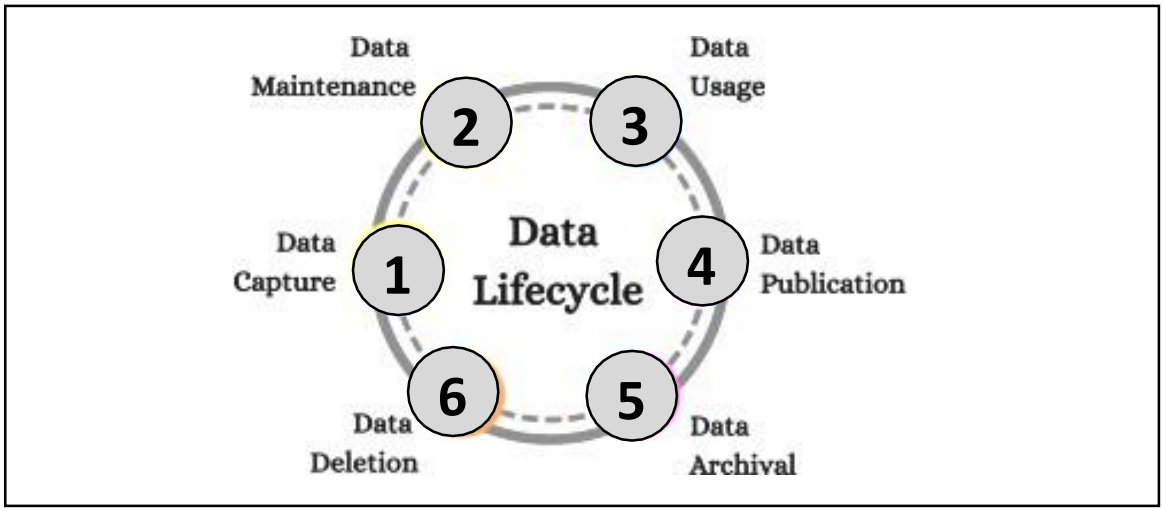
Unit IV: Data Life Cycle Management

1. Explain the various stages of Data Life Cycle

Data Life Cycle:

The data life cycle is the sequence of stages that a particular unit of data goes through from its initial generation or capture to its eventual archival and/or deletion at the end of its useful life.

The Stages of Data Life Cycle Management:



1. Data Capture(or) Data Creation

The first phase of the data lifecycle is the creation/capture of data. This data can be in many forms, *e.g.*, PDF, image, Word document, SQL database data. Data is typically created by an organization in one of the 3 ways:

- **Data Acquisition:** Acquiring already existing data which has been produced outside the organization.
- **Data Entry:** Manual entry of new data by personnel within the organization.
- **Data Capture:** Capture of data generated by devices used in various processes in the organization.

2. Data Maintenance/Storage

Once data has been created within the organization, it needs to be stored and protected, with the appropriate level of security applied. A robust backup and recovery process should also be implemented to ensure retention of data during the lifecycle.

3. Data Usage

During the usage phase of the data lifecycle, data is used to support activities in the organization. Data can be viewed, processed, modified and saved. An audit trail should be maintained for all critical data to ensure that all modifications to data are fully traceable. Data may also be made available to share with others outside the organization.

4. **Data Publication:** This is one way that data can leave your enterprise. Say you publish white paper that is downloaded by multiple companies, or you use data you've collected to send out invoices or investment statements to customers.

5. Data Archival

Data Archival is the copying of data to an environment where it is stored in case it is needed again in an active production environment, and the removal of this data from all active production environments.

A data archive is simply a place where data is stored, but where no maintenance or general usage occurs. If necessary, the data can be restored to an environment where it can be used.

6. Data Deletion/Destruction

The volume of archived data inevitably grows, and while you may want to save all your data forever, that's not feasible. Storage cost and compliance issues exert pressure to destroy data you no longer need. Data destruction or purging is the removal of every copy of a data item from an organization. It is typically done from an archive storage location. The challenge of this phase of the lifecycle is to ensure that the data has been properly destroyed. It is important to ensure before destroying data that the data items have exceeded their required regulatory retention period.

Having a clearly defined and documented data lifecycle management process is the key to ensuring Data Governance can be carried out effectively within your organization.

In conclusion, Data has evolved to be the new currency in the digital era that we live in. Hence, managing data has become an essential aspect in terms of data handling, giving rise to the concept of data lifecycle management.

2. Explain about Data in the organization

Every minute, we are sending over 204 million emails and sharing 2.5 million pieces of content via social networks. In the year 2020, the IoT(Internet Of Things) will comprise more than 30 billion connected devices generating 600 ZB(Zetta Byte(10^{21}) of data per year. The amount of stored and non-stored datais growing every day. To make sense of this data, reveal its patterns, and to be able to use it for operational improvements and more efficient strategic decision making is still a challenge.

Without the structure and clear purpose, data is unusable. It's nothing but tons of unrelated, chaotic information without any insights. One must understand data to create added value. Only when the structure of this data is thoroughly analyzed and understood, with a clear application vector assigned to it data can be transformed into intelligence.

Data is a valuable resource. Collected and stored in all areas of life and business, the global volumes of data are doubling every two years. Should we succeed in evaluating this data intelligently and comply with the principles of data diversity and data sovereignty, enormous potential for value creation can be generated.

Big players in the IT industry, research institutions, and numerous startups have been somewhat active in the data field. In many sectors, technologies for the economic utilization of the data collected are also very advanced, although, on a large scale, the exploitation of data is still in its infancy. Being the "digital capital" at the heart of digital transformation, data itself becomes

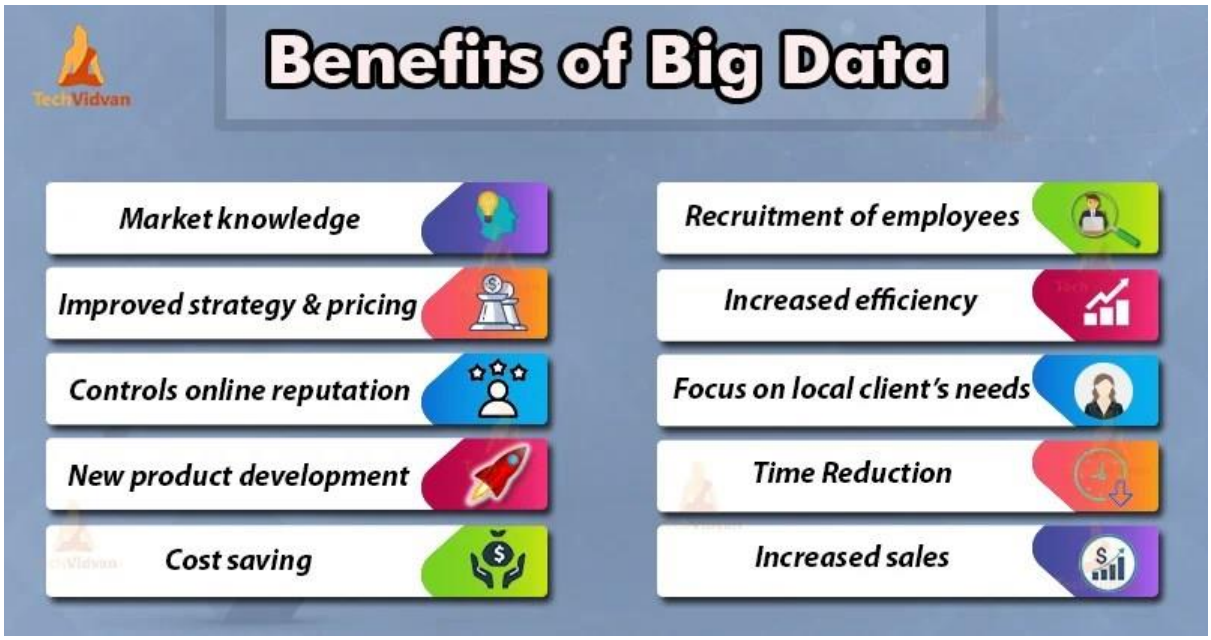
a critical capability and pathway to realizing value from it.

Today, every company has large datasets, but only a few of them are indeed able to convert their big data into smart data. Big Data means all data. But data as such is meaningless. For the companies to be able to obtain substantial benefits from this data, they have to turn it into actionable data.

Data is useful information that you collect to support organizational decision making and strategy.

The following are the some of the reasons why data is important, and what you can do with it in the organization:

- Improve people's livelihood
- Make informed decisions
- Get the results you want at proper time
- Find solutions to complex problems
- Back up the data units
- Make Strategic Decisions
- Keep track of it all
- Generate revenue
- Availability of resources all the time.



3. What are the various ways data can enter into the organization

(or)

Distinguish between ways that data enters the organization

Keyboards

Wired and wireless data keyboards break down linguistic input and software commands into individual units of meaning. You can type a single letter, combine alphanumeric keys with command keys to produce accented or specialized characters, and enter simple or complex instructions to control software processes. From the traditional QWERTY keyboard, adapted from the typewriter to layouts such as the keyboard/keypad which promotes faster typing, and adapted variations with additional keys, these devices translate finger-based key presses into data input.

Pointing Devices

Pointing devices translate taps, gestures and drawing input, either through wired or wireless interfaces. From clicking on a menu with a mouse or trackball, to tapping on a touchpad or touch screen, or drawing on a graphics tablet, these devices instigate commands and help retouch photographs, create illustrations and simulate the behavior of other devices, includingpaintbrushes and airbrushes. CAD pucks enable you to establish precise points of reference on an architectural plan or 3D rendering. Recreationally speaking, the world of computer games relieson devices that translate three-dimensional gestures into the world of a flight simulator, fight scene or race track, including joysticks, game pads and driving simulators.

Data Drives

Flash, optical and hard drives store the output of computer processes, but they also provide the input for other functions. From files that contain data to be manipulated to temporary data storage that provides input from a clipboard or program, these devices simplify and speed the act of making information available to software processes. Some of these devices contain moving platters, whereas others rely on solid-state NAND flash chips or inserted writeable/rewriteable optical discs. They may be installed as internal computer components or plugged in to a USB, FireWire or Thunderbolt port.

Audio/Video Devices

Computers accept a wide range of audio-data inputs. You can dictate audio input into a headset for use in a text-to-speech program that converts your words into word processing, sing or

play a musical instrument into a microphone for recording in an audio file, plug an electronic instrument into your computer for direct input of synthesized sounds or transfer audio from a recording device. To bring visual information into a new or existing document, you can digitize objects or printed information on a two- or three-dimensional scanner, transfer files from a digital camera or combined audio/video input from a camcorder, or record an individual or a surveillance scene captured on a webcam or security camera.

MIDI Devices(Musical Instrument Digital Interface)

Keyboards, synthesizers and other musical technologies provide computer input in the form of MIDI data. The Musical Instrument Digital Interface specification includes three types of data pathways, received and transmitted through MIDI IN, MIDI OUT and MIDI THRU ports. Instruments plug in to a MIDI IN jack on a computer interface, which plugs in to a USB or other standard data port. Instruments connect into series of daisy-chained devices by means of their MIDI THRU jacks, which enables the output of one piece of hardware to control or influence the behavior of another.

Specialized Hardware Devices

From testing and diagnostic equipment, laboratory measurement devices and manufacturing hardware to assistive technology that makes computers accessible to persons with disabilities, specialized input hardware solves equally specialized problems or overcomes specific challenges. Along with devices designed for medical, scientific, engineering and manufacturing use, these special-purpose input devices include gestural hardware that translates three-dimensional movement of the hands or body into character movement for animation or provides a substitute for conventional pointing devices.

4. How to identify the various forms of data to store and use within an organization.

Big Data; the Volume

The term Big Data was introduced by O-Reilly Media in 2005 to refer to "a set of data that is so large it just can't be managed and processed through traditional business intelligence tools". It's so voluminous and complex that it can't be analyzed with the standard data processing applications. Since then, the ever-increasing number of **companies** learn to deal with massive amounts of information to make better, more informed decisions.

Smart Data; the Value

In contrast to Big Data, **Smart Data** is actionable and does make sense and has a clear purpose. It's not about the volume of the data you are collecting – it's about the actions taken in response to that data. It's a concept that developed along with the development of algorithm- based technologies such as artificial intelligence and machine learning.

Smart Data is usually generated close to the data source, using edge computing technologies. So instead of collecting all data from a source, we process the data in the source to end dumping in our data lake only the valuable data; the Smart Data. They distinguish between two main types of smart data. The first type refers to the data collected by **smart sensors and associated with the Internet of Things** (IoT) movement. In this case, an intelligent data entry point captures information

and uses it for real-time decision making.

Dark Data

Dark Data is not just the small portion of big data; it's the most significant chunk of data with a massive amount of potential that is still ignored by the companies. **Dark Data** is the data that lies below the surface, hiding within the company's internal networks and holding piles of relevant information that can be moved to the data lake and generate vital business and operational insights.

Dark Data is the non-stored, and unstructured data organizations produce during their regular activities. It represents more than 80% of total data. Then, dark data is a great hidden resource that flows untapped through many organizations.