

# Zusammenfassung Data Warehousing FS2018

Alex Neher

March 8, 2018

## Contents

<b>1</b>	<b>Die Notwendigkeit von Data Warehouses</b>	<b>2</b>
1.1	Entscheidungsunterstützung (Buch S11) . . . . .	2
1.1.1	Expertensystem . . . . .	2
1.2	Ungenügen der "gängigen" Datenhaltung (Buch S13) . . . . .	3
1.3	Ungenügen der operativen Datenbanken für Entscheide (Buch S13) . . . . .	3
1.4	SQL-Abfragen für Management-Zwecke . . . . .	4
1.5	OLAP vs OLTP . . . . .	4
<b>2</b>	<b>Daten vs. Informationen vs. Wissen vs. Weisheit</b>	<b>5</b>
<b>3</b>	<b>Das Data Warehouse</b>	<b>5</b>
3.1	Definition Data-Warehouse (Buch S 33) . . . . .	6
3.2	Bestandteile eines Data-Warehouses . . . . .	7

# 1 Die Notwendigkeit von Data Warehouses

## 1.1 Entscheidungsunterstützung (Buch S11)

Es gibt vier Arten der Entscheidungsunterstützung:

**Modellbasiert:** z.B. Lineare Optimierung - Ein Mathematischer Ansatz basierend auf einem Modell  $\Rightarrow$  Abbildung der Realität

**Wissensbasiert:** z.B. Expertensysteme - Ansätze von Künstlicher Intelligenz

**Datenbasiert:** Basierend auf grossen Datenmengen  $\Rightarrow$  Data-Warehouse, OLAP oder Data-Mining

**KI:** Basierend auf Vorschlägen von Systemen, die Entscheidungen auf Basis von Daten und/oder gelernten Inhalten ( $\rightarrow$  Machine Learning)

### 1.1.1 Expertensystem

Ein Expertensystem (XPS oder ES) ist ein Computerprogramm, das Menschen bei der Lösung von komplexen Problemen wie ein menschlicher Experte unterstützen kann, indem es Handlungsempfehlungen aus einer Wissensbasis ableitet.

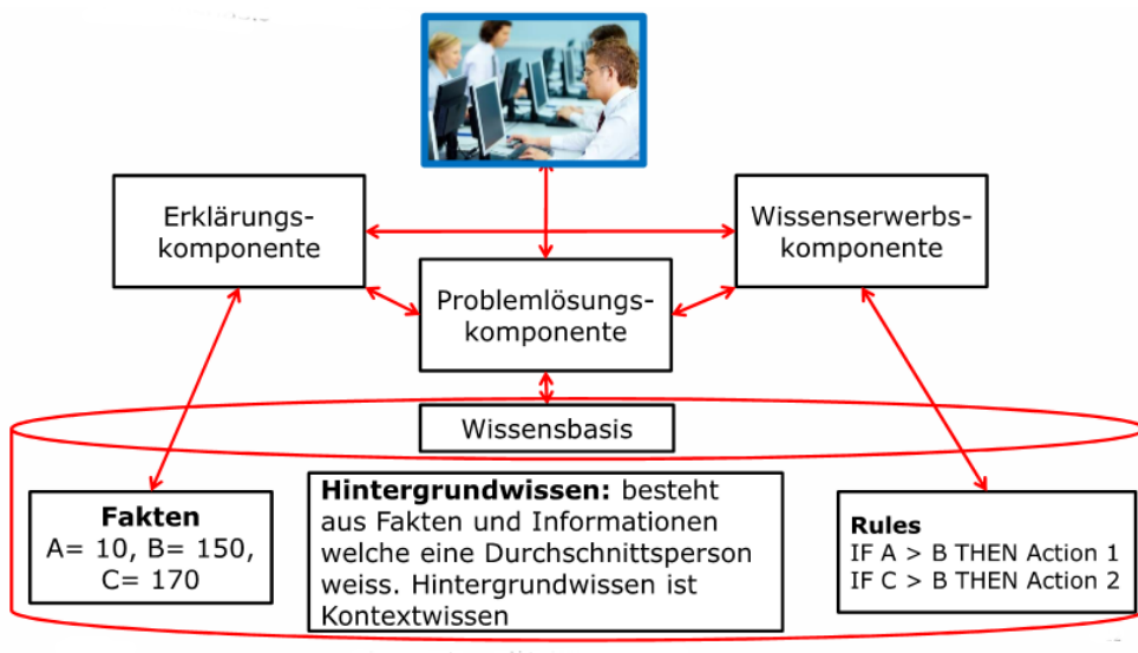


Figure 1: Beispiel eines Expertensystems

## 1.2 Ungenügen der "gängigen" Datenhaltung (Buch S13)

- Verschiedene Datenformate
- Verschiedene Werkzeuge
- Heterogenität der Daten

Technisch

Mainframe

DBMS

Flatfile

Logisch

Schemata

Formate

Darstellungen

Syntaktisch

Datum

Codierung

Währung

Qualitativ

Fehlende Werte

Falsche Werte

Doppelte Werte

Verfügbarkeit

Permanent

Periodisch

Temporär

Rechtlich

Datenschutz

Zugriffsverwaltung

Archivierung

→ Neuer Ansatz einer Datenaufbereitung muss her: **Homogenisierung**

## 1.3 Ungenügen der operativen Datenbanken für Entscheide (Buch S13)

"Reguläre" Datenbanken im Geschäftsumfeld sind zu fest mit gesellschaftsrelevanten Lese- und Schreiboperationen beschäftigt. Bei solchen Datenbanken spricht man von OLTP-System (Online Transactional Processing). Diese Datenbanken sind aus Performance-Gründen ziemlich schlecht geeignet für eine analytische, vorausschauende Bewirtschaftung.

Ausserdem liegen Daten in OLTP-Datenbanken meist in der 3. Normalform vor. Während dies eine sehr vernetzte und effiziente Art der Datenspeicherung ist, ist die 3. Normalform ein schlechtes Abbild des intuitiven Denkens eines Managers.

→ Neuer Ansatz einer Datenbank muss her: **analytische Datenbanken**

## 1.4 SQL-Abfragen für Management-Zwecke

Zusätzlich zu den vorhin genannten Gründen, sind Manager des SQL meist nicht mächtig. Sie wollen lieben "Drag and Drop" Interfaces, um sich ihre Daten "zusammenzuklicken" wie z.b. Microsoft Access.

→ Neuer Ansatz der Datenabfrage muss her: **OLAP**

## 1.5 OLAP vs OLTP

Merkmal	OLTP System	OLAP System
Ausrichtung auf	Programm, BWL Prozess	Mensch, Analyse
Zeitliche Reichweite	Taktisch	Strategisch
Entscheidungsstufe	Tief	Hoch
Zweck	Rationalisierung & Automatisierung	Planung & Entscheidungsfindung
Anwenderzahl	Hoch	Tief
Entscheidung	Deduktiv	Induktiv / Explorativ
Bewirtschaftung I	Ändernd	Befragend
Bewirtschaftung II	Auf Datensatzebene	Auf Aggregatzebene
Anwendungsmuster	Voraussehbar	Variierend
Befragungsmuster	Einfach	Komplex
Bearbeitung	Repetitiv	Ad hoc / unstrukturiert
Betriebliches Wissen	Verarbeitend	Generierend
Verteilungsgrad	Dezentral	Zentral
Performance-Bedarf	Durchgehend hoch	Variierend
Mehrbenutzersynchronisation	Hoch	Tief bis kein
Optimierung	Schneller Insert & Delete	Schnelles Lesen
Transaktionsdurchsatz	Hoch	Tief
Transaktionsdauer	Kurze Mutationen weniger Tupel	Lange Abfragen vieler Tupel
Abfragen	Häufige, einfache Abfragen	Weniger häufige, komplexe Anfragen
Antwortzeiten	(Mili)sekunden	Sekunden, Minuten, Stunden
Endbenutzerwerkzeug-Hersteller	DB-Hersteller	Marketing
Zeitbezug	Aktuell	Historisch
Zeitdimension	Zeitpunkt	Zeitraum
Beständigkeit	Dynamisch	Statisch
Granularität	Fein	Grob
Datenbestand	Vollständig	Lückenhaft
Redundanz	Normalisiert	Denormalisiert
Datenqualität / Aussagekraft	Tief	Hoch
Aufbereitung	Anwendungsneutral	Analyseorientiert
Aktualisierung	Laufend	Periodisch
Verarbeitungseinheit	Kein	Grob
Verteilungsgrad	Dezentral	Zentral
Datenquelle	Aktuelle Unternehmensdaten	Interne & externe Daten

## 2 Daten vs. Informationen vs. Wissen vs. Weisheit

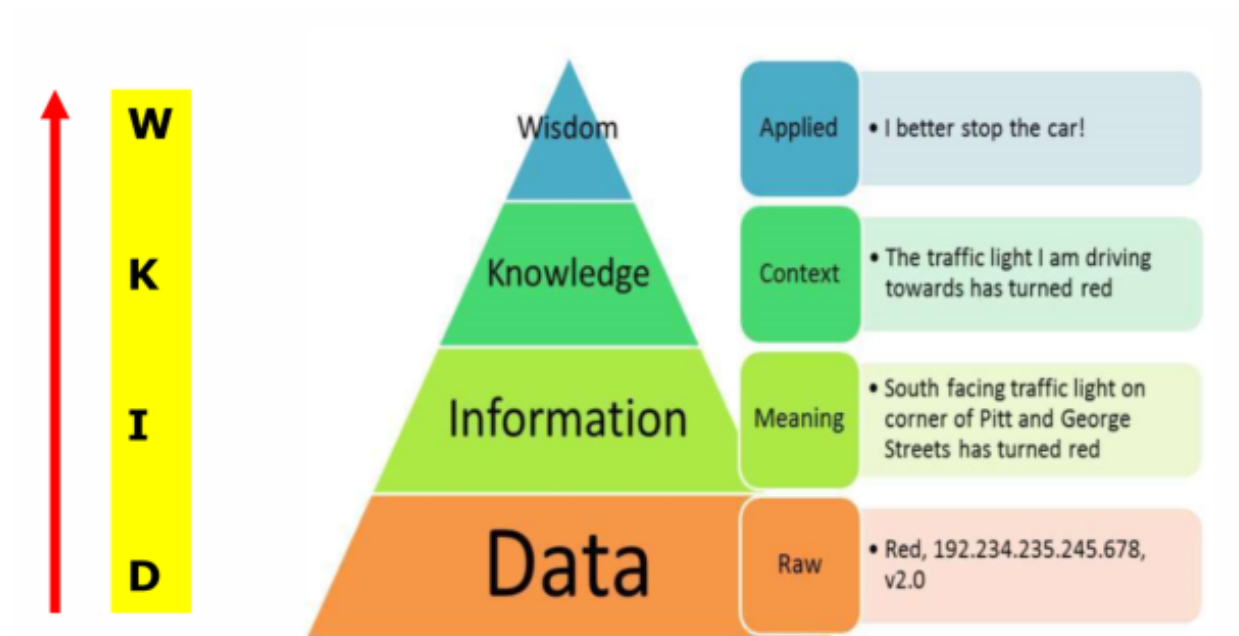


Figure 2: DIKW-Pyramid

Bei Entscheidungsfindungen muss unterschieden werden zwischen

- Daten
- Informationen
- Wissen
- Weisheit

Sprachregion	Jahr	Quartal	Kursgruppe	Teilnehmende
D	2011	q1	Informatik	1453
D	2011	q1	Sport und Freizeit	4783
F	2011	q1	Informatik	221
F	2011	q1	Sport und Freizeit	652
D	2011	q2	Informatik	1556
D	2011	q2	Sport und Freizeit	4512
F	2011	q2	Informatik	249
F	2011	q2	Sport und Freizeit	623

Figure 3: Beispiel Daten

**Daten** sind das, was in Datenbanken oder Excel-Tabellen gespeichert wird: Rohe Daten (z.B. Bild 3). Damit kann nicht viel angefangen werden. Diese Daten müssen zuerst gefiltert, sortiert, zusammengefasst etc. werden. Anschliessend hat man **Informationen** (z.B. Bild ?? )

## 3 Das Data Warehouse

In einer optimalen Welt würden Daten "perfekt" abgelegt werden, leicht zugänglich, platzsparend, sicher und für verschiedene Zwecke nützlich. Da wir aber leider nicht in einer optimalen Welt leben, ist dies nicht der Fall. (Buch S32)

Daten sind in der Praxis meist nicht optimal abgelegt. Daten existieren meist

- in unterschiedlichen Formaten (Excel, Access, DB etc)
- in unterschiedlichen DB-Strukturen
- in unterschiedlichen IT-Architekturen und -Systemen. Meist auch uralte Legacy-Systeme (Wie z.B. Cobol)
- zeit-aktuell und dynamisch
- zu detailliert und feingranular für wirk-same Management-Abfragen
- in einem Format, das für Änderungstransaktionen optimiert wurde (z.B. 3. Normalform)
- mit begrenzten Zugriffsrechten (z.B. aus Security-Gründen)
- in einem schlecht verfügbaren Zustand (Legacy-System, proprietäres Format, Security-Gründe)
- in einem Format, welches komplexe SQL-Queries verlangt, um an Informationen oder Wissen zu gelangen.

→ Lösung: **Data-Warehouse**

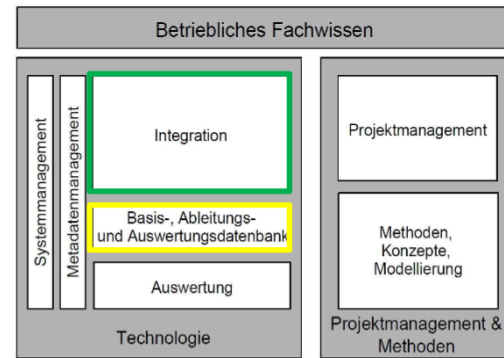


Figure 4: Aufbau eines Datawarehouses

### 3.1 Definition Data-Warehouse (Buch S 33)

A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains historic data derived from transaction data, but can include data from other sources. Data warehouses separate analysis workload from transaction workload and enable an organisation to consolidate data from several sources. - Oracle corp: Data warehousing Guide 11g (2007)

A data warehouse is a organisation's data with a corporate wide scope for use in decision support and informational applications. - IBM Corp: Enterprise Data Warehousing with DB2.9 - Redbook (2008)

Zusammengefasst kann man also sagen, ein Data Warehouse ist eine Datenbank, welche nicht (ausschliesslich) zur Speicherung von Informationen genutzt wird, sondern hauptsächlich als Hilfsmittel bei Entscheidungen eingesetzt wird (→ Experten-Systeme)

### 3.2 Bestandteile eines Data-Warehouses

**SSRS:** SQL Server Reporting Services

**SSAS:** SQL Server Analysis Services

**SSIS:** SQL Server Integration Services

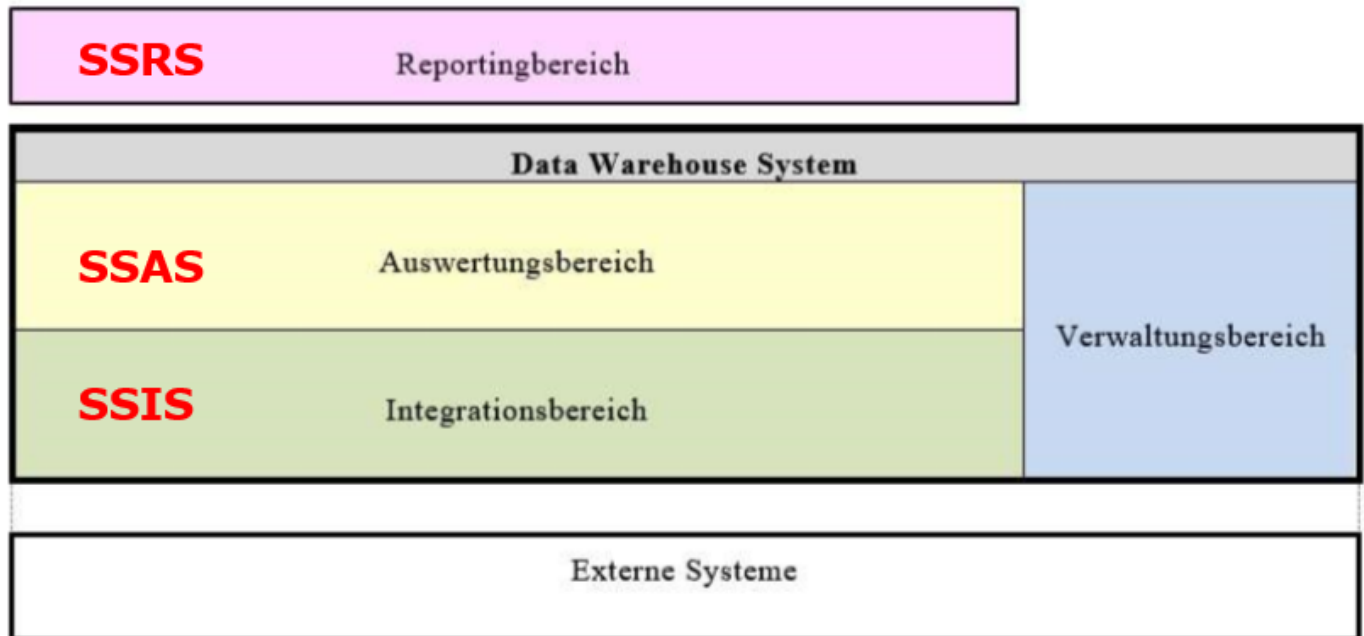


Figure 5: Bestandteile eines Data-Warehouses