

# LEC.1.



# Introduction to Data Science

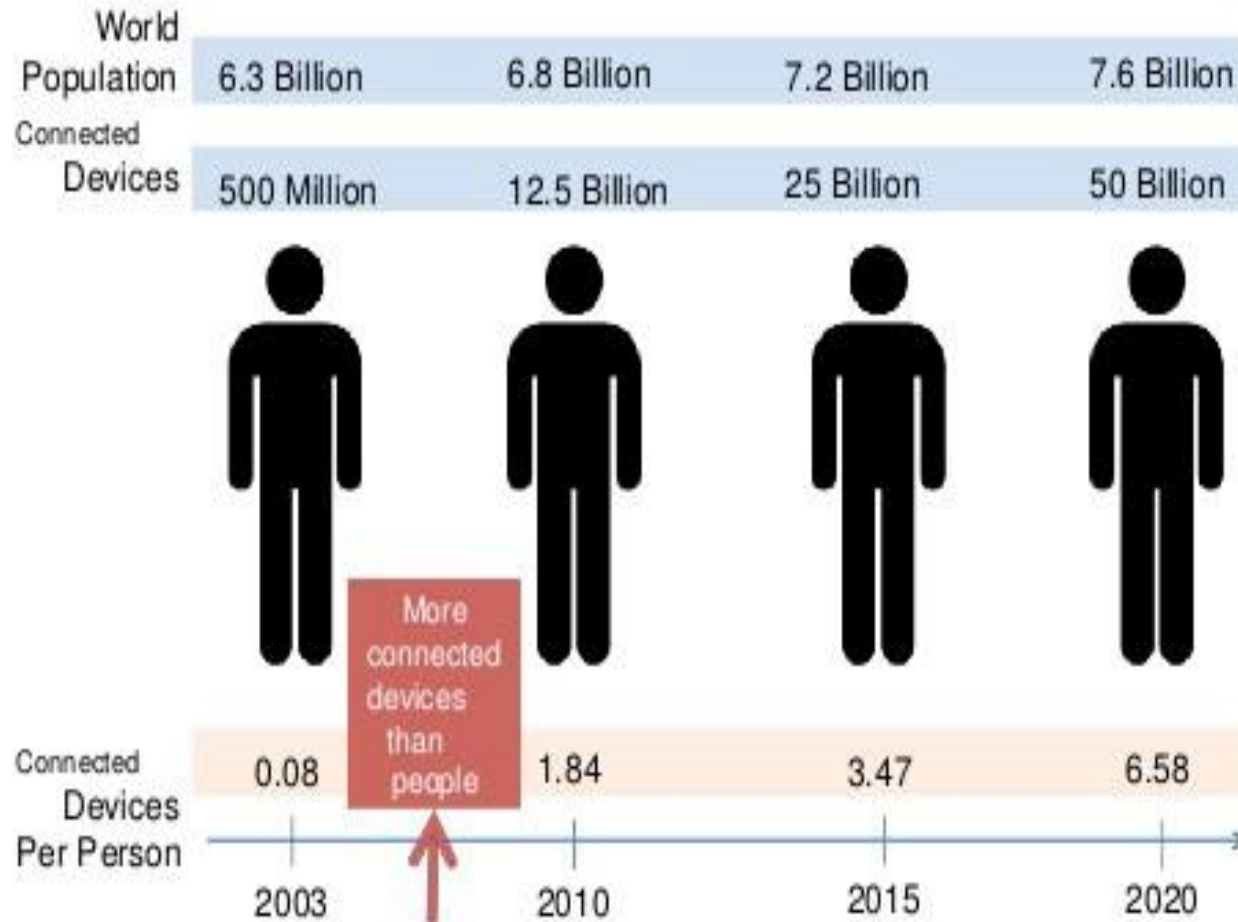
DATA IS THE NEW OIL

# Data All Around

- ▶ Lots of data is being collected and warehoused
  - Web data, e-commerce
  - Financial transactions, bank/credit transactions
  - Online trading and purchasing
  - Social Network



# More Connected Devices Than People

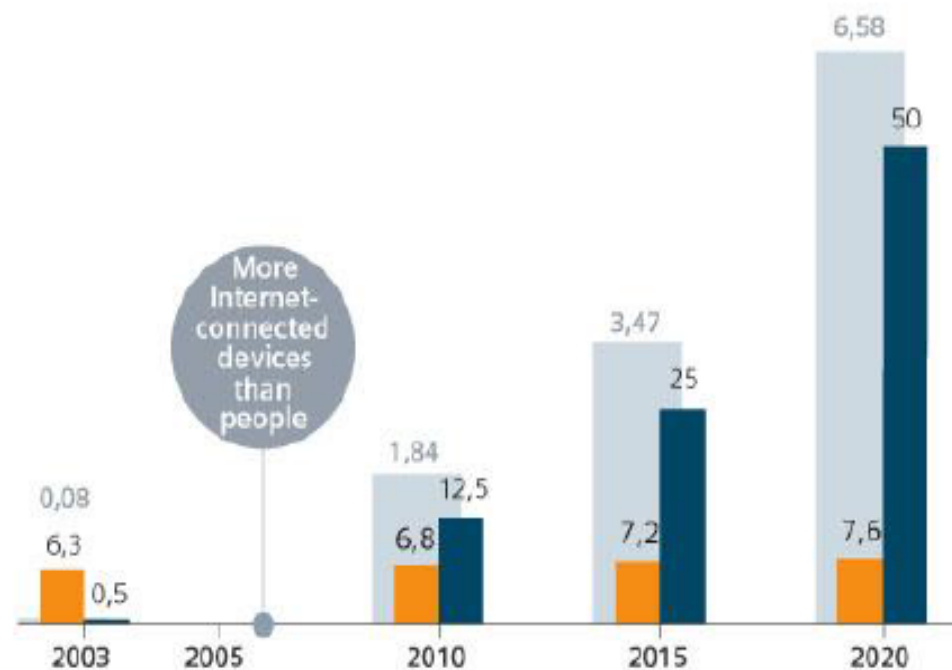


[Source: Cisco IBSG, April 2019]

# Devices vs. People

## Growth in Internet-Connected Devices by 2020

- World population (in billions)
- Internet-connected devices in (billions)
- Internet-connected devices per person



Source: Cisco IBSG, April 2011

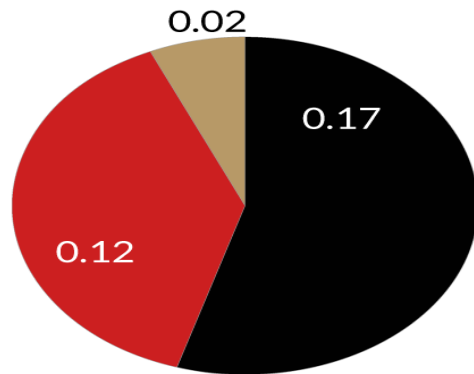
## Aggregate Time on Site in Billion Hours

Desktop

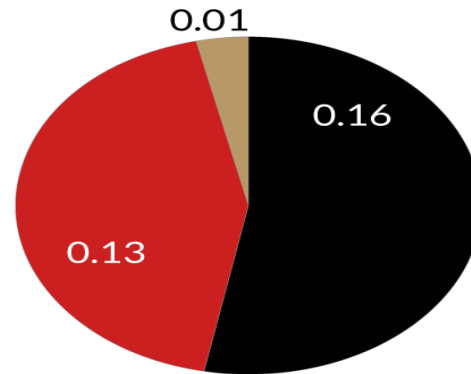
Mobile

Tablet

U.S. Only 2019



U.S. Only 2020



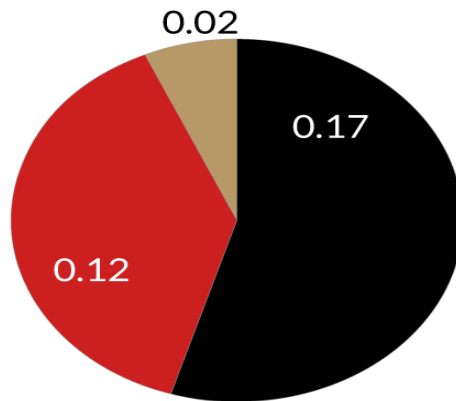
# Aggregate Time on Site in Billion Hours

Desktop

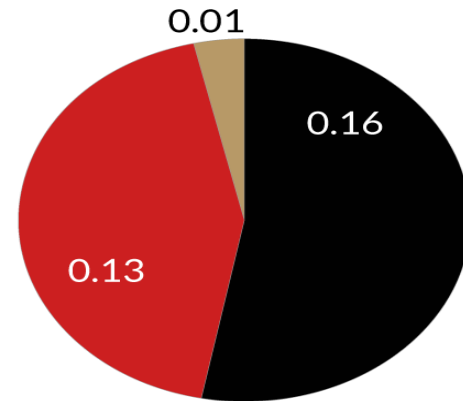
Mobile

Tablet

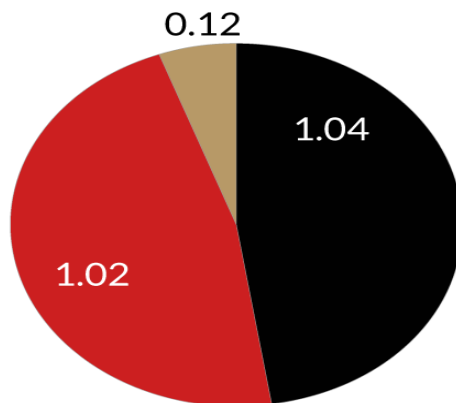
U.S. Only 2019



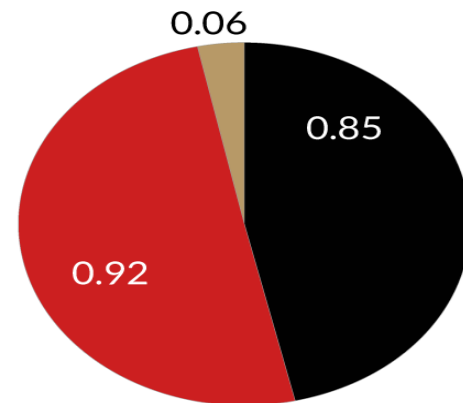
U.S. Only 2020



Global 2019

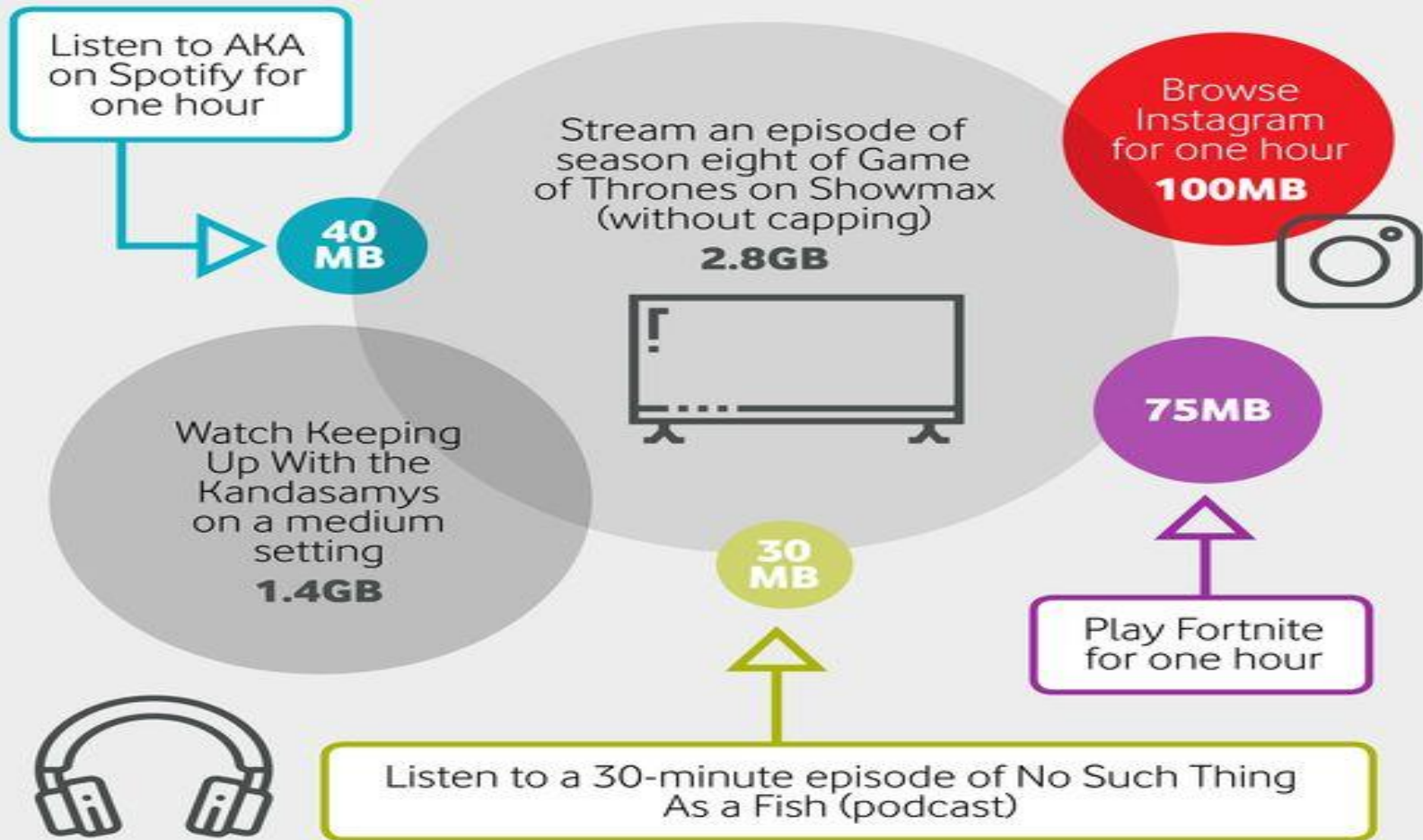


Global 2020





# HOW MUCH DATA DO YOU NEED TO...



**NOTE:** These figures are approximate, so monitor your habits before committing to a data plan. For example, gaming can use anywhere from 50MB to 100MB per hour.



NUMBER  
OF EMAILS  
SENT  
EVERY SECOND

2.9  
MILLION

DATA  
CONSUMED BY  
HOUSEHOLDS  
EACH DAY

375  
MEGABYTES

VIDEO  
UPLOADED TO  
YOUTUBE EVERY  
MINUTE

20  
HOURS

DATA PER  
DAY  
PROCESSED  
BY GOOGLE

24  
PETABYTES

# THE WORLD OF DATA



SOURCES: Cisco, comScore, NetScout Systems, Statista, YouTube

# THE WORLD OF DATA

TWEETS  
PER  
DAY

50  
MILLION

TOTAL MINUTES  
SPENT ON  
FACEBOOK  
EACH MONTH

700  
BILLION

DATA SENT  
AND RECEIVED  
BY MOBILE  
INTERNET USERS

1.3  
EXABYTES

PRODUCTS  
ORDERED ON  
AMAZON PER  
SECOND

72.9  
ITEMS



IN THE 21ST CENTURY, we live a large part of our lives online. Almost everything we do is reduced to bits and sent through cables around the world at light speed. But just how much data are we generating? This is a look at just some of the massive amounts of information that human

# Cognitive Computing



- People expect systems to behave like humans
  - ▣ Be Adaptive
    - Learn as information and goals change
  - ▣ Be Interactive
    - Interact easily with people and other systems
  - ▣ Be Contextual
    - Understand meaning, exploit additional sources of information

# Cognitive Computing

- People expect systems to behave like humans
  - ▣ Be Adaptive
    - Learn as information and goals change
  - ▣ Be Interactive
    - Interact easily with people and other systems
  - ▣ Be Contextual
    - Understand meaning, exploit additional sources of information
- Need to process large quantities of uncertain data of different types (text, speech, sensors, images etc.)

# Cognitive Computing in 5 Years



The image shows a screenshot of the Mashable website. The header is blue with the Mashable logo and navigation links: MUST READS, SOCIAL MEDIA, TECH, BUSINESS, ENTERTAINMENT, US & WORLD, and WATERCOOLER. Below the header, the 'Tech' section is highlighted. A row of article links is displayed: AdChoices, Smell Taste, IBM Computers, Taste Buds, and Smell Machine. The main article title, 'IBM: Computers Will See, Hear, Taste, Smell and Touch in 5 Years', is prominently displayed below the links.

**Mashable** MUST READS SOCIAL MEDIA ▼ TECH ▼ BUSINESS ▼ ENTERTAINMENT ▼ US & WORLD ▼ WATERCOOLER ▼

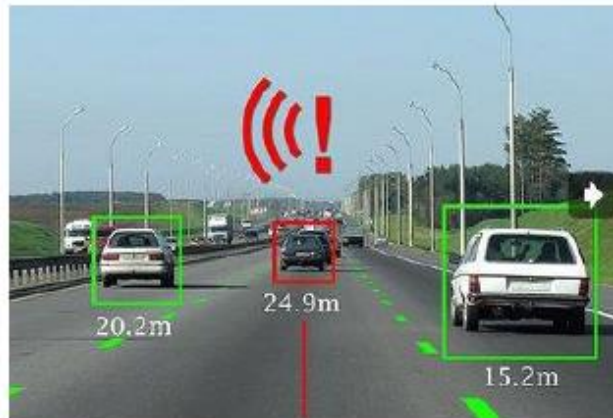
**Tech**

[AdChoices](#) ▶ [▶ Smell Taste](#) [▶ IBM Computers](#) [▶ Taste Buds](#) [▶ Smell Machine](#)

## IBM: Computers Will See, Hear, Taste, Smell and Touch in 5 Years



# Cognitive Computing



# Cognitive and Data Science



- People want their systems/devices to behave smarter
  - ▣ Personal devices
  - ▣ Industrial systems
- More data to acquire and analyze using more complex algorithms and technologies



# 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

# Course contents:

- ▶ Introduction
- ▶ Data analytic life cycle
- ▶ Data Visualization
- ▶ Clustering
- ▶ Classification
- ▶ Sql
- ▶ R