



Quelle: sapnews1 - FOTOLIA

SAND: Das SAP NetWeaver Nearline Storage Interface

Mehr „Ordnung“ im Data Warehouse

Eines der weniger bekannten Elemente der Data-Warehousing-Architektur SAP NetWeaver ist das Nearline Storage Interface. Mit dem wachsenden Einsatz von Transaktionssystemen wie mySAP ERP wurde auch der zunehmende Bedarf an einem umfassenderen Konzept für das Management von Unternehmensdaten immer deutlicher.

Von Roland Markowski, Managing Director Central Europe bei SAND Technology und Rainer Uhle, Produktmanagement NetWeaver, SAP

Neuere gesetzliche Vorgaben mit der Forderung, riesige Mengen an historischen Daten für die Analyse vorzuhalten, verschärfen das Archivierungsproblem zusätzlich. Mit NetWeaver 2004s hat SAP daher ein umfassendes Konzept für die Bewältigung wichtiger Anforderungen im Data Warehouse sowohl hinsichtlich Performance, als auch Skalierbarkeit und verbessertem Umgang mit den Datenvolumina herausgebracht. Die neu von SAP entwickelte Architektur, die z. B. in NetWeaver BI 2004s Release Notes SPS 7 ausführlich beschrieben wurde, bietet ein effektives Management sehr großer Datenvolumina auf der Basis des Konzepts des Information Lifecycle Management (ILM). Dabei werden Warehouse-Daten in drei Kategorien eingeteilt: aktuelle Daten, die online im Data Warehouse vorgehalten werden, weniger häufig genutzte Nearline-Daten sowie ältere Daten, die aus ökonomischen Gründen in eine Archivlösung verlagert werden.

Rainer Uhle, Produktmanager im Bereich NetWeaver, erläutert dies wie folgt: „Wichtiges‘ und ‚Unwichtiges‘ voneinander zu trennen, ist immer sinnvoll. Dies gilt auch für operationale Daten im Data Warehouse. Wichtige Daten sind solche, auf die sehr häufig zugegriffen wird. Für diese erwartet ein Großteil der Anwender Zugriffsmöglichkeiten mit hoher Performance, die wir heute über unseren BI Accelerator garantieren können. Solche ‚wichtigen‘ Daten gehören in die BI-Datenbank und als Replikate in den BI Accelerator. Unwichtigere Daten werden seltener angefordert und nicht mehr überschrieben (read only), aber dennoch als direkt zugreifbar erwartet, wenn auch ohne übermäßige Performanceerwartungen auf der Anwenderseite. Darüber hinaus würde ich noch die Kategorie ‚irrelevante‘ Daten anfügen. Dies sind zumeist alte Daten, die für Analysen und Ladeprozesse nicht mehr von Interesse sind und nur noch aufgrund rechtlicher Anforderungen

in klassischen Archivierungsumgebungen ohne Direktzugriffsmöglichkeiten gehalten werden sollten.“

Im NetWeaver 2004s BI können unwichtige und irrelevante Daten aus der Datenbank – bei gleichzeitiger Löschung dort – in Nearline-Storage-Umgebungen (NLS) ausgelagert werden. Aus dem NLS können diese Daten einerseits für Analyse- und Ladeprozesse im Direktzugriffsverfahren mit hoher Performance zur Verfügung gestellt, sie können andererseits aber auch in klassische Archivumgebungen ohne Direktzugriff verlagert werden. Somit kann insgesamt gesehen im SAP NetWeaver BI ein dreistufiges Konzept für Information Lifecycle Management angeboten werden, das wesentliche Kostensenkungen bei der Datenhaltung ermöglicht, ohne evidente Bedürfnisse im BI-Umfeld zu vernachlässigen. „SAND Technology nutzt für seine Lösungen die gleichen Technologiepotenziale (Grid-Computing, In-Memory-Processing, Column based Indices), wie sie im NetWeaver-Umfeld auch von SAP genutzt werden, und reiht sich somit nahtlos in unsere Strategie zur Kostensenkung im Storage-Bereich ein, ohne Abstriche bei der Integration und Interoperabilität machen zu müssen“, so Uhle weiter.

Unternehmen aller Branchen erleben derzeit eine wahre Explosion ihrer Daten. Gesetzliche Compliance-Anforderungen, SEC-Berichterstattung, FDA-Nachverfolgung, Basel II oder erweiterte Produktgewährleistungen sind nur einige der Triebkräfte hinter der rasanten Zunahme der Datenmengen, die Unternehmen vorhalten und verwalten müssen. Um den Anforderungen gerecht zu werden, müssen historische Daten für deutlich längere Zeiträume zur Verfügung gehalten werden. Dabei sollen diese weiterhin für die Analyse zur Verfügung stehen. Darüber könnte es nötig werden, sie auch nachträglich an spätere Ereignisse, wie Akquisitionen oder Übernahmen von Tochtergesellschaften, Veränderungen der Kontenpläne, Geschäfts- oder Fertigungsstrukturen usw. anpassen zu können. Audits könnten es erfordern, vorhandene Datensammlungen (wie InfoCubes) angesichts solcher Veränderungen neu zu berechnen und zu verifizieren sowie festzustellen, ob und wann ein solches Recasting durchgeführt werden sollte. Alle diese Aktivitäten erfordern das Vorhalten von Detaildaten, die in der Vergangenheit häufig einfach gelöscht wurden. Darüber hinaus ist es zunehmend üblich, dass Kunden große Volumina an Nicht-SAP-Daten in ihren NetWeaver BI-Systemen verwalten – sei es, um neue Konten-

pläne aus übernommenen Geschäftsbereichen mit anderen Systemen aufzubauen oder um Kundenverhalten anhand von Weblogs und Aufzeichnungen aus anderen Interaktionen, die außerhalb der SAP-Welt stattfinden, zu verstehen. Wenn neue Technologien wie z. B. RFID hinzukommen, ergeben sich Datenanforderungen von riesigen Ausmaßen. Dennoch haben viele SAP BI-Anwender absichtlich die Größe ihrer SAP-Data-Warehouses auf wenige Jahre der Reporting-Historie beschränkt, um diese verwaltbar zu halten. Mit den richtigen Funktionen zu einem annehmbaren Preis würden es viele SAP-Kunden jedoch sicher bevorzugen, ihr Data Warehouse auf das ganze Unternehmen auszuweiten.

Die Bereitstellung einer Nearline-Komponente in der NetWeaver-Architektur ist eine Antwort auf diese Belange des Informationsmanagements. Sie wurde entwickelt, um Read-only-Daten in einem Format vorzuhalten, in dem sie für Analyse und Berichterstellung zur Verfügung stehen, während die Herausforderung der Verwaltung großer Datenvolumina nicht außer Acht gelassen wird. Nun ist es nicht abwegig, bei der Planung für NetWeaver BI die Frage zu stellen, warum die herkömmliche Kombination aus Online-Data-Warehouse und Archiv nicht ausreicht, um alle Datenanforderungen im Unternehmen zu erfüllen. Im Folgenden soll daher das Nearline-Konzept und sein Verhältnis zum Haupt-Data-Warehouse und dem Archiv etwas näher erläutert werden.

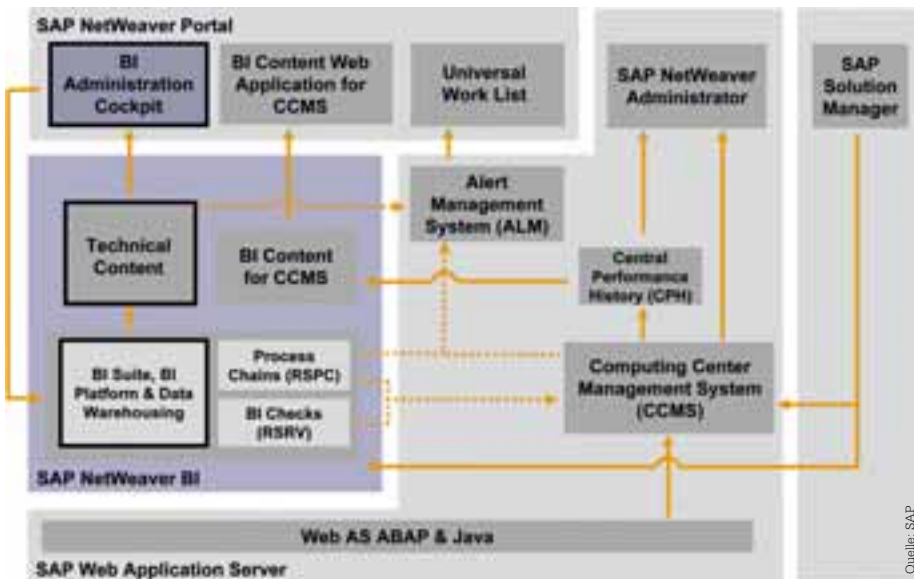
Warum nicht alle Daten ins Data Warehouse?

Warum nicht alle Daten ins Data Warehouse?

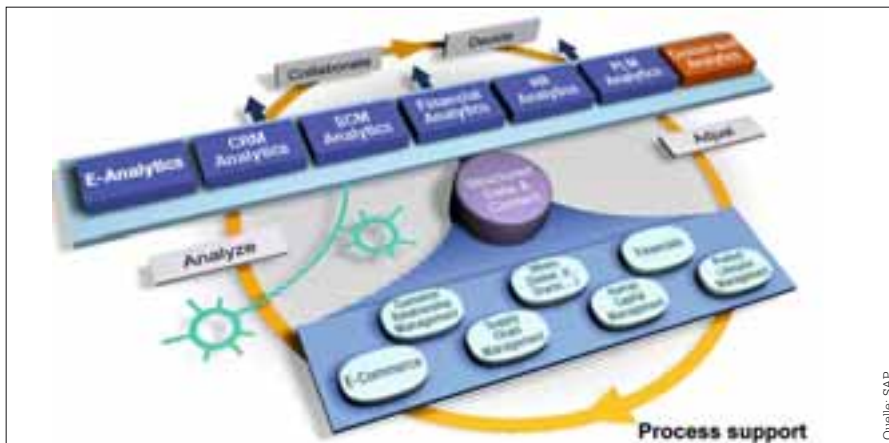
In einer typischen SAP NetWeaver BI-Umgebung gibt es eine Reihe von Faktoren, die eine Vergrößerung des Data Warehouses im Hinblick auf rasch wachsende Datenvolumina schwierig oder teuer machen könnten. So kann die SAP InfoCube-Architektur, auch wenn sie einfache multidimensionale Be-

richterstattung ermöglicht, zur beschleunigten Datenzunahme beitragen, da die vorkalkulierten Würfel u. U. viel Platz beanspruchen. Auch die Fähigkeit des BI Accelerator, den Aufbau und die Nutzung von Würfeln zu vereinfachen und zu skalieren, kann das Datenwachstum weiter anfeuern.

Werden alle Daten (strategische Daten, historische Daten, operationale Daten und Transaktionsdetails) im Data Warehouse vorgehalten, bedeutet dies, dass eine große Zahl weniger häufig benötigter Daten den gleichen Pflegeaufwand wie regelmäßig genutzte, aktuelle Daten für sich beansprucht – d. h. die gleiche Hochleistungs-Speicherumgebung und die gleiche Bearbeitungsfrequenz, um Verfügbarkeit und Integrität zu gewährleisten. Wenn das Data Warehouse wächst, kann dies die Kosten für die Bereithaltung eines angemessenen Service oder die Datenverfügbarkeit dramatisch nach oben treiben. Darüber hinaus umfassten Best Practices für



Mit NetWeaver 2004s hat SAP ein umfassendes Konzept für die Bewältigung wichtiger Anforderungen im Data Warehouse sowohl hinsichtlich Performance als auch Skalierbarkeit und verbessertem Umgang mit den Datenvolumina herausgebracht.

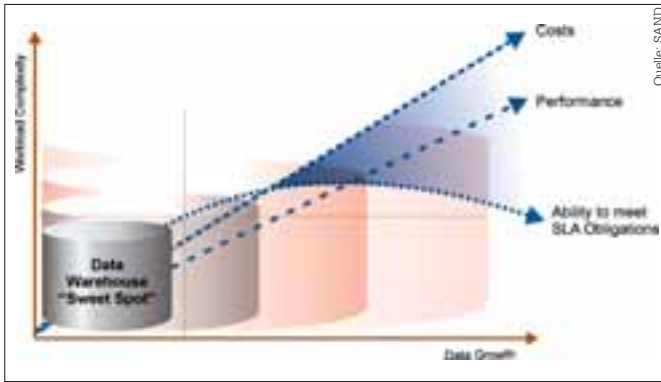


In einer typischen SAP NetWeaver BI-Umgebung gibt es eine Reihe von Faktoren, die eine Vergrößerung des Data Warehouses im Hinblick auf rasch wachsende Datenvolumina schwierig oder teuer machen könnten.

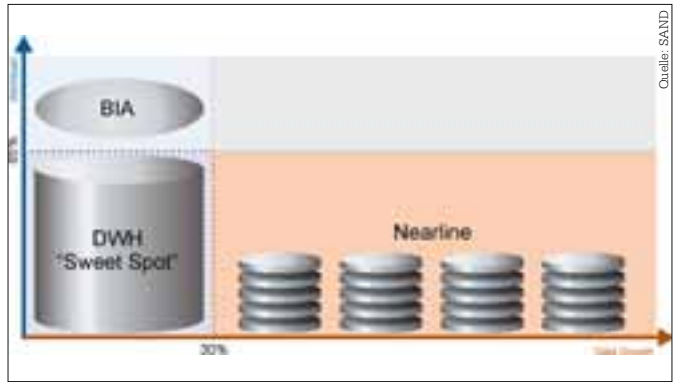
Nearline-Konzeptfragen

Um unmittelbaren Zugriff auf so viele historische Daten wie möglich zu bieten, hat die SAP-Entwicklung einen proaktiven Ansatz gewählt, um das aufkommende Konzept des Information Lifecycle Managements auch im BI-Umfeld konsequent umzusetzen. Das Ergebnis ist die Nearline-Komponente innerhalb der NetWeaver-Architektur. Zum Verständnis der Gründe für die Einführung des Nearline-Konzepts sind folgende Fragen hilfreich, vor denen ein Systemmanager regelmäßig steht:

1. Wie kann ich mehrere Terabyte an Informationen im SAP NetWeaver noch effektiver vorhalten und verwalten?
2. Was ist, wenn ich einen InfoCube wieder aufbauen oder verändern muss, um aufgrund von Änderungen im Unternehmen oder im Compliance-Reporting zusätzliche Daten zu integrieren?
3. Wie lange würde es dauern, meine Data-Warehouse-Umgebung aus detaillierten Informationen wieder herzustellen?
4. Was ist, wenn eine Analyse den Zugriff auf ältere historische Details erfordert, die nur unregelmäßig benötigt werden?
5. Wie hoch sind die Kosten, um mehrere Instanzen aller Data-Warehouse-Daten für neue Funktionen, Notfallwiederherstellung und Ausfallsicherung bereit zu halten?



Nearline-Komponenten wurden entwickelt, um ein Data Warehouse kostengünstig zu skalieren, um viele Terabytes und Petabytes an zugänglichen Daten vorzuhalten.



Werden alle Daten im DWH vorgehalten, bedeutet dies, dass weniger häufig benötigte Daten den gleichen Pflegeaufwand wie regelmäßig genutzte beanspruchen.

ein Data Warehouse häufig die Pflege mehrerer Kopien der Daten: Primäre Daten werden z. B. zur Erhöhung der Verfügbarkeit gespiegelt. Staging-Systeme für die nächsten „Go-live“-Änderungen erfordern voll befüllte Datensätze und auch die Notfallwiederherstellung fordert eine weitere Replikation der Daten offsite. Die Pflege aller Daten, einschließlich selten genutzter statischer Daten in einer hochleistungsfähigen Online-Umgebung, kann aufgrund der Beschränkungen herkömmlicher Datenbanken entweder unerschwinglich teuer oder gar physikalisch unmöglich sein. Die Implementierung einer Nearline-Komponente ermöglicht es, weniger häufig genutzte Daten, wie gealterte Informationen oder detaillierte Transaktionen zum Aufbau oder Recasting von InfoCubes, kostengünstiger vorzuhalten – in einigen Fällen mit deutlich weniger Platzbedarf und reduzierten Anforderungen hinsichtlich der Replikation. Darüber hinaus ermöglicht das Entfernen von Read-only-Daten aus dem Data Warehouse, dass Aufräumarbeiten deutlich flotter durchgeführt werden können und Daten schneller für Endanwender zur Verfügung stehen.

Nearline versus Archivierung

SAP war lange führend bei der Bereitstellung effektiver Archivierungswerkzeuge, um Daten sowohl entsprechend gesetzlicher Anforderungen, als auch für die Datenbedürfnisse des Unternehmens aufbewahren zu können. Warum ist ein solches Archiv dann nicht ausreichend, um das Data Warehouse von der Last der Read-only-Daten zu befreien?

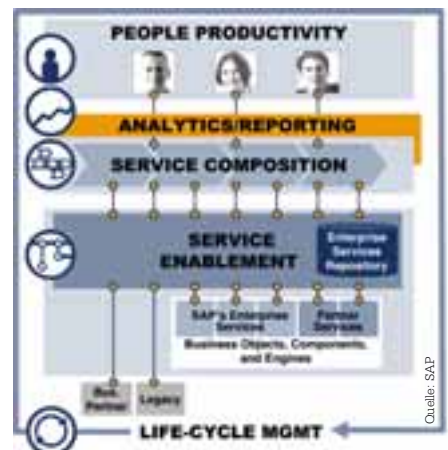
Die Antwort hat mit dem Zugriff zu tun. Wie der Name schon sagt, bieten Nearline-Lösungen raschen Direktzugriff auf die Read-only-Daten, die sie vorhalten, unabhängig von ihrem Volumen. Viele dieser Daten müssen unter Umständen nur unregelmäßig abgefragt werden. Im Fall des Falles müssen sie

dann jedoch ohne intensive Vorbereitung zu griffbereit sein. Ein hoher Aufwand, um archivierte Daten für die Analyse oder die Berichterstellung wiederherzustellen, kann extrem teuer oder zeitaufwändig sein und verpasste Geschäftschancen, hohe Geldbußen seitens der Behörden und verminderte Produktivität der Organisation nach sich ziehen. Die Tatsache, dass sich häufig nicht planen lässt, welche Daten Aktivitäten erfordern, macht die Einführung eines Archiv-Wiederherstellungsprozesses quasi unmöglich und das Maß der Ressourcen, die dies benötigen würde, lässt sich in der Regel schwer bestimmen. Das Vorhalten großer Mengen historischer und detaillierter Daten in einer Nearline-Komponente erleichtert dagegen den Zugriff auf diese Daten deutlich.

Nearline-Komponenten wurden entwickelt, um ein Data Warehouse kostengünstig zu skalieren, um viele Terabytes und sogar Petabytes an zugänglichen Daten vorzuhalten. Dennoch kann der Nearline-Speicher nicht als Beleg für Originaldaten dienen. In einigen Fällen liegen die Daten im Nearline-Speicher nicht in ihrer Originalform vor, sondern vielmehr als Darstellung oder Transformation der Originaldaten. In solchen Fällen wäre eine sichere Archivierung weiter erforderlich. Ebenso könnte es weiter erforderlich sein, die Wiederherstellung von Daten zu garantieren, die so alt sind, dass sie niemals wieder für Berichterstellung oder Analyse benötigt werden. In diesen Fällen kann eine Archivierung kostengünstiger sein als das Vorhalten der Informationen in zugänglicher Form. In NetWeaver BI 2004s ist die Nearline-Schnittstelle vollständig integriert. Für BW 3.x wurden zudem Nearline-Konzepte pilotiert, so dass Unternehmen die Vorteile der Nearline-Speicherung unmittelbar nutzen und nahtlos zur Implementierung von NetWeaver 2004s übergehen können. Dies bedeutet, dass bei gleichzeitigen Implementierungen oder bei neueren Systemen hohe Kos-

tenvorteile erzielt werden können, indem sichergestellt wird, dass die BI-Umgebung unbegrenzt skalieren kann, unabhängig davon, ob es um geplante oder ungeplante Aktivitäten im Unternehmen geht. Die Implementierung einer Nearline-Komponente ist in der Regel unkompliziert. Auch wenn es einige Planung erfordert, welche Daten von der Online-Umgebung migriert werden sollen, lässt sich dies gewöhnlich mit der Hilfe von Zeitregeln und Prozessketten zur automatischen Erzeugung der Nearline-Daten und Auflistung ihrer Verfügbarkeit für NetWeaver BI-Anwender einfach umsetzen. Nachdem dies erfolgt ist, integrieren sich die Nearline-Daten nahtlos in die Online-Data-Warehouse-Daten. Angesichts des wachsenden Bedarfs an Analysefähigkeiten bei immer mehr anwachsenden Datenmengen aus verschiedenen Quellen präferiert SAP einen erweiterten Ansatz für die optimierte Ablage und Verwaltung von Unternehmensdaten. Nearline-Storage-Lösungen können als Datenbank-Extensionen fungieren und so neue Möglichkeiten zum ökonomischeren Einsatz der SAP-Funktionalitäten im Unternehmen bieten.

Quelle: www.sandtechnology.de und www.sap.de



SAP präferiert einen erweiterten Ansatz für die Ablage und Verwaltung von Unternehmensdaten.