Vergleich von Spalten-basierten In-Memory-Datenbanken mit multidimensionalen OLAP-Systemen als Einsatz für BI von mittelständischen Unternehmen

Exposé

HUMBOLDT-UNIVERSITÄT ZU BERLIN

MATHEMATISCH-NATURWISSENSCHAFTLICHE FAKULTÄT II

INSTITUT FÜR INFORMATIK

eingereicht von:	Fabian Weber
geboren am:	16.08.1987
in:	Berlin
Betreuer:	Prof. Dr.-Ing. Ulf Leser (Humboldt Universität zu Berlin)

1. Überblick

Die Welt der Auswertung von Geschäftsprozessen und -daten entwickelt sich stets weiter. Es gibt eine Vielzahl von Technologien, Werkzeugen und Methoden um diese Auswertungen so effizient und komfortabel wie möglich durchzuführen. In dieser Arbeit wird eine neue Generation von Werkzeugen vorgestellt und mit derzeitig genutzten Instrumenten verglichen.

Zunächst werden die benötigten Begriffe und in dieser Arbeit verwendeten Technologien erläutert:

1.1. Business Intelligence

Laut Prof. Dr. Wilhelm Hummeltenberg, von der Universität Hamburg, beschreibt Business Intelligence (BI) „[…] die auf eine Unterstützung, Durchführung und Kontrolle betrieblicher Aktivitäten ausgerichtete Intelligenz (Einsicht) sowie die zu ihrer Erzielung eingesetzten Konzepte, Methoden und Informationssysteme. BI-Systeme sind informationsgetriebene Entscheidungsunterstützungssysteme zur Gewinnung und Verbreitung von Erkenntnissen für und über betriebliche Abläufe.“ [Humm14]. Zu Business Intelligence gehören auch die Konzepte, welche nachfolgend betrachtet werden und die fundamental für diese Arbeit sind.

1.2. OLAP-Systeme

OLAP steht für Online-Analytical Processing. OLAP repräsentiert Software für die Analyse von Unternehmensdaten. Sie dienen den Analysten dazu, große Datenmengen, auf Basis bestimmter Fragestellungen zu durchsuchen und die Resultate aufzubereiten [GlCh06].

OLAP-Systeme gibt es in mehreren Arten [ChDN11, p.92f]. In dieser Arbeit wird jedoch exemplarisch nur multidimensionales OLAP betrachtet, eines der etabliertesten Werkzeuge:

Multidimensionales OLAP (MOLAP)
verwendet multidimensionale Arrays für die Datenspeicherung, so genannte Würfel oder „Cube“s [KeBM04, p.95f]. Multidimensionalität ist hierbei die Hinterlegung der Daten in Klassen von logisch zusammenhängenden Informationen, den Dimensionen. Diese Dimensionen korrespondieren mit der Problemsicht des Analysten und werden dem entsprechend modelliert. Dimensionen haben charakteristischer Weise eine hierarchische Struktur [Oehl06, pp.136–142]. In einem Handelsunternehmen stellen z.B. Artikel eine

Hierarchieebene in einer Dimension „Artikel“ dar. Genauso würde auch eine mögliche höhere Hierarchieebene „Artikelgruppe“ zu dieser konstruierten Dimension gehören. Entlang dieser Hierarchieebenen können die quantitativen Werte verdichtet („Roll-Up“) oder aufgefächert („Drill-down“) werden. Ein MOLAP-Würfel wird meist so gestaltet, dass er komplett in den Hauptspeicher passt und trotzdem so viele bereits verdichtete Werte wie möglich enthält. Verdichtete Werte müssen zur Abfragezeit nicht mehr neu berechnet werden. Das Lagern aller Daten im Hauptspeicher hat den Vorteil, den Flaschenhals zu umgehen, welcher bei der Übertragung der Daten zwischen Festplatte und Hauptspeicher zu finden ist [Brow04]. Auswertungen können so schneller durchgeführt werden. Der MOLAP-Würfel speist sich in der Regel aus einem Relationalen Datenbankmanagementsystem (RDBMS).

Können nicht alle Daten in einen Würfel im Hauptspeicher geladen werden, die Daten werden jedoch für einige Anfragen benötigt, können hybride Techniken verwendet werden, hybrides OLAP (HOLAP). In HOLAP werden Daten, welche nicht im Würfel vorhanden sind aus dem RDBMS bezogen, wodurch allerdings die Geschwindigkeit für solche „Miss-Hits“ gering ausfällt.

Zusätzlich gibt es relationales OLAP, welches Daten nicht in multidimensionalen Arrays voraggregiert und Varianten bei denen ein Teil der Daten abgezogen und auf den Systemen des Nutzers analysiert wird, bekannt als Desktop OLAP (DOLAP) oder auch Self-Service-BI [BaGü04, p.242, ImWh11, Kaha06, chap.Desktop OLAP].

1.3. In-Memory-Datenbanken

Bei In-Memory-Datenbanken, wie bei MOLAP-Würfeln, verbleiben alle Daten für die Dauer des Betriebs komplett im Hauptspeicher. Dies ist erst möglich, seitdem durch die 64-Bit Technologie Hauptspeicher von Größen adressiert werden kann, welche eine komplette Datenbanken zu beherbergen imstande sind. Auch der stetig fallende Preis für RAM macht diese Art der Datenbanken mittlerweile wirtschaftlich [Yell10, p.3]. Wie bei MOLAP können, je nach Größe der Datenbank und des verfügbaren Hauptspeichers des Servers, evtl. nicht alle Daten im Hauptspeicher verbleiben. In diesem Fall können auch hier hybride Techniken zum Einsatz kommen, bei denen nur Teile der Daten im Hauptspeicher liegen, z.B. die Daten, auf die am häufigsten zugegriffen wird [Sche10]. Die Geschwindigkeit des Zugriffs sinkt dann

aber für den Teil der Daten, der nicht im Hauptspeicher liegt. Die Open-Source Datenbank MonetDB unterstützt z.B. diese Möglichkeit der Datenspeicherung.

1.4. Spalten-orientierte Datenbanken

In klassischen Zeilen-basierten Datenbanken müssen für Anfragen über wenige Spalten die jeweils ganzen Zeilen geladen werden. Dabei werden mehr Daten geladen, als letztendlich benötigt werden. Gerade bei der Berechnung von Kennzahlen wird, wie in Abschnitt 1.2 erwähnt, häufig über Spalten aggregiert. Will man nun z.B. die Summe der Umsätze berechnen, benötigt man hierfür nicht die korrespondierenden Produkte zu einer Transaktion.
Eine andere Herangehensweise ist, Tabellen nicht zeilenweise zu hinterlegen sondern spaltenweise. Der Vorteil ist, dass es bei der Selektion durch eine Anfrage an die Tabelle der Datenbank möglich ist, nur die Spalten zu laden welche wirklich benötigt [Losh09, p.2]. Zusätzlich bietet das Spalten-basierte Speichern die Möglichkeit einer besseren Kompression. Stavros Harizopoulos aus den HP Labs spricht hier von einer Kompression von 1:10 gegenüber 1:3 bei Zeilen-basierten Systemen [HaAB09]. Durch eine stärkere Kompression können mehr Daten in den Hauptspeicher geladen werden.

2. Motivation

Yellowfin, ein Hersteller von BI-Software, stellt fest: „Whilst this latency is generally acceptable for trend analysis and forecasting, traditional data warehouses [Data-Warehouses mit Datenspeicherung auf Festplatten] simply can't keep pace with today's BI requirements for fast and accurate data.“ [Yell10, p.6]. Es ist also anzunehmen, dass In-Memory-Datenbanken für BI-Anforderungen Geschwindigkeitsvorteile gegenüber anderen Methoden mit sich bringen. Große Software-Hersteller, wie z.B. SAP (HANA) und Microsoft (SQL Server 2014), werben mit ihren neu entwickelten In-Memory-Datenbanken [Dela14, Sapa11]. Laut den Analysten von Gartner, befindet sich die In-Memory-Database-Technologie auf der Spitze ihres Hypes [Gart13]. Vor diesem Hintergrund soll untersucht werden, in welchen Maßen diese neuen Technologien in einer speziellen Anwendung tatsächlich Vorteile bringen. Im Abschnitt 5 wird die spezielle Anwendung näher erläutert.

3. Forschungsstand

In-Memory-Datenbanken sind keine neue Technologie. Bereits 1984 schrieben Forscher der Universität Wisconsin über Techniken für Hauptspeicher residente Datenbanken [DKOS84]. Die Technologie wurde jedoch erst in jüngster Zeit wiederentdeckt, da Hauptspeicher immer günstiger und durch die 64-Bit Technologie in größerer Menge eingesetzt werden.
Das Spalten-basierte Speichern von Daten ist ebenfalls nicht neu, Sybase kommerzialisierte diese Technologie bereits 1990 und schon 2005 wurde durch die Universitätsprojekte C-Store (MIT) und MonetDB (CWI) ein Hype um Spalten-basiertes Speichern losgelöst [HaAB09].

Allerdings findet sich bis heute nur eine begrenzte Anzahl von Produkten, welche beide Technologien verbinden[1]:

Kommerzielle Produkte

- 1010datas Tenbase database
- EXASOL EXASolution
- IBM DB2 mit BLU Acceleration
- KDB
- MemSql (in der unveröffentlichten Version 3)
- Microsoft SQL Server 2014
- Oracle 12c Enterprise Edition mit In-Memory Option
- SAP HANA
- SAP Sybase IQ
- Vertica (C-Store kommerzialisiert)

Open-source Produkte

- Apache Cassandra (in DataStax)
- Apache Gora
- C-Store
- MonetDB

4. Ziel der Arbeit

In dieser Arbeit wird eine Auswahl der verfügbaren Spalten-basierten In-Memory-Datenbanken verglichen und bewertet. Daten und Workload für diesen Vergleichen werden von einem mittelständischen Betrieb übernommen, genauso wie dessen Ausführungsumgebung. Eine nähere Erläuterung des Betriebs und der Daten findet in Abschnitt 5 statt. Die spezifischen Daten und die Workload des Betriebs werden allerdings Auswirkungen auf die Abfragezeiten der einzelnen Datenbanken haben. Um diese

[1] Verwendet wurde die List von http://de.wikipedia.org/wiki/Spaltenorientierte_Datenbank und die Dokumentationen der jeweiligen Hersteller.

Auswirkungen zu extrahieren und so auch eine allgemeinere Aussage über die Software-Produkte treffen zu können, werden die Datenbanken in dieser Ausführungsumgebung zusätzlich mit synthetischen Daten und Workloads aus einem Benchmark verglichen. Um nun der Frage nachzugehen, ob diese Technologien einen Performanceschub bringen, werden alle Tests ebenfalls mit MOLAP-Servern durchgeführt, welche etabliert sind und so bei diesem oder ähnlichen Betrieben eingesetzt werden. Sollte die Bewertung der neuen Technologien positiv ausfallen, wäre dies ein Indikator dafür, dass das untersuchte mittelständische Unternehmen umdenken und ihr BI-System updaten sollte.

5. Testfall

Durch eine Kooperation mit der Firma Kantiko[2], einem Consulting Unternehmen für Planungssoftware, entsteht die Möglichkeit, In-Memory-Datenbanken mit realen Daten eines Kunden zu befüllen und die Datenbanken analog zu Kolodziej auf real eingesetzter Hardware testen zu können. Veröffentlichte Tests von Datenbankherstellern werden dagegen häufig mit synthetischen Daten, wie dem TPC-H, auf sehr potenter Hardware durchgeführt, um die Ergebnisse zu optimieren.
Der Kunde ist ein österreichisches Bauunternehmen mit Zimmereibetrieb, Bautischlerei und Bauschlosserei. Die Überwachung des Unternehmens wird durch BI unterstützt. Hierfür wird der Umsatz von z.B. einzelnen Baustellen, Bauleitern und Handwerkern verglichen. Die Geschäftsleitung hat tagesaktuell Zugriff auf Produktivitätsverläufe und ABC-Analysen, Einteilungen nach besten und schlechtesten Elementen. Die Geschäftsleitung und die Abteilungsleiter erhalten vom Controlling regelmäßig automatisch erzeugte und ihrem Funktionsbereich entsprechende Standardberichte. Auch Projektleiter haben Zugriff, jedoch dürfen diese nur sie betreffende Daten einsehen. Ein wichtiger Punkt für den Betrieb sind nicht abgerechnete Leistungen, also Leistungen, welche erbracht wurden, jedoch noch nicht entlohnt sind. Die Verantwortlichen können durch dieses System erkennen, wo die Schwachstellen liegen. Planung auf Basis von OLAP-Daten wird bei diesem Kunden nicht durchgeführt.

Datenstruktur des Kunden

Tabelle	Datensätze	Speicherverbrauch (KB)

[2] http://www.kantiko.com/

Leistungsart	591	80
Abteilung	12	8
Auftragsstatus	8	8
Mitarbeiter	139	16
Interne Kostenstelle Fein	97857	13408
Interne Kostenstelle Kategorien	8	8
Ist, Plan, Ziel	3	8
Buchungsstand	3	8
Typ Berechnete Daten	11	8
Kostenstellen-Mapping	64513	1088
Faktentabelle	76.903.169	4.172.608
Tagesweise Zuordnung (seit 2000)	5.818	544

Workload (exemplarisch)

- Leistung pro Monat pro Jahr pro Abteilung und Abweichung des aktuellen Monats zum Vorjahresmonat
- Leistung pro Monat pro Jahr pro Abteilung aufgespalten nach internen/externen Leistungen
- Leistung aufgespalten für einen Monat nach Kostenstellen
- Leistungen nach Kostenstellen im Jahresvergleich (Kategorien / fein)
- Leistungen nach Abteilung und Kostenstellen
- Nicht abgerechnete Leistungen pro Jahr pro Monat für eine Abteilung
- Top Aufträge mit nicht abgerechneten Leistungen
- ABC-Analyse der nicht abgerechnete Leistungen Aufträge / Verantwortlichen

Hardware der Testumgebung (Virtueller Server mit VMware)

- CPU: Intel Xeon E5-2620 @2,00GHz-2,40GHz (4 Prozessoren)
- RAM: 16 GB
- OS: Microsoft Windows Server 2008 R2 (64-Bit)
- HDD: 160 GB

6. Related Work

Es gibt Vergleiche von Datenbanken, die häufig von Herstellern betrieben werden, um ihr Produkt an anderen zu messen [Alti00, H00]. Eine vergleichbare Arbeit zu dieser findet sich in „Entwicklung eines Benchmarks für die Performance von In-Memory OLAP-Systemen“ von Christian Kolodziej [Kolo10]. Kolodziej vergleicht mehrere MOLAP-Server, Jedox Palo, IBM Cognos TM1 und Infor PM OLAP Server. Er analysiert hierfür Datenmodelle und Workloads von mehreren Kunden. Die Datenmodelle führt er in einem allgemeinen Sternschema zusammen, ein Modell, bei dem die Dimensionen nicht mehrstufig sind, also nicht mit weiteren Dimensionen verknüpft sind. Er generalisiert eine Workload für die Umsatzanalyse von Unternehmen und eine Workload für die Planung. Allerdings generiert Kolodziej für jedes Datenmodell die Daten nur synthetisch durch Selektion aus dem TPC-H, einem Datawarehouse-Benchmark. In dieser Arbeit sollen zusätzlich zu einem Benchmark auch das Verhalten mit realen Daten untersucht werden. Die eingesetzte Hardware ist ähnlich zu der, zu welcher in dieser Arbeit Zugriff besteht. Da die Arbeit von Christian Kolodziej in einer ähnlichen Umgebung und mit zu dieser Arbeit ähnlicher Workload durchgeführt wurde, kann sich an der Durchführung von Kolodziejs Arbeit orientiert werden. Im Abschnitt 5 gehen wir auf die eingesetzte Workload genauer ein. Es wäre möglich den gleichen Benchmark zu verwenden, der bereits von Kolodziejs verwendet wurde und so einen Vergleich der Abfragezeiten von Jedox Palo zu ermöglichen. Allerdings müsste für einen solchen Vergleich die komplette Test-Umgebung gleich sein, was nicht der Fall ist. Deshalb wird auf einen solchen Vergleich verzichtet und statt dem TPC-H Benchmark der, auf eher auf OLAP ausgerichtete, Star Schema Benchmark verwendet. Mehr zur Auswahl des Benchmarks befindet sich in Abschnitt 7.3.

7. Vorgehensweise

7.1. Auswahl Spalten-basierter In-Memory-Datenbanken

Die Datenbanksysteme aus Abschnitt 3 können nicht alle betrachtet werden. SAP HANA, EXASOL EXASolution und 1010datas Tenbase database bieten keine kostenlose Testversion an. MemSql, wie bereits erwähnt, hat die in Frage kommende Version 3 ihrer Software noch nicht veröffentlicht. Die Testversion von KDB läuft nur in einer abgespeckten 32-Bit Version. Dies ist zu wenig für die 16 GB RAM der Testumgebung. Im Falle von Vertica/C-Store würde die kommerzielle Variante verwendet werden, da C-Store nicht aktiv weiter entwickelt wird.

Die letzte Version 0.2 wurde 2006 veröffentlicht. Apache Gora ist theoretisch Windows kompatibel, allerdings wird es von den Herstellern auf dieser Plattform nicht getestet. Da das Testsystem ein Windows Betriebssystem hat, gilt es zu eruieren, ob Apache Gora eingesetzt werden kann. Alle anderen Systeme können verglichen werden. Um den Rahmen dieser Arbeit nicht zu sprengen, werden 2 kommerzielle und 2 Open-Source Produkte verglichen. Diese werden nach dem Popularitäts-Ranking von „DB-Engines“ ausgewählt[3].

Kommerzielle Produkte

- Microsoft SQL Server 2014
- Oracle 12c Enterprise Edition mit In-Memory Option

Open-source Produkte

- Apache Cassandra (in DataStax)
- MonetDB

7.2. Auswahl der MOLAP-Server

Als MOLAP-Server verwendet die Kantiko das bereits in Kolodziejs Arbeit verwendete Jedox Palo, sowie Prevero P7 und Microsoft SQL Server Analysis Services (MS SSAS). Diese Auswahl ist weder vollständig noch repräsentativ, spiegelt aber Software wieder, welche so tatsächlich bei mittelständischen Unternehmen im Einsatz sind. Aus dieser Selektion werden noch einmal, wie schon bei den In-Memory-Datenbanken, um den Umfang nicht zu groß zu gestalten, nur 2 Lösungen für die Analyse ausgewählt: Jedox Palo und MS SSAS in der Version 2008 RC2 um Überschneidungen bei der Technologie mit der neuen Version MS SQL Server 2014 zu vermeiden.

Die eingesetzten Datenbanksysteme und MOLAP-Server werden mit Werkseinstellungen verwendet. Christian Kolodziej erwähnt auch in seiner Arbeit, dass es schwer wäre, sollte eine Leistungs-Optimierung stattfinden, alle Systeme gleich zu behandeln.

7.3. Benchmark

Als Benchmark dient der Star Schema Benchmark (SSB), ein Benchmark für Datawarehouses basierend auf dem Standard-Benchmark für Datawarehouses, TPC-H [OnOC09].

[3] http://db-engines.com/en/ranking

7.4. Durchführung

Christian Kolodziej verwendet bei der Nebeneinanderstellung der OLAP-Systeme für das Einlesen der Daten in die Systeme und das Auswerten der OLAP-Anfragen die jeweils eigenen Schnittstellen der Software [Kolo10]. Dies ist sinnvoll, da die proprietären Schnittstellen auf das System abgestimmt sind. Jedoch verfügen nicht alle Systeme über dedizierte ETL-Werkzeuge. Um zu vermeiden, dass Ausnahmen beim ETL der verwendeten Systeme individuell behandelt werden müssen, wird ein freies Open-Source-Werkzeug verwendet: Pentaho Data Integration[4]. Dieses Werkzeug ermöglicht es die Datenintegration in alle Analyse-Systeme, welche in dieser Arbeit betrachtet werden, durch ein einzelnes Interface zu steuern und zu überwachen. Pentaho Data Integration kommuniziert dabei mit den Datenbanken über ihre vom Hersteller bereit gestellten nativen Datenbanktreiber. Dennoch wird die Leistung des Imports über dieses Werkzeug voraussichtlich schlechter sein, als mit proprietären Mitteln. Sollte sich ergeben, dass die Leistung durch Verwendung dieses systemfremden Werkzeugs so stark verliert, dass kein effizientes Arbeiten mehr möglich ist, muss doch analog zu Kolodziej gearbeitet werden.

Pentaho Data Integration ermöglicht es eine komplette Test-Orchestrierung mit allen Testsystemen und Datenquellen aufzusetzen, welche überschaubar, wartbar und replizierbar ist. In Abbildung 1 ist eine solche Orchestration dargestellt, wobei im oberen Teil der tatsächliche Verlauf abgebildet ist und im unteren Teil die Orchestrationen der einzelnen Teilschritte.

Analog zum Dateneinlesen gibt es Komplikationen beim Auswerten der Anfragedauer für die unterschiedlichen Systeme. Kolodziej verwendet für Infor PM sowie für Palo die Methode, die Anfragen über einen Proxy-Server zu leiten und hier die Zeiten zu überprüfen. Allerdings war dies für Cognos nicht möglich. Als Ausweg, wurde das Browser-basierte Werkzeug Cognos (Cube) Viewer verwendet und die Anfragezeiten über Firebug ausgewertet. Auch hier soll Pentaho Data Integration eine gleiche Ausführung für alle Systeme ermöglichen. Das Werkzeug bietet ein Logging[5], über das die Auswertungen festgehalten werden können. Auch im Fall der Auswertungen sei wieder angemerkt, dass eine Http-Request-Methode verwendet wird, wie Kolodziej sie eingesetzt, sollte diese Methode nicht effizient durchführbar sein.

[4] http://community.pentaho.com/projects/data-integration/
[5] http://wiki.pentaho.com/display/EAI/PDI+Logging

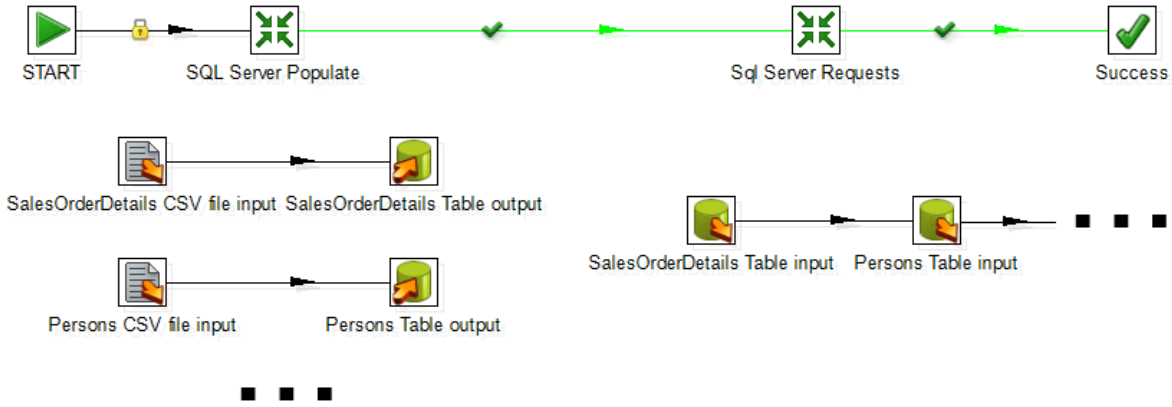

Abbildung 1: Orchestration des Tests

Es soll ermittelt werden, wie sich die Systeme bei unterschiedlichen Anfragen und Größe der Daten verhalten.

Die Workload des Kunden aus Abschnitt 6 beinhaltet bereits Anfragen mit unterschiedlicher Selektivität. Die Selektivität wird als Maßstab für die Komplexität der Anfrage dienen. Die Berechnung der Selektivität erfolgt in der eigentlichen Arbeit.

Die Workload des SSB ist bereits für Tests ausgelegt. Zusätzlich lassen sich hier unterschiedliche Datenbankgrößen erzeugen. Die Autoren des SSB, liefern ein Werkzeug mit, welches die Daten je nach Größe generiert.

8. Literatur

[Alti00] ALTIBASE: *Compare Altibase In-Memory Database vs. SAP HANA*. URL http://altibase.com/in-memory-database-comparisons/sap-hana/. - abgerufen am 2014-06-30

[BaGü04] BAUER, ANDREAS ; GÜNZEL, HOLGER: *Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung*. 2. ed. Heidelberg : dpunkt, 2004 — ISBN 3898645401

[Brow04] BROWN, ROBERT G.: Bottlenecks. In: *Engineering a Beowulf-style Compute Cluster*. Durham : Selbstverlag, 2004

[ChDN11] CHAUDHURI, SURAJIT ; DAYAL, UMESHWAR ; NARASAYYA, VIVEK: An overview of business intelligence technology. In: *Communications of the ACM* vol. 54, ACM (2011), Nr. 8, p. 88

[Dela14] DELANEY, KALEN: *SQL Server In-Memory OLTP Internals Overview. SQL Server Technical Article*, 2014

[DKOS84] DEWITT, DAVID J ; KATZ, RANDY H ; OLKEN, FRANK ; SHAPIRO, LEONARD D ; STONEBRAKER, MICHAEL R ; WOOD, DAVID: Implementation techniques for main memory database systems. In: *ACM SIGMOD Record* vol. 14, ACM (1984), Nr. 2, p. 1 — ISBN 0-89791-128-8

[Gart13] GARTNER: *Gartner's 2013 Hype Cycle for Emerging Technologies Maps Out Evolving Relationship Between Humans and Machines*. URL http://www.gartner.com/newsroom/id/2575515. - abgerufen am 2014-06-13

[GlCh06] GLUCHOWSKI, PETER ; CHAMONI, PETER: Analytische Informationssysteme. In: CHAMONI, P. ; GLUCHOWSKI, P. (eds.): *Analytische Informationssysteme. Business Intelligence-Technologien und –Anwendungen*. 3. ed. Berlin, Heidelberg : Springer Berlin Heidelberg, 2006 — ISBN 978-3-540-29286-9, pp. 143–176

[H00] H2: *Performance*. URL http://www.h2database.com/html/performance.html. - abgerufen am 2014-06-30

[HaAB09] HARIZOPOULOS, STAVROS ; ABADI, DANIEL ; BONCZ, PETER: Column-Oriented Database Systems. In: *VLDB'09*. Lyon, Frankreich, 2009

[Humm14] HUMMELTENBERG, WILHELM: *Business Intelligence*. URL http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/daten-wissen/Business-Intelligence. - abgerufen am 2014-07-08. — Enzyklopaedie der Wirtschaftsinformatik

[ImWh11] IMHOFF, CLAUDIA ; WHITE, COLIN: *Self-Service Business Intelligence - Empowering Users to Generate Insights*. Renton, USA, 2011. — TDWI Research, SAS

[Kaha06] KAHATE, ATUL: *Introduction to Database Management Systems*. 1. ed. : Pearson Education India, 2006 — ISBN 81-317-0078-X

[KeBM04] KEMPER, HANS-GEORG ; BAARS, HENNING ; MEHANNA, WALID: *Business Intelligence - Grundlagen und praktische Anwendungen - Eine Einführung in die IT-basierte*. Wiesbaden : Vieweg+Teubner Verlag, 2004

[Kolo10] KOLODZIEJ, CHRISTIAN: *Entwicklung eines Benchmarks für die Performance von In-Memory OLAP-Systemen*, Hochschule Karlsruhe, 2010

[Losh09] LOSHIN, DAVID: *Gaining the Performance Edge Using a Column-Oriented Database Management System*. Dublin, USA, 2009

[Oehl06] OEHLER, CARSTEN: *Corporate Performance Management mit Business Intelligence Werkzeugen*. Wien : Hanser, 2006

[OnOC09] O'NEIL, PAT ; O'NEIL, BETTY ; CHEN, XUEDONG: *Star Schema Benchmark*, University of Massachusetts Boston, 2009

[Sapa11] SAP AG: *SAP HANA™ for Next-Generation Business Applications and Real-Time Analytics Explore and Analyze Vast Quantities of Data from Virtually Any Source at the Speed of Thought*, 2011

[Sche10] SCHEUERMANN, BERND: Design of a Reconfigurable Hybrid Database System. In: *2010 18th IEEE Annual International Symposium on Field-Programmable Custom Computing Machines* : IEEE, 2010 — ISBN 978-1-4244-7142-3, pp. 247–250

[Yell10] YELLOWFIN INTERNATIONAL PTY LTD: *In-Memory Analytics*, 2010