



Hauptspeicher-Datenbanken: Denkgeschwindigkeit für alle?

Jens Lechtenböcker und Gottfried Vossen, Institut für Wirtschaftsinformatik,
Westfälische Wilhelms-Universität Münster

Eine Datenbank gilt einem Übersichtsartikel [1] aus dem Jahr 1989 zufolge als „Hauptspeicher-Datenbank“, wenn sich der primäre Datenbestand im Hauptspeicher befindet; Festplatten werden lediglich für Backup und Recovery verwendet.

Angesichts stetig steigender Hauptspeicher-Größen bei gleichzeitig stetig fallenden -Kosten gewinnen Hauptspeicher-Datenbanken seit einigen Jahren zunehmend an Popularität, wenn es um höchste Geschwindigkeitsansprüche geht. In der Tat gibt es gute Gründe für die Entwicklung spezialisierter Datenbank-Systeme, denn es wurde gezeigt, dass traditionelle, festplattenbasierte Datenbank-Systeme im populären TPC-C-Benchmark zur Leistungsmessung der Transaktionsverarbeitung, selbst wenn alle Daten vollständig in den Hauptspeicher passen, nur 10 Prozent der Zeit mit echter Transaktionsverarbeitung verbringen, während die verbleibenden 90 Prozent auf vier Quellen von Mehraufwand entfallen, nämlich die Verwaltung des Buffer-Pools (1), die Verwaltung des Multi-Threading (2), das Sperrverfahren auf Zeilenebene (3) und das Write-Ahead-Logging (4) [2]. Entsprechend liegt es nahe, spezialisierte Hauptspeicher-Datenbanksysteme zu entwickeln, in denen diese Arten von Mehraufwand eliminiert oder zumindest reduziert werden.

Mittlerweile bieten auch die großen Datenbank-Hersteller, darunter Oracle, hauptspeicherbasierte Produkte an und bewerben diese mit Geschwindigkeitsverbesserungen um Faktoren bis zu 100.000. Damit könnten Analysen oder Prognosen, deren Berech-

nungen bisher Tage dauerten, nun innerhalb von Sekunden durchgeführt werden, was neuartige Anwendungen ermöglicht. Vor diesem Hintergrund sprechen Plattner und Zeier [3] von einer Revolution für die Entwicklung betrieblicher Anwendungen.

Anwendungsbeispiele

Hauptspeicher-Datenbanken können enorme Geschwindigkeitsvorteile gegenüber traditionellen, festplattenbasierten Datenbanken ausspielen. Im Prinzip lässt sich eine festplattenbasierte Datenbank, die im Hintergrund einer zu langsamen Anwendung den Flaschenhals bildet, einfach durch eine Hauptspeicher-Datenbank austauschen, jedenfalls wenn sichergestellt ist, dass die Anwendung keinen herstellerspezifischen SQL-Dialekt voraussetzt, sondern eine Teilmenge des SQL-Standards nutzt, die von beiden Datenbanken verstanden wird. So berichtete Oracle bei der Vorstellung seiner hauptspeicherfähigen Datenbankversion 12c von Beschleunigungsfaktoren bis zu 3.500 für die hauseigene Unternehmenssoftware (Peoplesoft, JD Edwards), die in vielen Fällen zuvor langwierige Berechnungen zu interaktiven Vorgängen machten (siehe „<http://www.oracle.com/us/corporate/events/dbim/index.html>“).

Natürlich führt nicht jede Geschwindigkeitssteigerung direkt zur Denkgeschwin-

digkeit. Dennoch können Reduktionen der Berechnungszeiten von Stunden auf Minuten oder Sekunden zu deutlichen Geschäftsvorteilen führen. Als erstes Beispiel sei die Immonet GmbH genannt, ein Internet-Immobilienportal, das seine Auswertungen zu Markttrends, zu Kaufmustern von Kunden und zum Maklerverhalten auf die Exalytics In-Memory Machine X2-4 von Oracle umgestellt hat („siehe <http://www.oracle.com/us/solutions/fastdata/fast-data-customer-reference-2041391.pdf>“). Generell wurden dramatische Verbesserungen in der Ausführung von Abfragen und der Erstellung von Berichten erzielt, von Tagen oder Stunden auf Minuten oder Sekunden. In der Folge konnte das Unternehmen seine Suchmaschinenwerbung verbessern sowie die Kundenanfragen um 300 Prozent und die Umsatzerlöse um 200 Prozent steigern.

Weitere konkrete Anwendungsbeispiele wurden etwa im Rahmen der Oracle OpenWorld 2015 vorgestellt („siehe https://blogs.oracle.com/In-Memory/entry/oracle_openworld_2015_wrap_up“). Dort berichtete Swiss Mobiliar sowohl von der erfolgreichen Proof-of-Concept-Migration einer DB2/Netezza-Umgebung auf Oracle 12c In-Memory, die in drei Tagen durchgeführt werden konnte und lediglich eine einzige manuelle Tuning-Maßnahme erforderte, als auch von

der Umstellung des Risikomanagements von Oracle 11g auf Oracle 12c In-Memory, die zu durchschnittlichen Beschleunigungsfaktoren von 109 bis 525 führte. Für Beispiele anderer Hersteller von Hauptspeicherdatenbanken sei exemplarisch auf die SAP Customer Testimonials verwiesen, in denen Projekte von SAP-HANA-Anwendern vorgestellt sind („siehe <http://www.sap.com/customer-testimonials/index.html>“).

Losgelöst von konkreten Anwendungsszenarien hat der Geschwindigkeitsvorteil von Hauptspeicher-Datenbanken mehrere allgemeine Implikationen. Wir unterscheiden drei Arten von Vorteilen, die sich unterschiedlichen Anforderungen an Datenverarbeitung zuordnen lassen und nachfolgend kurz skizziert werden.

Erstens verspricht die Hauptspeicherverarbeitung generelle Geschwindigkeitssteigerungen. Im Idealfall arbeiten Anwendungen mit der von Plattner und Zeier in [3] postulierten Denkgeschwindigkeit (engl. „speed of thought“), die als Antwortzeit von etwa 550 bis 750 Millisekunden erreicht werden sollte, um flüssiges und effektives Arbeiten zu ermöglichen: Falls die Antwort auf eine Nutzer-Interaktion länger dauert als diese Denkgeschwindigkeit, wird die resultierende Pause als Wartezeit wahrgenommen und das Gehirn schweift von der eigentlichen Aufgabe zu anderen Dingen ab. Entsprechend ist dann kein konzentriertes Arbeiten mehr möglich. Durch den Einsatz von Hauptspeicher-Datenbanken ist zu erwarten, dass vermehrt Anwendungen diese Denkgeschwindigkeit erreichen können.

Zweitens können Geschwindigkeitsvorteile den Schlüssel zur Erschließung neuartiger (Big-Data)-Szenarien darstellen, bei denen Daten in bisher unzugänglichem Umfang und Detailgrad vorliegen oder produziert werden. Zum einen lassen sich größere Daten-Volumina in kürzerer Zeit auswerten, zum anderen nun auch Datenströme erfassen und auswerten, deren Geschwindigkeit zu hoch für die Verarbeitung in festplattenbasierten Systemen ist. Entsprechend werden Hauptspeicher-Datenbanken häufig mit Big-Data-Szenarien in Verbindung gebracht, die durch verschiedene „V“ charakterisiert sind [4]: Volumen oder Größe (engl. „volume“), (hohe) Geschwindigkeit (engl. „velocity“), (hohe) Vielfalt (engl. „variety“), (nicht immer gegebene) Präzision, Genauigkeit oder Vertrauenswürdigkeit (engl. „veracity“) und Wert (engl. „value“).

Drittens verspricht der Einsatz von Hauptspeicher-Datenbanken die Vereinfachung von IT-Landschaften, in denen bisher aus Leistungsgründen verschiedene Datenbank-Server für operative und dispositive Anwendungen eingesetzt sind. Dies trifft insbesondere auf die klassische Trennung von OnLine Transaction Processing (OLTP) zur Unterstützung von Geschäftsprozessen im Tagesgeschäft und Business-Intelligence-Anwendungen wie OnLine Analytical Processing (OLAP) für Auswertungszwecke zu. Aufgrund der gesteigerten Leistungsfähigkeit von Hauptspeicher-Datenbanken lassen sich auch komplexe OLAP-Auswertungen direkt auf operativen Detaildaten durchführen, ohne dass Optimierungstechniken wie

Voraggregation nötig wären und ohne operative Anwendungen auszubremsen. Neue Akronyme, die diese Kombination von OLTP und OLAP vor allem in Marketing-Materialien verdeutlichen, sind OLXP, OLTP oder HTAP (Hybrid Transactional/Analytical Processing).

Ausgewählte Systeme

In den letzten Jahren ist es zu einer immensen Steigerung von Datenbank-Systemen gekommen, die als Hauptspeicher-Datenbanken vermarktet werden. Diese unterscheiden sich teilweise erheblich, und nicht alle von ihnen sind echte Hauptspeicher-Systeme im Sinne der in der Einleitung zitierten Definition, die den Hauptspeicher als den primären Speicherort der Daten festlegt.

In ihrer Studie [5] verdeutlichen die Autoren (ohne Anspruch auf Vollständigkeit) die Bandbreite verfügbarer Systeme durch eine Auswahl einflussreicher akademischer (HyPer, MonetDB und VoltDB) und populärer kommerzieller Systeme (IBM DB2 BLU, MemSQL, Microsoft SQL Server 2014 Hekaton, Oracle TimesTen, Oracle 12c und SAP HANA DB). Eine längere Auflistung verfügbarer Systeme findet sich auch in Wikipedia (siehe „https://en.wikipedia.org/wiki/List_of_in-memory_databases“).

Fazit

Hauptspeicher-Datenbanken haben in den vergangenen Jahren dramatisch an Popularität gewonnen und zahlreiche Hersteller bieten mittlerweile Systeme an, die durch spezielle Anfrage- und Transaktionsverarbeitungs-Techniken die effiziente Datenverarbeitung im Hauptspeicher anstreben und damit die Festplatte als primären Speicherort der Daten durch den Hauptspeicher ersetzen, was Latenzen eliminiert und einen Leistungssprung in der Datenverarbeitung verspricht.

Mittlerweile zeigt sich ein genereller Trend zu Hauptspeicher-Datenbanken, nachdem die hauptspeicherbasierte Datenverarbeitung zunächst lange Zeit eine Domäne spezialisierter Datenbank-Systeme war: Die drei großen Hersteller traditioneller relationaler Datenbanken – IBM, Microsoft und Oracle – bieten für ihre Flaggschiffe mittlerweile den Betrieb als Hauptspeicher-Datenbank zumindest als Option an. Dementsprechend wird es immer einfacher, das Leistungsverhalten einer Anwendung bei Umstieg auf eine Hauptspeicher-Datenbank zumindest prototypisch zu testen. Für kos-

Acknowledgement

Im vergangenen Jahr wurde von der DBIS Group der Universität Münster eine Studie mit dem Titel „Hauptspeicher-Datenbanken – Denkgeschwindigkeit auch für KMU?“ erstellt. Die Inhalte des zugehörigen Abschlussberichts [5] sind in diesem Artikel in Auszügen wiedergegeben beziehungsweise zusammengefasst. Der vollständige Bericht stellt Einsatzmöglichkeiten, State-of-the-Art sowie mögliche Herausforderungen bei der Einführung von Hauptspeicher-Datenbanktechnologie herstellerneutral dar. Im Einzelnen sind Anwendungsbeispiele für die Nutzung von Hauptspeicher-Datenbanken skizziert, wobei auch

exemplarische Unternehmensprojekte vorgestellt werden. Es werden zentrale technische Grundlagen zusammengetragen, die für das Verständnis der im Zusammenhang mit Hauptspeicher-Datenbanken beworbenen Geschwindigkeitsvorteile notwendig sind. Aufbauend darauf werden ausgewählte Datenbanksysteme mit ihren Besonderheiten vorgestellt, wobei neben bekannten kommerziellen Systemen auch wegweisende universitäre Entwicklungen einbezogen sind. Zudem werden mögliche Herausforderungen bei der Einführung von Hauptspeicher-Datenbanken – insbesondere im Hinblick auf KMU – diskutiert.

tengünstige oder prototypische Neuentwicklungen kann zudem auf freie Software (MonetDB, Community Edition von VoltDB) zurückgegriffen werden.

Literatur

- [1] Eich, Margaret H.: Main Memory Database Research Directions. In Proc. IWDM '89, LNCS 368, Springer, 1989, S. 251–268
- [2] Harizopoulos, Stavros ; Abadi, Daniel J. ; Maden, Samuel ; Stonebraker, Michael: OLTP

through the looking glass, and what we found there. In Proc. SIGMOD 2008, S. 981–992

- [3] Plattner, Hasso ; Zeier, Alexander: In-Memory Data Management: An Inflection Point for Enterprise Applications. 2. Auflage, Springer, 2012
- [4] Vossen, Gottfried: Big Data: Der neue Katalysator für Business und andere Intelligenz. In: Schwarz, Torsten (Hrsg.): Leitfaden Marketing Automation. 2014, S. 1–12
- [5] Lechtenböcker, Jens ; Ling, Vanessa J. ; Vossen, Gottfried: Hauptspeicher-Datenbanken – Denkgeschwindigkeit auch für KMU? / Arbeitsberichte des Instituts für Wirtschaftsinformatik (136), Münster, 2015

Jens Lechtenböcker

lechtenboecker@wi.uni-muenster.de

Gottfried Vossen

vossen@wi.uni-muenster.de

Fünf Tipps von Oracle für eine Cloud-Strategie

Tipp 1: Einen Anwender mit möglichst umfangreichem Lösungsportfolio suchen

Das Lösungsportfolio sollte über alle drei Cloud-Layer (IaaS, PaaS und SaaS) verfügen. Der Grund: Wenn Anwender Cloud-Dienste unterschiedlicher Anbieter beziehen, sind diese mühsam zu integrieren, um zentral gesteuert werden zu können. Damit entstehen zusätzliche Schnittstellen und der Integrationsaufwand wächst. Das bindet Kapazitäten und erhöht die Kosten. Lässt sich ein breites Portfolio an Cloud-Lösungen aus einer Hand beziehen, sind sowohl der Einstieg als auch das Aufstocken leichter: Unternehmen können bereits aufeinander abgestimmte Cloud-Anwendungen ordern, die ohne Debugging, Redesign oder langwierige Testläufe funktionieren. Wenn der Provider auch noch in der Lage ist, von der Infrastruktur über die Plattformen bis hin zur Applikation den gesamten Cloud-Stack abzubilden, ist gewährleistet, dass sich sämtliche Cloud-Dienste nahtlos an die vorhandene IT-Infrastruktur anbinden lassen.

Tipp 2: Cloud-Lösungen sollten sich problemlos in eine bestehende IT-Infrastruktur integrieren lassen und einen flexiblen Wechsel zwischen Cloud und On-premise ermöglichen

Die wenigsten Unternehmen werden sofort eine vollständige Migration in die Cloud anstreben, sondern bestimmte Anwendungen weiterhin On-premise im eigenen Rechenzentrum betreiben wollen – eventuell auch dauerhaft. Möglicherweise geht es zunächst auch nur darum, Leistungsspitzen über die Cloud abzufedern. Laufende Lizenzverträge und bestehende Hardware sind weitere Gründe, warum der Wechsel in die Wolke

eher sukzessive verläuft. Für maximale Flexibilität beim Betrieb in der Cloud und On-premise müssen die Funktionalitäten und das „Look and Feel“ jedoch identisch sein – ansonsten entstehen komplexe mehrschichtige IT-Umgebungen, deren Management einen erhöhten Aufwand bedeuten. Genau dies befürchten laut einer aktuellen Studie von IDG Connect und Oracle (siehe „<https://go.oracle.com/privatecloud>“) auch zahlreiche Entscheider. Diesen Befürchtungen aktiv begegnen können Anbieter, die sowohl in der Cloud als auch On-premise 100 Prozent identische Services, Plattformen und Anwendungen bieten – wie Oracle. So bestimmt der Kunde allein das Tempo bei der digitalen Transformation.

Tipp 3: Marktposition des Anbieters prüfen

Wichtige Indikatoren in diese Richtung sind Branchenerfahrung und Innovationsstärke – die Voraussetzungen dafür wiederum eine langjährige Marktpräsenz und eine gesunde Finanzbasis. Wer einen finanzstarken Anbieter wählt, schließt Zukunftsrisiken aus und macht den Weg frei für eine langfristige enge Zusammenarbeit. Außerdem: Etablierte Technologieanbieter investieren in neue Partnerschaften, Technologien und Märkte und sind so in der Lage, schnelle Innovationszyklen zu realisieren und ihr Cloud-Angebot kontinuierlich weiterzuentwickeln. Das ist in einem schnelllebigen Marktsegment wie Cloud-Computing essenziell und ohne eine starke Finanzbasis nicht machbar.

Tipp 4: Flexibilität der Integration und Kompatibilität mit bestehenden Systemen und Anwendungen sicherstellen

Der Cloud-Anbieter sollte Lösungen auf Ba-

sis offener Standards anbieten, denn damit haben Unternehmen die freie Wahl, welche proprietären Systeme und Anwendungen sie weiter nutzen wollen. Zudem können sie auf dieser Grundlage eigenständig Plattformen erweitern, eigene Bausteine für die Cloud-Umgebung entwickeln oder Services von Drittanbietern nutzen. So vermeiden man eine exklusive Bindung an den Anbieter. Man kann die digitale Transformation im eigenen Tempo vollziehen – ohne sich Gedanken über die Integration neuer Systeme machen zu müssen.

Tipp 5: Auf Integrationsservices und -kompetenz achten

Unternehmen und ihre Cloud-Anbieter stehen vor der Herausforderung, Cloud-Lösungen in die bestehende unternehmensinterne IT-Infrastruktur einzubinden und die Services anschließend zentral zu verwalten. Wie wichtig insbesondere eine nahtlose Integration von Software-as-a-Service ist, zeigt eine Studie von Dynamic Markets: Demnach geben etwa 50 Prozent der Unternehmen SaaS-Projekte auf, weil sich Cloud-Anwendungen nicht in die vorhandene IT-Umgebung einbinden lassen. Und 54 Prozent der befragten Unternehmen hatten aus demselben Grund bereits Probleme, Projektlaufzeiten einzuhalten. Kommen Cloud-Lösungen unterschiedlicher Anbieter zum Einsatz, steigt die Komplexität. Der Integrationservice der Provider sollte in dem Fall Konnektoren für Applikationen, Transformation Mapper und Tools für das Monitoring der Integration von Private- und Public-Cloud-Diensten umfassen.