

Definition Data Warehouse

Ein Data Warehouse dient dazu, Daten aus unterschiedlichen internen und externen Quellen zusammenzuführen und zu speichern, um anschließend mithilfe unterschiedlicher Abfrage-, Analyse- und Auswertungsprogrammen neue Informationen zu gewinnen.

Worin besteht der Unterschied zwischen operativen & analytischen Daten?

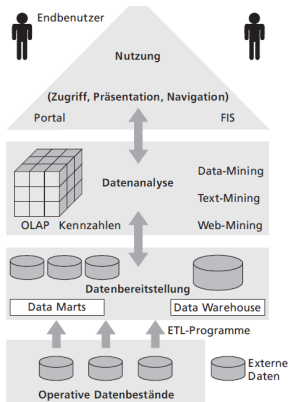
Kriterien	Daten für <i>operative</i> Anwendungen	Daten für <i>analytische</i> Anwendungen
Zweck	Unterstützung und Abwicklung operativer Geschäftsvorfälle	Informationen für das Management, Unterstützung von Entscheidungen, themenorientiert
Inhalt	detaillierte, aktuelle Daten über Geschäftsvorfälle, zeitpunktorientiert	verdichtete und bereinigte Daten, historische und zum Teil zukünftige Daten, zeitraumorientiert
Aktualität	hoch (Online, Realtime)	meist keine Tagesaktualität
Modellierung	Altdatenbestand oft nicht modelliert	themenbezogen modelliert, standardisiert, endbenutzertauglich
Zustand	redundant, inkonsistent, i. d. R. normalisiert	konsistent modelliert, kontrollierte Redundanzen, denormalisiert
Änderungen	laufend	automatische Fortschreibung, Beständigkeit des einmal übernommenen Datenbestandes
Abfragen	strukturiert, vordefiniert	Ad-hoc-Abfragen für komplexe, ständig wechselnde Fragestellungen, vordefinierte Standardauswertungen

Ziele eines Data Warehouse?

- Informationen für das Management
- Unterstützung von Entscheidungen
- zusammenführen unterschiedlicher Daten aus operativen Anwendungssystemen
- es werden (un-)strukturierte Daten übernommen
- Veränderung, Aggregation der Daten

Aufbau analytischer Informationssysteme

- **Zentrales DWH** enthält eine von den operativen Systemen isolierte Datenbank
- **Data Mart** ist ein subjektspezifisches oder abteilungsspezifisches DWH; entweder Datenbestände gleichzeitig an mehreren Orten schneller bereitzustellen oder einzelne Fachabteilungen spezifische Daten zu liefern



Unterschied Data Warehouse & Data Mart

Merkmale	Data Mart	Data Warehouse
Philosophie	anwendungsorientiert	anwendungsneutral
Adressat der Datenbereitstellung	Abteilung	Unternehmen
Vorherrschende Datenbanktechnologie	multidimensional	relational
Granularität (Detaillierungsgrad der gespeicherten Daten)	niedrig	hoch
Datenmenge	niedrig	hoch
Menge historischer Daten	niedrig	hoch
Optimierungsziel	Abfragegeschwindigkeit	Datenmenge
Anzahl	mehrere	eins bzw. sehr wenige
Datenmodell	je nach Data Mart verschieden	unternehmensweit

Ablauf ETL?

- Analyse und Dokumentation operativer und externer Datenquellen
- Extrahieren der ausgewählten Daten
- Transformation operativer Daten
- Bereinigung transformierter Daten
- periodisches Laden der Daten ins DWH

Extraktion

Unter Extraktion versteht man die Selektion der Daten aus den (zumeist) operativen Datenquellen und ihre anschließende Speicherung in einen Arbeitsbereich des DWH (Staging Area). Hier werden die Daten zwischengespeichert und transformiert bzw. bereinigt und im Anschluss in das DWH übertragen.

Wann wird die Extraktion durchgeführt?

- Periodisch
- Anfrage
- Ereignisgesteuert (wenn z.B. Werte unterschritten werden)
- Sofort (DWH hat die gleiche Aktualität wie die operativen Systeme)

Transformation

Transformation findet in der s.g. Staging Area statt und bereinigt bzw. transformiert die Quelldaten in das gewünschte Zielformat.

Qualitätsmängel der Quelldaten

- inkorrekte Daten (Eingabe-/Verarbeitungsfehler)
- logisch widersprüchliche Daten
- unvollständige, ungenaue, zu grobe Daten
- redundante Daten
- uneinheitliche Daten
- veraltete Daten
- irrelevante Daten
- unverständliche Daten (wegen qualitativ mangelhafter Metadaten)

Verfahren:

- Bereinigung
- Harmonisierung (betriebswirtschaftlich: Codierung, Schlüssel, Attribute)
- Verdichtung (für Analysezwecke aggregiert werden → Regionalzahlen usw.)
- Anreicherung (Ergänzung um errechnete Kennzahlen)

Bereinigung - Was ist zu beachten?

- Muss-Feld?
- Plausibilitätsprüfung bei der Eingabe?
- Wird das Feld gemäß der ursprünglichen Bestimmung genutzt?
- Wurde das Datenfeld nachträglich aufgenommen? (fehlt bei älteren Daten dann)
- Existieren konkrete Änderungspläne für die operativen Daten?

Daten-Mängel

Es werden **syntaktische** und **semantische** Mängel unterschieden.

Bereinigung	1. Klasse: Automatische Erkennung und Korrektur	2. Klasse: Automatische Erkennung und manuelle Korrektur	3. Klasse: Manuelle Erkennung und manuelle Korrektur
Syntaktische Mängel	Formatanpassungen	Erkennbare Formatinkompatibilitäten	–
Semantische Mängel	Fehlende Datenwerte	Ausreißerwerte, unstimmige Werte	Unerkannte semantische Fehler in operativen Datenquellen

Harmonisierung - Was wird getan?

- Vereinheitlichung unterschiedlicher Codierungen (z.B. männlich, m, 1, weiblich, w, 0)
- Synonyme und Homonymen (unterschiedliche Attributnamen mit gleicher Bedeutung z.B. vorname, vname, firstname)
- Harmonisierung von Schlüsseln und Kennzahlen

Verdichtung

Es werden Daten im DWH (Staging Area) auf verschiedenen Stufen aufsummiert.



Anreicherung

Es werden Berechnungen durchgeführt, die zusammen mit den übrigen analytischen Daten gespeichert werden, d.h. es werden konkrete Kennzahlen ermittelt basierend auf einem gegebenen Kennzahlensystem (z.B. DuPont-Schema \rightarrow ROI)

Vorteile der Anreicherung sind:

- kürzere Antwortzeiten bei späteren Anfragen da es sich um vorberechnete Werte handelt
- hohe Datenkonsistenz, da sie nach einem einheitlichen Algorithmus berechnet werden

Laden

Es wird unterschieden zwischen:

- Initialem Füllen aus den operativen Datenbanken
- Zyklischer Aktualisierung, neue Werte werden ergänzt, alte archiviert

Wenn die Daten zyklisch übernommen werden kann dies als:

- Kompletter Abzug (einfach aber zeitaufwendig)
- jeweilige Änderungen (geringe Datenmenge, aufwendig das Delta zu ermitteln, nur der letzte Stand wird ermittelt)
- Auswahl protokollierter Datenbanktransaktionen (auch Änderungen innerhalb des Deltas erfasst werden)

geschehen.

Metadaten

Metadaten sind Daten über Daten und enthalten Hintergrundinformationen über die im DWH gespeicherten Werte. Sie geben Aufschluss über:

- Umfang der verfügbaren Daten
- Datenstruktur und Beziehungen (Relationen)
- Herkunft der operativen Daten
- Speicherort im DWH
- Formate
- Zugriffsberechtigungen

In der Metadatenbank wird festgehalten:

- Welche Daten woher kommen
- Wie sie aufbereitet und verdichtet werden
- Wo sie gespeichert werden
- Welcher Anwender auf welche Daten Zugriff erhält

Archivierung

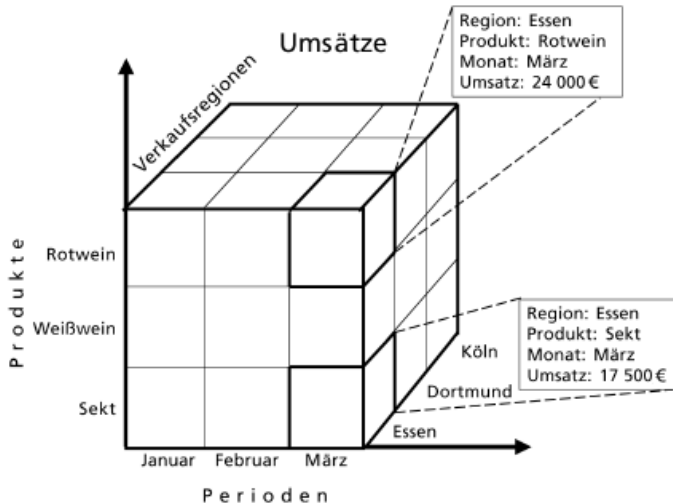
Es wird zwischen der:

- Datenarchivierung (auslagern auf Offlinedatenträger nach Zeit)
- Datensicherung (Dienen zur Wiederherstellung des DWH)

unterschieden.

OLAP

Der Begriff steht für Online Analytical Processing und umfasst alle Formen der *mehrdimensionalen Datenanalyse*. Im Focus stehen betriebswirtschaftliche Kennzahlen. Die Mehrdimensionalität wird durch s.g. *Datenwürfel* veranschaulicht.



Unterschied OLAP/TLTP

Merkmal	OLTP Operative IT-Systeme	OLAP Data Warehouse
Typische Datenstruktur	flache Tabellen	multidimensionale Strukturen
Datenmanipulation	<ul style="list-style-type: none">– Erfassung einzelner Datensätze– Update/Einfügen von Datensätzen zulässig	<ul style="list-style-type: none">– spezifische Analyse großer Datenmengen– nur lesender Zugriff möglich
Aktualisierung durch	Transaktionen	Batch-Läufe
Datenquelle	intern	intern und extern
Datenmenge pro Transaktion	klein	sehr umfangreich
Typische Betrachtungsebene	detailliert	aggregiert
Systemlast	vorhersehbar	ad hoc
geforderte Antwortzeit	2 – 3 Sekunden	mehrere Sekunden bis Minuten

Anforderungen an OLAP

- Mehrdimensionale konzeptionelle Sicht auf die Daten (Zeit, Produktgruppe, Region, Person, usw.)
- Transparenz (Anwender müssen keine technischen Details kennen)
- Zugriffsmöglichkeiten (auf möglichst viele heterogene und interne/externe Datenquellen)
- Stabile Antwortzeiten (möglichst schnell und vor allem gleichbleibend)
- Client-/Server-Architektur
- Gleichrangige Dimensionen
- Dynamische Handhabung dünn besetzter Matrizen (effiziente Speicherung trotz Lücken)
- Mehrbenutzerfähigkeit
- Unbeschränkt dimensionsübergreifende Operationen
- Intuitive Datenanalyse
- Flexibles Berichtswesen (Dokumentation in Form von Berichten und Grafiken)
- Unbegrenzte Anzahl von Dimensionen/Aggregationsstufen

Kritik: Die unscharfe Trennung zwischen fachlich-konzeptionellen Anforderungen und technischer Realisierung.

Alternativ **FASMI**:

- Fast (Antwortzeit max. 20s)
- Analysis (Anwender ohne technisches Wissen müssen auswerten können)
- Shared (Mehrbenutzer)
- Multidimensional
- Information (sämtliche benötigten Informationen können geliefert werden)

Mehrdimensionalität

Die Anzahl der Dimensionen lässt sich mit der Fakultät der Dimensionen berechnen:

$$2 \text{ Dim} = 1 \times 2 = 2 \text{ Sichten}$$

$$3 \text{ Dim} = 1 \times 2 \times 3 = 6 \text{ Sichten}$$

$$4 \text{ Dim} = 1 \times 2 \times 3 \times 4 = 24 \text{ Sichten}$$

Die verschiedenen Betrachtungsmöglichkeiten werden auch als **Slice and Dice** bezeichnet. *Slice* bedeutet das Herausschneiden von Scheiben aus dem Würfel. *Dice* bedeutet die Bildung von kleinen Würfeln aus dem Gesamtwürfel zur Einschränkung auf einen Wert bzw. Wertebereich.

Mittels **Drill down** ist es möglich von einer bestehenden Verdichtungsebene auf eine detaillierte Ebene zu wechseln. **Drill up** wechselt von einer Ebene auf eine verdichtere Ebene. **Drill across** ermöglicht zu einem anderen Wert auf der selben Ebene zu wechseln.

Wie können Daten verdichtet werden?

Wie die Daten verdichtet werden können hängt unmittelbar vom Typ ab.

- **Additive Daten** lassen sich beliebig aufsummieren (Umsatz in Kombination mit Produkten und Regionen)
- **Semiadditive Daten** lassen sich nicht über alle Dimensionen aufaddieren (z.B. bei Zeiträumen und Lagerbeständen)
- **Nichtadditive Daten** lassen keine sinnvolle Aufsummierung zu (Anteilswerte)

Modellierung

Die **multidimensionale Modellierung** ist in der Lage die konzeptionellen Datenstrukturen in physische Datenbank-/Speicherstrukturen umzusetzen. Im Falle eine *dünn besetzten Matrix* wird Speicherplatz vergeudet, da für nicht existente Werte gleich viel Speicher verbraucht wird, als wenn der Wert existiert.

Die **relationale Modellierung** setzt die mehrdimensionale Datenstruktur in einer relationalen Datenbankstruktur ab. Es wird dabei zwischen:

- Fakten (quantitative Kerndaten)
- Dimension (beschreibende Daten, kann beliebig viele Ausprägungen haben)

unterschieden.

Star-Schema

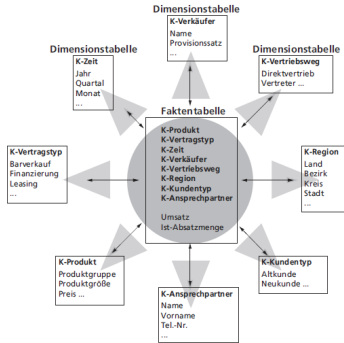
Für jede Dimension wird eine Tabelle eingerichtet. Die Dimensionstabellen sind nicht miteinander verknüpft, sondern stehen nur über die Faktentabelle miteinander in Beziehung. Der Primärschlüssel in der Faktentabelle setzt sich zusammen aus den Primärschlüsseln aller Dimensionstabellen.

Vorteil:

- einfaches, intuitive Datenmodell
- benötigt nur wenige Join-Operationen
- wenige Tabellen benötigt

Nachteil:

- durch die große Anzahl der Verknüpfungsmöglichkeiten kann es zu Performan-
ceproblemen kommen



Snowflake-Schema

Dieses Schema ist eine Erweiterung des Star-Schemas und wird durch die Normalisierung entwickelt.

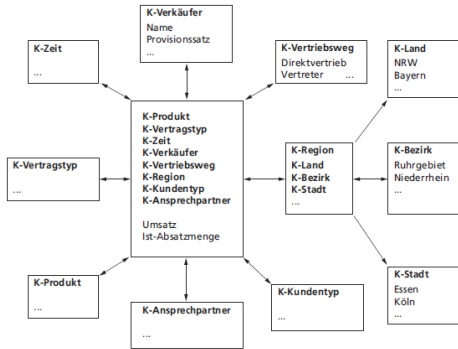
Vorteile:

- erleichtert die Aggregation
- keine Redundanzen

Nachteile

- höhere Anzahl JOINS
- höhere Anzahl Tabellen
- komplexere SQL-Abfragen

Achtung: Rückseite von # 26 ist zu voll.



Was ist Data-Mining?

Data-Mining bedeutet die Erforschung und Analyse großer Datenbestände hinsichtlich sinnvoller Muster, Zusammenhänge oder Regeln, mit dem Ziel sinnvolle Informationen heraus zu filtern. Die Besonderheit ist, das verborgene und bisher unbekannte Sachverhalte ermittelt werden sollen.

Eine alternative Vorgehensweise ist das *Hypothesis Testing* bei dem Annahmen bestätigt oder widerlegt werden. Im Idealfall generiert das Data-Mining-System die Hypothesen selbst und überprüft diese indem die Datenbank nach Mustern und Zusammenhängen durchsucht werden.

Merkmale/Hauptanwendungsgebiete von Data-Mining

Merkmale:

- Selbstständigkeit (soll autonom arbeiten)
- Mustergenerierung
- Interessantheit (Ergebnisse müssen von Bedeutung sein)

Anwendungsgebiete:

- Vorhersagen unbekannter oder zukünftiger Werte
- Entdeckung bislang unbekannter Zusammenhänge
- automatisierte Analyse und Klassifizierung

Alternative Verfahren zur Datenmustererkennung

Beim *Text-Mining* wird in unstrukturierten Datenbeständen gesucht. Problematisch ist die geringe innere Struktur der durchsuchten Dokumente.

Beim *Web-Mining* werden Daten durchsucht die durch die Nutzung des WWW entstanden sind (z.B. Logdateien). Problematisch ist das viele Zugriff über Proxies erfolgen, was eine Anonymisierung zur Folge hat.

Methoden des Data-Mining

Die *Klassifikation* ordnet einzelne Objekte (Datensätze) vorgegeben Klassen zu. Diese Zuordnung erfolgt anhand bestimmter und bekannten Eigenschaften (Attributwerten).

Die *Clusterbildung* soll ähnliche Objekte zu Gruppen (Segmente) zusammenfassen. Die Differenz zur Klassifikation besteht darin, dass die verwendeten Merkmale und die Bereiche ihrer Ausprägung nicht im Voraus festgelegt werden.

Die *Assoziation* sucht nach Verwandtschaft und Korrelation innerhalb des Datenbestands. Das Verfahren ermittelt Regeln über Zusammenhänge zwischen einzelnen Merkmalen und bewertet die gefundenen Muster mit Wahrscheinlichkeiten. Es entstehen s.g. Wenn-Dann-Regeln.

Welche Faktoren beeinflussen die Übersichtlichkeit von
Informationen?

- Format (Hoch-/Querformat → Präsentation/Fließtext)
- Gliederung (Inhaltsverzeichnisse, Absätze, Überschriften)
- Darstellungsformen (Schaubilder)

Zur Visualisierung werden häufig auch s.g. Metaphern verwendet.

- Landkarten-Metapher (Typisch im Marketing und Vertrieb)
- Wetterkarten-Metapher (Regionale Trends und Entwicklungen)
- Organigramm-Metapher (Kennzahlensysteme)
- Zeitungs-Metapher (Stark strukturierte Informationen)
- Unternehmensleitstand-Metapher (aktuelle Stati, Dashboard)

Welche Funktionen haben Kennzahlen?

- Vorgabefunktion als Zielgröße für Teilbereiche
- Anregungsfunktion zur Erkennung von Auffälligkeiten und Veränderungen
- Steuerungsfunktion zur Vereinfachung von Steuerungsprozessen
- Kontrollfunktionen zur laufenden Erfassung von Soll/Ist-Abweichungen
- Vergleichsfunktion (Benchmarking)
- Informationsfunktion in Frühwarnsystemen oder Berichtserstattung

Wie können Kennzahlen unterschieden werden?

Nach der *Reichweite*: unternehmensspezifisch, branchenbezogen, allgemeingültig

Nach der *zeitlichen* Gültigkeit: vergangenheits-, gegenwarts-, zukunftsbezogen

Nach der *Art*: statisch (Summen, Differenzen, Mittelwerte), relativ (Gliederungszahlen, Beziehungszahlen, Indexzahlenx)

Welche Berichte können unterschieden werden und was ist essentiell für diese?

Es wird zwischen *außerbetriebliche* (Geschäftsbereich) und *innerbetrieblichen* Berichten unterschieden.

Für beide ist es entscheidend, dass der Bericht:

- zum richtigen Zeitpunkt
- in der richtigen Detaillierung
- in der richtigen Form
- für die richtige Adressatengruppe

zur Verfügung gestellt wird.

Welche Berichtsarten können unterschieden werden?

- Standardbericht: periodisch, gleicher Aufbau
- Bedarfsbericht: Inhalt, Zeiträume und Formate sind Parameterisierbar und er wird fallweise angefordert
- Ausnahmeberichte: zeigen Abweichungen vom Normalzustand

Was ist ein Führungsinformationssystem (FIS)?

FIS werden auch als Executive Information Systems (EIS) bezeichnet. Für die Zielgruppe ist eine intuitive Bedienung eine managementgerechte Informationspräsentation notwendig. FIS werden unternehmensspezifisch aufgebaut. Die Anwender sollten selbstständig in der Lage sein das System zu bedienen.

Exception Reporting soll frühzeitig auf Abweichungen vom Sollzustand aufmerksam machen.

Historie Computerarbeit

Seit der Einführung der Computer in den Unternehmen hat eine Dezentralisierung stattgefunden. Zuerst musste der Computer-Nutzer nur einfache Masken bedienen, mit der Einführung des PCs in den Büros jedoch musste er zunehmend selbstverantwortlich mit dem Computer arbeiten. Speziell seit der Einführung von Tabellenkalkulationsprogrammen stehen ihm zum Teil eher Werkzeuge zum Selbstlösen von Aufgaben zur Verfügung als fertige Lösungen. Mit der Vernetzung der PCs und der wachsenden Bedeutung des Internets sowie der daraus resultierenden Gefahren stieg zudem die Verantwortung, da Fehlbedienungen an einem Rechner auch Auswirkungen auf andere Arbeitsplätze haben können.

Organisation von IT-Abteilungen

Innerhalb der Organisation eines Unternehmens unterscheidet man die Fachabteilungen sowie die IT-Abteilung(en). Für die organisatorische Eingliederung der IT-Abteilungen gibt es sehr unterschiedliche Varianten, die unter anderem von der Größe und technischen Ausrichtung des Unternehmens abhängen. Bei kleineren, eher nichttechnischen Unternehmen wird entweder ein vorhandener Mitarbeiter teilweise mit solchen Aufgaben betraut oder auch ein kleiner IT-Bereich geschaffen. Größere und grundlegende IT-Aufgaben werden in der Regel nach außen vergeben. Größere Unternehmen verfügen fast immer über eine eigene, zum Teil intern weiter strukturierte IT-Abteilung. Das schließt jedoch nicht aus, dass auch hier einzelne Tätigkeitsbereiche extern vergeben werden (Outsourcing). Bei sehr großen oder sehr technikorientierten Unternehmen werden oft zusätzlich bereichsspezifische IT-Abteilungen installiert.

Welcher rechtliche Rahmen gilt?

Unternehmen müssen bei der Nutzung von Computern zahlreiche rechtliche Vorschriften beachten. Die wichtigsten sind:

- Vermeidung von illegalen Software-Kopien
- Einbeziehung des Betriebsrates beim Einsatz von Computern
- Einhaltung der Vorschriften des Datenschutzes
- Langfristige Aufbewahrung und das Zugänglichmachen steuerlich relevanter gespeicherter Daten
- Vorbeugende Vermeidung rechtswidriger Handlungen der Mitarbeiter sowie Schädigung Dritter
- Regelmäßige Datensicherungen nach dem Stand der Technik

Was ist bei der Auswahl von Software zu beachten?

Bei der Auswahl von Office-Systemen sind folgende Komponenten zu berücksichtigen, die sich gegenseitig beeinflussen bzw. bedingen:

- Hardware (vor allem PC-Systeme und Mac)
- Betriebssystem (vor allem Windows, Linux und Mac OS)
- Anwendungssoftware

Bei der Entscheidung für ein System sind folgende Kriterien anzulegen: Sofern vorhandene Hard- und Software oder auch Schnittstellen zu anderen Stellen berücksichtigt werden müssen, ist unbedingt auf entsprechende **Kompatibilität** zu achten.

Beim **Leistungsumfang** sind zunächst die Systeme auszuschließen, die die K.-o.-Kriterien nicht erfüllen. Bei den verbleibenden Produkten empfiehlt sich eine Nutzwertanalyse, bei der die Leistungen in einzelnen Bereichen mit ihrer Wichtigkeit für die geplante Anwendung gewichtet werden. Neben den Eigenschaften der Produkte selbst ist auch deren Marktposition und die Unterstützung durch Dritte zu beachten.

Bei den Kosten ist nicht der Anschaffungspreis entscheidend. Vielmehr müssen die **Total Cost of Ownership** berücksichtigt werden, die unter anderem auch Arbeitskosten durch Installation, Schulung und Umstellung vorhandener Dokumente o.Ä. beinhalten.

Wie können Office-Anwendungen integriert werden?

Die Integration von Officeanwendungen bezieht sich auf zwei getrennte Aspekte: Zum einen geht es um die **Integration der Benutzeroberfläche**. Die einzelnen Programme besitzen dabei nicht nur eine einheitliche Gestaltung und Menüstruktur, sondern bilden im Idealfall eine einheitliche Applikation, in der die Grenzen zwischen den einzelnen Anwendungen für den Benutzer weitgehend aufgehoben sind.

Zum anderen geht es um die **Integration der Dokumente** unterschiedlicher Applikationen auf Daten- bzw. Dateiebene. In Windows wird dazu OLE verwendet, das zwei Integrationstechniken umfasst:

Beim **Verknüpfen** (Linking) bleiben die Dokumente in getrennten Dateien gespeichert und es wird im Zieldokument nur eine Referenz (Link) auf die Datei des Quelldokuments gespeichert. Vorteile sind dabei der geringe Speicherplatz und die ständige Aktualität. Nachteilig sind unbeabsichtigte Auswirkungen, wenn die Dokumente unabhängig voneinander inhaltlich oder bezüglich ihres Speicherplatzes verändert werden.

Beim **Einbetten** (Embedding) wird eine Kopie des Quelldokuments in die Datei des Zieldokuments eingefügt. Der Vorteil besteht darin, dass auf diese Weise Änderungen mit unbeabsichtigten Auswirkungen auf das Zieldokument ausgeschlossen sind und bei Dateioperationen die Einheit der Dokumente immer erhalten bleibt. Als Nachteil sind der (zum Teil zusätzliche) Speicherplatz innerhalb des Zieldokuments sowie die fehlende automatische Anpassung bei Aktualisierungen des originalen Quelldokuments zu nennen.

Wie können Office-Anwendungen konfiguriert werden?

Heutige Officeprogramme erlauben dem Nutzer vielfältige Konfigurationsmöglichkeiten, mit denen er das Programm seinen individuellen Bedürfnissen anpassen kann. Dabei unterscheidet man zwei Gruppen von Einstellungen:

Programmeinstellungen beziehen sich auf das Aussehen bzw. Verhalten des Programms selbst. Beispiele sind das Aussehen von Symbolleisten bzw. Menübändern sowie der Speicherort von Dateien.

In **Dokumentvorlagen** werden Einstellungen gespeichert, die sich auf neu erstellte Dokumente auswirken. Typische Vorgaben sind hier die Schriftart und Einstellungen für Ränder. Neben einer Standardvorlage, die automatisch für neu erstellte Dokumente verwendet wird, kann der Benutzer eine beliebige Anzahl weiterer Dokumentvorlagen für verschiedene Zwecke erstellen und beim Erzeugen eines neuen Dokuments auswählen.

Wie läuft das mit der Wartung?

Viele Programme müssen regelmäßig oder aufgrund bestimmter äußerer Ereignisse aktualisiert werden, um noch ausreichend nutzbar zu bleiben. Solche Aktualisierungen müssen vom Hersteller bezogen werden, der diese auf unterschiedliche Art zur Verfügung stellen kann:

- Fehlerbeseitigungen und das Schließen von Sicherheitslücken sollten grundsätzlich kostenlos erfolgen (in der Regel per Download aus dem Internet).
- Regelmäßige Aktualisierungen und Funktionserweiterungen werden – insbesondere im professionellen Bereich – über Wartungsverträge angeboten. Das gilt bei Unternehmen inzwischen auch für Office-Produkte.
- Mit dem Erscheinen einer neuen Version eines Programms stehen zum Teil neue Funktionen zur Verfügung, die einen Umstieg sinnvoll erscheinen lassen. Hier ist – sofern nicht über einen entsprechenden Wartungsvertrag abgedeckt – der Neukauf dieser Version notwendig, wobei es üblicherweise einen Rabatt für Besitzer der Vorversion(en) gibt.

Was ist individuelle Informationsverarbeitung?

Unter dem Begriff der Individuellen Informationsverarbeitung (IIV) versteht man, dass der Benutzer an seinem lokalen Arbeitsplatz mit dort verfügbaren Werkzeugen selbstständig Endbenutzeraufgaben mit arbeitsplatzbezogenen Daten wahrnimmt. Trotz der Bezeichnung individuell ist dies im betrieblichen Bereich in die Informations-Infrastruktur des Unternehmens eingebunden und damit stark standardisiert. Etwas einschränkender ist der Begriff der Officeanwendungen, der sich auf die Anwendungen im Büro beschränkt, wie sie über alle Branchen und Funktionen hinweg am Computer anfallen. Typische Programme, die auch den Kern von Officepaketen bilden, sind Textverarbeitung, Tabellenkalkulation und Präsentationsgrafik.

Was wird unter kooperativer Informationsverarbeitung
verstanden?

Die computerunterstützte Gruppenarbeit wird allgemein mit dem Begriff Computer Supported Cooperative Work (CSCW) bezeichnet. Unter Groupware versteht man die Systeme, mit denen dies unterstützt wird. Es werden meist zwei Schwerpunkte unterschieden:

Workflow Computing betrachtet arbeitsteilige Prozesse mit einer Abfolge einzelner Tätigkeiten, die meist von verschiedenen Personen in einer definierten Reihenfolge abgearbeitet werden. **Workgroup Computing** ist weniger strukturiert und stellt die Gruppe und ihre Kooperationsbeziehungen in den Mittelpunkt.

Eine weitere Unterscheidung bezieht sich einerseits auf den **Ort**, an dem sich die Teilnehmer aufhalten (gleicher Ort oder verschiedene Orte), andererseits auf den **Zeitbezug** der gemeinsamen Arbeit (synchron oder asynchron). Für den Austausch von Informationen existiert innerhalb eines Betriebes heute fast immer ein lokales Netzwerk (LAN), das die PCs miteinander mit hoher Geschwindigkeit verbindet.

Was ist ein Dokumentenmanagementsystem?

Dokumentenmanagementsysteme (DMS) dienen dazu, Dokumente in elektronischer Form zu erfassen, bereitzustellen, weiterzuleiten und dauerhaft zu archivieren.

Beim **Erfassen** der Dokumente werden diese in ein geeignetes elektronisches Format überführt, wobei man kodierte und nicht kodierte Informationen unterscheidet.

Bei der **Indizierung** werden den Dokumenten für die Speicherung Metainformationen hinzugefügt.

Bei der **Speicherung** der Daten ist ein schneller Zugriff ohne menschliches Zutun sicherzustellen. Wichtig ist außerdem die Vergabe angemessener Zugriffsrechte sowie die Absicherung gegen Datenverlust.

Mit der **Recherche** werden Dokumente nach bestimmten Kriterien gesucht. Dies geschieht über die bei der Indizierung vergebenen Metainformationen oder eine Volltextsuche.

Bei der **Verwendung** der gefundenen Dokumente kann man die Verarbeitung mit dem Originalprogramm des jeweiligen Formats oder einem speziellen Viewer unterscheiden.

Welche Arten von Datensicherung gibt es?

Unter Datensicherung (Backup) versteht man das Speichern von Daten, die noch aktuell auf dem Arbeitssystem gehalten werden, auf einem Sicherungsmedium. Datensicherungen müssen regelmäßig durchgeführt und die Sicherungsmedien besonders geschützt verwahrt werden.

Bei der **Vollsicherung** werden sämtliche relevanten Daten zu einem Zeitpunkt gesichert.

Bei der **inkrementellen Sicherung** werden nur die Daten gesichert, die sich seit der letzten Sicherung geändert haben.

Bei der **differenziellen Sicherung** werden nur die Daten gesichert, die sich seit der letzten Vollsicherung geändert haben.

Kartenübersicht ANS08

#	Karte	Notizen
1	Definition Data Warehouse	
2	Worin besteht der Unterschied zwischen operativen & analytischen Daten?	
3	Ziele eines Data Warehouse?	
4	Aufbau analytischer Informationssysteme	
5	Unterschied Data Warehouse & Data Mart	
6	Ablauf ETL?	
7	Extraktion	
8	Wann wird die Extraktion durchgeführt?	
9	Transformation	
10	Qualitätsmängel der Quelldaten	
11	Bereinigung - Was ist zu beachten?	
12	Daten-Mängel	
13	Harmonisierung - Was wird getan?	
14	Verdichtung	
15	Anreicherung	
16	Laden	
17	Metadaten	
18	Archivierung	
19	OLAP	
20	Unterschied OLAP/TLTP	
21	Anforderungen an OLAP	
22	Mehrdimensionalität	
23	Wie können Daten verdichtet werden?	
24	Modellierung	
25	Star-Schema	
26	Snowflake-Schema	
27	Was ist Data-Mining?	
28	Merkmale/Hauptanwendungsgebiete von Data-Mining	
29	Alternative Verfahren zur Datenmustererkennung	
30	Methoden des Data-Mining	
31	Welche Faktoren beeinflussen die Übersichtlichkeit von Informationen?	
32	Welche Funktionen haben Kennzahlen?	
33	Wie können Kennzahlen unterschieden werden?	

#	Karte	Notizen
34	Welche Berichte können unterschieden werden und was ist essentiell für diese?	
35	Welche Berichtsarten können unterschieden werden?	
36	Was ist ein Führungsinformationssystem (FIS)?	
37	Historie Computerarbeit	
38	Organisation von IT-Abteilungen	
39	Welcher rechtliche Rahmen gilt?	
40	Was ist bei der Auswahl von Software zu beachten?	
41	Wie können Office-Anwendungen integriert werden?	
42	Wie können Office-Anwendungen konfiguriert werden?	
43	Wie läuft das mit der Wartung?	
44	Was ist individuelle Informationsverarbeitung?	
45	Was wird unter kooperativer Informationsverarbeitung verstanden?	
46	Was ist ein Dokumentenmanagementsystem?	
47	Welche Arten von Datensicherung gibt es?	