

Big Data Overview

Agenda

- What is “Big Data” all about?
- The “Problem”: Relational Databases and Big Data
- The “Solution”: Scaling Techniques

Current Trends in Computing

- Massive Amounts of data being generated & collected
 - Terabytes (TB 10^{12}) PetaBytes (PB 10^{15}), ExaBytes (EB 10^{18})
- Explosion of Concurrent Users
 - Billions of people all over the world
- Lots of Different Kinds of data
 - Unstructured Text, Log Streams, Click Data, Mobile Devices
- Demands of the Marketplace
 - Competition is fierce
 - No tolerance for downtime, slow performance
 - Real-time data

Current Trends in Computing

- Major trend: **Volume** – Exabytes (EB 10^{18})
 – Generated & collected
- Explosion of **Concurrent Users**
 - Billions of people all over the world
- Lots of **Different Kinds** of data
 - Unstructured Text, Log Streams, Click Data, Mobile Devices
- **Demands of the Marketplace**
 - Competition is fierce
 - No tolerance for downtime, slow performance

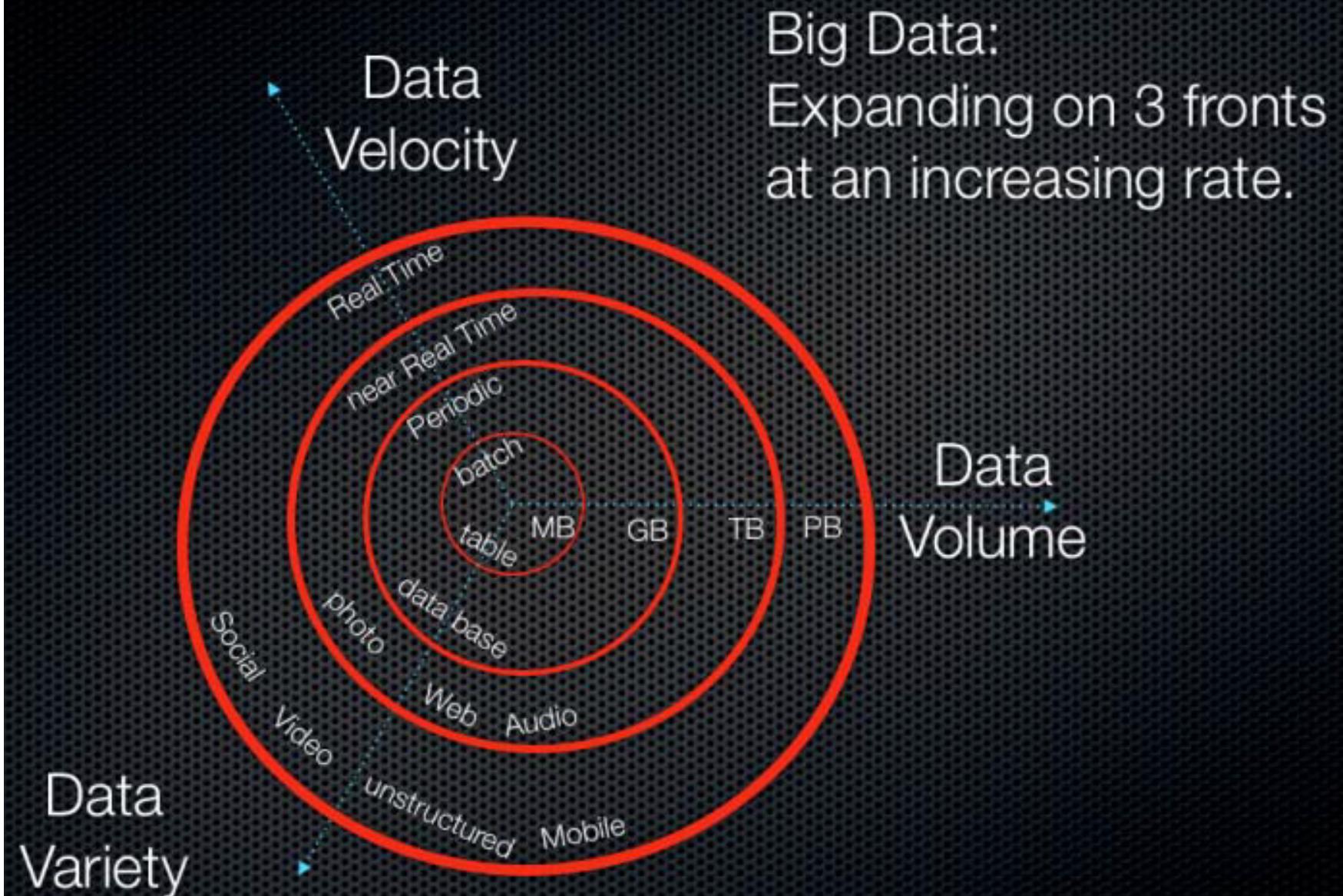
Current Trends in Computing

- Major trend: **Volume** – Exabytes (EB 10^{18})
 – Billions of concurrent users
 – Lots of **Variety**
 – Unstructured Text, Log Streams, Click Data, Mobile Devices
- Demands of the Marketplace
 - Competition is fierce
 - No tolerance for downtime, slow performance

Current Trends in Computing

- Mass of **Volume** data generated & collected
 - Petabytes (PB) to Exabytes (EB 10^{18})
- Explosion of **Variety** concurrent users
 - Billion+ types of data
- Lots of **Velocity**
 - Unstructured Text, Log Streams, Click Data, Mobile Devices
- Demands of **Velocity**
 - Competition
 - No tolerance for downtime, slow performance

Big Data Overview



Big Data Overview

A Few Amazing Facts ⁽¹⁾

- Annual global IP traffic passed the zettabyte threshold by the end of 2016, and will reach 2.3 ZB per year in 2020.
- Smartphone traffic exceeded PC traffic in 2016.
- Traffic from wireless and mobile devices will account for two-thirds of total IP traffic in 2020.
- The number of devices connected to IP networks will surpass more than three times the global population in 2020.

Global Data Created Each Day

44 billion GB of data was created per day in 2016.



+ IDC predicts growth to **463 billion GB** of data created per day in 2025

Big Data Overview

What is driving this data explosion?

- Smart Phones
- Apps
- The Internet of Things
- Digital Commerce
- Online Entertainment
- Cloud Computing
- Social Media

“Social media facilitate the most intimate of discussions among users, who seem to forget that they’re communicating on a public forum.”

Data Crush, Christopher Surdak

Smart Phones

- Smartphones -- Only here for a little over a decade: the first iPhone came out in January 2007
- In 2010, 4.5 B people owned a mobile phone
- By 2017, there were 6.8 B mobile phone users (90% of humanity)
 - Of which 2.5 B are “smartphones”
- The mobile phone has become the individual’s dominant point of interaction with society

Data Overview



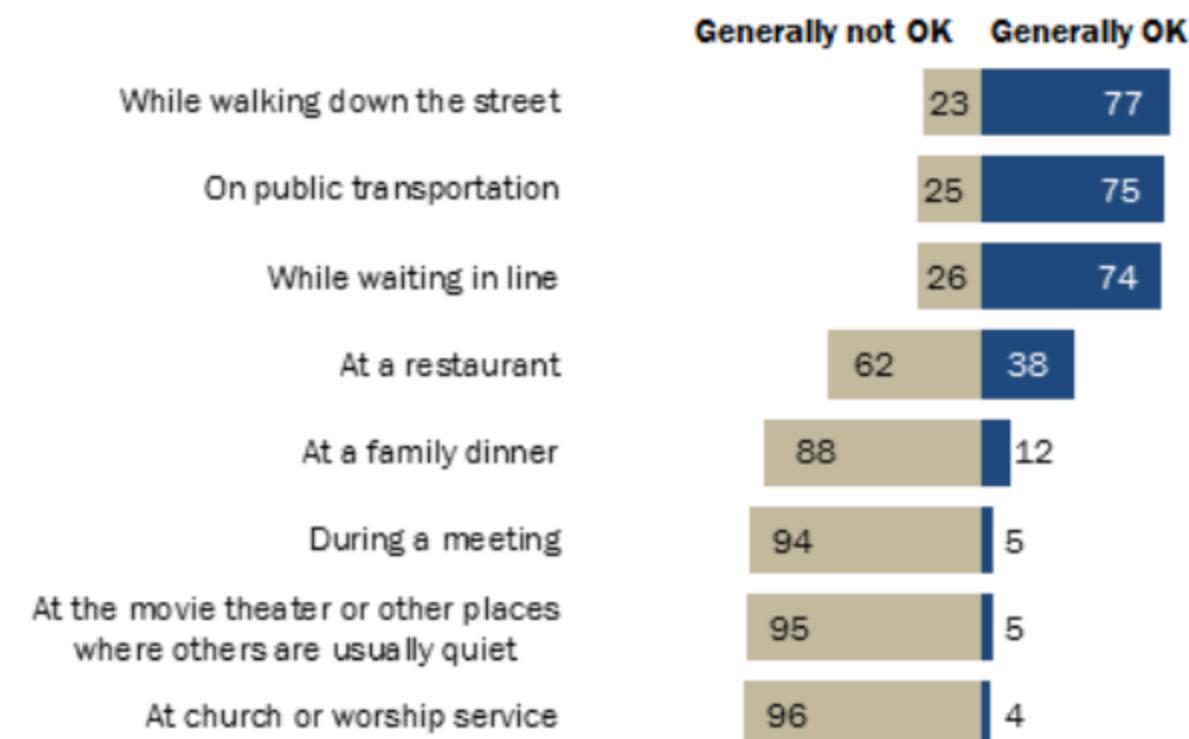
"On the Internet, nobody knows you're a dog."

©The New Yorker Collection 1993 Peter Steiner
From cartoonbank.com. All rights reserved.

Big Data Overview

People Have Varying Views About When It Is OK Or Not OK To Use Their Cellphones

% of adults who believe it is OK or not to use a cellphone in these situations



Source: Pew Research Center American Trends Panel survey, May 30-June 30, 2014.
N=3,217 adults.

Apps

- The smartphone “app” is 11 years old
 - Apple started selling apps in 2008, with 500 apps
- In 2020, 20 million registered iOS developers
- In 2020, 500 million weekly visitors to the App Store
- 2012 – Apple reached 25 B annual downloads of apps
- 2013 – Apple reached 40 B annual downloads of apps
- Android apps – 50 B downloads in 2015
- Android apps – 90 B downloads in 2016
- By December 2017 – combined downloads = 197 B
- Google Play provides over a million downloadable apps & games

IOT – The internet of things

- Internet connected devices that are collecting and transmitting data
 - TVs with wifi connectivity
 - Cars with OnStar
 - Appliances (refrigerators, thermostat, doorbell, etc.)
 - Medical devices
 - “Wearables”
 - FitBit
 - Apple Watch
 - Robots
 - Assistants (Alexa, Google Home, Siri)

eCommerce

1. Global Retail Ecommerce Sales Will Reach \$4.5 Trillion by 2021



Big Data Overview

eCommerce

Data from [Statista](#) forecasts a 246.15% increase in worldwide ecommerce sales, from \$1.3 trillion in 2014 to \$4.5 trillion in 2021.

Nearly threefold lift in online revenue.

Online Entertainment

- Online video has a 90 percent penetration rate among internet users on any device.
- In the United States, current data indicates more than 180 million digital video viewers with a penetration rate of approximately 75 percent of the online population.
- The majority of mobile video revenues are generated via subscriptions, followed by advertising.
- Current US mobile video subscription revenues have already surpassed 1 billion US dollars annually.

Cloud Computing

- Cloud computing is projected to increase from \$67B in 2015 to \$162B in 2020 attaining a compound annual growth rate (CAGR) of 19%.
- Gartner reports the worldwide public cloud services market grew 18% in 2017 to \$246.8B, up from \$209.2B in 2016.
- 74% of Tech Chief Financial Officers (CFOs) say cloud computing will have the most measurable impact on their business

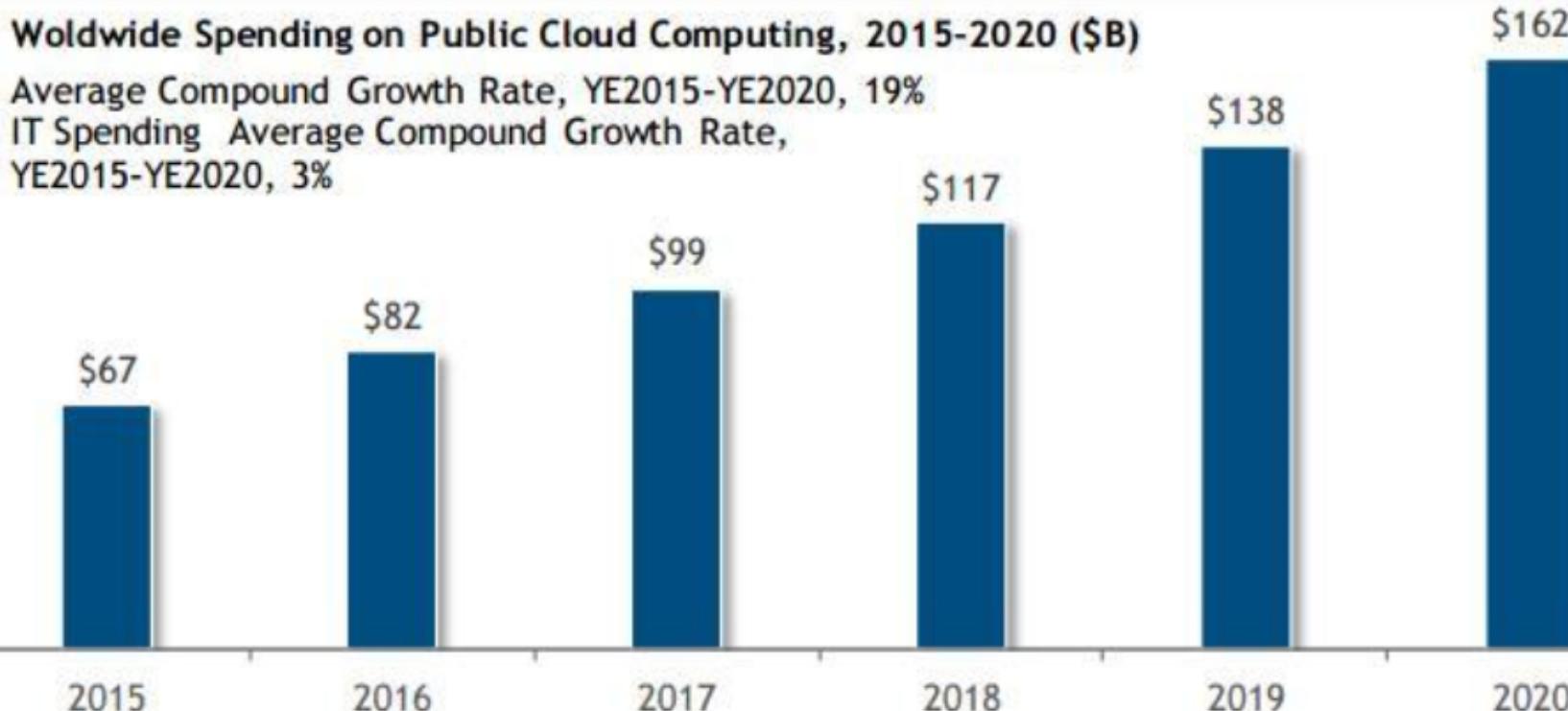
Cloud Computing

Cloud computing spending is growing at 4.5 times the rate of IT spending since 2009 and is expected to grow at better than 6 times the rate of IT spending from 2015 through 2020.

According to IDC, worldwide spending on public cloud computing will increase from \$67B in 2015 to \$162B in 2020 attaining a 19% CAGR.

Big Data Overview

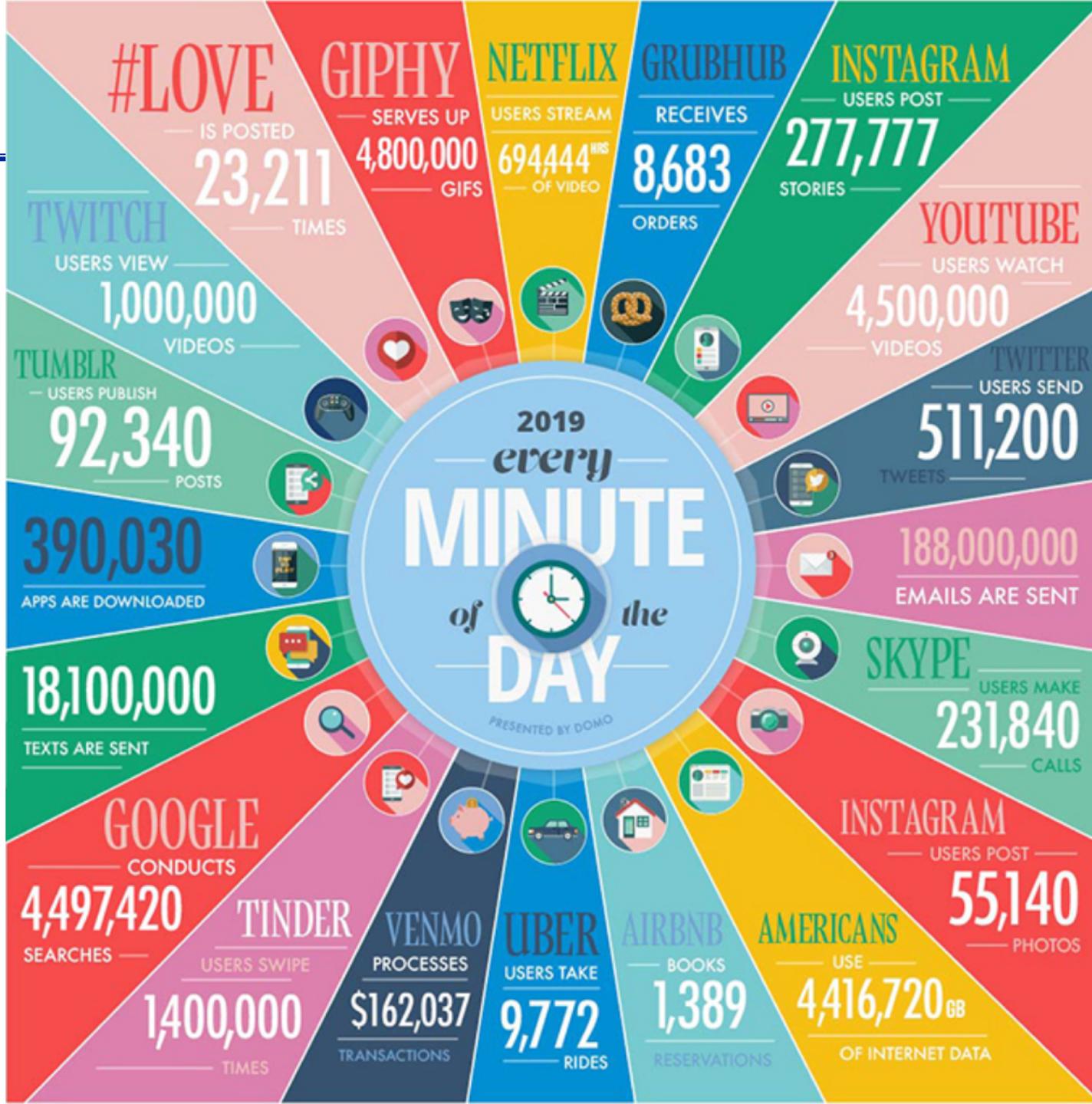
The Rapid Growth of Cloud Computing, 2015-2020



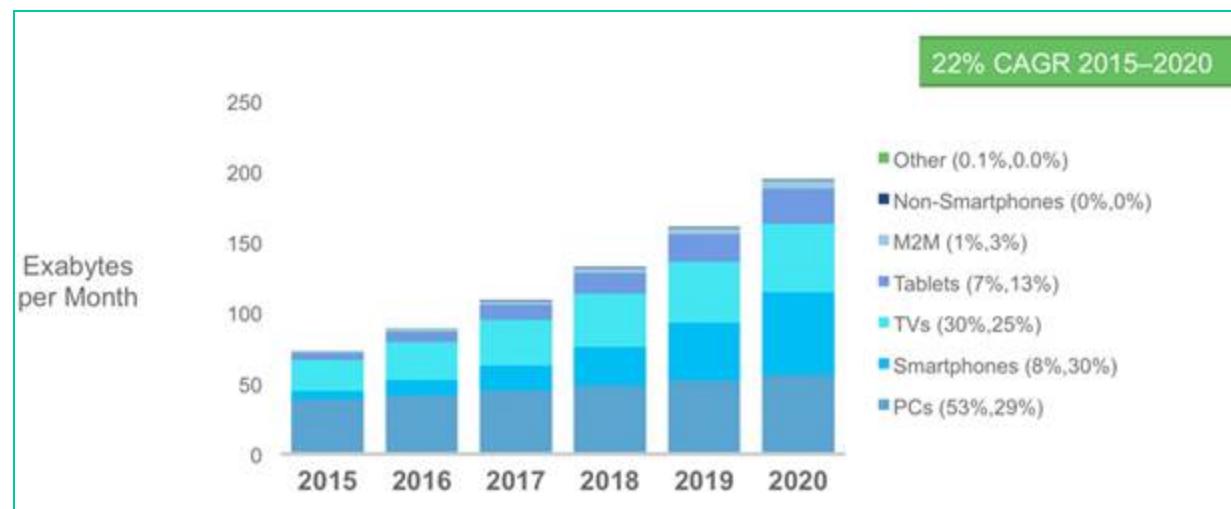
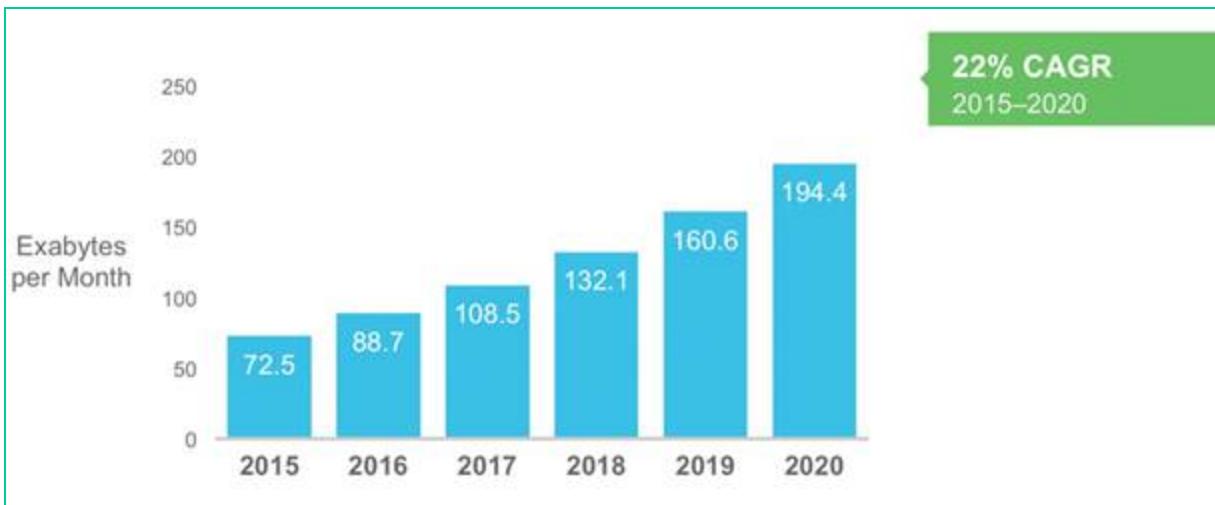
Source: IDC, 2016

Big Data Overview

- 90% of the data on the internet has been created since 2016.
- In 2014 there were 2.4 Billion internet users
 - By 2016, this grew to 3.4 Billion
 - By June 2019, this grew to 4.4 Billion
- More facts:
 - 1,209,600 new data producing social media users each day.
 - 656 million tweets per day!
 - More than 4 million hours of content uploaded to Youtube every day, with users watching 5.97 billion hours of Youtube videos each day.
 - 67,305,600 Instagram posts uploaded each day
 - There are over 2 billion monthly active facebook users, compared to 1.44 billion at the start of 2015 and 1.65 at the start of 2016.
 - Facebook has 1.32 billion daily active users on average as of June 2017
 - 4.3 BILLION Facebook messages posted daily!
 - 5.75 BILLION Facebook likes every day.
 - 22 billion texts sent every day.
 - 5.2 BILLION daily Google Searches in 2017.



Big Data Overview



Challenges

- How can we collect, process and store so much data?
- How can we wade through so much data in a timely manner?
- How can we analyze so much data and gain meaningful insights?
- Can my existing systems/architectures handle this?
 - Why not?
- Can my existing staff (skills & tools) make the transition?