

Datenmigration

Proseminar

Unternehmensübergreifende IT – Transformationen

SS 2011

Elvan Öztürk

elvan_16@hotmail.de

Abstract: In vielen Unternehmen besteht die IT aus einer gewachsenen Struktur heterogener IT-Systeme. Und in allen Unternehmen kommt irgendwann die Zeit, dass neue Applikationen eingeführt oder Daten in andere Systeme überführt werden müssen. Deshalb ist es wichtig, die Gründe und Vorgänge einer Datenmigration zu kennen und zu verstehen. Diese Seminararbeit beschäftigt sich im Speziellen mit den Ursachen, die zu solch einer Migration führen und mit dem exakten Ablauf einer Datenmigration. Das heißt, nach den Ursachen und Treibern für Datenmigrationsprojekten, wird kurz der Begriff „Datenmigration“ erläutert und daraufhin das gesamte Datenmigrationprojekt samt Teilprozess „Datenmigration“.

1 Ursachen und Treiber für Datenmigrationsprojekte

Unternehmenszusammenschlüsse

Im Zusammenhang mit Firmenübernahmen (Mergers & Acquisitions) oder die Abspaltung von Unternehmensteilen aus dem Mutterkonzern, müssen zum Einen, neue IT-Systeme in die eigene Infrastruktur eingebunden werden oder zum Anderen, vorhandene Strukturen in eine neue IT-Infrastruktur überführt werden.

Speicherengpässe

Wenn Unternehmen expandieren oder neue Kunden gewonnen haben werden ihre Datensätze größer. Das heißt, mehr Daten müssen gespeichert werden. Wenn die Kapazität des vorhandenen Datenspeichers erreicht ist, es jedoch noch zu speichernde Datensätze gibt, spricht man von einem Speicherengpass. Infolge dessen, müssen neue, größere Speichersysteme angeschafft werden oder die Daten müssen auf günstigere Speichermedien verschoben werden um wieder neue Kapazitäten zu schaffen.

Produktverbesserungen bei Hard- und Software

Aufgrund des exponentiellen Wachstums der von den operativen Unternehmensbereichen generierten Datenmengen aber auch durch laufende Produktverbesserungen bei Hard- und Software sind IT-Abteilungen einem ständigen Wandel unterworfen.

2 Begriffsdefinition

Als Datenmigration bezeichnet man den Transfer von Daten aus einer Umgebung in eine andere. Das heißt, es werden Daten von einem Quellsystem in ein Zielsystem überführt. Das Quell- und Zielsystem können sich strukturell unterscheiden, was den Transfer erschwert. Zum Einen, könnte das Zielsystem so aussehen, dass es existierende Aufgaben einer oder mehrerer anderer Systeme übernimmt oder zum Anderen, kann man das Quellsystem durch ein bereits produktives System ablösen.

Eine Datenmigration ist ein einmaliger Prozess und hat ein definiertes Ende, der Prozess kann jedoch über einen bestimmten Zeitraum stattfinden. Die einzelnen Prozessschritte werden jeweils durch bestimmte Tools unterstützt.

3 Ablauf einer Datenmigration [AI 07]

- Download
- Preprocessing
- Transfer/Merge
- Upload
- Processing

3.1 Download

Angefangen wird der Prozess der Datenmigration mit dem Download. Der Zweck dieses Schrittes ist es die

Verfügbarkeit der Daten in einem Transport- und Weiterverarbeitungsmedium sicher zu stellen. Dies kann beispielsweise durch sogenannte „Flat Files“ geschehen.

Die genaue Aufgabe dieses Prozess-Schrittes ist es die Daten aus dem Quellsystem zu exportieren. Dazu können Export Utilities der Quellsysteme benutzt werden. Ein weiteres Tool für diesen Export sind sogenannte Stored Procedures. In einer gespeicherten Prozedur können ganze Abläufe von Anweisungen unter einem Namen gespeichert werden, die dann auf dem Datenbankserver zur Verfügung stehen und ausgeführt werden können. Das heißt, sie kann eine Abfolge gespeicherter Befehle ausführen. Desweiteren können hierzu Datenbank Export Tools zum exportieren der Quelldaten herangezogen werden. Das wichtigste Werkzeug aber sind die ETL- Tools.

ETL-Prozess: Dieser Prozess führt Daten aus verschiedenen Quellsystemen zusammen. Aufgeteilt ist dieser Prozess in drei Schritte:

- Extraktion
- Transformation
- Laden

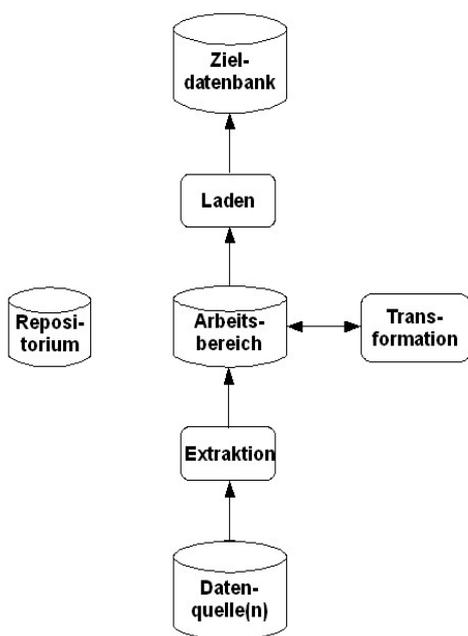


Abbildung 1: ETL-Prozess: Datenfluss und unterstützende Programme [Tz 09]

In der ersten Phase werden alle wichtigen Daten aus den Quellsystemen entnommen, daraufhin wird der Datensatz in das Format des Zielsystems transferiert und schließlich in die Zieldatenbank hochgeladen.

Für den ersten Schritt der Datenmigration, also dem Download ist auch der erste Schritt des ETL-Prozesses (Extraktion) relevant.

- Preprocessing

Das Ziel dieses Prozess-Schrittes ist es, ein einheitliches Format der Quelldaten zu erreichen, um diese einfacher weiterzuverarbeiten. Die Aufgaben sind abgesehen von der Überführung aller Quelldaten in ein gemeinsames (Zwischen-)Format, ist die Vorverarbeitung dieser Daten durch die Konsolidierung und Formatierung.

Die wichtigsten Werkzeuge hierfür sind Skript- und Programmiersprachen, Texteditormakros oder Excel.

- Transfer/Merge

In diesem Teilprozess ist das Ziel die Verfügbarkeit aller Daten für das Zielsystem sicher zu stellen. Der Transfer und das Zusammenführen der Daten aus den verschiedenen Quellsystemen sind die Hauptaufgaben dieses Schrittes.

Das Tool für den Transfer ist das FTP (File Transfer Protocol). Dieses Dateiübertragungsverfahren ist ein Netzwerkprotokoll zur Übertragung von Daten über IP-Netzwerke. Es wird benutzt, um Dateien vom Server zum Client (Herunterladen), vom Client zum Server (Hochladen) oder clientgesteuert zwischen zwei Endgeräten zu übertragen.

Für das Zusammenführen der Daten ist es möglich das Append/Copy-Verfahren oder einen einfachen Editor zu nutzen.

Upload

Die Bereitstellung der Daten in der Datenbank des Zielsystems ist das Hauptziel des Upload, indem die Daten in Schnittstellenstrukturen importiert werden. Werkzeuge hierfür sind wieder die ETL-Tools, diesmal jedoch der Prozessschritt des „Laden“, und die Datenbank API.

Processing

Das Ziel im Processing ist die endgültige Überführung der Daten in das Produktivsystem. Die Aufgaben sind hier unter anderem die Eliminierung doppelter Datensätze. Diese Aufgaben kann man durch Datenbankprozeduren oder durch Programmiersprachen wie Java,C oder Pascal realisieren.

4 Gesamtzusammenhang im Datenmigrationsprojekt

- Datensemantik
- Ablaufsteuerung
- Data-Cleansing
- Generalprobe
- Migration**
- Nachbereitung

Datensemantik

Aufgaben dieser Phase:

- Business rules
- Fachliche Rahmenbedingungen
- Enge Abstimmung zwischen Datenmodell-Experten

Ablaufsteuerung

Aufgaben:

- Konvertierungs- und Steuerungsprogramme
- Performance Maßnahmen (technische/organisatorische)
- Maßnahmen für die Qualitätssicherung (QS)

Data-Cleansing

Aufgaben:

- Wiederholte Testläufe mit Echtdate
- Datenqualität testen
- Testen der Migrationsskripte auf Qualität und Vollständigkeit

Ziel: alle Daten- und Skriptfehler finden und korrigieren

4.4 Generalprobe

Aufgaben:

- Finaler Testlauf
- Zusammensetzen aller bisher getesteter Verfahren

4.5 Migration

Dieser Prozess-Abschnitt wurde bereits unter 3. Ablauf der Datenmigration erläutert

- Nachbereitung

Aufgaben:

- Abnahme
- Nachbesserung von Fehlern, die die Produktion nicht verhindern
- Konsolidierung Dokumentation

5 Zusammenhang mit IT-Transformationen

6 Zusammenfassung

Literaturverzeichnis

Internetquellen