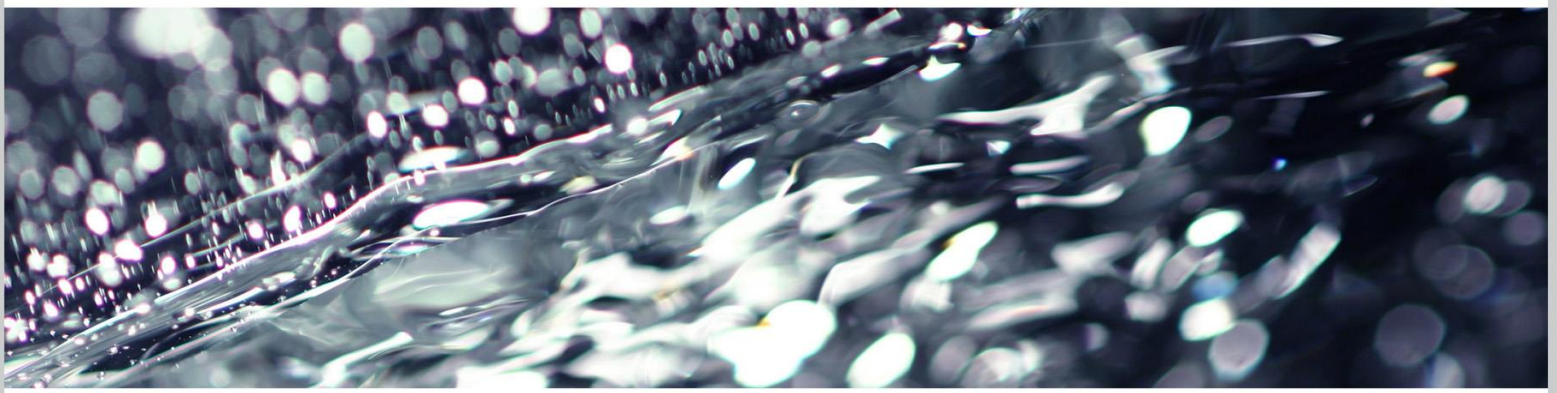
act on facts



Realtime Data-Warehouse-Prozesse mit Informatica PowerCenter und PowerExchange

Management Summary

Ziele

- … Minimierung der Last auf den Quellsystemen für die Datenbewirtschaftungsprozesse im Data Warehouse
- … Entkoppelung der Prozesslast zwischen Quellsystemen und Data Warehouse
- … Flexible Datenextraktionsmechanismen für wechselnde Zeitfenster aufgrund von Abhängigkeiten zu Batch-Prozessen in den Quellsystemen
- … Optimierung des Verarbeitungsvolumen durch changed data capture

Ansatz

- … Aufbau einer entkoppelten Prozessarchitektur mittels Informatica PowerExchange.
- … Koordinationsmanagement der Extraktions- und Ladeprozesse durch übergreifende Workflowsteuerung in Informatica PowerCenter

Ergebnis

- … Fflexible Datenübernahme
- … Entkoppelung von operativen Quellsystemprozessen

› Die Ausgangssituation

Die öffentliche Verwaltungseinrichtung verfügt über eine Vielzahl Quellsysteme und diverse Anforderungen zum Aufbau eines Data Warehouse im Bereich Melderegister und Mahnverfahren. Die inhaltliche Konsolidierung und Aufbereitung der unterschiedlichen Informationen stellt für die Behörde dabei nur eine von vielen Herausforderungen dar. Besonderheiten bestehen insbesondere hinsichtlich der Verfügbarkeit der Quellsysteme, deren Administration und Betrieb in der Verantwortung eines Rechenzentrums liegen, das keine verlässlichen Zeitfenster für die Datenextraktion ins DWH anbieten kann. Die Last auf den Quellsystemen durch ETL-Prozesse darf keinen kritischen Umfang annehmen. Und schließlich besteht auch noch die Anforderung, das DWH langfristig realtime-fähig zu machen.

Der Schlüssel für eine anforderungsgerechte ETL-Prozessarchitektur liegt in der automatisierten Bereitstellung veränderter Daten im Quellsystem und einer entkoppelten Verarbeitung innerhalb des DWH. Diese Fortschreibung muss flexibel ausführbar sein, inhaltliche Abhängigkeiten in der Informationsverarbeitung berücksichtigen und sowohl eine Gesamtverarbeitung im Sinne einer Initialisierung der Datenbestände als auch eine Deltaverarbeitung zulassen.

› Das Vorgehen

Als Werkzeuge zur technischen Umsetzung der Ladeprozesse wurden Informatica PowerCenter und PowerExchange gewählt. Diese Toolkombination bietet eine optimale Infrastruktur zur Entkoppelung der ETL-Schritte mit nahtloser Integrationsmöglichkeit. Das wesentliche Merkmal der genutzten PowerExchange-Komponente ist seine Implementierung auf der Quellsystem-Infrastruktur. Im vorliegenden Fall sind die Hauptquellsysteme auf einer DB2-Hostdatenbank implementiert. Mit PowerExchange ist es möglich, zusätzliche Jobs zu integrieren, die auf Basis der Datenbank-Logfiles Änderungen in der Datenbank extrahieren und persistent ablegen. Dabei können die benötigten Änderungssätze durch Auswahl von relevanten Entitäten und Attributen selektiv definiert werden.

Tests belegten, dass die zusätzliche Last durch die Log-Analyse keine wesentliche Beeinträchtigung des operativen Betriebes darstellt. Letztendlich stellen die von der Log-Analyse bereitgestellten Datenbestände den Dateninput für die nachfolgenden Ladeprozesse des DWH dar. Hierdurch wurde neben der technischen Entkoppelung auch eine scharfe Trennung der Verantwortlichkeiten erzielt.

Die PowerExchange-Komponente stellt sicher, dass jeweils nur neue Transaktionen, die als relevant definiert sind, bereitgestellt werden. Es entfällt somit ein häufig individuell umgesetzter Algorithmus für die Delta-Erkennung. Es waren allerdings auch Konstellationen zu berücksichtigen, die eine sehr genaue Konzeption dieser Schnittstellenmethodik erfordern. Darunter fallen insbesondere die Hand¬habung von Neuinitialisierungen des Datenbestandes sowie die Überwachung von datenbankspezifischen Vorgängen und deren Auswirkung auf die Extraktionsschnittstelle.

Für die initiale Beladung des DWH mussten separate Prozesse aufgebaut werden. Was sich zunächst nach Mehraufwand anhört, stellt sich in der Praxis im laufenden Betrieb des DWH meist als Vorteil heraus. In der Regel ist die Verarbeitung von initialen Beständen bei Weitem nicht so komplex wie die Delta-Verarbeitung. Vor allem aber erlaubt der Initial Load die Verwendung von Bulk-Load Mechanismen, welche die Ladezeit extrem verkürzen. Kann man, wie im vorliegenden Fall, auf Replikate der Quellsysteme zugreifen, bei denen zu bestimmten Zeitpunkten absolute Synchronität gewährleistet ist, so ist auch die Last für das operative System kein Problem.

“Data Warehouse Systeme müssen eine technische und organisatorische Trennung zu den operativen Quellsystemen zum Ziel haben, um ein störungsfreies Lademanagement mit klarer Verantwortungsabgrenzung zu gewährleisten - insbesondere bei der Integration von Rechenzentren.“

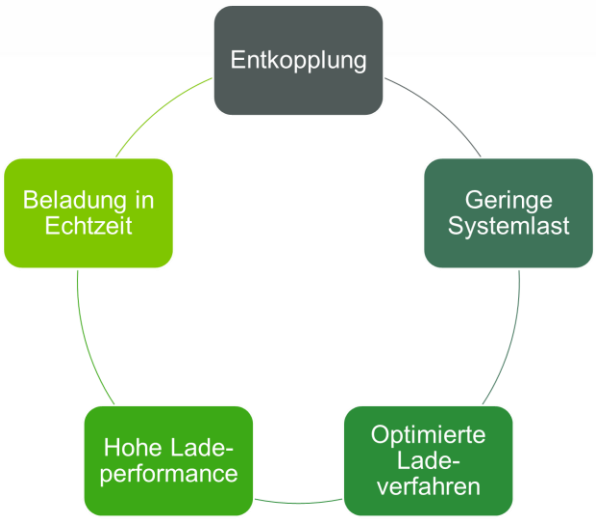
Thomas Weiler

Senior BI Architect, mayato GmbH

Ein weiterer sensibler Punkt bei der Nutzung des Delta-Verfahrens ist die Sicherstellung der nahtlosen Delta-Zeitreihe. Diese verwaltet zwar PowerExchange eigenständig, sorgfältige Tests zeigten aber, dass bei der Konfiguration der PowerCenter-Jobs sehr gründlich auf die Parametrisierung zu achten ist. Im vorliegenden Fall führte zum Beispiel die zyklisch durchgeführte Reorganisation von Tabellen in der Datenbank zu Verschiebungen des Delta-Zeitstempels. Da der Delta-Zeitstempel konfigurierbar ist, muss bei der Workflowsteuerung exakt definiert werden, unter welchen Umständen dieser bewusst manipuliert werden darf.

› Ziele erreicht

Aufgrund der Delta-Verarbeitung, die transaktionsgesteuert aus dem Quellsystem getriggert wird, ist die Verarbeitung im DWH nicht nur unabhängig vom Quellsystem, sondern per Definition auch real time-fähig. Die Transformationen sind einzelsatzbasiert, es werden immer ganze Transaktionen geliefert und die Transaktionsart ist bekannt.



Je näher der Aktualisierungszyklus ein einer Echtzeitbeladung ist, desto größer ist die Wahrscheinlichkeit, das fachliche Abhängigkeiten nicht sofort und vollständig aufgelöst werden können. Beispielsweise kommt es häufig vor, dass Vertrags- oder Meldeverfahren auf Teilnehmer referenzieren, die noch nicht im Partnersystem verarbeitet wurden. Im DWH kann dies konzeptionell beispielsweise durch Aussteuerungsbereiche berücksichtigt werden, in denen betroffene Datensätze bis zum Eintreten eines konsistenten Zustands werden.

› mayato Expertise

mayatos Berater blicken auf langjährige Erfahrungen bei Konzeption und Umsetzung von Business-Intelligence-und Corporate-Performance-Lösungen zurück. Ihr Wissen zahlt sich für Sie aus, wenn es darum geht, komplexe betriebswirtschaftliche und informationstechnische Anforderunrungen optimal durch den Einsatz von Data-Warehouse- und Business-Intelligence-Technologien zu erfüllen.

Als Analysten- und Beraterhaus ist mayato spezialisiert auf Lösungen für Business Intelligence und Business Analytics. In diesen Bereichen deckt mayato das komplette Spektrum an Dienstleistungen ab. Dazu gehören u.a. Toolauswahl, Strategien und Organisationskonzepte, Architektur und Design, Data-Warehouse-Modellierung und die Erstellung von Reports und Cockpits. Auch bei der der korrekten Interpretation von Informationen und der Vorhersage zukünftiger Ereignisse helfen mayatos Experten gerne mit Spezialknowhow in Statistik und Datenanalyse.

Als Think Tank analysiert mayato Trends und Innovationen, evaluiert Technologien und methodische Ansätze und unterzieht Werkzeuge intensiven Praxistests. Auf diese Weise sind mayato Berater immer up to date und können Ihren Kunden Dienstleistung auf höchstem Niveau vermitteln.