
# In-Memory-Datenbanken: Perfekte Add-on-Technologie für schnellere Analysen

Mathias Golombek, EXASOL AG

Dieser Artikel zeigt Einsatzzwecke und Vorteile der Daten-Analyse mithilfe von In-Memory-Technologie für datengetriebene Unternehmen auf. Darüber hinaus sind beispielhaft die In-Memory-Lösungen von Oracle und EXASOL sowie ihre Vor- und Nachteile für individuelle Unternehmensansprüche gegenübergestellt. Zusammen mit einem Anwenderbeispiel bietet der Artikel besonders Unternehmen eine gute Übersicht darüber, ob und in welcher Art sich der Einsatz einer analytischen In-Memory-Datenbank für sie lohnt.

Die Themen „Big Data" und „Daten-Analyse" sind deutlich mehr als der vielzitierte Hype, Big Data ist Realität: Innerhalb von nur 18 bis 24 Monaten verdoppeln sich produzierte Terabyte-Datenmengen in Unternehmen im Durchschnitt. Um von den Vorteilen der Daten-Analyse zu profitieren, sind häufig Unternehmens-, Produkt- oder Kundendaten für bestimmte Anwendungen im Subsekunden-Bereich zu analysieren. Für solche komplexen und schnellen Analysen sind In-Memory-Datenbanken eine ideale Lösung. Dieser Artikel geht auf die Besonderheiten analytischer In-Memory-Datenbanken als ergänzenden Performance-Layer für Oracle-Infrastrukturen ein.

## Analytik und die Vorteile von In-Memory

Der Einsatz analytischer Datenbanken nimmt zu, denn in den Unternehmen wächst die Nachfrage nach Analytik. Der Bedarf ist insbesondere im Marketing hoch, wenn es um die Steigerung der

Kunden-Profitabilität durch Kunden-Identifizierung in Echtzeit und intelligente Kunden-Interaktionen geht. In der Supply Chain stehen die Prozess-Optimierung durch bessere Planung sowie die Identifizierung und Vermeidung von Risiken im Vordergrund. Auf der Ebene der Unternehmenssteuerung geht es um rechtzeitiges Erkennen von Markt-Trends und Innovationspotenzialen. So durchdringt Analytik alle Unternehmensbereiche.

Analytische Datenbanken bieten zugleich große Vorteile: Sie verbessern die Skalierbarkeit und die Performance von analytischen Datenbank-Abfragen gegenüber traditionellen Datenbanken deutlich. Zusätzlich helfen sie auch, die Betriebskosten zu senken. Das beruht auf der Kombination von bekannten und neuen Technologien wie Spaltenorientierung, Komprimierung, speziellen intelligenten Zugriffsverfahren, massiv paralleler Verarbeitung sowie In-Memory-Technologien. Speziell bei relationalen SQL-Datenbanken mit einer spaltenbasierten In-Memory-Datenverarbeitung lassen sich sehr viele Daten in kurzer Zeit komprimiert hundertmal schneller wegschreiben als in herkömmlichen zeilenorientierten Datenbanken. Der Zugriff auf im Hauptspeicher liegende Daten ist um bis zu tausendmal schneller als der Zugriff auf Daten, die sich auf der Festplatte befinden. In den meisten Fällen sind analytische In-Memory-Datenbanken insgesamt dadurch um den Faktor fünfzig bis hundert schneller und ermöglichen fundierte Echtzeit-Reaktionen basierend auf großen Datenmengen, die zuvor undenkbar waren. So lassen sich Ad-hoc-Auswertungen in Sekundenschnelle auf Knopfdruck bereitstellen.

Datengetriebenen Unternehmen bieten In-Memory-Datenbanken zudem das benötigte Maß an Flexibilität: Die Lösungen sind leicht zu implementieren und skalierbar, was zu einer deutlichen Entlastung der eigenen IT-Ressourcen führt. Die Daten-Analyse vereinfacht sich außerdem erheblich, da strukturierte und unstrukturierte Daten aus verschiedenen Vorsystemen in einer In-Memory-Datenbank direkt ausgewertet werden können – ein vorheriges Aggregieren von Basisdaten ist nicht mehr zwingend erforderlich. Durch die zahlreichen Vorteile von In-Memory eröffnen sich Unternehmen so völlig neue Dimensionen der Daten-Analyse *(siehe Abbildung 1)*.

Zusätzlich zu den klaren Performance-Verbesserungen helfen In-Memory-Datenbanken auch, Investitionskosten zu minimieren. Big-Data-Analysen erfordern in der Regel keinen kompletten Umbau vorhandener Systeme. Oftmals geht es nur um gezielte Ergänzungen und Erweiterungen. Ausschlaggebende Kriterien für die Kosten der einzusetzenden Analyse-Systeme sind insbesondere die Menge der anfallenden Daten sowie die genutzte Datenbank-Technologie. Setzt ein Unternehmen auf eine analytische In-Memory-Datenbank, fallen die Lizenzkosten häufig nicht auf die gesamten Datenmengen an, sondern reduzieren sich auf den tatsächlich genutzten Arbeitsspeicher oder, wie bei der Oracle-In-Memory-Option, auf die Anzahl der CPUs. Hinzu kommen intelligente Kompressions-Algorithmen und vollautomatisierte Prozesse, die die Investitionskosten einer In-Memory-Datenbank im Vergleich zu konventionellen Datenbanken mehr als halbieren können. Auch Firmen mit beschränktem Budget sind damit in der Lage, durch flexibel skalierbare Software-as-a-Service- oder komplette Cloud-Lösungen Big-Data-Projekte erfolgreich und effizient umzusetzen. Hier empfiehlt sich ein Testlauf bei einem zeitlich begrenzten Projekt. Sind die Potenziale erkannt, können Unternehmen die temporären Lösungen Stück für Stück ausbauen und auf andere Bereiche und Projekte ausdehnen.

## In-Memory-Lösungen im Vergleich

Seit dem Jahr 2013 bietet Oracle mit seiner Database 12c Enterprise Edition eine eigene In-Memory-Option an. Vor diesem Zeitpunkt mussten sich Nutzer von Oracle extern nach einer In-Memory-Datenbank-Lösung umsehen: Der Katalog-Versender Quelle AG, Anfang der 2000er-Jahre eines der größten Data-Mining-Unternehmen in Europa, verwaltete seine enormen Datenmengen deshalb nicht nur mithilfe eines der größten Oracle-RAC-Systeme seiner Zeit, sondern nutzte ab dem Jahr 2004 zusätzlich die In-Memory-Lösung der EXASOL AG.

Mit dem Release der 12*c* Enterprise Edition müssen sich Nutzer nicht mehr zwingend nach einer externen In-Memory-Datenbank umsehen. Dennoch kann eine

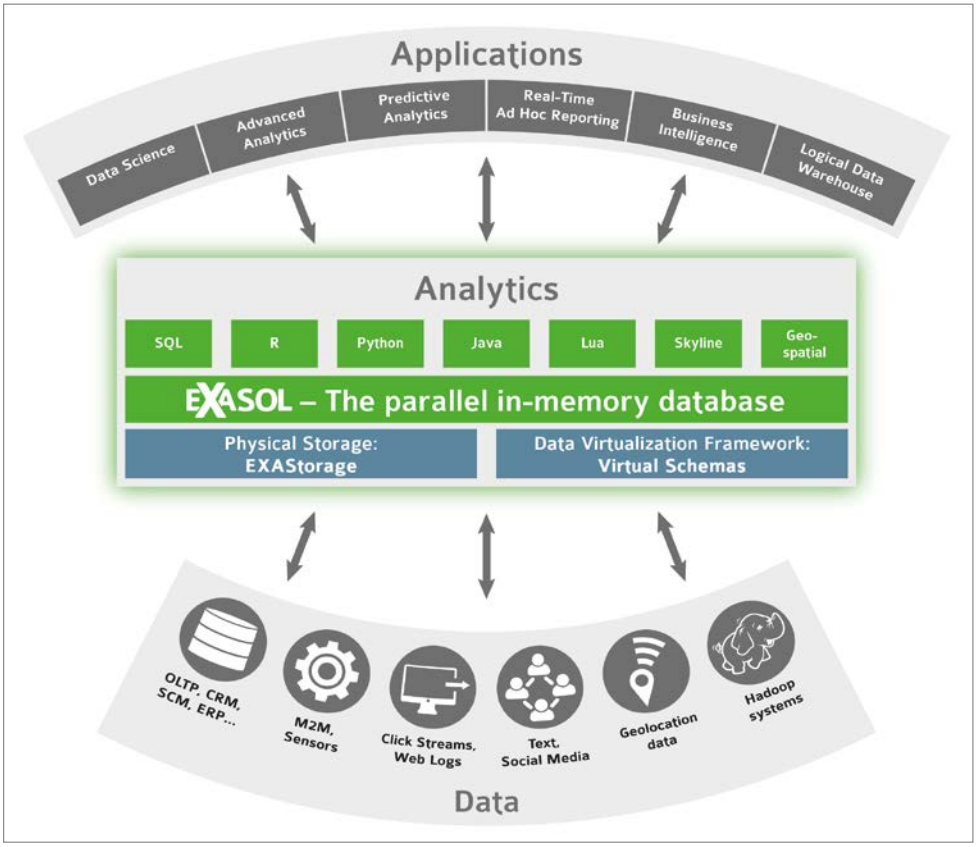


*Abbildung 1: Mit einer In-Memory-Datenbank lassen sich strukturierte sowie unstrukturierte Daten aus verschiedenen Vorsystemen ohne vorherige Aggregation direkt analysieren*

durchdachte Wahl klaren Mehrwert bieten: Sowohl die speziell für ein Oracle-Umfeld entwickelte Datenbank selbst als auch die auf ein solches Umfeld angepasste Datenbank von EXASOL bieten beispielsweise spezifische Eigenheiten, die unterschiedliche Vor- und Nachteile mit sich bringen. Die Entscheidung, welche Datenbank den eigenen Ansprüchen am besten gerecht wird, sollte daher individuell und gut durchdacht getroffen werden, wie nachfolgend in einem Vergleich zwischen den beiden Lösungen dargestellt ist.

Ein signifikanter Unterschied liegt bereits in der Speicherung der Daten. Bei Oracle 12*c* werden die Daten erst auf die Festplatte geschrieben und anschließend in den Hauptspeicher geladen. Dieses Vorgehen spart Platz im Hauptspeicher − das ist notwendig, da es zu Query-Abbrüchen kommen kann, falls der Hauptspeicher für die jeweilige Abfrage nicht ausreicht. Jedoch beeinflusst es die Performance der Datenbank: Der Zugriff auf das vergleichsweise langsame Transaktionslog sowie die Praxis, Daten erst auf die Festplatte zu schreiben und anschließend in den Hauptspeicher zu laden, verursachen vermeidbare Wartezeiten. Zusätzlich unterscheiden sich die Daten-Strukturen im Hauptspeicher von denen auf der Festplatte und müssen zwangsläufig konvertiert werden – ein weiterer Faktor, der Performance-Einbußen bei der Oracle-In-Memory-Lösung verursacht.

Bei EXASOL sind alle Daten automatisch im Hauptspeicher abgelegt. Die hierfür am besten geeigneten Kompressions-Algorithmen werden von der Datenbank automatisch gewählt. Das verschafft einen klaren Performance-Vorteil: Geänderte und neue Daten landen nicht zuerst auf der Festplatte, sondern direkt im Hauptspeicher und stehen somit nach einem Commit sofort zur Verfügung. Alle Daten-Objekte, auf die ein Zugriff erfolgt, werden in den Hauptspeicher geladen und bleiben dort für nachfolgende Abfragen bestehen. Trotzdem kommt es hier nicht zu Query-Abbrüchen; sollte der Hauptspeicher für die Abfragen nicht ausreichen, sinkt lediglich die Performance des Systems etwas.

Zusätzliche Geschwindigkeit gewinnt diese Datenbank-Lösung durch die Tatsache, dass die Daten-Strukturen im Hauptspeicher denen auf der Festplatte entsprechen und Daten somit hoch performant durch Memory-Mapping ein- und ausgelagert werden können. Der Prozessor wird hierbei nicht belastet. Dieser Ansatz verschafft mehr Performance – benötigt aber mehr Platz im Hauptspeicher.

Ein weiterer Unterschied ist die Tatsache, dass es 12*c*-Nutzern ermöglicht wird, die eigenen Daten durch eine Partitionierung in Hot- und Cold-Data zu unterteilen, wobei nur die Hot-Data-Partitionen in den Hauptspeicher geladen werden und der Rest auf der Festplatte verbleibt. Dank dieser Möglichkeit wird kein unnötiger Speicherplatz im Hauptspeicher blockiert. Bei EXASOL hingegen besteht die Möglichkeit zur Partitionierung nicht. Aus diesem Grund kann hier durch den selteneren Zugriff auf Cold-Data wichtigeres Datenmaterial aus dem Arbeitsspeicher verdrängt werden. Es muss so bei der nächsten Abfrage erst wieder eingelagert werden. Dieses Problem hat 12*c* nicht – aber die Möglichkeit zur Partitionierung bringt auch Nachteile mit sich: Bei Abfragen stehen nämlich nur die Daten-Objekte zur Verfügung, für die explizit festgelegt wurde, dass sie im Hauptspeicher gehalten werden sollen. Alle weiteren Daten müssen bei jeder Abfrage erneut von der Festplatte oder aus dem Cache eingelesen und verarbeitet werden, was die Bearbeitungszeit deutlich erhöht.

Die beiden Systeme unterscheiden sich unter anderem auch in ihren Lizenzierungsmodellen: Die Lizenzierung bei EXASOL ist ausschließlich RAM-basiert, während Oracle pro CPU berechnet. Oracle 12*c* ist eine sehr flexible Lösung, die gut an individuelle Bedürfnisse angepasst werden kann, dafür aber mit klarem manuellen Konfigurationsaufwand verbunden ist. Die Flexibilität, aber auch der manuelle Konfigurationsaufwand der Oracle-Datenbank sind relativ hoch und einige Fragen müssen wiederholt geklärt werden:

- Wie viel RAM soll/darf verwendet werden?
- Welche Tabellen und welche Spalten sollen im Speicher gehalten werden?
- Wie sollen diese komprimiert werden?
- Wann sollen die Daten in den Speicher geladen werden? („Lazy loading" bei Zugriff oder Laden beim Starten der Datenbank?)

Anwender können und müssen hier folglich viele Dinge selbst entscheiden und einstellen.

Bei EXASOL hingegen ist eine dezidierte Optimierung auf ein spezifisches Anwendungsszenario nicht möglich. Dafür entfällt aber auch der andauernde manuelle Konfigurationsaufwand: Dank der umfangreichen Oracle-Kompatibilität, dem nativen OCI-Support (Oracle Call Interface) und der Tatsache, dass die SQL-Abfragen beider Datenbanken syntaktisch nahezu „1:1" übereinstimmen, kann die Lösung einfach, schnell und mit geringem Aufwand in eine bestehende Oracle-Infrastruktur integriert werden und bearbeitet Anfragen danach automatisch und quasi ohne manuellen Aufwand.

## Welche Datenbank ist die bessere Alternative?

Die Einführung von Oracle 12c entbindet Oracle-Nutzer nicht zwingend von einer intelligenten Wahl der zu nutzenden In-Memory-Lösung. Wie am Beispiel von EXASOL beschrieben, bieten auch andere In-Memory-Lösungen anwendungsspezifische Vorteile und können, wie in diesem Fall, zum Beispiel komplementär als analytischer Performance Layer („Accelerator") oberhalb von Oracle verwendet werden und somit eine größtmögliche Geschwindigkeit innerhalb des Oracle-Umfelds garantieren oder alternativ für bestimmte Anwendungen als paralleles System zu einer Oracle-Umgebung eingesetzt werden *(siehe Abbildung 2)*.

Verschiedene In-Memory-Datenbanken bieten verschiedene anwendungsspezifische Vorteile und weisen gleichzeitig unterschiedliche Nachteile auf. So ist, wie auch John Appleby vom SAP-Beratungsunternehmen bluefin zu berichten weiß *(siehe „http://www.bluefinsolutions.com/blogs/john-appleby/october-2014/oracle-database-in-memory-faq")*, die In-Memory-Option von Oracle eher für einfachere Abfragen ausgelegt und kann bei komplexen analytischen Abfragen mit Performance-Einbußen zu kämpfen haben. Tatsächlich kann die Performance hier so weit abfallen, dass sie schlechter ist als bei Oracle-Systemen ohne In-Memory-Technologie. Zudem ist der manuelle Konfigurationsaufwand der Oracle-Lösung beträchtlich. Dafür besticht die Datenbank mit ihrer Flexibilität und Anpassungsfähigkeit an individuelle Bedürfnisse sowie mit der relativ geringen Hauptspeicherplatz-Belastung. Die Datenbank von EXASOL hingegen benötigt mehr

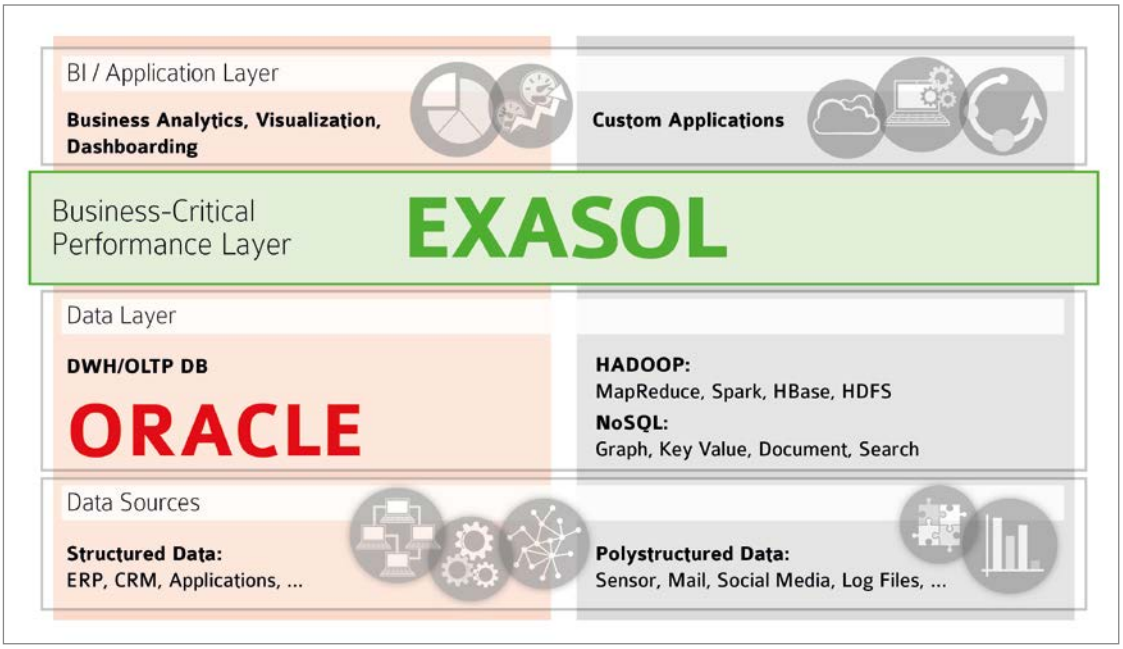


*Abbildung 2: EXASOL im Einsatz als Add-on-Technologie in einer Oracle-Umgebung zur Beschleunigung der Daten-Analyse*

Platz im Hauptspeicher und weist eine geringere Flexibilität auf als Oracle 12c. Dafür ist sie einfach implementiert und arbeitet fast ausschließlich selbstständig – der manuelle Konfigurationsaufwand entfällt hier. Außerdem beeindruckt sie durch enorme Performance und ist explizit für komplexe, analytische Abfragen ausgelegt.

Welche In-Memory-Lösung die richtige für das eigene Unternehmen ist, muss jeder selbst entscheiden – doch diese Entscheidung sollte durchdacht sein und am besten über eine Evaluierung der Bestandslage im Vorfeld einer geplanten Erweiterung oder Umstellung getroffen werden. Ein sogenannter „Proof of Concept" gibt Aufschluss über das Potenzial der In-Memory-Lösung im Unternehmen und plant das Projekt zudem personell und monetär ein. Die richtige Wahl der In-Memory-Datenbank, basierend auf den Bedürfnissen des Unternehmens und auf Informationen über die Vor- und Nachteile verschiedener Lösungen, hilft Unternehmen, sich den größten Mehrwert zu sichern, den Analytik und In-Memory zu bieten haben.

## Anwenderbeispiel: Accarda AG

Ein Anwender, der sich entschlossen hat, Oracle und EXASOL zu kombinieren, ist der Schweizer Finanzdienstleister Accarda. Um Umsatzverluste im Einzelhandel zu vermeiden, entwickelte Accarda in Zusammenarbeit mit der Manor AG das „Accarda Loss Prevention System", kurz ALPS. Diese Lösung ermittelt auffällige Trends, Unregelmäßigkeiten und Abweichungen in den Bezahlprozessen, um fehlerhafte Abrechnungen zu identifizieren. Hierzu müssen Ad-hoc-Analysen auf enormen Datenmengen ausgeführt werden. Aus diesem Grund benötigt ALPS eine hochleistungsfähige Datenbank im Hintergrund und setzt seit rund zwei Jahren auf die analytische In-Memory-Lösung aus Nürnberg.

Die Datenbank dient dem ALPS, um den Prozess flexibel und automatisch zu gestalten. Zu Beginn des Prozesses liefern Kunden ihre Kassendaten an. Accarda bereitet die Daten auf und analysiert sie hinsichtlich Strukturbrüchen, definierter Kennzahlen, Unregelmäßigkeiten oder Schwellenwertüberschreitungen. Registriert ALPS auffällige Kassentransaktionen oder Abläufe, so werden diese automatisch auf einem Report ausgewiesen und die zuständigen Stellen benachrichtigt. Die Reporte stehen täglich zur Verfügung. Zudem können Kunden vertiefende Ad-hoc-Analysen durchführen, um konkrete Nachforschungen anzustoßen oder Maßnahmen zur Prävention von Diebstahl einzuleiten. Damit diese Prozesse schnell ablaufen, bedarf es einer skalierbaren Lösung, die sämtliche Analysen auf Ebene der Einzeltransaktionen unterstützt und Daten „on-the-fly" analysiert. So profitiert der Finanzdienstleister in erster Linie davon, dass Daten-Aggregationen überflüssig werden beziehungsweise die Materialisierung der Daten entfällt. Lade- und Analysezeiten wurden erheblich beschleunigt. Die Handelspartner des Unternehmens profitieren wiederum von ständiger Verfügbarkeit und Analysen, die bis auf die Ebene der Einzeltransaktionen reichen. Laut Stefan Schurgast, dem Business Development und Innovation Manager der Accarda AG, war die Oracle-Kompatibilität von EXASOL das Hauptkriterium für die Auswahl dieser Datenbank. Sie unterstützt zudem Standard-SQL und läuft auf Standard-Hardware. Daher ist es möglich, auch nach der Integration bereits bestehende ETL-Tools weiterhin zu nutzen. Außerdem erklärt Schurgast, hat ihn auch die Vision mit den vier Eckpfeilern Clustering, MPP, spaltenorientierte Speicherung und intelligente Kompressionsalgorithmen sowie der In-Memory-Komponente überzeugt.

Mathias Golombek
mathias.golombek@exasol.com