

Bardo Gutberlet ist Leiter SAP Systementwicklung/-betreuung bei RI-Solution
E-Mail: Bardo.Gutberlet@ri-solution.com

Dr. Michael Hahne ist Product Manager Applications in der deutschen Niederlassung von SAND Technology. E-Mail: michael.hahne@sand.com

Die Guten ins Töpfchen...

SLAs effizient einhalten

Eine Nearline-Lösung ermöglicht im Rahmen von Information Lifecycle Management (ILM) die effiziente Speicherung von wenig genutzten Daten. Das zeigt ein Praxisbeispiel bei dem Münchener Dienstleistungsunternehmen RI-Solution GmbH. Das Problem ist virulent: Das exponentielle Wachstum von Unternehmensdaten – laut einer Studie der Analystengruppe IDC verdoppelt sich das Datenvolumen alle 18 Monate – und immer strengere gesetzliche Vorgaben für eine langfristige Datenspeicherung erschweren es Unternehmen mehr und mehr, die steigenden Anforderungen hinsichtlich Systemverfügbarkeit und Reaktionsgeschwindigkeit, vereinbart in Service Level Agreements (SLA), zu erfüllen. Hinzu kommt, dass Anwender immer höhere analytische Anforderungen stellen und einfachen Zugriff auf umfassende, aktuelle, detaillierte und hoch verfügbare Informationen über ihr Unternehmen fordern, um sinnvolle Entscheidungen treffen zu können. Abhilfe verspricht die Umsetzung einer ILM-Strategie (Information Lifecycle Management), die sich an Lebenszyklus und Nutzung der Informationen orientiert.

Niemand braucht alle Daten des Unternehmens zu jeder Zeit. Mit zunehmender Größe eines Data Warehouse sinkt der prozentuale Anteil der regelmäßig aktiv genutzten Daten; typischerweise hat jedoch ein zunehmendes Datenvolumen im Data Warehouse einen negativen Einfluss auf dessen Leistungsfähigkeit im Sinne der Geschwindigkeit von Analysen und Ladeprozessen. Demzufolge wird die Performance für die aktiv genutzten Daten durch die vorhandenen eher inaktiv genutzten Daten negativ beeinflusst. Deshalb empfiehlt sich eine Strategie, die Daten in drei Speicher aufzuteilen: einen Online- und einen Nearline-Speicher sowie ein (Offline-)Archiv (siehe Abbildung 1). Sinkt die Zugriffsfrequenz der Benutzer auf bestimmte Daten unter eine bestimmte Schwelle, werden diese physikalisch zunächst in den Nearline-Speicher und dann nach einigen Jahren in das Archiv verschoben. Damit wird der relationale Onlinebereich „schlank“ gehalten, aber gleichzeitig der Zugriff auf

langjährig benötigte Informationen für Analysen oder den Aufbau neuer Kennzahlen gewährleistet.

Benötigt wird hierfür ein effizienter Nearline-Speicher, der die großen Datenvolumina auf kleinstem Raum vorhalten kann und zwar in einem Format, das schnellen und einfachen Zugriff erlaubt. So ergaben erste Messungen bei Anwendern von Nearline-Speicherlösungen zum Beispiel eine Reduktion der Daten auf durchschnittlich weniger als fünf Prozent der Online-Daten. Damit lassen sich erhebliche Speicherkosten einsparen beziehungsweise vermeiden, wenn ein größeres Data Warehouse geplant wird.

	Online-Speicher	Nearline-Speicher	(Offline-) Archiv
Häufige Nutzung	✓	↔	
Seltene Nutzung	✓	✓	↔
Sehr seltene Nutzung	✓	✓	✓

Abb. 1: Klassifikation der Data Warehouse Daten nach ihrer Nutzungshäufigkeit

Wichtig ist jedoch auch die Ersparnis bei der personalintensiven Implementierung größerer Data Warehouses sowie bei deren Administration. Dieser Aufwand reduziert sich deutlich, da die Lösung eine Restrukturierung des Data Warehouse in der Regel minimiert und gleichzeitig aufgrund ihrer Wartungsfreundlichkeit den administrativen Aufwand in Grenzen hält. Ein weiterer positiver Effekt sind die zeitlichen Freiräume, die durch die starke Verkleinerung der Nearline-Daten entstehen, da sich die operativen Data-Warehouse-Prozesse wie Backup, Recovery oder tägliches Laden der neuen Daten deutlich verkürzen. Damit lassen sich auch Service Level Agreements effizient einhalten.

In der Praxis bewährt

Dies erkannte auch das Münchner IT-Dienstleistungsunternehmen RI-Solution GmbH, IT-Dienstleister unter anderem der BayWa AG. Das Unternehmen sah sich stark wachsenden Datenmengen im SAP NetWeaver BI gegenüber, die es in den Griff bekommen wollte, um die Service-Level-Vereinbarungen für seine Kunden mit weiterhin optimierten Ressourcen gewährleisten zu können. Mit zunehmendem Datenvolumen im Data Warehouse steigen auch die Anforderungen an die eingesetzte Backup- und Recovery-Lösung. Wie Abbildung 2 verdeutlicht, hat sich der Kapazitätsbedarf für Datensicherungen in diesem Jahr um 50% erhöht, aufgrund immer weiterer

Laufzeit 05	Laufzeit 06	Laufzeit 07	Tapes 05	Tapes 06	Tapes 07
100%	+50%	+100%	100%	+50%	+100%

Abb. 2: Prognostizierter Kapazitätsbedarf für Storage und Backup

Anwendungsbereiche und zunehmender Anforderungen an die Granularität der Data-Warehouse-Daten, wird für das kommende Jahr eine Verdopplung der benötigten Platten- und Bandkapazitäten prognostiziert.

Gesucht wurde daher eine adäquate Systemlösung, an die hohe Anforderungen gestellt wurden. Die Fachabteilung wollte sowohl Archivierung als auch Wiederherstellung der zentralen Data Warehouse Layer inklusive des Staging-Bereiches basierend auf InfoCubes, ODS-Objekten und Tabellen der Persistent Staging Area (PSA) sowie einen flexibleren Umgang mit archivierten Daten gewährleisten. Zudem war gewünscht, dass PSA-Objekte weiterhin als Datenquelle verwendbar sind und die Anwender transparent und ohne große Performance-Einbußen Zugriff auf archivierte Daten haben. Darüber hinaus sollte die Lösung selbstredend zu einer spürbaren Kostenersparnis führen, dabei möglichst geringe Investitionen in neue Hardware erfordern und minimalen Aufwand für die Administration aufwerfen.

Zusammen mit der BI-Fachabteilung nahm die Abteilung für SAP Systementwicklung/-betreuung verschiedene Archivierungslösungen unter die Lupe. Nach einer gründlichen Marktanalyse entschied sich RI-Solution schließlich zu einer Testinstallation mit SAND/DNA Access, die im Januar 2006 erfolgreich abgeschlossen wurde. Im April folgte die Installation der aktuellen Version, die etwa sechs Stunden für zwei Systeme in Anspruch nahm. Tests zeigten: Infocubes konnten wesentlich besser gewartet werden, durch geringeres Datenvolumen ergaben sich generell kürze Laufzeiten bei Standard-Jobs (Aufbau Indizes/Statistikdaten, Rollup/Aggregataufbauzeiten, Change-

Run). Weiter entstanden Vorteile durch die Archivierung von Staging-Daten: Neue Cubes konnten effizienter und schneller aufgebaut werden, der Zugriff auf aktuelle PSA-Datentabellen erfolgte wesentlich schneller und Fehler konnten effizienter und schneller bearbeitet werden.

Mit der eingesetzten Nearline Lösung ist man nun in der Lage, ausgewählte Daten sehr stark zu komprimieren und dennoch weiterhin zur Analyse oder als Grundlage für neue ODS-Objekte oder InfoCubes zur Verfügung zu stellen. Im Systembetrieb zeigte sich, dass der Speicherbedarf für die archivierten Daten um rund 85 Prozent niedriger lag und auch die Backup-/Restore-Zeiten deutlich kürzer waren. Dabei ging es insbesondere um die großen Volumina-Daten der Staging Area, die in SAP NetWeaver BI verwendet werden. Diese wären bei RI-Solution zum Teil nur sehr schwer beziehungsweise überhaupt nicht mehr aus den Quellsystemen wiederherzustellen, weil sie dort nicht mehr vorhanden sind oder die Ladezeiten einfach zu lang wären. Das Unternehmen benötigt sie jedoch noch für den Aufbau von neuen Analyseobjekten, weil immer wieder neue Anforderungen hinzukommen.

In der Praxis ist jene Lösung am besten geeignet, die einerseits das Datenvolumen im aktiven SAP NetWeaver BI deutlich reduziert und somit die Datenmanagement-Prozesse schlank bleiben lässt. Auf der anderen Seite werden diese Daten jedoch nicht gelöscht oder einfach nur archiviert, sondern die Systembetreuung kann weiterhin darauf zugreifen und somit dem Wunsch der Fachabteilung nach flexibel aufbereitetem Informationsgehalt effizient nachkommen.