



Der Host bedient das Data Warehouse

Die Hapag-Lloyd Container Linie setzt auf Informatica Lösungen, um das Data Warehouse „Compass“ mit Daten aus einem operativen Mainframesystem zu versorgen. Dabei kommt ein Change Data Capture-Verfahren zum Einsatz, das die Aktualität der Informationen erhöht und gleichzeitig die Ressourcen schont.

(erschieden in Computerwoche 09/2005, S. 28; Autor: Kai Leonhardt)

Seit mehr als 150 Jahren verbindet Hapag-Lloyd Kontinente, Länder und Kulturen und folgt damit dem Credo ihres einstigen Generaldirektors Albert Ballin: „Mein Feld ist die Welt“. Der Schwerpunkt der Hapag-Lloyd-Aktivitäten liegt in der internationalen Container-Linienschiffahrt und dabei vorwiegend auf den weltumspannenden Ost-/West-Routen. Anspruch von Hapag-Lloyd ist es, Trends in Transport und Logistik immer einen Schritt voraus zu sein. Neben Kundenorientierung, Qualitätsbewusstsein sowie motivierten Mitarbeiterinnen und Mitarbeitern ist zukunftsorientierte Informationstechnologie dabei ein Schlüsselfaktor.

Die Hapag-Lloyd Container Linie stützt sich dazu auf eine IT-Infrastruktur mit drei Säulen: das operative Fracht-Informationssystem FIS, für die finanziellen Belange vertraut man auf SAP und im Bereich „Business Intelligence“ wird auf das eigenentwickelte Data Warehouse „Compass“ gesetzt. Während SAP auf einem AIX-basierenden Server läuft, werden FIS und Compass auf Mainframes betrieben. Auf allen Serversystemen kommt als Datenbankplattform DB2 zum Einsatz. Sämtliche Server

sind im Rahmen eines „Strategic Outsourcing“ an ein Rechenzentrum der IBM in Schweinfurt ausgelagert worden.

Weil die Bewirtschaftung des enorm wachsenden Data Warehouse einen immer höheren Aufwand erforderte, wurde Anfang 2004 die Umstrukturierung der IT-Umgebung in Angriff genommen. Mitte 2005 soll das Data Warehouse auf AIX migriert werden und die Informatica-Lösungen PowerCenter und PowerExchange werden dann den effizienten Datenaustausch zwischen dem FIS-Mainframe, SAP und Compass übernehmen.

Alles im Fluss durch e-Business

Produktionsbetriebe investieren vorrangig in Forschung und Entwicklung, um neue Produkte zu entwickeln. Dienstleister wie die Hapag-Lloyd Container Linie benötigen hingegen zukunftsweisende Lösungen, die dazu beitragen, dass die Betriebsabläufe optimiert oder innovative Konzepte für Kunden und/oder die Umwelt entwickelt werden. Wer in diesem Bereich die Investitionen über einen längeren Zeitraum vernachlässigt, verliert schnell den

Anschluss an den Wettbewerb. In der Hapag-Lloyd Container Linie floss in der Vergangenheit deshalb ein erheblicher Teil der IT-Investitionen in den Bereich e-Business.

Ziel ist es, den gesamten Daten- und Informationsaustausch mit Kunden und Lieferanten elektronisch abzuwickeln. Der Vorteil: Der manuelle Aufwand aller an der Transportkette Beteiligten wird reduziert, die Qualität der Information steigt. Dieser Datenaustausch kann über die Website der Hapag-Lloyd Container Linie, die als virtuelles Büro gestaltet ist, beziehungsweise über EDI-Lösungen oder den klassischen Customer Service erfolgen, den der Frachtkunde telefonisch oder per Fax nutzen kann. Darüber hinaus steht das Internet-Portal „Intra“ zur Verfügung, an dem das Unternehmen beteiligt ist. Durch Automatisierung soll den Kunden zukünftig eine Schnittstelle für alle Transportbelange zur Verfügung stehen, einschließlich „Tracking & Tracing“ über das Internet. Aus den im FIS generierten, operativen Prozessen zieht die Hapag-Lloyd Container Linie dann Daten ab wie Umsatz, Kosten, Kunden- und Transportinformationen, Container-Spezifika oder die Art der La-

dung. Alle bei der Ausführung dieser Prozesse im FIS anfallenden Daten werden in das Data Warehouse geladen.

Datenintegration zwischen Mainframe und AIX

In der Vergangenheit wurde ein Code generierendes Tool eingesetzt, um die Daten aus dem operativen FIS-System abzuholen, zu transformieren und in das Data Warehouse zu laden sowie umgekehrt Daten aus dem Data Warehouse dem FIS-System zur Verfügung zu stellen. Dabei muss jeder einzelne Prozess – von der Erfassung von Auftragseingängen über das Orderverfahren bis hin zur Beauftragung von Subunternehmen - in so genannte Conversions aufgeteilt und dann einzeln umgesetzt werden. Mit dem eingesetzten Cobol-Generator musste dazu bei jedem Schritt ein Programmcode entwickelt, kompiliert und auf den Mainframe „hochgeladen“ werden - was sehr zeitaufwändig ist, denn in einer so komplexen Umgebung finden immer wieder Änderungen statt, die neue Eingriffe erfordern. Quellsysteme werden verändert, neue Objekte werden hinzugefügt oder entfallen. Weil die User immer mehr und möglichst aktuelle Infos benötigen, steigen die Anforderungen und der gewünschte Grad der Aktualität ständig an. Bislang wurden Daten im Batch-Verfahren bewegt, bei dem große Datenmengen in kurzer Zeit repliziert werden, um die Datenhaltung auf Quell- und Zielsystemen zu synchronisieren. Hapag-Lloyd hat diese Daten wöchentlich aktualisiert. Neben dem zeit-

aufwändigen Transfer der kompletten Daten lag dies auch daran, dass dieses Verfahren immer wieder Rechenzeit auf dem Host im Rechenzentrum erforderte und dadurch Kosten mit sich brachte. Im Zuge der Überarbeitung sollte daher eine moderne Datenintegrationsplattform eingeführt und Compass gleichzeitig vom Mainframe auf AIX migriert werden.

Hapag-Lloyd evaluierte mehrere Datenintegrationstools und entschied sich schließlich nach dem Motto „Best of Breed“ für Informatica PowerCenter. Seit Anfang 2004 ist PowerCenter im Einsatz. Mit Unterstützung der Professional Services von Informatica werden derzeit alle Prozesse, die in drei Jahren Entwicklungsarbeit mit der alten Lösung erstellt wurden, in PowerCenter umgesetzt. Diese Arbeit wird mehr als ein Jahr in Anspruch nehmen. Ab Juni 2005 wird PowerCenter dann als Datenintegrationsplattform die neue Infrastruktur unterstützen und die Rolle der Extraktion der Daten von FIS und SAP, die Transformation dieser Daten und das Laden in das Data Warehouse übernehmen. Je nach Analysebedürfnissen werden die Daten dabei umgewandelt und in verschiedenartige „Sichten“ gebracht, also je nach Bedürfnissen der Anwender in Mappings zusammengefasst und visualisiert. Der große Vorteil an Informatica PowerCenter ist, dass die Software einfacher und transparenter umzusetzen ist und als enginebasierende Technologie quasi codefrei, das heißt ohne Programmieraufwand auskommt. Für Renate Hein, Projektleiterin von Com-

pass bei der Hapag Lloyd Container Linie, war besonders wichtig, dass die Arbeit mit PowerCenter sowohl die Produktivität als auch die Transparenz steigert.

CDC schlägt Batch

Ein weiterer entscheidender Faktor in der neuen Infrastruktur ist der Einsatz von PowerExchange. PowerExchange fungiert als technischer Verbindungslayer zwischen den Quell- und Zielsystemen und nutzt PowerCenter als Engine für die Datenbearbeitung und –bewegung. In herkömmlichen Batch-Verfahren kommt irgendwann der Punkt, an dem bei wachsenden Datenvolumina die Batch-Bewegung der Daten nicht mehr in operationale Zeitfenster passt. Dies ist beispielsweise der Fall, wenn Daten vom operational eingesetzten Mainframe mit einer Webanwendung synchronisiert werden müssen. Bisher war Hapag-Lloyd auf rechenintensive Vorher-/Nachher-Vergleiche basierend auf „Timestamps“ angewiesen. Dabei wurde beispielsweise ein ganzer Datensatz repliziert, obwohl sich letztlich nur zwei Felder geändert hatten. Aber die Veränderungen konnten nicht lokalisiert werden, nur das geänderte Datum. PowerExchange schafft hier durch Einsatz eines „Change Data Capture“ (CDC)-Verfahrens Abhilfe, das ebenso wie DB2 auf dem Host läuft. Mit CDC werden ab einem definierten Zeitpunkt nur noch geänderte Daten geladen. Typischerweise geschieht dies in periodischen Abständen, etwa täglich. Es ist aber theoretisch

auch möglich, Änderungen sogar in Realtime zur Verfügung zu stellen.

Dieses Verfahren hat zahlreiche Vorteile: Informationen sind aktueller und die Änderungen können problemlos auch zeitgleich mehreren Anwendungen zur Verfügung gestellt werden. Ferner sinkt die Beanspruchung von Rechenleistung und Datentransfer drastisch. Mit CDC können Prozesse auch ganz anders angesteuert werden: Wenn bekannt ist, welches Feld sich wirklich verändert hat, können die entsprechenden Prozesse darauf geeicht, ein Trigger-Verfahren eingeführt und dadurch ein Großteil der Berechnungen, die ansonsten notwendig wären, eingespart werden.

Informationen beschleunigen das Business

Vorteile durch innovative Informationstechnologie müssen heute nicht immer in abstrakten ROI- oder TCO-Berechnungen gemessen werden, sondern können sehr konkrete Züge annehmen: Nur aufgrund der ausgefeilten IT-Anwendungen sowie des intern vorhandenen Know-hows war Hapag-Lloyd in der Lage, die anspruchsvollen Sicherheitsanforderungen des US-Zolls schnell und umfassend umzusetzen. Die US-Regierung legte im vergangenen Jahr fest, dass den Zollbehörden alle relevanten Transportdaten – Warenbeschreibungen, Herkunfts- und Bestimmungsort sowie Auftraggeber und Empfänger – 24 Stunden vor der Verladung im überseeischen Hafen elektronisch zu

übermitteln sind. Hapag-Lloyd gehörte zu den ersten Unternehmen, die diese Bedingungen erfüllten.

Mit Hilfe der Informatik-Lösungen kann die Hapag-Lloyd Container Linie in Zukunft noch schneller, flexibler und effizienter agieren und mit aktuellen Informationen weitere Schritte auf dem Weg zur „e-Company“ unternehmen. Mit der Migration des Data Warehouse vom Mainframe auf die AIX-Umgebung kann das Unternehmen in Zukunft unabhängiger und flexibler agieren. Der zweite Mainframe mit dem operativen System FIS wird im Zusammenspiel mit PowerCenter aber weiterhin die Infrastruktur beherrschen.