6.1.3 How data is generated

Use the link below to make notes on the 6 V’s of big data

<https://www.duckcreek.com/blog/the-6-vs-of-big-data-in-the-insurance-industry>

* **Volume** – Refers to the vast amount of data generated every second. This could range from terabytes to zettabytes of information.
* **Velocity** – The speed at which data is generated, collected, processed, and analyzed. Examples include real-time data from sensors or social media feeds.
* **Variety** – The diverse formats of data, such as structured data (databases), unstructured data (text, images, videos), and semi-structured data (XML, JSON).
* **Veracity** – The accuracy, quality, and trustworthiness of the data. This ensures that decisions based on the data are reliable.
* **Value** – The usefulness of the data in deriving insights, making decisions, and creating business opportunities.
* **Variability** – The inconsistency of data flows and formats, which can complicate data management and analysis.

Human Data:

Data generated everyday by people using digital systems

Suggest some methods in which data is generated by people

* Messages
* Social media posts
* Media (photos, videos, audio)
* Emails

AI and ML:

What is AI and ML?

* AI and ML are algorithms that are given mass amounts of data in order to “learn” the patterns within that data and can be used to make predictions
* AI is the science of training machines to perform human tasks
* ML is a subset of AI; it teaches a digital device to learn by looking for pattern in data to attempt to make a conclusion

What is a sensor?

A sensor is a piece of equipment that detects changes in the environment around it

Types of sensors:

* Optical sensor – measure light and convert it into electrical signals
* Gas sensor – keeps track of gas levels – particularly useful in ensuring gas levels aren't toxic for humans
* Light sensor – detect the presence of visible, infrared or ultraviolet light
* Analog sensor – measures a physical quantity or phenomenon and provides an output signal that is directly proportional to the quantity being measure

What is the Internet of Things (IoT)?

The Internet of Things (IoT) refers to devices connected via the internet that collect, share, and act on data, often without human input. Examples include smart home devices, wearables, and connected vehicles, enhancing automation and efficiency in various sectors.

Advantages and Disadvantages of using IoT in the home and business sectors:

|  |  |  |
| --- | --- | --- |
|  | **Advantages** | **Disadvantages** |
| **Home** | * Can be used as home security devices to protect against thieves * Can increase energy efficiency (e.g., smart lights can be powered on/off remotely) | * IoT devices are often vulnerable to hackers * IoT can be expensive to set up and some may require subscriptions to even function |
| **Business** | * Enhanced operational efficiency * Improved quality of life * Environmental benefits * Business innovation | * Increases data vulnerability * Incompatibility can lead to high maintenance costs * Reliant on network stability, can limit performance * Ethical concerns over privacy |

Smart cities:

Investigate the Smart City concept, including a real-world example. Discuss with the rest of your group the devices that could be connected, as well as the advantages and disadvantages of a smart city.

A smart city uses advanced technologies and interconnected devices (IoT) to improve quality of life and sustainability. This is achieved through data collection and analysis to optimize urban operations, including energy use, transportation, public safety, and waste management.

Real world example:

* Barcelona (Spain):
  + Smart Parking: Sensors guide drivers to available parking spaces, reducing congestion
  + Smart Lighting: Streetlights adjust brightness based on pedestrian activity/density
  + Connected Public Services: Bins with sensors notify when they need to be emptied, improving waste management efficiency

1. What data have you generated today? Make a list of all the data that you have personally created and compare with another student

* YouTube watch data
* Discord messages
* WhatsApp messages
* GitHub clones and commits
* Microsoft Office (college work)
* Firefox search history
* ChatGPT chat
* Google Drive (storing college work)
* Answering these questions
* Pi-hole logging DNS requests

1. Do you have a loyalty card? Make a list of what data the retailer collects about you each time you visit them. How do you feel about this method of data collection

* Items purchased (amount, price)
* Transaction cost
* Store location
* Number of times the card has been used
* Potential discounts used

1. Knowledge is challenging to store but is often considered the most valuable source of competitive advantage. How would you suggest that an organization you are familiar with stores and captures the “knowledge” inside the heads of experienced employees?