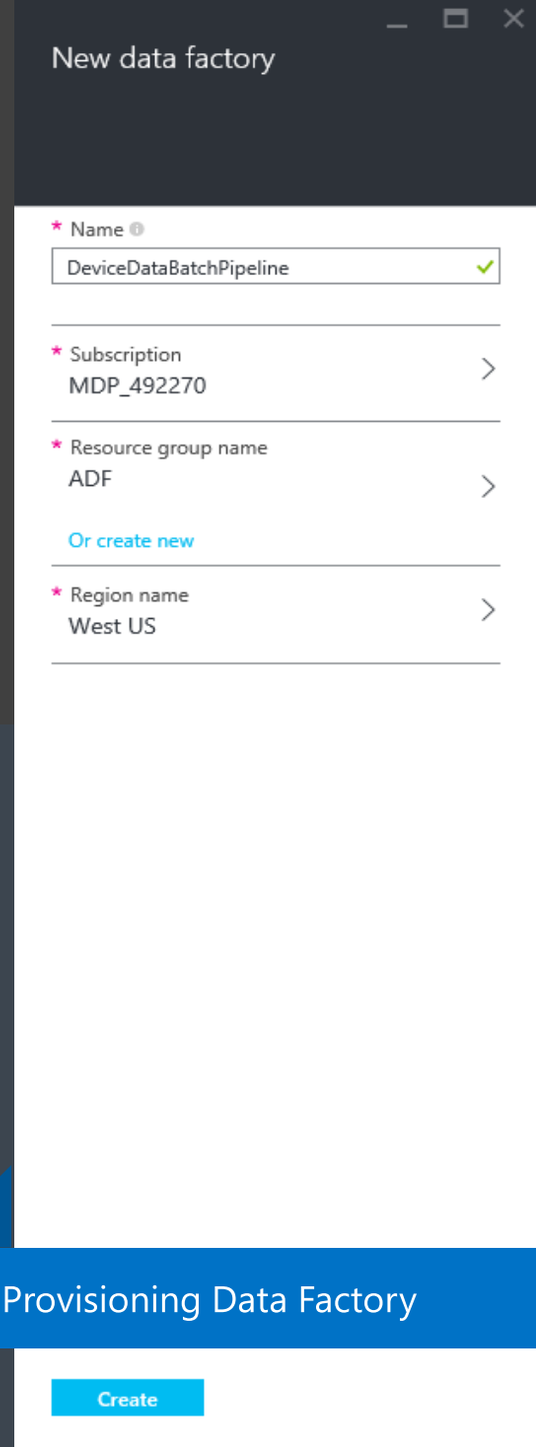


Provisioning Event Hub



Provisioning HDInsight Linux

Provisioning Stream Analytics



Provisioning Data Factory

2

Work with all types of data

Structured data
Unstructured data
At any scale

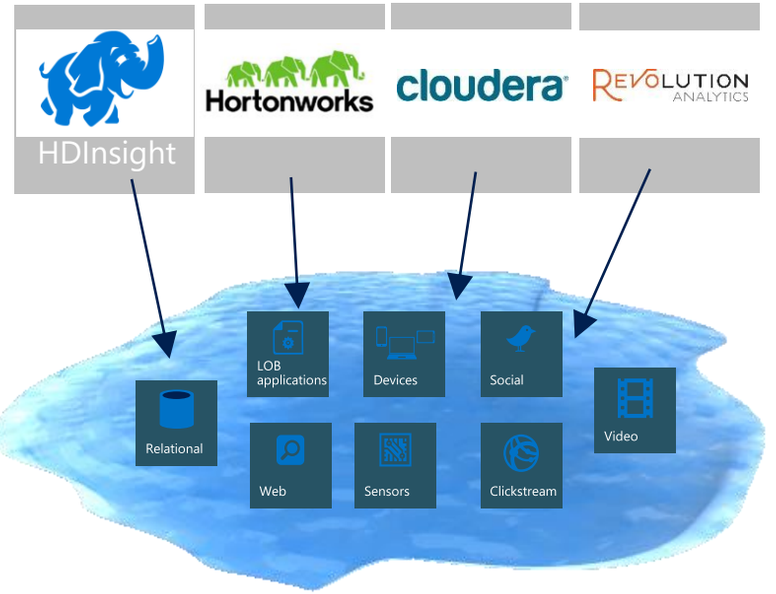
Azure Data Lake

Hadoop distributed file system for the cloud

Built from the ground-up as native HDFS

Integrated with HDInsight, Hortonworks, and Cloudera

Accessible to all HDFS compliant projects
(Spark, Storm, Flume, Sqoop, Kafka, R, etc.)



Built using open standards

3

Operationalize

Move from ad-hoc explorations to full operationalization using the same tools

Operationalize data movement training, scoring, and pipelines

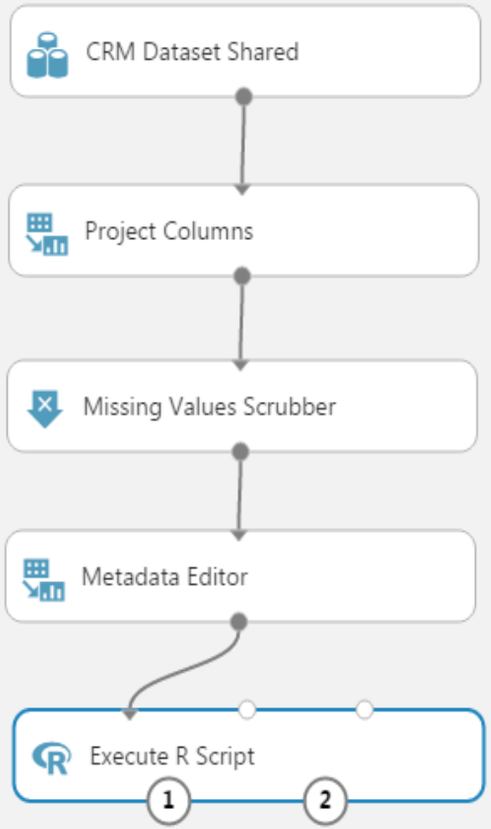
Using R for Performing PCA

In draft

Draft saved at 11:56:46 PM

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



Properties

Execute R Script

R Script

```
# @author weehyongtok
# @version 1.0
# @date 20150110
# Map 1-based optional input ports to variable
dataset1 <- maml.mapInputPort(1)

# Perform PCA on the first 190 columns
pca = prcomp(dataset1[,1:190])
top_pca_scores = data.frame(pca$x[,1:10])
data.set = top_pca_scores
plot(data.set)

# Select data.frame to be sent to the output Dataset port
maml.mapOutputPort("data.set");

```

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

4

Collaboration and sharing

Extract value from data requires full organizational participation

Tie disparate data ecosystems together

Share learnings

Provide access to data at any stage of the process

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 | Data Prof

Name:
hive

Friendly Name:
Vehicle Health Telemetry

Description:
The blobs data containing all vehic
telemetry data that is to be used fo
identifying all roadside asistance ca
vechile diagnostics and eco-driving

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 geographies.

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

Microsoft Azure Machine Learning | Home | Studio | Gallery PREVIEW

Browse | Search for entities by name, algorithms or tags

Discover. Learn. Share.

Learn more

EXPERIMENT

Experiment for the Guide to R in Azu

This experiment is developed i in Azure Machine Learning.

forecasting | predictive analytics | graphics

845 224 7 months ago

Stephen Elston

Machine Learning APIs

MACHINE LEARNING API

Text Analytics

MACHINE LEARNING API

Speech APIs

MACHINE LEARNING API

Recommendations

Refine by

Categories

- [x] Experiment
- [ ] Machine Learning API
- [ ] Tutorial

Show

- [ ] Microsoft content only

Tags

- [ ] R
- [ ] Classification
- [ ] classification
- [ ] Apply Transformation
- [ ] Template

Show all

Algorithms used

- [ ] Two-Class Boosted Decision Tree
- [ ] Two-Class Logistic Regression
- [ ] Two-Class Support Vector Machine
- [ ] Linear Regression

Results

You've selected: Experiment | Clear all

EXPERIMENT

Sample 1: Download dataset from UCI: Adult...

This sample demonstrates how to download a dataset from a http location, add column names to the...

reader | http reader input

42420 4982 25 days ago

Microsoft

EXPERIMENT

Binary Classification: Twitter sentiment analy...

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

Two-Class Support Vector Machine

Classification | Text Mining

8445 3061 25 days ago

Microsoft

EXPERIMENT

Regression: Demand estimation

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

Boosted Decision Tree Regression

demand estimation

5873 1886 25 days ago

Microsoft

Microsoft Azure Machine Learning | Home | Studio | Gallery PREVIEW

Browse | Search for entities by name, algorithms or tags

MACHINE LEARNING API

Customer Churn Prediction

by Microsoft April 22, 2015

SIGN UP

TRY IT NOW

2721 views

Tweet | Share

Description

What if you could know which of your customers are more likely to stop doing business with you next month?

Customer Churn Prediction is a service built with Azure Machine Learning. It's designed to predict when a customer (player, subscriber, user, etc.) is likely to end his or her relationship with a company or service. Being able to predict which customers have a high risk of leaving the relationship with a company provides the company with a window of opportunity to approach them and reduce the likelihood of their leaving.

How does this service work?

By providing historical data to the service, you enable it to learn and train a model that fits your business. After the model has learned this patterns, you will be able to provide recent history for a set of customers and will receive a prediction identifying the subset of customers that have high risk of leaving your business.

No coding required

Get started doing churn analysis today without having to write a single line of code by visiting the Customer churn prediction interactive experience.

0 Comments | Azure ML Gallery | Login

Recommend | Share | Sort by Best

Start the discussion…

Links

Documentation

Publisher Offer Terms

Publisher Offer Privacy Statement

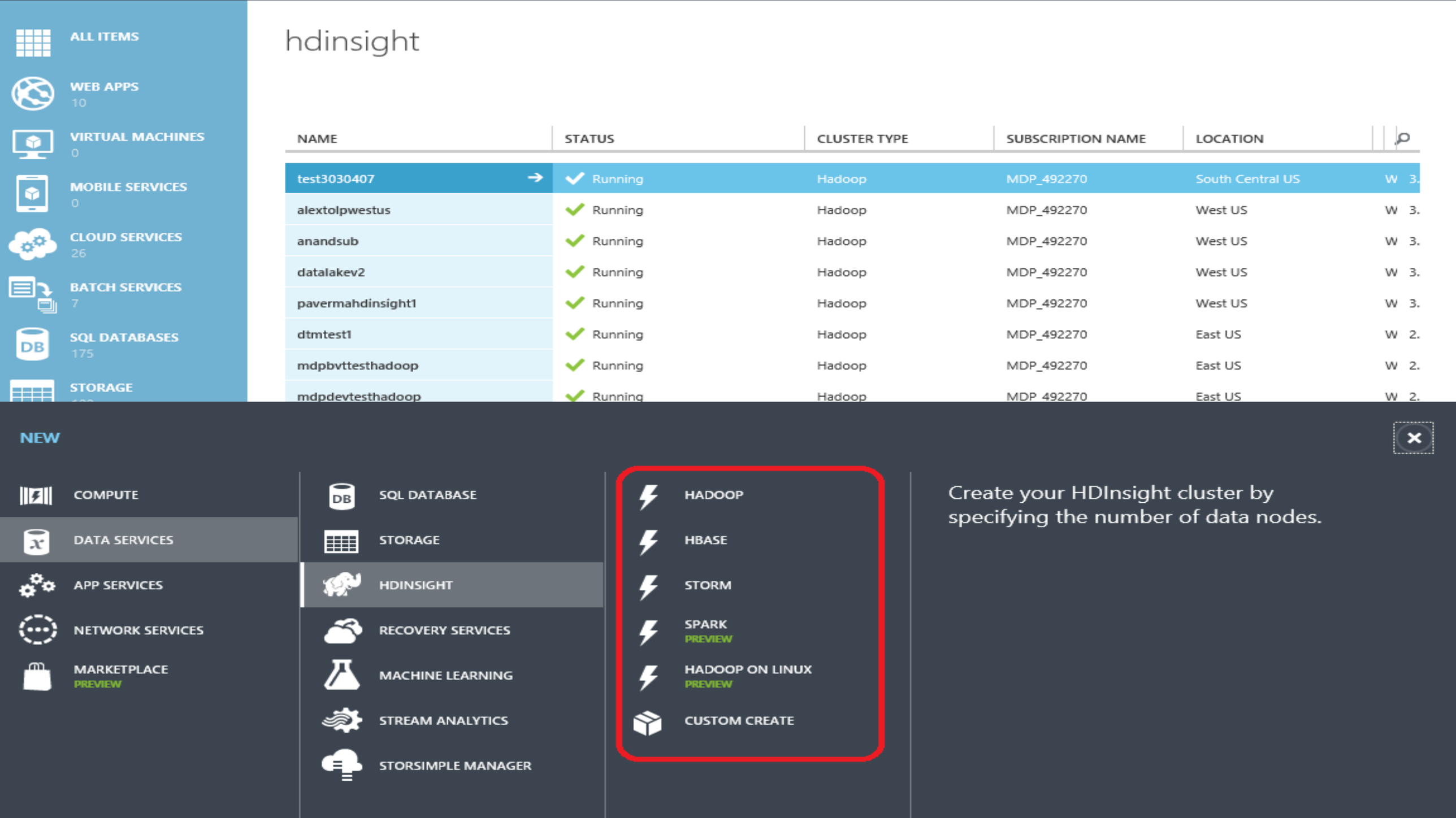
5

Openness

Embrace and extend

Agility is achieved when teams can pick the best-of-breed tools

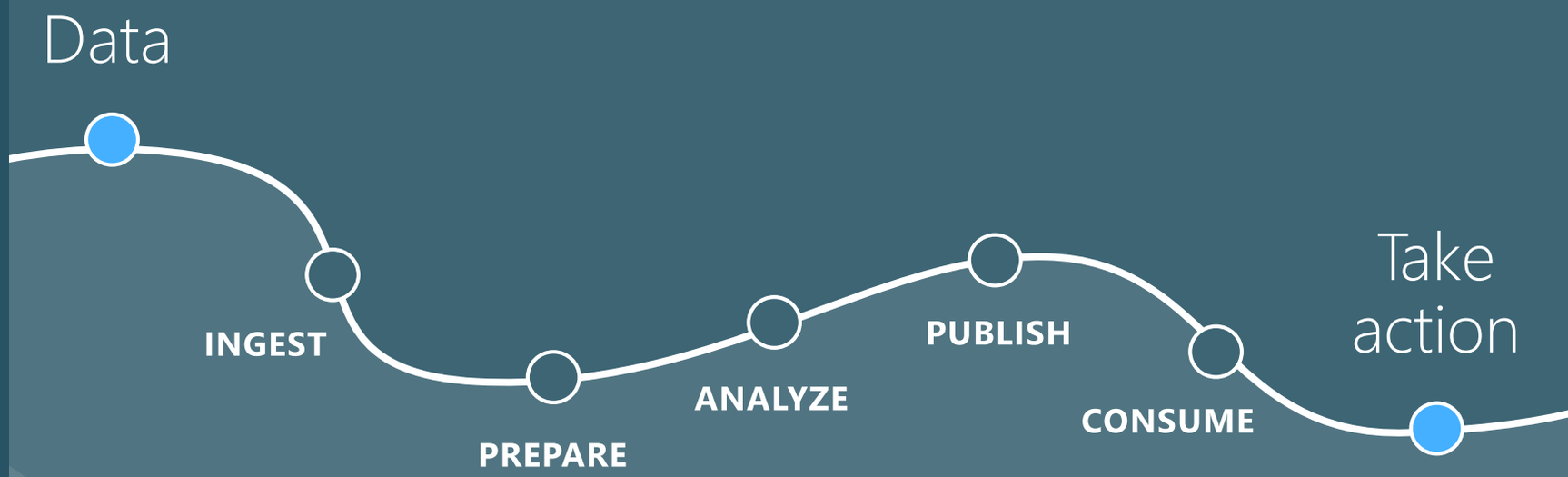
Leverage skills, extensibility and open ecosystem


ALL ITEMS
WEB APPS
10
VIRTUAL MACHINES
0
MOBILE SERVICES
0
CLOUD SERVICES
26
BATCH SERVICES
7
SQL DATABASES
175
STORAGE
hdinsight
NAME
STATUS
CLUSTER TYPE
SUBSCRIPTION NAME
LOCATION
test3030407
Running
Hadoop
MDP_492270
South Central US
alextolpwestus
Running
Hadoop
MDP_492270
West US
anandsub
Running
Hadoop
MDP_492270
West US
datalakev2
Running
Hadoop
MDP_492270
West US
pavermahdinsight1
Running
Hadoop
MDP_492270
West US
dtmtest1
Running
Hadoop
MDP_492270
East US
mdpbvttesthadoop
Running
Hadoop
MDP_492270
East US
mdpdevtesthadoop
Running
Hadoop
MDP_492270
East US
NEW
COMPUTE
DATA SERVICES
APP SERVICES
NETWORK SERVICES
MARKETPLACE
PREVIEW
SQL DATABASE
STORAGE
HDINSIGHT
RECOVERY SERVICES
MACHINE LEARNING
STREAM ANALYTICS
STORSIMPLE MANAGER
HADOOP
HBASE
STORM
SPARK
PREVIEW
HADOOP ON LINUX
PREVIEW
CUSTOM CREATE
Create your HDInsight cluster by specifying the number of data nodes.

6

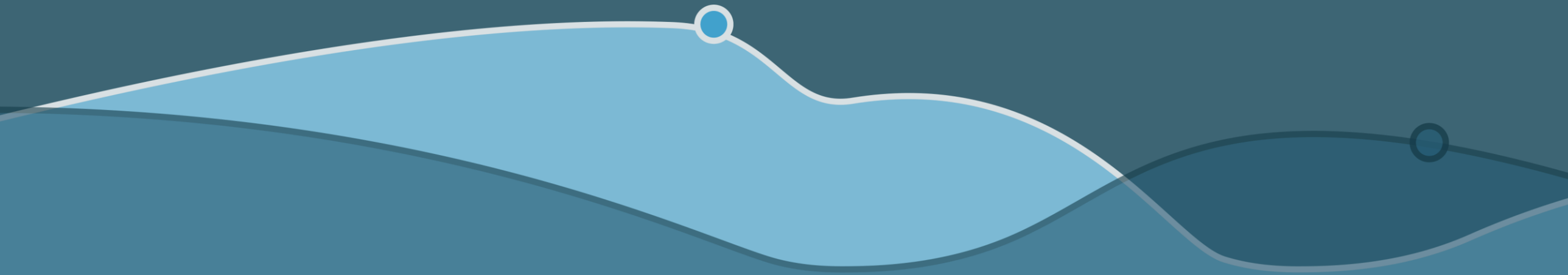
End-to-end

From data ingest to action and presentation



Cortana Analytics

Video



Key takeaways

FULLY MANAGED

Focus on delivering the solution, not the infrastructure

ALL TYPES

Work with all types and volumes of data

OPERATIONALIZE

Move from ad hoc explorations to full operationalization using the same tools

Operationalize data movement training, scoring, and pipelines

SHARE AND COLLABORATE

Tie the disparate data ecosystems together and share learnings

OPENNESS

Extend solutions using existing tools

Plug and play using the tools and languages you already know

END-TO-END

Enjoy an end-to-end experience from get data to presentation