



SAND/DNA: Nearline-Lösung optimiert Data-Warehouse-Ressourcen

Next Generation Enterprise Data Warehousing

Der effiziente Umgang mit Informationen ist für die meisten Unternehmen überlebenswichtig. Dennoch gibt es viele Organisationen, die sich aufgrund von nicht abgestimmten lokalen Prozessen oder BI-Anwendungen zahlreich verteilten und dabei inkompatiblen Systemen gegenüber sehen. Gemeinsame Standards fehlen und ein Überblick über das ganze Unternehmen ist quasi unmöglich.

Von Jürgen Haupt, Regional Implementation Group (RIG) und Dr. Michael Hahne, SAND Technology

Sicher können BI-Werkzeuge und lokale Data Marts, die aus bestimmten Reporting- und Analyse-Überlegungen heraus entstanden sind, dazu beitragen, taktische Ziele zu erreichen. Dennoch sind damit Probleme hinsichtlich der Konsistenz und Flexibilität verbunden, die sich in unkontrollierten Datenflüssen, wiederholten und redundanten Extraktionen der gleichen Daten, vielfältigen inkonsistenten Datenmodellen, hohen Entwicklungskosten, Einschränkungen bei der Erfüllung von Informationsbedürfnissen, Informationsinseln und einer insgesamt eher unzuverlässigen unternehmensweiten Datenbasis niederschlagen. Zu allem Übel stehen die IT-Abteilungen vor einem Flickenteppich aus Werkzeugen, der hohe Pflege- und Integrationsanstrengungen, sowie ständige Nacharbeit fordert. Dies treibt nicht nur die Gesamtkosten in die Höhe, sondern beeinflusst auch deutlich die Flexibilität im Unternehmen – mithin entscheidend für die Wettbewerbsfähigkeit.

Das Unvorhersehbare im Blick

Um bei abteilungsbezogenen BI-Implementierungen Datenkonsistenz und Reportingflexibilität zu gewährleisten, empfiehlt SAP daher die Einführung eines „Enterprise Data Warehouse Layers“, häufig auch „Corporate Memory“ genannt. Die Daten werden dort subjektbezogen und integriert, „unverfälscht“ – d. h. die Daten wurden nicht verändert, um den Zielen eines bestimmten Data Marts gerecht zu werden –, granular, historisch komplett und umfassend – mit mehr Attributen als aktuell von einem bestimmten Data Mart gefordert werden – gelagert. In den letzten Jahren wurde die Implementierung von SAP BI als Mehrebenen-Architektur mit einem zentralen Enterprise Data Warehouse Layer von mehr und mehr großen, aber auch mittelgroßen Kunden präferiert. Die Einrichtung unterscheidet sich zwar von Kunde zu Kunde, aber es lässt sich festhalten, dass stets zwei Haupt- bzw. Unterebenen bestehen:

- der „Data Acquisition Layer“ zur 1 : 1-Speicherung der Daten und
- der „Data Warehouse Integration“ oder „Propagation Layer“, der die Daten „gebrauchsfertig“ für die Data-Mart-Szenarien bereitstellt.

Während ein Unternehmen mit dem Data Acquisition Layer flexibel auf Anforderungen reagieren kann, die bereits geladene Daten betreffen, ist der Data Propagation Layer des EDW die Grundlage für Data Marts, die von neuen Geschäftsanforderungen initiiert werden. Weitere Ebenen, z. B. für Qualitätssicherung und Transformation, hängen von der jeweiligen Situation ab.

Die Vorteile einer SAP BI-Architektur:

1. Zuverlässigkeit und Reproduzierbarkeit

Der SAP Enterprise Data Warehouse Layer wird zentral administriert und nicht mehr für beliebige Projekte verändert. Auf diese Weise kann er als gemeinsame Quelle für alle Daten dienen, die zum Aufbau von Data Marts für den Zugriff durch Anwender im ganzen Unternehmen genutzt werden.

2. Konsistenz

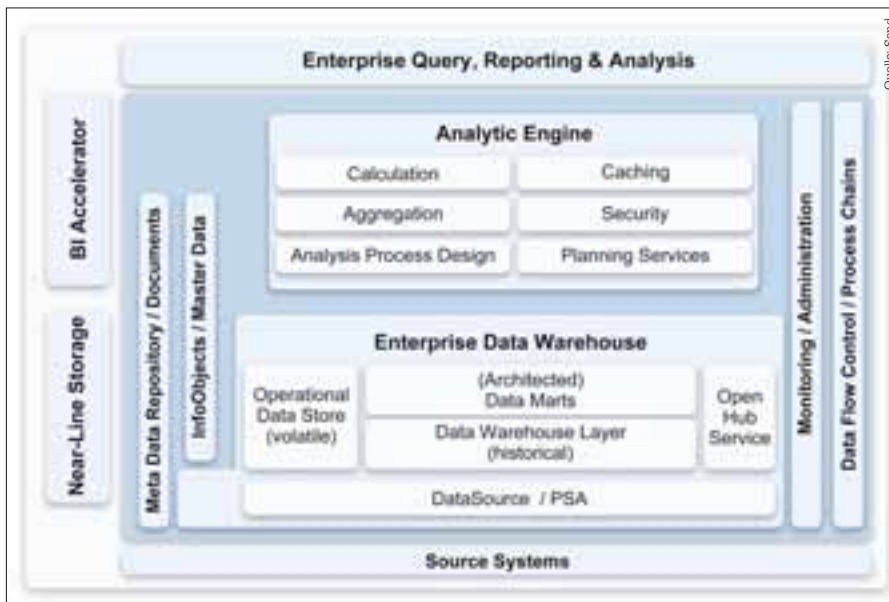
Damit werden auch Extraktion, Staging und Integration zentral verwaltet: Daten werden nur einmal extrahiert und redundante Prozesse vermieden. Dies trägt nicht nur zu Konsistenz der im Unternehmen genutzten Informationen bei, sondern reduziert auch Komplexität und Kosten für deren Verwaltung.

3. Harmonisierung von Geschäftsprozessen

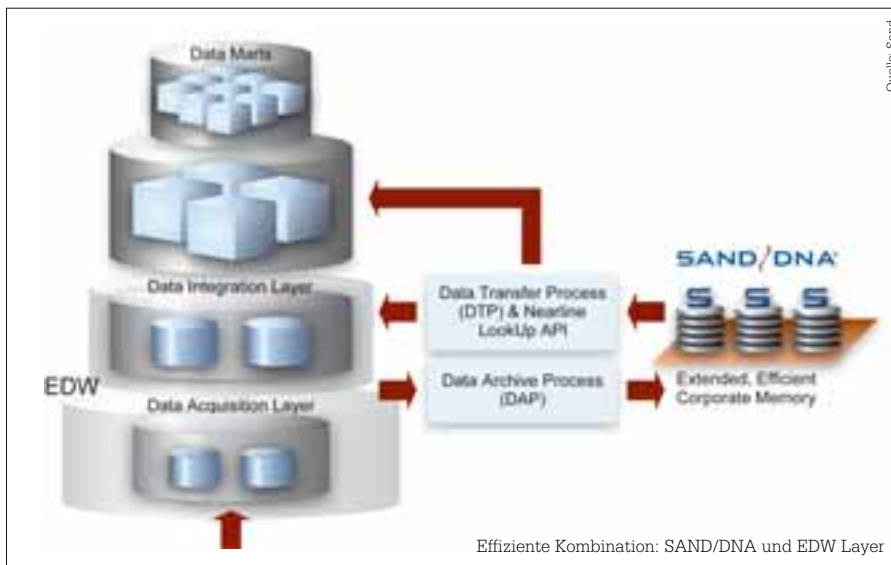
Eine gemeinsame Informationsquelle stellt sicher, dass Geschäftsprozesse durch konsistente Daten und ein gemeinsames KPI-System (Key Performance Indicators) unterstützt werden. Dashboard-Ansichten und Drill-downs werden besser unterstützt und eine schnellere Anpassung an neue Analysebedürfnisse, sowie Reaktion auf erweiterte Analyseanforderungen werden möglich.

Die neue EDW-Generation

Die Notwendigkeit, immer größere Datenmengen zu speichern und diese weiterhin für Analyse, Abgleich, Prüfung, Wiederherstellung von KPIs usw. zugänglich zu halten, bedeutet, dass herkömmliche Datenmanagementstrategien überdacht werden müssen. CIOs, die Service Level Agreements erfüllen müssen, benötigen eine Alternative zu den



Aufbau des Enterprise Data Warehouse



Effiziente Kombination: SAND/DNA und EDW Layer

üblichen Optionen, entweder hohe Investitionen in Hardware zu tätigen oder Daten aus dem Warehouse herauszunehmen und damit Kompromisse hinsichtlich der Qualität von Entscheidungen oder der Leistungsfähigkeit von BPM-Prozessen (Business Performance Management) einzugehen. Aus diesem Grund hat SAP – neben anderen nützlichen EDW-Funktionen, wie BIA (Business Intelligence Accelerator) zur Steigerung der Abfrage-Performance, neu gestaltete Datenübertragungsprozesse oder Echtzeit-Datenerfassung – in SAP NetWeaver BI 2004s das so genannte Nearline Interface eingeführt. Damit können Daten online, nearline und im Archiv gespeichert werden, da es wenig sinnvoll ist, große Datenvolumina in einer relationalen Datenbank vorzuhalten, wenn die Daten nicht regelmäßig gebraucht oder „nur für den Fall“ gespeichert werden. In Verbindung mit Nearline-Partnerlösungen kann das Nearline

Interface die relationale Hauptdatenbank in SAP NetWeaver BI von der „Bürde“ wachsender Volumina seltener genutzter Daten befreien. Dies sorgt für eine leistungsfähigere Online-Datenbasis und optimiert die Systemverfügbarkeit.

Mehr Hardware ist keine Lösung

Man mag anführen, dass die Kosten für Festplattenspeicher ständig sinken, während deren Leistung steigt, so dass man einfach mehr Speicher anschaffen könnte, um dieses Problem zu lösen. Aber Datenvolumina wachsen sogar noch schneller, als sich das Preis-Leistungsverhältnis der Speichertechnologie verbessert. Darüber hinaus wird geschätzt, dass jeder Euro, der in Festplattenspeicher investiert wird, fünf bis zehn Euro Betriebskosten während der Laufzeit nach sich zieht, so dass dies eine sehr teure Alternative ist. Noch

deutlicher spricht die Problematik der mangelnden Systemverfügbarkeit, die sich aus großen Datenmengen ergibt: Diese kann nicht durch mehr Speicherkapazität gelöst werden, da die operativen Flaschenhalse (z. B. langwierige Backup- und Recovery-Zeiten) weiterhin Probleme verursachen, wenn die Datenmengen weiter wachsen.

Effiziente Erweiterung des EDW Layers

Dank neuer technologischer Entwicklungen ist es jedoch möglich, diese Herausforderungen zu lösen bzw. diesen proaktiv vorzubeugen. So hat SAP gemeinsam mit dem Software- und Entwicklungspartner SAND Technology die Lösung SAND/DNA for SAP BI als Nearline-Speicherkomponente für das oben erwähnte Nearline Interface zertifiziert. Die Lösung ermöglicht die Implementierung eines effizienten Nearline-Speichers, der sehr große Volumina von seltener genutzten Daten, z. B. aus DataStore Objects des Data Acquisition Layers und des Data Propagation Layers, auf kleinstem Raum vorhalten kann und zwar in einem Format, das schnellen und einfachen Zugriff erlaubt. So wurden in produktiven Umgebungen mit SAND/DNA for SAP NetWeaver BI (SAP BI 3.1/3.5 und BI 2004s) bereits Kompressionsraten zwischen 85 und 95 Prozent erreicht, abhängig von InfoObject-Typ, Datenvolumen und -diversität. Administratoren können DataStore Objects aus dem Data Acquisition Layer und Data Integration Layer in SAND/DNA verlagern, dort komprimiert speichern und dies über die Nearline-Integration-Anwendung von SAP NetWeaver BI 2004s steuern. Die ausgelagerten Daten werden dann in SAP NetWeaver BI 2004s gelöscht. Aufgrund der hohen Kompression und der Möglichkeit, die DataStore Objects weiterhin über DTPs (Data Transfer Processes) für den Aufbau neuer DataStore Objects oder InfoCubes heranzuziehen, bietet sich SAND/DNA in Verbindung mit den Nearline-Interface-Funktionen als effiziente Erweiterung des Enterprise Data Warehouse Layers in SAP NetWeaver BI 2004s an.

Die Vorteile der Nearline-Speicherung:

1. Optimiertes Service Level Management

Der Einsatz von SAND/DNA in Verbindung mit einem SAP NetWeaver BI 2004s Enterprise Data Warehouse Layer verringert das Datenvolumen der DataStore Objects deutlich. Dies verkürzt die Zeiten für Backups und das Laden der Daten und erleichtert damit das Einhalten von Service Level Agreements. Darüber hinaus reicht es, eine kleine Menge

der komprimierten Daten zu lesen, um bestimmte Informationen aufzurufen. Ressourcenintensives Staging erübrigt sich, was die Data-Warehouse-Abläufe deutlich rationalisiert.

2. Niedrigere Gesamtkosten

Die Einführung effizienter Nearline-Speicherung mit SAND/DNA senkt die Aufwendungen für Speichermedien, Hauptspeicher und CPUs. Administrationskosten sinken ebenfalls dank schnellerem, einfacherem Software- und Versionsmanagement und kürzeren Backup- und Recovery-Zeiten.

3. Historie und Corporate Memory vollständig und zugänglich

Wenn bereits effizienter und sicherer Speicherplatz vorhanden ist, erweitert SAND/DNA den SAP NetWeaver BI 2004s Enterprise Data Warehouse Layer durch die Speicherung von Daten mit höherem Detaillierungsgrad. Dies umfasst unbearbeitete Details, mehr Granularitätsebenen, unregelmäßig genutzte Daten, Daten, die „nur für den Fall“ vorgehalten werden, ältere und historische Daten sowie Altdaten.

4. Mehr Flexibilität bei unvorhersehbaren Anforderungen

Ein effizienter Nearline-Speicher mit integrierten detaillierten Daten als Teil der physikalischen Implementierung von SAP BI EDW erlaubt die einfache Rückspeicherung der Daten in NetWeaver 2004s oder deren Nutzung als sehr leicht zugängliche Datenquelle. Dies bedeutet, dass neue InfoProvider (Data Marts) ohne Verzögerung erstellt werden können, auch wenn sie historische Daten oder neue Extraktionen umfassen.

Darüber hinaus ermöglicht SAND/DNA durch den direkten Zugriff auf die ausgelagerten Daten der DataStore Objects die schnelle Reaktion auf unvorhersehbare, einmalige Anwenderanfragen, ohne persistente Strukturen aufbauen oder aktivieren zu müssen. „Mit der Nearline-Lösung SAND/DNA können wir die Explosion der Daten im NetWeaver BI unter Kontrolle halten. Die Lösung optimiert die Nutzung der Data-Warehouse-Ressourcen und ermöglicht es uns, unsere Service Level Agreements unabhängig von der beteiligten Datenmenge zu erfüllen“, sagt Bardo Gutberlet, Leiter SAP-Systementwicklung und -betreuung bei RI Solution GmbH.

Quelle: www.sand.com

Zu **Virtual Infrastructure 3**, kurz VI3, gehören die Komponenten:

- VMware ESX Server 3
- VMware VirtualCenter 2 Management Server und den aktuellen Optionen
- VMware VMotion
- VMware VMware Distributed Resource Scheduler – DRS
- VMware High Availability – HA
- VMware Consolidated Backup – VCB