Information Technology Institute

Operating Systems

Lectures 6

Inter-process Communication

A process can be of two types:

- Independent process.
- Co-operating process.

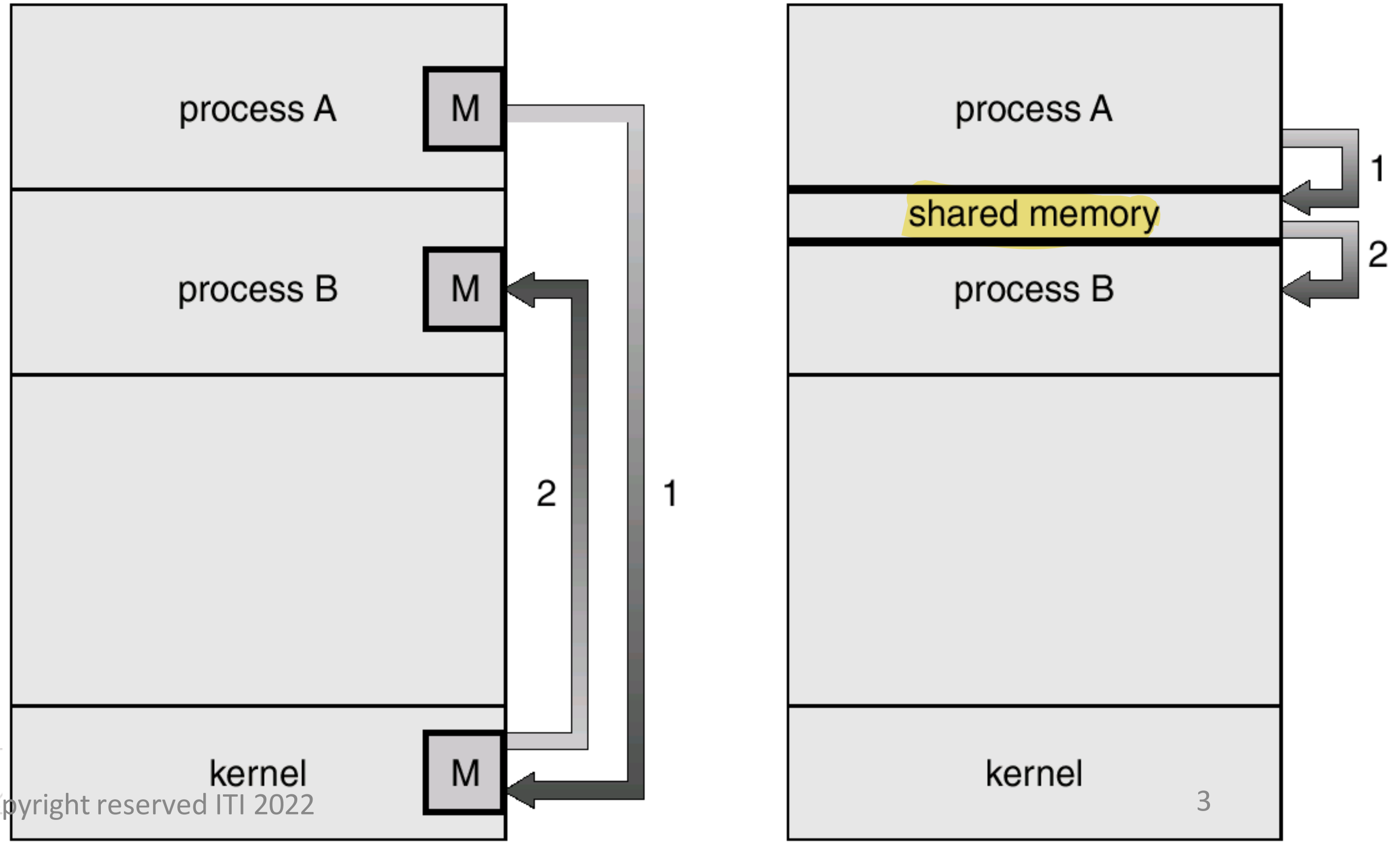
An independent process is not affected by the execution of other processes while a co-operating process can be affected by other executing processes.

Inter-process communication (IPC) is a mechanism that allows processes to communicate with each other and synchronize their actions.

Inter-process Communication

- The communication between these processes can be seen as a method of co-operation between them. Processes can communicate with each other through both:

- Shared Memory
- Message passing



Shared Memory Method

- When two or more processes need to communicate with each other, they may create a shared memory area that is accessible by both processes.
- Then, one of the processes may act as the *producer of data*, while the other could act as the *consumer of data*.
- The ***memory acts as the communication buffer*** between these two processes.
- This is a very common mechanism to communicate between processes.
- Note: this method need a way of management when the two processes need to save in the shared memory at the same time, it is called ***Synchronization***

Shared Memory Method

- Process1 generates information about certain computations or resources being used and keeps it as a record in shared memory.

- When process2 needs to use the shared information, it will check in the record stored in shared memory and take note of the information generated by process1 and act accordingly.

- Processes can use shared memory for extracting information as a record from another process as well as for delivering any specific information to other processes.

Message Passing Method

- In this method, processes communicate with each other without using any kind of shared memory. If two processes p1 and p2 want to communicate with each other, they proceed as follows:

- **For examples** – chat programs, ***TCP/IP communication,*** print server

A print server is a software application, network device or computer that manages print requests and makes printer queue status information 6

- Establish a communication link (if a link already exists, no need to establish it again.)
- Start exchanging messages using basic primitives.
  We need at least two primitives:
  – **send**(message, destination) or **send**(message)
  – **receive**(message, host) or **receive**(message)

Message Passing Method

Message Passing through Communication Link.
Direct and Indirect Communication link

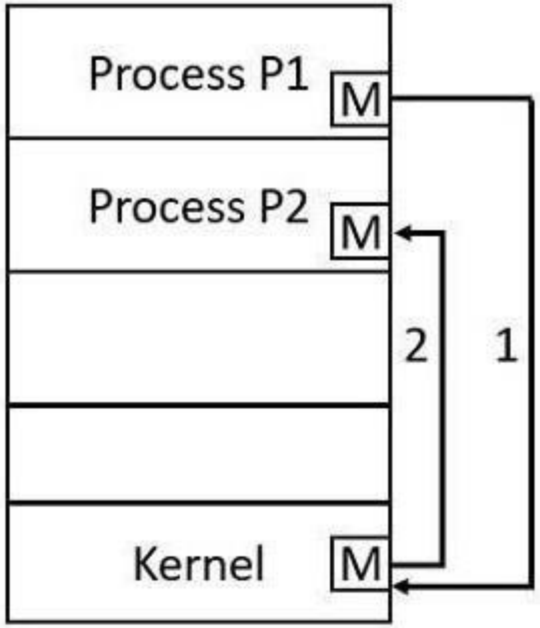
In Direct message passing, The process which wants to communicate must explicitly name the recipient or sender of the communication.
e.g. **send(p1, message)** means send the message to p1.
Similarly, **receive(p2, message)** means to receive the message from p2.

In indirect Communication is done via a shared mailbox (port), which consists of a queue of messages. The sender keeps the message in mailbox and the receiver picks them up.

The standard primitives used are: **send(A, message)** which means send the message to mailbox A. The primitive for the receiving the message also works in the same way e.g. **received (A, message)**.

Message Passing Method (chat program)

- **Step 1** – Message passing provides two operations which are as follows –
- Send message
- Receive message
- Messages sent by a process can be either fixed or variable size.
- **Step 2** – For fixed size messages the system level implementation is straight forward. It makes the task of programming more difficult.
- **Step 3** – The variable sized messages require a more system level implementation but the programming task becomes simpler.
- **Step 4** – If process P1 and P2 want to communicate they need to send a message to and receive a message from each other that means here a communication link exists between them.
- **Step 5** – Methods for logically implementing a link and the send() and receive() operations.



Message Passing System

File Systems

Content

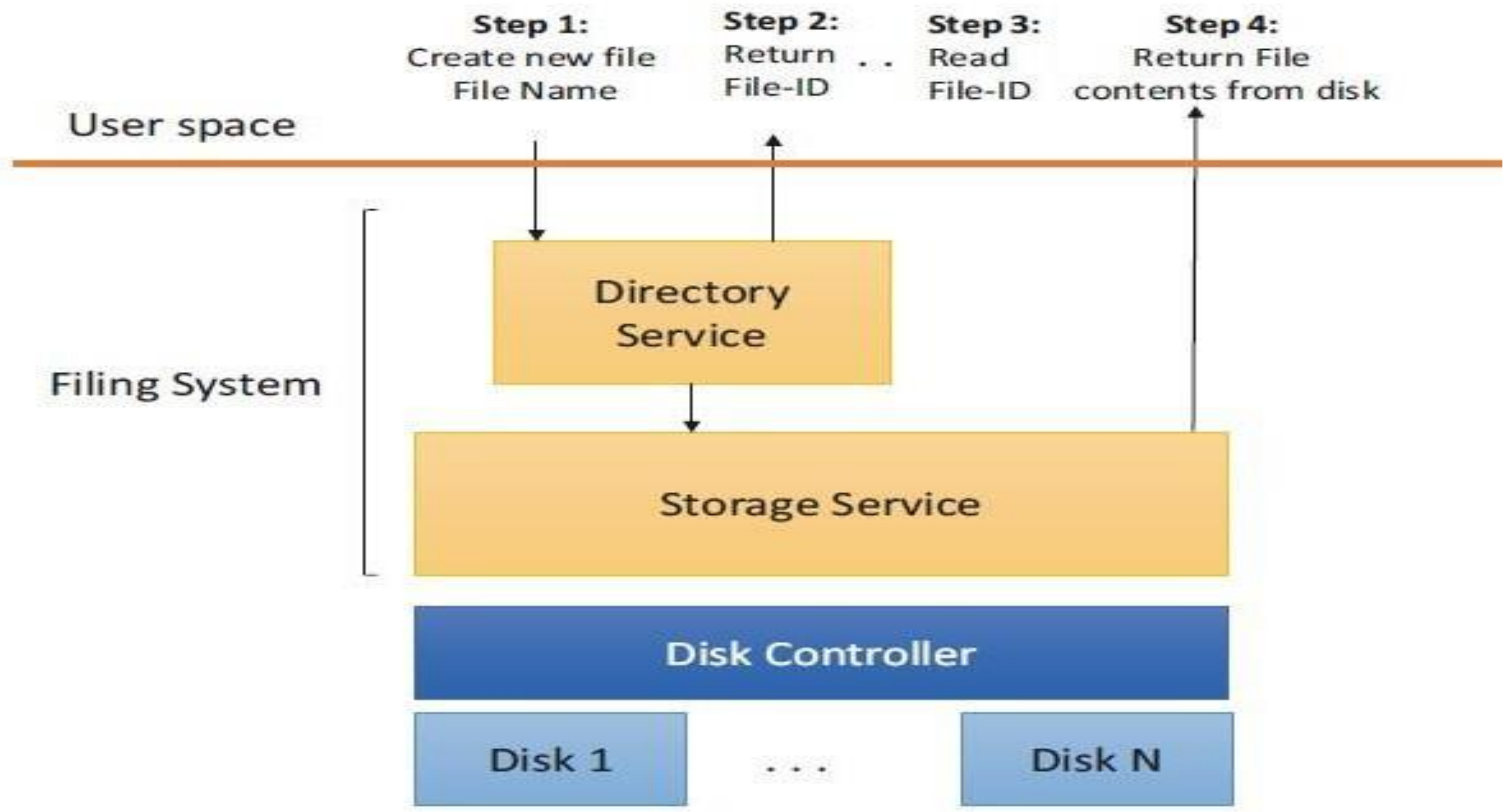
- Need for File Systems
- File Concept
- Directory Name Space
- Access Control
- Concurrency

Course Outlines

Need for File Systems

- Applications often need to read and write files to achieve their goals. We leverage the OS to create, read, and write such files on the system. We depend on the OS to maintain and manage files on the system.
- OS file systems have two main components to facilitate file management:
  - ***Directory Service***: There is a need to uniquely manage files in a structured manner, manage access, and provide **Read-Write-Edit controls** on the file system. This is taken care by a layer called as the directory service.
  - ***Storage Service***: There is a need to **communicate to** the underlying **hardware such as the disk**. This is managed by a storage service that abstracts different types of storage devices on the system.

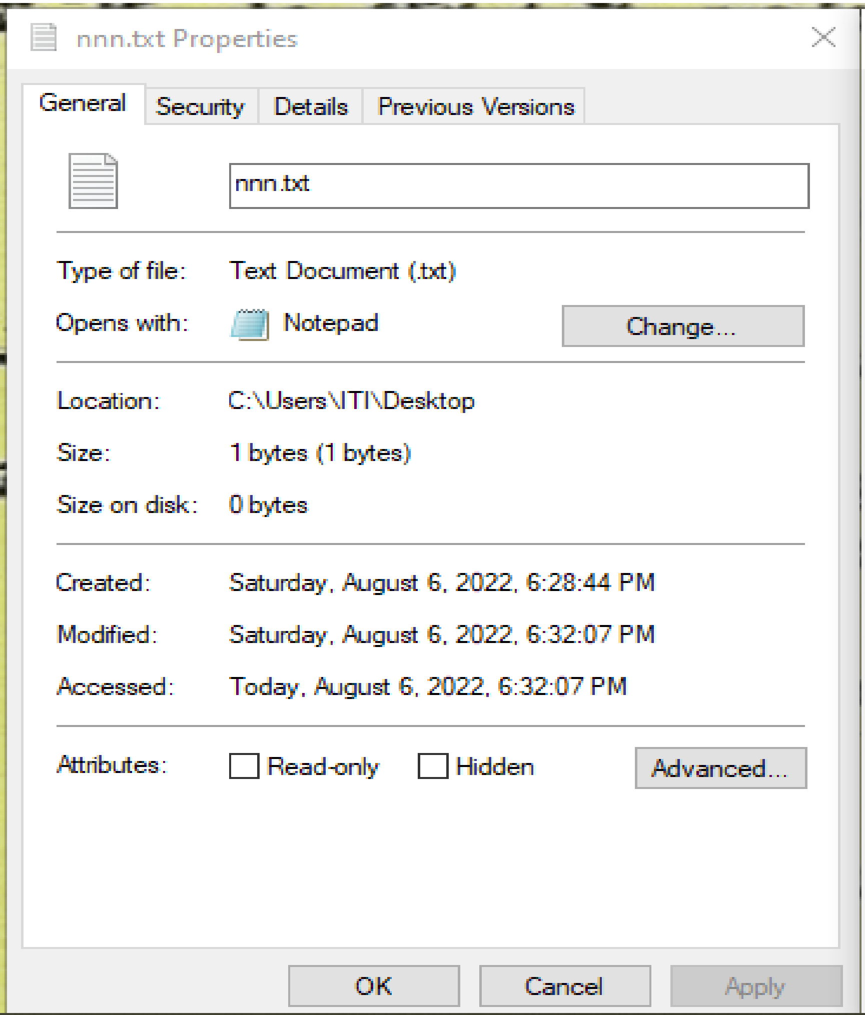
Need for File Systems



File Concept

Example on different formats (txt file, word file and PDF file) with the same content (one letter “a”) with different sizes.

if have different file format (.txt .pdf) and each file has same co
but each file has different size



nnn.txt Properties

nnn.txt

Type of file: Text Document (.txt)

Opens with: Notepad

Location: C:\Users\ITI\Desktop

Size: 1 bytes (1 bytes)

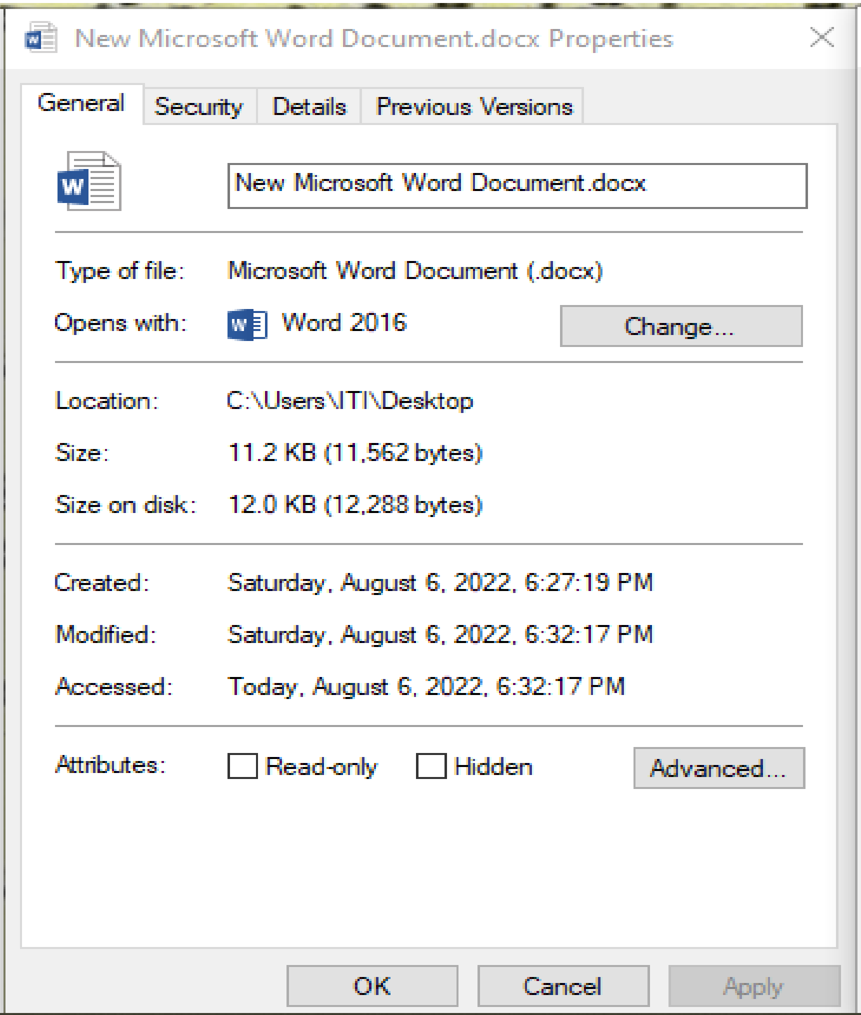
Size on disk: 0 bytes

Created: Saturday, August 6, 2022, 6:28:44 PM

Modified: Saturday, August 6, 2022, 6:32:07 PM

Accessed: Today, August 6, 2022, 6:32:07 PM

Attributes: ☐ Read-only ☐ Hidden



New Microsoft Word Document.docx Properties

New Microsoft Word Document.docx

Type of file: Microsoft Word Document (.docx)

Opens with: Word 2016

Location: C:\Users\ITI\Desktop

Size: 11.2 KB (11,562 bytes)

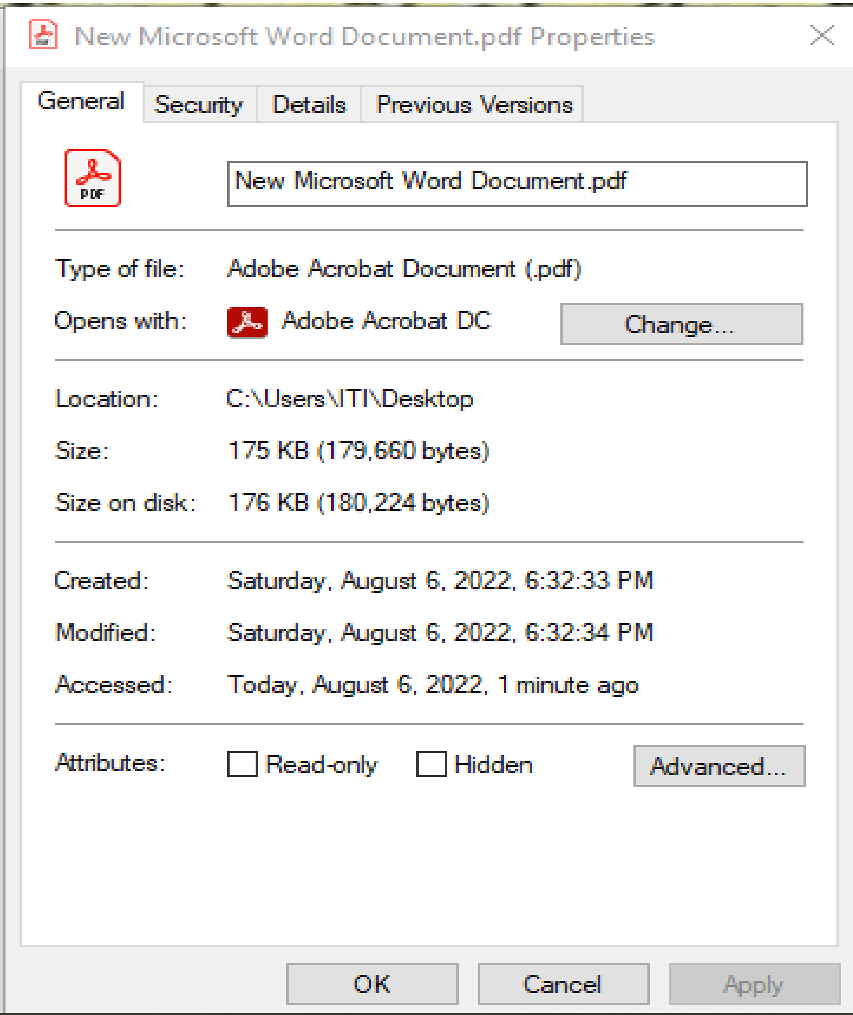
Size on disk: 12.0 KB (12,288 bytes)

Created: Saturday, August 6, 2022, 6:27:19 PM

Modified: Saturday, August 6, 2022, 6:32:17 PM

Accessed: Today, August 6, 2022, 6:32:17 PM

Attributes: ☐ Read-only ☐ Hidden



New Microsoft Word Document.pdf Properties

New Microsoft Word Document.pdf

Type of file: Adobe Acrobat Document (.pdf)

Opens with: Adobe Acrobat DC

Location: C:\Users\ITI\Desktop

Size: 175 KB (179,660 bytes)

Size on disk: 176 KB (180,224 bytes)

Created: Saturday, August 6, 2022, 6:32:33 PM

Modified: Saturday, August 6, 2022, 6:32:34 PM

Accessed: Today, August 6, 2022, 1 minute ago

Attributes: ☐ Read-only ☐ Hidden

File Concept

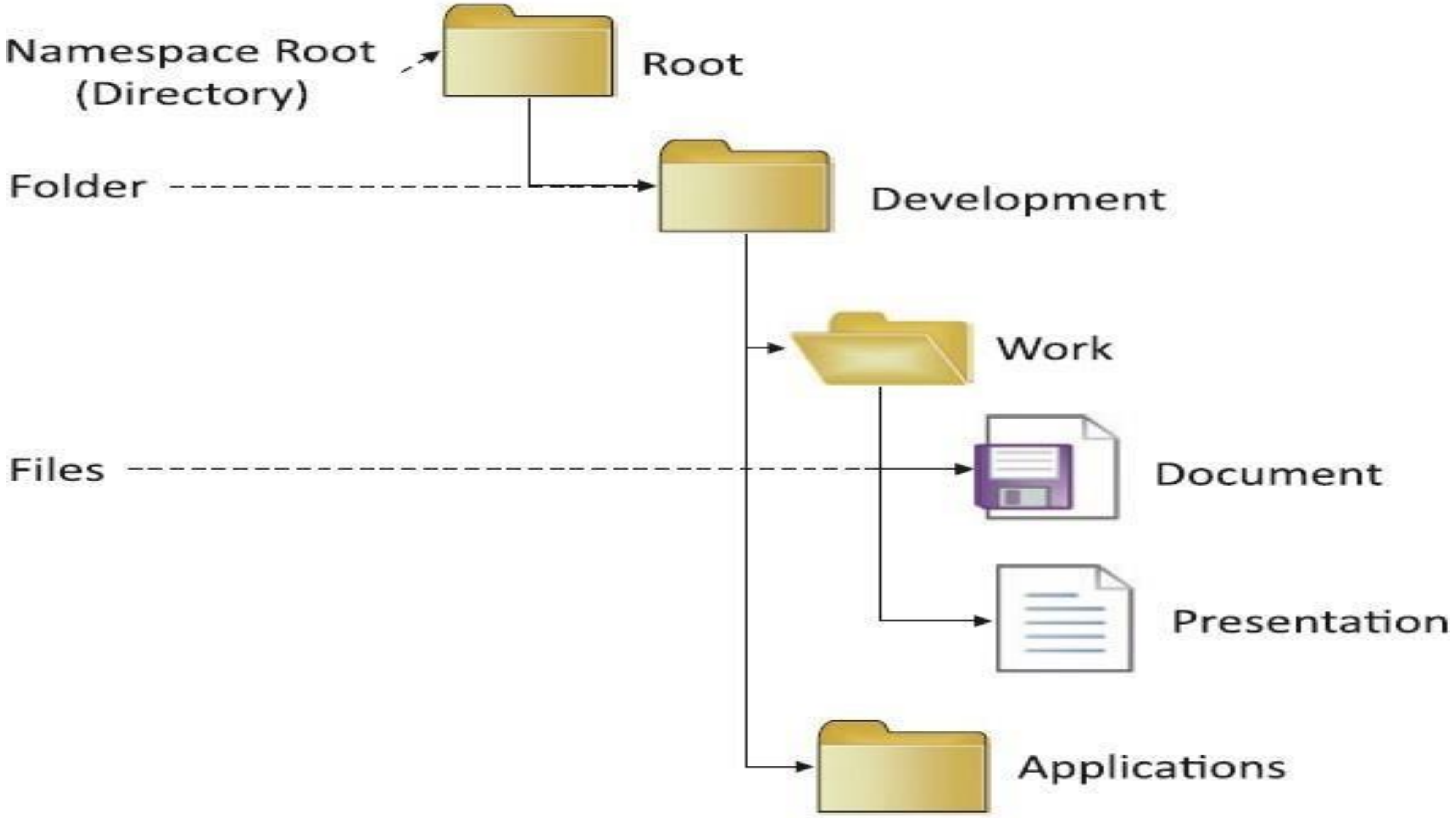
- From the perspective of the user, a file is a collection of related data that is stored together and can be accessed using a unique file ID usually referred as the file name.
- These files can be represented internally by different methods. For example, there could be **.bin** files in Windows, which only represent a sequence of bytes.
- There are also many **application-specific files**, with their own formats. It is up to the programmer to define and identify if they require a custom file format for their application or if they can leverage a standard or common file format such as the JavaScript Object Notation (**JSON**) or the Extensible Markup Language (**XML**).

Ref: https://www.uspto.gov/ebc/portal/infofilesize.htm

Directory Name Space

- The operating system defines a logical ordering of different files on the system based on the usage and underlying storage services. One of the criteria most OSs adopt is to structure their directory service to locate files efficiently.
- Most OSs ***organize*** their files in a ***hierarchical*** form with files organized inside ***folders***.
- Each folder in this case is a directory. This structure is called as the ***directory namespace***.
- The directory service and namespace have additional capabilities such as searches by size, type, access levels, and so on.
- The directory namespaces can be ***multileveled*** and adaptive in modern OSs as we can see in the following folder structure with folders created inside another folder

Directory Name Space



Access Control

- There are different access levels that can be applied at file and directory levels.
- The OS provides different access control ***IDs*** and ***permissions*** to different users on the system.
- Also, each file may also have ***different levels of permissions*** to Read, Write, Modify, and so on.
- For example, there may be specific files that we may want anyone to be able to access and Read but not Write and Modify.
- The file system provides and manages the controls to all files when accessed at runtime.
- These may also be ***helpful*** when ***more than one user*** is using the same system.

Concurrency and Cleanup Control

- There are many cases when the OS needs to ***ensure that a file is not moved or deleted when it is in use***.
- For example, if a user is making changes to a file, the OS needs to ensure that the same file cannot be moved or deleted by another application or process. In this case, the OS would cause the attempt to move or delete the file to fail with an appropriate error code.
- *As a programmer, it is appropriate to access a file with the required* ***access level*** *and* ***mode*** *(Read/Write). This also helps to be in line with the concurrency needs of the OS and guards against inconsistent updates.*

Concurrency and Cleanup Control

- The OS needs to be able to ***periodically clear temporarily created files*** that may no longer be required for the functioning of the system.
- This is typically done using a ***garbage collector on the system***.
- Many OSs ***mark unused files over a period of time*** and have additional settings that are exposed, which the user can set to clean up files from specified locations automatically.

Virtualization and User Interface

A hypervisor is software that manages multiple virtual machines (VMs) — not just a single one.
It allocates hardware resources (CPU, memory, storage) among them and isolates them to prevent inte

There are two main types:

Type 1 (Bare-metal) – runs directly on hardware (e.g., VMware ESXi, Microsoft Hyper-V).

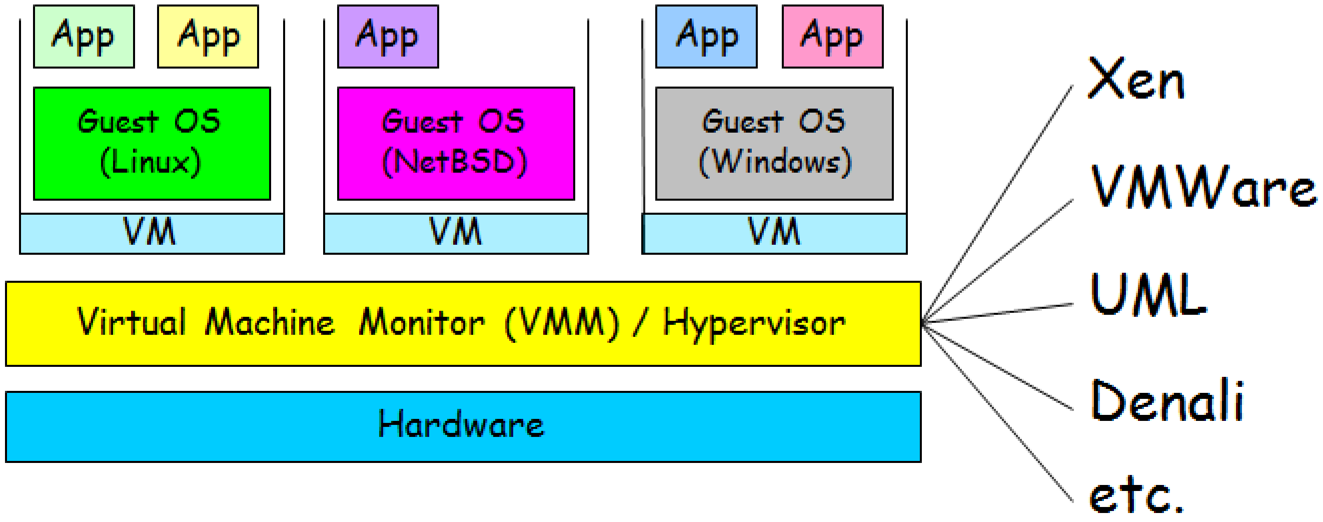
Type 2 (Hosted) – runs on top of an operating system (e.g., VirtualBox, VMware Workstation).

7.1 Virtualization

- Operating systems and modern hardware provide a feature called virtualization that ***virtualizes the hardware*** such that each calling environment believes it has the dedicated access it needs to function.
- Virtualization is delivered via so-called ***virtual machines*** (VMs).
- A VM has its own ***guest OS***, which may be the same as or different from the underlying ***host OS***.
- A user can launch a VM, much like running any other program, and log into the guest OS.
- The ***host OS*** provides a ***hypervisor***,which manages the access to the hardware.
- The guest OS is usually unaware of the internals and passes any resource/hardware requests to the host OS.
- The user can completely ***customize their VM and perform all their actions*** on this VM ***without affecting the host OS*** or any other VM on the system.

Virtualization

- VM technology allows multiple virtual machines to run on a single physical machine



7.1 Virtualization

- At a high level, VMs help effectively utilize the hardware resources and are used heavily in server and cloud deployments.
- Advantages of virtualization:
  - Run operating systems where the ***physical hardware is unavailable***.
  - Easier to create ***new machines***, ***backup machines***, etc.
  - ***Software testing*** using “clean” installs of operating systems and software.
  - ***Emulate more machines*** than are physically available.
  - ***Debug problems*** (suspend and resume the problem machine).
  - ***Easy migration of virtual machines*** (shutdown needed or not).
  - Run ***legacy systems***

7.3 User Interface

- Although the ***user interface*** (UI) is not part of the OS kernel itself.

- There can be multiple user interfaces for the OS all being implemented either as **command line** interface or a ***graphical-based*** interface

- The **graphical user interface** is the rich set of graphical front-end interfaces and functionalities provided by the OS for the user to interact with the computer.

- There could be an alternate simpler interface through a **command line** that most OSs also provide for communication. This is a text-based interface.

- *It is common for programmers to use the shell interface instead of the GUI for quickly traversing through the file system and interacting with the OS.*

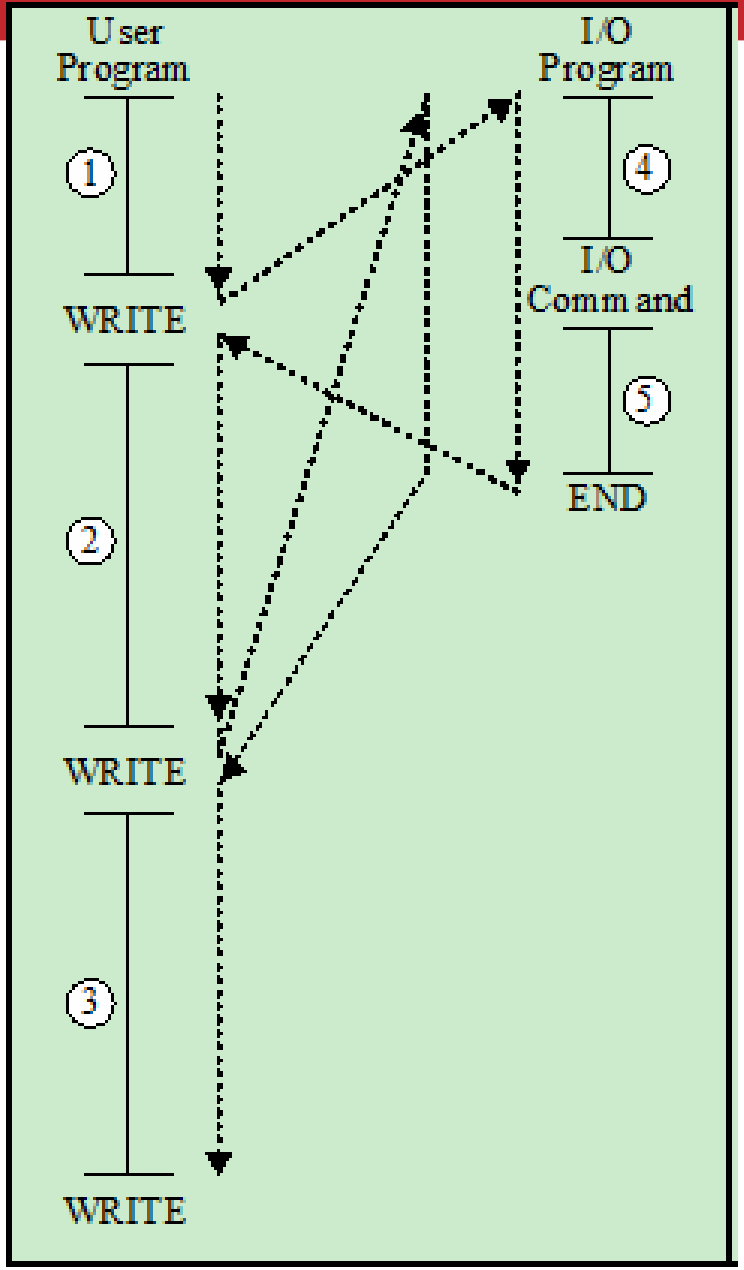
Executing instructions

- Instructions are executed one by one
- You must wait the first instruction to finish

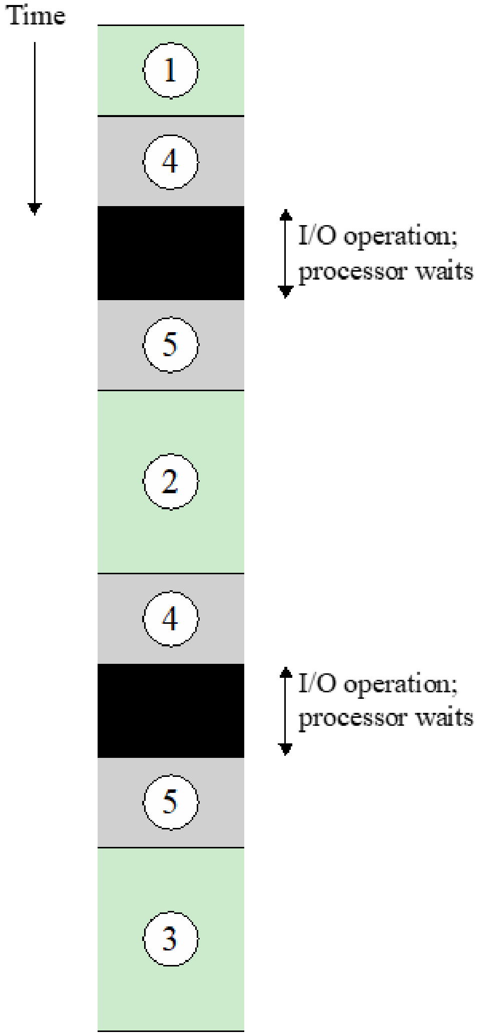
Interrupts

I/O Operation

- Prepare commands
- Send commands to device
- Wait the device to be ready
- Start operation



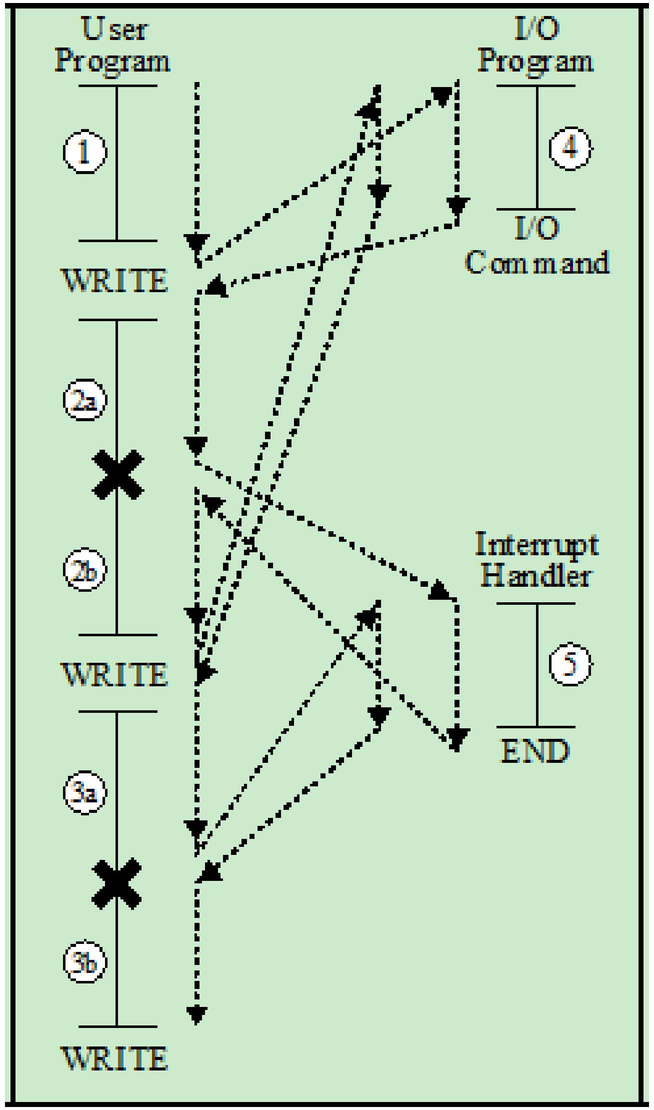
(a) No interrupts



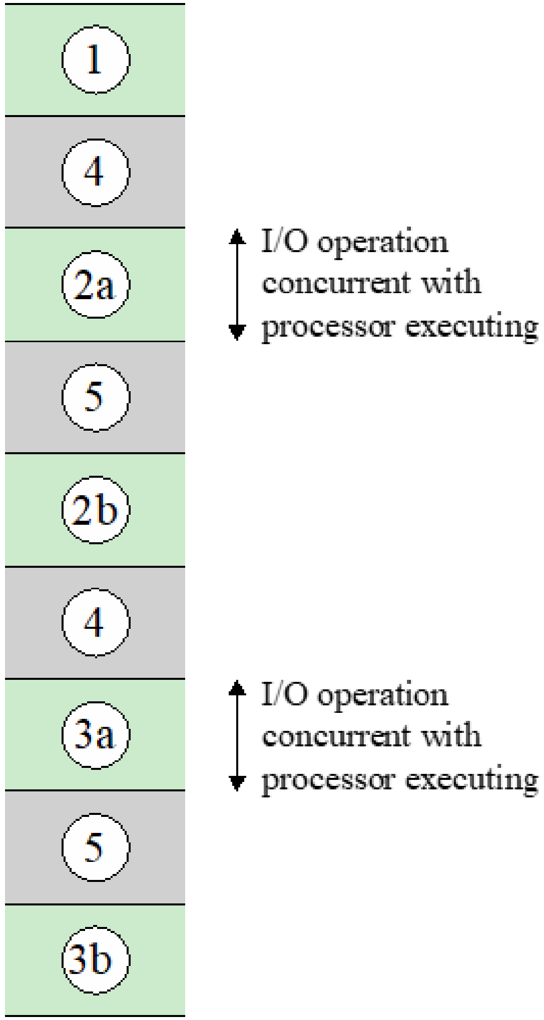
Time
1
4
I/O operation; processor waits
5
2
4
I/O operation; processor waits
5
3

(a) Without interrupts

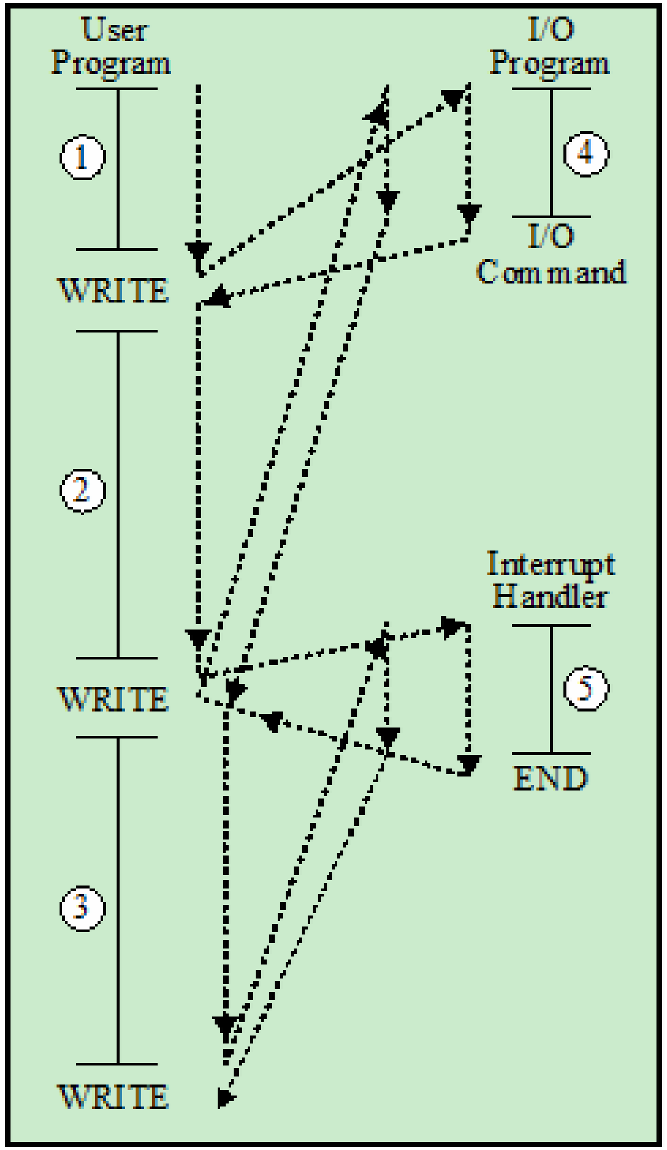
- Normal execution of programs may be preempted if some device requires urgent servicing.
- The normal execution of the current program must be interrupted – the device raises an *interrupt* signal.



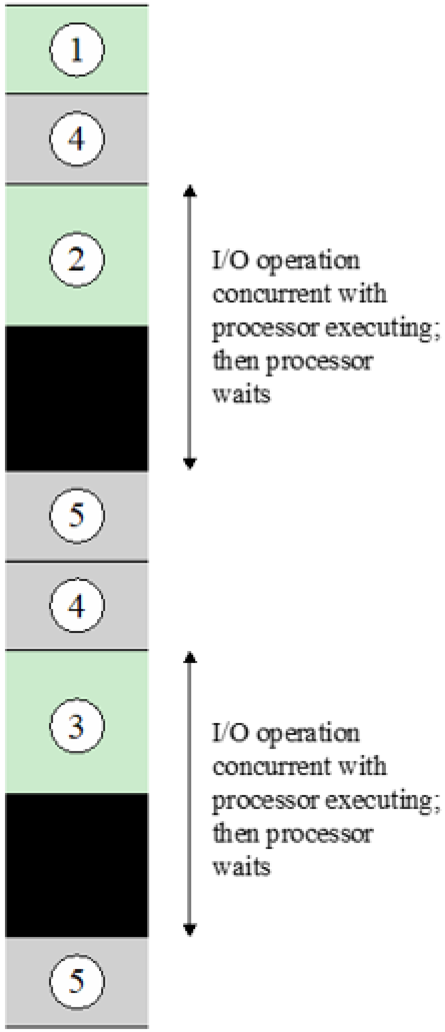
(b) Interrupts; short I/O wait


1
4
2a
5
2b
4
3a
5
3b
I/O operation concurrent with processor executing
I/O operation concurrent with processor executing

(b) With interrupts

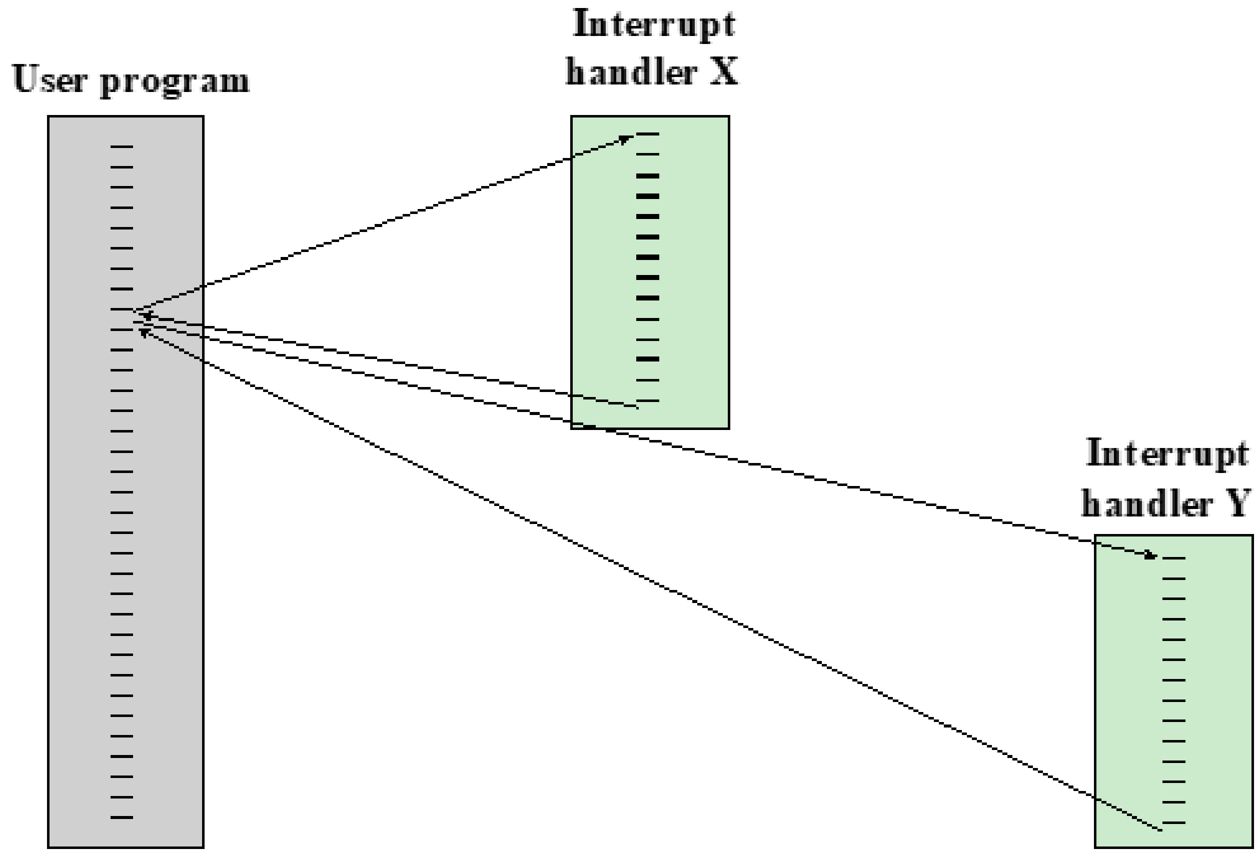


(c) Interrupts; long I/O wait


1
4
2
I/O operation concurrent with processor executing; then processor waits
5
4
3
I/O operation concurrent with processor executing; then processor waits
5

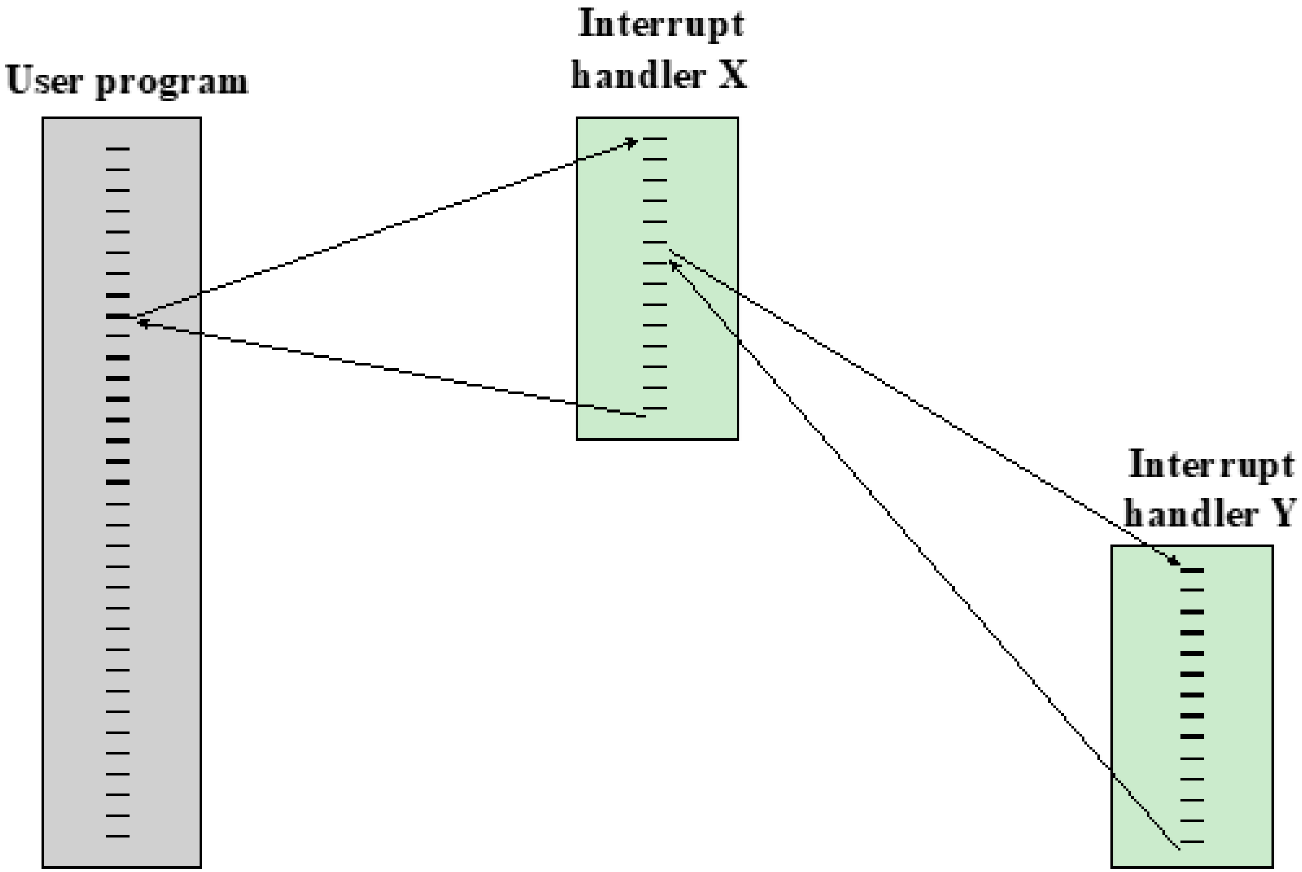
(b) With interrupts

Transfer of control with multiple interrupts



(a) Sequential interrupt processing

Transfer of control with multiple interrupts



(b) Nested interrupt processing

Example

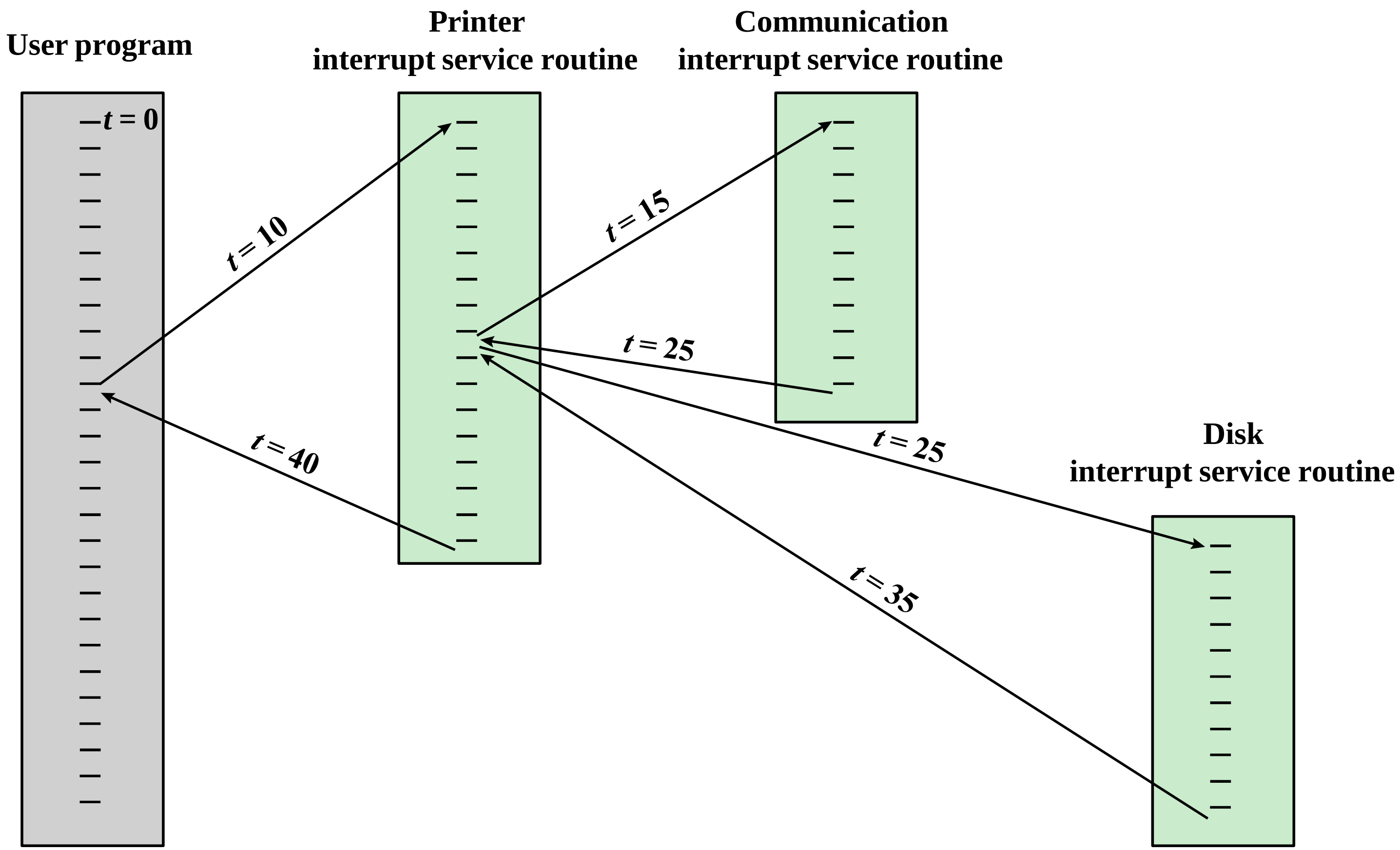

User program
Printer interrupt service routine
Communication interrupt service routine
Disk interrupt service routine
t = 0
t = 10
t = 15
t = 25
t = 25
t = 35
t = 40

Figure 3.14 Example Time Sequence of Multiple Interrupts

Save context

- Save current state

  PC, Registers

Thank You

What is the

Big Data Analysis?

What is the Data?

What is the Data? What is the dataset?

What is the Data?

measurements (unprocessed or processed) represented as text, numbers, or multimedia.

What is the Dataset?

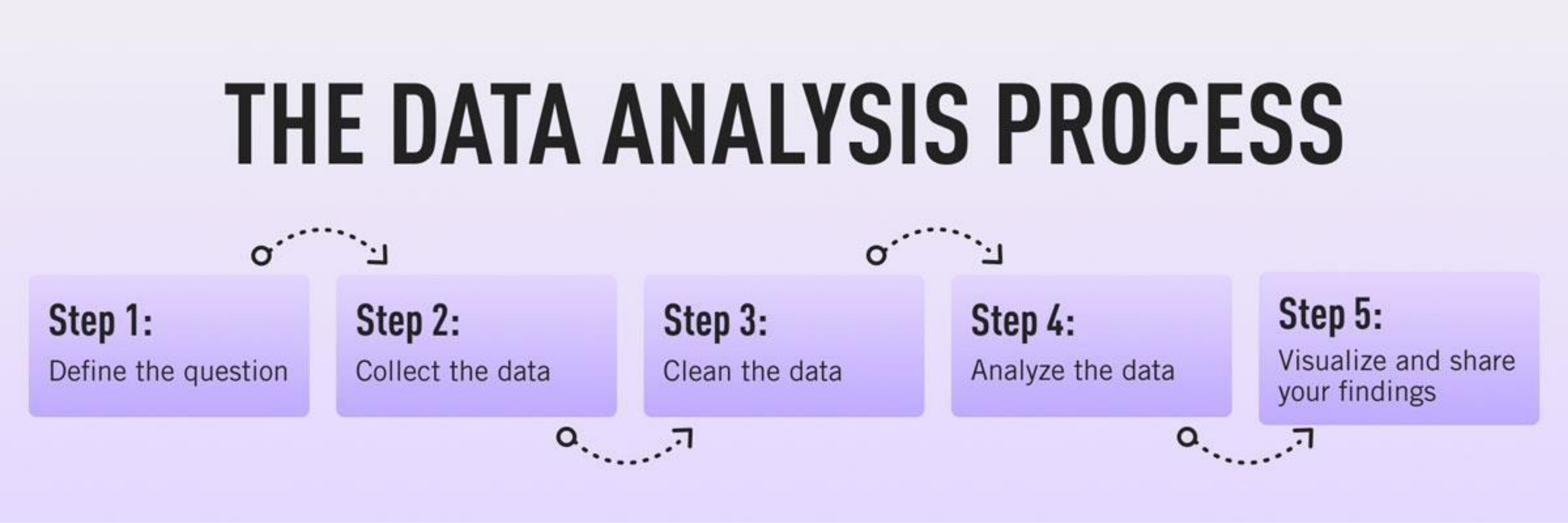
A data set is a collection of numbers or values that relate to a particular subject. For example, the test scores of each student in a particular class is a data set.

What is Data analysis?

- It is a process of inspecting, cleansing, transforming, and modeling data.

- this process is done to discover useful information, informing conclusions, and support decision-making.

What is Data Analysis Process?



THE DATA ANALYSIS PROCESS

Step 1: Define the question

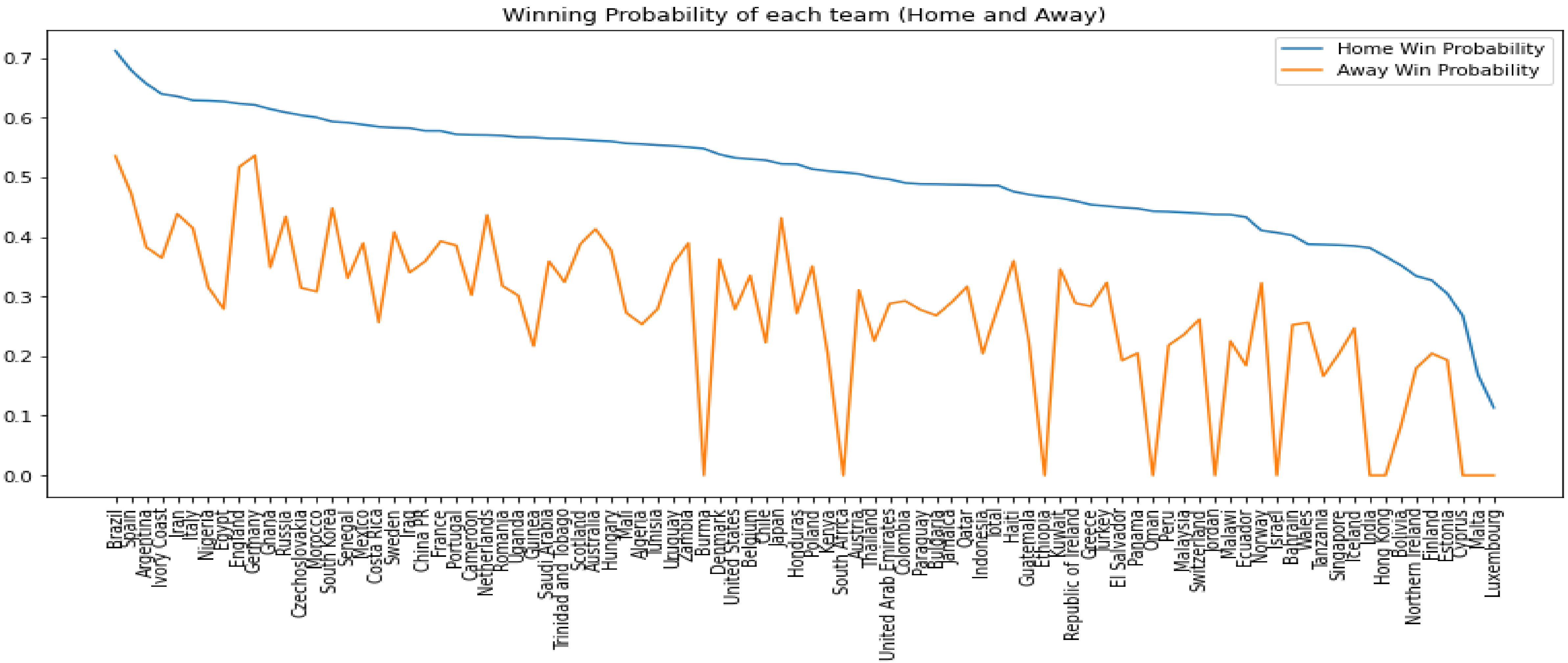
Step 2: Collect the data

Step 3: Clean the data

Step 4: Analyze the data

Step 5: Visualize and share your findings

Data Analysis Process (Data Visualization Example)



What is Big Data?

No standard definition:

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy.

Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so on."

Who is generating Big Data?

Social

Twitter for Business

Small business advertisers, learn how to get $100 in free Twitter advertising from American Express.

User Tracking & Engagement

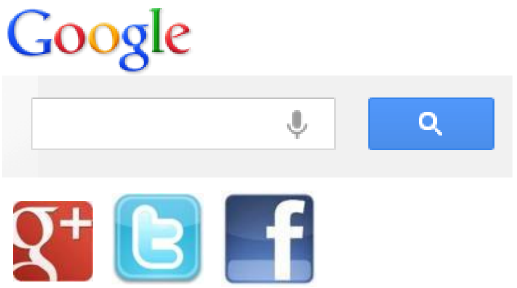
Homeland Security

U.S. DEPARTMENT OF HOMELAND SECURITY

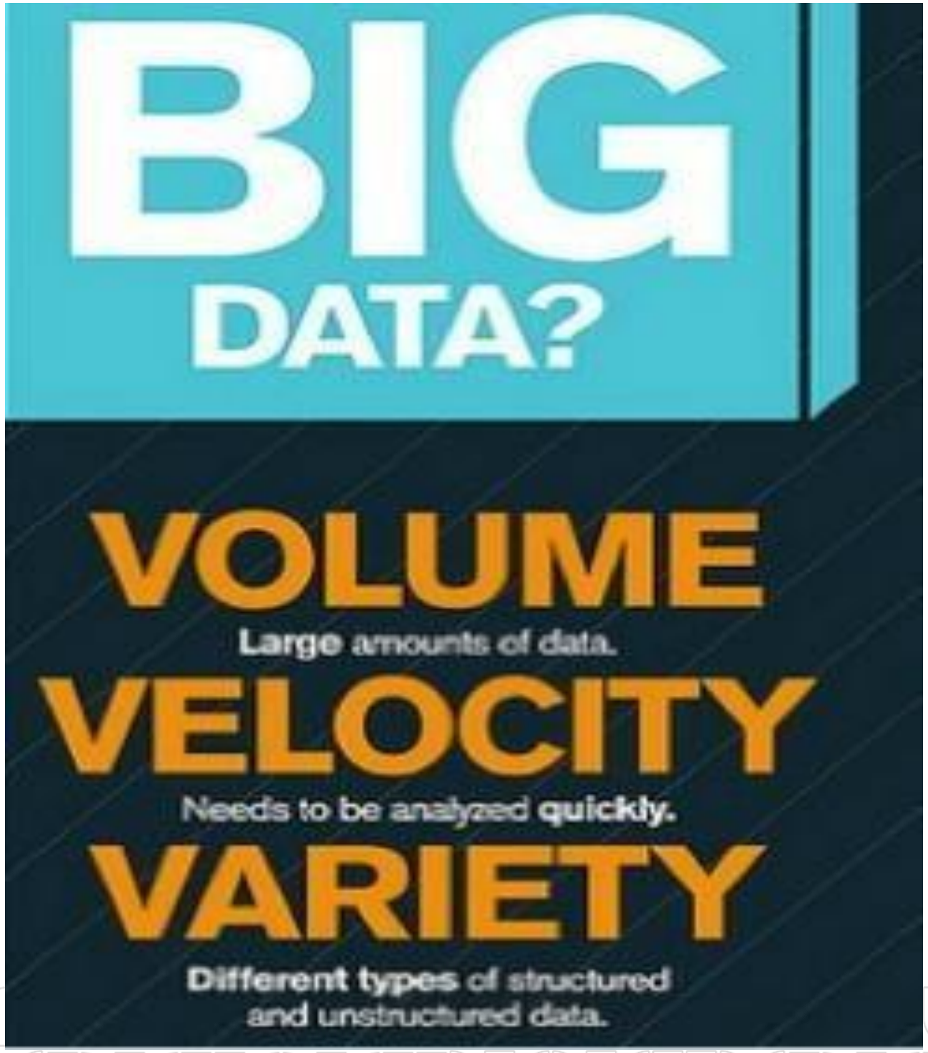
eCommerce

Financial Services

Real Time Search



Big Data Characteristics: 3V



BIG DATA?

VOLUME

Large amounts of data.

VELOCITY

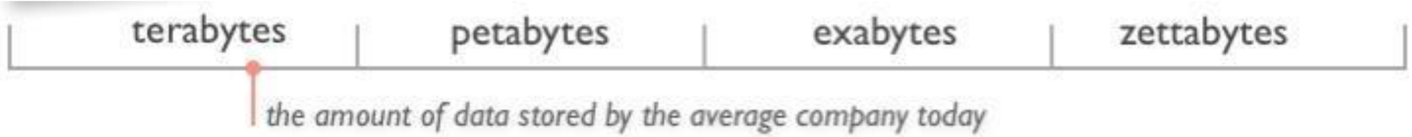
Needs to be analyzed quickly.

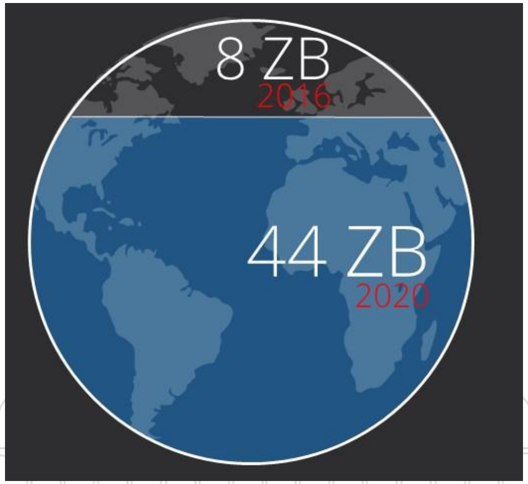
VARIETY

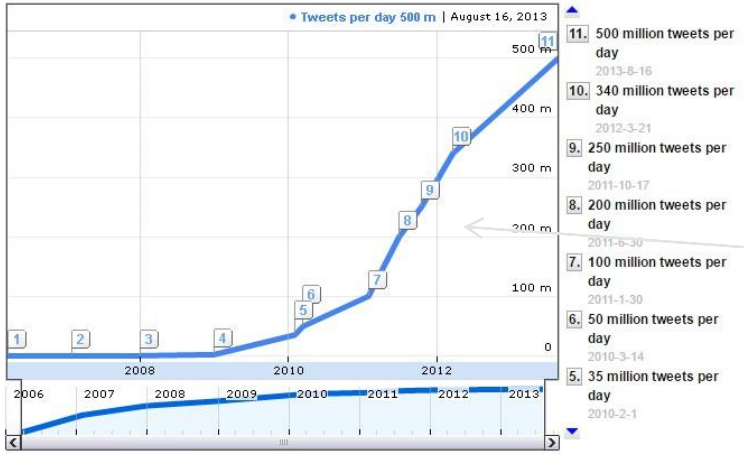
Different types of structured and unstructured data.

Big Data Characteristics: 3V

- Data Volume
  - Growth 40% per year
  - From 8 zettabytes (2016) to 44zb (2020)
- Data volume is increasing exponentially

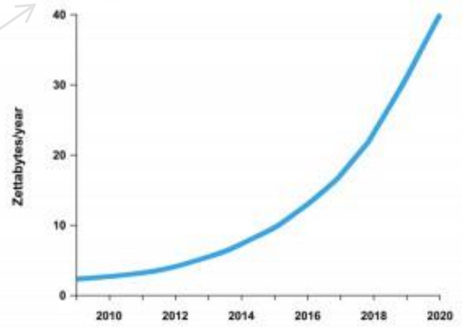






Exponential increase in collected/generated data

Size of the Digital Universe – Annual Data Created & Consumed



Data Source: IDC Digital Universe 2013

Variety

- Different Types:
  - Relational Data (Tables/Transaction/Legacy Data)
  - Text Data (Web)
  - Semi-structured Data (XML)
  - Graph Data
    - Social Network, Semantic Web (RDF), …
  - Streaming Data
    - You can only scan the data once
  - A single application can be generating/collecting many types of data
- Different Sources :
  - Movie reviews from IMDB and Rotten Tomatoes
  - Product reviews from different provider websites

Velocity

- Data is begin generated fast and need to be processed fast
- Online Data Analytics
- Late decisions ➔ missing opportunities
- Examples
  - **E-Promotions:** Based on your current location, your purchase history, what you like ➔ send promotions right now for store next to you
  - **Healthcare monitoring:** sensors monitoring your activities and body ➔ any abnormal measurements require immediate reaction
  - **Disaster management and response**

Extended Big Data Characteristics: 5V

- Volume: In a big data environment, the amounts of data collected and processed are much larger than those stored in typical relational databases.
- Variety: Big data consists of a rich variety of data types.
- Velocity: Big data arrives to the organization at high speeds and from multiple sources simultaneously.

- Veracity: Data quality issues are particularly challenging in a big data context.
- Value: Ultimately, big data is meaningless if it does not provide value toward some meaningful goal.

Veracity (Quality & Trust)

- *Data = quantity + quality*
- When we talk about big data, we typically mean its quantity:
  - What capacity of a system provides to cope with the sheer size of the data?
  - Is a query feasible on big data within our available resources?
  - How can we make our queries tractable on big data?
  - . . .
- Can we trust the answers to our queries?
  - Dirty data routinely lead to misleading financial reports, strategic business planning decision ⇒ loss of revenue, credibility and customers, disastrous consequences
- *The study of data quality is as important as data quantity*

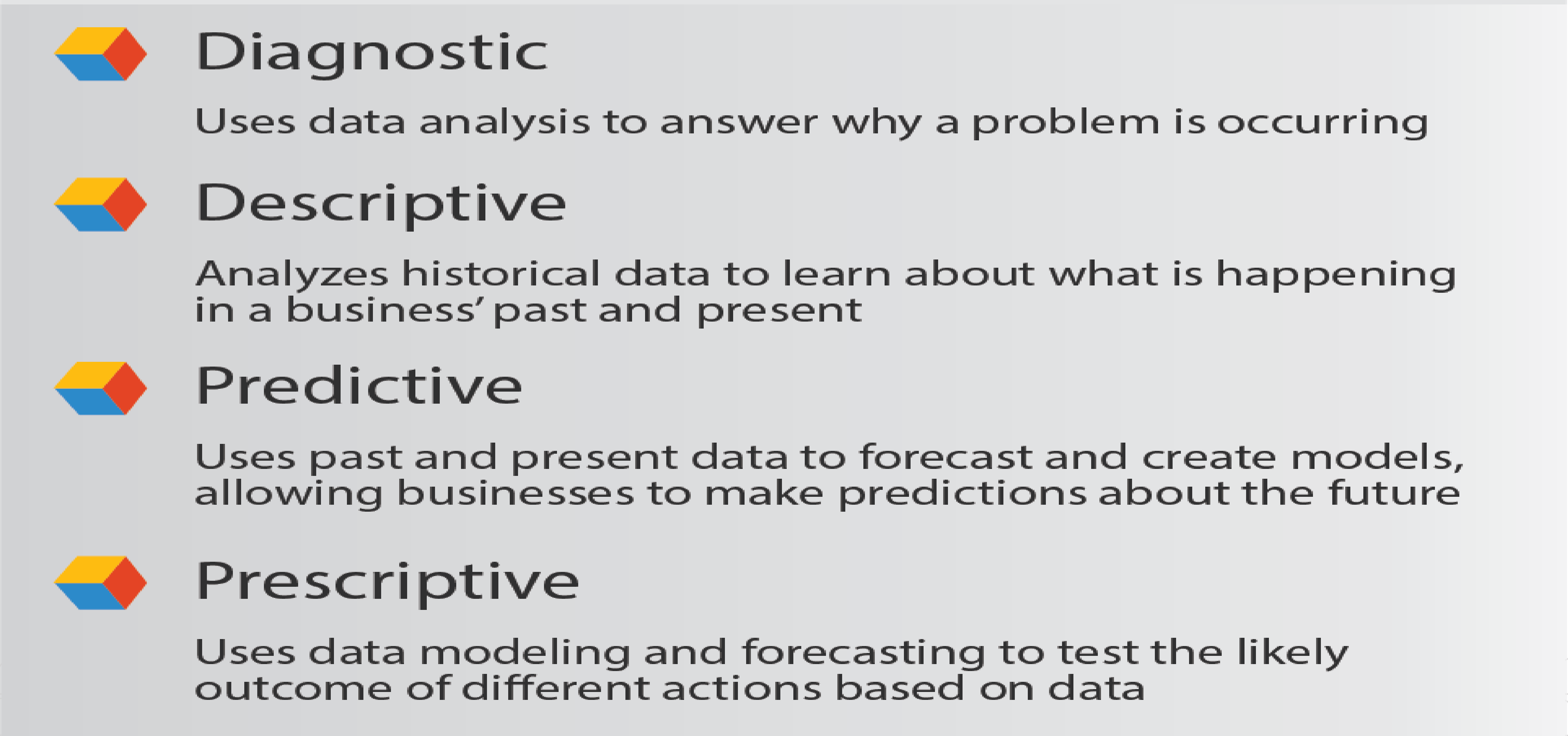
Value

- Big data is meaningless if it does not provide value toward some meaningful goal

What is the Big Data Analysis?

- Big data analysis is the process of storing, managing, processing large sets of data to generate outputs. They help the businesses to understand the trends, patterns, user's preferences.
- Big data platforms help to analyze large chunks of data, in multiple formats, at a very high speed and efficiency.

Types of Big Data Analysis



Diagnostic

Uses data analysis to answer why a problem is occurring

Descriptive

Analyzes historical data to learn about what is happening in a business’ past and present

Predictive

Uses past and present data to forecast and create models, allowing businesses to make predictions about the future

Prescriptive

Uses data modeling and forecasting to test the likely outcome of different actions based on data

Types of Big Data Analysis: Diagnostic Analytics

Why did it happened ?

It is used to determine why something happened in the past.

Like a doctor investigating a patient’s symptoms, they aim to understand the underlying issues and determine why an issue is happening.

Some questions that would have to be addressed with diagnostic analytics include:

- Why have sales increased without any increased marketing attention for a certain region?
- Why did employee performance fall during this month?

Types of Big Data Analysis: Descriptive Analytics

What is happening now based on the incoming data ?

It provide insight into the past and answer: “What has happened?”

The descriptive analytics does exactly what the name implies they “describe” or summarize raw data and make it interpretable by humans.

The aim of these types of analytics is to learn from the past. One common example is analyzing seasonal purchasing trends to determine the best time to launch a new product.

Types of Big Data Analysis: Predictive Analytics

What might happen in the future if … ?

It uses statistical models and forecasts techniques to understand the future and answer: “What could happen?”

Predictive analytics makes predictions about future events by applying data modeling, artificial intelligence, **machine learning algorithms and deep learning algorithms** and statistics **to current and historical data**.

Example: Cancer Prediction and epileptic seizure prediction

Types of Big Data Analysis: Prescriptive Analytics

What action should be taken in this case?

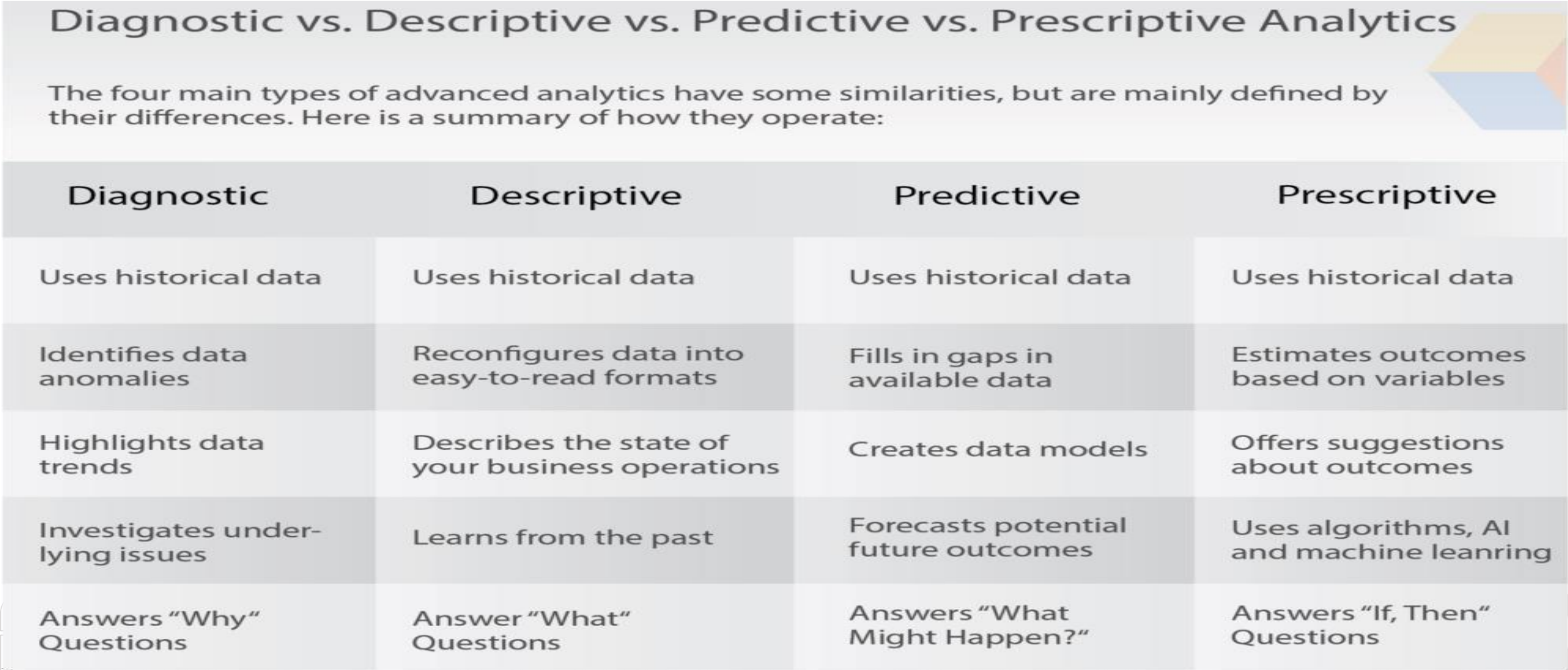
It uses optimization algorithms to advice on possible outcomes and deals more with trial and error

It allows users to “prescribe” a number of different possible actions and guide them towards a solution. this analytics is all about providing advice.

If predictive analytics answers, “What might happen?” then prescriptive analytics answers, “What do we have to do to make it happen?” Prescriptive and has a bit of a hypothesis-testing nature to it.

For example: Self Driving cars

Types of Big Data Analysis



Diagnostic vs. Descriptive vs. Predictive vs. Prescriptive Analytics

The four main types of advanced analytics have some similarities, but are mainly defined by their differences. Here is a summary of how they operate:

Diagnostic	Descriptive	Predictive	Prescriptive
Uses historical data	Uses historical data	Uses historical data	Uses historical data
Identifies data anomalies	Reconfigures data into easy-to-read formats	Fills in gaps in available data	Estimates outcomes based on variables
Highlights data trends	Describes the state of your business operations	Creates data models	Offers suggestions about outcomes
Investigates under-lying issues	Learns from the past	Forecasts potential future outcomes	Uses algorithms, AI and machine leanring
Answers "Why" Questions	Answer "What" Questions	Answers "What Might Happen?"	Answers "If, Then" Questions

Types of Big Data Analysis

At the end of the day, diagnostic, descriptive, predictive and prescriptive analytics solutions work together to build a story.

It’s a story about what your business has, what it needs, and what it could achieve.

So you can make decisions that are fully informed by your data.

Ref: https://www.edureka.co/blog/big-data-analytics/
Ref: https://www.selecthub.com/business-intelligence/predictive-descriptive-prescriptive-analytics/

Dealing with AWS – Big Data services

Firstly we go to AWS pricing calculator through the following link?:

https://calculator.aws/#/
To create estimate.

And then select Amazon EMR Amazon EMR (Elastic MapReduce) which is the cloud big data platform

To select the numbers of core nodes and its type (EC2 instance) “Elastic Compute Cloud”.
So to know more info about the **Amazon EC2 Instance Types, we go to this link: (https://aws.amazon.com/ec2/instance-types/) and also select (**Compute Optimized**) to deal with** high performance processors or (Memory optimized) to deal with large amount of memory and so on.

Thank You