

**Bereich:** Business Intelligence  
**Workshop:** Data Vault: Konzepte, Architektur und Modellierung  
**Dauer:** 2 Tage

## Beschreibung des Workshops

---

In Data Warehouse Systemen sind für anfrageorientierte Schichten (Datamarts) multidimensionale Modellierungsmethoden State of the Art. Bei den vorgelagerten Integrationsschichten (Core Warehouse, Enterprise Data Warehouse) kommen hingegen klassische Modellierungsmethoden (ER-Modellierung, Normalisierung) zum Einsatz, die ursprünglich für transaktionsorientierte Systeme entwickelt wurden. Diese Kern-Technologien und Standard-Vorgehensweisen sind langjährig erprobt und haben sich in der Praxis bewährt.

Vor dem Hintergrund stetig wachsender Anforderungen an Data Warehouse Systeme (z.B. hinsichtlich kürzerer Aktualisierungsintervalle, höherer Ladeperformance bei wachsender Datenmenge und agile Adaptierbarkeit) treten Schwächen transaktionaler Modellierungsverfahren allerdings immer stärker zu Tage. Hier müssen neben neuen technologischen Pfaden (z.B. InMemory-Datenbanken, Big Data Technologie) auch konzeptionell und methodisch neue Wege beschritten werden.

Ein Lösungsansatz ist die von Dan Linstedt entwickelte Modellierungsmethode Data Vault, die seit vielen Jahren erprobt ist und zunehmend auch im deutschsprachigen Raum Verwendung findet.

Data Vault bietet

- ein einfaches Grundmodell mit wenigen Basiskonzepten
- Struktur-Entkopplung und Impact-Isolation für Modelländerungen und -erweiterungen
- massiv parallelisierbare Ladeprozesse durch strukturelle Entkopplung
- flexible Strukturерweiterung bei gleichzeitiger Historisierungsoption

Im Seminar lernen Sie die wesentlichen Grundbausteine des Data Vault-Modells und die Hauptargumente für den Einsatz kennen. Zusätzlich werden Ihnen grundsätzliche Vorgehensweisen zur Umsetzung von ETL-Prozessen und modernen Data Warehouse Architekturen vorgestellt. Nachdem die Grundlagen gelegt wurden, lernen Sie am zweiten Seminartag erweiterte Konzepte und Best Practices kennen, die es Ihnen ermöglichen, konkrete Problemstellungen individuell zu meistern.

## Zielgruppe

---

Datenmodellierer, Designer, Architekten, Projektleiter

## Voraussetzungen

---

Grundlagenwissen Datenbanken, relationale Datenmodellierung, Grundkenntnisse DWH-Architekturen und ETL-Prozesse

## Ziele

---

Nach dem Seminar kennen Sie - neben den grundlegenden und fortgeschrittenen Konzepten - die Basiskomponenten und Architekturvarianten eines Data Vaults, sowie die Vorgehensweisen zur Umsetzung der ETL-Prozesse. Dies ermöglicht es Ihnen zu bewerten, ob Data Vault für Ihre Projekte in Betracht kommt und die beste Möglichkeit zur Umsetzung zu identifizieren. Durch Data Vault sind Sie in der Lage, Ihr Data Warehouse schrittweise aufzubauen, schneller zu beladen und gleichzeitig mit einer agilen Entwicklung Schritt zu halten.

## Inhalt

---

### Tag 1

#### Einführung und Terminologie

- Typische Probleme in klassischen Data Warehouse Systemen
- Gründe für den Einsatz von Data Vault
- Was ist Data Vault (Begriffe/Definition, Kernkonzepte, Historie, DV 2.0)

#### Data Vault Basiskomponenten

- Konzept der Dekomposition
- Hubs
- Satellites
- Links

#### Modellierung eines Data Vault

- Modellierungsprozess
- Entwurfsregeln für Hubs, Links und Satellites

#### Lade- und Abfrageprozesse eines Data Vaults

- ETL-Logik für das Beladen von Hubs, Links und Satellites
- Übersicht über die Ladeprozesse

#### Data Vault Architekturen

- Klassische Architektur vs. Data Vault Architektur
- Raw Vault/Business Rule Vault
- Weitere Konzepte erklärt: Error Vault, Meta Vault, Metrik Vault, Operational Vault

## Tag 2

### Data Vault Generatoren

- Generierung von Datenmodellen und ETL-Prozessen
- Vorhandene Lösungen
- Vorstellung Quipu

### Advanced Concepts

- Hubs
- Links
- Satellites
- Hilfstabellen

### Best Practices

- Auflösen von Link-to-Link Beziehungen
- Gruppieren von Attributen
- Umgang mit unvollständigen Daten
- Model Driven vs. Data Driven Structure
- Anchor Modelling
- Namenskonventionen