

Eine Data Warehouse-Landschaft zieht um.

IBM/Power auf HP-RHEL

**Heiko Becker
comdirect bank AG
Quickborn**

Schlüsselworte

Migration, Projektbericht, IBM, PowerPC, x86, HP, Linux, Little Endian, Big Endian, Oracle, Data Warehouse

comdirect Data Warehouse-Landschaft

Das comdirect Data Warehouse ist über Jahre gewachsen und hatte seinen Ursprung im Marketing. Mittlerweile hat es sich zu einer firmenweiten Lösung weiter entwickelt.

Insbesondere technologisch hat sich die Welt bewegt und dabei hat dieses sich auch bei der comdirect gezeigt.

Landschaften lassen sich aus verschiedenen Perspektiven betrachten. So ist es möglich, die comdirect Data Warehouse Landschaft aus technologischer, fachlicher oder organisatorischer Perspektive zu betrachten. Dabei stehen die unterschiedlichen Perspektiven in einer Wechselbeziehung, z.B. fachliche Anforderungen wirken sich auf technologische Fragestellungen aus und organisatorische Aspekte ergänzen die fachlichen Anforderungen.

Die fachliche Perspektive zeigt dabei Lösungen im Bereich Marketing & Vertriebsunterstützungen, Finanz & Controlling und insbesondere im Bereich der bankfachlichen Themen, wie Risikobewertungen und einer Vielzahl von regulatorische Anforderungen.

Die organisatorische Perspektive muss die Prozesse zum Anforderungsmanagement, der Anwendungsentwicklung und zur Anwendungssteuerung unterstützen. Dabei geht es um Prioritäten und die strategische Weiterentwicklung der analytischen Systeme. In der comdirect bank AG wurde in diesem Umfeld eine BI Organisation entwickelt, die einen regelmäßigen Austausch zwischen den unterschiedlichen beteiligten Fachgebieten gewährleistet und Verantwortlichkeiten regelt.

Wird der Blick auf die technologische Perspektive gerichtet, stehen die Komponenten der Infrastrukturen im Vordergrund. Die Infrastruktur des comdirect Data Warehouse basiert auf den Vorgaben der IT-Strategie. Dabei wurde im Datenbankumfeld auf IBM-Server mit Power-Prozessoren und auf auf Oracle gesetzt. Ein Vorteil der bisherigen Plattform war die Partitionierung der Hardware und dynamische Anpassungen und Verteilung von Hardwareressourcen bei einem dynamischen Geschäftsumfeld. Im Middletier-Bereich wird bereits seit Jahren auf eine Farm aus Intel x86 Serverblades gesetzt.

Grundlegende Richtungsänderungen in einem existierendem Umfeld

Nichts ist so beständig wie der Wandel. Die technologische Weiterentwicklung und Entwicklungen am Markt beeinflussen die in der Vergangenheit getroffenen Entscheidungen. Um mit diesem Wandel Schritt zu halten, bedarf es einer wiederkehrenden Überprüfung der vorhandenen Strategien und Lösungen. Wobei in kürzeren Zyklen einzelne Inseln und in größeren Zyklen die Gesamtstrategie hinterfragt werden sollte. Die kürzeren Zyklen entstehen häufig aus Risikoabwägungen und geplanten Investitionsrechnungen für einzelne Systembestandteile.

Bei diesem Projekt war das Ziel die Anwendungsarchitektur beizubehalten, insbesondere die Kompatibilität der Bewirtschaftungsprozesse und implementierten fachlichen Anwendungen.

Somit fokussierte sich dieses Projekt auf den Austausch der Serverplattform, welche in den vergangenen Jahren auf einem IBM pSeries p595-System basierte.

Bereits in wenigen Workshops konnten die möglichen Lösungen für die Zukunft auf die folgenden Szenarien reduziert werden:

- Fortsetzung mit IBM PowerPC basierten Lösungen, in diesem Fall pSeries p-770-Systeme.
- Wechsel zu HP ProLiant, Intel x86 und Red Hat Enterprise Linux (RHEL) Lösungen.
- Einführung der Appliance Lösung Exadata von Oracle.

Der Vergleich von Systemen ist eine immer wiederkehrende Aufgabe in der Informationstechnologie. In einigen Fällen gibt es keine Historie, also Erfahrungen mit einem potenziellen System, was einen Vergleich erschwert. Bei vorhandenen Systemen wachsen das Vertrauen und die Erfahrung mit der Zeit die ein System in einem Unternehmen genutzt wird. Dieses Ungleichgewicht bedingt auch Aspekte eines Change-Managements.

Um aber die Fragestellung, „Welches der möglichen Szenarien ist die optimalste Lösung für das gegebene Umfeld?“, zu beantworten ist ein Vergleich zwischen der vorhandenen und den möglichen neuen Optionen notwendig. Für einen Vergleich bietet sich die Definition von Kriterien und von Testfällen an. Dabei ist auch die Ausführung der Testfälle auf dem vorhandenen System relevant. Eine solche Messung kann als Baseline bezeichnet werden, gegen welches sich alle folgenden Szenarien vergleichen lassen müssen.

Aber ist es denn überhaupt notwendig Testfälle, zu erstellen?¹ Andere Güter werden nach bekannten oder zu definierenden Kriterien ausgewählt, Beispiele sind in der folgenden Tabelle dargestellt.

Auto	PS/KW, Höchstgeschwindigkeit, Anzahl Sitze, Kosten
Fahrrad	Gewicht, Bereifung
Lautsprecher	Frequenzbereich

Auch Server lassen sich entsprechend vergleichen, dabei sind es zum Beispiel Faktoren wie die Taktungsrate der CPU, die maximalen Übertragungsraten zwischen CPU und RAM und Transferleistung zwischen Speichersystem via Bussystem zum RAM.

Fakt ist aber, dass die oben genannten Güter auch Prüfungen unterzogen werden. So werden Probefahrten unternommen, um subjektive Eindrücke zu gewinnen und bisher nicht betrachtete oder definierte Kriterien zu prüfen.

Ein Testfall dient einer Normierung der Prüfung, mit dem Ziel eine Vergleichbarkeit zu ermöglichen. Die Normierung spielt auch bei einigen weniger wichtigen Kriterien eine Rolle, bei einem Auto würde die Distanz die mit dem vollgetankten Auto zurück gelegt werden könnte, keine Aussage ermöglichen, wohl aber der Verbrauch pro 100 km. Bezogen auf den Vergleich der Server wird der für den Testfall zur Verfügung stehende Speicher (RAM) festgelegt, also begrenzt, und in jedem Fall gleich gehalten. Im Vergleich der Systeme spielt das Kriterium RAM dennoch eine Rolle und zwar in der maximalen Ausbaugröße und deren Kosten.

¹ Vergleiche <http://www.inout.ch/produkte/>

Für die Evaluation in diesem Projekt wurde ein Datenmodell aus dem Data Warehouse herausgelöst, welches es ermöglicht, Ladeprozesse und Auswertungsabfragen zu simulieren. Das Datenvolumen entsprach dabei ca. 250 GB, und es wurden 12 Test SQL Skripte entwickelt.

In den beiden folgenden Abbildungen sind die drei Systeme gegenübergestellt. Während bei den Kosten ein signifikanter Unterschied zu erkennen ist, sind bei dem Data Warehouse Test keine signifikanten Unterschiede durch die Testfälle ersichtlich. Im Vergleich zur Baseline war dennoch ein Faktor von drei zu sehen.

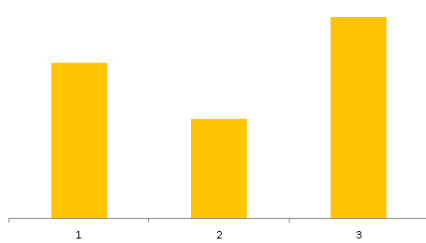


Abb. 1: Kosten inkl. Energie

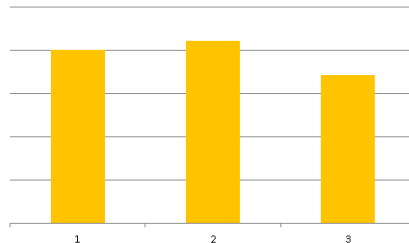


Abb. 2: Laufzeittest

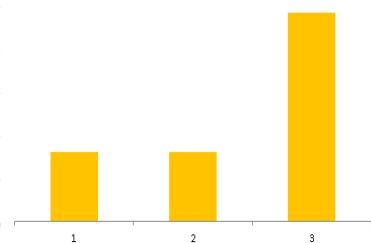


Abb. 3: Kriterien

Vorbereitung

Nach einer entsprechenden Entscheidung für eine entsprechende Richtungsänderung, steht die Planung für eine entsprechende Umstellung an.

Bei Aufgaben, bei denen die Erfahrungen im eigenen Haus fehlen und die auch keinen größeren Mehrwert für die Zukunft bieten, ist es eine sinnvolle Maßnahme, externe Unterstützung einzubeziehen. In diesem Fall boten sich Dienstleister an, welche die Umgebung schon kannten und entsprechende Expertisen vorzuzeigen hatten.

Neben der Auswahl der Mannschaft wird bereits in dieser Phase über mögliche Alternativen bei der Umsetzung diskutiert. Dabei wurden folgende Umsetzungsalternativen thematisiert:

- Oracle Data Guard
- Oracle GoldenGate
- Quest Shareplex
- dbVisit
- Oracle Transportable Tablespaces
- Oracle RMAN convert
- Oracle Data Pump

Bei der ausgewählten Richtungsänderung, von der Power Prozessorarchitektur wegzugehen, hin zu der Intel Prozessorarchitektur, kommt die Endian² Problematik zum Tragen.

Die Power Prozessorarchitektur sieht bei der Speicherung von Multi-Byte Informationen vor das Byte mit den höchstwertigen/signifikantesten Bit zuerst zu speichern. Bei der Intel x86 Architektur wird dagegen das Byte mit den niederwertigen Bits an der kleinsten Speicheradresse gespeichert. Dieses hat zur Folge, dass die physische Speicherung von Datenbankblöcken umgewandelt werden muß. Diese technische Gegebenheit beeinflusst unter anderem die möglichen Umsetzungsalternativen.

Weitere Aspekte, die die zur Auswahl der für dieses Projekt möglichen Alternativen beeinflussten, sind:

- Abkehr von einer Multi-Block-Konfiguration der Data Warehouse Datenbank.
- Reorganisation – oder warum ein Lager voll zu seinen scheint, obwohl es halbleer ist.
- Kosten
- Downtime – Nicht-Verfügbarkeit der Anwendungen.
- Gewinnung von Erfahrungen für weitere Systeme und deren Umstellung.

² Endian / Byte-Reihenfolge siehe Wikipedia - <http://de.wikipedia.org/wiki/Byte-Reihenfolge>

Bei der Multi-Block-Konfiguration wurde vor Jahren die Hypothese aufgestellt, eine Optimierung der Leistung durch die Verwendung von 8KB, 16KB und 32KB Blöcken und damit unterschiedliche Buffer-Caches zu erreichen, sowie Einfluß auf das IO Verhalten zu erhalten. Entsprechende Vorteile waren nicht mehr erkennbar, wobei Nachteile, z.B. ungünstige Ausnutzung vom RAM, offensichtlicher waren. Eine entsprechende Änderung auf die eine Standardgröße von 8KB bedingte aber auch eine Reorganisation. Nach mehr als 10 Jahren stand eine vollständige Reorganisation auf der Wunschliste aller Beteiligten. Diverse Tablespace beinhalteten Freiräume bzw. Luft, welche sich auch nicht durch regelmäßiges und einfaches „Housekeeping“ befreien lassen wollte. Was hilft eine Richtungsänderung, wenn sich diese nicht auch mit sinnvollen Kosten erreichen lässt. Unter Berücksichtigung dieser Aspekte fokussierten die Projektmitarbeiter die Oracle Data Pump Lösung.

Die Migration

Ein wesentlicher Punkt ist, neben den üblichen Artefakten wie Pflichtenheft, Konzeption und Sourcecode, das sogenannte Drehbuch. In der Umsetzungsphase wird dieses entwickelt und verfeinert. Es beschreibt den zeitlichen Fahrplan und einzelne Arbeitsschritte der Migration. Enthalten sollten folgende Aspekte sein:

- Kommunikation, wer, wen und wann – damit alle im Bild sind, wie es gerade läuft.
- Eskalation
- Quality Gates³, Punkte, an denen Entscheidungen auf Testergebnissen fallen für weitere Schritte oder sogenannte Fallbacks.
- Fallbacks – Was passiert, wenn noch unbekannte und nicht lösbare Probleme auftreten?

Während in der Umsetzungsphase das Drehbuch verfeinert wird und Skripte zur Automatisierung von Arbeitsschritten entwickelt werden, stehen in der Projekttest oder Integrationstestphase die Probeläufe bzw. die Generalprobe an. Die Produktionseinführung folgt dem Drehbuch und ermöglicht eine kontrollierte Einführung und Umstellung.

Eine Herausforderung an das Projekt ist es, dass die Migration nicht im sogenannten In-Place Verfahren möglich ist, d.h. die Daten müssen mindestens einmal zusätzlich vorgehalten werden. In diesem Projekt bedeutet dieses, die Daten werden via Data Pump in eine Datenbank auf einem „temporären“ Speichersystem migriert.

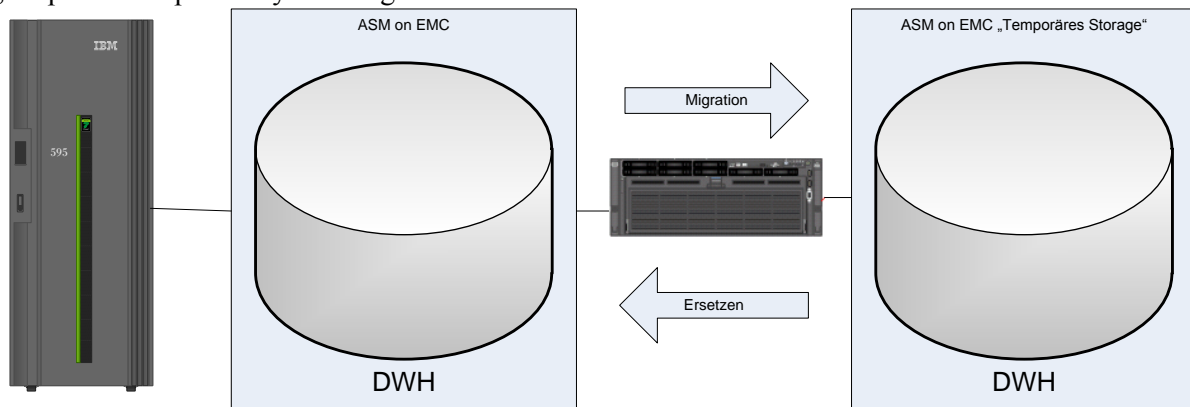


Abb. 4: Daten Rotation

Nach der Produktionseinführung wird das „alte“ Speichersystem bereinigt und im Anschluß können dann vom „temporären“ Speichersystem die Daten, in das vorhandene und bereinigte Speichersystem, überführt werden.

³ Quality Gate siehe Wikipedia - http://de.wikipedia.org/wiki/Quality_Gate

Neben der Migration der Daten sind auch Anwendungen und deren Bestandteile, die auf Basis von dem comdirect Data Warehouse entstanden sind, zu überführen. In diesem Fall sind das z.B.:

- ETL – eine Oracle Warehouse Builder Umgebung mit mehr als 1000 Ladeprozessschritten.
- Shell-Skripte
- Data Mining / SAS Integration
- Address- und Dublettenprüfungen
- Kampagnen-Management
- Oracle Essbase
- Schnittstellen an andere Systeme

In den meisten Fällen ist mit wenig Risiko zu rechnen, aber in einigen Fällen ist es notwendig, neue Softwareversionen einzuführen, so im Fall von der Address- und Dublettenprüfung oder dem Oracle Warehouse Builder. Ziel in allen Fällen ist es, dieselben Ergebnisse zu erreichen wie im bisherigen System.

Fazit

Selten gibt es die eine Lösung für den Umzug einer ganzen Data Warehouse Landschaft, welche nur Vorteile und keine Nachteile hat. Um eine Migration erfolgreich durchzuführen, ist es ein großer Vorteil, eine Vielzahl von Optionen zu haben und ein strukturiertes Vorgehen anzuwenden. Die Vielzahl von Umsetzungsalternativen ermöglicht im Problemfall alternative Lösungswege zu beschreiten und nicht in einer Sackgasse fest zu sitzen.

Ein Vergleich zwischen den durch eine Evaluation entstandenen Erwartungshaltung und der nach einer vollständigen Umstellung gewonnenen Kenntnisse ist erst nach einer längeren Betriebsphase möglich. Aber in Einzelfällen sind Verbesserungen auch schon während der Projektdurchführung ersichtlich. So lassen sich folgende positive Aspekte auflisten:

- Reorganisation brachte signifikante Reduktion beim Volumen
- Verbesserungen bei Abfragezeiten
- Kostenreduktion
- Vermeidung des sogenannten Vendor-Lock-In.

Die comdirect bank AG ist ein Unternehmen der comdirect Gruppe und steht für das Geschäftsfeld B2C. Als führende Direktbank und Marktführer im Online-Wertpapiergeschäft für moderne Anleger bietet comdirect Brokerage, Banking und Beratung. Darauf vertrauen mehr als 1,6 Millionen Privatkunden.

www.comdirect.de ist mit monatlich rund 200 Millionen Seitenaufrufen eine der meistbesuchten Finanz-Websites in Deutschland.

Kontaktadresse:

Heiko Becker
comdirect bank AG
Pascalkehre 15
D-25543 Quickborn

Dr. Gernot Schreib
b.telligent GmbH
Georg-Brauchle-Ring 54
D-80992 München

Telefon: +49 (0) 4103-704 1730
E-Mail heiko.becker@comdirect.de
Internet: www.comdirect.de

Telefon: +49 (0) 89-1222 811 10
E-Mail gernot.schreib@btelligent.de
Internet: www.btelligent.de