

Business Intelligence

# Aus Daten entscheidungs- relevante Informationen gewinnen



Im globalen Wettbewerb ist seit Jahren eine Beschleunigung der Geschäftsprozesse wahrzunehmen. Die Marktbewegungen auf der Beschaffungs- und Absatzseite sind durch eine immer höhere Dynamik geprägt und erfordern eine Verkürzung der Analyse- und Entscheidungsprozesse. Sowohl die volatilen Märkte als auch die verkürzten Produktzyklen erhöhen den Handlungsdruck auf das Management, Unternehmensdaten heranzuziehen, um fundierte Informationen für anstehende Entscheidungen zu generieren.

Es gibt somit gute Gründe, Softwarelösungen zur Auswertung der Unternehmensdaten zu nutzen. Software-Anbieter haben diese Problematik erkannt und bieten mittlerweile Lösungen, um aus der Fülle von Daten entscheidungsrelevante Informationen zu gewinnen. Solche Softwarelösungen werden unter dem Begriff Business Intelligence (BI) subsummiert. BI ist keine Erfindung unserer Tage. Schon in den 70er Jahren des letzten Jahrhunderts gab es Management-Informationssysteme für ein datenunterstützendes Berichtswesen, später kamen Decision-Support-Systeme zur Unterstützung von Planungs- und Entscheidungsprozessen hinzu. Die Komplexität und Reduktion auf bestimmte Anwendergruppen verhinderte jedoch den großen kommerziellen Durchbruch.

In den 90er Jahren kamen die Begriffe Data Warehousing, Data Mining und Online Analytical Processing (OLAP) auf. Die Idee bestand darin, auch externe Datenströme durch die Nutzung des Internets in die Analyse einzubeziehen. Inzwischen wurden nicht nur die Begriffe durch Industrie 4.0, Big Data und Machine Learning erweitert, sondern auch um die dahinterliegenden Verfahren, Algorithmen und Verfügungsoptionen (Cloud versus On Premises). Die Zielsetzung hat sich ebenfalls erweitert und umfasst nun komplette Produktionsketten und Geschäftsprozesse, die einer digitalen Datenanalyse unterzogen werden – »to help enterprises to make better business decisions« [1].

Für viele CEOs ist es eine überaus faszinierende Vorstellung, das gesamte Unternehmen zentral vom Schreibtisch aus zu monitoren und steuern zu können. Genau dies ist seit fünf Jahrzehnten eines der Versprechen der IT und der betrieblichen Datenverarbeitung. Business Intelligence und Machine Learning sind die heutigen Metaphern, dieses Versprechen einzulösen.

Das bedeutet auch, dass der Begriff Business Intelligence, der primär die Informationsbereitstellung, also die Selektion, Aufbereitung und Präsentation der Daten in einer für den Anwender verständlichen Form zum Ziel hat, heute über unterschiedliche Konzepte der Datenanalyse und -modellierung realisiert wird.

Prinzipiell existieren 2 verschiedene Arten von Rohdaten für eine Analyse. Strukturierte Datenbestände, die in wie auch immer gearteten Datenbanken erfasst werden, und nicht strukturierte Daten, die mittlerweile rund 80 Prozent aller Unternehmensdaten umfassen, zum Beispiel Texte, E-Mails, Logfiles oder textbasierte Maschinendaten.

**Strukturierte Daten.** Strukturierte Datensammlungen hatten bisher den Vorteil, über relationale Verknüpfun-

gen der Daten eigene Hypothesen verifizieren beziehungsweise falsifizieren zu können, ohne dass der Anwender spezielle Kenntnisse zur Datenprogrammierung benötigt. Ein Problem hierbei stellt allerdings die Datenaggregation dar, da alle Daten zunächst normiert werden müssen, um diese dann datenbankkonform in das Datenrepository einzulesen. Ferner ist es notwendig, im Vorfeld ein Early Binding vorzunehmen. Damit ist die Definition der abzufragenden Parameter gemeint, die in der Datenbank erfasst werden sollen. Abfragen zu Parametern, die zuvor nicht im Early Binding definiert wurden, sind somit nicht möglich. Praktisch sind alle Data-Warehouse- beziehungsweise Data-Mart-Anwendungen auf strukturierte Datenbasen angewiesen und bedürfen der Extraktion, Transformation (Normalisierung) und des Loadings der Daten (ETL-Prozess)

**Unstrukturierte Daten.** Um große Mengen unstrukturierter Daten in kurzer Zeit zu durchsuchen, hat Google 2004 das sogenannte MapReduce-Programmiermodell entwickelt [2].

Inzwischen setzen eine ganze Reihe von Herstellern auf dieses NoSQL-Verfahren, darunter auch das von der Open Software Foundation Apache ins Leben gerufene Projekt »Hadoop«. Daten werden dabei nicht mehr in eine Datenbankstruktur eingelesen, sondern nach Ereignissen getrennt in große Filesysteme abgelegt, wobei jedes Ereignis einen Index erhält. Damit ist kein Early Binding mehr notwendig und es können auch Abfragen erfolgen, die zu Beginn der Datensammlung noch gar nicht auf der Tagesordnung standen. Der Anwender ist dabei allerdings mehr gefordert, da er komplette Use-Case-Szenarien zunächst erarbeiten und dann in die spezifische Suchsprache der Anwendung umsetzen muss. Allerdings schreitet die Entwicklung sehr schnell voran. Inzwischen gibt es auch hier viele vordefinierte Abfragen für spezifische KPI eines Unternehmens. Die Vorteile dieses Ansatzes liegen auf der Hand, da sowohl textbasierte Daten aus der klas-

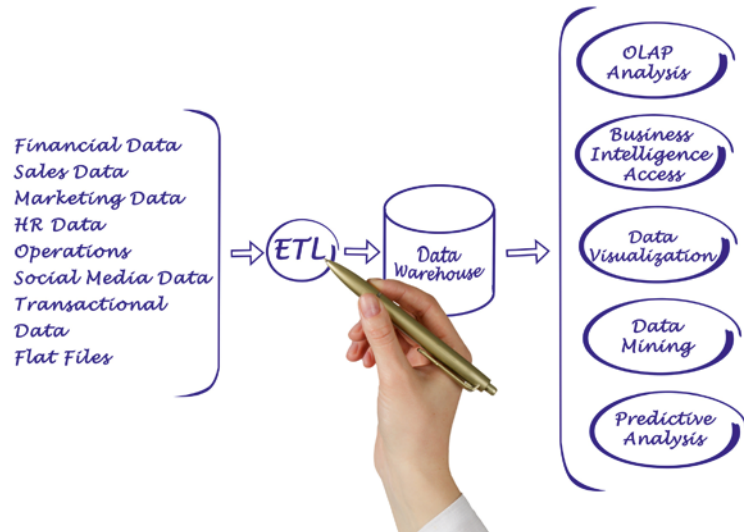
sischen IT als auch Maschinendaten aus der Produktion für eine Analyse einbezogen werden können.

Da die Datenvolumen exponentiell wachsen, die Vielfalt der Datenformate und -strukturen zunehmen und Analysen auf alle relevanten Unternehmensdaten angewendet werden sollten, stoßen rein datenbankorientierte Ansätze an ihre Grenzen. Auch hierauf hat der Markt reagiert. Die Firma Splice Machine wirbt beispielsweise damit, Datenbanksysteme mit ihrer SQL-Technik mit In-Memory- und Hadoop-Konzepten zu verknüpfen und damit die Skalierung bestehender Datenbanksysteme zu erhalten [3].

Anbieter von Analysewerkzeugen, die sich auf textbasierte, nicht strukturierte Maschinendaten, fokussiert haben, ermöglichen wiederum die Einbeziehung der klassischen Data-Mining- und Data-Mart-Konzepte und erweitern die Möglichkeiten um die

### Der ETL-Prozess

Quelle: Fotolia 2016



Data-Warehouse- und Data-Mart-Anwendungen sind auf strukturierte Datenbasen angewiesen und bedürfen der Extraktion, Transformation und des Loadings der Daten.

## » Zunehmend halten auch Machine Learning und Predictive-Analytic-Verfahren Einzug, um businessrelevante Informationen aus den Daten zu extrahieren. «

Echtzeitfähigkeiten der indexbasierten Suchalgorithmen – und beziehen somit auch Maschinendaten aus der Produktion in die Analyse ein.

**Machine Learning und Predictive Analytic.** Zunehmend halten auch Machine Learning und Predictive-Analytic-Verfahren Einzug, um businessrelevante Informationen aus den Daten zu extrahieren. Diese Verfahren nutzen sowohl strukturierte als auch spezifisch aufbereitete, unstrukturierte Daten zur Mustererkennung, unter anderem für Text- oder Transaktionsanalysen.

Machine Learning ist ein Oberbegriff, der zunächst nur die künstliche Generierung von Wissen aus Erfahrung beschreibt. Das System lernt aus Beispielen der Vergangenheit. Algorithmen, die auf Mustererkennung speziali-

siert sind, versuchen aus den zugrundeliegenden Daten, Vorhersagen für die Zukunft zu extrapolieren (Predictive Analytics). Die hohe Komplexität der »Real World Daten-Gemengelage« macht es aber notwendig, sich beispielsweise auf folgende Themenfelder zu fokussieren und hierfür spezifische Algorithmen zu entwickeln.

- || Betrugserkennung bei Transaktionen
- || Intelligente Verkehrssteuerungskonzepte
- || Logistikplanung (Problem des Handlungsreisenden)
- || Analysen zur Effektivität von Support, Marketing- und Vertriebsmaßnahmen
- || Maintenance-Sicherstellung – Vorhersagen für den Ausfall von Komponenten in der Produktion oder in der IT-Infrastruktur

Für jeden Bereich gibt es mittlerweile spezifische Toolboxes, die auch über Cloud-Dienste angeboten werden (AWS, IBM, Google, Microsoft, SAP), um die gängigsten statistischen Verfahren zur Mustererkennung zur Verfügung zu stellen:

- || Support Vector Machines
- || Principal Components Analytics
- || Regressionsverfahren
- || Komplexitätsanalyse durch Multilabel-Klassifizierung

Man darf sich aber nicht der Illusion hingeben, dass die Einführung eines dieser Werkzeuge automatisch zu einem besseren Verständnis der Geschäftsprozesse beiträgt beziehungsweise die KPIs mit einer höheren Qualität ausstattet. Die Nutzung dieser Verfahren setzt hohe fachliche Kompe-

tenz in jeder Abteilung des Unternehmens voraus, da diese die Quellen für die Daten bereitstellen müssen. Auch die Interpretation der in Dashboards aufbereiteten Informationen erfordert die Einbeziehung der Fachabteilungen, um Plausibilität und Relevanz der Informationen zu qualifizieren – besonders bei der Auswertung von Maschinendaten.

Darüber hinaus ist es wichtig, Werkzeuge zu nutzen, die für jede Hierarchieebene im Unternehmen separate Datenanalysen ermöglichen. So lässt sich jede Entscheidungsebene mit den für sie relevanten Informationen und angepassten Dashboards versorgen. Die damit einhergehenden Drill-down-Fähigkeiten erlauben es, von einer Management-Sichtweise aus schnell in tiefe, technisch spezifischere Ebenen zu gelangen, um beispielsweise eine schnelle Fehlereingrenzung bei Ausfällen geschäftskritischer Anwendungen zu gewährleisten. Rollenkonzepte, die die Active-Directory-Struktur nachbil-

den, gestalten den Zugriff auf sensible Daten so, dass das »Need-to-Know«-Prinzip nicht aufgeweicht werden muss.

Auch die Betrachtung von Cloud versus On-Premises muss im Vorfeld genau diskutiert werden. Unternehmensdaten sind kritische Daten und es ist ratsam, klar zu definieren, welche Analysen in der Cloud und welche auf jeden Fall innerhalb des eigenen Unternehmens erfolgen sollen.

**Fazit.** Unternehmen stehen bereits heute ausgereifte Business Intelligence-Technologien zur Verfügung, um fundierte Informationen für anstehende Entscheidungen zu generieren und somit Geschäftsprozesse an neue Herausforderungen anzupassen. Die zentrale Fragestellung für die Geschäftsführung lautet, auf welche Bereiche sich die Analyse zunächst konzentrieren soll. Ein zu großer Scope und nicht genau definierte Use Cases führen bei der Implementierung von

Business Intelligence schnell zu einer Überforderung der Fachabteilungen. In jedem Fall ist es empfehlenswert, für eine erfolgreiche Einführung von BI einen erfahrenen Dienstleister in das Projekt einzubeziehen. Als Systemintegrator und Managed Service Provider mit themenübergreifendem Know-how und langjähriger Erfahrung ist Controlware hier der ideale IT-Partner.

*Rainer Funk*



Rainer Funk,  
Solution Manager  
Information Security,  
Controlware GmbH  
[www.controlware.de](http://www.controlware.de)

- [1] Datawarehousing and Knowledge Discovery 7th International Conference
- [2] Google Inc. MapReduce: Simplified Data Processing on Large Clusters
- [3] Splice Machine: The First Hybrid, In-Memory RDBMS – Whitepaper