IMD0105 - Special Issues in Information Technology VI

Big Data Foundations

Natal-RN
Fevereiro de 2017

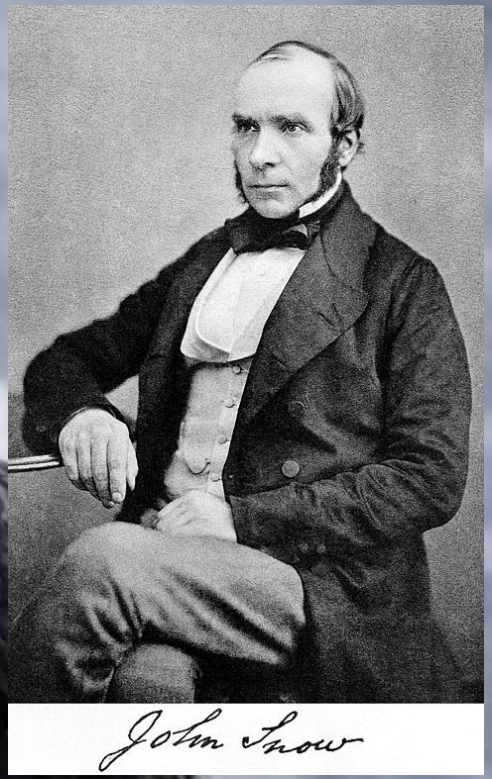
Agenda

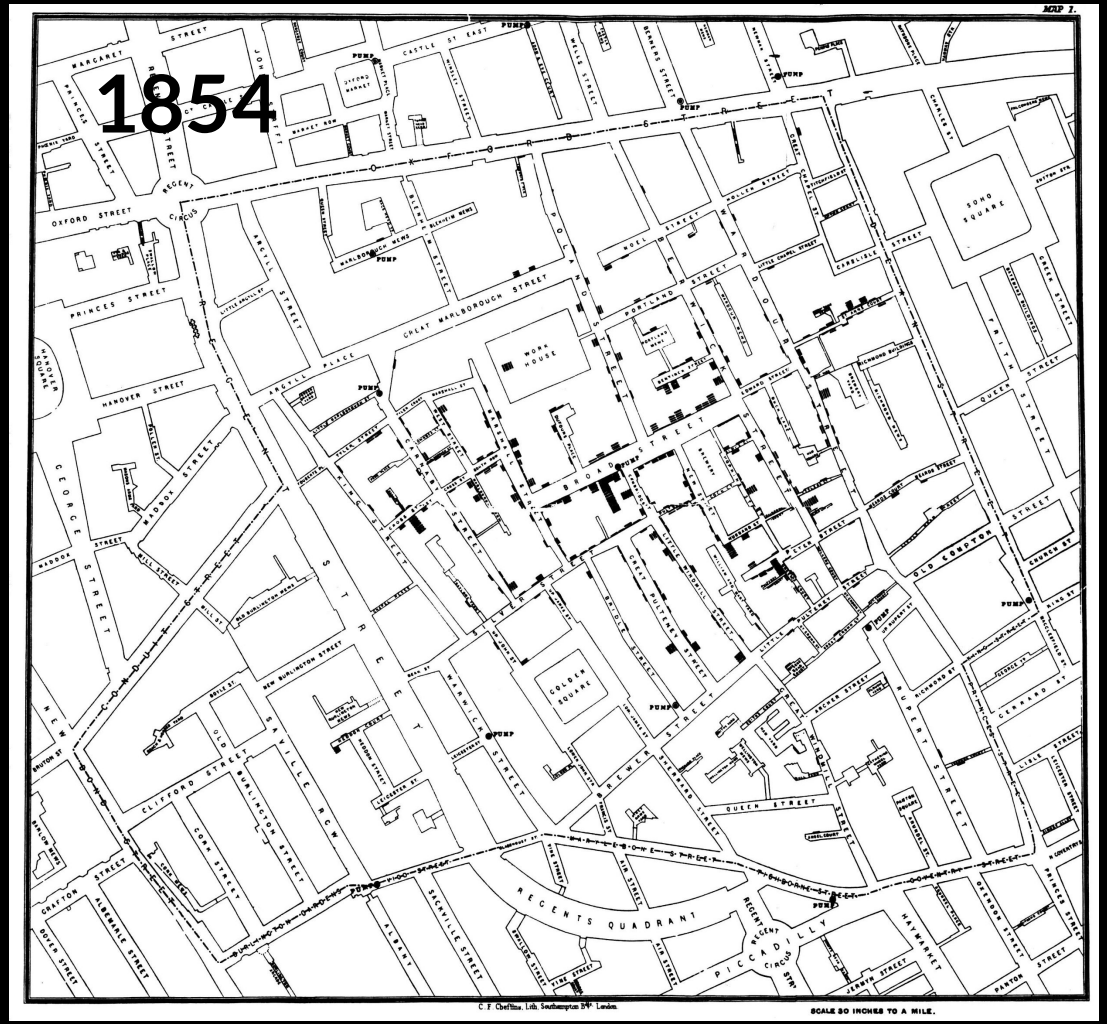
Brief introduction about Big Data

Previously on last class (...)

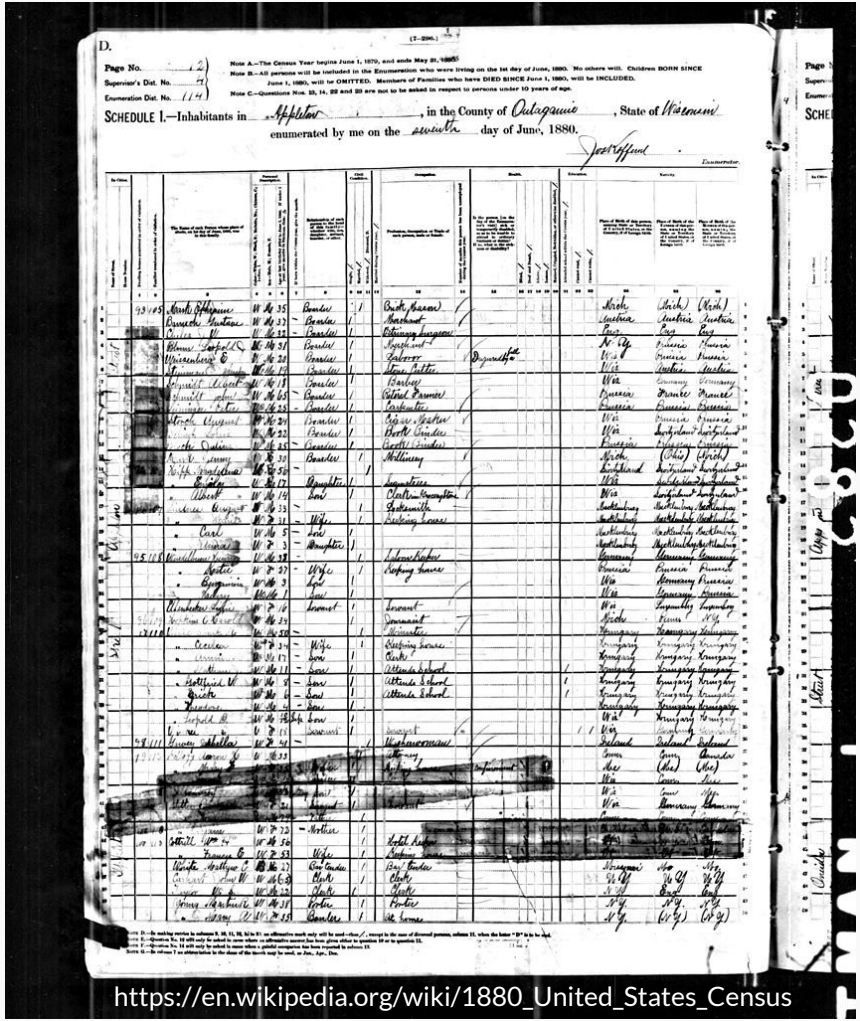
http://www.winshuttle.com/big-data-timeline/

John Snow, London, 1854





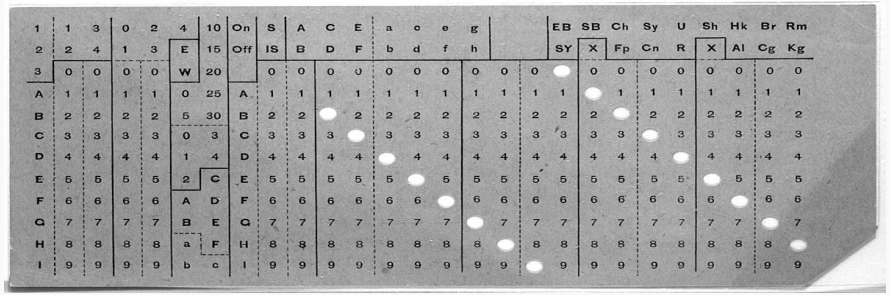
https://en.wikipedia.org/wiki/John_Snow



U.S. census (1880)

- 50 million people
- Age, sex, occupation, education level
- It took 8 years to be tabulate

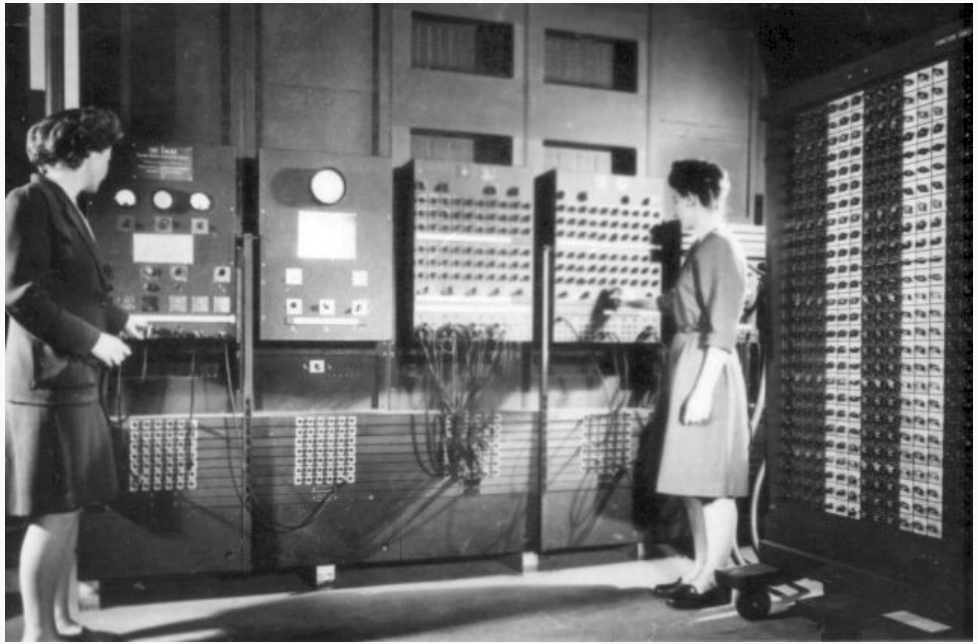
https://en.wikipedia.org/wiki/Tabulating_machine



https://en.wikipedia.org/wiki/1890_United_States_Census

U.S. census (1890)

- Tabulating machine.
- 63 million people
- The results were announced after only six weeks of processing.
- Without this invention, experts had estimated, the 1890 census would have taken 13 years to fully tabulate.



The population boom (1932)

- 123 million people.
- Information overload continued with the boom in the US population
- It demanded more thorough and organized record-keeping.

The Effect on Libraries (1940)

- Libraries, the original source of data organization and storage, had to adapt their storage methods to meet the quickly increasing demand of new publications and research.

The First Warning of Data's Storage (1944)

- American University Libraries were doubling in size every sixteen years.

- Yale Library (2040)
  - 200,000,000 volumes,
  - 6,000 miles of shelves
  - 6,000 persons.

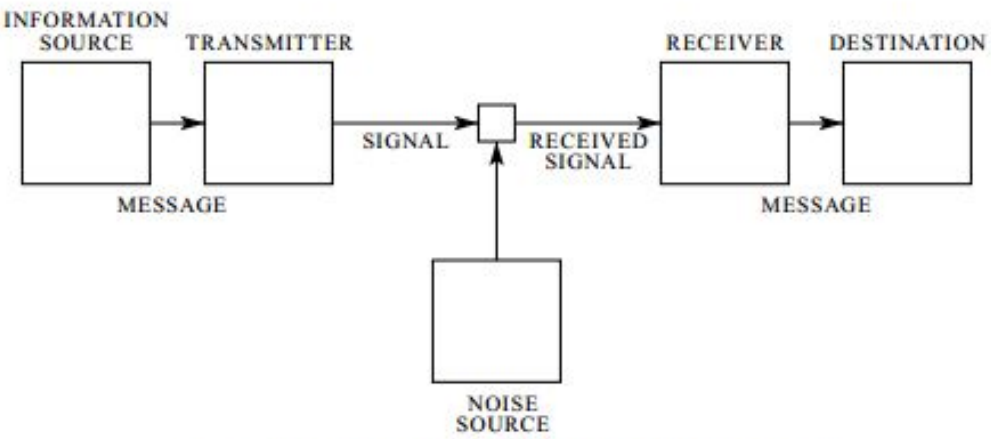
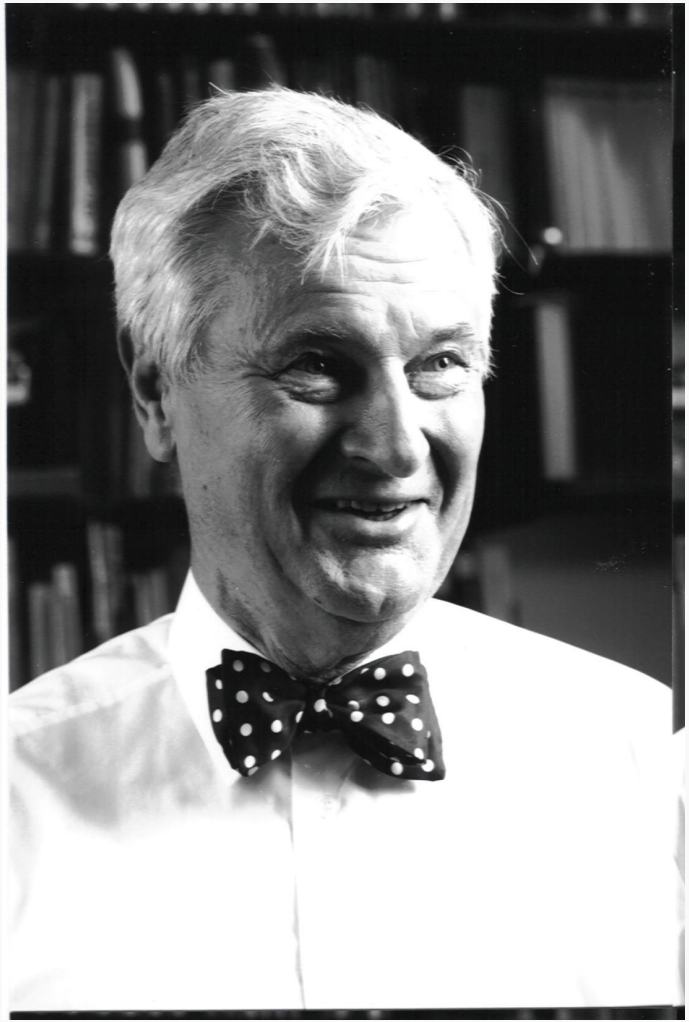


Fig. 1—Schematic diagram of a general communication system.

Shannon's Information Theory (1948)

- Claude Shannon published a framework for determining the minimal data requirements to transmit information over noisy (imperfect) channels.
- Without this understanding, data would be "bigger" than it is today



Virtual Memory (1956)

- The concept of virtual memory was developed by German physicist Fritz-Rudolf Güntsch as an idea that treated finite storage as infinite.

Scientific Knowledge Expands (1962)

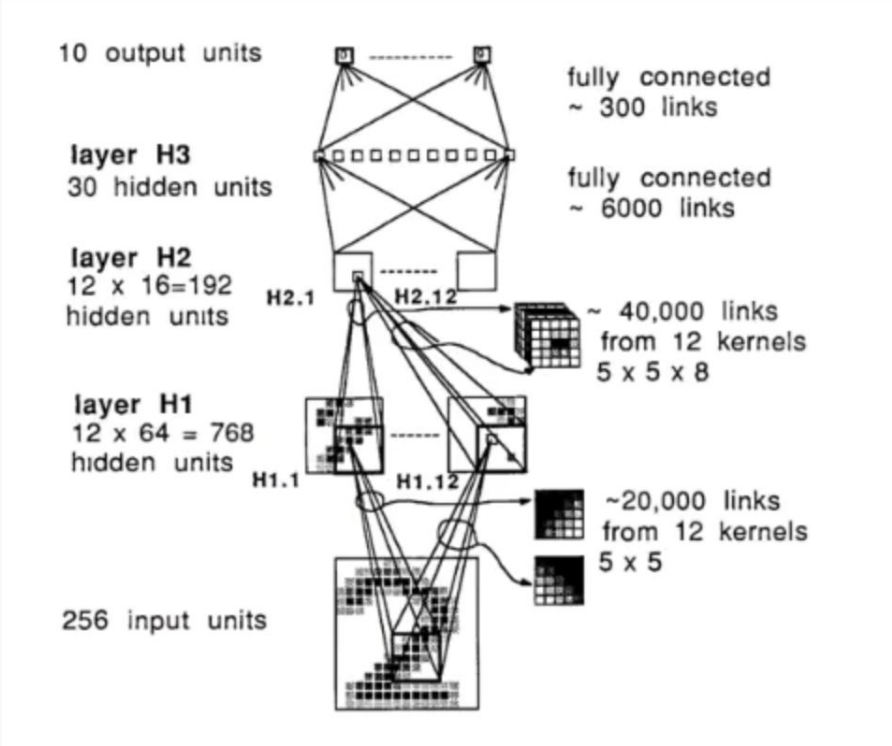
- Derek Price publishes "Science Since Babylon"
- He concludes that the number of new journals has grown exponentially rather than linearly.
- This is now better known as the "law of exponential increase".

The Rise in Two-Way Communication (1975)

- The Information Flow Census, conducted by the Ministry of Posts and Telecommunications in Japan, started tracking the volume of information circulating in that country
- The study found that information supply greatly exceeded information consumption, and the demand for one-way communication had stagnated.

Information Growth and the Broadcasting Industry (1983)

- Companies were beginning to use data to provide answers for better business decisions.
- The massive information growth was credited to the expansion of the broadcasting industry.



Neural Networks debuted (1990)

Matan, Ofer, et al. Handwritten character recognition using neural network architectures. Proceeding of the 4th USPS Advanced Technology Conference.

The bird of the Internet (1991)

- Courtesy of Sr Tim Berners-Lee, data and information can now be posted online for the first time.

The future of data storage (1996)

- Digital storage became more cost-effective for store data than paper.
- Michael Lesk published How much information is there in the world? So in only a few years, we will be able [to] save everything

Google

Founded in 1996

The first time (1997)

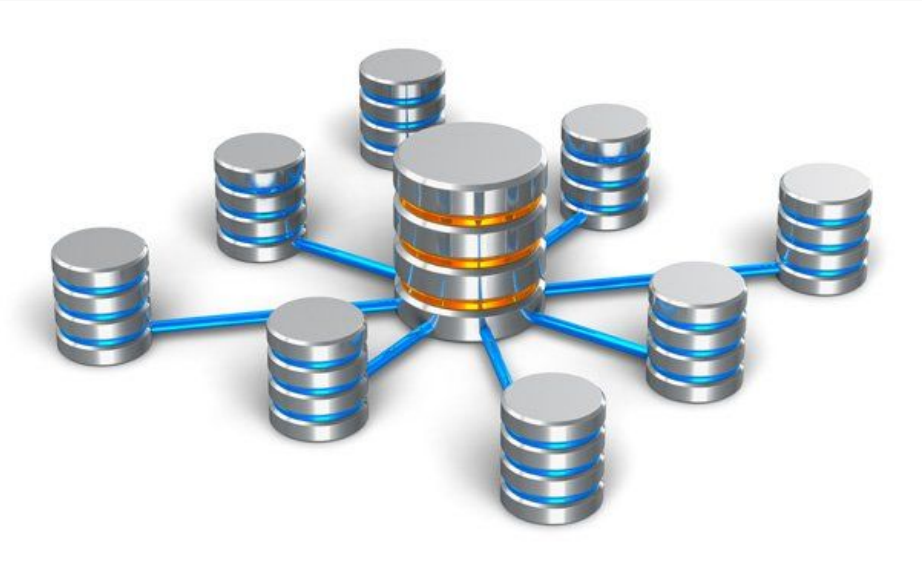
- The term "Big Data" was used for the first when researchers M. Cox and D. Ellworths wrote an article identifying that the rise of data was becoming an issue for current computer systems.

Internet of Things (1999)

- The term “Internet of Things” was coined by British entrepreneur Kevin Ashton, Co-Founder of the Auto-ID Center at MIT, during a presentation linking the idea of Radio Frequency Identification (RFID) in supply chain to the internet world.

Software as a Service (SaaS) 2001

- The SaaS acronym first appears in an article in the Software & Information Industry's (SIIA) eBusiness Division.



The 3Vs (2001)

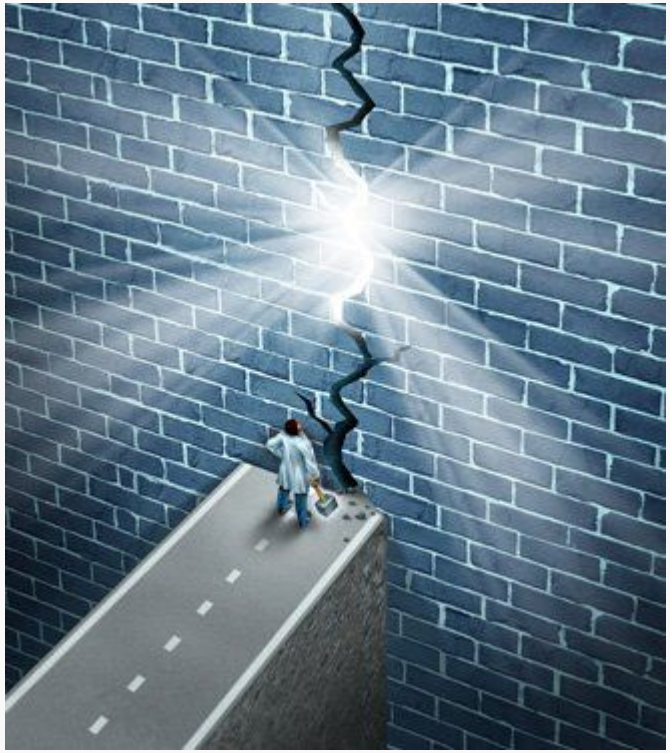
- Gartner Analytst, Doug Laney, published a research paper titled 3D Data Management: Controlling Data Volume, Velocity, and Variety. Even today, the “3Vs” are the generally-accepted dimensions of big data.

Web 2.0 (2005)

- "What is Web 2.0" was published by writer Tim O'Reilly, in which he stated that "data is the next Intel Inside and SQL is the new HTML".

An Open Source Solution to the Big Data Explosion (2006)

- Hadoop was created in 2006 out of the necessity for new systems to handle the explosion of data from the web.
- Free to download, use, enhance and improve, Hadoop is a 100% open source way of storing and processing data that enables distributed parallel processing of huge amounts of data across inexpensive, industry-standard servers that both store and process the data, and can scale without limits”.

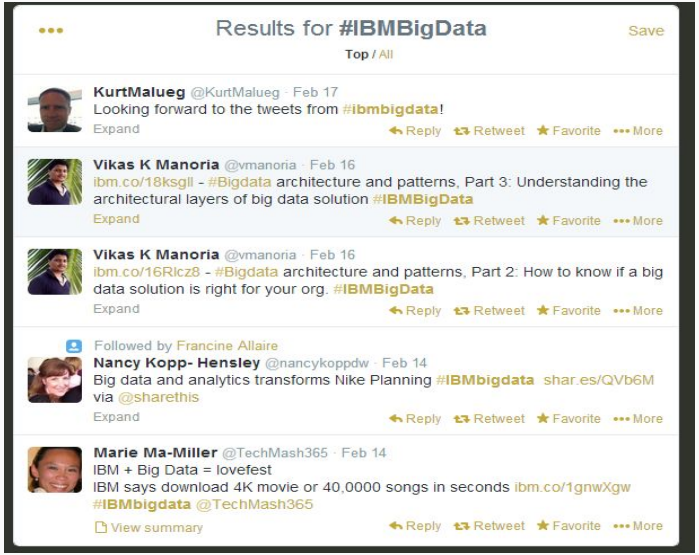


Revolutionary Breakthroughs (2008)

A group of computer science researchers published a paper titled Big Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science, and Society.

Big-data computing is perhaps the biggest innovation in computing in the last decade

A modest investment by the federal government could greatly accelerate its development and deployment.



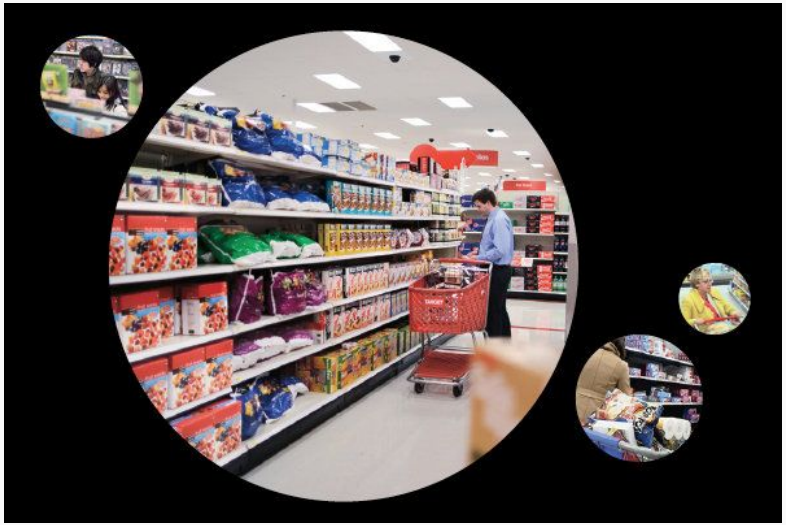
#IBMBigData (2011)

IBM introduced a Twitter hashtag, #IBMbigdata which expanded on their Big Data themed website that they built in 2008 in an effort to integrate it into their marketing.

http://www.zdnet.com/article/google-brain-simulator-teaches-itself-to-recognize-cats/

Deep Learning experiment at Google (2012)

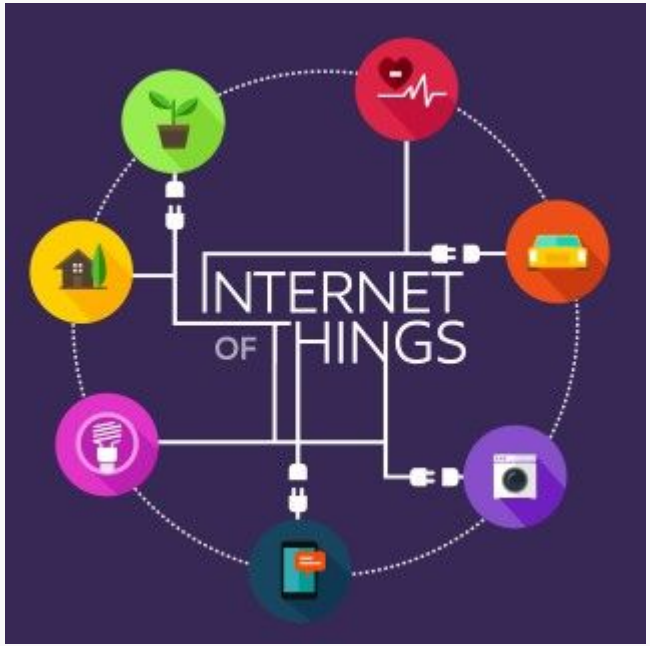
After seeing 10 million images from YouTube videos within three days, the 16,000-computer network, which had one billion connections, began to recognize cats, even though it had never been taught what a cat looked like.



http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp

How Companies Learn Your Secrets (2012)

"My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"



#The Year of the Internet of Things (IoT) 2014

The IoT has become a powerful force for business transformation, and its disruptive impact will be felt across all industries and all areas of society.

According to Gartner, there were 3.7 billion connected "things" in use in 2014

Calvin Klein

Knits in myriad varieties, including curly mohair coats and hand-stitched, multipaneled sweaters, were the focus of a disciplined but cozy collection in a soothing palette of earth tones and snow.

Mixed-knit jumpers blocked out in white, black and gray

Turtleneck sweater tops with chunky, sampler-scarf knit panels

Ralph Lauren

A collection distinguished by its sure-handed hybrid of refinement and ease, its muted pastels and buttery fabrics arguing for opulence but in no way overstating the case.

Read more: **Ralph Lauren Plays Polo**

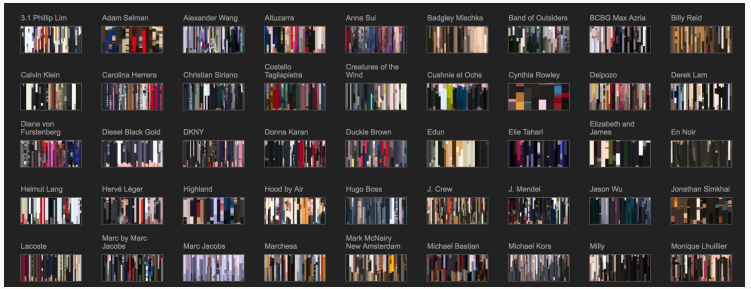
An upscale hippie dress under a distressed bomber

A massive, asymmetrical cape in soft pink

BIG DATA

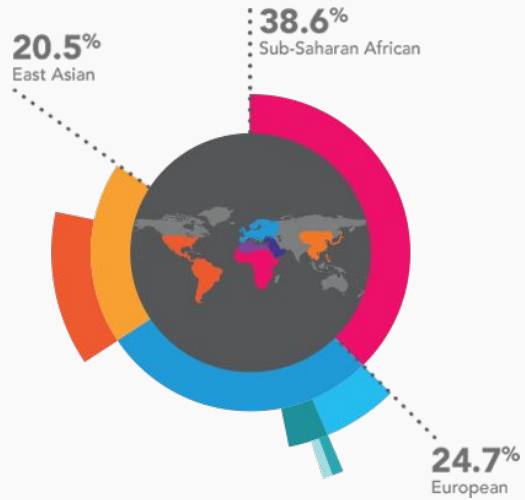
2014

FASHION WEEK MEN



23andMe

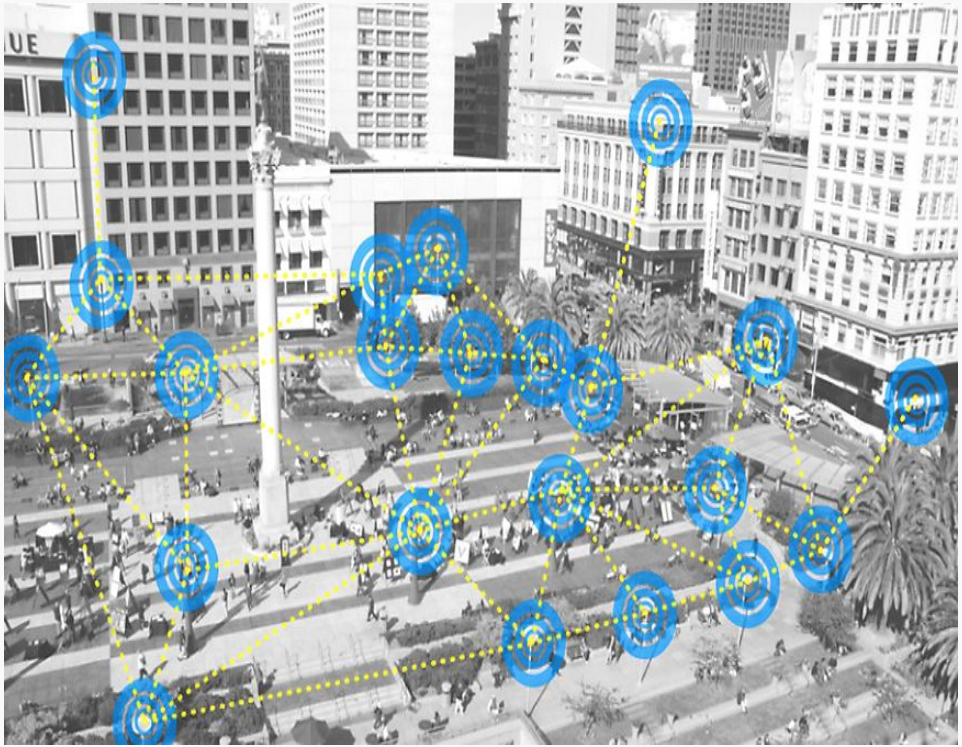
https://www.23andme.com/



23andMe has been selling a product with both ancestry and health-related components in Canada since October 2014

http://www.telegraph.co.uk/technology/news/10959864/Germanys-World-Cup-tactics-shaped-by-data.html

Germany’s 12th Man at the World Cup (2014)



http://www.forbes.com/sites/danielnewman/2016/08/15/big-data-and-the-future-of-smart-cities/#16c07f1d3f2d

Connect To The Cities Of The Future (2015)

With the growth of our population and the advent of ideas such as big data and the Internet of Things, the natural step cities will take is to become more interconnected.

Not only will this result in brighter streets, but the new lights will also be an interconnected system that will inform the city of each bulb's status.

2016

http://www.pokemongo.com/

https://www.youtube.com/watch?v=NrmMk1Myrxc

Why Netflix thinks its personalized recommendation engine is worth $1 billion per year (2016)

“Consumer research suggests that a typical Netflix member loses interest after perhaps 60 to 90 seconds of choosing"

The user either finds something of interest or the risk of the user abandoning our service increases substantially

https://datafloq.com/read/the-top-7-big-data-trends-for-2017/2493

An app for blind people identifies and reads out objects in their surroundings

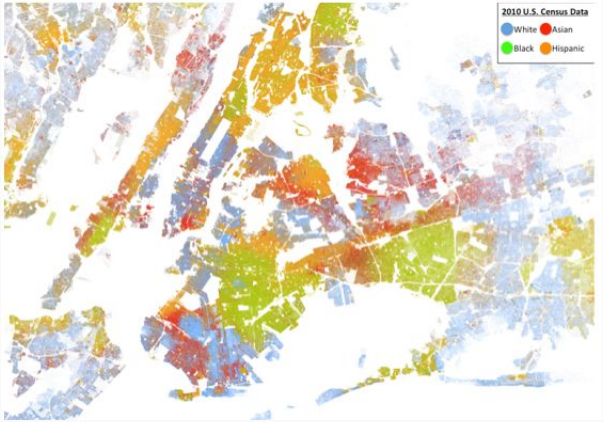
https://goo.gl/J9O1me

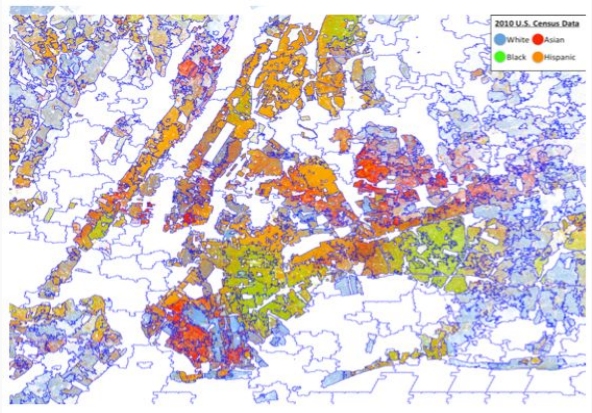
PathoMAP

http://www.pathomap.org/

TODAY!

Patho map to human ancestry pipeline in Smart City

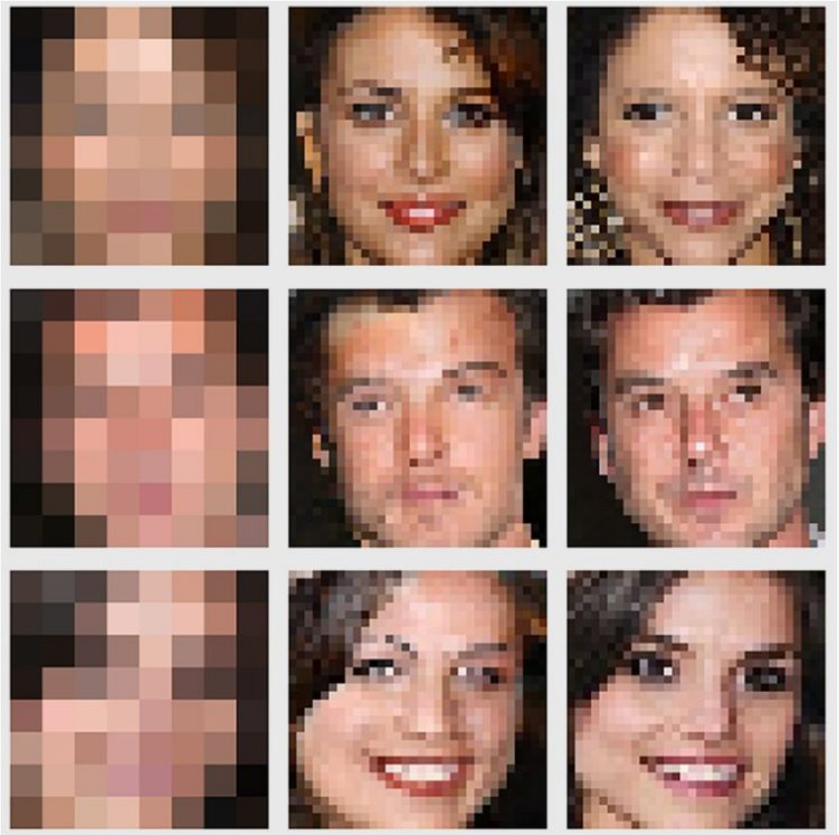




Vid2Speech: Speech Reconstruction from Silent Video

https://github.com/arielephrat/vid2speech

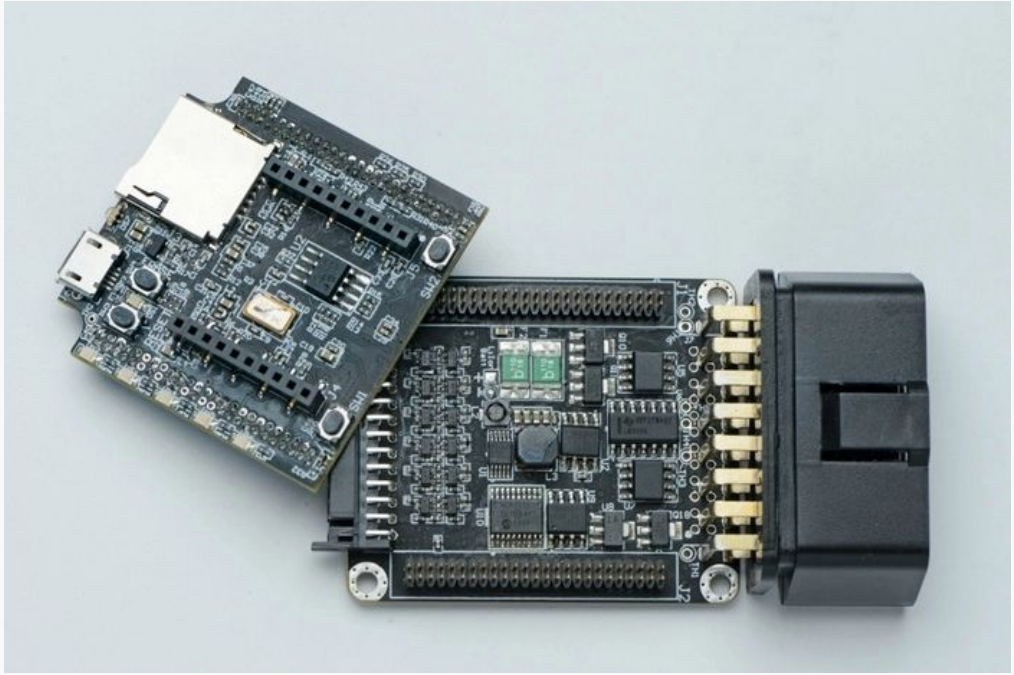
https://goo.gl/HPYCbK



Google Uses AI to ‘Rebuild’ a Portrait from an 8×8-Pixel Image

http://www.theverge.com/2017/2/7/14532206/google-brain-research-neural-networks-zoom-and-enhance-pixelated-images

IoT + Car Hacking + Big Data



http://hackaday.com/2017/02/21/first-look-macchina-m2/

KICKSTARTER

https://www.kickstarter.com/projects/1029808658/macchina-the-ultimate-tool-for-taking-control-of-y?token=e40d156d

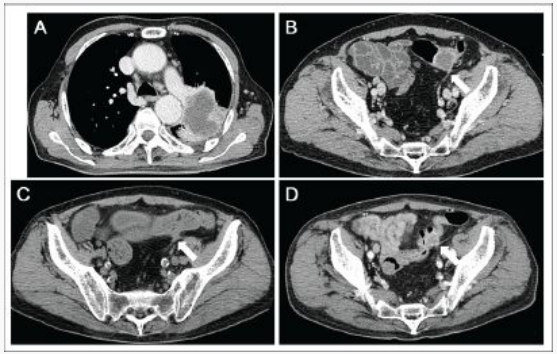
FISH BOWL

Models


Volume

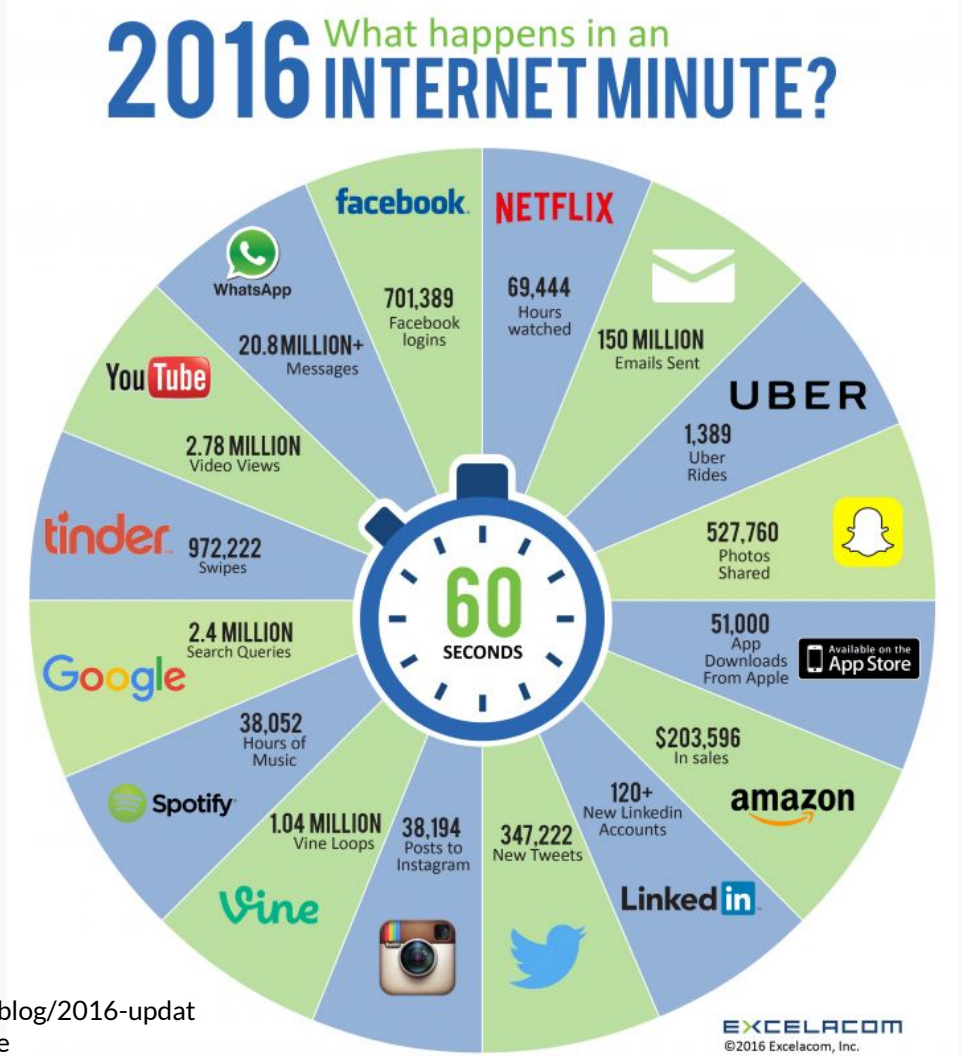
90% of the data in the world today
has been created in the last two years alone

Variety



YouTube

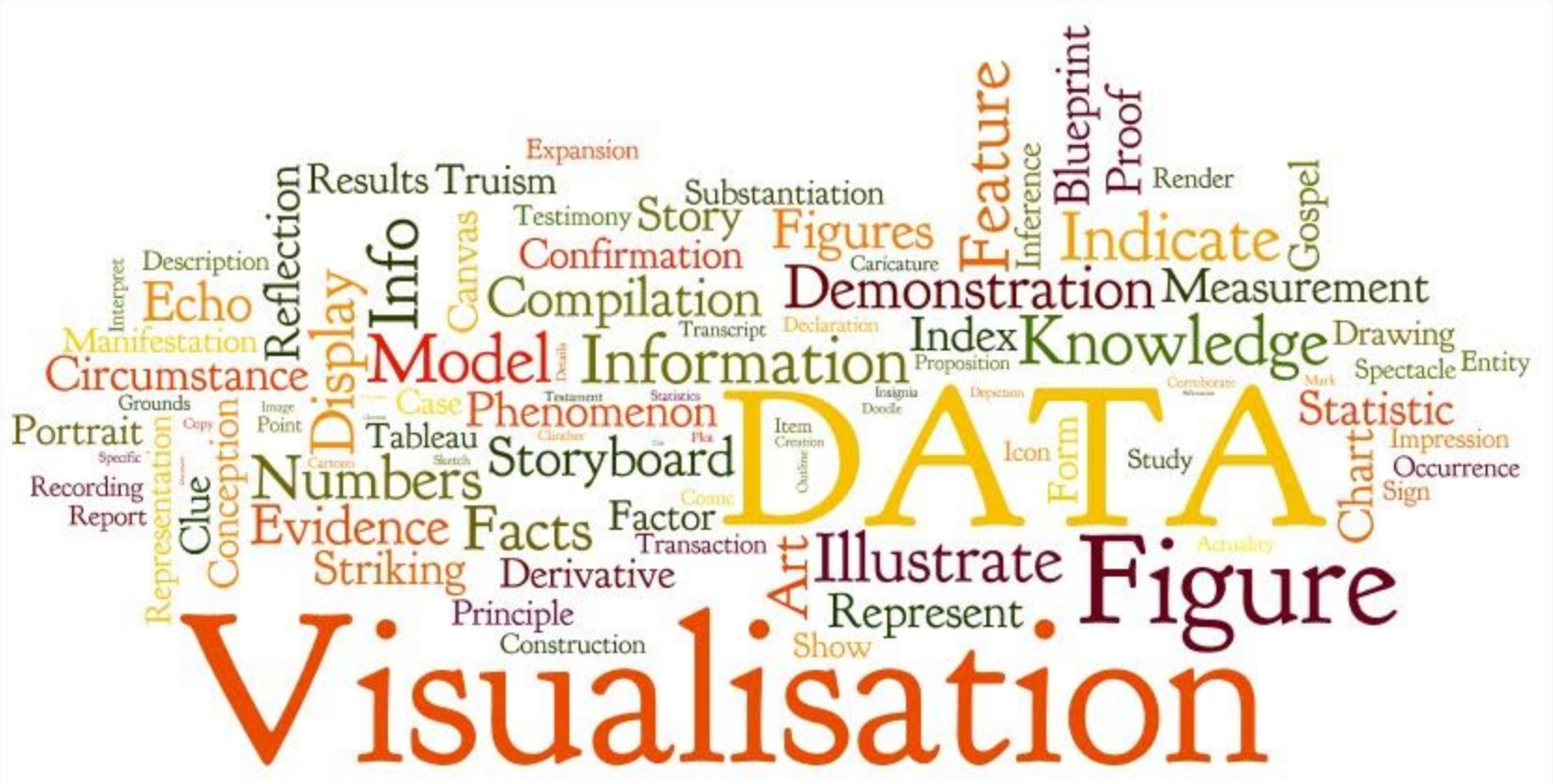
Velocity



http://www.excelacom.com/resources/blog/2016-update-what-happens-in-one-internet-minute


Veracity

Boatos.org


Visualisation
DATA
Figure
Illustrate
Represent
Information
Knowledge
Model
Display
Info
Evidence
Facts
Numbers
Storyboard
Phenomenon
Demonstration
Measurement
Compilation
Indicate
Figures
Feature
Blueprint
Proof
Render
Gospel
Inference
Story
Testimony
Substantiation
Results
Truism
Expansion
Reflection
Echo
Description
Interpret
Manifestation
Circumstance
Grounds
Portrait
Recording
Report
Representation
Clue
Conception
Point
Image
Tableau
Case
Striking
Derivative
Principle
Construction
Art
Show
Factor
Transaction
Item
Creation
Icon
Form
Study
Chart
Statistic
Impression
Occurrence
Sign
Actuality
Index
Proposition
Drawing
Spectacle
Entity
Declaration
Transcript
Canvas
Confirmation
Caricature
Insignia
Doodle
Testament
Statistics
Comic
Outline
Plot
Sketch
Specific
Copy

http://staging.kapowsoftware.com/assets/infographics/variety-of-big-data-sources-kapow-software-lg.png

Search and add articles

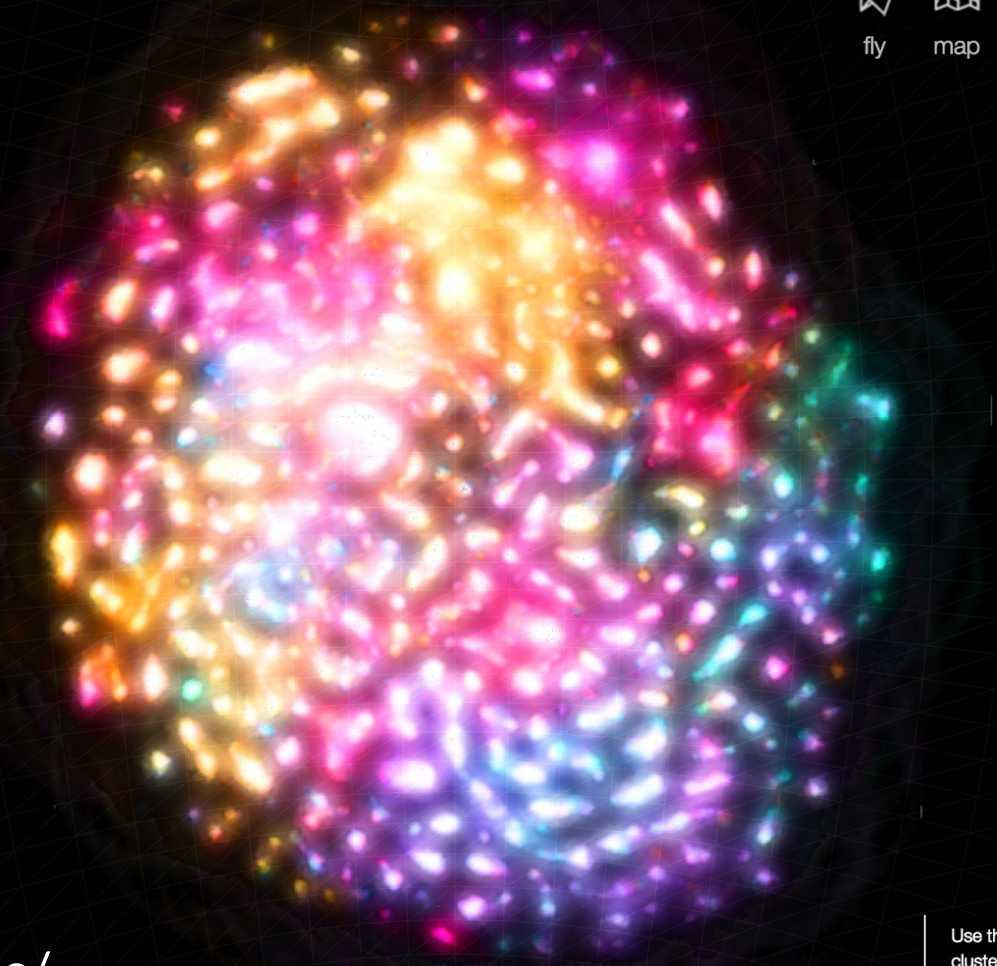
fly map home help about UI off full

Article links

Pick an article

Welcome.

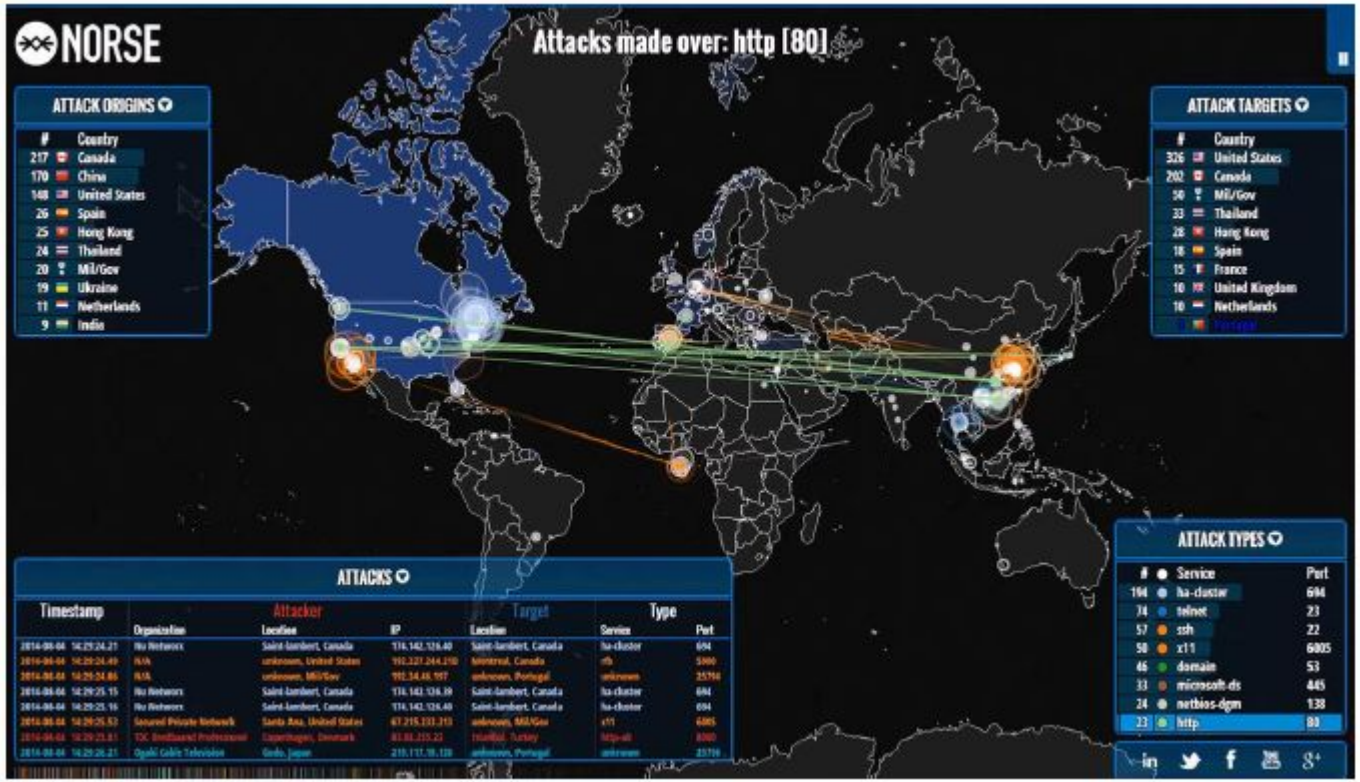
WikiGalaxy is a 3D web experiment that visualizes Wikipedia as a galactic web of information. With it I aim to show the world the beauty and variety of knowledge that is available at our fingertips.
I used 100,000 of 2014's most popular articles, all clustered with hyperlinks. In this world Wikipedia articles are stars, interests are nebulas and you are on a journey through knowledge.



http://wiki.polyfra.me/

Use the mouse to see a preview of articles in each cluster
Click anywhere on the map to fly there

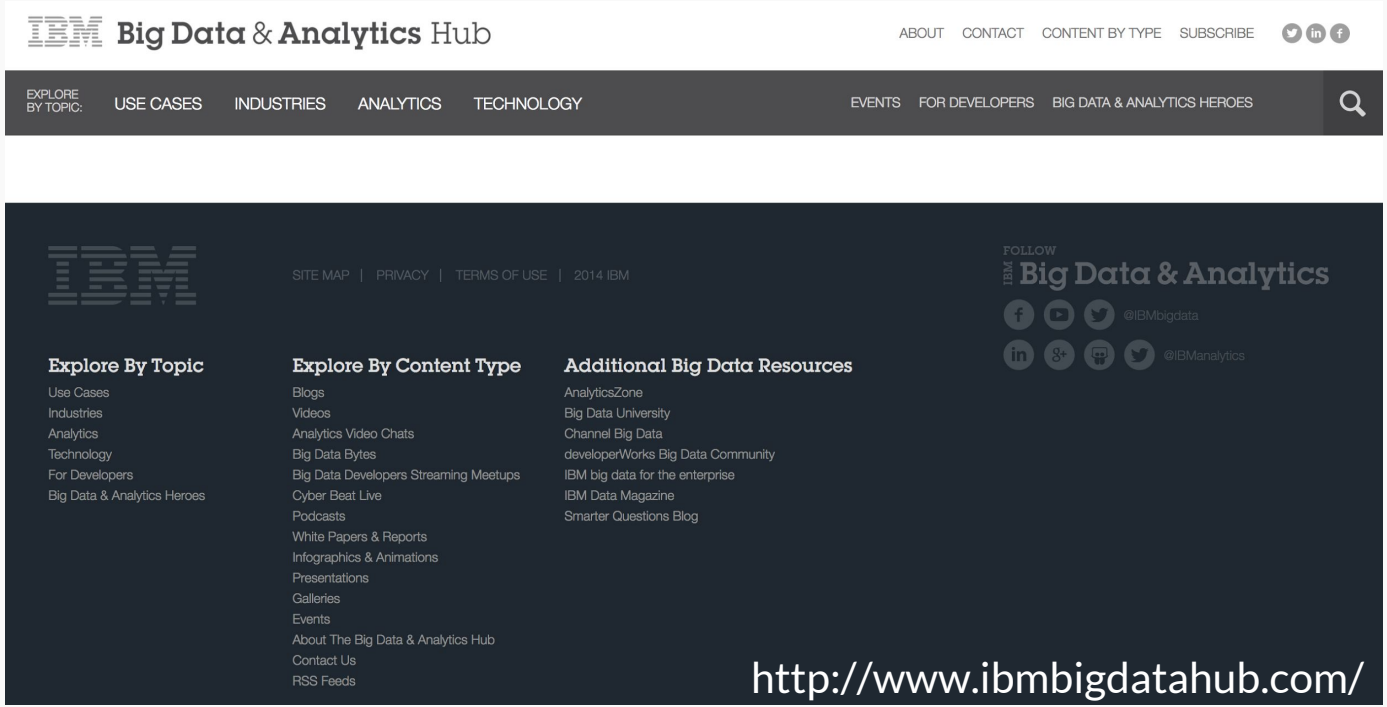
http://map.norsecorp.com/



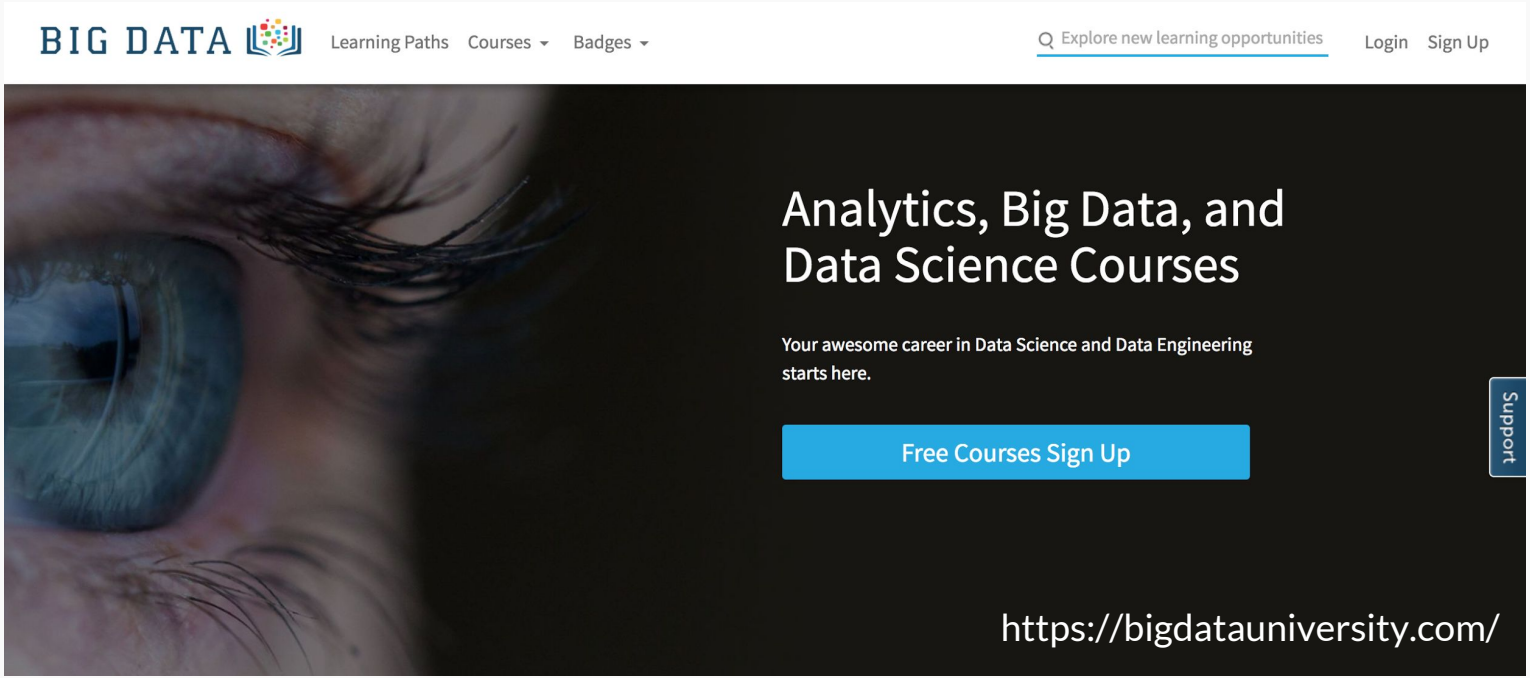
Gift to Carnival!!!!

- https://play.google.com/store/apps/details?id=com.sololearn.python&hl=en

References


IBM Big Data & Analytics Hub
ABOUT
CONTACT
CONTENT BY TYPE
SUBSCRIBE
EXPLORE BY TOPIC:
USE CASES
INDUSTRIES
ANALYTICS
TECHNOLOGY
EVENTS
FOR DEVELOPERS
BIG DATA & ANALYTICS HEROES
SITE MAP | PRIVACY | TERMS OF USE | 2014 IBM
FOLLOW
IBM Big Data & Analytics
@IBMbigdata
@IBManalytics
Explore By Topic
Use Cases
Industries
Analytics
Technology
For Developers
Big Data & Analytics Heroes
Explore By Content Type
Blogs
Videos
Analytics Video Chats
Big Data Bytes
Big Data Developers Streaming Meetups
Cyber Beat Live
Podcasts
White Papers & Reports
Infographics & Animations
Presentations
Galleries
Events
About The Big Data & Analytics Hub
Contact Us
RSS Feeds
Additional Big Data Resources
AnalyticsZone
Big Data University
Channel Big Data
developerWorks Big Data Community
IBM big data for the enterprise
IBM Data Magazine
Smarter Questions Blog
http://www.ibmbigdatahub.com/

References


BIG DATA
Learning Paths
Courses
Badges
Explore new learning opportunities
Login
Sign Up
Analytics, Big Data, and Data Science Courses
Your awesome career in Data Science and Data Engineering starts here.
Free Courses Sign Up
Support
https://bigdatauniversity.com/