Fundamentals

① Data – Data is generally something that describes the entity about which it is collected in the first place. Concept of data is so abstract that it is not informative to the user. Only when data is analyzed, the analysts learn something about the entity for which data has been collected, and it becomes information about the entity. Further principles based on the information yield knowledge about the entity.

In context of computers, data is information possessed or stored by a computer, ex. documents, images, audio clips, software programs etc. A computer represents data (both structured like tabular data and unstructured data like images, sounds) in form of bits. This binary representation allows storage, manipulation, sharing of such data.

Data, in general, is different from information and knowledge. For instance, numbers about age and height of people is data. The information that height increases with age, then stops increasing after a certain point can be gained by analysing those numbers. And knowledge would be linking this information to information about growth hormones to better understand Nature

② Science – is the systematic process of building information from observations around us. It involves constructing a hypothesis to explain an observation, testing hypothesis on known results, and ~~impro~~ improving it until it predicts correctly unknown phenomena. For instance, natural sciences seek to gain knowledge

about Nature in general. This process involves making observations and thinking of hypothesis that might generalize those observations. A systematic approach of developing experiments to confirm those hypothesis is applied, and theories are developed. These theories are then used to predict future phenomena. This is the essence of science.

---

③ Computers – Most generally, a computer has five components – input, output, memory, arithmetic operations, and communication between components. Any electronic device that takes raw data as input from the users and processes the raw data under a set of instructions to produce some output is a computer. A computer may even store the outputs to prevent repeated processing of the inputs.

The ability to process information in the desired way (ability to get programmed) is what separates a computer from other electronic devices. In modern times, most of the electronic devices surrounding an average individual have some parts, at least, working as computers. By this definition, smart phones, calculators, smart watches, modern washing machines etc. are all computers (taking in user input and directing the hardware based on some programmed computation performed).

It is common to refer to desktop workstations as 'computers'. This is however a misnomer, as desktop workstations are only a 'type' of computer and not 'the' computer.

④ Data Science – Data science is the merger of data and science. It involves the scientific process to acquire knowledge about an entity through analysis of acquired data. Data science involves the study of various types of data such as structured, semi-structured, and unstructured (where structure normally refers to a tabular arrangement of data). Multidisciplinary approaches such as data mining, feature engineering, machine learning, deep learning etc. (all involving significant statistics, modelling, and model training) form a part of this collective process to understand data in its entirety.

Data science reflects science at its very core (hence the name). For instance, we might want to look at patterns in stock market prices. We collect numbers about past prices (this is the unorganised, unanalysed data) and develop some hypothesis that might explain its variation over time (this is science). We might come to accept or reject hypothesis based on experiments (this is science), and if found acceptable, these hypothesis help predict future prices (something scientific hypothesis/theories do).

⑤ Computer Science – Merger of computers and science, implying the scientific process of studying computers' functional aspects. Computers have

various aspects attached to them - computing power, memory, clock rates, number of transistors on a chip etc; and performance boost (response time and CPU execution time) is always desirable. So a scientific process of investigation in computer architecture, design, software, programming languages, compilers etc. is needed to achieve optimal performance. Study of all these subfields comes under the discipline of computer science.