

Big Data – Begriffsdefinition

Der Begriff Big Data kommt aus dem Englischen und lässt sich mit „Datenmassen“ übersetzen. Er bezeichnet Datenmengen, die so komplex sind, dass sie neue computergestützte Formen der Verarbeitung benötigen. Dem stehen „Small Data“ gegenüber, womit Datenmengen gemeint sind, die ein Mensch alleine oder mithilfe eines einfachen Computerprogramms verstehen und verarbeiten kann, oder die sich nur auf eine einzelne Person oder einen eng begrenzten Sachverhalt beziehen. Big Data ist auch ein Oberbegriff, der unterschiedliche Tätigkeiten zusammenfasst, die mit den enormen Datenmassen in Verbindung stehen.

Big Data bedeutet,

1. sehr viele unterschiedliche Daten zu sammeln. Dafür erfassen und verarbeiten zahlreiche Sensoren und Programme Informationen. Diese sind z. B. in Tablets, Smartphones oder Computern verbaut bzw. installiert. Liegen die gesammelten Daten analog vor, werden sie digitalisiert.
2. die digitalisierten Informationen als Daten möglichst strukturiert zu speichern. So entstehen komplexe Datenbanken. Sie werden gebraucht, um die Datenmassen computergestützt zu verarbeiten.
3. die Daten zu verknüpfen und über die Datenauswertung neue Zusammenhänge zu erkennen. Durch die Strukturierung ist es möglich, Datenbestände gezielt zu befragen und auf ganz bestimmte Aspekte hin miteinander zu vergleichen. Oft sind es Annahmen oder Theorien, die anhand der Datenauswertung bestätigt oder widerlegt werden sollen.

Die drei Vs und ihre Bedeutung

Fachleute beschreiben Big Data häufig mit den sogenannten drei Vs, die vor allem das „Big“ im Namen näher bestimmen und den Begriff präzisieren sollen. Dabei handelt es sich um drei Begriffe, die im Englischen alle mit dem Buchstaben V beginnen.

Volume

steht für die Datenmenge, die verarbeitet wird. Mit modernen Methoden ist es möglich, sehr große Datenmengen zu bewältigen. Dafür ist eine Vielzahl von Computern oder – neutraler ausgedrückt – von Knoten notwendig. Nur so können die Datenmassen strukturiert verarbeitet werden, um schließlich sinnvolle Ergebnisse auszugeben.

Velocity

steht für die Geschwindigkeit, mit der die großen Mengen an Daten verarbeitet werden. Oft ist ein Big-Data-Verfahren nur dann von Nutzen, wenn das Ergebnis ausreichend schnell ausgegeben wird.

Variety

steht für die Vielfalt der Daten. Sie werden miteinander verglichen oder verknüpft. Sie müssen z. B. aus vielen unterschiedlichen Quellen stammen, zu unterschiedlichen Zeitpunkten erfasst werden oder an unterschiedlichen Orten auf der Welt entstanden sein. Nur dann besitzen sie eine breite Gültigkeit.