

# CIO Summit: *Accelerating Innovation w/ Data Science*



**Carlo Appugliese**

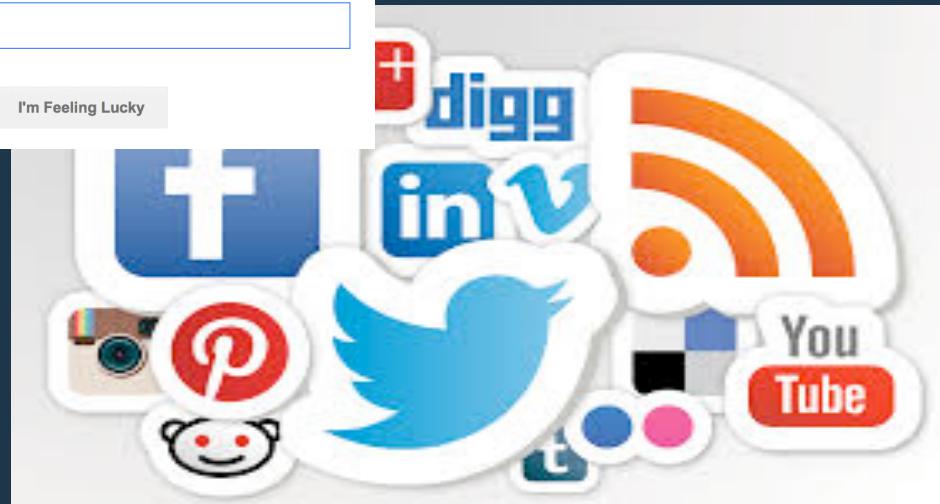
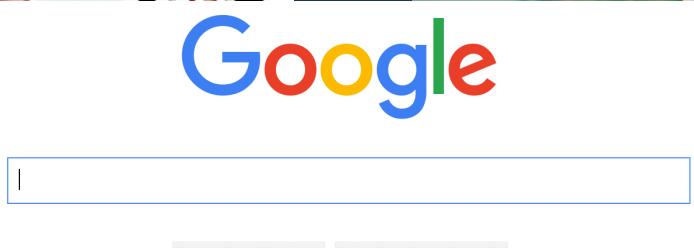
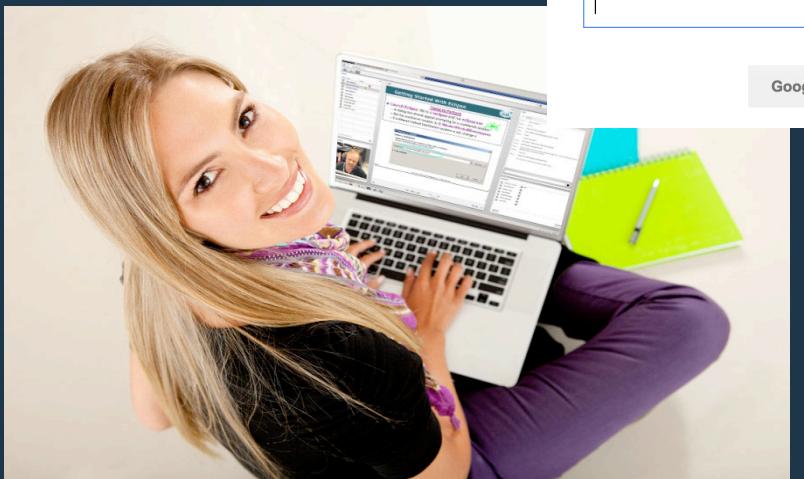
Data Science Evangelist

IBM Analytics, Watson and Cloud

May 2017

IBM

# The digital age has changed the way we Live, Play, Learn and Work...



# Transformation is Critical

Digital businesses  
are disrupting  
industries and  
professions



# Alibaba

## E-Commerce

Alibaba.com – B2B

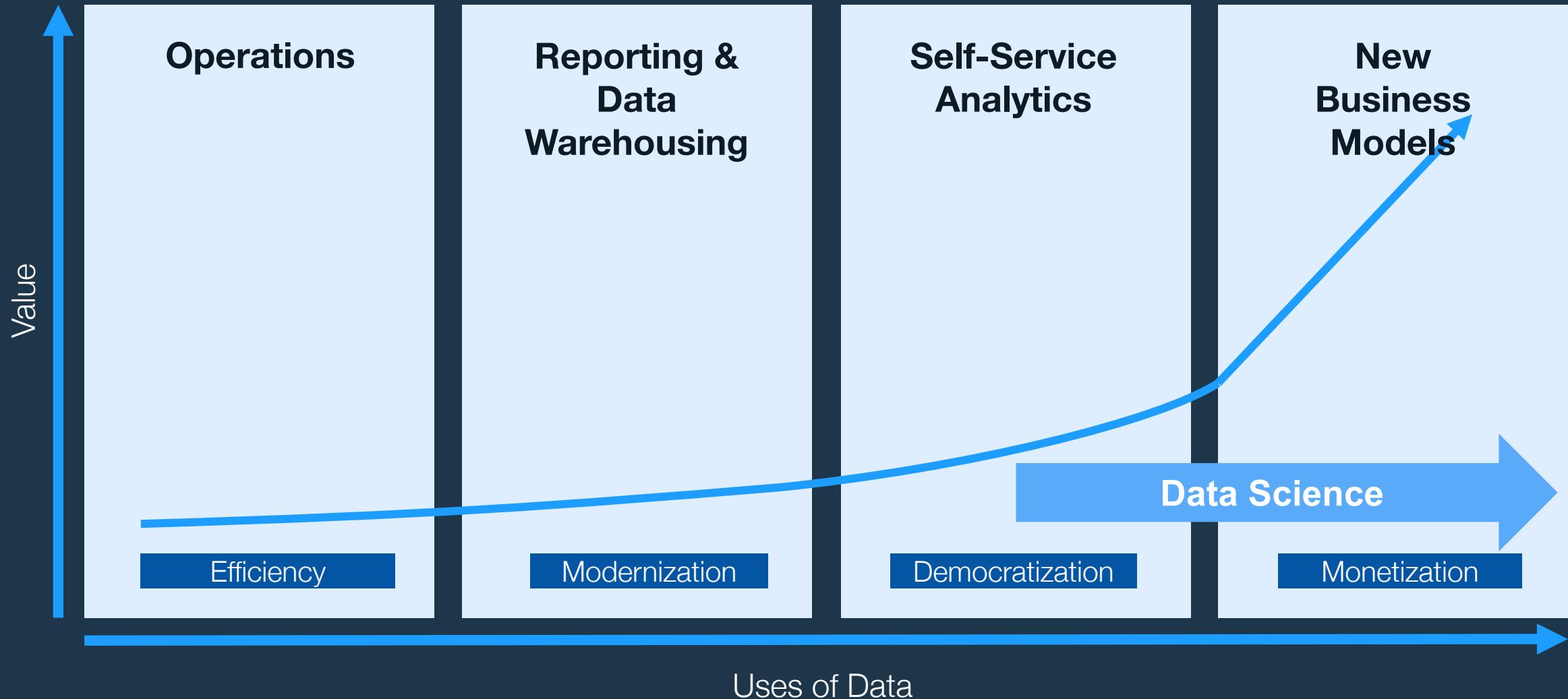
Taobao - C2C

1688.Com – B2B

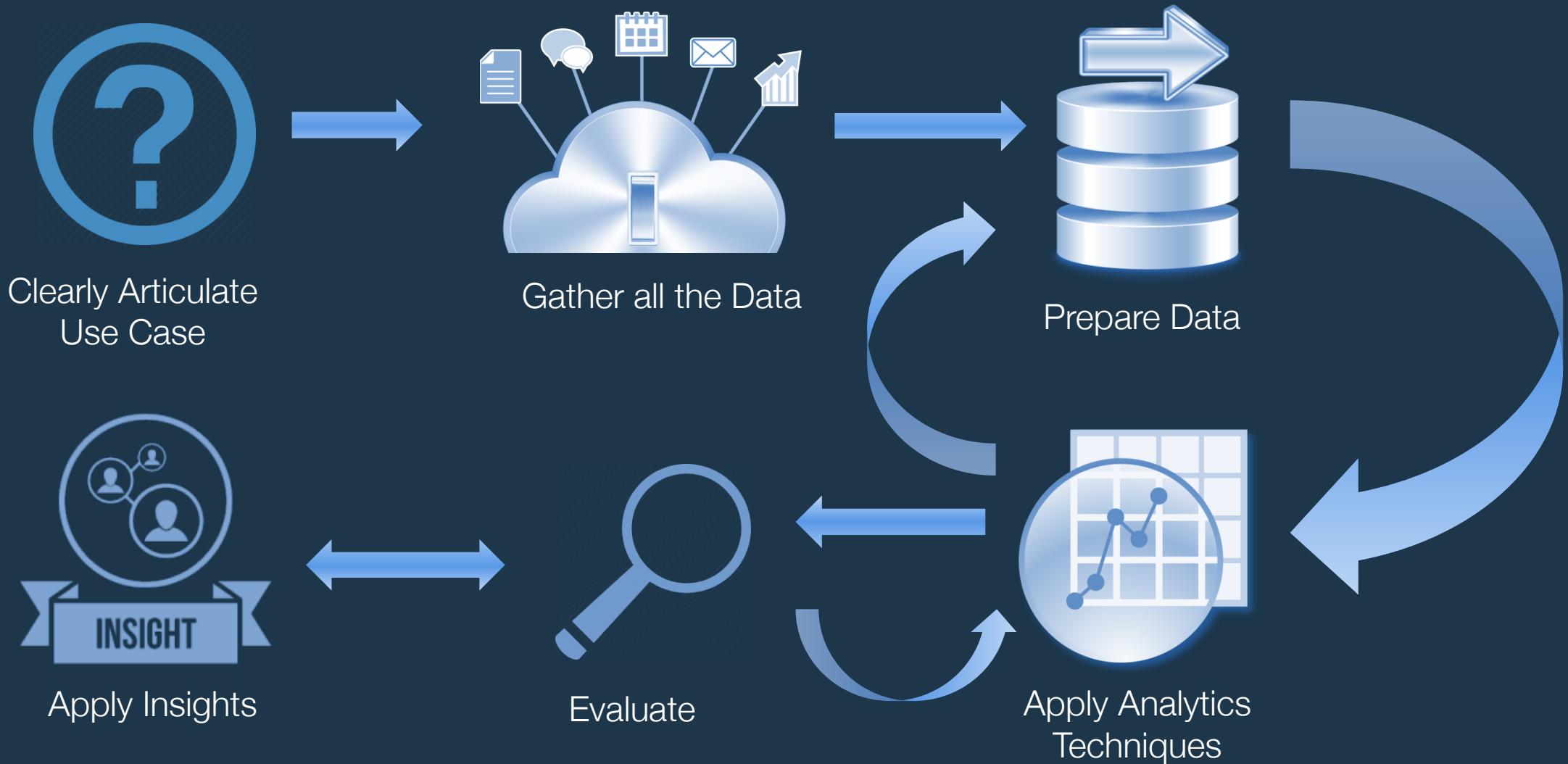
Tmall – B2C



To Transform, you have to make the shift to a  
**Data-Driven** Organization and **Unleash** your data...



# Steps to Transformation



# But Familiar Approaches **Hold The Business Back**

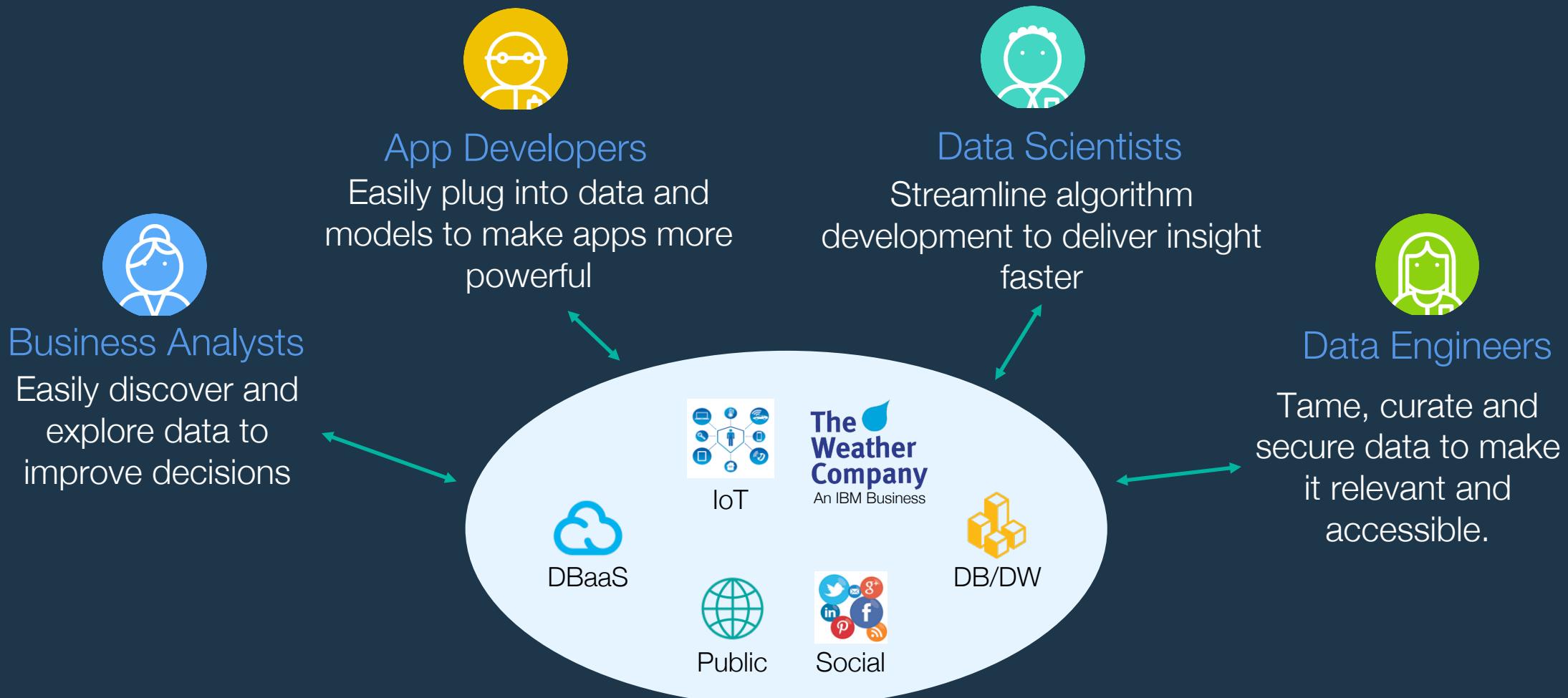
	<b>Approach Found Lacking</b>	<b>What's Needed Now</b>
	Limited to traditional data repositories	Ubiquitous data access
	Ungoverned and unknown data environments	Trusted data recognition
	Silo data professionals	Collaborative user interaction
	Silo analytic techniques	Analytics everywhere
	Slow deployment of new insights to the business	An “InsightOps” approach

# Our **Mission**

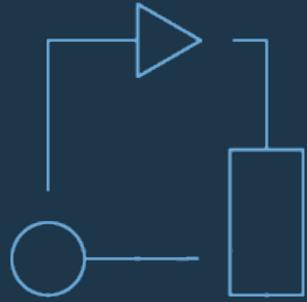
Make data  
simple and  
accessible to  
all



# The Data Professional is at the forefront of transformation to a Data-Driven Organization



# Our Strategy



## Platform

Leveraging Apache Spark

Composable, hybrid-cloud services

Intelligence to accelerate insight

Self-service for all user types

The IBM  
**Watson Data Platform**



## Ecosystem

Best-in-class partners

Open source-based  
services

Open Standards  
Community Accepted  
Trusted **Partners**



## Method

Understand what works and  
what doesn't

Get started fast

Immediate business value

The IBM  
**DataFirst Method**

A New **Platform**



[ibm.co/MakeDataSimple](http://ibm.co/MakeDataSimple)

# IBM Watson Data Platform



## Intelligent by Design

with intelligence services ranging from cognitive APIs to automatic entity analytics infused in every aspect of the making/creation process



## Lead the way in collaboration

so data scientists, developers and data engineers are natively supported in their tasks and are working together to deliver an intelligent application



## Self-service trusted access to data

giving data professionals the freedom of access to the data they need with the trust that the business expects



## Streaming with real-time analytics

to support modern application demands with first class streaming analytics that augments batch



## Open and Extensible

achieving scale and the network effect by attracting partners to build on and extend through APIs and toolkits

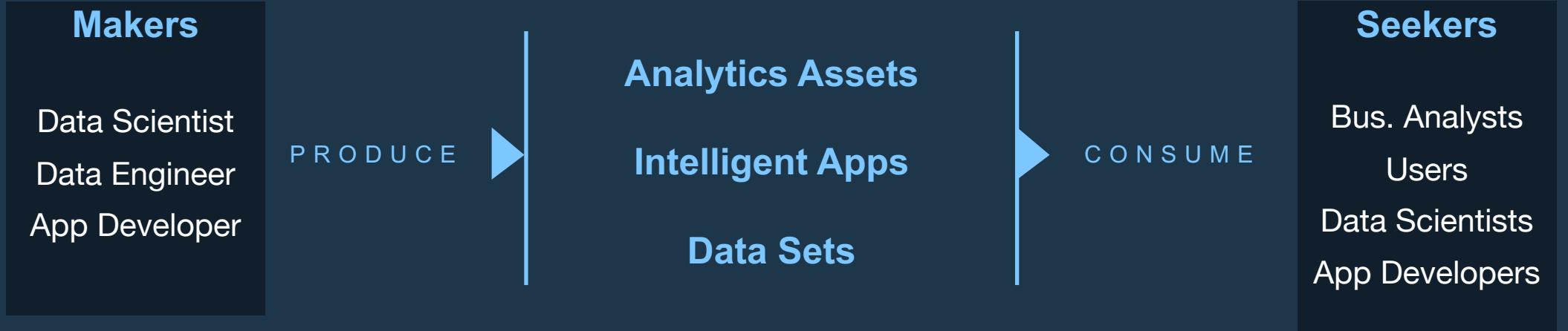


## Be the premier content hub

of open, rich 3<sup>rd</sup> party content and curated IBM assets that enhances the types of insights users are able to derive

# IBM Watson Data Platform

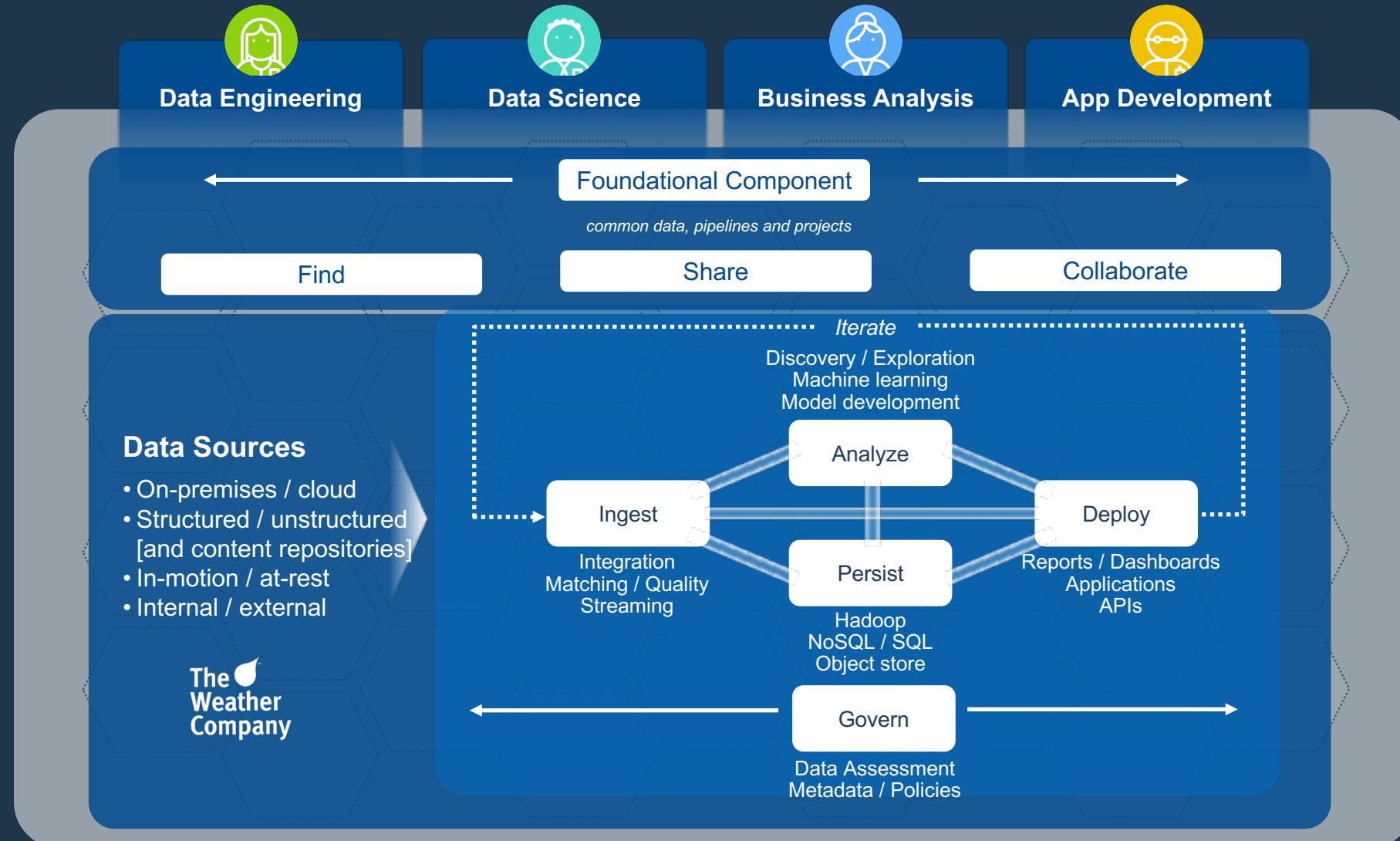
Where teams put data to work.



The **Watson Data Platform** is a fully managed data and analytics platform that enables makers of analytics assets and intelligent applications to efficiently create valuable assets that can be commercialized and adopted. It facilitates the consumption of these assets and enhances the making process through critical content and intelligent services.

Watson Data Platform enables the ingest of data from any source, the persistence and management of the full lifecycle of data, the processing of the data to derive insight and intelligence, that will power the next generation of cognitive systems.

# IBM Watson Data Platform



# A comprehensive **enterprise cloud platform**, differentiated with **Watson** and vertical capabilities

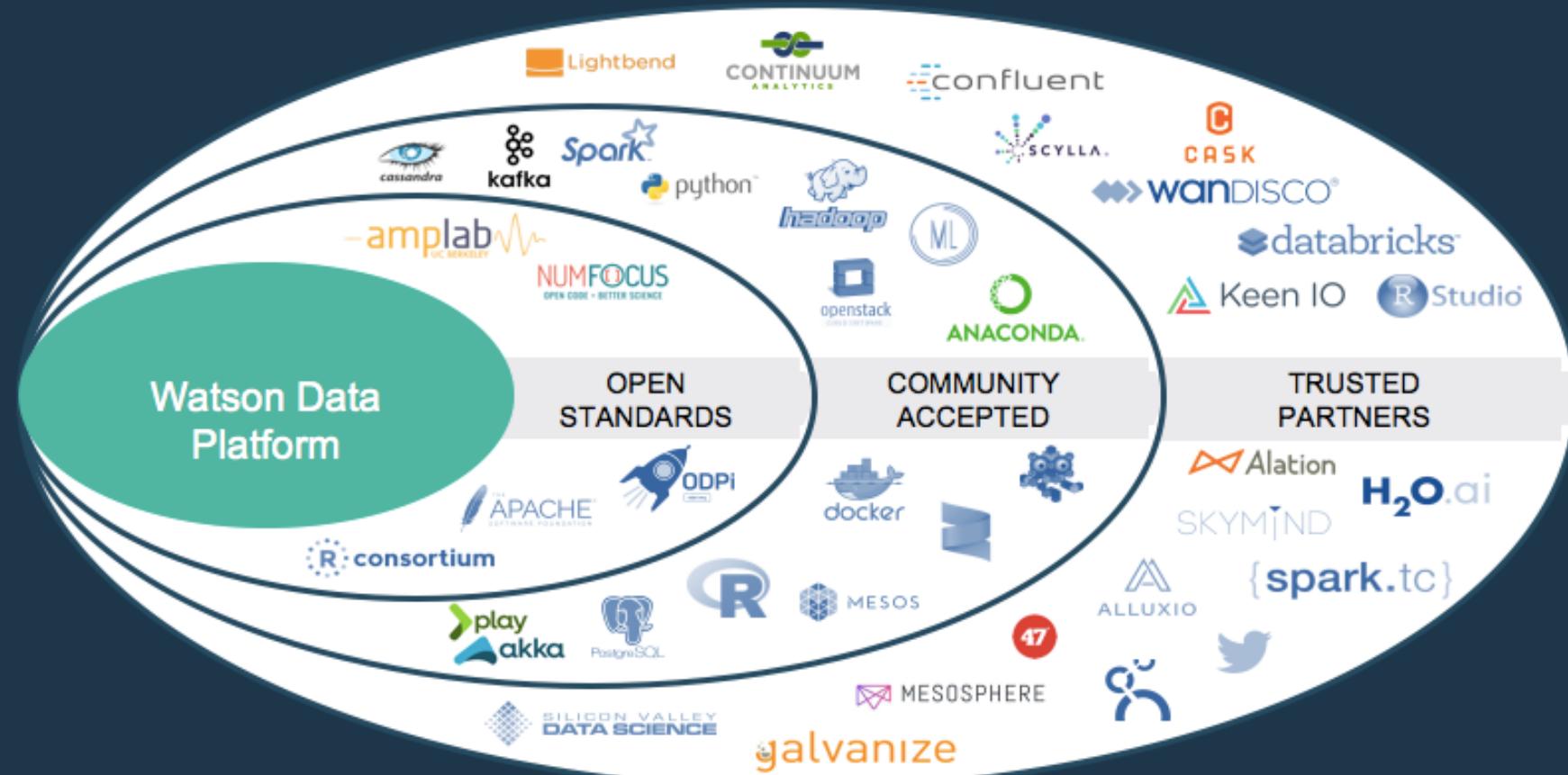


# A New **Ecosystem**

The background features a dark teal gradient. Overlaid on it are two glowing blue bands. The upper band is a dense grid of small, glowing blue dots, while the lower band is a solid, translucent blue color. These bands curve across the frame, creating a sense of depth and motion.

# The Ecosystem

## Open Standards, Open Community

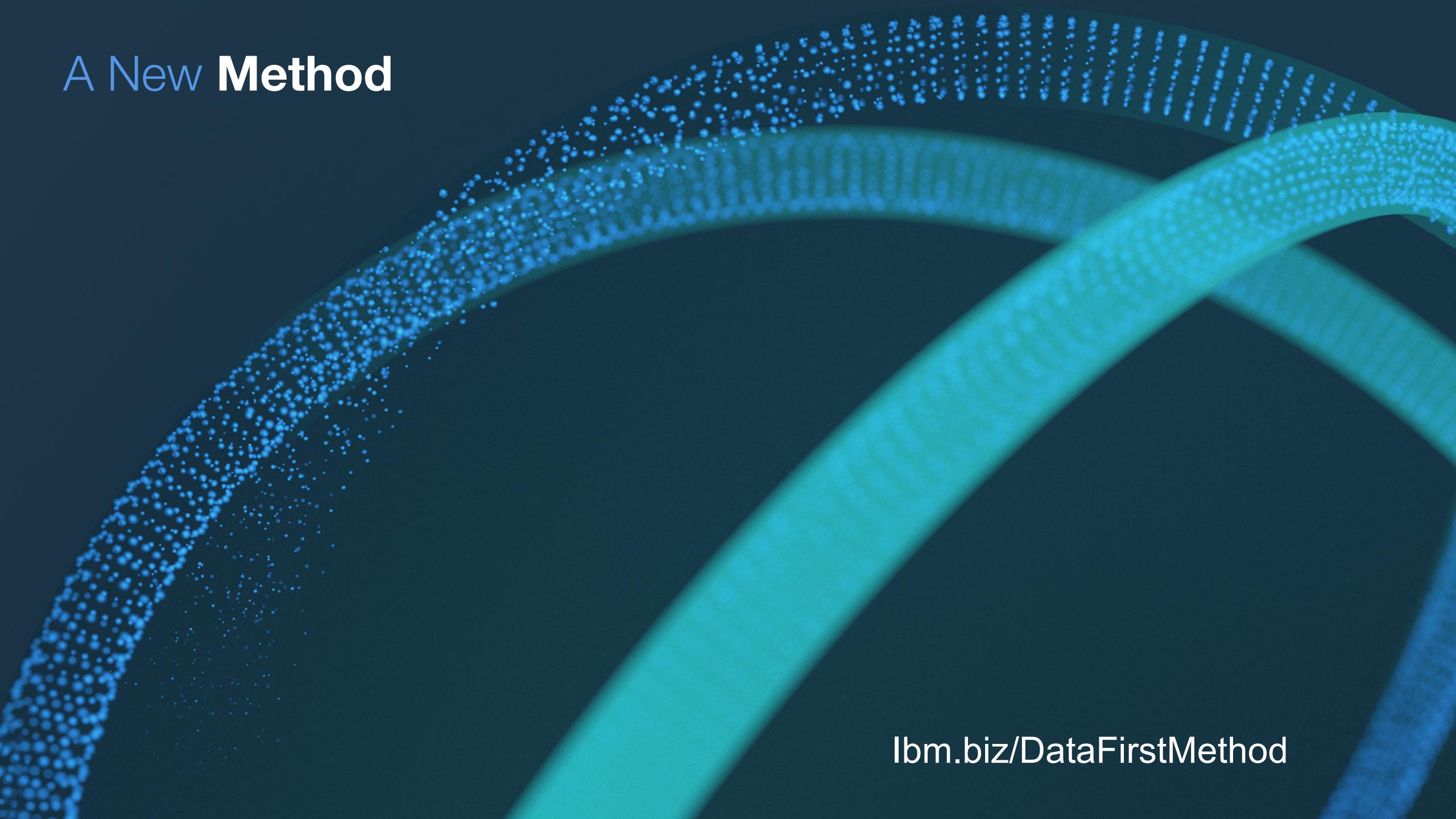


# IBM Open Source

- **Founding Apache Software Foundation member**
- **Contributions, contributions, contributions...**
  - #1 Cloud Foundry and LARGEST world deployment
  - #1 in Spark machine learning and #2 in Spark overall
  - #4 OpenStack
  - #3 Linux
  - #1 Eclipse open source IDE
  - Founding Member: AMPLab, Node.js Foundation
  - Extensive investments in Spark and Hadoop
  - Spark Technology Center @ [www.spark.tc/](http://www.spark.tc/)
- **Decades of lineage and contributions**
  - CouchDB
  - Linux, Derby, Geronimo, Jakarta, etc..
  - Eclipse founded by IBM
  - Toree, Quarks, +++
  - Significant Lucene contributions (ILEL), Text (UIMA)
  - DRDA, XQuery, SQL, XML4J, XERCES, Java, Linux, +++



# A New **Method**

The background of the slide features a dark teal gradient. Overlaid on this are two prominent, glowing blue bands. These bands are composed of numerous small, white, glowing dots, giving them a textured, digital appearance. One band curves from the bottom left towards the top right, while the other is positioned above it, curving in a similar but slightly different path. The overall effect is one of motion and data flow.

[ibm.biz/DataFirstMethod](http://ibm.biz/DataFirstMethod)

The IBM **DataFirst** Method helps you transform your business on your journey to become a cognitive enterprise by gaining the most value from your data.

# The IBM DataFirst Method: **What**



## **Focuses on analytics & cloud journey**

- Identifying and documenting gaps and plan to address (products, services)

## **Designs a solution to a specific business problem**

- With a roadmap to the future
- Leveraging IBM's proven workshops, assets and expertise

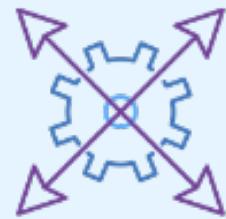
## **Drives to a successful production implementation**

- Leveraging IBM Analytics offerings with IBM & delivery partner expertise
- Delivering and providing value against the established business case

# The IBM DataFirst Method: **How**

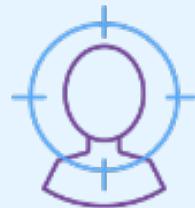


## Start Anywhere



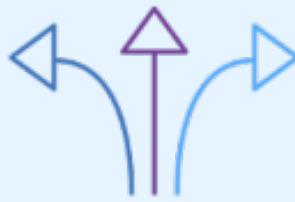
Focus on your  
largest business  
opportunities

## Fill the Gaps



Strategy.  
Expertise.  
Skills.

## Build Value at Every Step



Become a data  
driven culture, one  
initiative at a time.

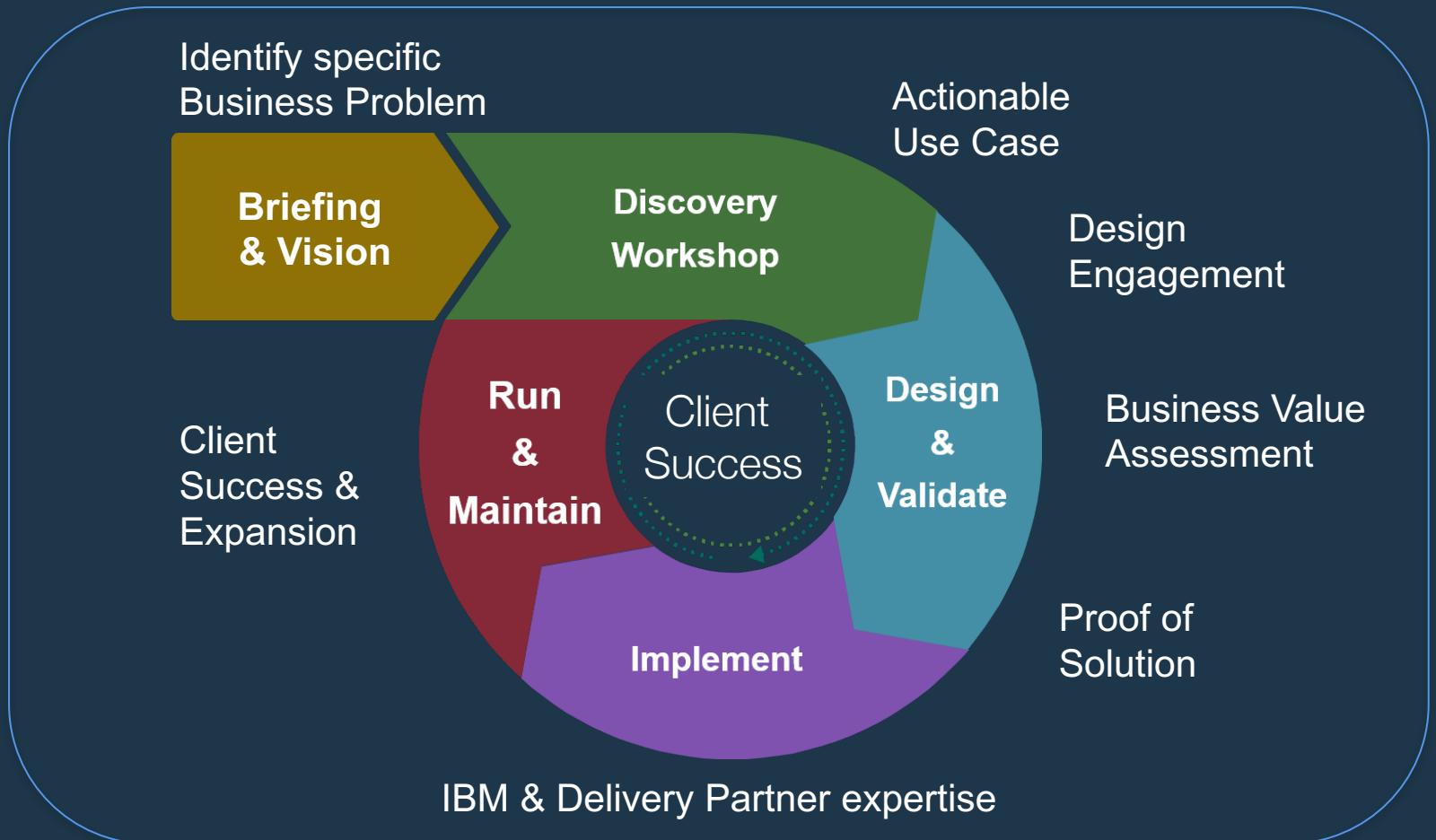
# The IBM DataFirst Method: **Why**



**The DataFirst Method will...**

1. Accelerate time to value through **agile sprints**
2. Mitigate risk for cloud-first **solution design** and **implementation**
3. Establish business case for making an investment to advance **data-driven transformation**

# The IBM DataFirst Method: **Process**



# Runkeeper sets the pace for a personalized experience



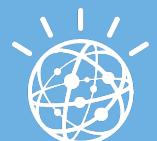
## Cloudant & Graph

Deliver a personalized customer experience



## DashDB

Uncover insights to deliver new services



## Watson Analytics

Gain a deeper understanding of your community



## Watson Personality Insights

Leverage cognitive capabilities to improve customer relations



*Runkeeper is now part of the ASICS family*

# Runkeeper

Fitness App Pumps up Customer Experiences

## Company Background

- iOS/Android mobile fitness app that guides and tracks user activity
- Provides motivation and allows users to monitor progress toward goals
- >50 million users, 3 million routes, ~200 countries, >100 third-party integrations

## Success Criteria

- Build more personalized interactions for users based on wealth of data
- Gain deeper understanding of personalities and preferences of a global community
- Increase customer loyalty and engagement to continue growing user base

## Solution & Results

- They created a more meaningful, immersive experiences for individual users:
  - **Cloudant** captures location and integrates with **Graph** to personalize route recommendations
  - **dashDB** uncovers insights using multiple sources to optimize app performance and drive new services
  - **Watson Analytics** provides deeper understanding of user communities based on social data
  - **Watson Personality Insights** uses social data to create personality profiles for market segmentation
  - **The Weather Company data** enhances context for the optimal conditions to run specific routes

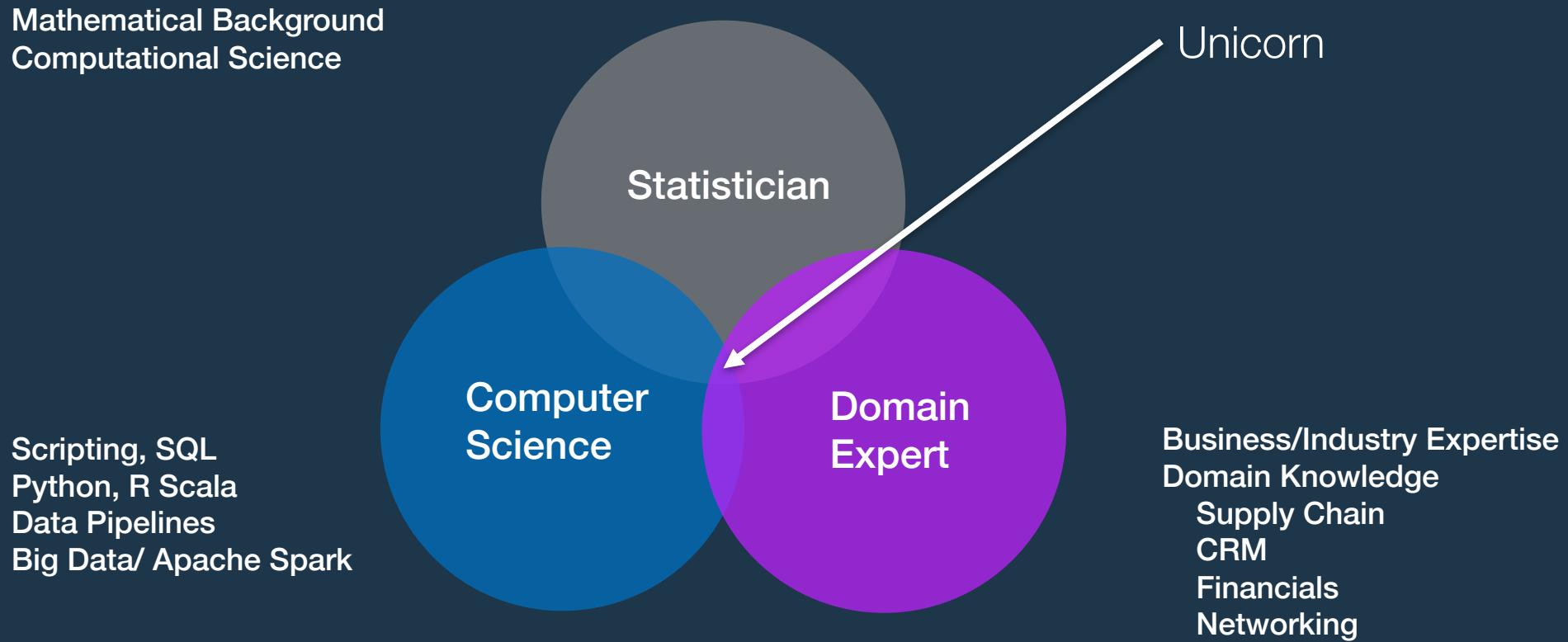
# Data Science

## A New Experience

[ibm.DataScience.com](http://ibm.DataScience.com)

# Data Science Professional

*Data Scientists Combine Skills across areas of Expertise*



*A Data Science Professional vary in a combinations of these skills*

# Data Science Work

**1) Exploration:** We want to understand the structure present in the data. Are there clusters? Non-obvious relationships? Any Correlation?

**Often referred to as “unsupervised learning”**

-E.g., K-means clustering

Use Cases -> behavior segmentation, inventory categorization, etc...



**2) Prediction:** The data contains a particular attribute (called the target attribute) and we want to learn how the target attribute depends on the other attributes.

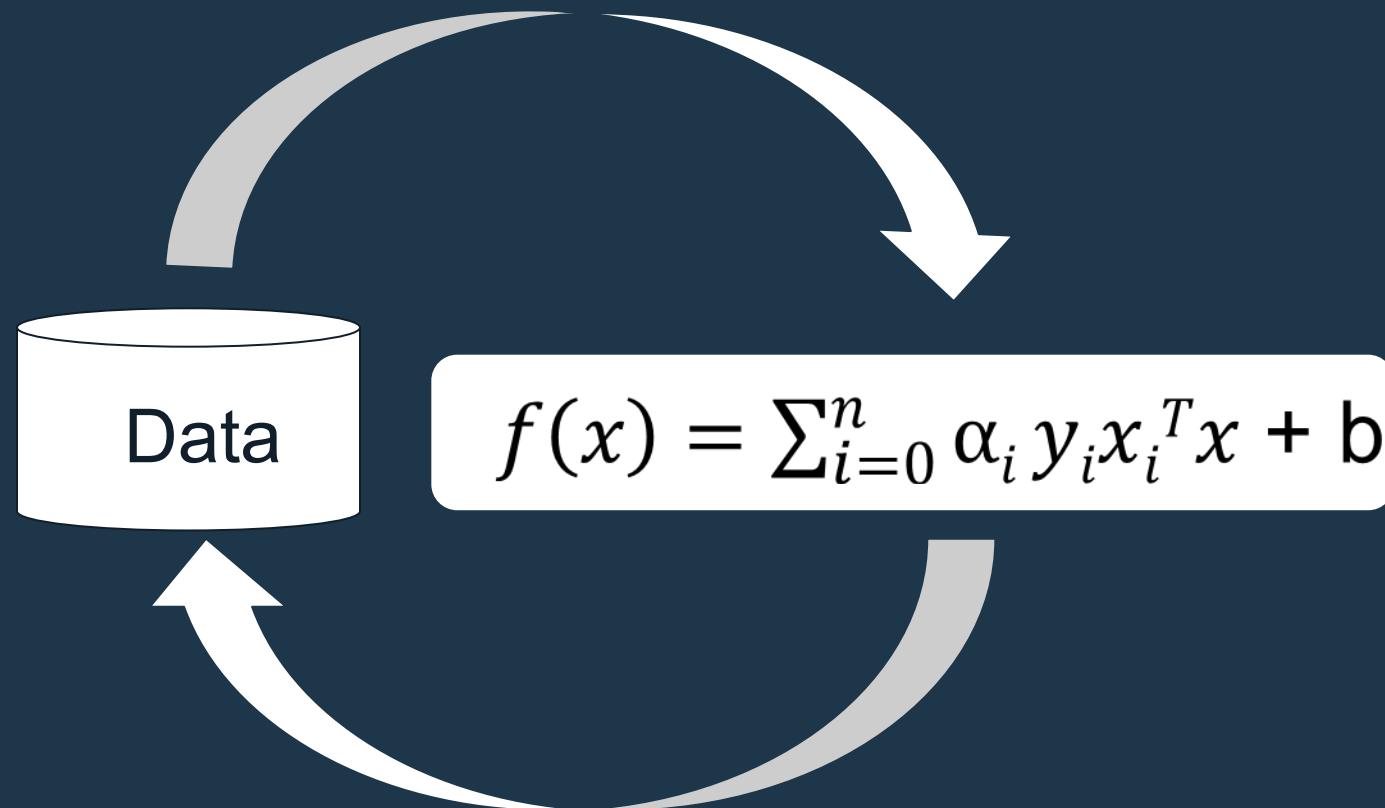
**Also referred to as “supervised learning”**

E.g., Support vector machines

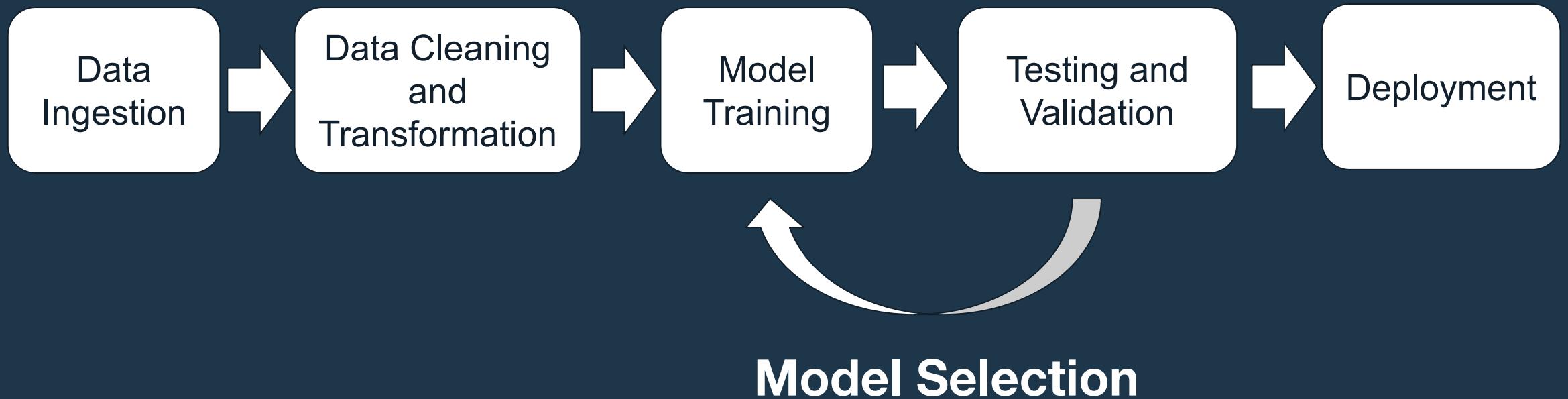
Use Cases -> Anomaly detection, Predict customer churn, fraud, etc...

# Machine Learning

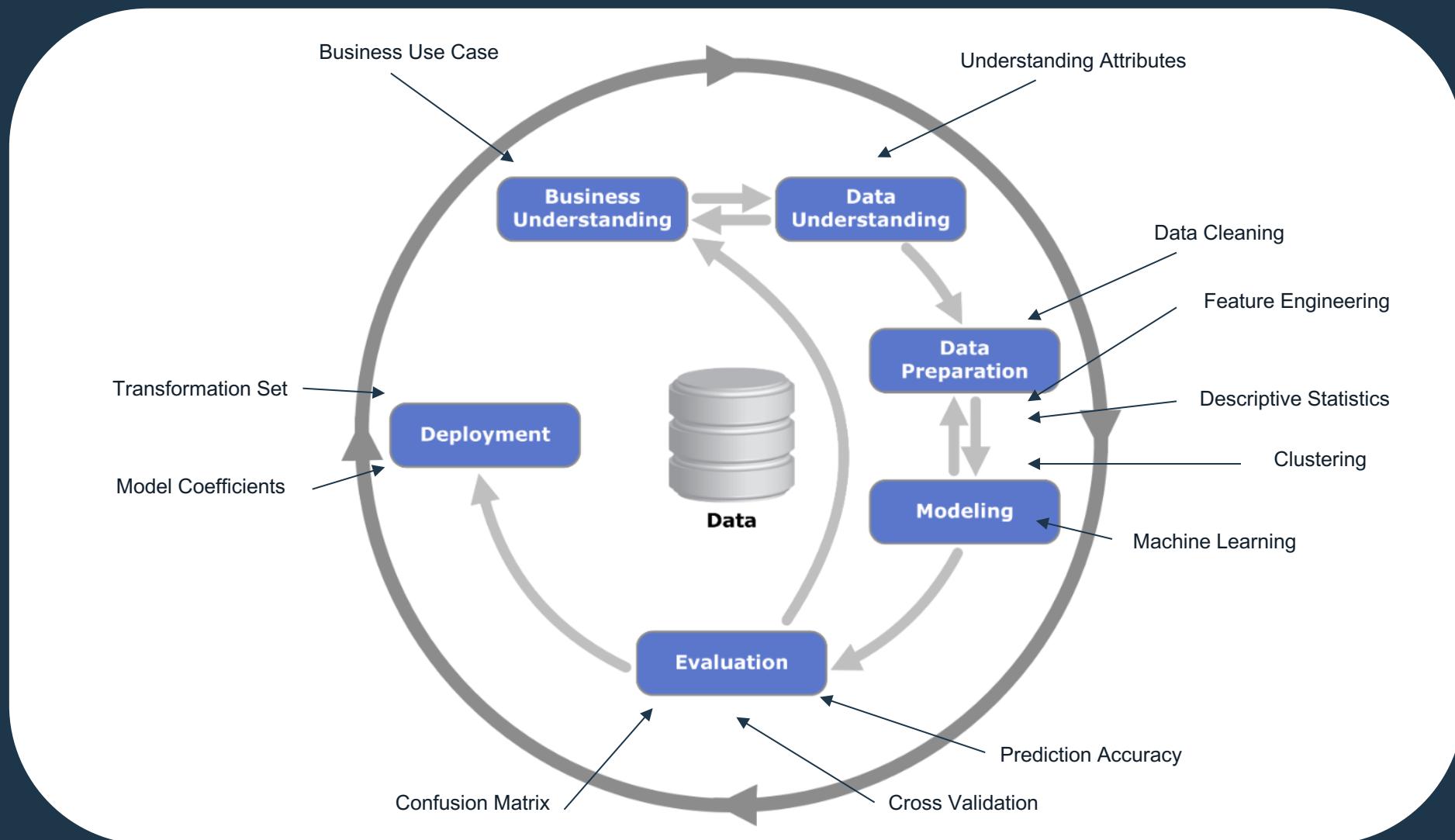
- “The science of getting computers to act without being explicitly programmed”
- “Systems that can learn from data”



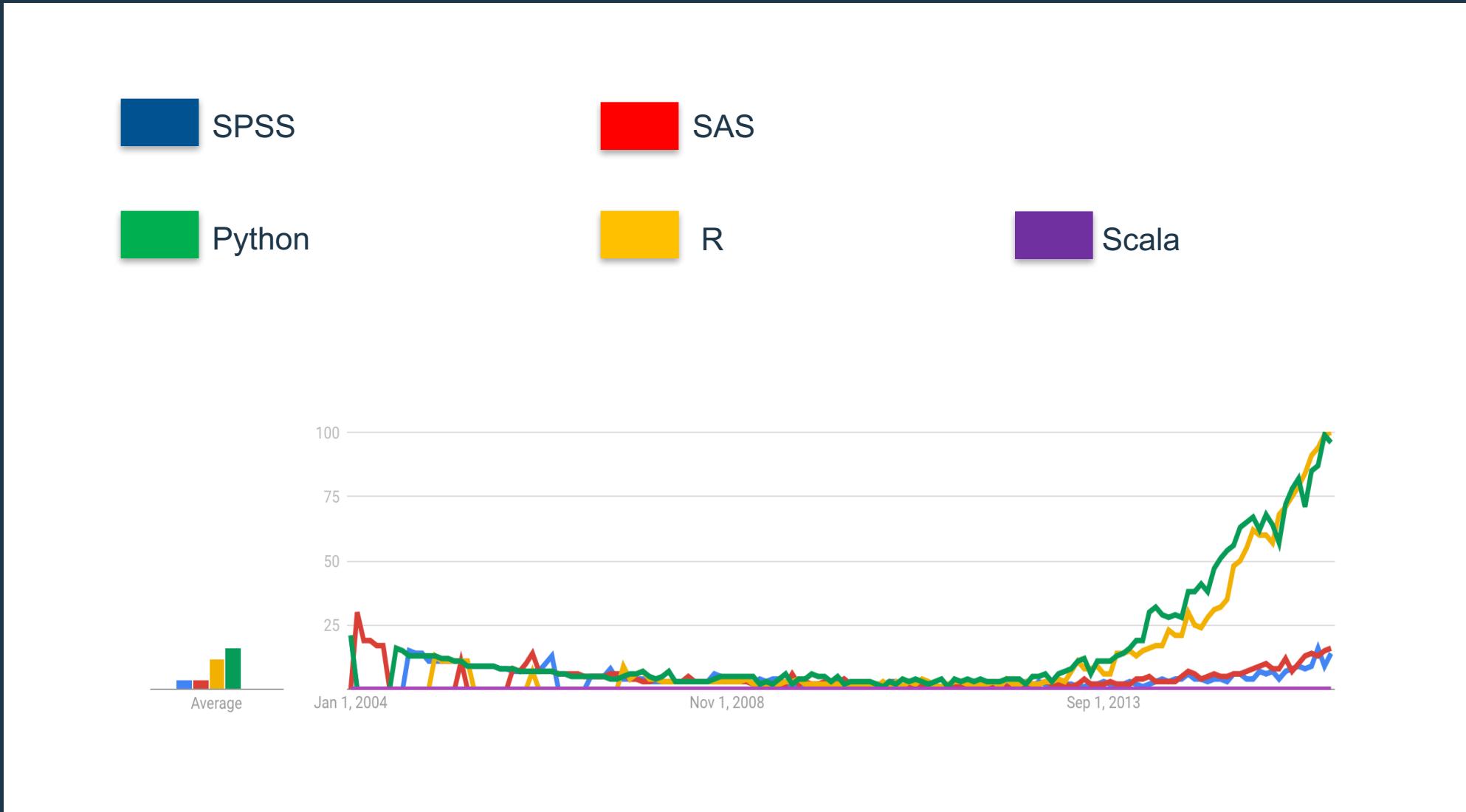
# Data Science - Process



# Data Science – Process



# Google Trends - *Data Science Languages*



# Open Source Innovation - Data Science Technologies



# Data Science – Challenges

## Rigid toolset

- Have to choose only one approach
- Cannot easily connect all of the capabilities needed
- Difficult to navigate between the various tools used

## Fragmented and time consuming

- Using multiple disjointed environments
- Separate on-ramp/community for each tool/environment
- Does not have meta data or data lineage

## Analytical Silo

- Difficult to maintain and version control project assets
- Limited means of collaborating with team
- Results are difficult to share



# Data Science Platforms

## Gartner February 2017 Magic Quadrant



# Introducing IBM Data Science Experience



**ALL YOUR TOOLS IN ONE PLACE**

IBM Data Science Experience provides an environment that brings together everything that a Data Scientist needs. It includes the most popular Open Source tools and IBM unique value-add functionalities with community and social features, integrated as a first class citizen to make Data Scientists more successful.



## Learn

Built-in learning to get started or go the distance with advanced tutorials

## Create

The best of open source and IBM value-add to create state-of-the-art data products

## Collaborate

Community and social features that provide meaningful collaboration

URL: <http://DataScience.ibm.com>

# Core Attributes of the Data Science Experience

## IBM Data Science Experience



### Community

- Find tutorials and datasets
- Connect with Data Scientists
- Ask questions
- Read articles and papers
- Fork and share projects

### Open Source

- Code in Scala/Python/R/SQL
- Jupyter and Zeppelin\* Notebooks
- RStudio IDE and Shiny apps
- Apache Spark
- Your favorite libraries

### IBM Added Value

- Data Shaping/Pipeline UI \*
- Auto-data preparation and modeling\*
- Advanced Visualizations\*
- Model management and deployment\*
- Documented Model APIs\*

Powered by IBM **IBM Watson Platform** in the Cloud

\* DSX product roadmap items



**IBM Data Science Experience**  
<https://youtu.be/1HjzkLRdP5k>

# Helpful Resources

Big Data University <https://bigdatauniversity.com/>

IBM Big Data & Analytics <http://www.ibmbigdatahub.com/>

Data First Method <https://ibm.biz/BdidFq>

IBM Bluemix Cloud <https://console.ng.bluemix.net/>

IBM Data Science Experience <http://datascience.ibm.com/>

# Questions?

**Carlo Appugliese**

[carlo.appugliese@us.ibm.com](mailto:carlo.appugliese@us.ibm.com)



The **Watson & Cloud** Platform

