

Business Intelligence | Funktionsweise und technische Grundlagen

Whitepaper 1/5

HINTERGRUND

Die richtige Information zur richtigen Zeit abrufen zu können kann für ein Unternehmen entscheidend sein. Doch um an den PCs auf Führungsebene wirklich effektiv mit den vielen verschiedenen Daten und Formaten zu arbeiten, ist Business Intelligence nötig: ein System, das aus digitalen Daten relevante, einfach verfügbare Informa-

tionen macht. Technisch gesagt: die unternehmensweite, maßgeschneiderte Optimierung von Datenspeicherung, -aufbereitung und -bereitstellung. Wie das geht und welche technischen Grundlagen und Funktionsweisen hinter Business Intelligence Systemen stehen, das erfahren Sie auf den folgenden Seiten.

1. Architektur

Klassische BI lässt sich in einem 3-Ebenen-Modell darstellen: Datenspeicherung, Informationsgenerierung und Zugriff.

Die Abbildung stellt die generelle Architektur eines Business Intelligence Systems auf einem hohen Abstraktionslevel dar. Die genaue Funktionsweise wird in den folgenden Unterkapiteln beschrieben.

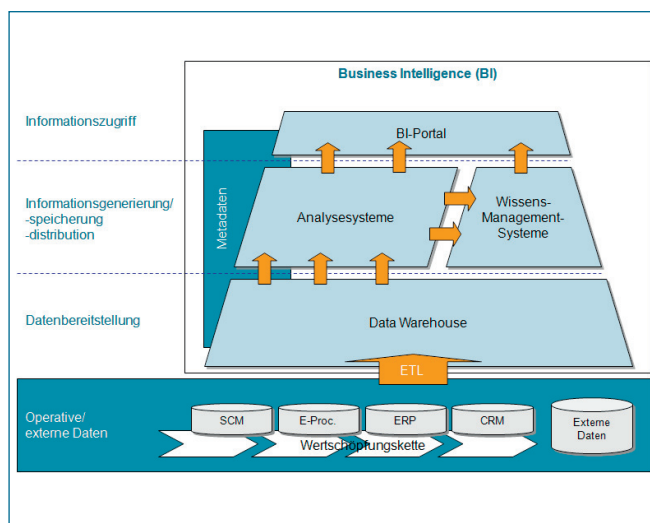


Abb. 1: Architektur von Business Intelligence [KeMU06]

1.2 Operative und externe Daten

Der Erfolg eines Unternehmens ist von der Verfügbarkeit der Informationen und dem daraus ableitbaren Wissen abhängig. Technisch gesprochen also von den Daten. Eine erfolgreiche Anwendung dieser Daten im BI-Umfeld benötigt kon-

sistente und stimmige Daten als Grundlage. [KeMU06] Dabei sind die Effizienz und die Effektivität der Datenquellen und der darin enthaltenen Information von zentraler Bedeutung. Als Basis dienen interne und externe Daten aus mehreren verschiedenen operativen und inhomogenen Vorsystemen, welche sich über die gesamte Wertschöpfungskette des Unternehmens verteilen, wie zum Beispiel Supply Chain Management (SCM), Electronic Procurement (E-Proc.), Enterprise Resource Planning (ERP) oder Customer Relationship Management (CRM), sowie Daten aus anderen externen Quellen.

Bei der Bestimmung der Datenquellen sollten folgende Faktoren bei der Auswahl beachtet werden:[Baue04]

- Zweck des Data Warehouses (DWH)
- Qualität der Quelldaten
- Verfügbarkeit (rechtlich, sozial, organisatorisch, technisch)
- Preis für den Erwerb (speziell bei externen Daten)

Grundsätzlich erfolgt die Unterscheidung in Primär- und Metadaten. Beide Datenarten existieren im Unternehmen unabhängig von ihrer Herkunft.

Bei Primärdaten handelt es sich um effektive Daten, die in den operativen Systemen erzeugt werden. Sie dienen zur Steuerung und Überwachung des Unternehmens und sollen die Unternehmensleitung bei ihrer operativen und strategischen Entscheidungsfindung unterstützen. Um dies zu ermöglichen werden die operativen und meist transaktionsorientierten Primärdaten in dispositive Daten überführt. Die typischen Merkmale operativer und dispositiver Primärdaten lassen sich aus folgender Tabelle entnehmen:

Merkmal	Operative Daten	Dispositive Daten
Ziel	Abwicklung der Geschäftsprozesse	Entscheidungsfindung und Managementunterstützung
Ausrichtung	tagesaktuelle Geschäftsprozesse	verdichtete, transformierte Daten
Zeitbezug	zeitpunktbezogen	zeitraumbezogen
Zustand	redundant, inkonsistent	bereinigt, konsistent
Update	laufend	fortschreitend
Queries	strukturiert, statistisch im Programmcode (meist)	Ad-Hoc: für wechselnde und komplexe Fragestellungen, für vorgefertigte Standardauswertungen

Abb. 1: Architektur von Business Intelligence [KeMU06]

Metadaten hingegen sind beschreibende Daten. Sie beinhalten alle Informationen, die den Aufbau, die Administration und die Wartung des Systems vereinfachen und die Informationsgewinnung ermöglichen. Somit sind Metadaten Informationen ÜBER bestimmte Daten, die aus anderen Quellen kommen, wie z. B. aus dem DWH, aus der Basisdatenbank oder Daten über physische Speicherinformationen, Qualitätssicherung, Zugriffsrechte, konzeptuelle und logische Datenbankschemata und Dokumentation.[Baue04]

Um die Metadaten abzulegen, wird ein Repository des Data Warehouses angelegt, ein verwaltendes Verzeichnis zur Speicherung und Beschreibung von digitalen Objekten. [Ortn99] Ein Repository reduziert den Aufwand für den Aufbau und den Betrieb des Systems. Darüber hinaus wird allen Anwendergruppen ein optimaler Gewinn an Informationen aus dem DWH ermöglicht.[Baue04]

1.3 Data Warehouse (DWH)

Die Speicherung der Daten erfolgt, redundant zu den operativen Systemen, themenspezifisch und dauerhaft im Data Warehouse. Dieses DWH bildet das funktionale Fundament einer jeden BI-Lösung, bestehend aus der Basisdatenbank und dem Repository, welches die Metadaten beinhaltet. [Baue04] Im Gegensatz zu den transaktionsorientierten Vordersystemen richtet sich der Fokus auf einen Zeitraumbezug wie Tage, Wochen oder Monate bis zu einem Horizont von zehn Jahren. Das Data Warehouse gewährleistet die Datenkonsistenz zwischen dem operativen Tagesgeschäft und den resultierenden Analysen und Reporting. In einem Unternehmen kommen je nach Anforderungen mehrere DWHs zum Einsatz.

Ein Data Warehouse ist also der zentrale Punkt einer Business Intelligence Lösung und darum der Ort im Unternehmen mit der höchsten Datenqualität. An das Konzept eines DWH werden folgende Kriterien gestellt:[Schr05]

- **kein direkter Zugriff auf die Produktionsdaten**
Aus Gründen der Sicherheit bzw. der Performance erfolgt kein direkter Zugriff auf die echten operativen Daten im Unternehmen. Der Eingriff in die Kerninformationsstruktur des Unternehmens kann empfindlichen Einfluss auf die operativen Abläufe nehmen. Die Daten werden für den Input der Analysen über einen Transfervorgang in einen zweiten Speicherort redundant geladen. (Dieser Vorgang wird als „Staging“ bezeichnet.)

- **Integration von verteilten Datenbeständen**
Die zu Grunde liegenden Daten werden in der Praxis in mehreren verteilten Datenbanken gehalten. Im DWH werden diese isolierten Daten zusammengeführt, was eine Sicht auf alle für die Analyse relevanten Daten ermöglicht und diese miteinander in Beziehung setzt.

- **Verarbeitung und Konsolidierung der Daten**
Eine Verfälschung der Analyseergebnisse wird durch die Prüfung der Daten auf bestimmte Qualitätsmerkmale ausgeschlossen. Um unnötige Rechenleistung bei der Datenauswertung zu vermeiden, werden diese auf den entsprechenden Stufen verdichtet. Diese Transformation findet während der Übertragung der operativen Daten in das Data Warehouse statt.

Das Aufnehmen aller relevanten Daten aus den operativen Vordersystemen kann Dimensionen annehmen, die nicht immer leicht in den Griff zu bekommen sind und hohe Ansprüche an die Performance stellen. Aus diesem Grund werden oft so genannte Data Marts gebildet. Data Marts bestehen aus einer Teilmenge des gesamten Data Warehouses zu einem bestimmten Themengebiet, wie beispielsweise einer Abteilung oder einer Produktgruppe.[NABD05]

1.4 ETL-Prozess

Die Bereitstellung und Überführung der operativen Daten in das Data Warehouse ist von zentraler Bedeutung und geschieht im so genannten ETL-Prozess (Extrahieren-Transferieren-Laden). ETL bezeichnet die in folgender Abbildung dargestellten Phasen:

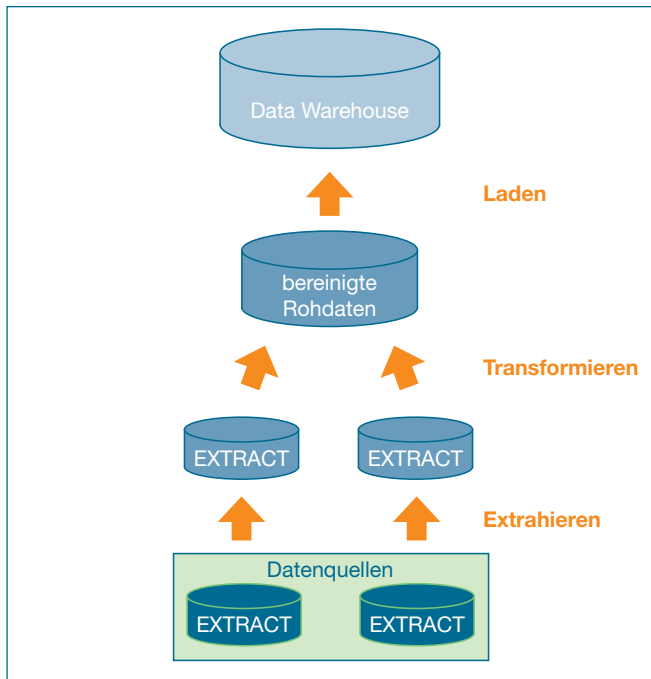


Abb. 2: Der ETL-Prozess [Manh08]

Zu Beginn werden die relevanten Rohdaten ausgewählt und aus den internen und externen Datenquellen extrahiert. In der darauf folgenden Phase erfolgt die Transformation der extrahierten Daten nach den in Tabelle 2 abgebildeten aufeinander folgenden Schritten.

Schritt	Erklärung
Filterung	Bereinigung syntaktischer und inhaltlicher Defekte der Daten
Harmonisierung	Betriebswirtschaftliche Abstimmung der gefilterten Daten
Aggregation	Verdichtung der gefilterten und harmonisierten Daten
Anreicherung	Berechnung und Speicherung betriebswirtschaftlicher Kennzahlen

Tab. 2: Transformationsschritte des ETL-Prozesses [Manh08]

In der letzten Phase des ETL-Prozesses werden die transformierten, bereinigten Daten periodisch oder ereignis- bzw. abfragegesteuert in das Data Warehouse transferiert. Die Höhe des Aufwandes ist abhängig von der formalen und inhaltlichen Heterogenität der Vorkomponenten.[BaNK03] Die transformierten, verdichteten und hoch aggregierten Daten werden für verschiedene Steuerungs-, Kontroll- und Analysezwecke zur Entscheidungsfindung vom Management eingesetzt. Der ETL-Prozess ist bei Anbietern von kompletten BI-Lösungen ein Teil der in den Unternehmen eingesetzten BI-Plattform und für die automatische Bereitstellung der relevanten Daten zur Unternehmenssteuerung zuständig.

1.5 OLAP

Online Analytical Processing (OLAP) ist eine komplexe, interaktive Analyse „in Form einer mehrdimensionalen Sicht auf vorhandene Datenbestände“[Pepe05]. Im Gegensatz zu Data Mining als analytisches Informationssystem zählt OLAP zu den hypothesengestützten Analysemethoden. „OLAP ist eine Kategorie von Anwendungen und Technologien zum Sammeln, Verwalten, Bearbeiten und Darstellen von multidimensionalen Daten für Analyse- und Verwaltungszwecke.“ [Pend08]

OLAP, als multidimensionales Konzept, setzt sich aus den Merkmalen Dimension, Measure, Hierarchie und Cube zusammen:[NABD05]

- **Dimensionen** bilden sich aus den einzelnen in ihnen definierten Elementen und beschreiben die Daten, wie beispielsweise Dimension Produkt: zusammengesetzt aus Produkt A, Produkt B, Produkt C und Produkt D.
- **Measures** sind die einzelnen oder zusammengefassten (aggregierten) Werte, auf welche über die Dimensionen im Cube zugegriffen wird.
- **Hierarchien** beschreiben die innerhalb einer Dimension definierten Elemente. Die Aggregation der Measures erfolgt über die einzelnen Hierarchie- beziehungsweise Verdichtungsstufen. Als Ebenen bieten sich beispielsweise für die Dimension Produkt die Stufen Produkt - Produktgruppe - alle Produkte an oder Tag - Monat - Quartal - Jahr für die Dimension Zeit, über welche die Verdichtung der Werte erfolgt. Klassifikationsknoten bilden als Elemente die Verdichtungsstufe einer Dimension.

• Inhalt und Struktur einer multidimensionalen Datenmenge definieren sich durch Dimensionen und Measures, welche den so genannten **Cube**, auch OLAP-Würfel oder OLAP-Tabelle genannt, aufspannen. Die Struktur eines Würfels muss mindestens drei Dimensionen betragen und ist nach oben offen. Aus Gründen der Übersichtlichkeit bietet sich eine Begrenzung zwischen maximal 15 und 23 Dimensionen an. Die nachstehende Abbildung verdeutlicht grafisch den Zusammenhang der beschriebenen Merkmale eines OLAP-Cubes.

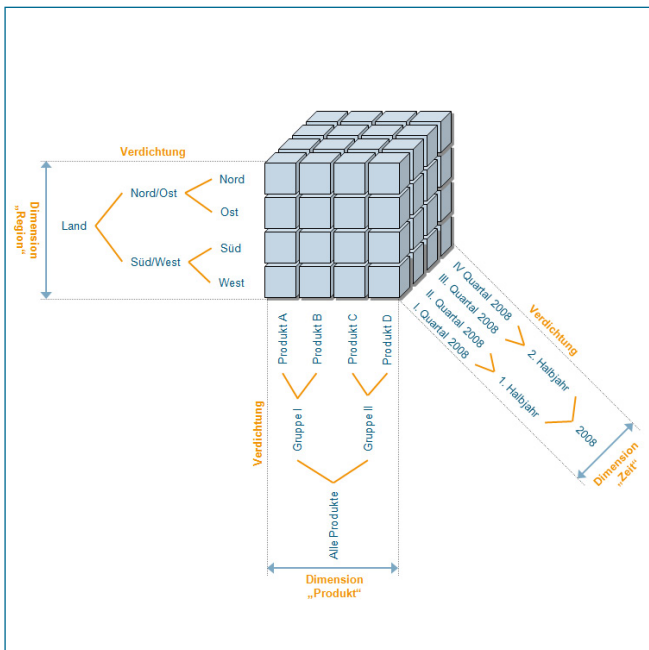


Abb. 3: Beispiel eines 3-dimensionalen OLAP-Cubes mit Verdeutlichung der Verdichtungsstufen

1.6 Informationszugriff

OLAP setzt als Analysewerkzeug auf den im Data Warehouse konsolidierten Daten, die aus den verteilten Vorsystemen stammen. Der Anwender kann mit spezifischen Operatoren dynamisch durch die multidimensionale Struktur der Daten navigieren und diese aus verschiedenen betriebswirtschaftlichen Blickwinkeln betrachten:

Operation	Erläuterung	Darstellung
Pivotierung (Rotation)	Drehung des Würfels um seine Achse durch Vertauschen der Dimensionen	
Roll-up (Komplement zu Drill-down)	Aggregation der Daten entlang des Konsolidierungspfades	
Drill-down (Komplement zu Roll-up)	Herunterbrechung der verdichteten zu den detaillierten Daten	
Drill-across	Wechsel des Würfels zu einem anderen Würfel	
Slice	Aggregation von Kenngrößen über einen Klassifizierungsknoten einer Dimension. Betrachtung einer einzelnen Scheibe des Würfels	
Dice (Ad-Hoc)	Betrachtung eines Teilwürfels für bestimmte Kombinationen von Klassifizierungsknoten.	

Tab. 3: Multidimensionale Operatoren im Überblick [Baue04]

2. Anforderungen an das Unternehmen

Von der fachbereichsspezifischen BI-Lösung bis zur unternehmensweiten Analyse, für das erfolgreiche Implementieren und Betreiben von Business Intelligence muss ein Unternehmen folgende Voraussetzungen schaffen:[Zirk07]

- Vermeidung inkonsistenter Informationsgrundlagen und Gewährleistung der Datenqualität
- Flexible IT-Architekturen zur Kostenreduzierung
- Berücksichtigung der Zunahme von BI-Aktivitäten zu ihrer späteren organisatorischen Bewältigung
- Vermeidung von Wildwuchs in den Dimensionen, Fachlichkeit, Technik und Organisation des BI-Systems
- Reduzierung der Abhängigkeit von externen BI-Lieferanten durch unternehmensinternes Know-how BI 360° Blickwinkel

Die Akzeptanz des Einsatzes von BI-Werkzeugen durch die Anwender sowie die Stellung der Informationen im Kontext der Geschäftsprozesse sind als wesentliche Erfolgsfaktoren einer Business Intelligence Lösung anzusehen. Führungskräfte und Prozessverantwortliche können die aktuelle Situation eines jeden Prozesses verstehen und bei Risiken und Problemen rechtzeitig die hierfür notwendigen Maßnahmen treffen. [ISRe] Business Intelligence ist mehr als IT und bildet als ganzheitlicher Ansatz zur Verbesserung der Entscheidungsqualität die Grundlage für den Unternehmenserfolg.

Autor:

Andreas Raaz | Consultant | raaz@pst.de
PST Software & Consulting GmbH

Literatur- und Quellverzeichnis

[Baue04] Bauer A.: Data Warehouse Systeme, Architektur, Entwicklung, Anwendung, dpunkt, Heidelberg (2004)

[ISRe] o.V.: Handeln in Zeiten des Anbieterdarwinismus, Inwieweit können Anwender von der Marktkonsolidierung profitieren?, in: isreport, 12. Jahrgang (2008)

[KeMU06] Kempfer H.-G., Mehanna W., Unger C.: Business Intelligence – Grundlagen und praktische Anwendungen, 2. Auflage, Vieweg, Wiesbaden (2006)

[Manh08] Manhart K.: BI-Datenmanagement (Teil1): Datenaufbereitung durch den ETL-Prozess, Grundlagenserie Business Intelligence, 17.03.2008, Online im Internet: URL: http://www.tecchannel.de/server/sql/1746250/bi_datenmanagement_teil_1_datenaufbereitung_durch_den_etl_prozess/ [Stand: 09.11.2008]

Manhart K.: BI-Datenmanagement (Teil1): Datenaufbereitung durch den ETL-Prozess, Grundlagenserie Business Intelligence, 17.03.2008, Online im Internet: URL: <http://images.tecchannel.de/images/tecchannel/bdb/365030/original.jpg> [Stand: 09.11.2008]

Manhart K.: Business Intelligence (Teil2): Datensammlung und Data Warehouses, Grundlagenserie Business Intelligence, 23.01.2008, Online im Internet: URL: http://www.tecchannel.de/server/sql/1739205/business_intelligence_teil_2_datensammlung_und_data_warehouses/index2.html [Stand: 18.11.2008]

[NABD05] Neumann B., Avevedo P., Brosius G., Dehnert S., Scheerer B.: Business Intelligence und Reporting mit Microsoft SQL Server 2005, Microsoft Press, Unterschleißheim (2005)

[Ortn99] Ortner E.: Repository Systems. Teil 1: Mehrstufigkeit und Entwicklungsumgebung, in: Informatik Spektrum, Ausgabe 4, Springer Berlin Heidelberg, (1999)

[Pepe05] Pepels W.: What do all the TLAs and jargon really mean?, 23.10.2008, Online im Internet: URL: www.olapreport.com/glossary.htm [Stand: 14.11.2008]

[Schr05] Schrödl H.: Business Intelligence mit Microsoft SQL Server 2005, Carl Hanser Verlag, München Wien (2006)

[Zirk07] Zirkel M.: BI-Strategie, Der Weg zum maximalen Return-On-BI-Invest, Sterie Mummert Consulting AG, München (2007)

© PST Software & Consulting GmbH, Alle Rechte vorbehalten. Alle Angaben können sich ohne Vorankündigung ändern. 03/2010