

In-Memory Technologie braucht eine strenge Produktauswahl



Foto: Fotolia

In-Memory-Lösungen bringen einen massiven Umbruch. Weil Hersteller unterschiedliche Wege gehen, finden IT-Verantwortliche nur schwer die richtige Technologie und den besten Zeitpunkt für den Einstieg.

Alle großen Hersteller von Business Software – am sichtbarsten SAP – betreiben aktuell einen großen Entwicklungs- und Marketingaufwand, um ihren In-Memory-Lösungen im Markt zum Durchbruch zu verhelfen. Glaubt man den Aussagen von Anbietern, Analysten und Consultern, steht die IT-Welt gerade vor einem Paradigmenwechsel bei Geschäftsanwendungen. Seinen Ausgang nahm dieser Umbruch im Segment der Business-Intelligence(BI)-Anwendungen, wo *QlikView*, *Tableau* und andere Produkte mit hoher Analysegeschwindigkeit und traumhafter Flexibilität zum Liebling der Fachbereiche wurden und den arrivierten Anbietern – und vielen IT-Abteilungen – das Leben schwer machten. Die großen Software-Anbieter, darunter auch BI-Spezialisten wie *MicroStrategy* und *SAS*, zogen in den vergangenen Jahren alle nach, so dass heute kein ernstzunehmender Hersteller mehr auf eine In-Memory-Komponente verzichtet. Selbst Open-Source-Suiten wie *Jaspersoft* bieten solche Optionen. Das Spektrum der Lösungen ist jedoch weit gefächert: von der eher einfacheren Lösung, die man früher als OLAP-Cache (Online-Analytical Processing) bezeichnet hätte, bis hin zu spezialisierten Appliances, die auch größere Datenbestände im Hauptspeicher halten. Die Idee der Nutzung des Arbeitsspeichers als performante Datenablage existiert seit vielen Jahrzehnten. Selbst *Microsoft Excel* beruhte lange ausschließlich auf dem Prinzip, die zu analysierenden Daten vollständig in den Arbeitsspeicher zu ziehen, um die Verarbeitungsgeschwindigkeit zu steigern. Auch BI-Werkzeuge wie *TM1*, das über *Cognos* zwischenzeitlich im Produktportfolio der IBM gelandet ist, waren bereits in den 80er Jahren echte In-Memory-Anwendungen.

Niedrige Hauptspeicher-Preise geben den Startschuss

Für den Einsatz im Unternehmensmaßstab konnten sich In-Memory-Ansätze jedoch lange Zeit nicht durchsetzen. Zu groß waren die Einschränkungen hinsichtlich des bearbeitbaren Datenvolumens. Als wesentlicher Einflussfaktor gelten oft die Kosten für Arbeitsspeicher innerhalb von Rechnerarchitekturen. Tatsächlich konnten bis vor wenigen Jahren nur wenige Unternehmen oder Behörden große Mengen an Arbeitsspeicher bezahlen. Zu den finanziellen Einschränkungen kamen technische Hindernisse: So ermöglichen es erst die heute marktüblichen 64-Bit-Architekturen sowie die Parallelisierung in Hardware und Software, ein ausreichend großes Speichervolumen jenseits von zwei Gigabyte anzusprechen. Die Speicherung großer Datenmengen in festplattenbasierten Dateien und Datenbanken wurde für mehrere Generationen zum Standard, und der In-Memory-Gedanke geriet in Vergessenheit. Das sogenannte I/O-Problem (Input/Output), also der Laufzeitverlust durch die häufigen Lese- und Schreibzugriffe auf die Festplatte, wurde zum vertrauten, unvermeidlichen Begleiter von Entwicklern und Administratoren, die sich zu dessen Minimierung eine Vielzahl an teils sehr aufwändigen Optimierungen und Workarounds einfallen ließen. Beispiele sind intelligente Caching-Verfahren, aber auch spezialisierte Datenstrukturen, wie etwa OLAP-Cubes, Aggregate und andere Formen der redundanten Datenhaltung. Obwohl auch transaktionale Anwendungen unter der beschriebenen I/O-Problematik litten, gelangten Analyseappli-

kation wie Data-Warehouse-Systeme oder OLAP-Lösungen bereits früher an ihre Grenzen. Optimierungen und Workarounds machten die Architekturen nämlich unflexibel, komplex, fehleranfällig und kostspielig. Viele wünschenswerte Analysen scheiterten an technischen Hürden, weswegen wichtige Unternehmensentscheidungen weiter aus dem Bauch getroffen wurden. Fachbücher über das Scheitern von BI-Lösungen fanden im vergangenen Jahrzehnt ein breites Publikum. Zudem traten viele Wunderheiler mit vermeintlich richtigen BI-Konzepten auf den Plan.

Schnelle Antworten, steigende Flexibilität und sinkende Kosten

Dass einige der fundamentalen Probleme klassischer BI-Architekturen in der festplattenbasierten Datenablage und den daraus resultierenden Performance-Engpässen wurzeln und dass dort auch die Lösung liegen könnte, drang nur selten ins Bewusstsein. Genau hier setzen In-Memory-Architekturen an, und deswegen geht ihr Nutzen auch weit über blanke Performance-Verbesserungen hinaus. In-Memory-Konzepte lösen die Starre vieler BI-Lösungen auf, reduzieren deren Komplexität und machen sie flexibler, wartbarer und letztlich – trotz der Investition in neue Hard- und Software – oft kostengünstiger. Zufriedene Fachbereiche und eine stringente IT-Strategie erscheinen plötzlich nicht mehr als gegensätzliche Ziele.

In-Memory-Architekturen bieten – je nach Ansatz in unterschiedlichem Maße – folgende Nutzenpotenziale:

- Dank der Verkürzung der Antwortzeiten können Daten häufiger ausgewertet werden. Der Nutzen der Anwendung fällt somit häufiger an.
- Daten lassen sich dank der erhöhten Flexibilität der In-Memory-Lösung von unterschiedlichen Seiten beleuchten, zum Teil auch simulativ und explorativ unter Einsatz von hypothesenfreien Algorithmen wie im Data Mining. So entstehen neue Erkenntnisse.
- Auch sehr aktuelle Daten können analysiert werden. Im Extremfall erfolgt die Analyse in Echtzeit und lässt sich zur Steuerung von Prozessen nutzen.
- Es steht ein breiteres Spektrum an Daten (weitere Datenquellen, länger zurückreichende Historien) für die



Hohe Geschwindigkeit und Flexibilität sowie niedrigere Kosten stellen die primären Nutzenpotenziale von In-Memory-Architekturen dar. Quelle: Mayato

Analyse zur Verfügung.

- Die Daten können feingranularer analysiert werden, bis hin zum Zugriff auf Einzelbelege und Vorgänge im Mikrobereich.
- Durch die Reduktion der Komplexität und die Vermeidung von aufwändigen Optimierungen und Workarounds entstehen teilweise massive Kosteneinsparungen.

Da für die Anschaffung einer In-Memory-Lösung für Hardware und Software sowie gegebenenfalls auch die Umstellung bestehender Anwendungen bei größeren Lösungen schnell Kosten in sechs- bis siebenstelliger Höhe anfallen, ist eine genaue Berechnung des Business Case unverzichtbar. Die Bewertung der Erwartungen kann im Einzelfall durchaus schwierig sein. Für die spätere Einschätzung des Erfolgs der Lösung und nachgelagerter Entscheidungen über einen Rollout

des gewählten Vorgehens sollte diese Betrachtung überaus gründlich durchgeführt und dokumentiert werden.

Analyse-Anwendungen profitieren vom hohen Tempo am meisten

Nach wie profitieren vor allem die analytischen Applikationen von In-Memory-Ansätzen. Es ist jedoch zu beobachten, dass es vielfach heute nicht mehr dabei bleibt, die Antwortzeiten interaktiver Analysen zu beschleunigen. Damit hatten Werkzeuge wie *QlikView* vor gut zehn Jahren ihren Siegeszug angetreten. Heute geht es oft um ausgefeiltere Szenarien, in denen komplexe Fragestellungen und Algorithmen, etwa Predictive Analytics, Data Mining sowie Simulation und Planung, in den Mittelpunkt rücken.

Viele der Daten, die heute von In-Memory-Lösungen verarbeitet wer-



**Je nach Anwendung ergeben sich spezifische Anforderungen an In-Memory-Werkzeuge.
Nur ein intensiver Produktvergleich hilft bei der Auswahl. Quelle: Mayo**

den, stammen aus Datenquellen, die für Business-Intelligence-Anwendungen relativ neu sind. Beispiele hierfür sind Soziale Netzwerke, Mobilgeräte, RFID-Chips (Remote Frequency Identification) oder Sensoren in Fahrzeugen und Maschinen. Technisch stellen diese Daten alleine aufgrund ihres Volumens eine Herausforderung dar. Darüber hinaus sind sie vielfach nicht vollständig strukturiert und auch nicht in relationalen Datenbanken abgelegt, sondern sie kommen als kaum oder nicht strukturierte Textdateien oder auch in Form von Bild, Ton oder Video daher. Teilweise müssen diese Daten von der Festplatte, teilweise aber auch aus spezialisierten Datenablagen wie Hadoop gelesen werden. Eine In-Memory-Lösung muss daher offen

sein für das Einlesen von Daten verschiedenster Quellen und für die Verarbeitung von nicht oder semistrukturierten Daten.

Durch die Möglichkeit, mittels intelligenter Verfahren auch große Datenmengen in sehr kurzer Zeit automatisch nach nützlichen Informationen zu durchleuchten, entstehen mehr und mehr Ideen für Echtzeitanwendungen. In-Memory-Lösungen rücken damit immer näher an die transaktionalen Prozesse, die durch intelligente Entscheidungsunterstützung direkt beeinflusst und gesteuert werden. Realtime Analytics bietet gigantische Nutzenpotenziale durch unzählige neue Optimierungsmöglichkeiten. Sie stellt In-Memory-Lösungen aber auch vor Herausforderungen: die Tools müs-

sen kontinuierlich neue Daten entgegennehmen können, wo vorher die einmalige oder periodische Beladung über Nacht ausreichte.

Auch an die Verfügbarkeit von Systemen und an die Ausfallsicherheit stellen solche Szenarien hohe Anforderungen. Sind diese erfüllt, ist die IT nicht mehr weit entfernt vom letzten Schritt, der vollständigen Abbildung des transaktionalen Prozesses auf der In-Memory-Lösung. Spätestens dann müssen diese Lösungen allerdings auch alle funktionalen Anforderungen abbilden, die heute an relationale Datenbanken gestellt werden. Allen voran sind dies klassische ACID-Anforderungen: Atomarität, Konsistenz, Isolation und Dauerhaftigkeit von Transaktionen.

Der Anwendungszweck bestimmt die Technologiewahl

Die meisten In-Memory-Lösungen können zumindest auf einem der genannten Anwendungsgebiete eine sinnvolle Investition sein. Trotzdem ist der Vergleich verschiedener Produktangebote unerlässlich, da bis heute kein Produkt alle oben genannten Anforderungen in gleicher Weise – geschweige denn vollständig – abdeckt. Die Technologien im Markt sind sehr unterschiedlich leistungsfähig und bringen unterschiedliche Features und Einschränkungen mit sich, zum Teil schlicht deshalb, weil sie auch für unterschiedliche Anwendungszwecke gedacht sind.

So kommen beispielsweise sehr unterschiedliche Komprimierungstechniken zum Einsatz. Starke Kompression der Daten steigert die Effizienz der Nutzung des verfügbaren Arbeitsspeichers. Dies wirkt sich oftmals auch auf die Lizenzkosten aus, die nicht selten vom komprimierten Datenvolumen abhängen. Da spaltenbasierte Datenbanken sich tendenziell stärker komprimieren lassen als zeilenbasierte, ist dieser Ansatz inzwischen weit verbreitet. Spaltenbasierte komprimierte Daten sind für viele Fragestellungen – vor allem für starke Aggregation bei gleichzeitig selektiver Filterung – tatsächlich massiv schneller als reihenbasierte Ablagen, wie sie fast durchgehend in den herkömmlichen festplattenbasierten, relationalen Datenbanken zum Einsatz kommen. Spaltenbasierte Datenbanken bringen in vielen Fällen aber auch Nachteile, beispielsweise beim Update von Daten und bei Abfragen, die sehr viele Einzelsätze vollständig zurückliefern. Hier wird deutlich, dass die zu realisierende Anwendung maßgeblich die Auswahl der richtigen In-Memory-Lösung bestimmt.

Ein In-Memory-Werkzeug muss sich auch daran messen lassen, wie einfach sich dort Daten integrieren und organisieren lassen. Bei vielen in BI-Suiten integrierten Komponenten lassen sich lediglich die zugrundeliegenden Datenwürfel einer sekundären Datenbank selektiv in den Arbeitsspeicher replizieren. Diese Konfiguration ist relativ einfach und benötigt keine Programmierung. Der Nachteil: andere Datenstrukturen lassen sich damit nicht abbilden und Einschränkungen der sekundären Datenbank hinsichtlich nutzbarer Datenquellen vererben sich

auf die In-Memory-Lösung. Wer Daten aus sozialen Netzwerken, Hadoop und anderen Big-Data-Datenquellen benötigt, sollte daher in der Werkzeugauswahl auch diesbezügliche Schnittstellen berücksichtigen.

In-Memory hat unterschiedliche Modellierungswerkzeuge

Auch die Modellierungswerkzeuge für den Aufbau der Datenstrukturen innerhalb des In-Memory-Tools unterscheiden sich deutlich. Bei proprietären Komponenten innerhalb von BI-Suiten haben Administratoren und Anwender meistens keine wirklichen Einflussmöglichkeiten, aber eben auch keine Arbeit mehr. Selbst für In-Memory-Lösungen, die auf Datenbeständen anderer Technologieherkunft operieren, existieren durchaus intelligente Ansätze. *Qlikview* beispielsweise ermittelt Assoziationen zwischen verschiedenen Datentabellen automatisch und macht so – zumindest theoretisch – eine explizite Modellierung überflüssig. Solche Automatismen sind aber nur so gut wie die impliziten Metadaten der jeweiligen Quellen. *Qlikview* & Co zwingen folglich zu Systematik und Disziplin bei der Bereitstellung und beim Import der Daten.

Ein weiteres Unterscheidungsmerkmal für In-Memory-Werkzeuge liegt in der Verfügbarkeit von komplexen Analyseverfahren (Data Mining, Predictive Analytics). SAS als langjähriger Marktführer in diesem Segment bietet mittlerweile mehrere In-Memory-Lösungen an, mit *SAS Visual Analytics* als Flaggschiff. Andere BI-Hersteller versuchen ihre In-Memory-Komponenten ebenfalls mit Analyse-Intelligenz auszustatten. Einige integrieren die populäre Open-Source-Plattform für Analysesoftware R. Diese Plattform bietet auch eine Vielzahl an Datenschnittstellen zu den wichtigsten modernen Datenquellen, allen voran den Sozialen Netzwerken. Trotzdem dürften SAS und IBM (mit SPSS) auf dem Feld der Predictive Analytics noch auf lange Sicht einen deutlichen funktionalen Vorsprung vor allen anderen BI-Anbietern haben.

Schwer zu durchschauen ist auf den ersten Blick, wie Anwendungen die im Arbeitsspeicher verfügbaren Daten nutzen. Details können hier den Performancenutzen massiv beeinflussen. Wenn beispielsweise die Anwendungslogik einer Applikation

weitgehend unverändert außerhalb der In-Memory-Lösung abläuft und für diesen Zweck zur Laufzeit große Datenmengen zwischen der In-Memory-Datenbasis und der Anwendungsschicht transportiert werden müssen, kann hierbei leicht ein großer Teil des Performancegewinns wieder aufgefressen werden. Die meisten Hersteller unternehmen bereits Anstrengungen zur Optimierung der Anwendungslogik auf In-Memory-Lösungen und zu deren Integration in die In-Memory-Komponente. Viele Anbieter nutzen festplatten- und arbeitsspeicherbasierte Datenbanken aber heute auch noch ähnlich wie bisher, auch um dadurch weitgehend plattformunabhängig zu bleiben.

An den Technologiegrenzen wird dieses Problem besonders offenkundig. Wo unter bestehende Anwendungen des einen Herstellers lediglich zur Laufzeitverbesserung eine In-Memory-Datenbank eines anderen Anbieters geschoben werden soll, führt dies für manche Anwendungen eventuell nicht zu den gewünschten Effekten. Die Ursache liegt darin, dass an den Technologiegrenzen überwiegend die Standards SQL (Structured Query Language), ODBC (Open Database Connectivity) und JDBC (Java Database Connectivity) – vereinzelt auch MDX (Multidimensional Expressions) und ODBO (OLE DB for OLAP) – eingesetzt werden. Alle Standards sind heute noch vom klassischen Modell der festplattenbasierten Datenbank geprägt und kennen viele Funktionen nicht, die beim Einsatz von In-Memory-Lösungen nützlich wären. Allerdings darf man davon ausgehen, dass die Standards sich zukünftig diesbezüglich weiterentwickeln und gegebenenfalls auch neue Standards entstehen. Insofern sollten sich IT-Architekten bis dahin auch nach der Unterstützung zusätzlicher Programmierschnittstellen, vorhandenen Zertifizierungen und Referenzprojekten mit ähnlichen Anforderungen erkundigen.

Manche Schnittstellen arbeiten Hersteller-exklusiv

Hersteller, die sowohl In-Memory als auch Anwendungskomponenten anbieten, setzen bewusst auf proprietäre Schnittstellen, die exklusiv die eigenen Anwendungen bedienen. Für Unternehmen, die gerne alle Komponenten aus einer Hand nehmen, ergibt

dies den Vorteil einer potenziell besser integrierten Gesamtlösung. So bieten die Walldorfer *SAP HANA* seit 2012 integriert mit ihrem Data-Warehouse-Produkt *SAP BW* und seit kurzem auch als Datenbank unter der *SAP Business Suite*. Die Reife dieser Kombilösungen lässt sich trotz ermutigender Erkenntnisse aus den ersten Projekten zu *SAP BW on HANA* und des Feedbacks von anderen SAP-HANA-Pilotkunden noch nicht abschließend beurteilen. In jedem Fall kann man davon ausgehen, dass SAP noch erheblich in die Weiterentwicklung von *SAP HANA* und die Anpassung ihrer Anwendungen auf In-Memory-Abläufe investieren muss. Die meisten Kunden des Walldorfer Konzerns haben das Vertrauen, dass sich dort am Ende ein Erfolg einstellen wird.

Oracle bietet mit *Exalytics* eine Paketlösung aus spezieller SUN-Hardware, der In-Memory-Datenbank *Oracle Times Ten* und der *Oracle BI Foundation Suite* einschließlich der multidimensionalen In-Memory-Datenbank *Oracle Essbase*. Die Kalifornier fokussieren ihr In-Memory-Offering damit bisher noch auf die klassische Funktion der Beschleunigung von BI-Applikationen. Für Bestandskunden von Oracle ist dieses Angebot durchaus interessant. Für Unternehmen, die bisher noch keine BI-Lösung von Oracle im Einsatz und auch keinen Wechsel planen, entfaltet dieser Ansatz allerdings lediglich eingeschränkte Attraktivität.

Oracle und SAP liefern sich seit geraumer Zeit intensive Marketingschlachten um ihre Appliances. Tatsächlich eignen sich beide In-Memory-Datenbanken für schnelle interaktive Datenanalyse. Es gibt jedoch eine Reihe nennenswerter Unterschiede zwischen *Oracle Exalytics* und *SAP HANA*:

- *Oracle Exalytics* ist noch nicht auf inkrementelle Deltabeladungen, auf Schreiboperationen und den Einsatz in transaktionalem Kontext ausgelegt. SAP verfolgt mit *SAP HANA* hingegen einen sehr breiten Ansatz, der eines Tages im Prinzip alle oben genannten Anwendungen unterstützen soll und es bereits heute an vielen Stellen tut.
- Der Einsatz von *Oracle Exalytics* setzt *Oracle Exadata* als sekundäre Datenbank voraus, wo größere Volumina verbleiben, während im Arbeitsspeicher lediglich bis zu einem

Der Autor



Foto: Mayato

Marcus Dill, Geschäftsführer vom Analysten- und Beraterhaus Mayato, ist Experte für Business Intelligence und Customer Relationship Management.

Terabyte Arbeitsspeicher verfügbar sind. *Oracle Exalytics* umgeht diese Beschränkung dadurch, dass lediglich die am häufigsten beziehungsweise die zuletzt genutzten Daten In-Memory gehalten werden. *SAP HANA* hingegen kommt ohne sekundäre Datenbank aus und hält sämtliche Daten im Speicher.

- Den Scale-Out-Fall behandeln Oracle und SAP unterschiedlich. Dieser Fall tritt dann ein, wenn die Grenzen einer Server-Hardware erreicht sind und weitere Server in ein Netzwerk von mehreren Knoten verschaltet werden müssen. Für *Oracle Exalytics* ist dieser Scale Out noch nicht wirklich vorgedacht. Man kann lediglich mehrere *Exalytics*-Appliances parallel und unabhängig voneinander betreiben. SAP hingegen musste den Scale-Out-Fall angesichts des zuvor betrachteten Aspekts zwangsläufig berücksichtigen, muss hier allerdings auch noch den Nachweis erbringen, dass die hierfür entwickelten Konzepte in der Praxis tragen.

Oracle Times Ten bieten die Kalifornier auch als eigenständige Datenbank an – mit einem im Vergleich zu herkömmlichen Datenbanken eingeschränktem Funktionsumfang – und als zusätzlichen Cache für Datenbanken. Im letztgenannten Szenario werden gro-

ße, Performance-kritische Tabellen in *Oracle Times Ten* ausgelagert und so Anwendungen – auch transaktionale – beschleunigt. Im Gegensatz zum Pendant *IBM SolidDB*, das als In-Memory-Cache beliebige SQL-Datenbanken als Quelle unterstützt, ist *Oracle Times Ten* beschränkt auf die Nutzung in Kombination mit Oracle Datenbanken. Laut Analysteneinschätzungen werden sowohl Oracle als auch IBM in nächster Zeit ihre In-Memory-Funktionalitäten auch als Bestandteil ihres jeweiligen Datenbankangebots verfügbar machen. Damit würden sie mit *SAP HANA* vergleichbar.

Big Blue verfügt neben *IBM SolidDB* über weitere In-Memory-Lösungen in seinem BI-Portfolio, darunter *IBM TM1*. *Cognos Dynamic Cubes* ähnelt funktional eher dem SAP-HANA-Vorläufer *SAP BW Accelerator*, mit dem SAP schon seit einigen Jahren erfolgreich sein Data Warehouse beschleunigt. Gleiches gilt für den *Informix Warehouse Accelerator*. Beide Angebote von Big Blue werden jedoch unabhängig von der Hardware vertrieben, bieten den Kunden also in dieser Hinsicht etwas mehr Flexibilität.

Pauschalempfehlungen sind gegenwärtig nicht sinnvoll

Der Markt für In-Memory-Werkzeuge ist stark in Bewegung und wird seine Dynamik wohl in den nächsten Jahren weiter aufrechterhalten. In der Praxis müssen Konzepte erprobt und verfeinert werden. Manches wird dabei sicher auch verworfen. Da gleichzeitig neue Ideen für Anwendungen und damit auch neue Anforderungen entstehen, sind Pauschalempfehlungen für das eine oder andere Werkzeug gegenwärtig nicht sinnvoll. Unternehmen müssen ihre Investitionsentscheidungen trotzdem sorgfältig und individuell absichern. Erfahrungen im Einsatz mit In-Memory-Werkzeugen zu sammeln, ist sicherlich für jedes Unternehmen ein guter Rat. Das defensive Vorgehen, zu warten bis andere durch Evaluationen und Pilotinstallationen die Spreu vom Weizen getrennt haben, kann leicht zu einem großen Rückstand gegenüber Konkurrenten führen. Der Einsatz von In-Memory-Technologie dürfte sich nämlich in vielen Bereichen sehr schnell zum Standard entwickeln. Wer nicht rechtzeitig mitmacht, gerät vielleicht bald ins Hintertreffen. *jj*