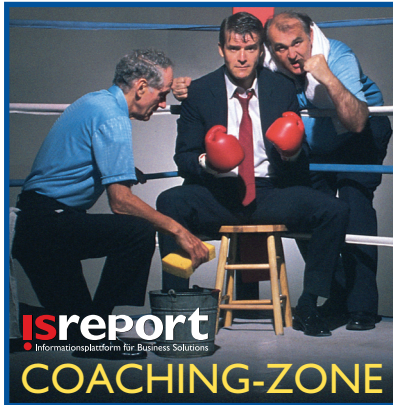


Big Data ohne Big Budget realisieren

Schnelle, skalierbare, fast grenzenlose Datenanalyse ist kein Privileg der großen Unternehmen. Wer aus riesigen Datenbeständen einen Geschäftswert erzielen will, muss sich mit vier Schlüsseltechnologien auseinander setzen.



DAUERHAFTES Datenwachstum hat zwei maßgebliche Auswirkungen auf Unternehmen. Zum einen erhöht sich der Druck auf die bestehende IT Infrastruktur sowie deren Budgets. Zum anderen bieten die sich auftürmenden Datenmengen die Möglichkeit, neues Geschäft zu generieren. Folgende vier Technologien haben einen hohen Nutzen im Kontext von Big Data:

- **Integration**

In den meisten Fällen verteilen sich Daten über verschiedenen Silos wie Datenbanken verbunden mit Enterprise Resource Planning (ERP) oder Customer Relationshipmanagement (CRM) Unternehmenslösungen oder lokale Excel-files. Oft werden Datenquellen außerhalb der Unternehmensgrenzen genutzt, welche das Internet generiert. Hierzu gehören Social Networks, Blogs und weitere Plattformen zur Meinungsäußerung. Die damit einhergehende Komplexitätserhöhung macht es unerlässlich, Daten von einer Quelle in die andere zu bewegen. ETL (Extract, Transform, Load) kann hierbei als Prozess genutzt werden, bei dem Daten aus mehreren gegebenenfalls unterschiedlich strukturierten Datenquellen in einer Zieldatenbank vereinigt werden.

- **In-Memory**

Anwender wollen immer schneller mit Daten beliefert werden, um Ihre Geschäftsanalysen durchzuführen und

qualitativ hochwertige Informationen zu erhalten. Somit ist die zeitnahe Verfügbarkeit von Daten ein Wettbewerbsvorteil, da hierdurch richtungsweisende Ableitungen und Maßnahmen getroffen werden können. Eine der effektivsten und schnellsten Technologien, um Daten zu prozessieren, ist In-Memory. Eine solche Lösung lädt die Zieldaten direkt in den Arbeitsspeicher (Random Access Memory) eines Servers oder Desktop Computers. Damit unterscheidet sich das In-Memory System von herkömmlichen Systemen, die dazu Festplattenlaufwerke verwenden. Die Zugriffsgeschwindigkeiten sind höher und die Algorithmen für den Zugriff sind einfacher. Unternehmen, die diese Technologie einsetzen, müssen zwar größere Datenvolumen betrachten, aber können auch einen viel höheren Anteil der eigenen Daten analysieren und bewerten. Die Wahrscheinlichkeit, eine Fragestellung richtig zu beantworten, stellt sich als viel größer dar.

- **Struktur**

Neben strukturierten Daten aus Datenbanken ist die große Stärke des Big Data Ansatzes eine Nutzung von unstrukturierten und semi-strukturierten Daten aus Dokumenten, Social Networks, E-Mails, News Feeds, Bildern, Videos oder anderen Rich-Media-Inhalten. Mit Apache Hadoop als open-source (Java) Framework lässt sich hier eine skalierbare, verteilt arbeitende Software zum Datenmanagement nutzen. Diese basiert auf dem bekannten MapReduce-Algorithmus von Google und ermöglicht intensive Rechenprozesse mit großen Datenmengen auf Computerclustern. Der frei verfügbare Code, der auf Standardhardware läuft, belastet entsprechend wenig das Budget.

- **Visualisierung**

Sobald eine wertvolle Information aus den Datenbeständen abgeleitet oder identifiziert wurde, muss diese leicht

zugänglich anderen Nutzergruppen zur Verfügung gestellt werden. Dies bedarf im besten Falle einer eingängigen Visualisierung in Form von Reports oder Dashboards. Eine solche Visualisierung sollte sich automatisch updaten, wenn der Nutzer Parameter ändert oder Drill-Downs auf zugrunde liegende Daten macht. Ein deutlich ablesbarer Trend im Business Intelligence Umfeld geht weg von statischen Reporting aus den Händen der IT und hin zu mehr Selbstständigkeit des Nutzers von Daten. Dies wird im Big Data Umfeld sogar noch verstärkt, da die entstehende Komplexität den IT Ressourcen alles abverlangt und das Management von Business Intelligence durch die IT mehr und mehr unattraktiv wird. Lösungen, die es den Endanwendern direkt ermöglichen, Daten zu sondieren erhöhen die Wahrscheinlichkeit, die Fragen zu beantworten, die zu Wettbewerbsvorteilen führen.

Zusammenfassend kann gesagt werden, dass Big Data Initiativen nicht nur den global agierenden, großen Unternehmen vorbehalten ist. Auch schon mit schmalen Budgets unter zu Hilfenahme der oben beschriebenen Technologien können die Schätze gehoben werden, die für das Unternehmen wertvoll sind. Heute verfügbare Datenmanagement-Systeme vereinen häufig die vier aufgeführten Technologien. ◀hei

Der Coach



Andreas Auer
ist Leiter Vertrieb bei Oraylis