

Es entstehen immer größere Datenmengen, die aus immer unterschiedlicheren Formaten und aus immer mehr Datenquellen gespeist werden. Die Erkenntnisse, die aus diesen Daten gewonnen werden können, sind das Gold des Informationszeitalters.

# Aktuelle Trends bei Business Intelligence und Data Warehouse

Klaus Rohrmoser, data2fact

Soziale Medien wie Facebook oder Google stellen nicht umsonst kostenlos ihre Dienste zur Verfügung. Daten intelligent zu verwenden, wird künftig immer stärker im Fokus stehen. Hierbei gibt es verschiedenste Methoden und Technologien, die unter den Begriffen „Business Intelligence“ (BI) und „Data Warehouse“ zusammengefasst werden können. Dieser Artikel gibt einen Überblick über aktuelle Trends.

Business Intelligence und Data Warehouse werden oft als Synonym verwendet, sie unterscheiden sich jedoch grundsätzlich. Während der Begriff „Data Warehouse“ Technologien zur optimierten Datenspeicherung umfasst, ist Business Intelligence als ein Prozess zu verstehen, der relevante Informationen aus Daten gewinnt und diese an operative Systeme zurückspielen kann (Closed Loop). Business Intelligence kann IT-Systemen und -Anwendern ein Lernen aus verfügbaren Daten ermöglichen, um effizientere Entscheidungen zu treffen.

## Self Service Business Intelligence

In den meisten Unternehmen ist die IT an vorgegebene Release-Zyklen, einzuhaltende SLAs und Kostenlimits gebunden. Fachanwender möchten jedoch eine schnelle und dynamische Umsetzung ihrer Anforderungen, um ihr Geschäft zu betreiben. Dieser Widerspruch lässt sich durch Self Service BI auflösen, sofern einige wichtige Aspekte beachtet werden.

Mit Self Service BI kann mehr Agilität in der Analyse und Auswertung von Unternehmensdaten generiert werden. Reporting-Anwender können Anforderungen mit Ad-hoc-Analysen und Reports schnell und flexibel mit den vorhandenen Reporting-System(en) umsetzen. Damit sind Unternehmen in der Lage, schneller auf Anforderungen der Kunden und des Geschäftsumfelds zu reagieren. Die wichtigsten Aspekte von Self Service BI sind:

rungen mit Ad-hoc-Analysen und Reports schnell und flexibel mit den vorhandenen Reporting-System(en) umsetzen. Damit sind Unternehmen in der Lage, schneller auf Anforderungen der Kunden und des Geschäftsumfelds zu reagieren. Die wichtigsten Aspekte von Self Service BI sind:

- *Benutzerfreundlichkeit*  
Einfache Bedienbarkeit des Reporting-Systems bei der Erstellung von Ad-hoc-Reports oder Analysen. Möglichkeit zur Zusammenarbeit zwischen Benutzern (Collaboration). Integration von eigenen, oft Excel-basierten Auswertungen in ein bestehendes Dashboard.
- *Rollenverständnis*  
Fachanwender lieben Excel und können damit ausgefeilte Auswertungen erzeugen. Daten-Analysten sind SQL-affin, verstehen Datenmodelle und können komplexe Analysen erstellen. Die IT muss eine Infrastruktur finden, um beiden Rollen einen Mehrwert bieten zu können.
- *Data Governance*  
Qualität, Herkunft und Aktualität der Daten sowie Kennzahl-Definitionen sind entscheidende Aspekte, damit bei der Anwendung von Self Service BI keine Äpfel mit Birnen verglichen werden, sondern verlässliche und nachweisbare Informationen entstehen.
- *Agilität und Business Intelligence*  
Ein ständiger Austausch zwischen Fachanwendern, Daten-Analysten und IT, um mithilfe der gefundenen Erkenntnissen IT-Systeme weiterzuentwickeln und/oder Geschäftsprozesse zu verbessern.

- *Gemeinsame Datenbasis*  
Eine über mehrere Quellen integrierte Datenbasis für Reporting Stakeholder, die in einem Data Warehouse vorhanden sein kann. Flexibel auswertbar und um zusätzliche (etwa Fachbereich-bezogene) Datenquellen erweiterbar, unter Einhaltung der Data Governance.
- *Sandboxing*  
Durch die einmalige Bereitstellung von produktiven Daten zur Daten- und Anforderungs-Analyse können Erkenntnisse sowohl für das operative Geschäft als auch für künftige Anforderungen gewonnen werden.

Nicht alle Anwender wollen SQL verstehen, sondern vielmehr schnell Antworten auf ihre Fragen erhalten. Zum Erfüllen ihrer Kernaufgaben im Unternehmen benötigen sie Informationen. Um Daten lesen und verarbeiten zu können, braucht man technisches Wissen über SQL und Datenmodelle, erst dann entstehen wertvolle Analysen. Rollenverständnis und Data Governance sind zwei wesentliche Aspekte, um erfolgreich Self Service BI im Unternehmen anzuwenden.

Die IT sollte sich darauf konzentrieren, die Schnelligkeit und Qualität ihres Lösungsportfolios stetig zu verbessern, und die Fachseite sollte Data Governance als wichtigen Bestandteil von Business Intelligence anerkennen. Gemeinsam kann den Anwendern dadurch ermöglicht werden, ihre Kernaufgaben im Unternehmen wahrzunehmen.

## Trends bei Datenbank-Technologien

Aufgrund der steigenden Datenmenge und der höheren Ansprüche an Anzahl

und Dauer von Zugriffen entwickeln sich neue Datenbank-Technologien, die als Alternative zu relationalen Datenbanken eingesetzt werden können. Wichtig ist, die Eigenschaften dieser neuen Technologien im Kontext des Gesamtsystems, dessen Bestandteil die Datenbank ist, zu verstehen.

**In-Memory-Datenbanken** (IMDB) halten Daten primär im Hauptspeicher eines Rechners, um dadurch schnellere Zugriffszeiten auf die gespeicherten Daten zu ermöglichen, da ein Hauptspeicher in der Regel wesentlich höhere Zugriffsgeschwindigkeiten und effizientere Zugriffsalgorithmen als eine Festplatte vorweisen kann. Jedoch gehen bei einem Systemausfall Daten im Hauptspeicher verloren. Ein wesentliches Merkmal von Datenbanken ist Transaktions-Konsistenz, die durch Persistenz der Daten erreicht wird. IMDBs stellen dazu folgende Methoden bereit:

- Zustände der Daten werden in Zeit-Intervallen erfasst und auf persistente Speichermedien geschrieben (Snapshots). Daten, die zwischen diesen Zeit-Intervallen anfallen, können verloren gehen.
- Transaktions-Protokolle werden aus-gelesen und auf persistente Speicher-medien geschrieben (Replikation):
  - Bei einer asynchronen Replikation sind Transaktion und Persistenz getrennt, was eine hohe Performanz und auch ein hohes Risiko des Datenverlusts mit sich bringt.
  - Bei einer synchronen Replikation sind Transaktion und Persistenz zusammengeführt, was bei Schreibzugriffen zu längeren Laufzeiten führen kann, dafür hohe Datensicherheit durch eine ACID-konforme Transaktionssteuerung sicherstellt (siehe <http://de.wikipedia.org/wiki/ACID>).
- Der Einsatz von NVRAM-Speicher („Non-Volatile Random-Access Memory“, Hauptspeichermodule mit eigener Stromversorgung), der bei Systemausfällen den letzten Daten-zustand wiederherstellen kann.
- Hybride Ansätze, die die oben ge-nannten Methoden kombinieren, um damit ACID und Hochverfüg-

barkeit bei gleichzeitig schnellen Datenzugriffen zu erreichen.

**NoSQL-Datenbanken** sind strukturierte Datenspeicher, die keine relationalen Algorithmen wie Datenmodelle in der dritten Normalform verfolgen und teilweise auch auf SQL oder ACID verzichten. Sie skalieren horizontal und arbeiten verteilt, damit werden Ziele wie schnelle Zugriffe, Hochverfügbarkeit oder niedrige Hardwarekosten erreicht. Folgende Ansätze werden als NoSQL-Datenbanken bezeichnet:

- „Key-Value-Paare“ sind einfache Lookup-Strukturen, wobei die Schlüssel je nach Datenbank-Hersteller auch gruppiert oder erweitert werden. Diese können In-Memory oder auf Festplatte gespeichert sein.
- „Columnar“ speichern Daten spaltenorientiert und besitzen die Eigenschaften von NoSQL. Dies unterscheidet diesen Ansatz von den relational spaltenorientierten Datenbanken.
- „Document“ verwenden auch den Key-Value-Ansatz, wobei der Wert ein Dokument darstellt.
- „Graph“ modelliert die Verbindungen zwischen Daten-Objekten, die wiederum als Key-Value-Paare dargestellt sind.

NoSQL-Datenbanken sind für einfache, schnell zugängliche Datenstrukturen bei gleichzeitig hohen Datenvolumen und einer hohen Anzahl an Zugriffen optimiert. Es gibt wenige Restriktionen bei Datenmodellen, Änderungen sind damit auch bei großen Datenmengen leicht umsetzbar.

In **relational spaltenorientierte Datenbanken** sind die Daten spaltenorientiert in Blöcke geschrieben, mit dem Ziel, weniger I/O zu generieren und damit einen schnelleren Zugriff auf Daten zu erreichen. Konventionelle Datenbanken schreiben Daten zeilenorientiert in Blöcke.

Beim Auslesen der Daten werden in spaltenorientierten Datenbanken nur die Spalten gelesen, die in der Query enthalten sind, bei den zeilenorientierten Datenbanken werden hingegen Zeilen – und damit alle Spalten

– gelesen. Dies kann bei der hohen Spaltenzahl in Star-Schemata einen entscheidenden Unterschied darstellen.

Zudem komprimieren spaltenorientierte Datenbanken besser als zeilenorientierte, da direkt auf komprimierte Daten zugegriffen wird, ohne diese über die CPU zu dekomprimieren. Durch interne Indizes wird ein schneller Zugriff bei Filtern mit „n>1“ Spalten erreicht, oft werden CPU-Algorithmen direkt von der Datenbank effizient eingesetzt. Spaltenorientierten Datenbanken generieren damit oft weniger I/O als zeilenorientierte, insbesondere bei Reporting und Analyse. Die Eigenschaften der relational spaltenorientierten Datenbanken sind:

- spaltenorientierte Speicherung
- Komprimierung der Daten
- effizienter CPU-Einsatz
- Sortierung der Daten
- hohe Zugriffsgeschwindigkeiten

Das Beispiel in Tabelle 1 verdeutlicht die Speicherung, wobei am Bonus auch die Komprimierung nachvollzogen werden kann.

Zeilenorientierte Speicherung:

Eine Zeile ist zusammenhängend in „n>0“ Blöcken gespeichert:  
1,Müller,7000; 2,Meier,3000; 3,Berg,3000

Spaltenorientierte Speicherung:

Eine Spalte ist zusammenhängend in „n>0“ Blöcke gespeichert, wobei die Daten sortiert nach Zeilen abgespeichert sind:  
1,2,3; Müller,Meier,Berg; 7000,3000,3000

Im Unterschied zu „NoSQL Columnar“ enthalten die relational spaltenorientierten Datenbanken eine wichtige Eigenschaft: Sie bilden Daten relational ab. Dies unterstützt insbesondere Auswertungen und Analysen mit konventionellen SQL-Abfragen, auch die meisten Reporting- und Analyse-Anwendungen unterstützen SQL.

**Big Data** bedeutet, große Datenmengen in unterschiedlichsten Datenformaten (Bilder, Dokumente,

Geo data etc.) ausreichend schnell zugreifbar und auswertbar zu halten. Man spricht auch von den drei „V“s: „Volume“, „Variety“ und „Velocity“. Relational spaltenorientierte Datenbanken, NoSQL-Datenbanken, „Massive Parallel Systems“ und flexible Daten-Schemata sind Bestandteile der Technologie von Big Data.

ID	Nachname	Bonus
1	Müller	7000
2	Meier	3000
3	Berg	3000

Tabelle 1: Beispiel

**Weitere Trends**

Weitere, hier nur stichwortartig angesprochene Trends sind:

- *Predictive Analytics*  
Computergestützte, statistische Vorhersagemodelle
- *Mobile BI*  
Verteilung von Analysen und Reports auf mobile Endgeräte
- *Business Intelligence 3.0.*  
Kollaboration, interaktive und leicht bedienbare Reporting-Systeme, Vorhersagemodelle, Cloud

**Fazit**

Business Intelligence wird bei zunehmender Dynamik in der Geschäftswelt und mit einer stetig steigenden Datenmenge immer wichtiger in Unternehmen. Der Artikel hat anhand der zwei Themen „Business Intelligence“ und „Data Warehouse“ einige Trends vorgestellt, mit der diese Herausforderungen gemeistert werden können.



Klaus Rohrmoser  
klaus.rohrmoser@data2fact.de

**PROLICENSE®**  
OPTIMIZING SOFTWARE ASSETS  
Kompetent – Unabhängig – Erfolgsbasiert

# ORACLE-SOFTWARE IST JEDEN CENT WERT!

Unsere Mandanten zahlen trotzdem weniger.  
Sprechen Sie mit uns!

Wir sind nur unseren Mandanten verpflichtet.

- > **Compliance sichern**
- > **Audit vermeiden**
- > **Kosten senken**

**ProLicense GmbH**  
Friedrichstraße 191 | 10117 Berlin  
Tel: +49 (0)30 60 98 19 230 | www.prolicense.com