

Big Data definition



Open your mind. LUT.
Lappeenranta University of Technology

Big data can be defined in terms on 3+2V's: **Volume, Variety, Velocity + Value, Veracity**; it is characterised by uncertainty and has certain present value as well as unexpected future values and possible uses. We can further clarify it by explaining our understanding of V's and value.

1. **high volume** - massive amount of data and metadata, terabytes or even petabytes, using “all” data
2. **high variety** - unstructured data, secondary use of data, many sources of data
3. **high velocity** - speed with which data is generated and used. Batch, real time, near time, streams.
4. **high value** - primary data has latent value, more value through secondary use of data, information unveiled can be of great value for companies. Companies with big data sets can also price and sell them.
5. **high veracity** - messiness of the data, data can have errors but errors usually average out in big data set

Big data can be used as a source of innovation by uncovering data patterns that can be combined to facilitate new knowledge. One of the most important functions of Big Data technologies is increasing business intelligence. Currently, big data can be used to find correlation, but not causation. The value of big data encourages more to datafication, collecting data from every possible source that is ready to be analyzed.

Known challenges for Big Data include capture, curation, storage, search, sharing, transfer, analysis, and visualization.