

# Machine Learning vs Data Analytics vs Data Engineering

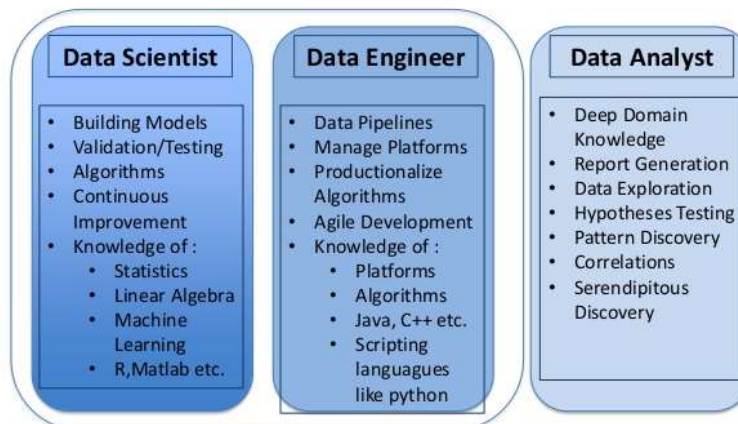
DataCrux Insights @ 2018 ALL RIGHTS RESERVED

## Data Mining, Machine Learning and Deep Learning

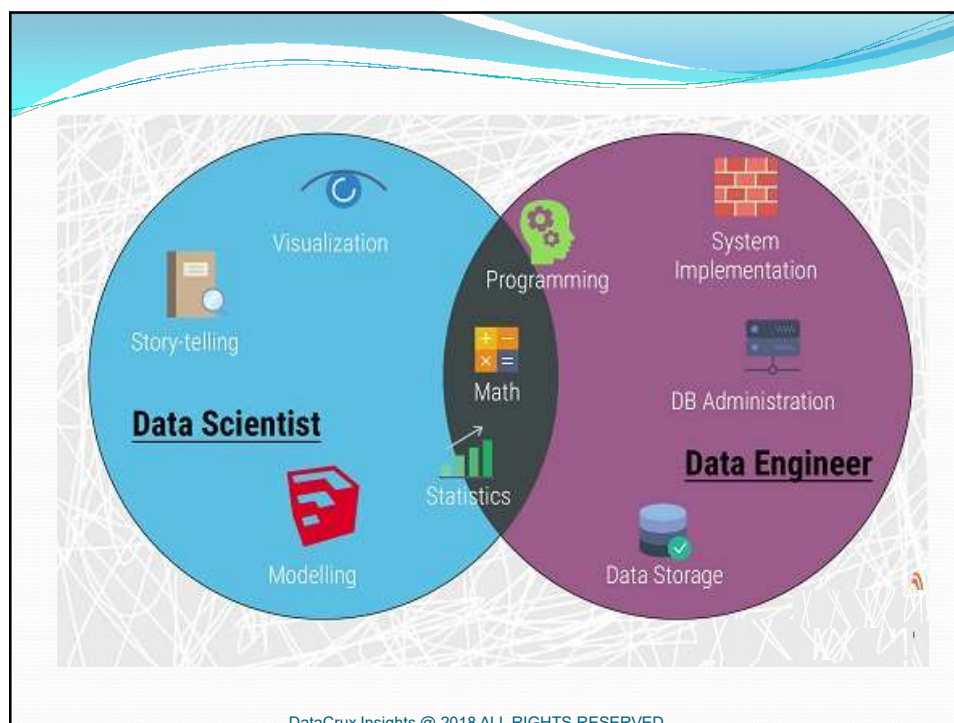
- same algorithms and techniques
- predictions vary
- data mining discovers previously unknown patterns and knowledge
- machine learning reproduces known patterns and knowledge, and further automatically apply that to data, and also to decision making and actions.

DataCrux Insights @ 2018 ALL RIGHTS RESERVED

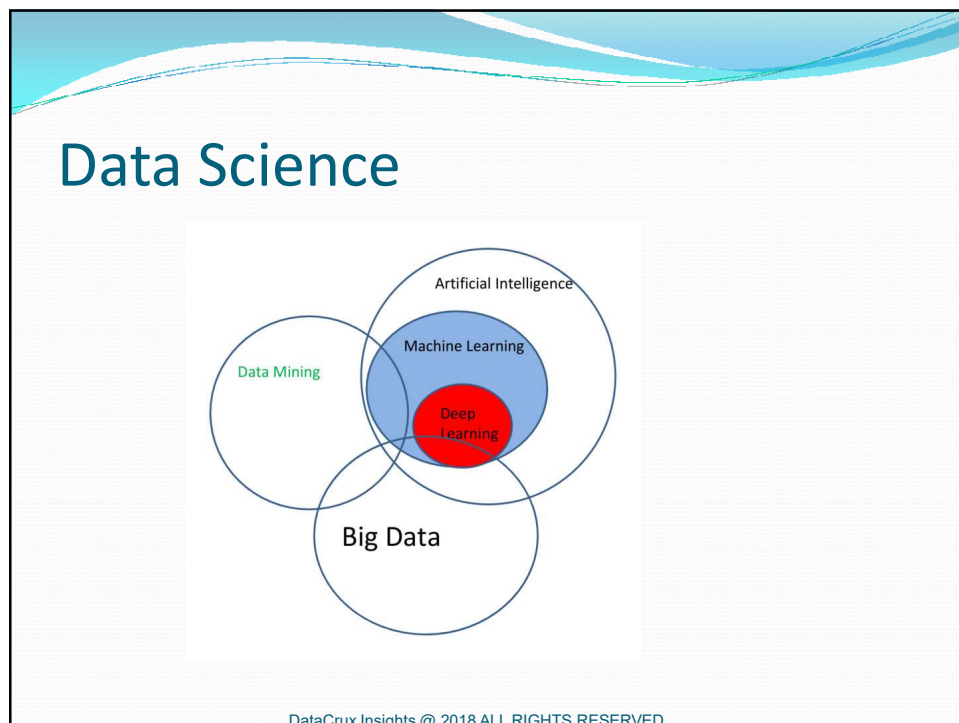
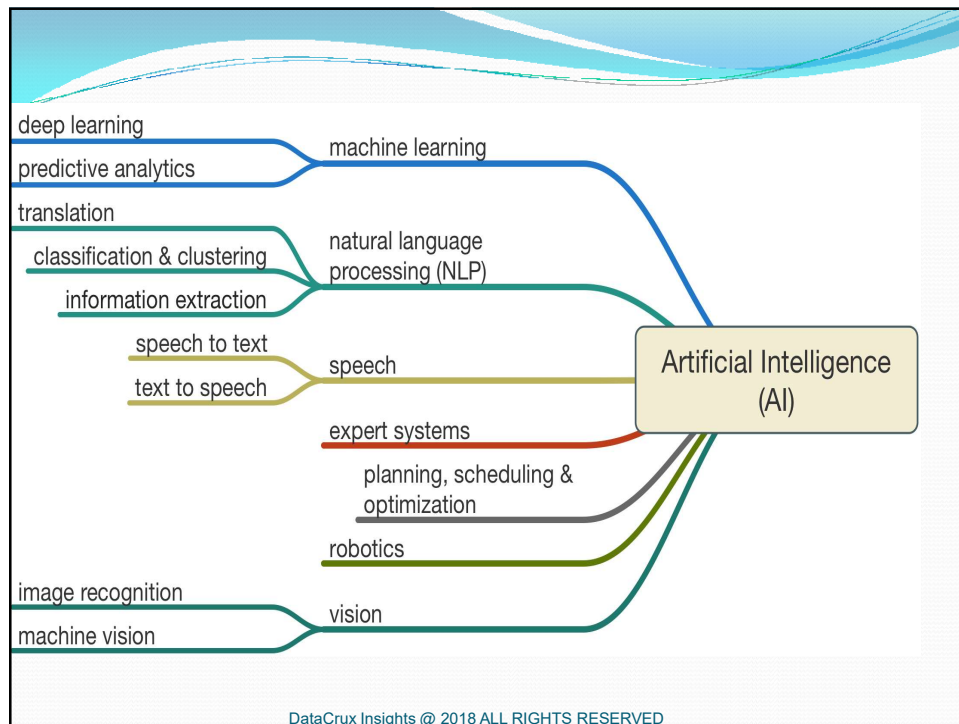
## Life in the Big Data Universe



DataCrux Insights @ 2018 ALL RIGHTS RESERVED



DataCrux Insights @ 2018 ALL RIGHTS RESERVED



## Data Science vs. Data Engineering

	Data Science	Data Engineering
Approach	Scientific (Exploration)	Engineering (Development)
Problems	Unbounded	Bounded
Path to Solution	Iterative, exploratory, nonlinear	Mostly linear
Education	More is better (PhD's common)	BS and/or self-trained
Presentation Skills	Important	Not as important
Research experience	Important	Not as important
Programming skills	Not as important	Important
Data skills	Important	Important

DataCrux Insights @ 2018 ALL RIGHTS RESERVED

### Data Engineer

Data Storage/Management  
Big Data Infrastructure  
Data Pipelines  
Persistence Technology

### Data Scientist

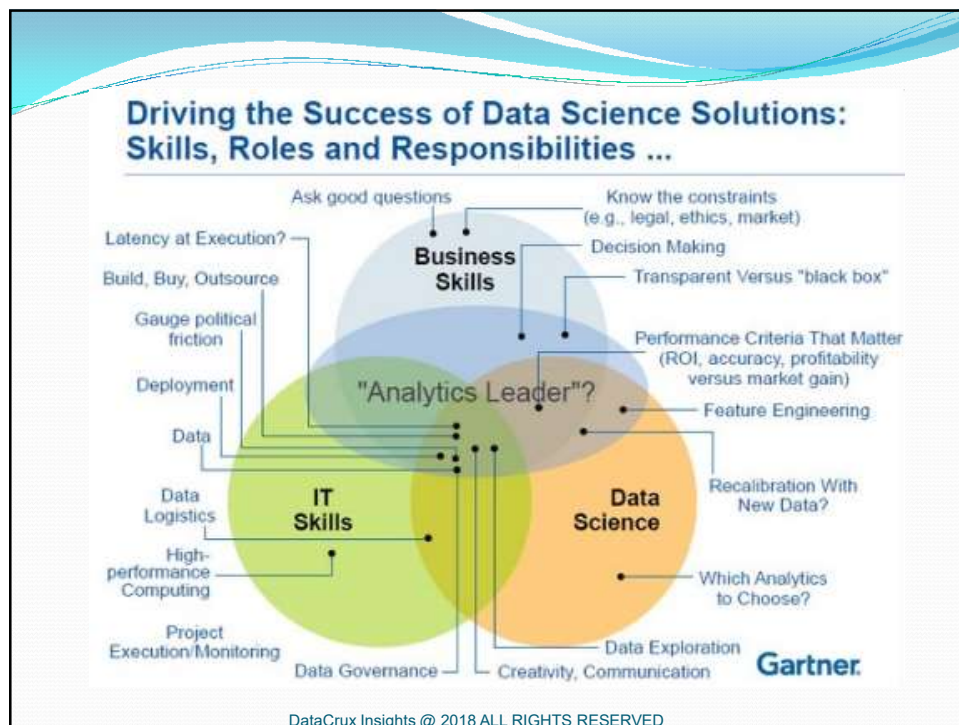
Visualization  
Storytelling  
Model Building  
Business Savvy

### Cross-Functional

Programming  
Data Munging  
Data Applications  
Stakeholder Engagement

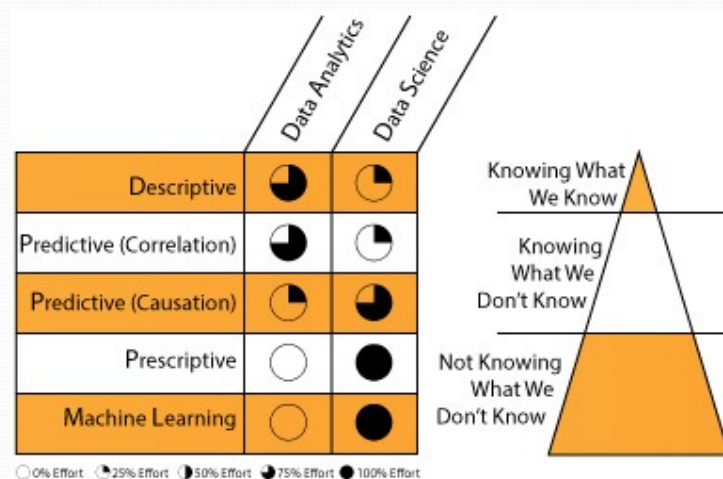
DataCrux Insights @ 2018 ALL RIGHTS RESERVED







## Data Analytics vs Data Science



DataCrux Insights @ 2018 ALL RIGHTS RESERVED

## A GLOSSARY OF ARTIFICIAL-INTELLIGENCE TERMS

### ■ ARTIFICIAL INTELLIGENCE

AI is the broadest term, applying to any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning).

### ■ MACHINE LEARNING

The subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with experience. The category includes deep learning.

### ■ DEEP LEARNING

The subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing multilayered neural networks to vast amounts of data.

DataCrux Insights @ 2018 ALL RIGHTS RESERVED

## How concepts are named?

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
large grant = \$1,000,000	large grant = \$50,000

DataCrux Insights @ 2018 ALL RIGHTS RESERVED