

## Big data – Definition

The term big data means “masses of data”. It refers to data volumes that are so complex that they require new computer-based forms of processing. This is the opposite of “small data”, which means data volumes that a person can understand and process without help or with the aid of a simple computer program, or data volumes that refer to only one individual person or a narrowly restricted set of facts. Big data is also an umbrella term encompassing various activities associated with the enormous masses of data.

### Big data means...

1. gathering a great deal of different data. Numerous sensors and programs capture and process information. These sensors and programs are built into or installed on tablets, smartphones, or computers, for example. If the collected data are analog, they are digitalized.
2. saving the digitalized information as data in as structured a manner as possible. This results in complex databases, which are used to process the masses of data using computers.
3. linking the data and identifying new correlations by means of data analysis. Thanks to the structuring, it is possible to make targeted queries in databases and compare very specific aspects of databases. Often, assumptions or theories are confirmed or disproved based on the data analysis.

## The three Vs and their meaning

Experts frequently describe big data with the three Vs, which are intended primarily to specify the “big” in the name in more detail and define the term more precisely. These three terms begin with the letter V.

### Volume

This is the amount of data that is processed. With modern methods it is possible to manage enormous volumes of data. A vast number of computers or – expressed more neutrally – nodes is necessary for this purpose. This is the only way the masses of data can be processed in a structured manner to ultimately yield useful results.

### Velocity

This is the speed at which the large volumes of data are processed. Often, a big-data method is beneficial only if the result is output sufficiently quickly.

### Variety

This refers to the different types of data. They are compared or linked with each other. They must originate from many different sources, for example, be captured at different times, or be generated at different locations around the world. Only then do they have broad validity.