Lecture -7

PYTHON INSTITUTE
Open Education & Development Group

Programing in Python

Instructor : AALWAHAB DHULFIQAR

Advisor : Dr. Tejfel Mate

What you will learn:

differentiate between open data and private data
introduce big data
differentiate between structured data and unstructured data
to describe two basic data management techniques

Chapter 1: Data and the Internet of Things

Big Data & Analytics

Cisco | Networking Academy®
Mind Wide Open™

Chapter 1 - Sections & Objectives

- 1.1 Value of Data
  - Demonstrate the value of data.
- 1.2 Data and Big Data
  - Explain the concept of big data.
- 1.3 Managing Big Data
  - Demonstrate knowledge of data management approaches in the IoE.

1.1 The Value of Data

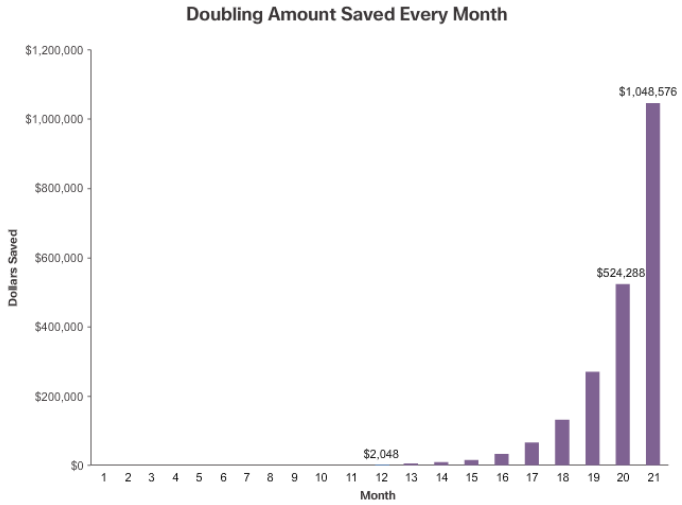
Cisco | Networking Academy®
Mind Wide Open™

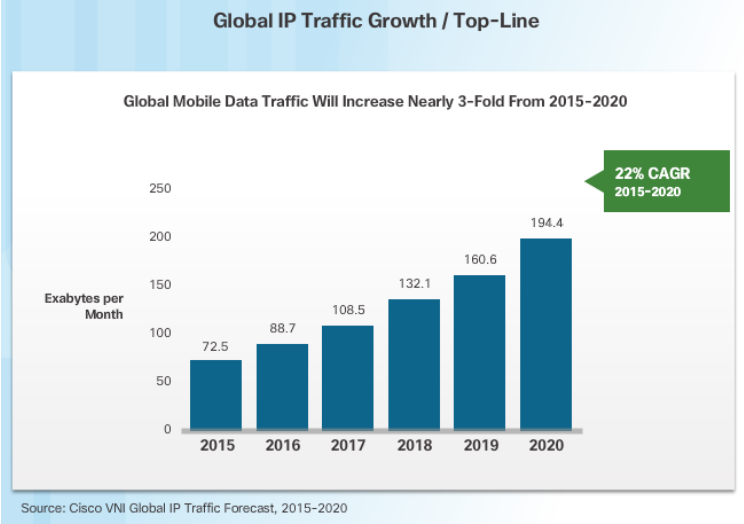
The Data Aspect of a Connected World

- The Value of Data
  - The amount of data to be stored and analyzed is expanding.
  - The variety of data will reach new areas.
  - The digital transformation will impact three elements of our lives: business, social, and environmental.
- What is Data?
  - Data can be many things.
    - Words in a book, article, or blog
    - Contents of a spreadsheet or database
    - Pictures or video
    - A stream of measurements from a device
  - Useful data is information.
  - Determine the amount of data to be collected.
  - Not all data can be used as-is.
  - Data analysis provides useful information and/or trends.

Data is Growing Exponentially

- Estimating Exponential Growth
  - Two types: linear and exponential
  - Exponential growth is more dramatic.
- Growth of Data
  - Today's data is growing exponentially.
  - Sample data growth forecast for 2015 to 2020 from Cisco's Visual Networking Index (VNI)
    - Consumer mobile data traffic will reach 26.1 exabytes per moth in 2020.
    - IP traffic will reach 194.4 exabytes per month in 2020.
    - 64% of all global Internet traffic will cross content delivery networks in 2020.





Data Growth Changes Our Lives

- Data Growth Impact
  - Fueled by the proliferation of IoT devices
  - Including sensors, wireless end devices, and mobile networks
- Business Example: Kaggle
  - Kaggle is a platform that connects businesses and other organizations that have questions about data to the people who know how to find the answers.
  - Kaggle runs online competitions.
- Social Example: DrivenData
  - Brings cutting-edge practices in data science and crowdsourcing to people and organizations that are addressing these challenges
- Environmental Example: Climate Change
  - NASA and Cisco partnership – Planetary Skin
  - Online collaborative global monitoring platform
  - Captures, collects, analyzes and reports data on environmental conditions

1.2 Data and Big Data

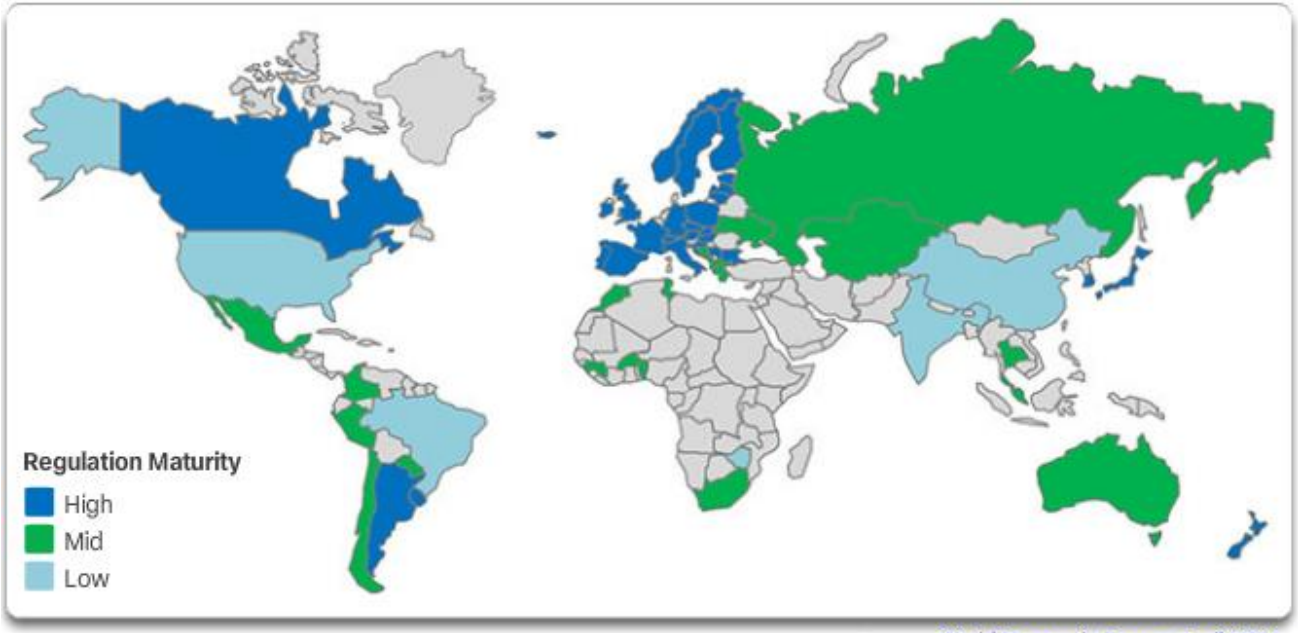
Cisco | Networking Academy®
Mind Wide Open™

Where Does Big Data Come From

- Defining Big Data
  - Data that is so vast, fast, or complex that it becomes impossible to store, process, and analyze using traditional data storage and analytics applications
- Big Data Characteristics
  - 4 big Vs of Big Data: volume, velocity, variety, and veracity
  - Volume – amount of data
  - Velocity – rate data is generated
  - Variety – type of data
  - Veracity – preventing inaccurate data from spoiling a data set
- How much Data is Big Data
  - IBM's Paul Zikopaulos stated it takes 200 to 600 Terabytes to qualify as Big Data

Open Data and Private Data

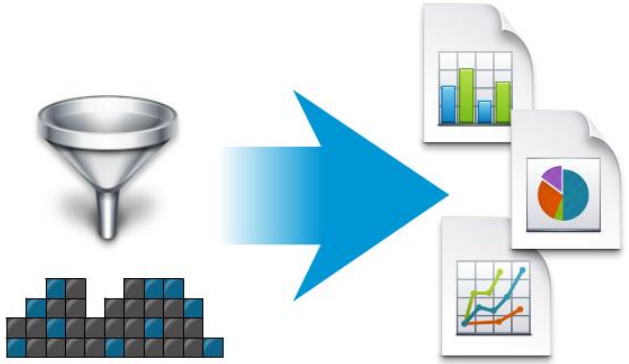
- Open Data
  - The Open Knowledge Foundation describes Open Data as “any content, information or data that people are free to use, reuse, and redistribute without any legal, technological, or social restriction.”
- Private Data
  - Data related to an expectation of privacy and regulated by a particular country/government



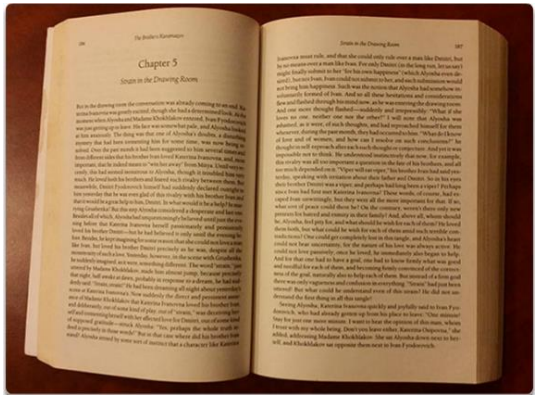
World Economic Forum, April 2014

Structured and Unstructured Data

- Structured Data
  - Data entered and maintained in fixed fields within a file or record
  - Easily entered, classified, queried, and analyzed
  - Relational databases or spreadsheets

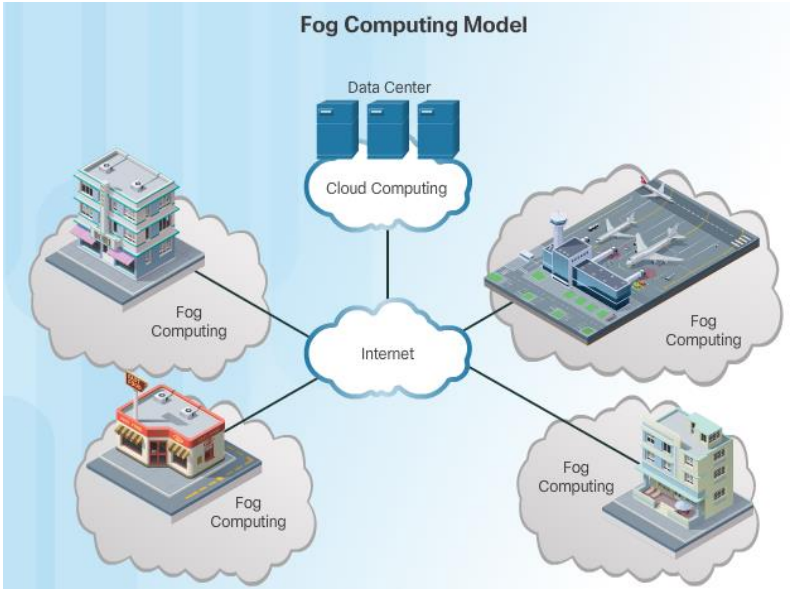


- Unstructured Data
  - Lacks organization
  - Raw data
  - Photo contents, audio, video, web pages, blogs, books, journals, white papers, PowerPoint presentations, articles, email, wikis, word processing documents, and text in general



Data at Rest and Data in Motion

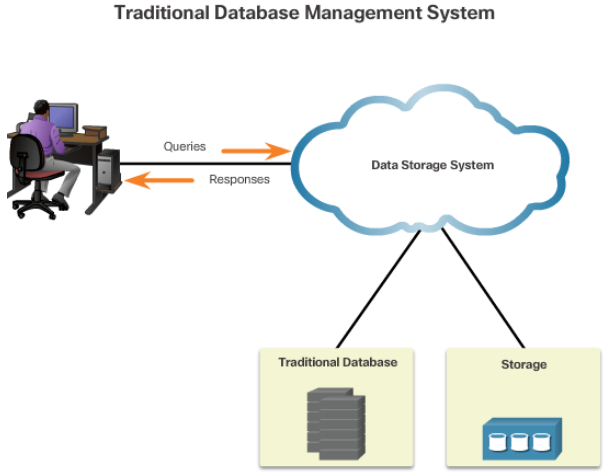
- Data at Rest
  - Data stored in a physical location such as a server hard drive or within a data center
  - Follows the traditional data analysis flow of **Store > Analyze > Notify > Act**
- Data in Motion
  - Dynamic data that requires real-time processing before the data becomes irrelevant or obsolete
  - Analysis and action happen sooner rather than later
  - Data analysis flow is **Analyze > Act > Notify > Store**



1.3 Evolution to Big Data

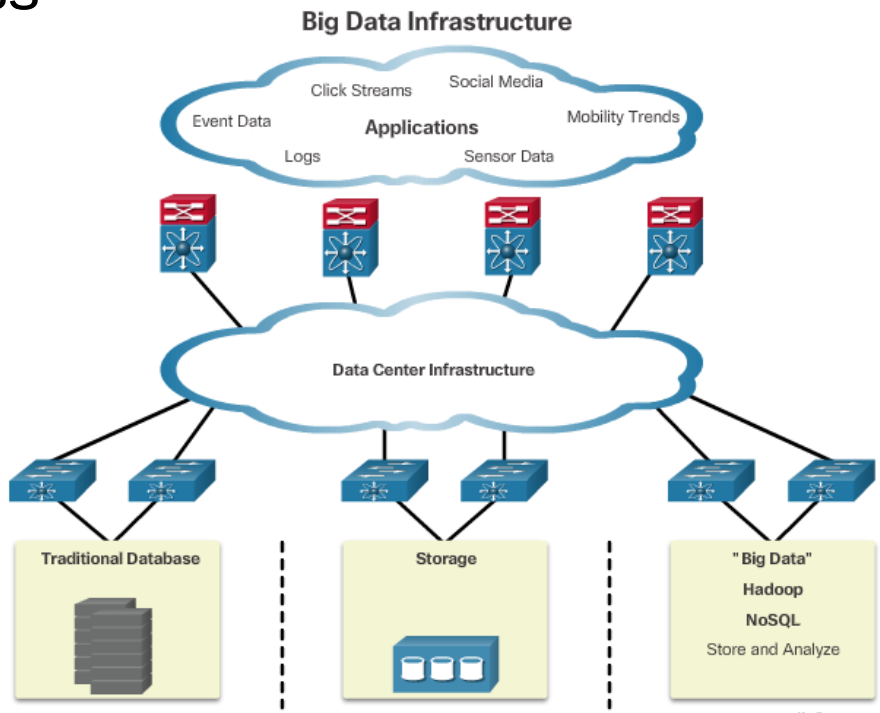
Cisco | Networking Academy®
Mind Wide Open™

Evolution to Big Data



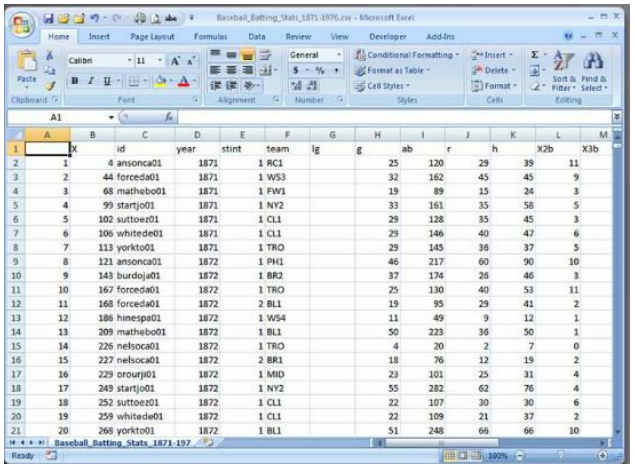
Traditional to Big Data Infrastructure

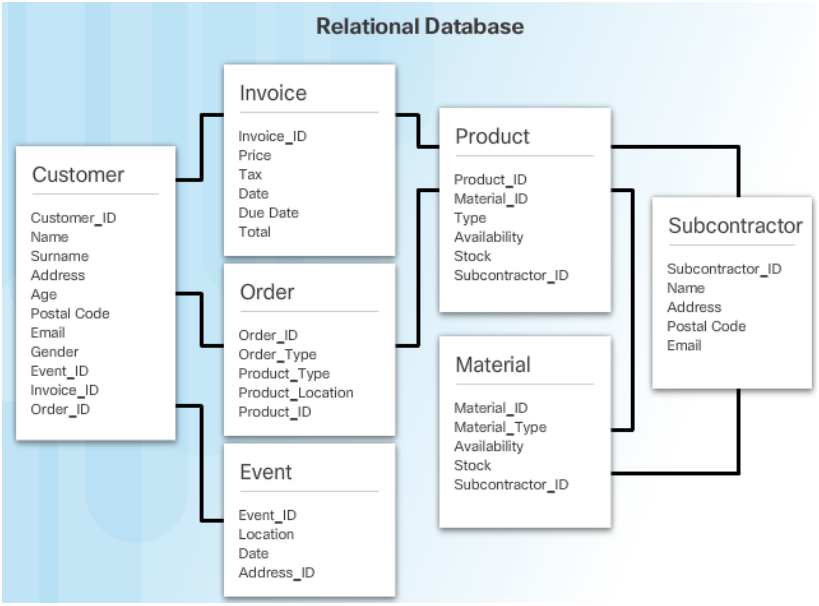
- Database servers and traditional data processing tools
- Distributed data systems across horizontally coupled, independent resources to achieve the scalability needed for the efficient processing of extensive data sets
- Onsite and cloud computing solutions



Basic Data Management Technologies

- Flat file database – stores records in a single file with no hierarchical structure such as a spreadsheet
- Relational database – capture relationships between different sets of data, creating more useful information





Basic Data Management Technologies

- Relational Database Management System is the dominant database technology with no challenge for over 30 years.
- Big Data analytics becomes increasingly difficult to manage with a relational database management system (RDBMS)
- Hadoop Distributed File System (HDFS) is a distributed, fault tolerant file system created to deal with big data volumes.
- NoSQL database structure created to make database design simpler with faster. Meets the demands of Web applications.
- SQLite – simple and easy to use SQL database engine that is the most widely deployed database in the world.

1.4 Summary

Cisco | Networking Academy®
Mind Wide Open™

Summary

- Data can be words in a book, contents of a spreadsheet, photos, files, or streams of measurements sent by a device.
- Data growth can be linear and exponential. Exponential is a more dramatic increase.
- Four Vs of Big Data are volume, velocity, variety, and veracity.
- Structured data is data entered in fixed fields within a database file or record. Unstructured data does not have a fixed schema that identifies the type of data.
- Data at rest is static data stored in a physical location.
- Data in motion analyzes and extracts value from the data before it is stored.
- A flat file database is like a spreadsheet storing records in a single file with no hierarchical structure.
- A relational database captures the relationships between different data sets and can provide more useful information.

Summary

- Hadoop was created to deal with big data volumes.
- A NoSQL database stores and accesses data differently than relational database.
- SQLite is a simple and easy to use SQL database engine that is the most widely deployed database in the world.