



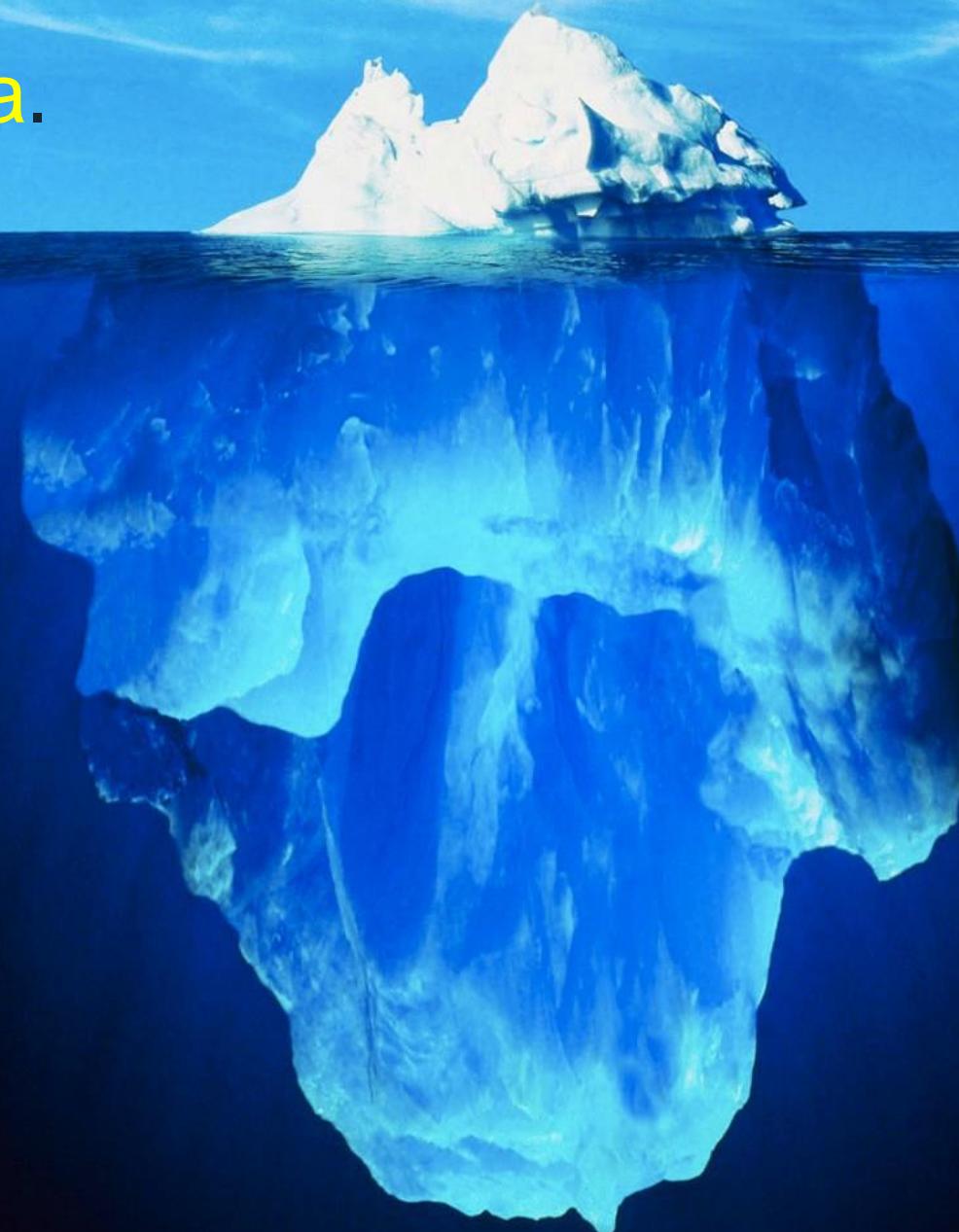
#ibmdevconnect

## **Watson Discovery – Actionable insights**

# Agenda

- Introduction
- Watson Discovery – Why, When?
- How? - Walkthrough – Exploring Discovery - Bluemix Console – Discovery Tool and API
- Where? Possible uses
- Summary

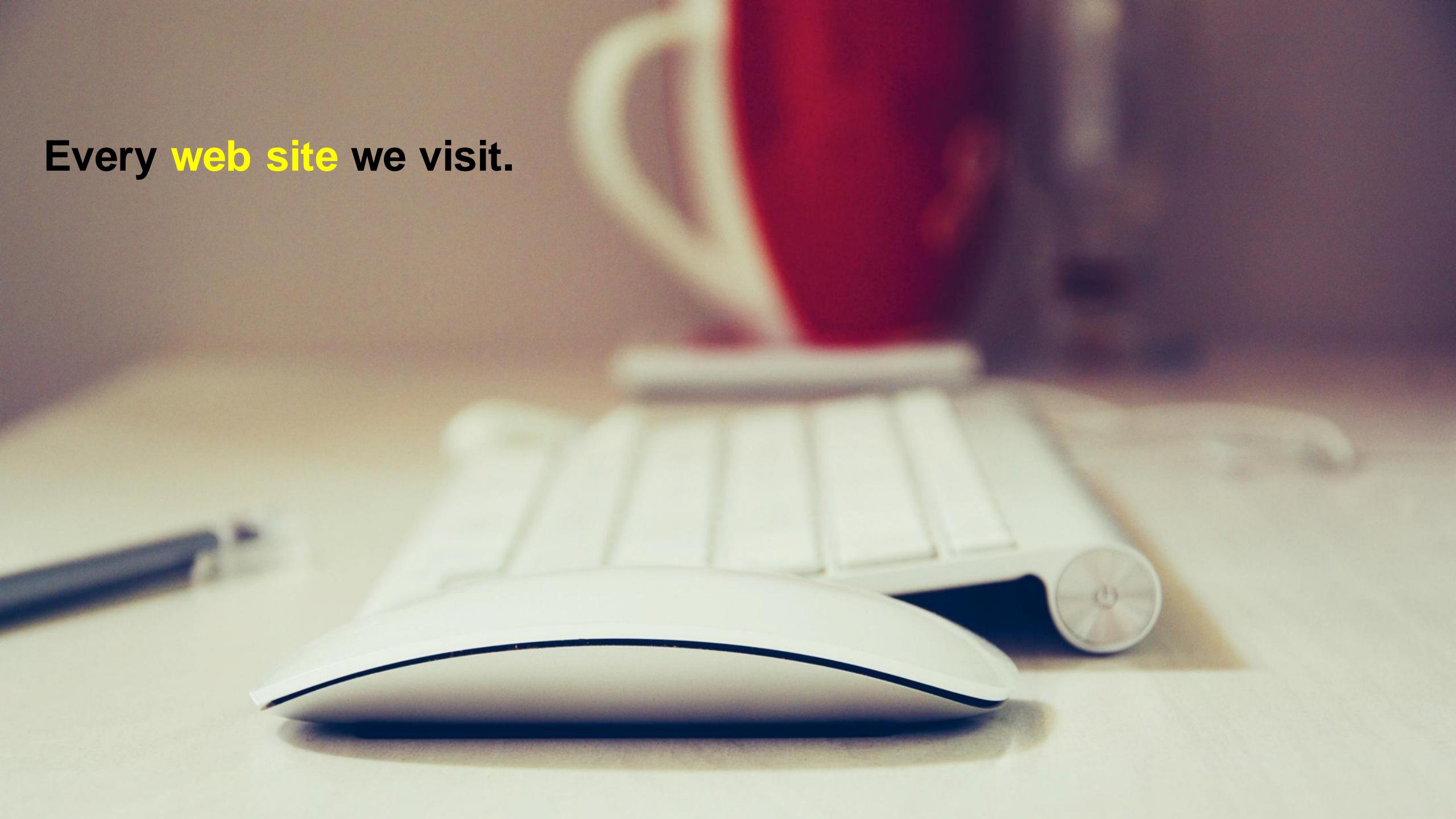
The world is  
swimming in data.



Nearly everything we do  
generates or consumes  
data.



**Every web site we visit.**



Every **item** we purchase.



Every navigation route  
we take.



# We're accumulating more data than ever before



Yet the vast majority of  
data is **locked away** in  
silos.



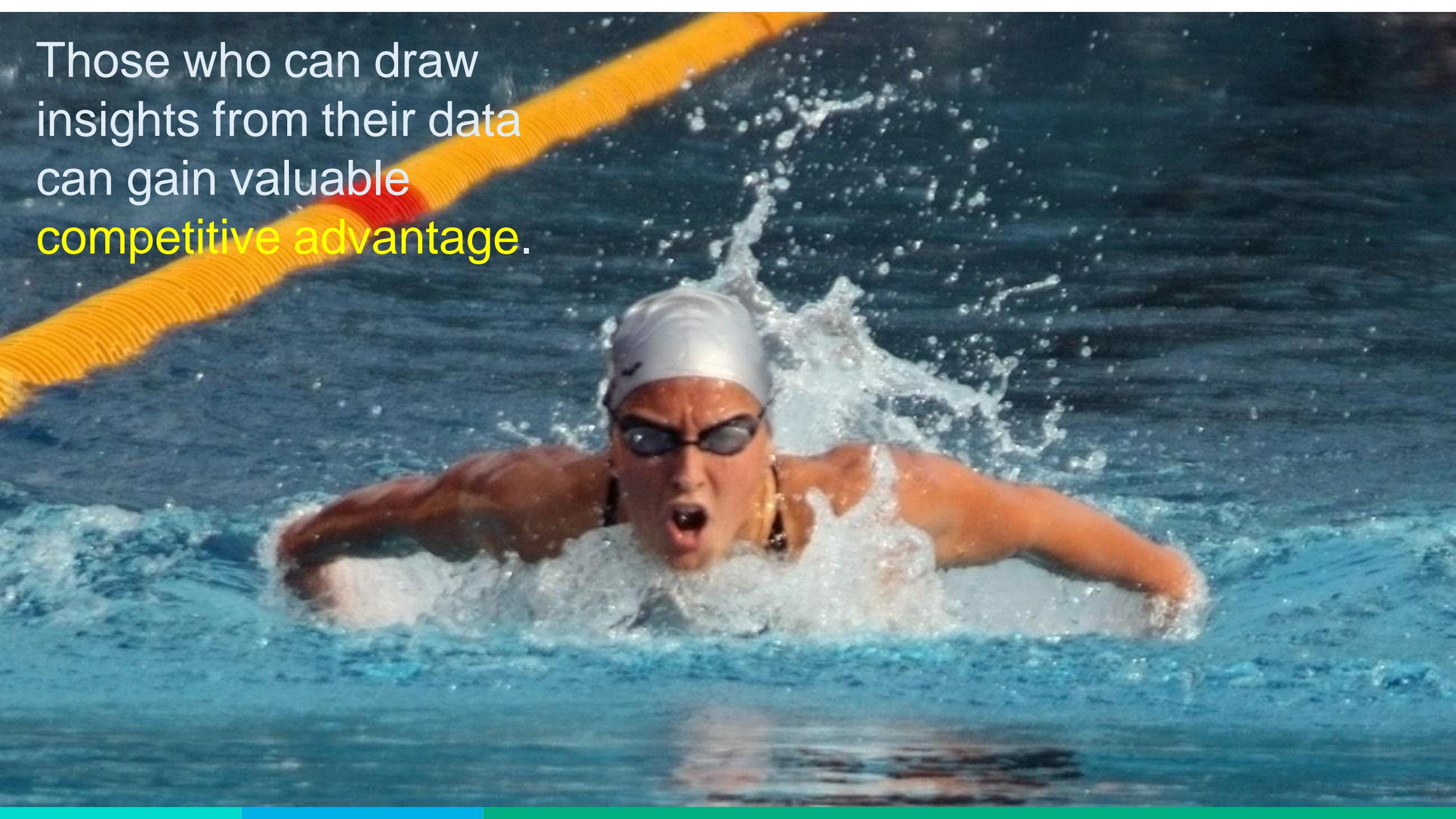


Worse yet,  
most data is  
unintelligible to  
traditional BI  
systems.  
It is unstructured.

Watson **sifts** through  
massive amounts of data.



Those who can draw insights from their data can gain valuable competitive advantage.



Cognitive  
technologies unlock  
insights in your data.

IBM Watson



## Discovery – Why and When?

Have **large volumes** of **unstructured data** with valuable insights locked away?

Need **instant access** to pre-enriched **data sources**?

Need a solution that matches your **cloud based** business strategy?

Need **cognitive insights** like **sentiment analysis** integrated to **company's applications**?

Want to **free up** your **data analysts** from tedious curation work?

Want to add **deep insight answers** to your **chatbots**?

Introducing Watson  
Discovery Service.



# Watson Discovery Service

A powerful set of APIs built to help developers find value in your data, fast.



Many data scientists spend as much as **80 percent\*** of their time on so-called “data janitor” work – **collecting, cleaning and organizing** data sets.

**Developers** face numerous **barriers** that slow them down when building applications to leverage unstructured data:

Massive amounts of diverse and dispersed unstructured content

**NLP APIs** are **difficult** to integrate

Analysis requires **complex queries** and manual filtering.

Watson Discovery packages cognitive technologies together in an easy-to-use end-to-end system to focus on specific business problems.

It **simplifies** the infrastructure, scale and algorithm challenges associated with enriching and analyzing large data sets.



With Watson Discovery, go from **data** to **insights rapidly** and with less effort

Questions

Documents/Data



??



Data Sources

- Your private data
- Internet sources
- 3<sup>rd</sup> party sources

**“What are the top terms mentioned by customers in support interactions”**

**“How do I setup a new connection for a customer with a locked accounts”**

**“How many times was my campaign mentioned over the last 30 days”**

Apply [cognitive enrichments](#) to your data and extract insights using a powerful and flexible query language in real time



## Automated Data Ingestion

Automate ingestion using the APIs, web upload, or data crawler, and feed through Document Conversion to deal with multiple file types.

## Integrated Enrichment

Rapidly setup advanced Natural Language Processing steps and custom models to extract entities, relations, keywords, sentiment, etc.

## Diverse Content Sources

Create one-of-a-kind applications that leverage data from the pre-enriched news sources and private content from across the enterprise.

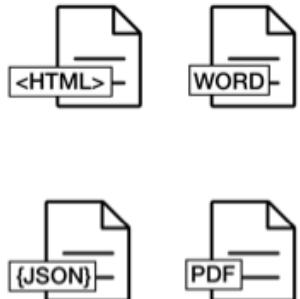
## Simplified Query Language

Perform multiple query types including Boolean, filter, and aggregation queries to discover patterns, trends, and answers.

# The Automated Data Pipeline

## | Data

Private data



## | Ingestion

**Convert and enrich** by leveraging Watson APIs to add NLP meta data to your content, making it easier to explore and discover insights

**Clean and normalize** through an automated processing of NLP results, improving data quality

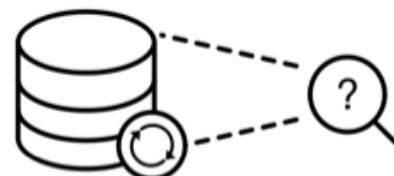
## | Storage

Normalized data is **indexed into a collection** as part of your **environment** in the cloud



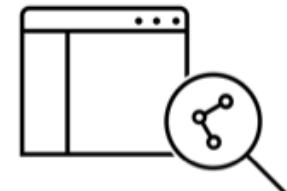
## | Query

**Understand** data faster, create better hypothesis and deliver better outcomes



## | Output

Actionable insights into your app



Discovery Service lets you build an automated data pipeline once, and then add it to existing solutions

Watson Discovery  
enriches your  
documents with natural  
language understanding.



Watson Discovery  
**indexes** your data,  
making it **searchable**.

Watson **Discovery** gives  
developers the  
**fastest path** to build cognitive apps  
that extract value  
from structured and unstructured data.



A new lens for  
understanding  
your **data assets**.



We are at the beginning of  
the Cognitive Era.

Watson Discovery Service  
is available via the [Watson](#)  
[Developer Cloud](#).

Try it today!



# Simple Exploration & Discovery Examples



## Problem

When you're discovering a new domain - like "Machine Learning"-can be difficult to find the right resources. More particularly, reference **textbooks** are tricky to identify because you have to find one that is reliable, but that also corresponds to your level of understand and that would help you grow and learn.

Also, the format of the book itself can be mismatching your specific learning style.

## Solution

Using **comments** left online on **textbooks**, how could you identify the right textbook for you?

Also, how could you let someone different than you identify its best textbook?

What if we could mine those comments to **extract features** of these textbooks, but also what **people have felt** using them,

how **positive** they are about these features?

**Data:** Provide a collection of 321 amazon reviews on 35 different textbooks.  
These reviews are only a very small extract from the [Amazon Reviews dataset](#).

# Steps

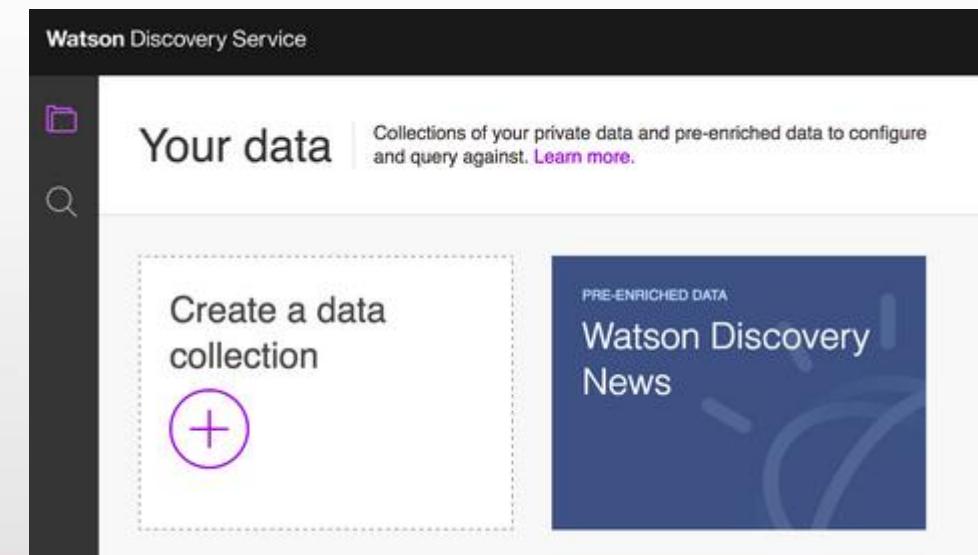
1. Introduction to Bluemix
2. Create an instance of Discovery
3. Create, configure and upload a collection of documents
4. Write your first queries using Discovery

# Create collection

Upload documents to collection. Each document is a JSON format.

```
{  
  "asin": "0070428077",  
  "price": 50.27,  
  "title": "Machine Learning",  
  "reviewText": "There is no real formatting in this book, just one big slob of text. It is very unreadable and boring"  
}
```

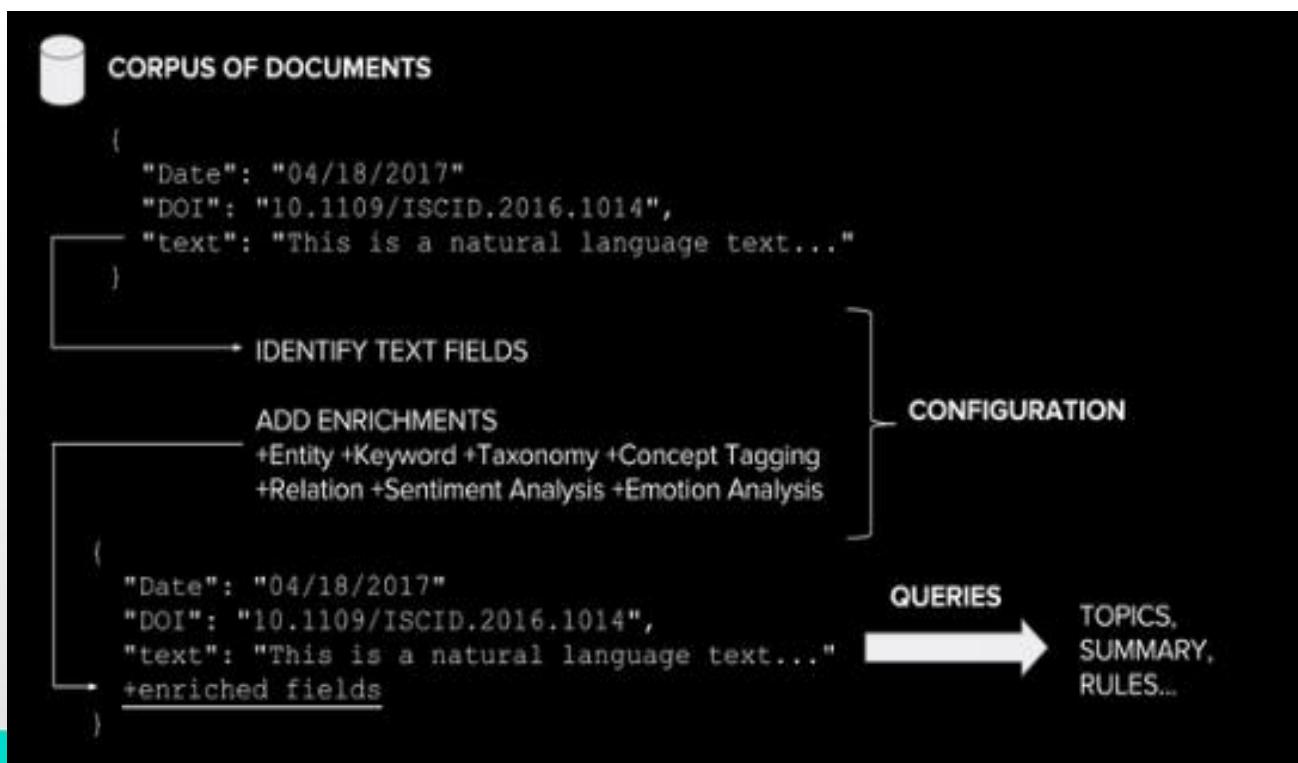
These are set of collection of Amazon reviews of different textbook.



# Create new configuration

The three steps of customizing a configuration file are: **Convert**, **Enrich**, and **Normalize**.

- 1 [Converting sample documents](#)
- 2 [Adding enrichments](#)
- 3 [Normalizing data](#)



<b>Entity Extraction</b> Extracts people, companies, organizations, cities, geographic features, and more from this field. <a href="#">Learn more</a> Added!	<b>Keyword Extraction</b> Determines important keywords in this field, ranks them, and optionally detects the sentiment. <a href="#">Learn more</a> Added!
<b>Taxonomy Classification</b> Classifies this field into a hierarchy of categories that's five levels deep. <a href="#">Learn more</a> Added!	<b>Concept Tagging</b> Identifies general concepts that aren't necessarily directly referenced in this field. <a href="#">Learn more</a> Added!
<b>Relation Extraction</b> Parses sentences into subject, action, and object form and returns additional semantic information. <a href="#">Learn more</a> Added!	<b>Sentiment Analysis</b> Identifies the overall positive or negative sentiment within this field. <a href="#">Learn more</a> Added!
<b>Emotion Analysis</b> Analyzes the emotions (anger, disgust, fear, joy, and sadness) in this field. <a href="#">Learn more</a> Added!	

# Upload Documents in collection

From the original documents formatted in JSON,

Discovery has enriched each review by 7 fields:

concepts, entities, relations, taxonomy, keywords, docSentiment, docEmotion.

## Queries

Time to query the collection of documents/reviews processed by Discovery.

INPUT	TEXT
query	...
count	...
filter	...
aggregations	...
return	...

INPUT	TEXT
query	machine learning
count	...
filter	enriched_text.docSentiment.type:positive
aggregations	term(title)
return	title

# Using API's to access

APIs ▾ Docs Developer Tools Starter Kits Community

Delete a configuration  
Test your configuration...  
Test your configuration  
Collections  
Create a collection  
List collections  
List collection details  
Update a collection  
List collection fields  
Delete a collection  
Documents  
Add a document  
Update a document  
List document details  
Delete a document  
Queries  
Query your collection  
Query ingestion noti...  
Training data  
Add a query to the tr...

REQUEST

Name	Description
environment_id	The unique identifier for this environment. query string required
collection_id	The unique identifier for this collection. query string required
version	The release date of the version of the API you want to use. Specify dates in YYYY-MM-DD format. The current version is 2017-06-25. query string required
query_options	Specifies the parameters of the query. In Java and Node, specify this as a query string. In Python, specify this as a Python dictionary. query string   dictionary

Curl Java Node Python

Example request

```
Discovery discovery = new Discovery("2017-06-25");
discovery.setEndPoint("https://gateway.watsonplatform.net/discovery/api/v1");
discovery.setUsernameAndPassword("{username}", "{password}");
String environmentId = "{environment_id}";
String collectionId = "{collection_id}";

QueryRequest.Builder queryBuilder = new QueryRequest.Builder(environmentId,
, collectionId);
queryBuilder.query("{field}:{value}");
QueryResponse queryResponse = discovery.query(queryBuilder.build()).execute();
```

# Watson Discovery Service – Where Possible uses



## Customer Service

Add an app to your support workflow that finds and delivers answers to complex customer questions, in context



## Social Media and News Channels

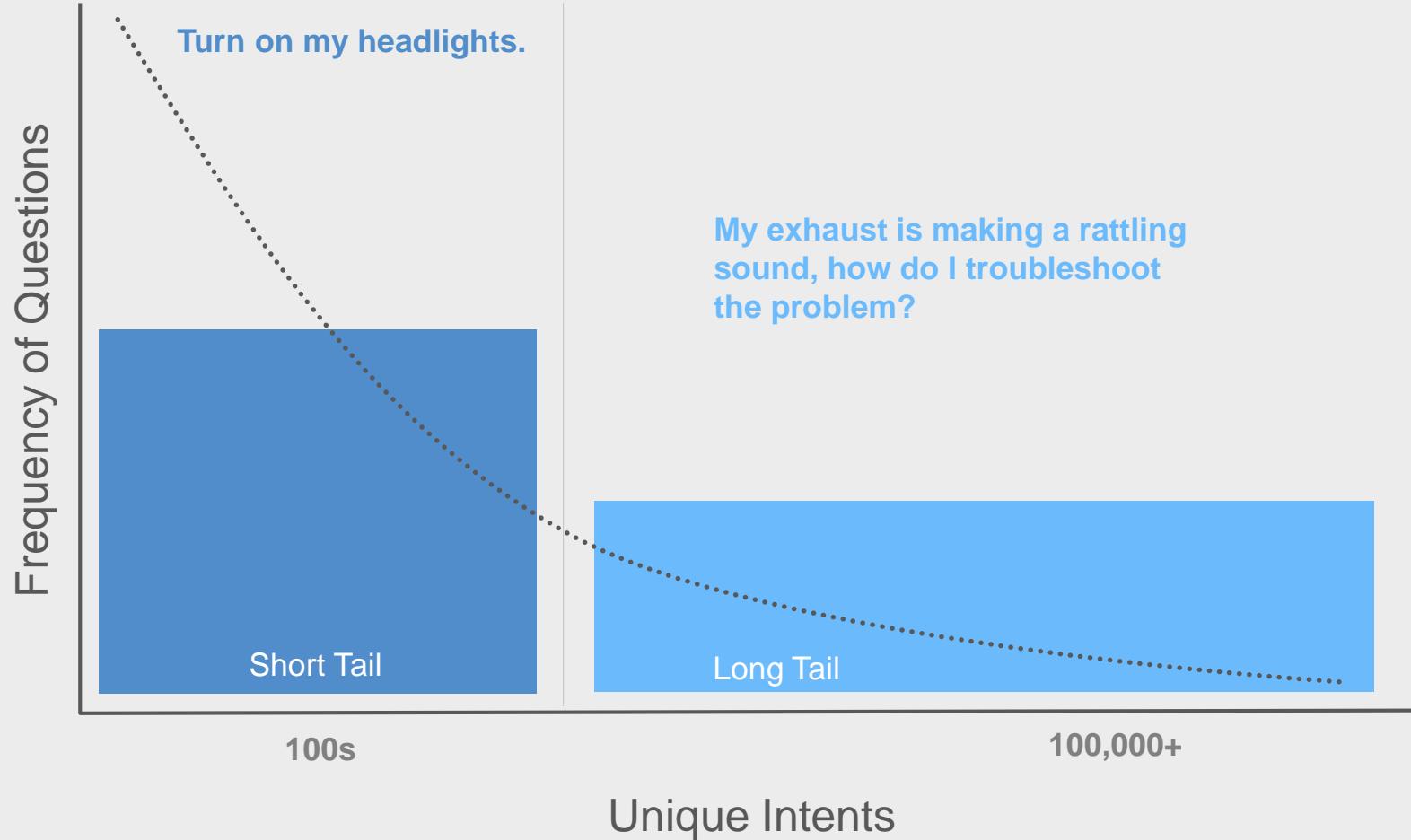
Add an app to get real-time visibility to perceptions about products, services and brands, and then correlate today's results with historical results



## Research Insights

Add an app to extract consolidated industry and domain-specific insights from vast numbers of research and other documents

# Combine Watson Discovery with Watson Conversation for question distribution



## Watson Conversation



Here Watson uses reasoning strategies that focus on the language and context of the question.

## Watson Discovery Service



Here Watson uses reasoning strategies that focus on identifying the most appropriate answer.

## Summary - Watson Discovery Service

- (Unstructured Data) - Explore and **discover insightful information** hidden in **unstructured data**.
- (Data Sources) - Data can come from public, third party or your own proprietary sources
- (Cloud) - WDS is a cloud service available on IBM Bluemix.
- (Applications) Leverage Watson SDKs, web-based tooling and the WDS APIs to bring cognitive insights to your custom applications.
- (Enrich) Let WDS crawl, convert, enrich and normalize data- spend your time exploring.

WHAT  
DO  
YOU  
THiNK?



Please give us your valuable feedback  
<http://ibm.biz/devconnect17>



# Stay Connected and continue coding !



Code & instructions available here

<https://github.com/IBMDevConnect17/>



Join developerWorks India Community

<https://developer.ibm.com/in/>

Check out the cool developer journeys

<https://developer.ibm.com/code/>



Join our Slack team and stay in touch with the experts

<https://ibmdevconnect.slack.com>

Send in your request to -

<http://ibm.biz/slackrequest>



**Join** our Meetup groups

Mumbai :

<https://www.meetup.com/Cloud-Mumbai-Meetup/>

Hyderabad:

<https://www.meetup.com/Hyderabad-Cognitive-with-Cloud>

Bangalore :

<https://www.meetup.com/IBMDevConnect-Bangalore>