Data Science Applications & Use Cases

Instructor: Ekpe Okorafor

1. Accenture – Big Data Academy
2. Computer Science African University of Science & Technology

Objectives

Objectives

- **Understand Big Data Challenges**
- **What exactly is Data Science and what do Data Scientists do**
- **Data Science contrasted with other disciplines**
- **Case Study & Use Cases**

Outline

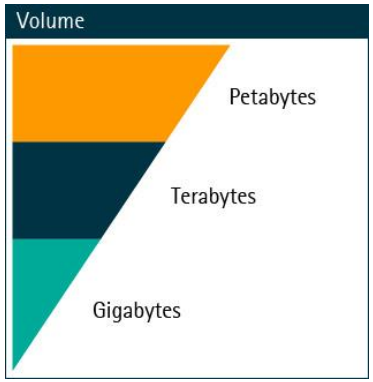
- **Big Data & Challenges**
- **What is Data Science**
- **Data Science & Academia**
- **Data Science & Others**
- **Case Studies**
- **Essential points**
- **Conclusion**

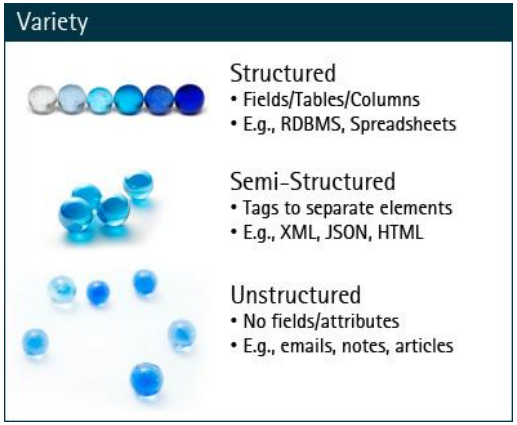
Data All Around

- **Lots of data is being collected and warehoused**
  - Scientific Experiments
  - Internet of Things
  - Web data, e-commerce
  - Financial transactions, bank/credit transactions
  - Online trading and purchasing
  - Social Network
  - ……many more!

Big Data

- Big Data are data sets so large or so complex that traditional methods of storing, accessing, and analyzing their breakdown are too expensive. However, there is a lot of potential value hidden in this data, so organizations are eager to harness it to drive innovation and competitive advantage.
- Big Data technologies and approaches are used to drive value out of data rich environments in ways that traditional analytics tools and methods cannot.





What To Do With These Data?

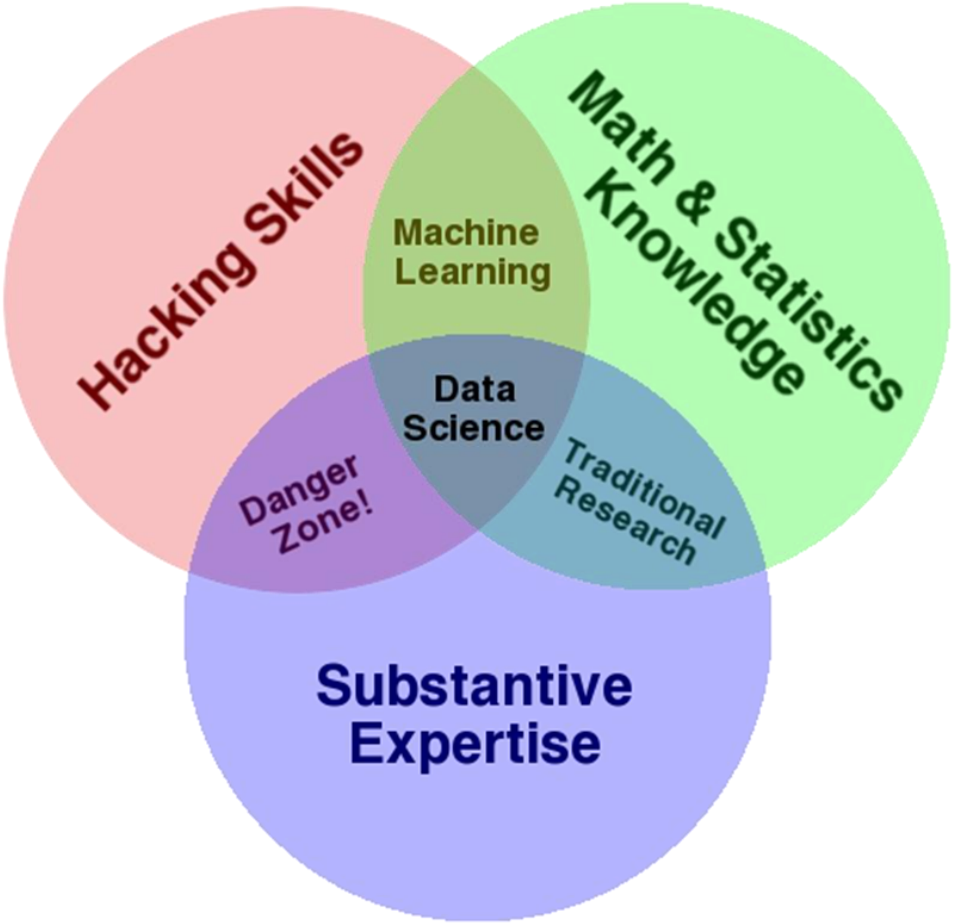
- Aggregation and Statistics
  - Data warehousing and OLAP
- Indexing, Searching, and Querying
  - Keyword based search
  - Pattern matching (XML/RDF)
- Knowledge discovery
  - Data Mining
  - Statistical Modeling
- Data Driven
  - Predictive Analytics
  - Deep Learning

Big Data & Data Science

- "… the sexy job in the next 10 years will be statisticians," Hal Varian, Google Chief Economist
- The U.S. will need 140,000-190,000 predictive analysts and 1.5 million managers/analysts by 2018. McKinsey Global Institute's June 2011
- New Data Science institutes being created or repurposed – NYU, Columbia, Washington, UCB,...
- New degree programs, courses, boot-camps:
  - e.g., at Berkeley: Stats, I-School, CS, Astronomy…
  - One proposal (elsewhere) for an MS in "Big Data Science"
  - Plans for Data Science Stream at AUST
  - RDA-CODATA School of Research Data Science

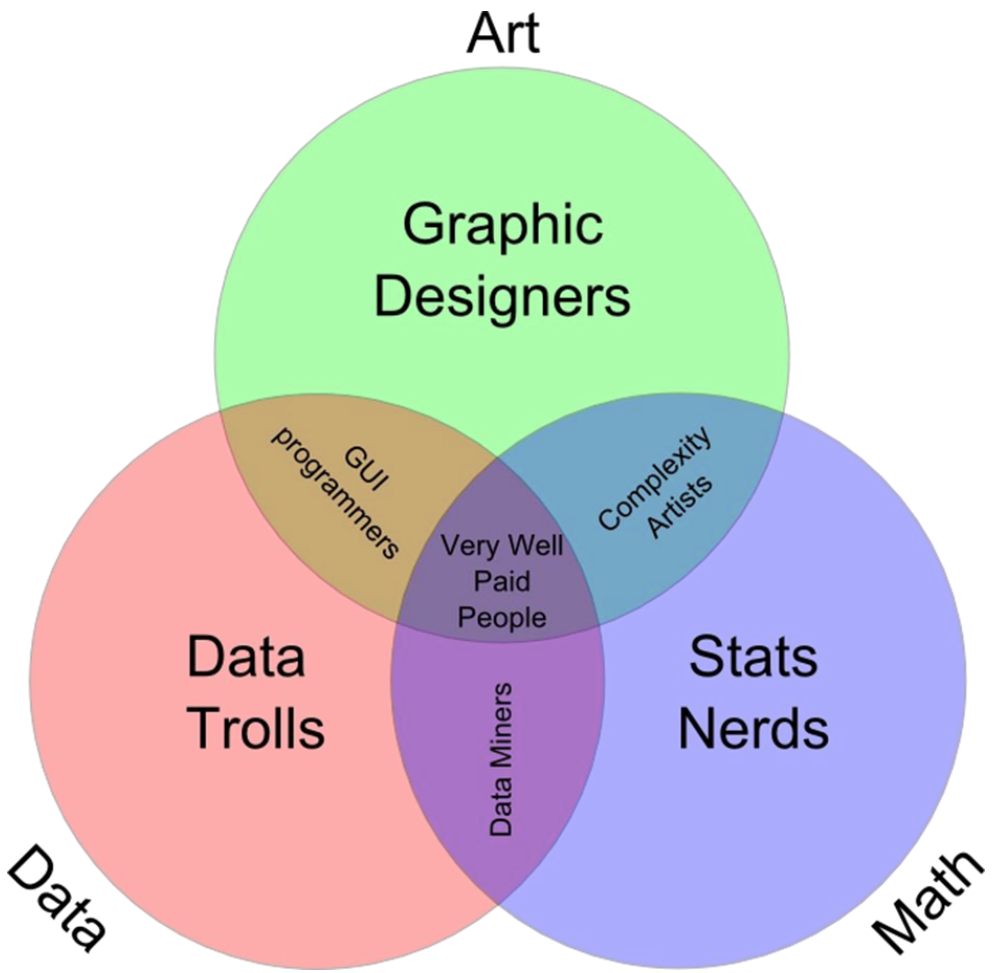
What is Data Science?

- Some definitions link computational, statistical, and substantive expertise.



What is Data Science?

- Other definitions focus more on technical skills alone.



What is Data Science?

- An area that manages, manipulates, extracts, and interprets knowledge from tremendous amount of data
- Data science (DS) is a multidisciplinary field of study with goal to address the challenges in big data
- Data science principles apply to all data – big and small

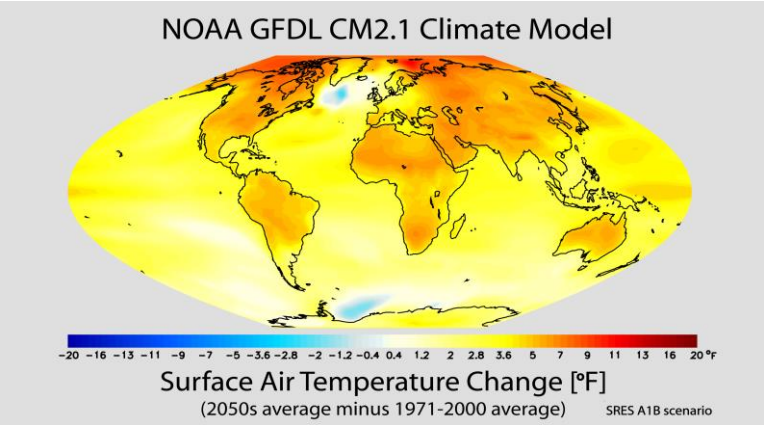
What is Data Science?

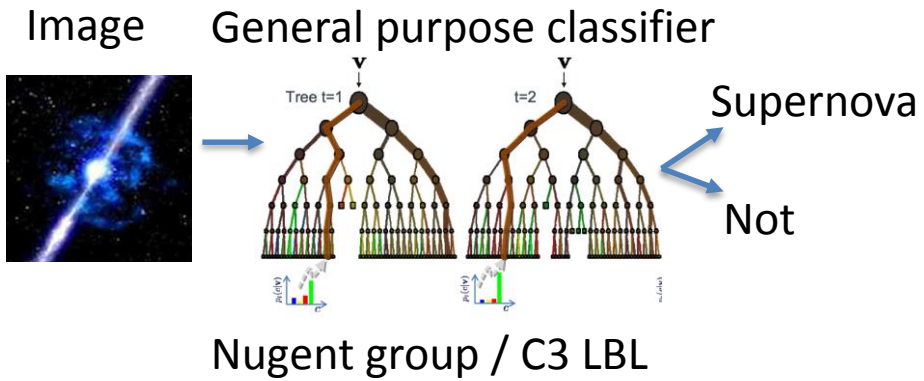
- Theories and techniques from many fields and disciplines are used to investigate and analyze a large amount of data to help decision makers in many industries such as science, engineering, economics, politics, finance, and education
  - Computer Science
    - Pattern recognition, visualization, data warehousing, High performance computing, Databases, AI
  - Mathematics
    - Mathematical Modeling
  - Statistics
    - Statistical and Stochastic modeling, Probability.

Data Science Vs Analysis Vs Software Delivery

Component	Traditional Analysis	Traditional Software Delivery	Data Science
Tools	SAS, R, Excel, SQL, in-house tools	Java, source control, Linux, continuous integration, unit testing, bug reports and project management	R, Java, scientific Python libraries, Excel, SQL, Hadoop, Hive, Pig, Mahout and other machine learning libraries, github for source control and issue management
Analytical Methods	Regressions, classifications, measuring prediction accuracy and coverage/error, sampling	N/A	Classification, clustering, similarity detection, recommenders, unsupervised and supervised learning, small- and large-scale computations, measuring prediction accuracy and coverage/error
Team Structure	Statisticians, Mathematicians, Scientists	Developers, Project Managers, Systems Engineers	Mathematicians, Statisticians, Scientists, Developers, Systems Engineers
Time Frame	Either: • Usually on-going research and discovery within a team in the organization Or: • Specific project to determine answers	Regular software release cycle, continuous delivery, etc.	Either: • Discovery/learning phase leading to product development Or: • On-going research and product invention/improvement

Contrast: Scientific Computing

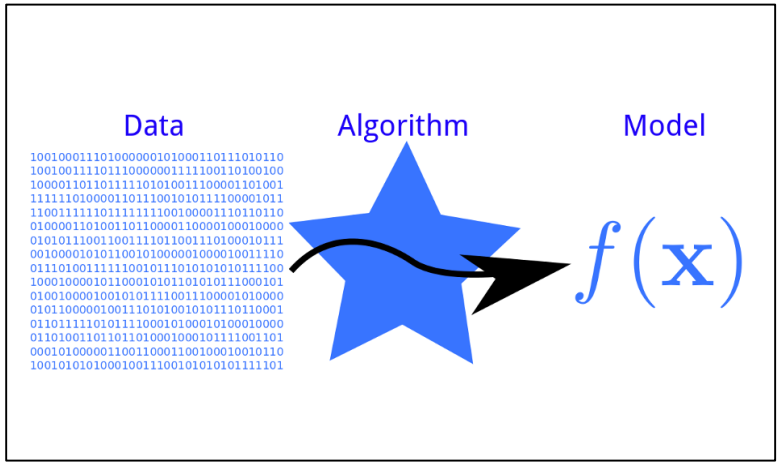




Scientific Modeling
Physics-based models
Problem-Structured
Mostly deterministic, precise
Run on Supercomputer or High-end Computing Cluster

Data-Driven Approach
General inference engine replaces model
Structure not related to problem
Statistical models handle true randomness, and **un-modeled complexity**.
Run on cheaper computer Clusters (EC2)

Contrast: Machine Learning



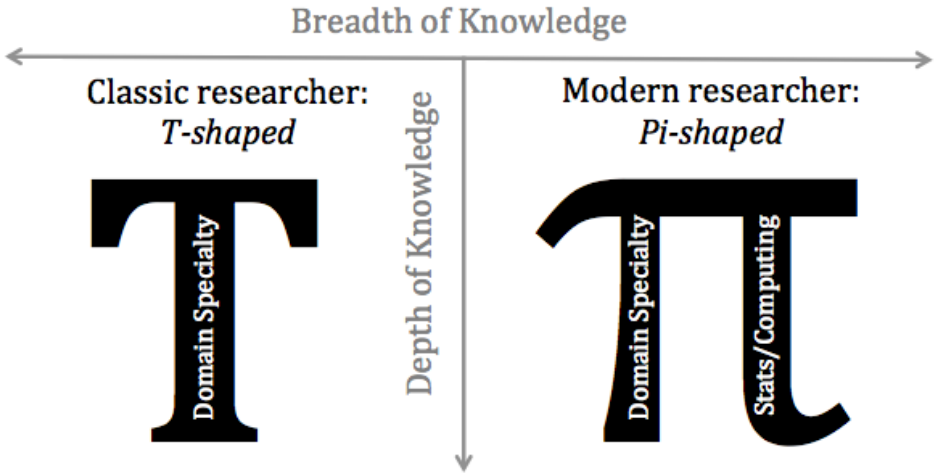
Machine Learning	Data Science
Develop new (individual) models	Explore many models, build and tune hybrids
Prove mathematical properties of models	Understand empirical properties of models
Improve/validate on a few, relatively clean, small datasets	Develop/use tools that can handle massive datasets
Publish a paper ☺	Take action!

Contrast: 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

Data Science & Academia

- In the words of Alex Szalay, these sorts of researchers must be "Pi-shaped" as opposed to the more traditional "T-shaped" researcher. In Szalay's view, a classic PhD program generates T-shaped researchers: scientists with wide-but-shallow general knowledge, but deep skill and expertise in one particular area. The new breed of scientific researchers, the data scientists, must be Pi-shaped: that is, they maintain the same wide breadth, but push deeper both in their own subject area and in the statistical or computational methods that help drive modern research:



Data Science & Academia

- In a post by Jake Vanderplas in 2014 related to SciFoo discussion on: ***Academia and Data Science***, the following questions below were discussed.
- I encourage you to develop your own thoughts on them and come up with your assessment

  - Where does Data Science fit within the current structure of the university & research institutions?
  - What is it that academic data scientists want from their career? How can academia offer that?
  - What drivers might shift academia toward recognizing & rewarding data scientists in domain fields?
  - Recognizing that graduates will go on to work in both academia and industry, how do we best prepare them for success in both worlds?

Data Science Applications

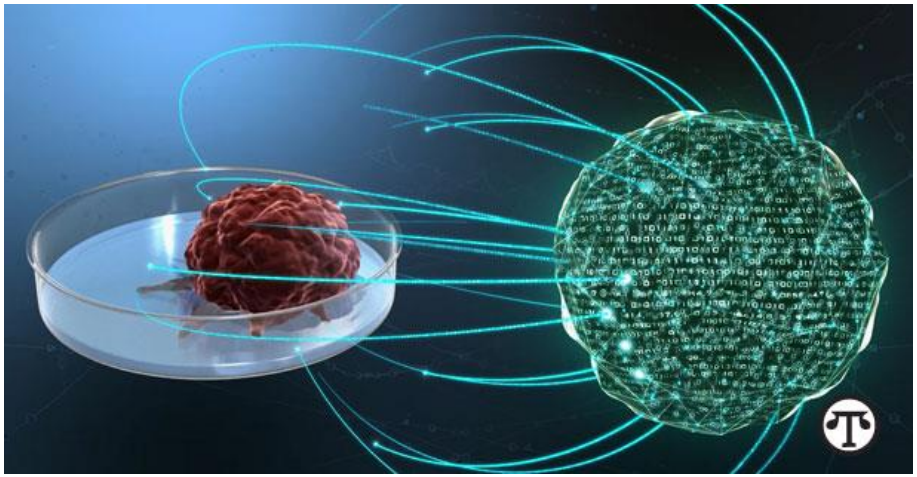
	Business	Health Care	Urban Leaving
Summary	From car design to insurance to pizza delivery, businesses are using data science to optimize their operations and better meet their customers' expectations.	Tomorrow's healthcare may look more efficient thanks to things like electronic health records. It also may look a lot more effective. Reduced readmissions, better care, and earlier detection are on the horizon.	For the first time in human history, more people live in cities than in suburban or rural areas. An emerging field called "urban informatics" combines data science with the unique challenges facing the world's growing cities
What is happening?	Two-Way Street for the Ford Focus Electric Car	Reducing Hospital Readmissions	Taking on Megacity Traffic
	Better Fraud Detection Boosts Customer Satisfaction	Better Point-of-Care Decisions	Fighting Crime with Data "predictive policing"
	E-Commerce Insights: Domino's Secret Sauce		
What is possible	Using Social Data to Select Successful Retail Locations .	Medical Exams by Bathroom Mirrors	Instrumenting cities

Contrast: Computational Sciences

- Is there a contrast between Data Science and Computational Science?

Data Science: Case Study
Cancer Research

- Cancer is an incredibly complex disease; a single tumor can have more than **100 billion cells**, and each cell can acquire mutations individually. The disease is always changing, evolving, and adapting.
- Employ the power of big data analytics and high-performance computing.
- Leverage sophisticated pattern and machine learning algorithms to identify patterns that are potentially linked to cancer
- Huge amount of data processing and recognition



Data Science: Case Study
Health Care

Stanford MEDICINE | News Center

FRONT PAGE ALL NEWS TOPICS MULTIMEDIA

Email Share 73 Tweet

Search News... Go

Stanford Medicine, Google team up to harness power of data science for health care

Stanford Medicine will use the power, security and scale of Google Cloud Platform to support precision health and more efficient patient care.

AUG 8 2016

Stanford Medicine and Google are working together to transform patient care and medical research through data science.

The new collaboration combines Stanford Medicine's excellence in health-care research and clinical work with Google's expertise in cloud technology and data science. Stanford's forthcoming Clinical Genomics Service, which puts genomic sequencing into the hands of clinicians to help diagnose disease, will be built using Google Genomics, a service that applies the same technologies that power Google Search and Maps to securely store, process, explore and share genomic data sets.

Lloyd Minor, dean of the School of Medicine, says the collaboration with Google marks a "milestone for the future of patient care and research."
Glenn Matsumura

Stanford Medicine includes the Stanford School of Medicine, Stanford Health Care and Stanford Children's Health. Together, Stanford Medicine and Google will build cloud-based applications for exploring massive health-care data sets, a move that could transform patient care and medical research.

"Stanford Medicine and Google are committing to major investments in preventing and curing diseases that afflict ordinary people worldwide. We're proud to be setting this milestone for the future of patient care and research," said Lloyd Minor, MD, dean of the School of Medicine.

The agreement — considered key to Stanford Health Care's development of the Clinical Genomics Service — makes Google Inc. a formal business associate of Stanford Medicine. As such, Google and Stanford will both comply with the Health Insurance Portability and Accountability Act, a federal law that regulates the privacy and security of medical information. HIPAA requires that Stanford Medicine patient data stored on Google Cloud Platform servers stay private. Patient information will be encrypted, both in transit and on servers, and kept on servers in the United States.

LEADING IN PRECISION HEALTH

Stanford Medicine is leading the biomedical revolution in precision health, defining and developing the next generation of care that is proactive, predictive and precise.

Learn more

A LEGACY OF INNOVATION

Stanford Medicine's unrivaled atmosphere of breakthrough thinking and interdisciplinary collaboration has fueled a long history of achievements.

View timeline

- Stanford Medicine, Google team up to harness power of data science for health care
- Stanford Medicine will use the power, security and scale of Google Cloud Platform to support precision health and more efficient patient care.
- Analyzing genetic data
- Focusing on precision health
- ***Data as the engine that drives research***

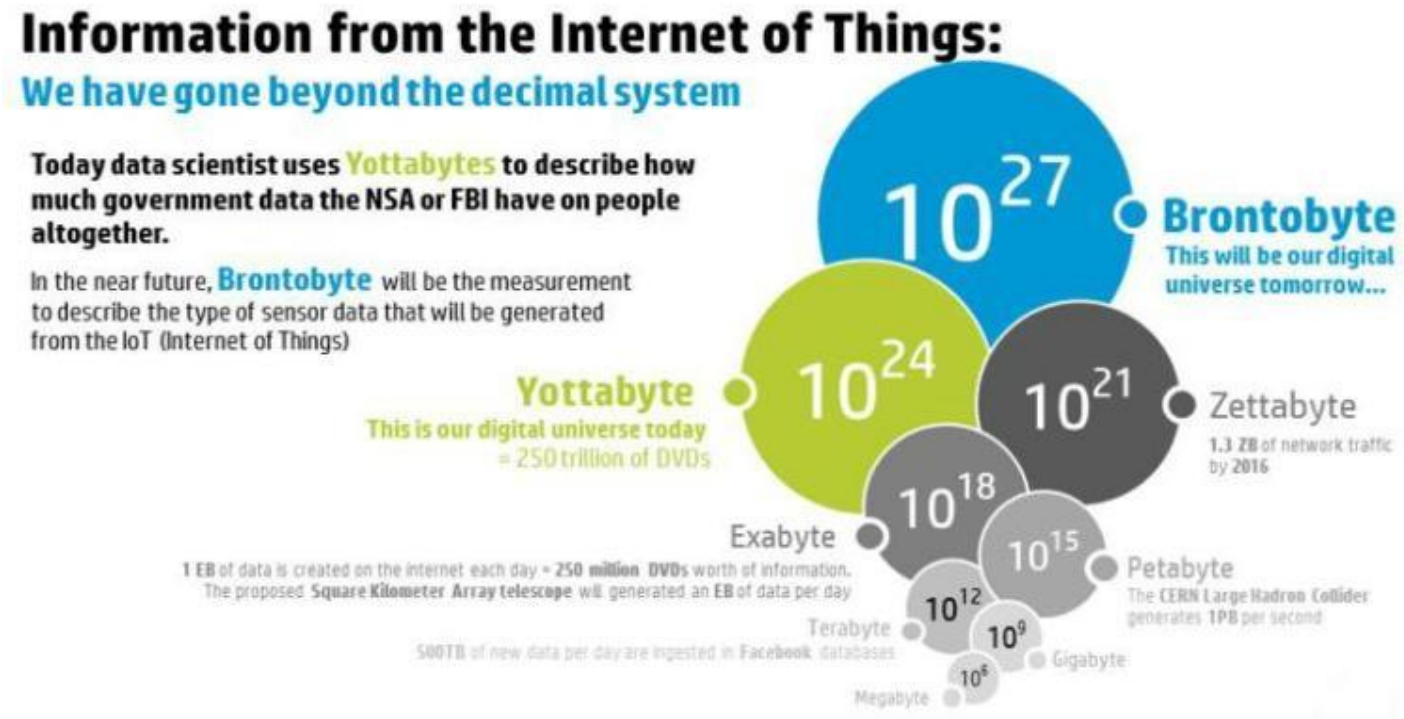
http://med.stanford.edu/news/all-news/2016/08/stanford-medicine-google-team-up-to-harness-power-of-data-science.html

Data Science: Case Study Elections

- The Obama campaigns in 2008 and 2012 are credited for their successful use of social media and data mining.
- Micro-targeting in 2012
  - http://www.theatlantic.com/politics/archive/2012/04/the-creepiness-factor-how-obama-and-romney-are-getting-to-know-you/255499/
  - http://www.mediabizbloggers.com/group-m/How-Data-and-Micro-Targeting-Won-the-2012-Election-for-Obama---Antony-Young-Mindshare-North-America.html
- Micro-profiles built from multiple sources accessed by aps, real-time updating data based on door-to-door visits, focused media buys, e-mails and Facebook messages highly targeted.
- 1 million people installed the Obama Facebook app that gave access to info on “friends”.

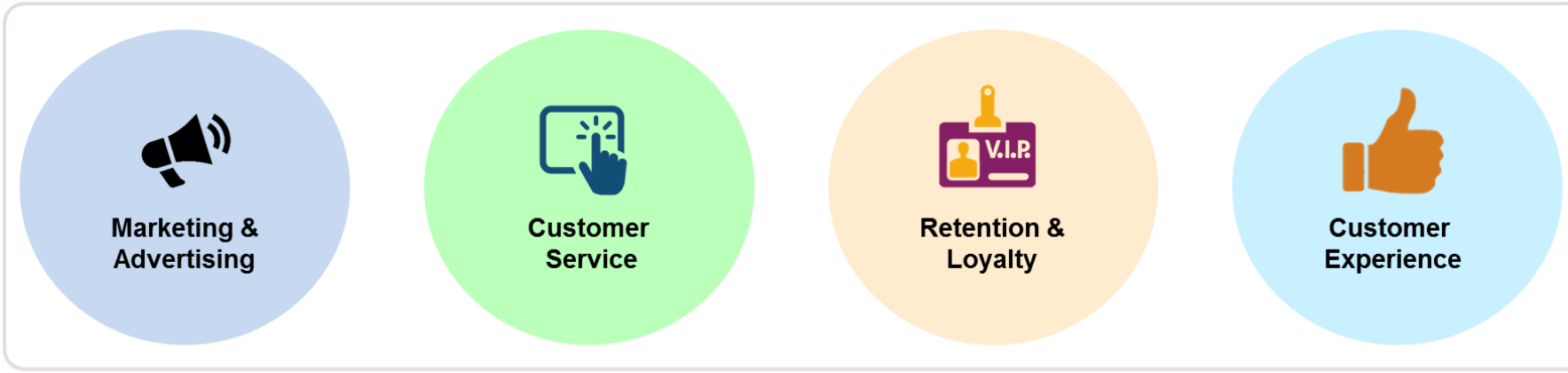
Data Science: Case Study
Internet of Things (IoT)

- The Internet of Things is rapidly growing. It is predicted that more than 25 billion devices will be connected by 2020.



- The Internet of Things (IOT) will soon produce a massive volume and variety of data at unprecedented velocity. If "Big Data" is the product of the IOT, "Data Science" is it's soul.

Data Science: Case Study
Customer Analytics



Leveraging customer data to move ever closer to the elusive goal of truly personalized marketing: the right offer, at the right time, in the right location and context, to the right person.

By capturing and analyzing the data from customer touch points within an organization, companies can identify customer pain points and issues proactively and update their customer service FAQs or other communications with existing customers.

Using customer data and analytics, these companies deploy and refine predictive models that help them retain customers with proactive approaches. Investments, in terms of offers and upgrades, can be made at the right time to increase the likelihood of retaining desirable customers.

The experience that customers have with companies matters a great deal. Other recent research has highlighted the critical connection between experience and company financial performance.

Essential Points

- **Big Data has given rise to Data Science**
- **Data science is rooted in solid foundations of mathematics and statistics, computer science, and domain knowledge**
- **Sexy profession – Data Scientists ☺**
- **Not every thing with data or science is Data Science!**
- **The use cases for Data Science are compelling**

Conclusion

In this section you have learned

- **What Big Data Challenges are**
- **What exactly is Data Science and what do Data Scientists do**
- **Data Science contrasted with other disciplines**
- **Case Study & Use Cases**

Questions?

http://www.ign.com/articles/2015/12/16/star-wars-the-force-awakens-review