

“Dearly beloved,  
we are gathered here today  
to get through this thing  
called life.” **Big Data**

—PRINCE

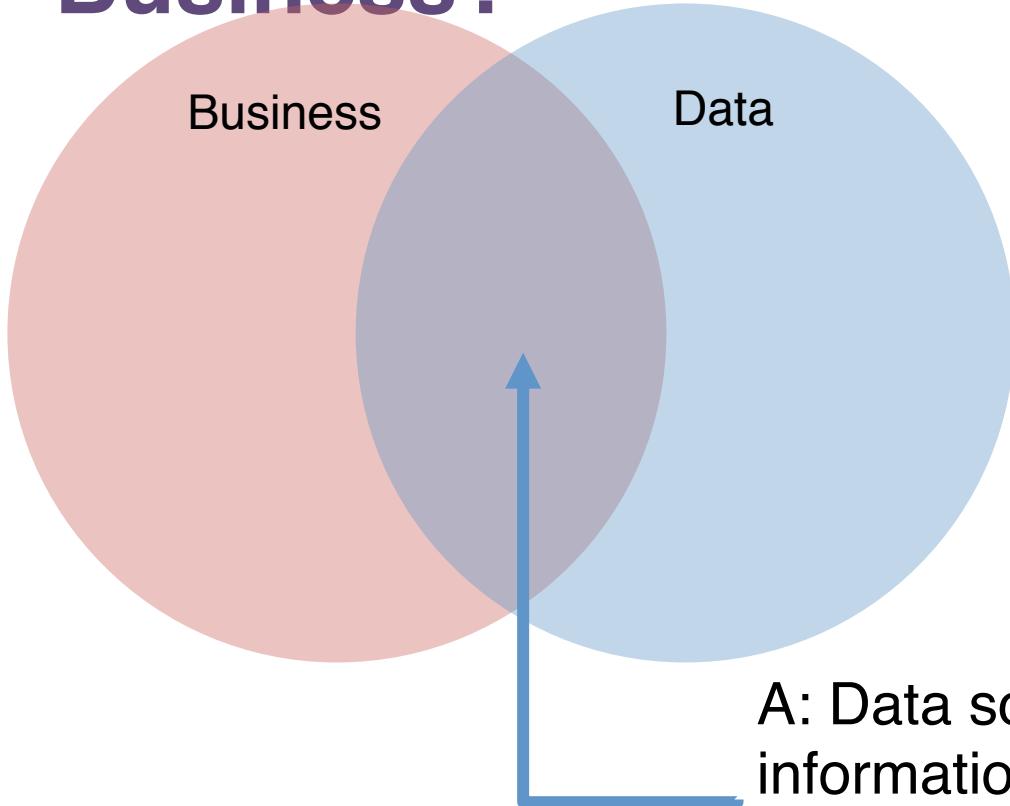
# Agenda

- Introduction
- Data Science Primer
- Working with Data Scientists
- Decoding the Data Science Lingo
- Q&A

# Data Science Primer

- What can Data Science do for the Business?
- Applications of Data Science
- Data-Driven Decisions
- What does a Data Scientist do?
- Data Science Skills

# What can Data Science do for the Business?



A: Data science! Extracting useful information and knowledge from large volumes of data in order to **improve business decision-making** or providing the business insights to make **data-driven decisions**

# What can Data do?

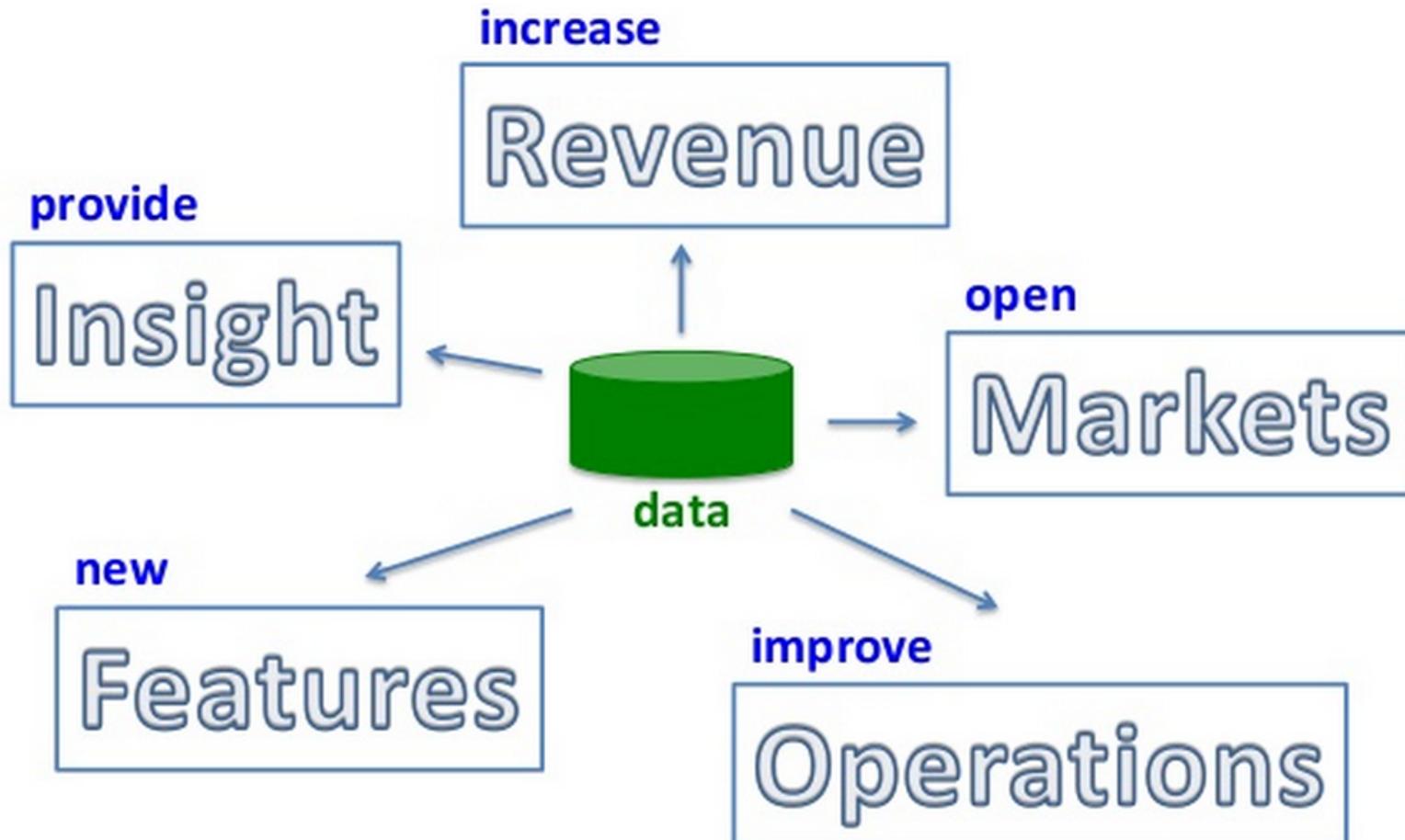


Image: <http://www.slideshare.net/andrewgardner5811/big-data-and-the-art-of-data-science>

# Applications of Data Science

fitbit



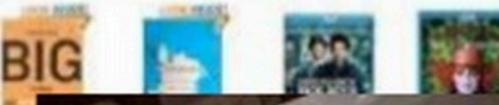
flu tracker

amazon.com

amazon

Recommended for You

Amazon.com has new recommendations for you based on items you purchased or told us you own.



SIRI

www.joshuareina.com

**Hotels in Sonoma County**  
Great food, wineries, redwoods,  
Beaches and more (we take  
pets!)  
[www.SonomaOrchidInn.com](http://www.SonomaOrchidInn.com)

AdChoices ▾

ads

# Data-Driven Decisions

- Practice of basing decisions on data, rather than purely on intuition
- There is evidence that data-driven decision making and big data technologies substantially improve business performance

# The Art and Science of Data Science

- Discover **unknowns** in data
- Obtain predictive, **actionable** insights
- **Communicate** business data stories
- Build **confidence in decision** making
- Create valuable **Data Products** that has business impacts

**I WANT ANSWERS**



**CALL THE DATA SCIENTIST.  
RIGHT NOW!**

# What does a Data Scientist do?

- Data curiosity. Explore data. Discover unknowns
- Understand data relationships
- Understand the business, has domain knowledge
- Can tell relevant stories with data
- Holistic view of the business
- Knows machine learning, statistics, probability
- Can hack and code
- Define and test an hypothesis, run experiments
- Asks good questions

# Data Science Skills

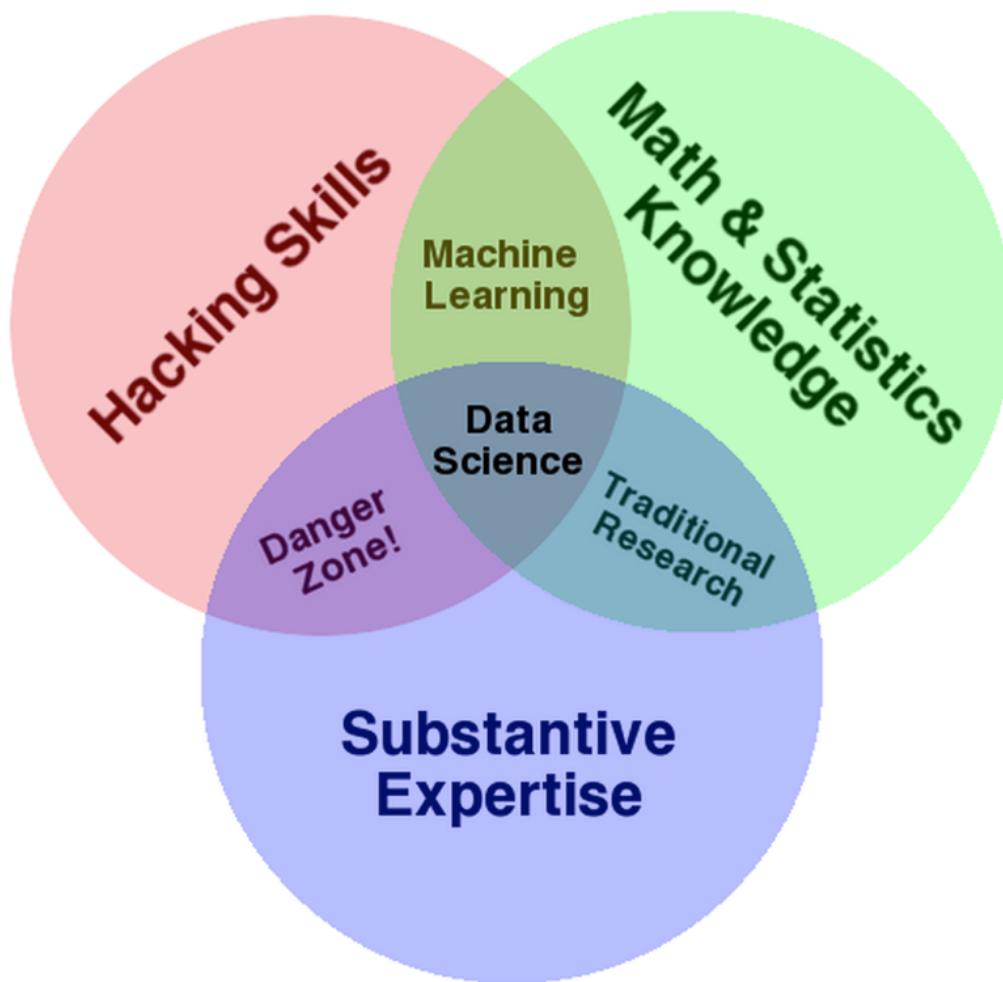
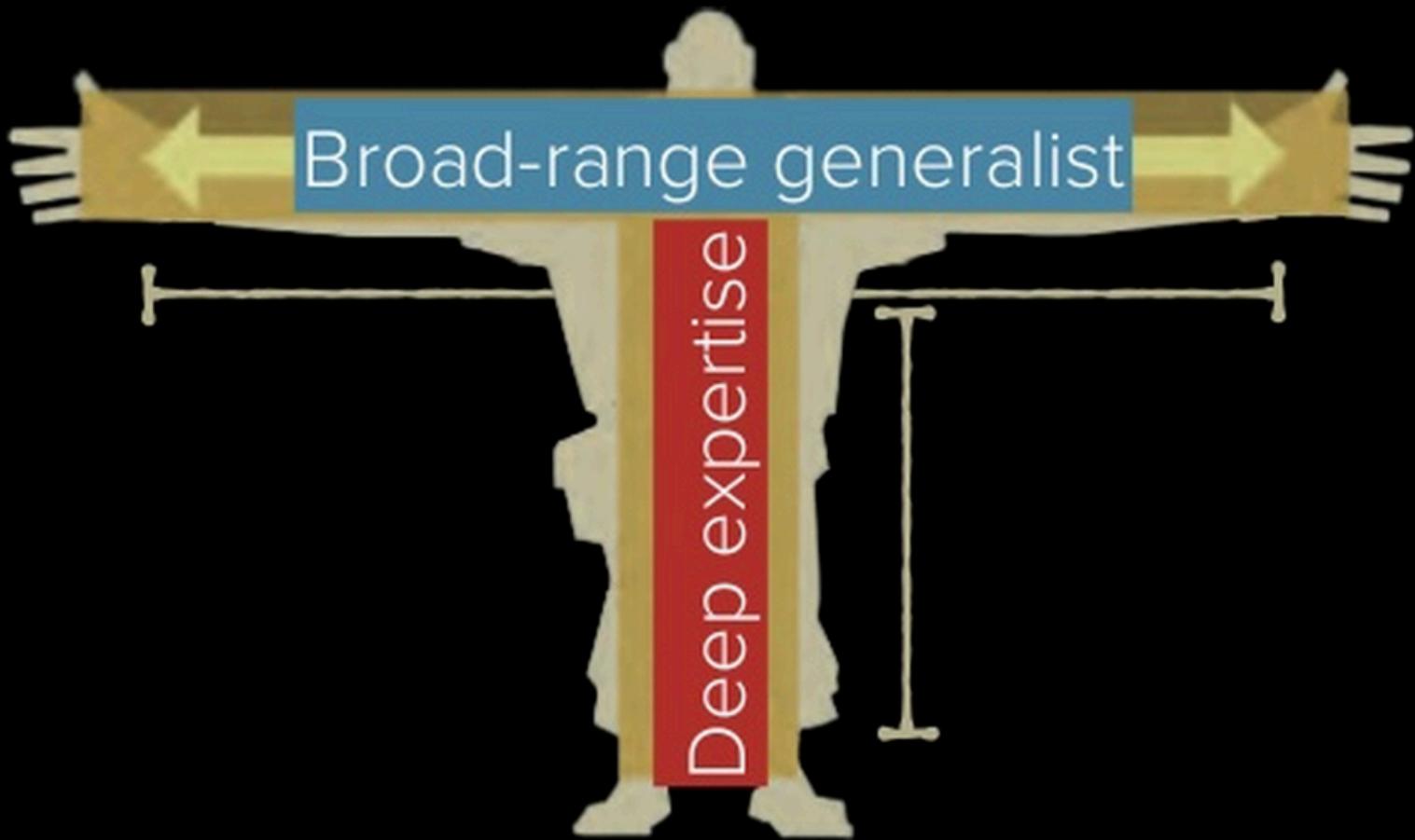
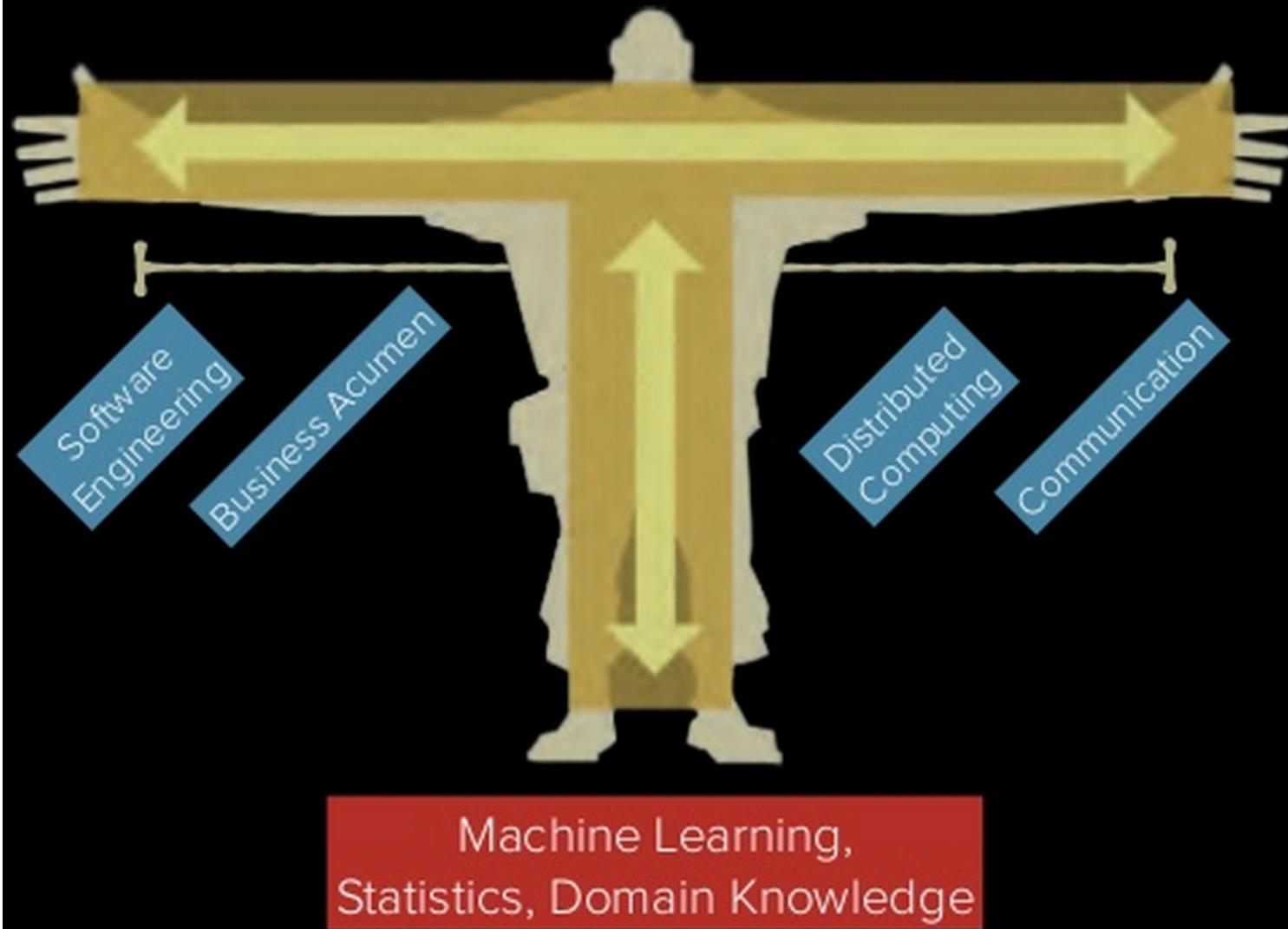


Image: <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

# T-Shaped Skillset



# T-Shaped Skillset



# Working with Data Scientists

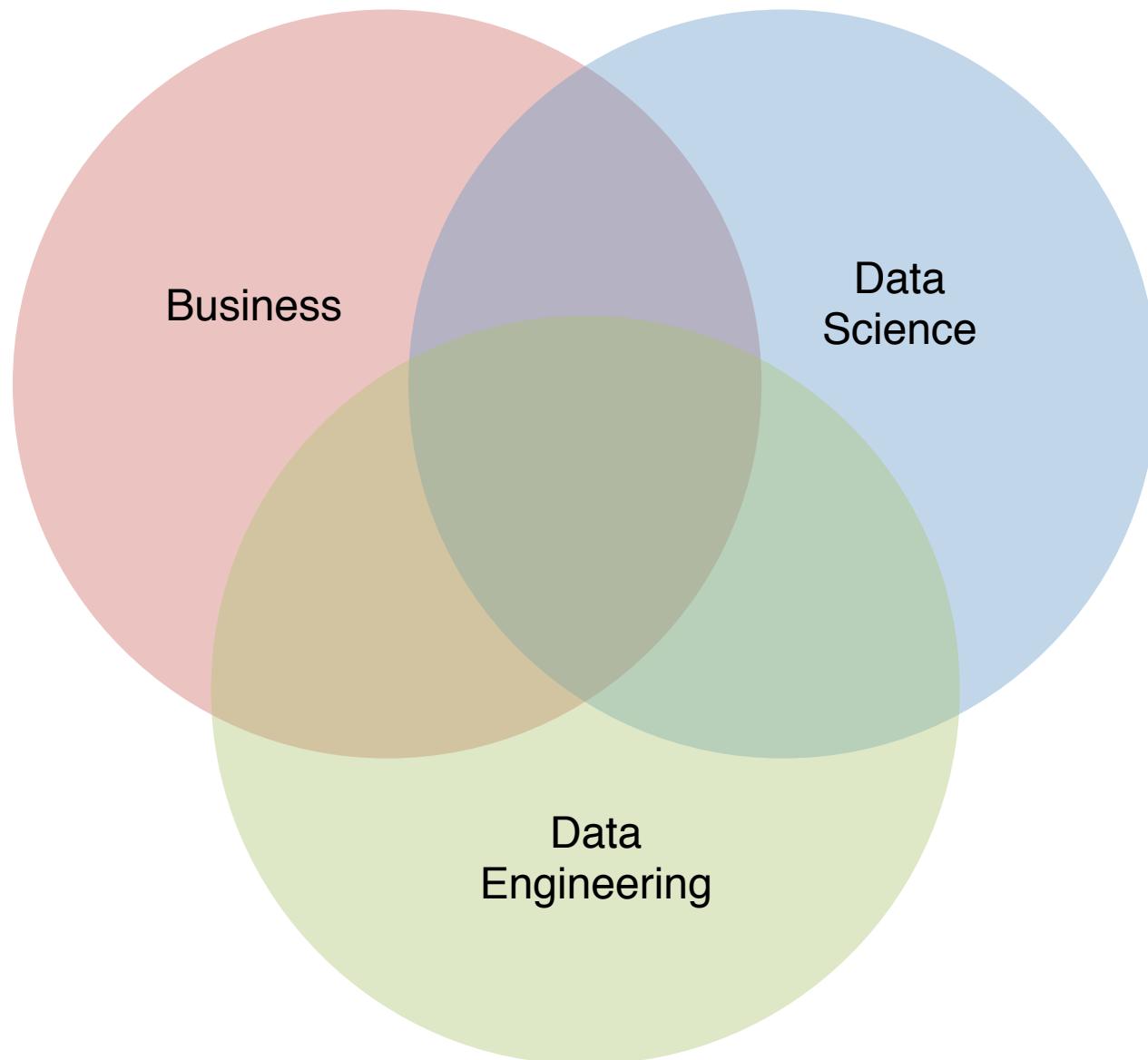
- Collaboration
- Data Science Cycle
- Organizational Models for Data Science Teams

**I WANT ANSWERS**

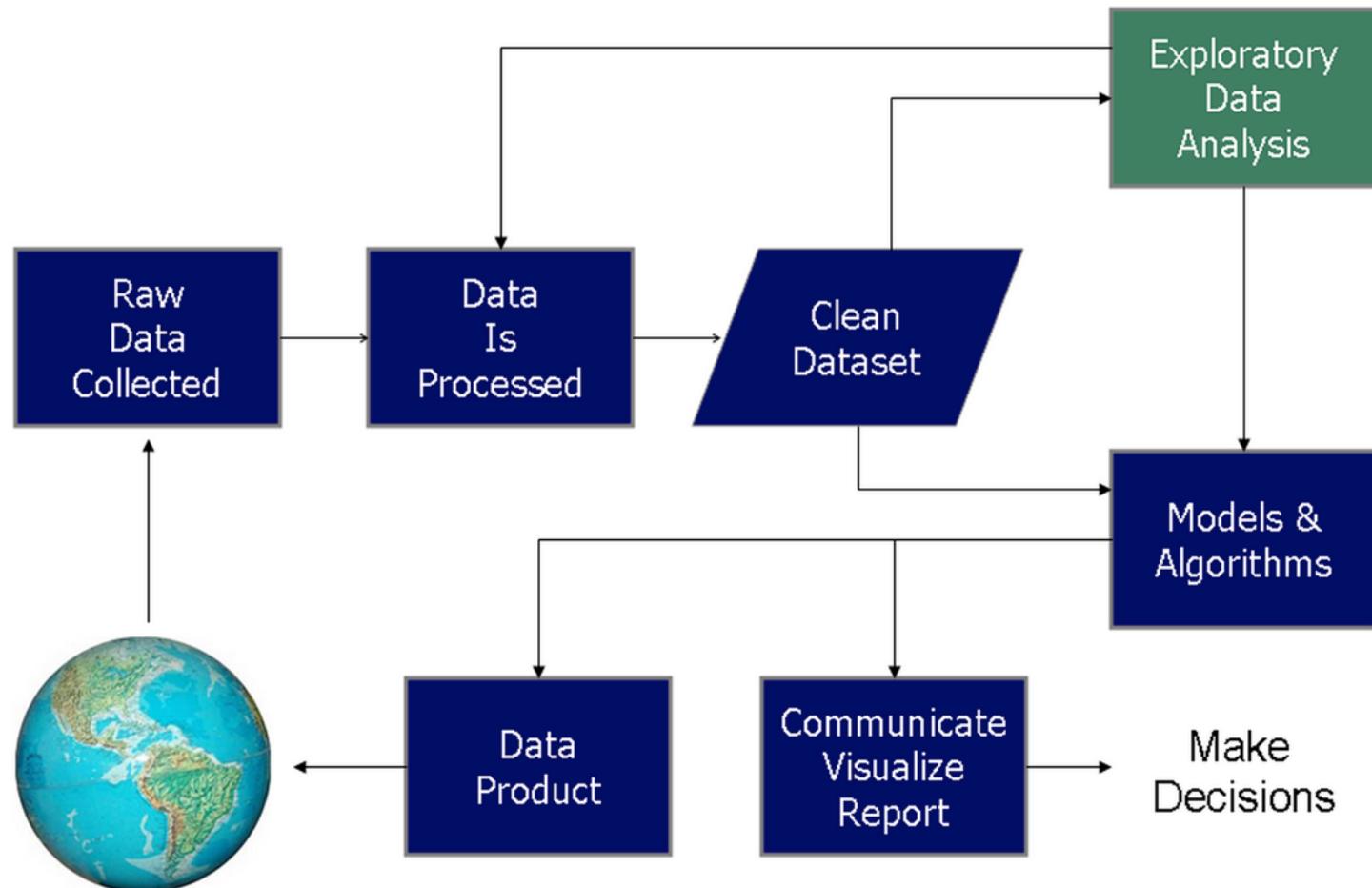


**CALL THE DATA SCIENTIST.  
RIGHT NOW!**

# Working with Data Scientists



# Data Science Cycle



# Organizational Models for Data Science Teams

## Centralized



The Data Science Team Functions As A Hub And Spoke Model, In Which They Are A Central Provider Of Analytics To Multiple Business Units

## Hybrid



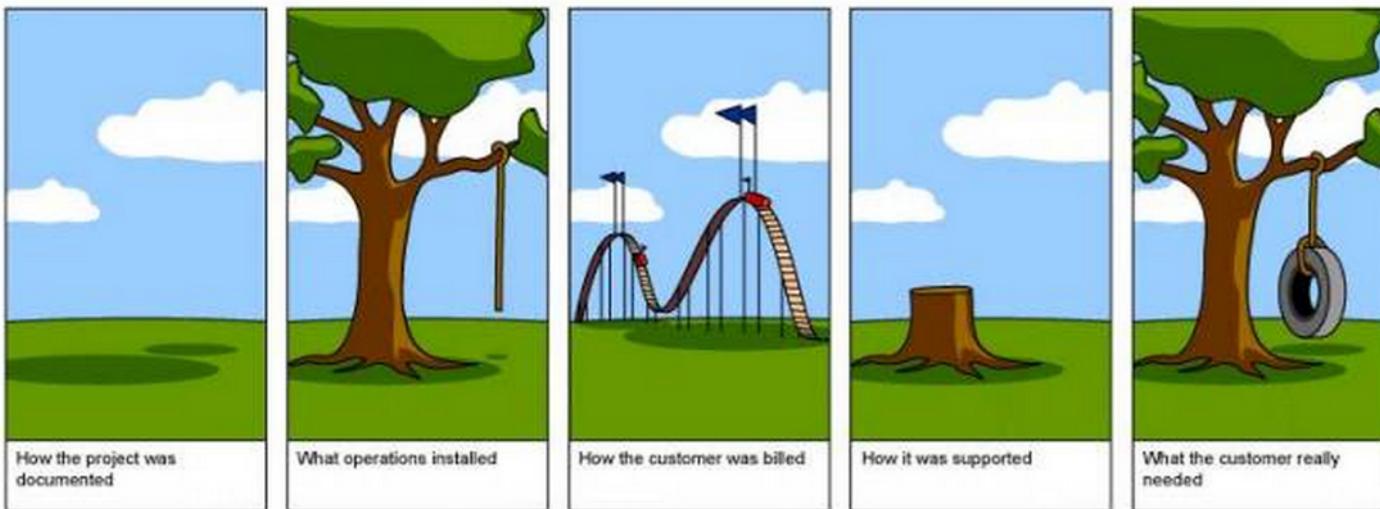
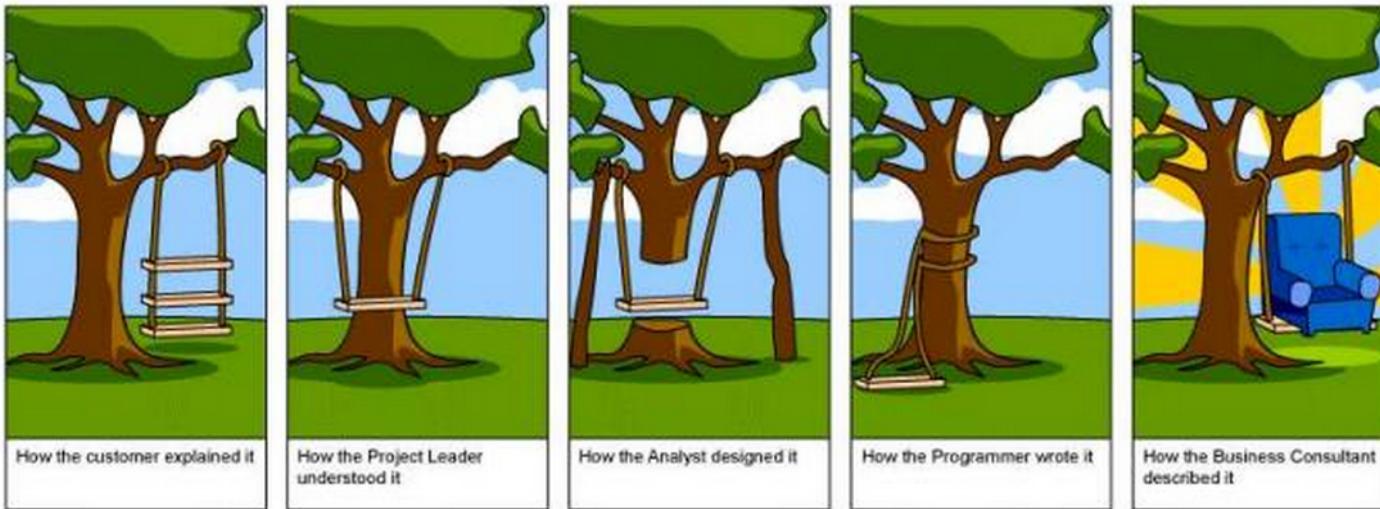
There Is A Centralized Data Science Team, But Business Units Also Have Data Science Capabilities

## Decentralized



Each Business Unit Has Its Own Data Science Capabilities

# Decoding the Data Science Lingo



# Machine Learning

## Data Science Definition:

- A subfield of computer science and artificial intelligence (AI) that focuses on the design of systems that can learn from and make decisions and predictions based on data.
- Machine learning enables computers to act and make data-driven decisions rather than being explicitly programmed to carry out a certain task.
- Machine Learning programs are also designed to learn and improve over time when exposed to new data.

## Business Application:

- Everything!

# Unsupervised Learning

## Data Science Definition:

- Where a program, given a dataset, can automatically find patterns and relationships within the dataset.
- The business will decide how deeply or many categories there are.
- Clustering or grouping of like data.
- *Examples:* k-means clustering, hierarchical clustering

## Business Application:

- Customer segmentation
- Understanding users and behaviors
- Classifying unknown and pre-defined images into categories

# Supervised Learning

## Data Science Definition:

- Where a program is “trained” on a pre-defined dataset.
- Based off its training data the program can make accurate decisions when given new data.

## Business Application:

- Classifying Twitter sentiments
- Recommender systems

# Score

## Data Science Definition:

- Number of ways to evaluate how well the model assigns the correct class value to the test instances.

## Business Application:

- Confidence gauge

# Classification

## Data Science Definition:

- Sub-category of Supervised Learning
- Classification is the process of taking some sort of input and assign a label to it. The predictions are discrete, categories, or “yes or no” nature.
- *Examples:* Logistic Regression, Random Forest

## Business Application:

- What customers should a company target with its marketing campaigns?
- Is this Nigerian prince committing fraud? (Spam classification)
- Is this actually Barack Obama’s Facebook profile and review on Amazon? (Fraud detection)

# Regression

## Data Science Definition:

- Sub-category of Supervised Learning
- Regression is a type of algorithm that predicts a continuous values.

## Business Application:

- How much would a user spend on a mobile game like CandyCrush?
- How much would someone spend on healthcare out of pocket?
- How many attendees will come to this event based on past registration?

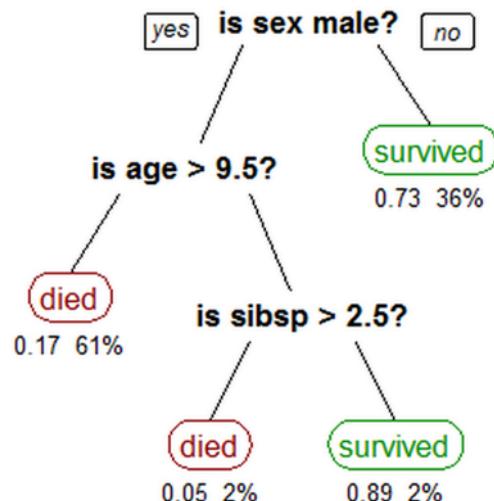
# Decision Trees

## Data Science Definition:

- Using a tree-like graph or model of decisions and their possible consequence.

## Business Application:

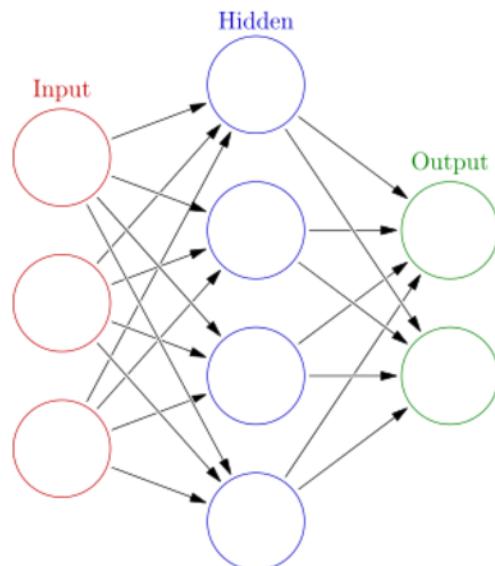
- Medical Testing (e.g. health incidences, etc.)
- Genealogy breakdowns (e.g. eye color, blood type, etc.)



# Deep Learning

## Data Science Definition:

- A category of machine learning algorithms that often use Artificial Neural Networks to generate model.



## Business Application:

- Image classification
- Language processing
- Audio processing
- Outlier and fraud detection