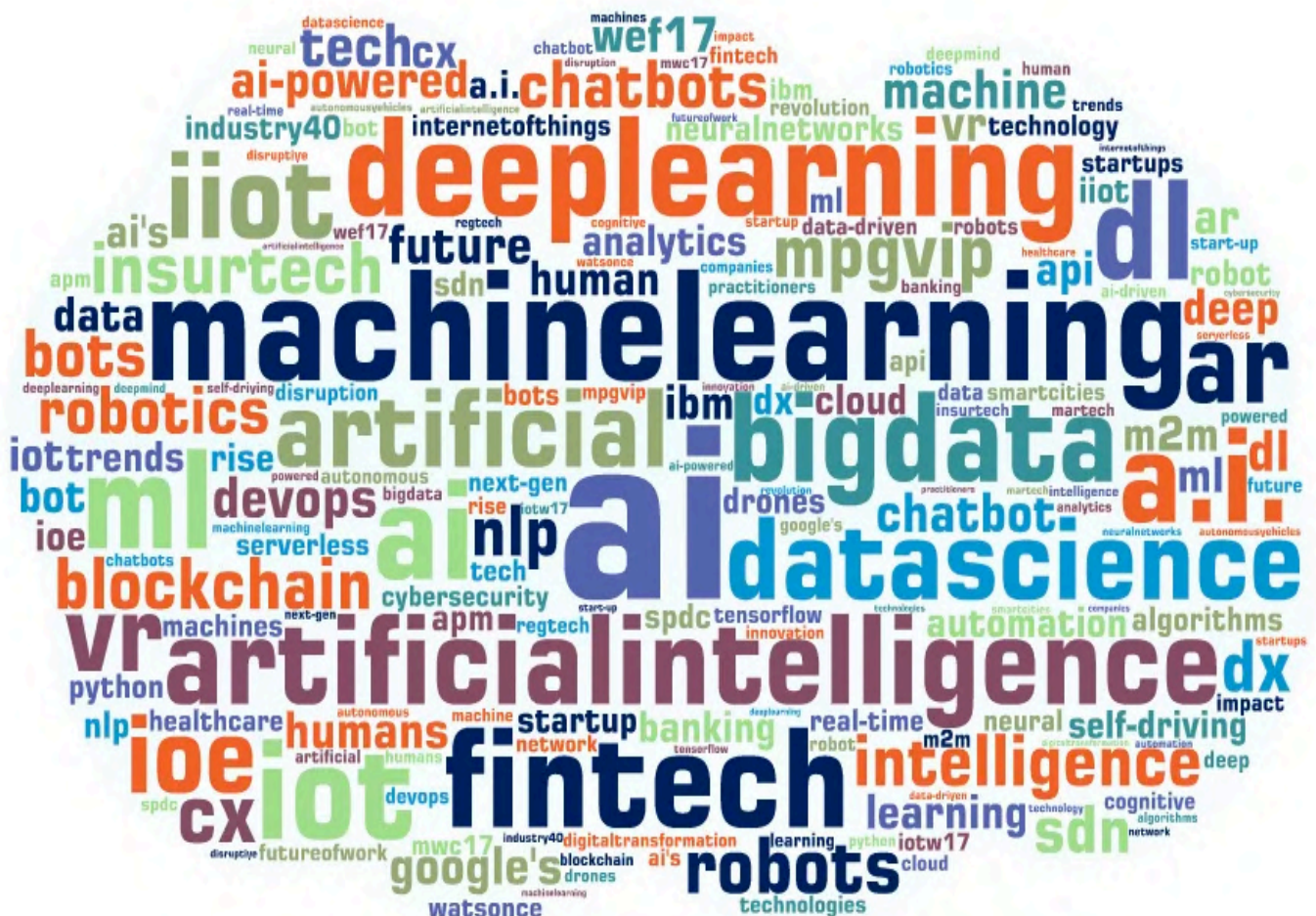# Introduction to Data Science

Mehul Motani
Electrical & Computer Engineering
National University of Singapore
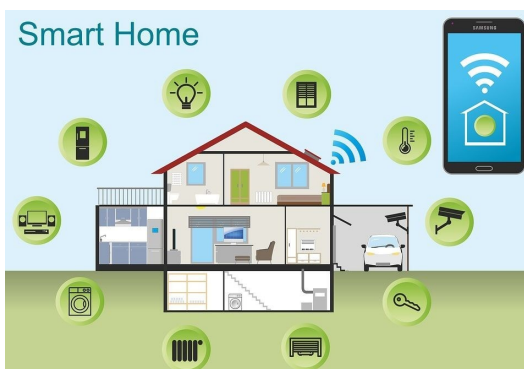Email: motani@nus.edu.sg

What is the Internet of Things (IoT)?
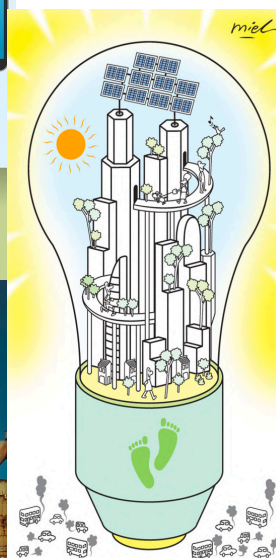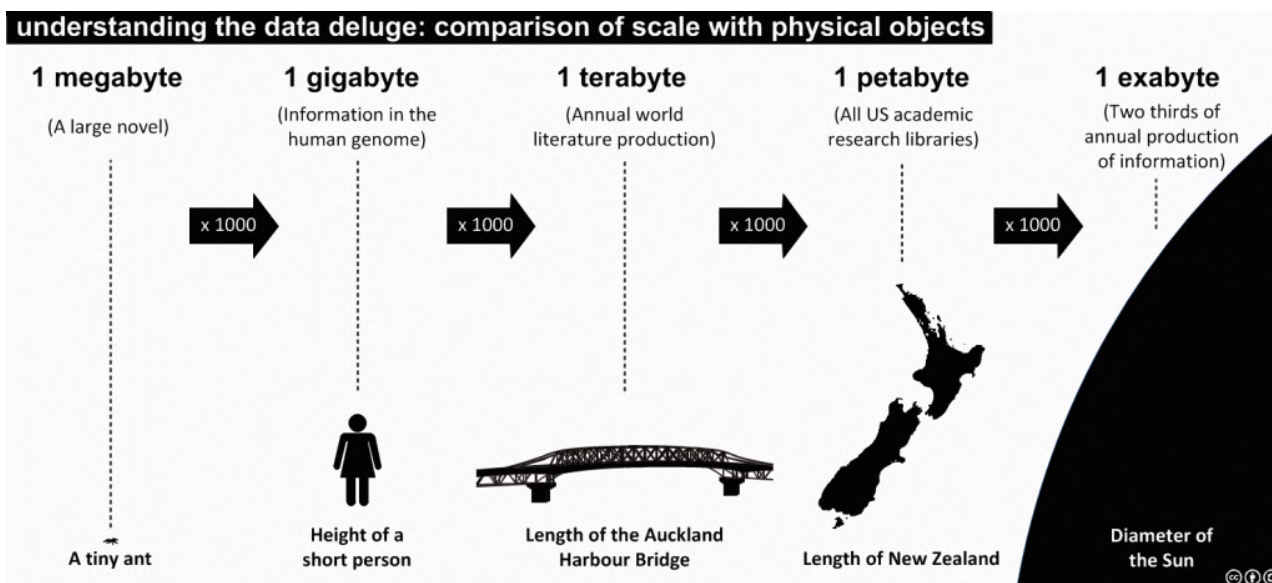
- The IoT is the network of machines, mobile phones, computers, vehicles, home appliances, and other physical and virtual devices embedded with intelligence, sensors, and actuators which collect, share and analyze data streams.

Applications of IoT

Smart Home

Smart Nation

Smart Healthcare

Industrial IoT

THE SMART CITY

The Data Deluge

understanding the data deluge: comparison of scale with physical objects

1 megabyte	1 gigabyte	1 terabyte	1 petabyte	1 exabyte
(A large novel)	(Information in the human genome)	(Annual world literature production)	(All US academic research libraries)	(Two thirds of annual production of information)
x 1000	x 1000	x 1000	x 1000	
A tiny ant	Height of a short person	Length of the Auckland Harbour Bridge	Length of New Zealand	Diameter of the Sun

Source: ritholtz.com

Sources of Big Data

Healthcare   Transportation

Finance   Social Media
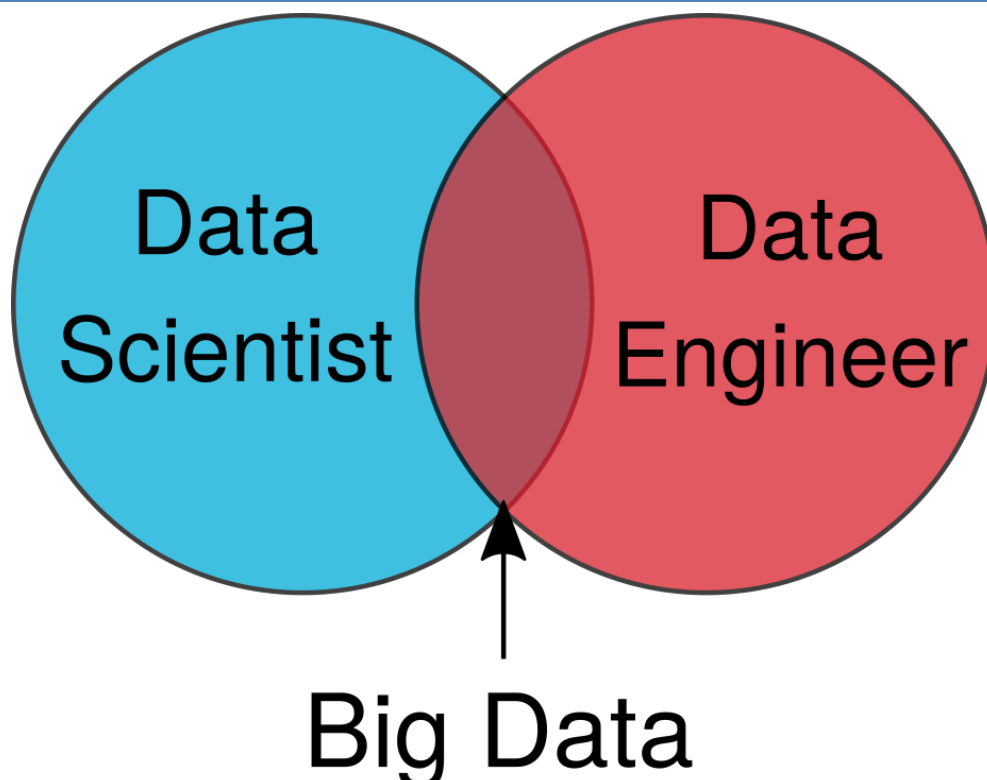
Big Data Explosion

The Data Explosion

Data Science and Data Engineering

- <u>Data science</u> is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured.

- <u>Data engineers</u> build tools, infrastructure, frameworks, and services which allow them to tease insights from the myriad of data streams being generated. They have the ability to collect, curate, analyze and visualize data in all its forms for intelligent decision making and sense-making.

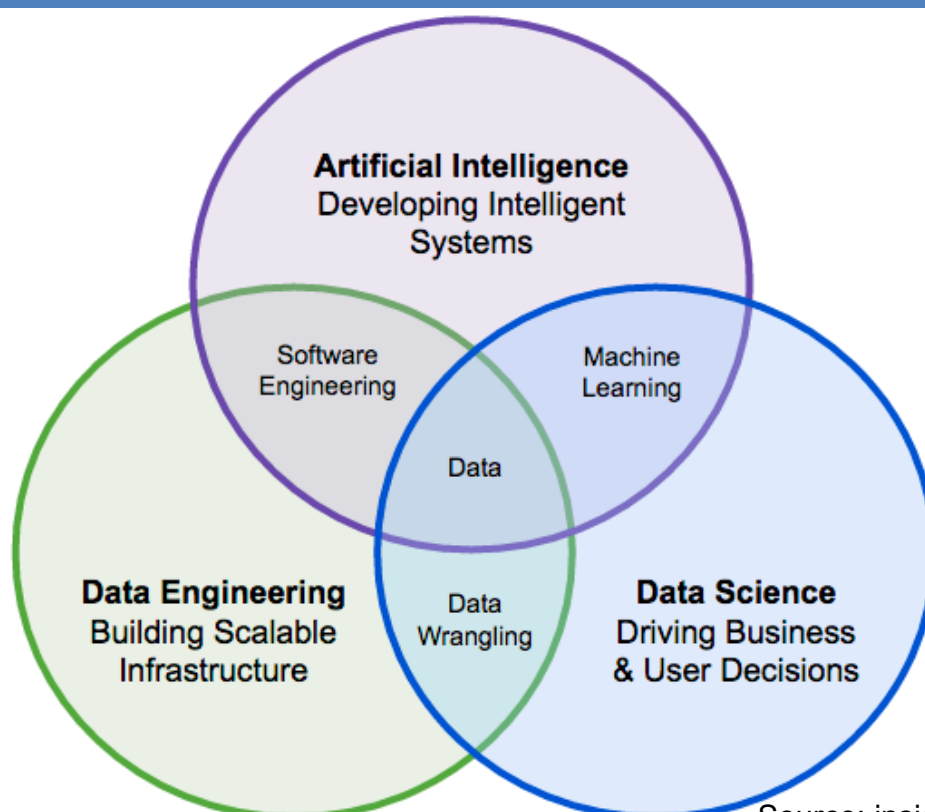# Big Data: Data Science and Data Engineering

Illustration by Jesse Anderson, Oreilly.com
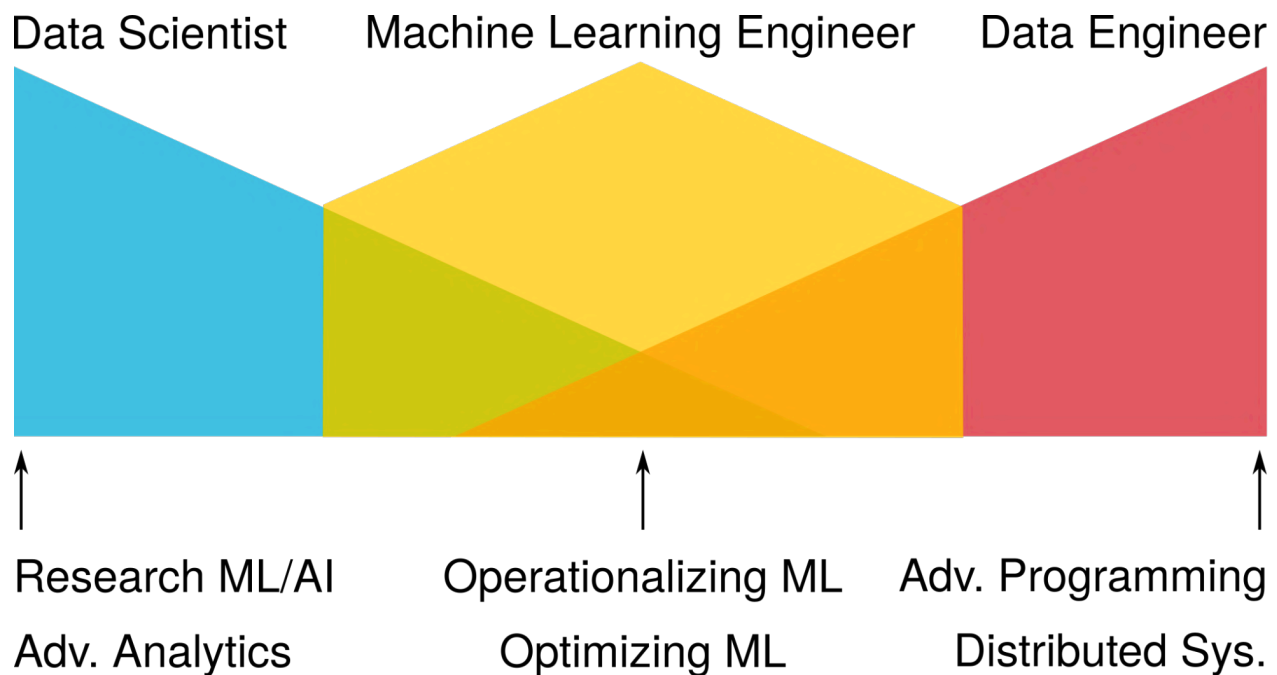
Some Terminology

- <u>Data Science</u> – interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from both structured and unstructured data. -Wikipedia

- <u>Artificial Intelligence</u> – Intelligence demonstrated by machines

- <u>Machine Learning</u> – A set of techniques to give computers/machines the ability to learn.

- <u>Data Engineering</u> – the ability to collect, manipulate, analyze and visualize data in all its forms.

- <u>Neural networks</u> – A computer system modelled on the human brain. Early efforts started in 1940s. First real unsupervised network in 1975.

- <u>Deep Learning</u> – The rebirth of neural networks!

How everything fits together

Source: insightdatascience.com

Getting a job

Data Scientist   Machine Learning Engineer   Data Engineer

Research ML/AI   Operationalizing ML   Adv. Programming

Adv. Analytics   Optimizing ML   Distributed Sys.

Illustration by Jesse Anderson and the Big Data Institute


Historical Perspective

- 1943 – Foundations of neural networks established by McCulloch and Pitts.

- 1950 – Alan Turing introduces the Turing test as a way of testing a machine's intelligence.

- 1952 – Claude Shannon demonstrates machine learning in the 1950s!

  Footage courtesy of AT&T Archives and History Center, Warren, NJ

  https://youtu.be/_9_AEVQ_p74

Claude Shannon demonstrates machine learning
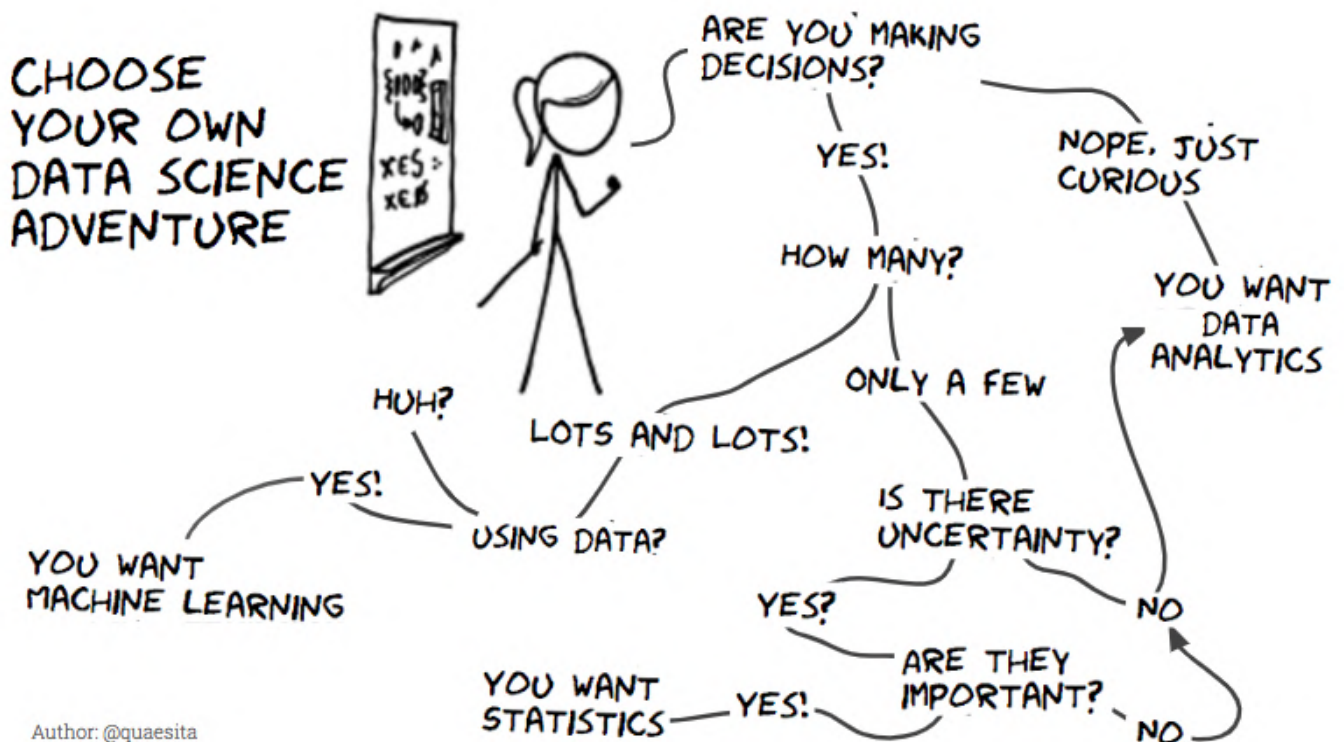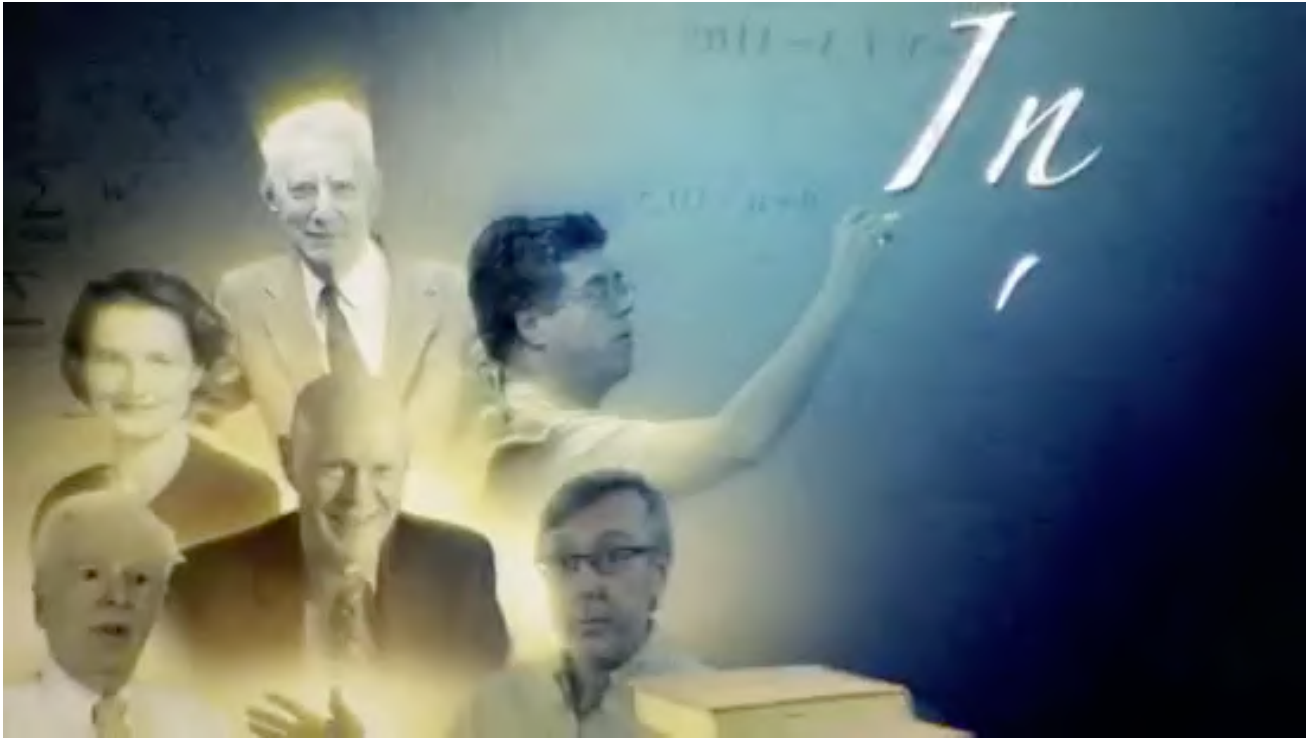
Pause here if you need more time. 00