

Big-Data-Mythen

Auch Unternehmen, die keine enormen Datenmengen haben und nicht über spezialisierte Datenwissenschaftler und ein riesiges Budget verfügen, können erfolgreich in Big-Data einsteigen. Wenn man die Wahrheit hinter den drei großen Big-Data-Mythen analysiert, kann man auch klein einsteigen und groß herauskommen.

Die Vorteile von Big-Data sind unbestritten: Neukundengewinnung und Kundenservice lassen sich effizient und erfolgreich gestalten, Geschäftsprozesse optimieren und völlig neue Märkte erschließen. Kurz: Big-Data verwandelt den Überfluss an Informationen in zusätzlichen Gewinn.

Viele Entscheider meinen aber, dass ihr Unternehmen noch nicht in Big-Data einsteigen kann. Drei häufig vorgebrachte Argumente sind: Wir haben gar nicht so viele Daten, wir haben kein Budget und uns fehlen Datenwissenschaftler. Diese drei Einwände erweisen sich jedoch als Mythen, wenn man die ausgereifte Technologie des Ökosystems „Open-Source Hadoop“ in Betracht zieht.

Mythos Nr. 1: Bei uns gibt es gar keine Big-Data

Viele Unternehmen haben Schwierigkeiten mit dem Daten-Management, selbst wenn der Datenbestand nur etwa 10 bis 50 Terabyte umfasst. Das Problem kann darin bestehen, dass vorhandene Datenbanken an ihre Kapazitätsgrenzen stoßen oder ältere Daten in einem dunklen Archiv schlummern, wo sie nicht abgefragt oder ausgewertet werden können. Viele Unternehmen schöpfen das Potenzial an Daten nicht aus, die sie am Point of Sale, bei Online-Transaktionen oder durch Sensoren an Maschinen generieren. Möglicherweise werden potenziell wertvolle Daten gar nicht gesammelt oder zu früh gelöscht.

Zweitens verfügt ein durchschnittliches Fortune-1000-Unternehmen zwar über einen Gesamtdatenbestand in Petabyte-Höhe, doch diese Daten werden in getrennten Silos aufbewahrt und nicht gesamtheitlich professionell bewirtschaftet.

Drittens, und dies ist der wichtigste Punkt: Viele Unternehmen sammeln beziehungsweise unterhalten allein deshalb keine großen Datenmengen, weil schon die zugrundeliegende IT-Architektur zu teuer ist. Die Anschaffungskosten eines herkömmlichen Enterprise-Data-Warehouses

(EDW) belaufen sich auf 20.000 bis 80.000 Dollar pro Terabyte. Das bedeutet: Das Budget des Unternehmens ist schnell erschöpft.

Mythos Nr. 2: Big-Data erfordert ein großes Budget

Kern und Ursache dieses Mythos bilden die hochskalierten In-Memory-Lösungen von Anbietern wie Oracle oder SAP. Mit Marketingbudgets in zweistelliger Millionenhöhe wurde vielen Kunden die SAP-Lösung HANA als erste und einzige Wahl für Big-Data angepriesen. Tatsächlich ist HANA jedoch eine teure aber zielgerichtete Lösung dort, wo die Geschwindigkeit der Datenverarbeitung zur Entscheidungsfindung in Echtzeit im Vordergrund steht.

HANA ist pro Terabyte etwa 10.000-mal teurer als Hadoop. Daher hat SAP Hadoop mittlerweile mit HANA kombiniert, um die Investition im Bündel attraktiver zu machen. Denn Hadoop punktet auch in Situationen,



lösung liegen im niedrigen sechsstelligen Bereich. Wer 1 Million Euro investiert, kann ein Cluster in Petabyte-Größe erwerben.

Mythos Nr. 3: Wir haben keine Datenwissenschaftler

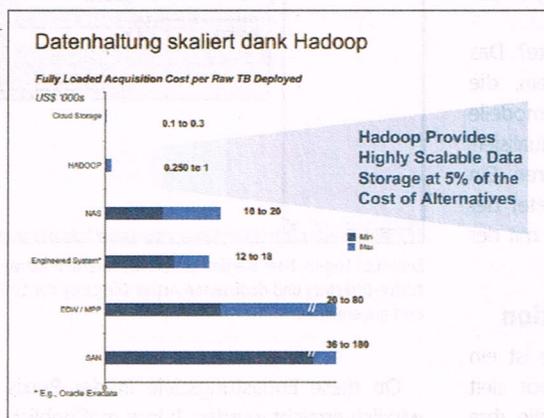
Datenwissenschaftler sind wie weiße Einhörner: seltene Wesen mit sagenumwobenen Kräften. Ein Datenwissenschaftler verbindet die Fähigkeiten und Erfahrungen eines Software-Entwicklers mit denen eines anwendungsbezogenen Wissenschaftlers. Unternehmen, die auf Big-Data verzichten, weil ihnen die erfahrenen Spezialisten fehlen, zäumen jedoch das Pferd von hinten auf.

Die meisten Unternehmen setzen Hadoop zunächst als günstige, hoch skalierende Datenquelle für existierende Business-Intelligence-Tools ein. So können etablierte Fragen, ausgedrückt als Structured-Query-Language (SQL), auf größeren Datenbeständen laufen.

Einige fangen im zweiten Schritt an, neue analytische Anwendungen für Hadoop zu entwickeln, meist um aus neuen, unstrukturierten Datenquellen Wert zu schöpfen. Erst wenn das Unternehmen beträchtliche Kosten eingespart und Erfahrungen mit Hadoop gesammelt hat, ist es sinnvoll so genannte Data-Scientists einzustellen. Dieser neue Beruf hat zur Aufgabe, fortgeschrittene Auswertungen mittels statistischer Algorithmen zu erstellen.

Unternehmen, die Big-Data schrittweise einführen, stellen fest, dass sich die Investition selbst trägt. Um nur ein Beispiel aus der Telekommunikations-Branche anzuführen: Wenn die jährlichen Investitionskosten für den Ausbau des 4G-Mobilfunknetzes mehr als eine Milliarde Euro betragen, und ein einzelner Datenwissenschaftler in der Lage ist, diese Investitionen um nur einen Prozentpunkt zu senken, dann hat er in diesem Jahr bereits mehr Geld eingespart, als seine eigene Stelle über 20 Jahre hinweg kostet.

Grafik: TechAlpha



Die Anschaffungskosten verschiedener Datenmanagement-Lösungen pro Terabyte

wo Datenvolumen, Datenvielfalt sowie die explorative Suche nach Zusammenhängen in den Daten im Vordergrund stehen.

Mit einer Hadoop-Lösung können Unternehmen jeder Größe eine preisgünstige Landebasis für ihre gesamten Daten anlegen und für zunehmende Datenmengen skalieren. Zum Einstieg reicht bereits ein Budget von 20.000 Euro, die Kosten für eine Groß-