

Wie führt man Daten sinnvoll zusammen? Big-Data-Infrastrukturen unterscheiden sich im Aufbau von konventionellen Business-Intelligence-Systemen durch die Art und Weise, wie Daten zusammengeführt werden. Die grosse Datenmenge und die hohe Geschwindigkeit bei der Erzeugung neuer Daten bedingen spezielle Technologien und Ansätze, die es im Einzelfall auszuwählen und zu bewerten gilt. Die sogenannte «Lambda-Architektur» zum Beispiel sieht aus diesem Grund eine Trennung in der jeweiligen Vorverarbeitung vor: einerseits in eine skalierbare Batchverarbeitung und andererseits in eine ebenso skalierbare Echtzeitverarbeitung. Die Resultate beider Verarbeitungen werden synchronisiert und zusammengeführt. Sind die Daten einmal zusammengeführt – es wird auch von einer Datenfusion gesprochen –, so steht eine Vielzahl neuartiger Analyse-, Mining-, Statistik- und Visualisierungstechniken zur Verfügung. Sie alle werden unter dem Begriff Data Science zusammengefasst und umfassen zusätzliche Aspekte wie Speicher- und Datenbanktechnologien, erweiterte mathematische Grundlagen, Engineering-Prinzipien und Sicherheitsaspekte.

Sinnige und unsinnige Korrelationen. Auch wenn die Mehrheit aller Unternehmen davon ausgeht, dass mehr Informationen auch mehr Produktivität bedeutet, heisst das noch lange nicht, dass jede neue Korrelation auch neue unternehmensrelevante Weisheiten offenbart. Im Gegenteil, insbesondere die Kombination interner und externer Daten erfordert grösste Sorgfalt bei der Interpretation. Der Satz von Data-Mining-Pionier Nicolas Bissantz «Daten bleiben scheu und grausam. Es ist schwierig, aus ihnen zu lernen, und noch schwieriger, das Gelernte umzusetzen» gilt beim Umgang mit Big Data in besonderer Masse, da nicht aus jeder technisch ableitbaren Wechselbeziehung auch eine wirtschaftlich relevante Ursache abgeleitet werden kann.

Wie lernt man Big Data? Big-Data-, Big-Data-Analytistics- oder Data-Science-Kurse werden an vielen Hochschulen angeboten. Beispielsweise als Zertifikats- oder Diplomlehrgänge der Zürcher Hochschule für Angewandte Wissenschaften, der Hochschule für Wirtschaft Zürich, der Hochschule Luzern oder der Berner Fachhochschule. Es stehen jedoch auch Online-Lehrgänge oder Kurse von privaten Anbietern zur Verfügung. Der typische Aufbau eines Big-Data-Kurses umfasst die Themen Speicherarten, deren Einsatzgebiete und Auswahl, Verarbeitungsmöglichkeiten sowie die neuen Analysetechniken und mögliche Umsetzungsszenarien. Ein normaler Data-Science-Kurs umfasst Aspekte der Manipulation grosser Datenmengen sowie deren Analyse und Resultataufbereitung. Der Bedarf an Fachkräften in diesem Bereich ist nach Aussagen von Analysten bereits heute sehr gross und wird in Zukunft stark ansteigen.

Datenqualität. Die Zunahme von Daten bedeutet auch immer eine mögliche Zunahme von fehlerhaften Daten. Unternehmensrelevante Entscheide auf Basis falscher Fakten zu treffen, hatte in vielen Fällen bereits verheerende Konsequenzen. Gemäss einer Studie des Marktforschungsunternehmens

Kelton Research gaben mehr als 45% der über 500 befragten Unternehmen an, bereits Fehlentscheide aufgrund falscher Daten gefällt zu haben. Moderne Datenanalysen können zwar mit fehlerhaften oder unvollständigen Daten umgehen, sie sind jedoch auf belastbare Qualitätsaussagen angewiesen. Datenqualitätsaspekte sind aus diesem Grund spätestens bei der Einführung einer Big-Data-Lösung unternehmensweit aufzuarbeiten. Dabei ist immer zu definieren, welche Informationen für eine Firma wichtig und damit wertvoll sind. Daraus wird der Qualitätsanspruch an die entsprechenden Daten abgeleitet. Und es sind Verantwortliche zu definieren, die die fachliche und technische Qualität beurteilen und verbessern helfen.

Echtzeitanalyse. Die Echtzeitanalyse oder auch die möglichst zeitnahe Aufbereitung von Daten spielt eine zunehmend grössere Rolle. Zum Beispiel können durch die laufende Überwachung von Börsentransaktionen in Sekundenbruchteilen Risiken für Anleger wie beispielsweise unsere Pensionskassen minimiert werden. Darüber hinaus ergibt die Möglichkeit, möglichst schnell die richtigen Informationen richtig aufbereitet und analysiert am richtigen Ort zur Verfügung zu haben, eine Vielzahl von Anwendungsmöglichkeiten. Sie reichen von Vermeidung der Verschwendung natürlicher Ressourcen über optimale Einsatzplanungen für Güter, Fahrzeuge und Personal bis hin zu schnellen nachfrageorientierten Produktionsanpassungen. Gerade für hoch spezialisierte Unternehmen, die integrierter Bestandteil einer grösseren Wertschöpfungskette sind, eröffnet Big Data eine Vielzahl von höchst interessanten neuen Geschäftsmöglichkeiten.

Zu hoher Komplexitätsgrad? Die zunehmend vernetzte Welt bietet einen immer umfassenderen Zugriff auf eine immer grössere Menge an Daten. Was also tatsächlich passiert, wird immer genauer dokumentiert. Daraus können wir bessere Informationen gewinnen. Informationen, die bei der Steuerung eines Unternehmens wertvolle Dienste leisten können. Allerdings bedingt dies den klugen Einsatz innovativer Technologien und Methoden. Wie jede unternehmerische Tätigkeit folgt auch diese einem Zyklus der ständigen Verbesserung. Nur so wird aus moderner Alchemie alltagsfähiges Werkzeug.



DANIEL LIEBHART

ist Dozent für Informatik an der ZHAW (Zürcher Hochschule für Angewandte Wissenschaften), Experte für Enterprise-Architekturen und Solution Manager der Trivadis AG. Er ist Autor und Co-Autor verschiedener Fachbücher.