



Von der Lagerhalle zum Markt

DWH-Architektur Grundlagen

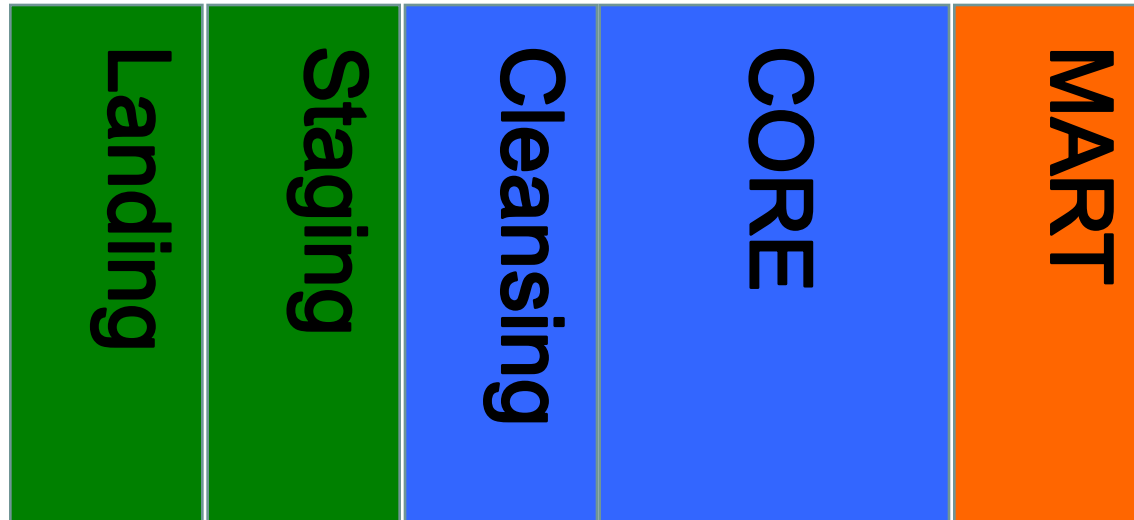
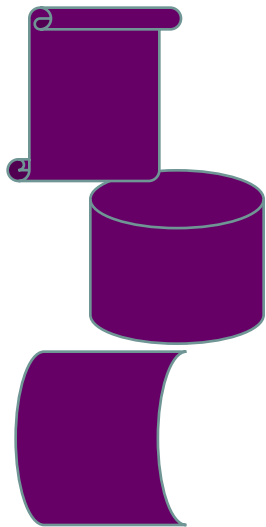
Dr. Andrea Kennel
InfoPunkt Kennel GmbH

Juni 2016

Agenda

- **Gemeinsames Verständnis von DWH**
- Modellierung mit ERM
- Modellierung mit Star/Snowflake
- Modellierung ADAPT
- Puzzle: was eignet sich wo?
- Wo gehört Data Vault hin?

Das typische DWH



Agenda

- Gemeinsames Verständnis von DWH
- **Modellierung mit ERM**
- Modellierung mit Star/Snowflake
- Modellierung ADAPT
- Puzzle: was eignet sich wo?
- Wo gehört Data Vault hin?

Entity-Relationship-Modell

- Das Entity-Relationship-Modell (kurz auch nur ER-Modell oder ERM genannt) dient primär der Beschreibung von Daten in einem Anwendungsbereich. Man beschreibt also mit Hilfe eines ER-Modells einen Ausschnitt der realen Welt. Da das Entity-Relationship-Modell bestimmten Regeln unterliegt und leicht verständlich ist, dient es daher als gute Kommunikationsbasis zwischen Anwendern und Entwicklern bzw. anderen beteiligten Parteien. Dadurch hat sich das ER-Modell zum De-facto-Standard für die Datenmodellierung etabliert.

- Quelle <http://www.datenbank-grundlagen.de>

Entity-Relationship-Modell

- Entität: Repräsentiert eine Gruppe von realen Objekten, die in der Regel als Tabellen umgesetzt werden
 - Projekt; Kunde
- Beziehung: Beziehung zwischen Entitäten mit der Kardinalität
 - Ein Projekt gehört zu genau einem Kunden
 - Ein Kunde kann mehrere Projekte haben

Übung ERM

- **Stundenerfassung (DidIT)**
 - Mit DidIT sollen alle Mitarbeitenden ihre täglich geleisteten Stunden erfassen. Dabei sollen sie angeben, für welche Projekte sie diese Stunden eingesetzt haben. Ein Projekt muss einem Kunden zugeteilt werden. Es gibt auch interne Projekte, die der eigenen Firma zugeteilt sind.
 - **Aufgaben:** Zeichnen Sie ein Datenmodell



Agenda

- Gemeinsames Verständnis von DWH
- Modellierung mit ERM
- **Modellierung mit Star/Snowflake**
- Modellierung ADAPT
- Puzzle: was eignet sich wo?
- Wo gehört Data Vault hin?

Star = Sternschema

- Das **Sternschema** ist eine besondere Form eines Datenmodells, dessen Ziel nicht die Normalisierung ist, sondern eine Optimierung auf effiziente Leseoperationen. Hauptanwendungsfeld sind Data-Warehouse und OLAP-Anwendungen
- Die Bezeichnung Sternschema rührt daher, dass die Tabellen sternförmig angeordnet werden: Im Zentrum steht eine Faktentabelle, um die sich mehrere Dimensionstabellen gruppieren
- Quelle: de.wikipedia.org



Snowflake= Schneeflockenschema

- Das **Schneeflockenschema** ist eine Weiterführung des Sternschemas, das beim OLAP und Data-Warehousing eingesetzt wird.
- Beim Sternschema liegen die Dimensionstabellen denormalisiert vor, was eine bessere Verarbeitungsgeschwindigkeit zu Lasten der Datenintegrität und des Speicherplatzes mit sich bringt. Im Gegensatz dazu werden beim Schneeflockenschema die einzelnen Dimensionstabellen verfeinert, indem sie klassifiziert oder normalisiert werden. Durch diese Weiterverzweigung des Datenmodells entsteht die Form einer Schneeflocke, woher die Bezeichnung dieses Entwurfsmusters stammt

Übung Star/Snowflake

■ Produktivität

- Das DWH soll folgende Auswertung erlauben:
- Geleistete und verbucht Stunden mit Angaben zu Projekt, Kunde und Datum
- Aufgaben: Zeichnen Sie ein Datenmodell

Agenda

- Gemeinsames Verständnis von DWH
- Modellierung mit ERM
- Modellierung mit Star/Snowflake
- **Modellierung ADAPT**
- Puzzle: was eignet sich wo?
- Wo gehört Data Vault hin?

ADAPT

- Die ADAPT Notation ist eine Designmethode, die sich für die Erstellung dimensionaler Datenmodelle sehr gut eignet. ADAPT (Application Design for Analytical Processing Technologies) ermöglicht die Darstellung von Dimensionen mit Hierarchien und Hierarchiestufen, sowie Fakten und deren Zuordnung zu Dimensionen. Das ADAPT-Diagramm kann nicht nur die hierarchischen Zusammenhänge innerhalb der Dimensionen sehr gut darstellen, sondern auch die Granularitätsstufe der Fakten grafisch veranschaulichen.
- Quelle: <http://www.datenbanken-verstehen.de>

Übung ADAPT

■ Produktivität

- Das DWH soll folgende Auswertung erlauben:
- Geleistete und verbucht Stunden mit Angaben zu Projekt, Kunde und Datum
- Es soll auch möglich sein, die Zahlen für die ganze Firma und je Kunde auf Stufe Monat und Jahr auszuwerten.
- Aufgaben: Zeichnen Sie ein Datenmodell

Agenda

- Gemeinsames Verständnis von DWH
- Modellierung mit ERM
- Modellierung mit Star/Snowflake
- Modellierung ADAPT
- **Puzzle: was eignet sich wo?**
- Wo gehört Data Vault hin?

Agenda

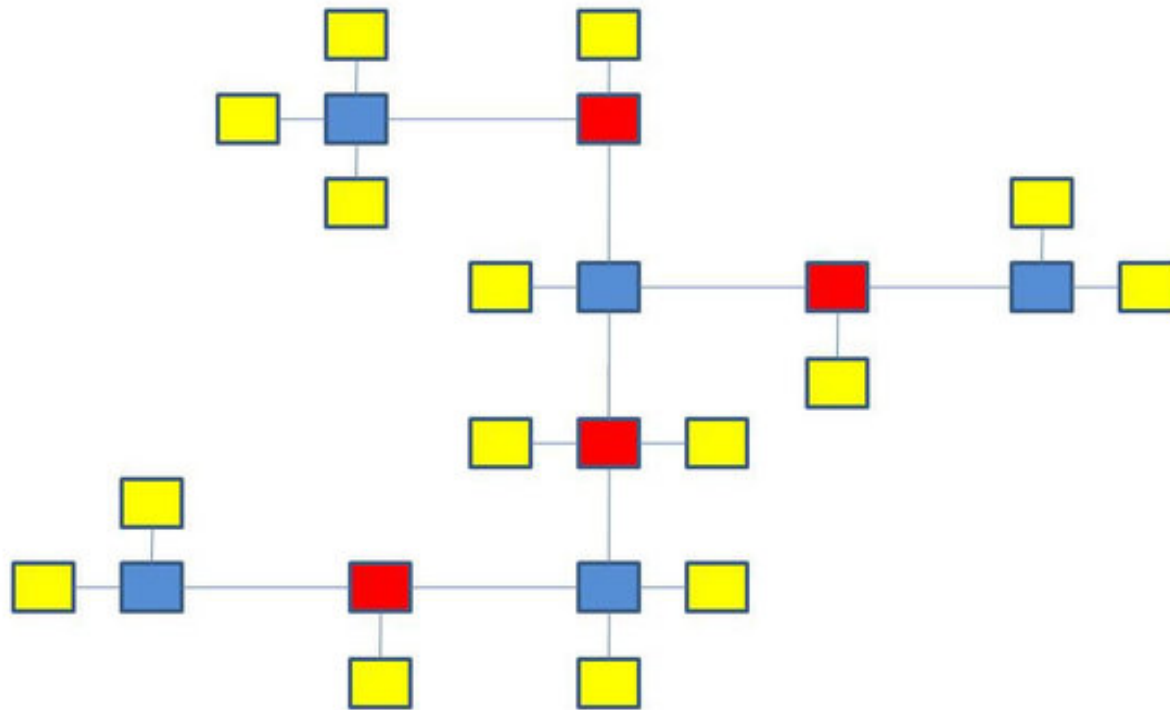
- Gemeinsames Verständnis von DWH
- Modellierung mit ERM
- Modellierung mit Star/Snowflake
- Modellierung ADAPT
- Puzzle: was eignet sich wo?
- **Wo gehört Data Vault hin?**

Wo gehört Data Vault hin?

- **Data Vault** ist eine Modellierungstechnik für Data Warehouses, die insbesondere für agile Data Warehouses geeignet ist. Sie bietet eine hohe Flexibilität bei Erweiterungen, eine vollständige Historisierung der Daten und erlaubt eine starke Parallelisierung der Datenladeprozesse.
- Entwickelt wurde die Data-Vault-Modellierung seit 1990 von Daniel Linstedt, wobei erst im Jahr 2000 die erste Veröffentlichung dazu erschien. Sie ist eine Kombination aus der relationalen Datenbankmodellierung mit der dritten Normalform (3NF) und dem Sternschema.
- Quelle: de.wikipedia.org

Wo gehört Data Vault hin?

Data Vault – Hubs / Links / Satellites



■ Quelle: hanshultgren.wordpress.com

Diskussion

?

!

Gediegener Tisch

- Die Firma “Gediegener Tisch” ist in verschiedenen Bereichen unterwegs
 - Holzkonstruktionen
 - Dekoration
 - Bewirtung
- Es wird Software genutzt für
 - Stundenerfassung (**DidIT**)
 - Buchhaltung (**Finz**)

Anforderungen

- Es soll ein kleines DWH gebaut werden mit folgenden Zielen
 - Bessere Übersicht über Finanzen
 - Einfache Debitorenkontrolle
 - Projektkontrolle
 - Produktivität der Mitarbeitenden

Anforderungen

ID	Report	Beschreibung
R1	Projektreport	Je Monat und Kunde alle Projekte mit geleisteten Stunden und verrechneten Beträgen
R2	Finanzen	Je Monat unsere zwei zentralen Finanzkennzahlen
R3	Debitorenspiegel	Liste der Debitoren, aufgeschlüsselt nach offen, gemahnt und bezahlt
R4	Produktivität	Geleistete und gebuchte Stunden je Mitarbeiter und Monat

Anforderung R1

Periode	Kunde	Projekt	Status	Stunden PL	Stunden MA	Verrechnete Leistung	Verrechnetes Material
Jan 16	Blumenladen	Schrankwand	in Arbeit	4	4	0	120
Jan 16	Blumenladen	Schaufenster	fertig	2	8	1200	700
Jan 16	Brauerei	Küche	in Planung	8	0	0	0
Jan 16	Brauerei	Farbkonzept	in Arbeit	16	0	0	0
Jan 16	Restaurant	Verkaufsunterstützung	in Arbeit	1	32	1000	20
Jan 16	Total	Total		31	44	2200	840
Feb 16	Blumenladen	Schrankwand	fertig	2	40	3600	1500
Feb 16	Brauerei	Küche	in Arbeit	0	24	400	0
Feb 16	Brauerei	Farbkonzept	fertig	8	8	1600	500
Feb 16	Restaurant	Verkaufsunterstützung	in Arbeit	1	32	1000	20
Feb 16	Total	Total		11	104	6600	2020
Alle	Blumenladen	Total		8	52	4800	2320
Alle	Brauerei	Total		32	32	2000	500
Alle	Restaurant	Total		2	64	2000	40
Alle	Total	Total		42	148	8800	2860



Was wird gebraucht, wo ist die Quelle?

Was	Beschreibung	Quelle
Periode	Monat und Jahr	DidIT, Finz
Kunde	Name des Kunden	DidIT, Finz
Projekt	Name des Projekts	DidIT
Status	Status des Projekts	DidIT
Stunden PL	Stunden, die der PL am Projekt gearbeitet hat	DidIT
Stunden MA	Stunden, die alle MA am Projekt gearbeitet hat	DidIT
Verrechnete Leistung	Betrag, der für Stunden verrechnet wurde	Finz
Verrechnetes Material	Betrag, der für Material verrechnet wurde	Finz

Anforderung R2

Periode	Fremdkapital/Eigenkapital	Liquidität 1
Jan 16	54%	92
Feb 16	58%	105

Anforderung R3

Datum Rechnung	Fälligkeit	Betrag	Kunde	Status
10.01.2016	09.02.2016	1020	Restaurant	bezahlt
01.02.2016	02.03.2016	2020	Blumenladen	bezahlt
09.02.2016	10.03.2016	1020	Restaurant	gemahnt
01.03.2016	01.04.2016	5100	Blumenladen	offen
01.03.2016	01.04.2016	2500	Brauerei	offen
10.03.2016	09.04.2016	1020	Restaurant	offen



Anforderung R4

Periode	MitarbeiterIn	Projekt-Stunden	Interne Stunden	Abwesenheiten	Weiterbildung
Jan 16	AAA	12	25	4	0
Jan 16	BBB	105	5	0	24
Jan 16	CCC	189	11	0	0
Jan 16	Total	306	41	4	24
Feb 16	AAA	0	31	0	8
Feb 16	BBB	150	15	0	0
Feb 16	CCC	120	9	32	0
Feb 16	Total	270	55	32	8
Alle	AAA	12	56	4	8
Alle	BBB	255	20	0	24
Alle	CCC	309	20	32	0
Alle	Total	576	96	36	32