

In dieser Ausgabe:

- ⇒ Datenmigration - Ein Überblick
- ⇒ Durchführung und Test bei der Datenmigration
- ⇒ Literatur und Zitate

Datenmigration - Ein kurzer Überblick

Unter **Migration** versteht man die Umstellung eines wesentlichen Teils der eingesetzten Software in eine andere Umgebung (z.B. von Windows XP auf Windows Vista oder von Windows auf Linux) beziehungsweise den Transfer von Daten aus einer Umgebung in eine andere (z.B. altes Datenbanksystem auf ein neues) sowie die Umstellung von Hardware einer alten Technologie in neue Technologien. Von der Vorgehensweise grundsätzlich zu unterscheiden sind daher die „Software-Migration“, „Hardware-Migration“ und „Daten-Migration“.

Die **Daten-Migration** ist das Überführen der Daten in einem bestimmten Zustand und System zu einem neuen Zustand bzw. System und kann sowohl die Datenformate (z.B. von einer Dateispeicherung hin zu einer relationalen Datenbank oder von flachen ASCII-Dateien hin zu XML-Strukturen) als auch die Datenstrukturen wie das Datenbank-Schema betreffen (z.B. redundante Speicherung wird übergeführt in eine normalisierte Tabellen-Struktur oder zwei Tabellen werden zu einer verschmolzen, ...).

Im ersten, deutlich einfacheren Fall spricht man auch von Daten-Konvertierung, im zweiten ist eine Umformung der Daten oder Schema-Konvertierung notwendig.

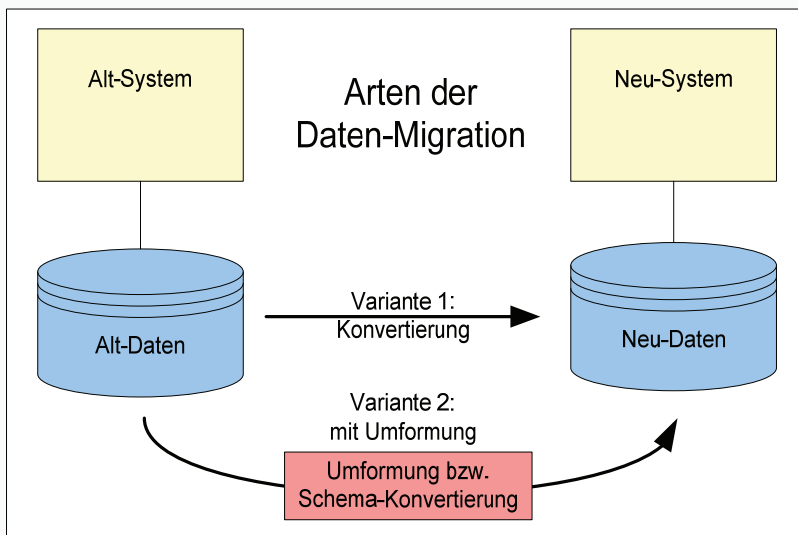


Abb. 1: Varianten der Datenmigration
Quelle: J. Bergsmann

Diese beiden Varianten unterscheiden sich ganz wesentlich sowohl in der Durchführung der Datenmigration als auch beim Test der Migration.

In diesem Knowledge-Letter werden einige wesentliche Aspekte der Datenmigration und des Tests der Daten-Migration aufgezeigt.

Weiterführende Informationen erhalten Sie gerne auf Anfrage.



Wer denkt schon an Datenmigration?

Im Rahmen der Durchführung von Projektaudits oder bei projektbegleitendem Qualitätsmanagement stelle ich immer wieder fest, dass es zwar detaillierte Projektpläne gibt, die alles mögliche enthalten, aber auf die Datenmigration vergessen.

Vor allem auch bei Standard-Software-Projekten ist auffälligerweise oft so. Dabei ist die Datenmigration ein Teil im Projekt, der zu enormen Aufwänden führen kann.

Wenn dann die Projektbeteiligten auf das Fehlen der Datenmigration angesprochen werden, kommen unterschiedliche Reaktionen:

- ⇒ Der Auftraggeber hat nicht daran gedacht, weil er selbstverständlich meint, dass sich der Auftragnehmer darum kümmert und das alles im Projektpreis enthalten ist.
- ⇒ Der Auftragnehmer will sich mit dem Thema nicht auseinandersetzen, weil das nur viel Staub aufwirbelt und die Angebotssumme erhöht. Oft gibt es auch die Aussage: „Darum kümmern wir uns im Rahmen der Implementierung“.

Leider ist es jedoch so, dass dieser Aufwandsbrocken unausweichlich irgendwann im Projekt umzusetzen sein wird, auch wenn er nicht im Projektplan aufscheint.

Und erst kurz vor dem Go-Life darüber nachzudenken, wie man denn nun auch seine alten Daten in das neue System hinein bekommt (ganz abgesehen davon wie man testet, ob die Migration auch erfolgreich war), ist sicherlich kein guter Projektansatz.

Dipl.-Ing. Johannes Bergsmann

Staatl. befugter und beideter Ingenieurkonsultent für Informatik

Der Quality-Knowledgeletter ist ein periodisches Informationsmedium von Software Quality Lab und dessen Partnern mit den Schwerpunkten IT-Qualitätsmanagement, Projekt- und Prozess-Management.

Inhalt: fachliche Beiträge und Schwerpunktthemen, Vorstellung neuer Produkte und Leistungen, neue wissenschaftliche Erkenntnisse und andere Fachbeiträge aus unseren Themenbereichen.

Aktuelle Fach- und Forschungsbeiträge sind willkommen. Einsendungen an info@software-quality-lab.at.

Weitere Infos zu diesem und anderen Themen finden Sie auf <http://www.software-quality-lab.at>.

Qualitätssicherung im Rahmen der Daten-Migration

von Dipl.-Ing. Bernhard Meingassner und Dipl.-Ing. Johannes Bergmann

Bei der Umstellung auf eine neue Software ist die verlustfreie Konvertierung der Daten aus dem „alten“ System in das neue System von großer Bedeutung. Selbst bei kleinen Systemen wächst der Datenbestand im Laufe der Jahre auf einen oft so großen Umfang an, dass es nicht mehr möglich ist, jeden einzelnen Datensatz auf korrekte Konvertierung zu überprüfen.

Daher muss hier ein entsprechend methodischer Testansatz angewendet werden, damit die Testaufwände nicht explodieren aber trotzdem eine ausreichende Sicherheit und Qualität der Datenmigration überprüft werden kann.

Dieser Artikel soll einen kurzen Überblick über Vorgangsweisen und Testmethodiken bei der Datenmigration geben.

Planung — Die Datenmigration als eigenes Projekt

Das Thema Datenmigration wird in den meisten Projekten nicht rechtzeitig, sondern erst am Ende der SW-Entwicklung oder gar des Softwaretests betrachtet und damit hat meist jede Verzögerung in diesem Bereich sofort unmittelbare Auswirkung auf den Endtermin des gesamten Projekts!

Aus dieser Sicht ist das Teilprojekt Datenmigration sogar ein zeitlich kritisches Teilprojekt. Die Datenmigration sollte schon bei der Konzeption der Software als eigenständiges Projekt angesehen werden, damit kann der gesamte Testprozess explizit betrachtet und auch eingeplant werden.

Die Teilbereiche des Datenmigrationstests wie Test-Planung, Test-Spezifikation, Test-Umgebung, Test-Durchführung und Test-Ende-Bewertung können sich z.B. in der Qualifikation der Testteams, der zur Verfügung stehenden Testumgebung (Testdatenbanken, Parametrierdatenbanken) oder Testmethodik unterscheiden.

Es kann damit schon am Projektbeginn auf diese geänderten Bedingungen Rücksicht genommen werden:

Benötigt man mehrere spezielle leistungsfähige Datenbankserver, ausgebildete Datenbankspezialisten, sind spezielle Abnahmemetriken für die Datenmigration erforderlich, ...?

Warnung / Hinweis:

In Projekten, in denen Altsysteme abgelöst werden sollen und daher meistens auch verschiedene Daten in das neue System übertragen werden müssen, sollte der Projektteil „Datenmigration“ inkl. Test der Datenmigration jedenfalls schon in der Angebotsphase berücksichtigt und zeit- und kostenmäßig eingeplant werden.

Manche Anbieter haben diesen Aufwandspunkt gar nicht in ihrem Angebot oder Projektplan.

Anbieter oder Berater, welche die Datenmigration nicht als eigenen Punkt im Angebot bzw. der Projektplanung angeführt und zumindest grob einen Aufwand geschätzt haben, oder die auf die Frage nach der fehlenden Datenmigration mit der folgenden so oder ähnlich formulierten Aussage „Um die Datenmigration kümmern wir uns im Rahmen der Umsetzung.“ antworten, sind jedenfalls mit großer Vorsicht zu bewerten!

Die Tragweite und kostenmäßige Auswirkung der Datenmigration wird hier ev. bewusst ignoriert oder mangels Erfahrung des Lieferanten nicht angesprochen.

Kundenorientierte, ehrliche und praxiserfahrene Lieferanten und Berater werden diesen Punkt jedenfalls immer ansprechen und im Rahmen der Kunden-Analyse auch versuchen, dies kostenmäßig abzuschätzen.

In der Praxis haben sich folgende Methodiken bewährt:

Risikoklassenbildung

100% Testen bzw. eine vollständige Überprüfung aller Datensätze ist bei sehr großen Datenbeständen eventuell wirtschaftlich nicht sinnvoll und teilweise auch technisch nicht durchführbar (weil z.B. der vollständige Vergleich aller Datensätze in der Datenbank zu lange dauern würde — z.B. wenn ein vollständiger Vergleich eines Datenbestands von 10 Millionen Datensätzen 4 Tage in Anspruch nehmen würde, jedoch für die Datenmigration nur ein „Fenster“ von 1 Tag zur Verfügung steht, weil die Produktion eventuell keine längere Ausfallszeit zulässt).

In diesen Fällen empfiehlt sich eine risikoorientierte Vorgehensweise. Dabei werden beim Testen alle Tabellen bzw. Felder der Datenbank in Risikoklassen eingeteilt.

Alle Felder mit höheren Risikoklassen werden vorrangig bzw. eventuell auch nach einer speziell für diese Risikoklasse festgelegten Testmethodik getestet.

Diese Vorgangsweise wird in der Testspezifikation berücksichtigt und die dafür notwendigen Testszenarien an die Methodik geknüpft.

Die Testtiefe der Testspezifikation wird über die Risikoklasse gesteuert: Je risikoreicher die Datensätze sind, desto intensiver wird getestet (bis hin zu einem vollständigen Datensatzvergleich aller Hochrisiko-Datensätze) und je unkritischer die Daten sind, desto weniger intensiv wird in diesem Bereich die Datenübernahme getestet (z.B. nur Zählen der migrierten Datensätze in dieser Risikoklasse).

Es empfiehlt sich auch, für Felder und Tabellen mit hoher bzw. mittlerer Risikoklasse Äquivalenzklassen zu bilden:

Äquivalenzklassen/Grenzwertanalyse

Bei der Äquivalenzklassenbildung werden die möglichen Werte einer Eingabengröße in Klassen eingeteilt – jede Äquivalenzklasse steht für eine Ausprägung des Datenfeldes (z.B. für das Feld Geschlecht gibt es typischerweise 2 gültige Äquivalenzklassen „m“ und „w“).

Zur Überprüfung der Richtigkeit der Datenmigration wird z.B. die Anzahl der Datensätze jeder Äquivalenzklasse im Altdatenbestand mit der Anzahl der Datensätze jeder Äquivalenzklasse dieses Feldes im neuen System verglichen.

Wenn z.B. im Altsystem 120324 männliche und 42334 weibliche Personen eingetragen sind, so muss im Neusystem

(Fortsetzung auf Seite 3)

(Fortsetzung von Seite 2)

genau die selbe Anzahl männlicher und weiblicher Personen vorhanden sein —ausgenommen natürlich, wenn im Rahmen der Datenmigration eine Datenbereinigung erfolgte und daher aus diesem Grund die Datensatzanzahl gewollt unterschiedlich ist.

Anmerkung: Auf diesen Fall wird im Rahmen dieses Artikels nicht eingegangen, er muss jedoch beim Datenmigrations-Test ebenfalls berücksichtigt werden.

Ist für die getestete Größe auch ein Grenzwert entscheidend, so wird die dafür notwendige Grenzwertanalyse bei der Auswahl der zu testenden Klassen mit einbezogen.

Zusätzlich dazu wird dort, wo dies sinnvoll ist, eine eigene Klasse für ungültige Bereiche (Negativtest) gebildet.

In vielen Fällen ist es jedoch nicht ausreichend, wenn der Test der Datenmigration nur bezogen auf die Äquivalenzklassen eines Feldes bzw. Tabelle durchgeführt wird.

Vor allem bei risikoreichen Datenbankfeldern werden auch Kombinationen von Äquivalenzklassen sinnvoll sein (z.B. alle Personen, die „HIV-infiziert“ sind und jünger als „20 Jahre“ sind - hier werden Äquivalenzklassen der Felder „HIV-Infektion“ und „Alter“ kombiniert).

Die durch Kombinatorik von Äquivalenzklassen ableitbaren Testfälle können jedoch schnell eine praktisch nicht mehr durchführbare Anzahl von Testfällen überschreiten.

Eine Reduktion der Testfälle ist hier notwendig. Ein Verfahren, die Testfälle zu reduzieren und trotzdem noch eine verhältnismäßig gute Test-Qualität der Migration zu erzielen ist der Pairwise-Ansatz, der im Folgenden erklärt wird:

Pairwise-Methode

Untersuchungen haben ergeben, dass ein sehr hoher Anteil der Fehler beim Testen von Wertepaaren entsteht.

Eine Methode, die sich dieses Ergebnis zu nutze macht, ist die Pairwise-Methode.

Hier werden Testfälle, bei denen einzelne Wertepaare schon einmal mit einem Testfall überprüft wurden, nicht nochmals getestet.

Gibt es zum Beispiel in einer Tabelle drei Felder, die jeweils drei verschiedene Ausprägungsgruppen (Äquivalenzklassen - „ÄK“) annehmen können, so müssten für eine komplette Abdeckung aller möglichen Kombinationen der Äquivalenzklassen 27 Tests durchgeführt werden:

Nr	ÄK 1	ÄK 2	ÄK 3
1	A	a	1
2	A	a	2
3	A	a	3
4	A	b	1
5	A	b	2
6	A	b	3
7	A	c	1
...
26	C	c	2
27	C	c	3

Zur Reduktion der Zeilen kann folgende Vorgangsweise gewählt werden:

Man geht von der untersten Zeile der Tabelle beginnend aus und streicht jede Zeile, in der die vorkommende Kombinationen von jeweils zwei Werten bereits in einer der bisher nicht gestrichenen Zeilen vorkommt. Dies setzt man so lange fort, bis keine Zeile mehr gelöscht werden kann.

Mit dieser Vorgehensweise kann die Anzahl der Tests auf 9 reduziert werden:

Nr.	ÄK 1	ÄK 2	ÄK 3
2	A	a	2
6	A	b	3
7	A	c	1
13	B	b	1
15	B	b	3
17	B	c	2
19	C	a	1
23	C	b	2
27	C	c	3

ACHTUNG: Fehler, die nur bei bestimmten Kombinationen von mehr als zwei Feldern auftreten (sogenannte Multi-Mode-Faults), werden beim Einsatz der Pairwise-Methode unter Umständen übersehen!

Daher sollten in der Risikoanalyse entschieden werden, welche Risikoklassen mit der Pairwise-Methode reduziert getestet werden (z.B. nur Äquivalenzklassen mit mittlerer Risikoklasse). Für andere Risikoklassen (z.B. hohes Risiko) sollten dann gegebenen Falls auch vollständige Tests durchgeführt werden, die möglicherweise auch Kombinationen von mehr als 2 Feldern berücksichtigen.

Vollständiger Datensatzvergleich

Falls das Risiko einer fehlerhaften Datenübernahme für einzelne Felder sehr groß ist (z.B. bei hochkritischen oder sensiblen Datensätzen), sollte ein vollständiger Datensatzvergleich durchgeführt werden.

Speziell in diesem Bereich sollte sehr genau geprüft werden, in welchem Verhältnis sich Risiko und Aufwand bewegen.

Ein vollständiger Datensatzvergleich ist vor allem bei komplexen und großen Datenbeständen bzw. auch im Falle von notwendigen Umformungen von Daten ein sehr zeitaufwendiger und damit zeitkritischer Faktor im Testzyklus und muss in der Testplanung entsprechend berücksichtigt werden.

Prüfsummen

Eine Möglichkeit der Überprüfung der Datenmigration ist die Prüfsummenbildung.

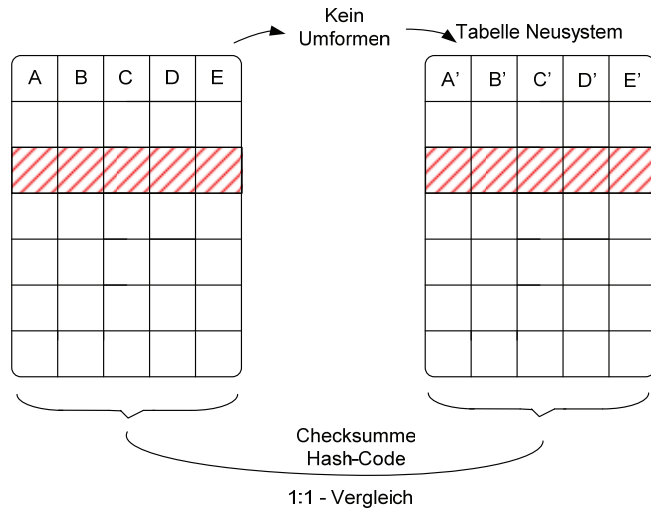
Dieses Verfahren zur Fehlererkennung, das z.B. in vielen Codierungs-Systemen implementiert wird, kann im einfachsten Fall durch Bildung der Quersumme eines Datensatzes errechnet werden.

Bei anderen Verfahren kommen kompliziertere mathematische Algorithmen (z.B. HASH-Funktionen) zum Einsatz.

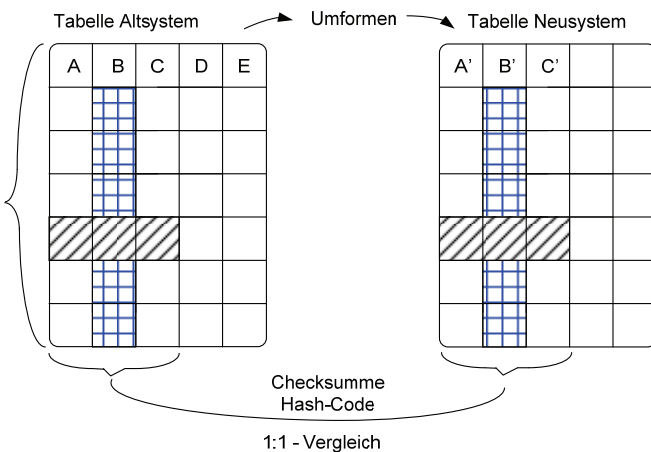
(Fortsetzung auf Seite 4)

(Fortsetzung von Seite 3)

Im Bereich der Datenmigration ist es wichtig, zu unterscheiden, ob bei der Konvertierung der Daten vom Altsystem in das Neusystem eine Umformung durchgeführt werden muss.



Ist keine Umformung notwendig, so kann die Prüfsumme eines ganzen Datensatzes mittels einer Prüfroutine ermittelt werden. Es müssen in diesem Fall die Prüfsummen der beiden Datensätze (Alt-Daten und Neu-Daten) übereinstimmen, in dem oben angeführten Beispiel muss also die Prüfsumme alle Felder A,B,C,D,E mit der Prüfsumme A',B',C', D',E' übereinstimmen.



Findet eine Umformung zwischen den Datenbanken statt, so können nur diejenige Felder, bei denen keine Umformung stattfindet, direkt geprüft werden – die Prüfsumme der Felder A,B,C muss mit der Prüfsumme der Felder A',B',C' übereinstimmen. Dies stellt auf Ebene eines Datensatzes sicher, dass die Felder A,B,C korrekt in das Neusystem übertragen wurden.

Eine weitere Möglichkeit der Prüfung besteht darin, dass die Prüfsumme über für eine Spalte (z.B. alle Datensätze für das Feld B) gebildet wird und mit der Prüfsumme der entsprechenden Spalte (z.B. B') verglichen wird. Dies stellt sicher, dass eine komplette Spalte korrekt in das Neusystem übertragen wurde.

Die Testdurchführung der Migration von Datensätzen über **Prüfsummen (Funktion „Checksum“)** bietet sich vor allem bei Datenbanken an, die diese Prüfsummen als Funktion direkt anbieten (z.B. Oracle, MS SQL-Server). Für andere Datenbanken muss eine eigene Checksum-Routine programmiert werden.

Neben den genannten Möglichkeiten für eine Überprüfung der Datenmigration gibt es noch weitere Möglichkeiten. Außerdem müssen abhängig von den jeweiligen Rahmenbedingungen noch weitere Aspekte (z.B. Datenqualität, Datenbereinigung, ...) berücksichtigt werden.

Weitere Infos auf Anfrage.

Literatur und Links

- ⇒ Datenmigration in SAP, Galileo Press, 2007 ISBN 3-8362-1037-1

Leistungen

Im Rahmen der **Datenmigration** unterstützen die Experten von Software Quality Lab als neutrale Berater durch folgende Leistungen:

- ⇒ **Planung und Aufwandsabschätzung** der Datenmigration
- ⇒ Unterstützung beim Erstellen von ausreichenden **Spezifikationen für die Datenmigration**
- ⇒ Erstellen von **Testplänen** sowie den entsprechenden **Testspezifikationen** für die Datenmigration
- ⇒ Aufsetzen von passenden **Testumgebungen**.
- ⇒ **Durchführen von Datenmigrations-Tests**
- ⇒ **Automatisierung** der Datenmigrations-Tests
- ⇒ Begleitendes **Qualitätsmanagement** in größeren Projekten
- ⇒ **Coaching** von Projektleitern.
- ⇒ ...

Zitate

„If you're going to get bad news on a software project, it's best to get it early because of the upstream/downstream effect on project costs.“

Steve MC Connell, Software Project Survival Guide

„The problem with quick and dirty, as some people have said, is that dirty remains long after quick has been forgotten.“

Steve MC Connell, Software Project Survival Guide

„Viel Glück bei Ihrem nächsten Projekt - aber verlassen Sie sich nicht darauf!“

Tim Lister