

Inhaltsverzeichnis

Herzlich willkommen	4	Beispiele von Geschäftsanalysen und Data Mining	46
Wissen managen	6	Typische Reports.....	46
Daten, Informationen, Wissen.....	6	... Zeitreihenanalyse.....	46
Controlling - Anforderungen an betriebliche Informationssysteme.....	7	... berechnete Kennziffern	47
Data Warehouse - die zentrale Sammelstelle für Daten	9	... Absatzweganalyse	47
Der Begriff.....	9	... grafische Analysen.....	48
Operative und informative Systeme.....	10	... Top 10 Analysen	49
Extraktion, Transformation und Laden	11	... Vertreteranalyse.....	50
Die OLAP-Technologie	12	Kundenklassifikation	50
Der multidimensionale Datenwürfel	13	... 80-20-Klassifikation.....	50
Die FASMI-Anforderungen.....	14	... Histogrammfunktion	53
Organisationsformen von Data-Warehouse-Systemen – Data Mart	14	Data Mining	55
Meta-Datenbank	16	... Kaufverhaltensanalyse.....	55
Archivierungssystem.....	16	... Warenkorbanalyse	59
Analysewerkzeuge.....	16	Plan-Ist-Vergleich	63
Coodsche Regeln	17	Die Implementierung eines Data-Warehouse-Systems	67
Data Mining – das Graben nach Kundeninformationen	19	Ordne deine Informationslandschaft.....	67
Der Begriff.....	19	Plane dein Data-Warehouse-Projekt	67
Aufgaben und Methoden.....	19	Zum Schluss noch für Interessierte	70
... Vorhersage.....	19	Ein Data-Warehouse-Portal	70
... Segmentierung.....	20	Wir suchen die Grenze	70
... Abhängigkeitsentdeckung	20	Das wär's	73
Einige nützliche Statistikfunktionen aus Excel	22	Dateiverzeichnis	74
SUMMEWENN-Funktion.....	22	Glossar	75
RANG-Funktion	24	Stichwortverzeichnis	77
PEARSON-Funktion	24		
Histogramm	25		
Unsere Beispielfirma	27		
Was ist klein, mittelgroß oder groß bei Firmen? ..	27		
Was macht unsere Firma?	27		
Welche Daten haben wir?	28		
Wir bauen unser eigenes Datenlager mit Access und Excel	30		
Betrachte deinen Bauplan.....	30		
Bereite deine Daten auf.....	30		
Modelliere deine Daten	35		
Rufe deine Daten mit Microsoft Query ab	37		
Arbeite mit dem OLAP-Cube-Assistenten.....	39		
Ohne Berichte geht es nicht	42		
Der PivotTable-Bericht.....	42		
Dein erster Bericht – die Vorlage für deine Geschäftsanalysen	42		
Pivot-Bericht ist nicht gleich Pivot-Bericht.....	44		

110 interessante KnowWare-Titel!

4,- KnowWare bringt jeden Monat mehrere Computerhefte auf den Markt, von A wie Access über H wie Homepage und P wie Programmierung bis W wie Word oder X wie XML. Bitte blättere zum Bestellschein auf Seite 79–80 oder schau dich auf www.knowware.de um! Hier eine Auswahl:

- Excel 2000–2003 im Schnellkurs
- Excel 2002 für Einsteiger
- Excel 2003 (2002/XP) für Fortgeschrittene
- Diagramme mit Excel 2003/2
- Word für Profis
- Office 2003 für Einsteiger
- Access 2003/2002 leicht & verständlich



Herzlich willkommen

... in der Data-Warehouse-Welt. Lass dich anstecken vom Schnüffeln nach Daten, grabe wie ein Bergmann in den Tiefen der Erde nach Informationen im Datenwust, betreibe dein eigenes Data Mining.

Schreck lass nach: Wieder so ein englischer Begriff ... Aber solche Begriffe haben sich in der Informationswelt eingebürgert und lassen sich durch deutsche Wörter nur schwer ersetzen.

Der Begriff Data Warehouse beschreibt nicht etwa ein „Daten-Warenhaus“: Es geht eher um so etwas wie ein „Daten-Zentrallager“ – und was ein Lager ist, davon hat wohl jeder eine gewisse Vorstellung.

Data Mining, also die Suche nach Daten wie in einer Kohlengrube, ist zwar ein schöner bildhafter Begriff – aber seien wir ehrlich: Wer bekommt hier wohl inhaltliche Assoziationen? Wie immer man diesen Begriff nun definiert – es geht im Grunde um das Erkennen bzw. Freilegen von versteckten Informationen und Mustern aus umfangreichen Datenbeständen.

Und wem nützt das eigentlich? Nun – nicht nur Weltkonzerne und Großbetriebe haben ein Interesse an einer allumfassenden Informationsversorgung; auch in kleinen und mittleren Betrieben ist es wichtig, dass Informationen jederzeit aktuell sind, wenn Entscheidungsträger und Mitarbeiter rechtzeitig und angemessen reagieren sollen.

Das Problem liegt nicht etwa darin, Daten zu erzeugen. Softwaresysteme für Fakturierung, Lagerhaltung oder Buchhaltung liefern bei jeder geschäftlichen Transaktion eine Unmenge an Daten. Es geht vielmehr darum, diese Datenflut sinnvoll zu nutzen. Und eben hier schlägt die Stunde des Data Warehouse: Es handelt sich dabei nämlich um ein Datenbanksystem, das getrennt von den operativen Datensystemen Daten sammelt und Entscheidungsprozesse unterstützt.



Es gibt eine Vielzahl von Softwareanbietern, die umfangreiche Lösungen liefern, beispielsweise SAP, Cognos, Business Objects oder MicroStrategy. Solche professionellen Lösungen sind unbedingt leistungsfähig – sie haben aber auch große Nachteile: Sie

- kosten bis zu hunderttausenden von Euros,
- verlangen oft neue zusätzliche Hardware,
- sind mit teuren Berater- und Spezialistenleistungen verbunden und
- erzeugen zusätzliche Betriebskosten durch Lizenz- und Wartungsverträge.

Außerdem sind diese professionellen Werkzeuge so leistungsstark, dass ihre Möglichkeiten im betrieblichen Alltag vielfach nur zu einem geringen Teil genutzt werden können.

Kleine Unternehmen verfügen nicht über die notwendigen Gelder. Aber auch hier gibt es Hilfe, und zwar in der Form von Data-Warehouse-Systemen mit Online Analytical Processing-Technologien – kurz OLAP genannt.

Benötigt werden hierfür nur Excel und Access in der Version 2000 oder neuer – und natürlich der zukünftige Data Warehouse-Manager.



Microsoft Excel



Microsoft Access

Du solltest gewisse Vorkenntnisse auf diese Reise mitnehmen; dazu gehören ein grundlegendes Datenverständnis und ein vertrauter Umgang mit den Grundfunktionen von Access und Excel. Das reicht aber im Grunde völlig – wir wollen dir diese Welt praktisch nahe bringen.

Hast du dieses Heft durchgearbeitet, weißt du dann hoffentlich, was ein Data Warehouse ist – und auch die OLAP-Technologie ist für dich kein Fremdwort mehr.

Du hattest bisher noch nichts mit Datenmodellen zu tun? Kein Problem – in den Kapiteln zu unserer Beispielfirma erhältst du die grundlegenden Informationen zu diesem Thema.

Außerdem lernst du, wie du mit dem Excel-Cube-Assistenten umgehst und auf dieser Basis Pivot-Berichte erstellst. Schließlich kannst du dein eigenes Tun in die Business Intelligence einordnen – dabei handelt es sich um eine Strategie zum Umgang mit Informationen.

Auch für dieses Heft haben wir eine Serviceseite erstellt – du findest sie hier:
www.bilke.de/dwh.

Auf dieser Seite stehen auch die im Heft vorgestellten Beispieldateien zum Download bereit.

Viel Spaß beim Aufbau deines Data Warehouses wünschen dir

Petra Bilke und Steffen Bilke.

Wissen managen

Daten, Informationen, Wissen

Daten, Informationen und Wissen – diese Begriffe nutzen wir jeden Tag, auf aktive oder passive Weise. Meist tun wir das sogar intuitiv richtig – und machen uns um die exakte Begriffsbeschreibung kaum Gedanken.

Aber worum geht es hier eigentlich ganz genau? Sehen wir uns die Sache an einem oft benutzten Beispiel an; dabei geht es um Uhrzeit und Entfernung.

Daten symbolisieren Sachverhalte; sie werden auf der Grundlage von bestimmten Ordnungsregeln aus Zeichen gebildet.

Beispiel: 16.00 Uhr; 20 km

Informationen entstehen, wenn Daten in der Form von Strukturen miteinander verknüpft werden. Sie erhalten dadurch eine Bedeutung und einen Zweckbezug.

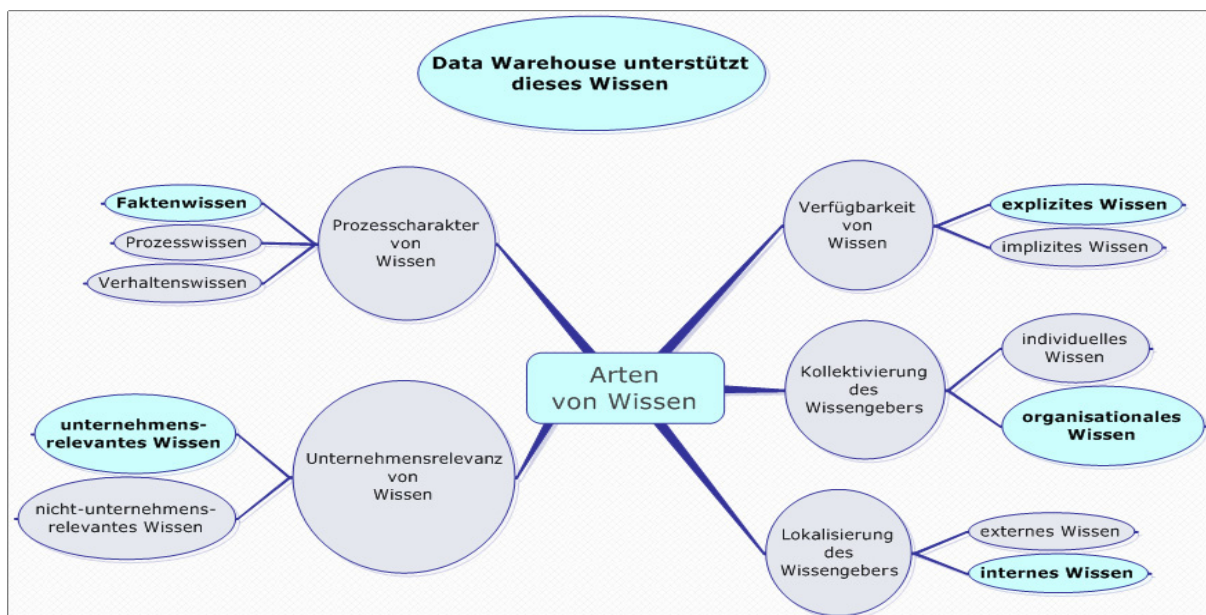
Beispiel: Um 16.00 Uhr wollen wir Paul treffen – und müssen dazu einen Weg von 20 km zurücklegen.

Bei unseren Handlungen verarbeiten wir die vorliegenden Informationen; und das führt wiederum zu Wissen. Wissen ist also eine individuelle Fähigkeit, durch die du eine bestimmte Aufgabe ausführen kannst.

Beispiel: Wenn wir um 16.00 Uhr bei Paul sein wollen, müssen wir 20 km im Berufsverkehr fahren; also müssen wir, wenn wir pünktlich sein wollen, um 15.00 Uhr losfahren.

Und damit können wir nun zusammenfassen:

Wissen ist das Produkt aus Informationen und persönlichen Erfahrungen sowie Fertigkeiten. Es äußert sich in Form von Kenntnissen und Fähigkeiten.



Arten von Wissen

Will man heutzutage im Wettbewerb bestehen, kommt dem Wissen eine zentrale Bedeutung zu. Und Wissen kann, wie unsere Abbildung zeigt, viele verschiedene Formen annehmen.

Dabei solltest du bezüglich des Wissens unterscheiden zwischen

- Unternehmensrelevanz,
- Prozesscharakter und
- Verfügbarkeit.

Weiteres Wissen resultiert aus

- Lokalisierung und
- Kollektivierung

des Wissensgebers.

Jede Firma verfügt über viel *unternehmensrelevantes Wissen*. Das Data-Warehouse-Konzept bietet die Möglichkeit, das *unternehmensrelevante Wissen* zu erweitern.

Der Wissensgeber – das Data Warehouse – ist in der Firma lokalisiert; also können wir dieses Wissen als *internes Wissen* der Firma bezeichnen.

Der Kern dieses Wissens ist das *Faktenwissen*, also das Wissen über reale Sachverhalte.

Nun verhält es sich mit diesem Wissen so, dass es, da es sozusagen implizit in den Köpfen der Mitarbeiter gespeichert ist, nicht unmittelbar zugänglich ist. Das Data Warehouse bringt *explizites Wissen* hervor und verbessert somit seine Verfügbarkeit. Im Unterschied zur impliziten Art lässt es sich somit leicht übertragen, vervielfältigen und speichern.

Das individuelle Expertenwissen der Mitarbeiter nimmt einen etwas untergeordneten Platz ein, und das *organisationale Wissen* des Unternehmens tritt mit Hilfe des Data Warehouses in den Vordergrund.

Es stimmt – das hört sich alles etwas theoretisch an. Aber im Verlauf unseres Heftes wollen wir zeigen, dass man mit einem Data-Warehouse-System einen schnellen Zugang zu viel Wissen erlangen kann, das normalerweise nicht so ohne weiteres zur Verfügung steht.

Controlling – Anforderungen an betriebliche Informationssysteme

Wissen managen – das bedeutet letztendlich Transparenz oder Klarheit über die vorhandenen Daten und Informationen sowie deren Aufbereitung zu schaffen. Es geht darum, Mitarbeitern und Entscheidungsträgern in einem Unternehmen schnellen Zugriff auf diesen Informations- und Wissenspool zu ermöglichen.

Ein altbekannter Spruch sagt: „Vertrauen ist gut, Kontrolle ist besser.“ Schön und gut – nur ist diese Kontrolle grundsätzlich rückwärts gerichtet; sie untersucht, wie etwas gelaufen ist.

Wenn wir die eben erwähnte Transparenz und Klarheit in unserem Betrieb erreichen wollen, müssen wir aber auch vorwärts planen – und natürlich kontrollieren; und um diese beiden Arten der Übersicht voneinander zu unterscheiden, bezeichnen wir dies als *Controlling*.

**Kontrolle analysiert das Vergangene;
Controlling kontrolliert im Voraus.**

Nun ist dieses Controlling keineswegs eine neue Erfindung, und ebenso wenig ist es etwas, das nur für Großunternehmen Bedeutung hat. Auch Kleinstunternehmen betreiben es, und dabei wenden die Unternehmer oft Controllingmethoden an, ohne dass ihnen dies bewusst ist.

Das Controlling lässt sich auf zweierlei Art anwenden:

- Beim *strategischen* Controlling handelt es sich um die langfristige Unternehmensplanung. Sie beschreibt, wie sich ein Unternehmen entwickeln soll. Hier geht es um die dauerhafte Sicherung des Unternehmens.
- Beim *operativen* Controlling geht es darum, kurzfristige Maßnahmen durch ein funktionsübergreifendes betriebliches Steuerungssystem abzusichern.

Dies alles erfordert ein komplettes Informationssystem, das bestimmten Ansprüchen genügen muss. Die nach unserer Meinung wichtigsten zählen wir hier kurz auf:

1. Hohe Informationsqualität

Die Daten müssen sich aus verschiedenen Gesichtswinkeln betrachten lassen, um so betriebswirtschaftliche Ad-hoc-Abfragen zu ermöglichen, also Abfragen, die sich aus der jeweiligen Situation ergeben.

2. Hohe Verfügbarkeit

Sollen Informationen jederzeit aktuell verfügbar sein, dürfen sie nicht an unterschiedlich gearteten Orten und in unterschiedlichen Formaten gespeichert sein.

3. Anspruchsvolle Entscheidungsunterstützung

Sollen die Daten unterschiedlichen Informationsbedürfnissen entsprechen, müssen sie entsprechend aufbereitet und verdichtet werden.

4. Hochentwickeltes Berichtswesen

Sollen die Informationen einleuchtend und ansprechend präsentiert werden, müssen die Daten in Standardberichten wie auch in individuell gestalteten Analysen darstellbar sein. Die Organisationsform des Unternehmens muss die Grundlage für das Berichtswesen bilden, und intuitive Erlernbarkeit ist eine wichtige Voraussetzung.

Wir hoffen, dass du mit unserer Umsetzung des Data-Warehouse-Konzepts den kompletten Werkzeugkasten für ein betriebliches Informationssystem erhältst.

Data Warehouse – die zentrale Sammelstelle für Daten

Der Begriff

Wir wollen den Begriff Data Warehouse, also Datenlager, kurz und knapp definieren, ohne dabei zu sehr ins Detail zu gehen – das überlassen wir lieber den Theoretikern.

Ein Data Warehouse ist eine Sammlung von unternehmensspezifischen Daten aus unterschiedlichen Quellen zur Unterstützung von Managemententscheidungen.

Sollen diese Daten dem Anspruch eines Data Warehouses entsprechen, müssen sie bestimmte Kriterien erfüllen; sie müssen

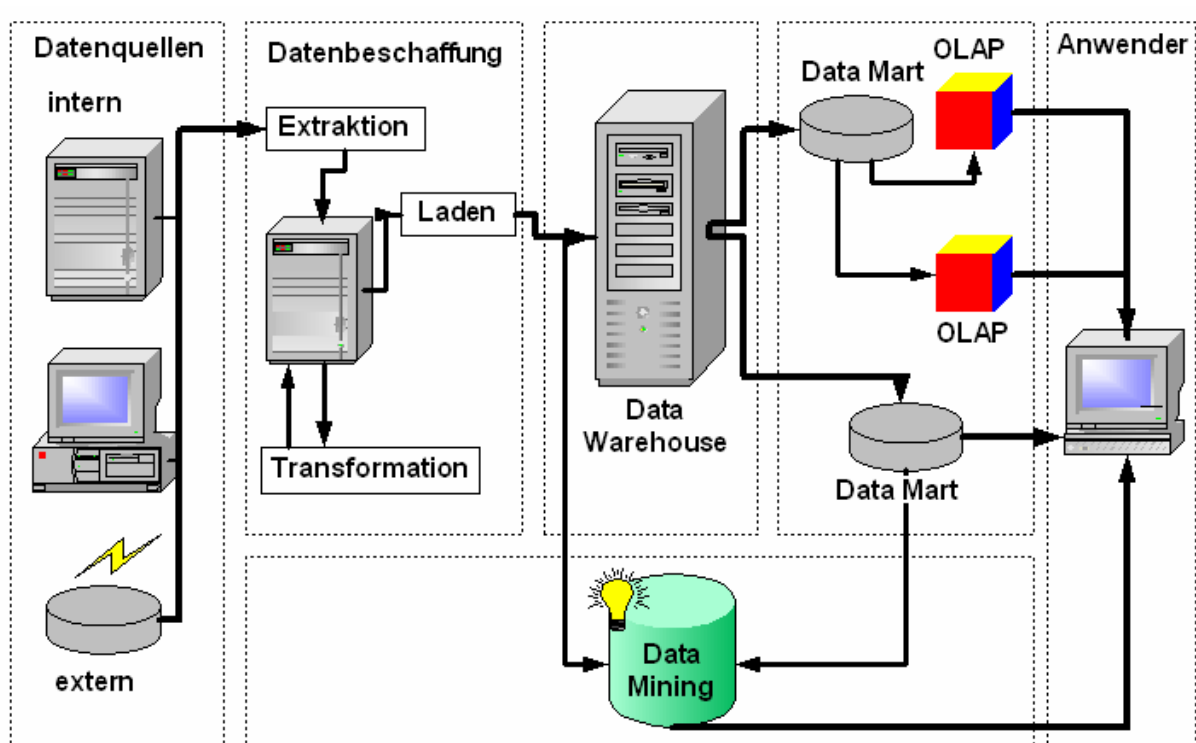
- fachorientiert sein; d.h. sie unterliegen speziellen Anwendungszielen;
- integriert sein; d.h. sie stammen aus verschiedenen Quellen und werden in einer Datenbank zusammengeführt;

- historisch sein; d.h. sie beziehen sich auf bestimmte Zeitpunkte, und das ermöglicht wiederum Zeitreihenanalysen und Trendberechnungen;
- unveränderlich sein; d.h. sie werden im Allgemeinen nicht geändert oder gelöscht. Ausgenommen sind natürlich Daten-Updates.

Wie bei einem richtigen Warenlager geht es also in der Praxis um die Bewirtschaftung und Nutzung – nur dass es sich hier nicht um Waren, sondern um Daten handelt. Das erfordert die Komponenten Datenbeschaffung, Speicherung und Analyse. In ihrer Gesamtheit bilden diese Komponenten das Data-Warehouse-System.

Die ursprünglichen Datenquellen und ihre Systeme werden nicht als Bestandteil des Data-Warehouse-Systems angesehen.

In den folgenden Abschnitten wollen wir uns genauer mit den einzelnen Komponenten befassen.



Data-Warehouse-Architektur

Operative und informative Systeme

Betriebswirtschaftliche Lösungen, wie Fakturierungs-, Lagerhaltungs- oder Buchhaltungsprogramme, um nur einige Beispiele zu nennen, werden auch ERP-Lösungen genannt. Das Kürzel steht für Enterprise Resource Planning.

Der Oberbegriff ERP umfasst unternehmensübergreifende Software-Lösungen, die zum Optimieren von Geschäftsprozessen eingesetzt werden.

Ein typischer Vertreter dieses Software-Typs ist das Produkt SAP R/3. Das Unternehmen SAP AG entwickelt „Systeme, Anwendungen und Produkte in der Datenverarbeitung“; das „R“ steht für „Realtime“.

Die enorme Produktpalette der Firma könnte man getrost als Software-Bibliothek bezeichnen. Informationen hierzu gibt dir das Hilfe-Portal der Firma unter der Adresse help.sap.com.

The screenshot shows the SAP Help Portal interface. The top navigation bar includes 'Home', 'SAP', 'SAP Service Marketplace', and a search field. Below the navigation bar, there's a section titled 'SAP Business Information Warehouse'. A warning message states: 'This area relates to SAP Business Information Warehouse 3.5.1 Add-On and earlier. Newer releases of this SAP component are now part of SAP NetWeaver® 04. Go to http://help.sap.com/new04 to find all relevant documentation for the latest version.' Below this, there are three release entries:

Release	SAP BI Content 3.5.1 Add-On
File Type	HTML
Language	English German
Material No.	500 680 46 (DVD)
Release Date	April 2004

Release	SAP BW 3.1/BI Content 3.3 Add-On
File Type	HTML
Language	English German
Material No.	500 653 21
Release Date	October 2003

Release	SAP BW 3.1/BI Content 3.2 Add-On
File Type	HTML
Language	English German
Material No.	500 629 74
Release Date	May 2003

Release	SAP BW 3.1 Content
---------	--------------------

Das SAP-Hilfe-Portal

Hier findest du auch Informationen zur Data-Warehouse-Lösung der Firma SAP, genannt „Business Information Warehouse“. Nimm dir ruhig die Zeit, einmal durch diese wahre Informationsflut zu surfen – hier erhältst du, zumindest in den allgemeinen Abschnitten, garantiert interessante Hinweise.

Sehen wir uns nun am Beispiel von R/3 die Hauptkomponenten eines ERP-Systems an. Die allgemeinen Funktionen sind in zwei Komponenten enthalten:

- BC – Basis (engl.: Basic Components)
- CA – Anwendungsübergreifende Komponenten (engl.: Cross Applications)

Die drei großen Anwendungsbereiche lauten:

- AC – Rechnungswesen (engl.: Accounting) mit Finanzwesen, Controlling, Unternehmenscontrolling, Investitionsmanagement und Treasury
- HR – Personalwirtschaft (engl.: Human Resources) mit Personalmanagement, Veranstaltungsmanagement und Personalzeitwirtschaft
- LO-Logistik (engl.: Logistics) mit Allgemeiner Logistik, Materialwirtschaft, Instandhaltung, Produktionsplanung und -steuerung, Projektssystem, Qualitätsmanagement und schließlich dem wichtigen Vertrieb.

In diesem System werden alle Transaktionen abgebildet. Es geht um die operativen Geschäftsprozesse, die im Unternehmen anfallen. Für diese Funktionalität gibt es den Begriff des Online Transaction Processing (OLTP).

OLTP-Systeme sind vorrangig darauf ausgerichtet, operative Daten effizient und zuverlässig zu verwalten.

Die Vorgänge in den operativen OLTP-Systemen stellen die Grundlage für die Daten dar, die in einem informativen System von der Art eines Data Warehouses gespeichert werden.

Wir haben also zwei unterschiedliche Systeme der Informationserfassung. Sehen wir uns nun in der folgenden Tabelle ihre unterschiedlichen Charakteristika an.

	Operative Systeme (Enterprise Resource Planing)	Informative Systeme (Data Warehouse)
Ziele	Automatisierung der Geschäftsprozesse	Erzeugen von Wissen
Prioritäten	Hohe Verfügbarkeit, hoher Datendurchsatz	Einfache Benutzung, flexibler Datenzugriff
Datensicht	detailliert	verdichtet, aggregiert
Antwortzeiten	sehr schnell, Sekundenbereich	normal, bis zu mehreren Minuten
Verarbeitungseinheit	Datensatz, eindimensional	Matrizen, multidimensional
typische Datenstrukturen	relational	multidimensional
Datenoperationen	Einfügen, Verändern, Löschen, Lesen	Lesen
Abfragen	vorhersehbar, periodisch	unvorhersehbar, ad hoc
Datenalter	aktuell	historisch
Datenaktualität	ständig gegeben	nur nach Updates
Update	permanent	niedrige Frequenz, teilweise monatlich
Daten aus anderen Anwendungen	geringe Integration	umfassende Integration
Zeithorizont	bis ca. 2 Jahre	2 bis 10 Jahre
Archivierung	ja	ja

Es gibt also sehr unterschiedliche technische und betriebswirtschaftliche Anforderungen an operative und informative Systeme. Daher sollte man diese Systeme auch deutlich voneinander trennen. Die notwendige Verbindung muss anschließend durch entsprechende Schnittstellen und Prozesse erstellt werden.

Extraktion, Transformation und Laden

Hast du dich mit einer echten Lagerwirtschaft beschäftigt, kennst du sicher den Begriff „LTU-Prozesse“. Allerdings handelt es sich hier nicht um den bekannten Reiseanbieter, sondern um „Lager-, Transport- und Umschlagsprozesse“. Das Data-Warehouse-Konzept kennt eine ganz ähnliche Abkürzung – auch was den Inhalt betrifft: die so genannten ETL-Prozesse. Das Kürzel steht für Extraktion, Transformation, Laden. Softwaretools, die diesen Prozess der Datenbeschaffung unterstützen, nennt man auch ETL-Tools.

Zunächst ermittelst du deine Datenquellen, z. B. Exceltabellen und die Dateien des Fakturierungsprogramms, mit dem Rechnungen in deinem Betrieb geschrieben werden. Diese Daten müssen nun also in unser Data Warehouse geladen werden.

Und wie sieht das nun im Einzelnen aus?

- Bei der *Extraktion* definierst du, welche Felder der einzelnen Datenquellen in dein Datenlager gelangen sollen.
- Beim *Transformieren* geht es darum, die unterschiedlichen Datenstrukturen und Dateninhalte zu konsolidieren; es geht also um die technische Aufbereitung von Daten aus unterschiedlichen Systemen. So können z. B. Kundennummern in dem einen System in einem numerischen Feld gespeichert sein, während das andere System diese Daten alphanumerisch speichert. Diese Unterschiede müssen einander angeglichen, also *homogenisiert* werden.
- Daneben kann ein Cleansing, also eine Reinigung erforderlich sein. Data-Cleansing-Verfahren werden verwendet, um „verdreckte“ Daten zu reinigen oder von Fehlern zu befreien. „Schmutzige“ Daten enthalten Duplikate, veraltete Werte usw. – das hängt jeweils vom konkreten Fall ab.

Möchtest du z. B. Feldinhalte aufspalten, weil das für die spätere Auswertung sinnvoll erscheint, sollte das hier erfolgen.

Nehmen wir an, du hast eine Datumsangabe „12.03.2004“ in einem Feld in der Form „040312“. In deiner späteren Auswertung möchtest du jedoch zwischen Jahr, Monat und Tag navigieren.