

DataVault in der Praxis – nur ein Werbeversprechen?

Marcel Aretz
OPITZ CONSULTING Deutschland GmbH
Hamburg

Schlüsselworte

DATA VAULT / ZEITSTEUERUNG / MODELLIERUNG / EXADATA / ORACLE 12C

Einleitung

Im Rahmen eines Kundenprojektes soll ein neues Datawarehouse (DWH) auf Basis einer Oracle-Exadata X6 entstehen. Das bestehende DWH ist über 15 Jahre gereift. Sowohl die zahlreichen Entwicklerdialekte, als auch die fehlenden Standards für die Entwicklung, haben unter anderem dazu geführt, dass Wartung und Anpassung des bestehenden DWH an neue und geänderte betriebliche Abläufe zunehmend komplizierter und damit aufwändiger werden. Die in allen Schichten verteilte Logik erschwert zudem die Suche und Beseitigung von Fehlern.

Erwartungshaltung

Das neue DWH soll Vieles besser machen. Direkt im Fokus steht dabei ein agiles Projektvorgehen. Beginnend mit grundlegenden Geschäftsobjekten wie Artikeln, Kunden und Aufträgen aus dem primären ERP-System sollen Schritt für Schritt weitere Geschäftsobjekte und Quellsysteme integriert werden. Dabei werden idealerweise in jedem Inkrement die bereits bestehenden Komponenten nicht erneut angefasst. Zudem werden Stück für Stück die alten und neuen Anforderungen der Fachbereiche in das neue DWH einfließen.

Bezüglich der Modellierung des DWH-Core fällt die Wahl schnell auf DataVault. Bereits die ersten beiden Sätze bei Wikipedia verstärken den Eindruck, dass DataVault die ideale Modellierungstechnik für dieses DWH ist:

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.

Zudem ergibt sich durch die Trennung in RawDataVault (RDV) und BusinessDataVault (BDV) eine saubere Abgrenzung zwischen Rohdaten und durch Geschäftsregeln modifizierten, gefilterten und transformierten Daten.

Außerdem eignet sich DataVault ideal für Automatisierung, da es nur drei grundlegende Tabellentypen gibt und damit nur eine kleine Auswahl an Implementierungsmustern gibt.

Zielsetzung

Im Laufe dieses Projektes, aber auch in zahlreichen anderen Projekten, in denen DataVault zum Einsatz kommt, entwickeln sich während der Implementierung immer ähnliche Probleme und Fragestellungen.

Dieser Vortrag beleuchtet die im o.g. Projekt aufgetretenen Komplikationen bei der Einführung von DataVault. Hierbei werden auch Ansätze vorgestellt, wie diese und ähnliche Probleme vermieden oder zumindest abgeschwächt werden können.

Abschließend erfolgt im Fazit eine kritische Gegenüberstellung von o.g. Erwartungen und den gemachten Erfahrungen.

Themenschwerpunkte

- Da für das RDV-Modell idealerweise über klare Regeln wie PrimaryKeys / ForeignKeys / UniqueKeys in der Quelle bestimmt werden kann, welche Hubs / Links / Satellites benötigt werden, kann dies auch vollständig generiert werden.
→ Vor- und Nachteile eines solchen RawDataVaults
- In zahlreichen Fällen werden vom Fachbereich Informationen aus Masken im ERP-System angefordert, die nicht zwingend in der Quelle persistiert sind, sondern komplexe Funktionen enthalten.
→ Umgang mit komplexen Kalkulationen im Quellsystem (z.B. EK-/VK-Preisermittlung)
- DataVault-Modelle zeichnen sich durch eine im Vergleich zum 3NF-Modell deutlich erhöhte Anzahl von Tabellen aus. Diese Tabellen müssen auch bewirtschaftet werden. Unabhängig vom ETL-Tool oder SQL-Dialekt ist hier eine Automatisierung sinnvoll, wenn nicht gar erforderlich.
→ Grenzen der Automatisierung
- Die Vielzahl der zu bewirtschaftenden Tabellenobjekte erfordert auch eine gute Orchestrierung der Ladeprozesse. Ein DataVault-Modell bietet durch die Nutzung von Hashes sehr gute Voraussetzungen für die Parallelisierung der unabhängigen Prozesse.
→ Orchestrierung
- In zahlreichen Projekten werden Tabellen aus dem ERP benötigt, die rein technischer Natur sind, um Logiken / Regeln / Vertextungen abzubilden. Die Objekte stellen jedoch keine echten Geschäftsobjekte dar (z.B. „95% Cotton / 5% Elasthan“)
→ RDV um jeden Preis?

Kontaktadresse:

Marcel Aretz
OPITZ CONSULTING Deutschland GmbH
Standort Hamburg
Landwehr 2
22087 Hamburg

E-Mail marcel.aretz@opitz-consulting.de
Internet: www.opitz-consulting.de