



Datenmigration für Unternehmen - Herausforderungen und Lösungen mit Snowflake

1. Datenmigration für große Retailer - Herausforderungen und Bedeutung	2
2. Die Wahl von Snowflake - eine zukunftssichere Lösung für Retailer	2
3. Der Lift-&-Shift-Ansatz - eine schonende Migration ganzer Data-Warehouses	2
4. Der Migrationsprozess im Detail - von Teradata zu Snowflake	3
5. Data Migration führt zu Kostenoptimierung und Kostentransparenz	4
6. Fristgerechte Datenmigration in die Cloud	5

Einführung

Daten spielen eine entscheidende Rolle im modernen Geschäftsumfeld, insbesondere für große Retailer und Unternehmen, die eine Fülle von Informationen verarbeiten müssen, um fundierte Entscheidungen zu treffen. Angesichts der wachsenden Datenmengen und der Notwendigkeit, eine agile und flexible Infrastruktur zu haben, sehen sich Unternehmen oft mit der Herausforderung konfrontiert, ihre Daten in eine neue Umgebung zu migrieren. Die Datenmigration ist ein kritischer Prozess, der eine sorgfältige Planung, technische Expertise und

effektive Lösungen erfordert, um den reibungslosen Übergang zu gewährleisten.

Hier erläutern wir die Herausforderungen der Datenmigration für große Retailer und Lösungen auf Basis von Snowflake, einer leistungsstarken Cloud-Data-Plattform. Unsere Fallstudie beschreibt die Bedeutung dieser Prozesse für Retailer im Allgemeinen und die erfolgreiche Migration einer der größten Teradata-Datenbanken Europas in die Cloud.

1. Datenmigration für große Retailer – Herausforderungen und Bedeutung

In einer wettbewerbsintensiven Branche wie Retail, in der Kundenanforderungen sich ständig ändern und die Geschwindigkeit der Entscheidungsfindung entscheidend ist, hängt der Erfolg eines Retailers stark von der Verfügbarkeit

und Genauigkeit der Daten ab. Allerdings sind viele Retailer mit Datenbanken konfrontiert, die gerade bei Flexibilität und Skalierbarkeit auf Grenzen stoßen und nicht mehr den Anforderungen einer datengetriebenen Welt gerecht werden.

2. Die Wahl von Snowflake – eine zukunftssichere Lösung für Retailer

Die Migration auf eine moderne Cloud-Data-Plattform wie Snowflake bietet die Möglichkeit, diese Herausforderungen zu bewältigen und eine Reihe von Vorteilen zu erzielen. Snowflake ermöglicht eine flexible, skalierbare und zukunftsorientierte Datenverarbeitung. Damit arbeiten Retailer agiler und effizienter als davor. In diesem Fall war es dem Kunden besonders wichtig, Performance-Einbrüche bei Zugriffsspitzen, also Inflexibilität in Bezug auf Concurrency (gleichzeitige

Zugriffe), zu reduzieren bzw. zu vermeiden. Dazu bietet Snowflake weitere, für Retailer interessante Möglichkeiten: Sicherheit, Zugriffsrechte, Datenverfügbarkeit, Data Sharing zwischen Fachbereichen und erweiterte Funktionalitäten. Auch das flexible Kostenmodell, nach dem die Betriebskosten der Plattform abgerechnet werden, überzeugten unseren Kunden. Mit Snowflake haben wir den „sweet spot“ zwischen Kostenoptimierung, Skalierbarkeit und Flexibilität getroffen.

3. Der Lift-&-Shift-Ansatz – eine schonende Migration ganzer Data-Warehouses

Für unseren Kunden war es wichtig, vorhandene Daten und Prozesse in eine neue Umgebung zu übertragen, ohne die bestehenden Strukturen zu ändern. Der Lift-&-Shift-Ansatz bietet hierfür eine schonende Lösung. Damit migrierten wir zwei voneinander unabhängige Business Entitäts vollständig in die Cloud. Es wurden sämtliche Datenstrukturen und -objekte in die Cloud übertragen, ohne dass dabei Änderungen an den bestehenden Strukturen

vorgenommen werden mussten. Die bestehenden Prozesse und Datenintegrität blieben erhalten. Für Retailer ist dies von entscheidender Bedeutung, da eine Unterbrechung der Geschäftsprozesse vermieden und der Übergang nahtlos gestaltet wird. Während der Migration der einzelnen Fachbereiche haben diese ihre Arbeit mit der Datenbank nahtlos fortgesetzt. Unsere Arbeit erfolgte dabei vollständig „unter der Haube“.

Tools, die Daten zur Datenbank liefern, waren z. B. bereits festgelegt. Hier haben wir an die Schnittstelle zu Ab Initio, einer Datenmanagement-Plattform, angedockt, um Daten nach Snowflake zu

laden. Wir haben Snowflake an MicroStrategy, ein Dateianalyse-Tool, angeschlossen. Damit wurden für das Reporting in MicroStrategy die Daten aus Snowflake und nicht mehr aus Teradata verwendet.

4. Der Migrationsprozess im Detail - von Teradata zu Snowflake

Was bedeutet Lift-&-Shift?

Bei einer Lift-&-Shift-Migration werden sämtliche Strukturen und Daten in die Cloud migriert, ohne große Änderungen an den Objekt-Strukturen vorzunehmen. Das bedeutet für die Migration, dass keine der aktuellen Prozesse oder Objekte strukturell angepasst werden dürfen. Dies ist besonders

herausfordernd, da die Architekturen und Prozesse jahrelang für das alte onPrem-System optimiert wurden. Es waren nur minimale Anpassungen erlaubt, um in Snowflake ein mindestens genauso performantes Data Warehouse aufzubauen.

Welche Schritte wurden während der Migration unternommen?

Für die Migration wurden die vorhandenen Daten und Prozesse in fachlich zusammengehörige Pakete aufgeteilt. Anschließend wurde jedes Paket einzeln migriert. Dafür wurden Datenstrukturen, ETL-Prozesse, Schnittstellen und natürlich die Daten in die für

Snowflake benötigte Struktur gebracht und anschließend inhaltlich validiert. Zudem wurden die Prozesse und Zugriffe hinsichtlich der Performance optimiert. Parallel dazu wurde der Snowflake-Account inklusive Rechte, Warehouses, Parameter etc. eingerichtet.

Welche Tools wurden für Data Loading & Reporting verwendet?

Die Daten wurden mit dem Tool Ab Initio geladen. Dieses wurde auch schon vorher für die Datenbeladung auf der Teradata genutzt. Entsprechend mussten die von Ab Initio getriggerten Prozesse nach Snowflake migriert werden. Als Reporting-Tool wurde MicroStrategy verwendet. Das Tool greift auf den View Layer zu, der nach

Snowflake migriert wurde. Es wurde für ein performantes Reporting ein optimales Setting in MicroStrategy ausgearbeitet und zusätzlich eine Warehouse-Architektur und Settings entwickelt. Dadurch wurde trotz sehr großer Datenmengen und Vielzahl an Zugriffen ein performantes Reporting aufgebaut.

Welche Verbesserungen in der Performance haben wir durch Prozessoptimierung und Verbesserung der Abfragemotik erreicht?

Durch den gewählten Lift-&-Shift-Migrationsansatz und die auf Teradata optimierten Prozesse standen wir während der Migration vor der Frage, wie ohne große Änderungen an den Objekten ein performantes Reporting in Snowflake aufgebaut werden kann. Als Lösung wurde ein

automatisiertes Materialisierungsframework entwickelt. Mit diesem wurden Ergebnisse vorberechnet und automatisiert zwischengespeichert. Dadurch konnten wir die erforderliche Performance erreichen, ohne große Umbauten an den Datenstrukturen vornehmen zu müssen.

Wie standardisiert das ACE-Deploy-Framework die technischen Komponenten und Prozesse auf Snowflake?

Im von initions entwickelten Deployment-Framework ACE Deploy haben wir ein Werkzeug geschaffen, das ein standardisiertes Deployment sicherstellt. Mit dem Framework finden Änderungen an den Objekten des Data Warehouses kontrolliert statt und es entsteht für die Produktion keine Gefahr durch ungetestete Änderungen. Dafür sind in vordefinierten

Prozessen automatisierte Checks, Approval-Prozesse und ein automatisiertes Deployment im Framework enthalten. Zudem werden durch das Framework sämtliche Datenbankstrukturen in einem Git Repository abgelegt, wodurch Änderungen an den Strukturen komplett nachvollziehbar sind.

Individuelle Berechtigungskonzepte als Best Practice

Für das zukünftige Cloud Data Warehouse wurde ein für Snowflake optimiertes Berechtigungskonzept entwickelt, um alle aktuellen und zukünftigen Zugriffe einfach kontrollieren und managen

zu können. Dafür wurden die Best Practices aus anderen Projekten genutzt und an die Gegebenheiten des Kunden angepasst.

Testautomatisierung für Performance-Tests und Validierung einer erfolgreichen Datenmigration

Da ein manuelles Testing von Performance und Inhalt alles andere als effizient ist, wurde für der Migration ein Test-Framework entwickelt. Dieses Python-basierte Framework ermöglichte uns, automatisiert Inhalte abzugleichen und Performance-Vergleiche

zwischen Teradata und Snowflake durchzuführen. Durch die Automatisierung der Tests wurde viel Zeit gespart, da nur die fehlerhaften Objekte oder langsame Querys manuell analysiert werden mussten.

Kostentransparenz von Snowflake

In Snowflake wurde eine Warehouse-Architektur aufgebaut, in der die Workloads möglichst kosteneffizient auf einzelne Warehouses für einzelne Bereiche aufgeteilt wurden. Dadurch bekommt jeder Bereich die Rechenleistung die er benötigt

und es gibt weniger gemeinsam genutzte Ressourcen. So können die entstehenden Kosten viel einfacher den entsprechenden Bereichen zugeordnet werden. So zahlen die Bereiche nur das, was sie wirklich an Kosten verursacht haben.

5. Data Migration führt zu Kostenoptimierung und Kostentransparenz

Die erfolgreiche Migration war nach 17 Monaten mit einer harten Deadline termingerecht abgeschlossen. Durch unsere intensive Zusammenarbeit mit dem Kunden bei der Projektplanung und beim Reporting haben wir einen soliden Migrationsplan erstellt und umgesetzt. Die Performance-Ziele wurden eingehalten. Das Problem der zeitweisen Nichtverfügbarkeit von Teradata wurde durch die

Migration behoben. Das neue, auf Snowflake basierende Data Warehouse schafft Kostentransparenz, z. B. zwischen Fachbereichen. Es besteht eine größere Skalierbarkeit in Bezug auf Datenmengen und Concurrency-Belastungsresistenz bei mehr Usern, Fachbereichen oder Daten. Zukünftig kann die Datenbank flexibel skaliert oder weiterentwickelt werden und dabei performant bleiben.

Etablierung eines neuen Deployment-Frameworks

Wir haben Ace Deploy, ein von initions entwickeltes Acceleration-Framework beim Kunden etabliert. Diese technischen Komponenten und Prozesse sind bei Snowflake für die gesamte Plattform

standardisiert. Mit diesem Deployment-Framework sind saubere Prozesse für die Entwicklung möglich. Das Framework haben wir dem Kunden für eine künftige Entwicklung zur Verfügung gestellt.

Auf den Kunden angepasste Berechtigungskonzepte

Das Berechtigungsmanagement haben wir auf den Kunden zugeschnitten. Dafür wurden spezielle technische Rollen entwickelt, die bestimmte Rechte bündeln. Damit regulieren Retailer ganz

einfach die Berechtigungen auf ihre Daten. Die technischen Rollen werden genutzt, um den Datenzugriff für die Fachbereiche zu konfigurieren.

Tool für die Testautomation

Wegen ungewöhnlich großer Datenmengen und zur Wiederverwendbarkeit bei Massentests haben wir ein Tool gebaut, mit dem wir Performance Tests zur inhaltlichen Validierung der Migrationsreports

durchgeführt haben. Nach der Migration eines jeden Fachbereichs haben wir einen Test durchgeführt, der die Übereinstimmung der Daten vor und nach der Migration bestätigte.

6. Fristgerechte Datenmigration in die Cloud

Nach 17 Monaten schließt initions erfolgreich und fristgerecht die Migration einer der europaweit größten Teradata-Datenbanken in die Cloud ab. Die Datenbank-Struktur darf nicht verändert werden, also wurde ein Lift-&-Shift-Verfahren durchgeführt. Die Verfügbarkeit von Daten und Zugriffsgeschwindigkeit wird verbessert. Da

Snowflake nach dem Headless-Prinzip arbeitet, ist die neue Lösung für unseren Kunden beliebig erweiterbar und an zukünftige Bedürfnisse anpassbar. Das Kostenmodell der Cloud-Datenbank richtet sich nach den Zugriffsspitzen. Dies ermöglicht dem Kunden die Nachverfolgung von Zugriffen und damit weitere Kostenoptimierung.



André Paul Henkel

Managing Director
mindcurv: Data & AI
Hamburg

+49 (0)40 8221 71-300