Entwicklung eines Benchmarks für die Performance von In-Memory OLAP-Systemen

Zu meiner Person

- Christian Kolodziej
- Informatik- und Wirtschaftsinformatik-Studium an der HS Karlsruhe
- seit 1.1.11 BI-Consultant bei der Infomotion GmbH, Frankfurt/Main

Zu meiner Person

- Christian Kolodziej
- Informatik- und Wirtschaftsinformatik-Studium an der HS Karlsruhe
- seit 1.1.11 BI-Consultant bei der Infomotion GmbH, Frankfurt/Main

OSBI Blog
Open Source und Business Intelligence

Entwicklung eines Benchmarks für die Performance von In-Memory OLAP-Systemen

Entwicklung eines Benchmarks für die Performance von In-Memory OLAP-Systemen

OLAP-Systeme

- OLAP-Systeme definieren sich über die FASMI-Definition
- multidimensionale Speicherung

	OLTP	***OLAP***
Daten-Eigenschaften	nicht abgeleitet, zeitaktuell, autonom, dynamisch	abgeleitet, konsolidiert, historisiert, integriert, stabil
Datenvolumen einer Abfrage	wenige Datensätze	viele Datensätze
Datenvolumen	Megabyte - Gigabyte	Gigabyte - Terabyte

Quelle: BG09, Seite 10f

Entwicklung eines Benchmarks für die Performance von **In-Memory** OLAP-Systemen

In-Memory-Technologie

- Preisverfall und 64-Bit-Systeme sorgen für neuerliche Nutzung
- schnellerer Datenzugriff durch Datenhaltung im Arbeitsspeicher
- geringere Anforderungen an Infrastruktur
- *Wichtig:* Ausreichend großer Arbeitsspeicher vorhanden

Entwicklung eines Benchmarks für die Performance von In-Memory OLAP-Systemen

Performance

- Performance in doppelter Hinsicht wichtig
  - fördert Akzeptanz durch Benutzer
  - Skalierbarkeit hängt eng mit der Performance zusammen
- Datenmengen steigen weiter an
- Performance alleine kann BI-Projekte zum Scheitern bringen

Entwicklung eines **Benchmarks** für die Performance von In-Memory OLAP-Systemen

APB-1

- Ziel: allgemeine Aussage über die Performance eines OLAP-Servers
- Kritikpunkte
  - nicht alle Zeiten wurden berücksichtigt
  - umfangreicher Audit-Report
  - Datenmodell nicht repräsentativ
  - Lediglich relativ einfache Abfragen

TCP-H

- Ziele: Alltagsprobleme der Nutzer aufzeigen
- Probleme
  - zu große Beeinflussung durch Hersteller
  - überdimensionierte Test-Systeme fernab jeglicher Praxisrelevanz
- Fazit
  - Ergebnisse ohne Nutzen für den Endanwender
  - Hersteller nutzen TCP-H, um eigene Datenbank zu optimieren

Rückschlüsse für die eigene Arbeit

- Ladegeschwindigkeit berücksichtigen
- repräsentatives Szenario
- Planungskomponente
- Trennung der einzelnen Teilbereiche
- Abstraktion von der Hardware

Entwicklung eines Benchmarks für die Performance von In-Memory OLAP-Systemen

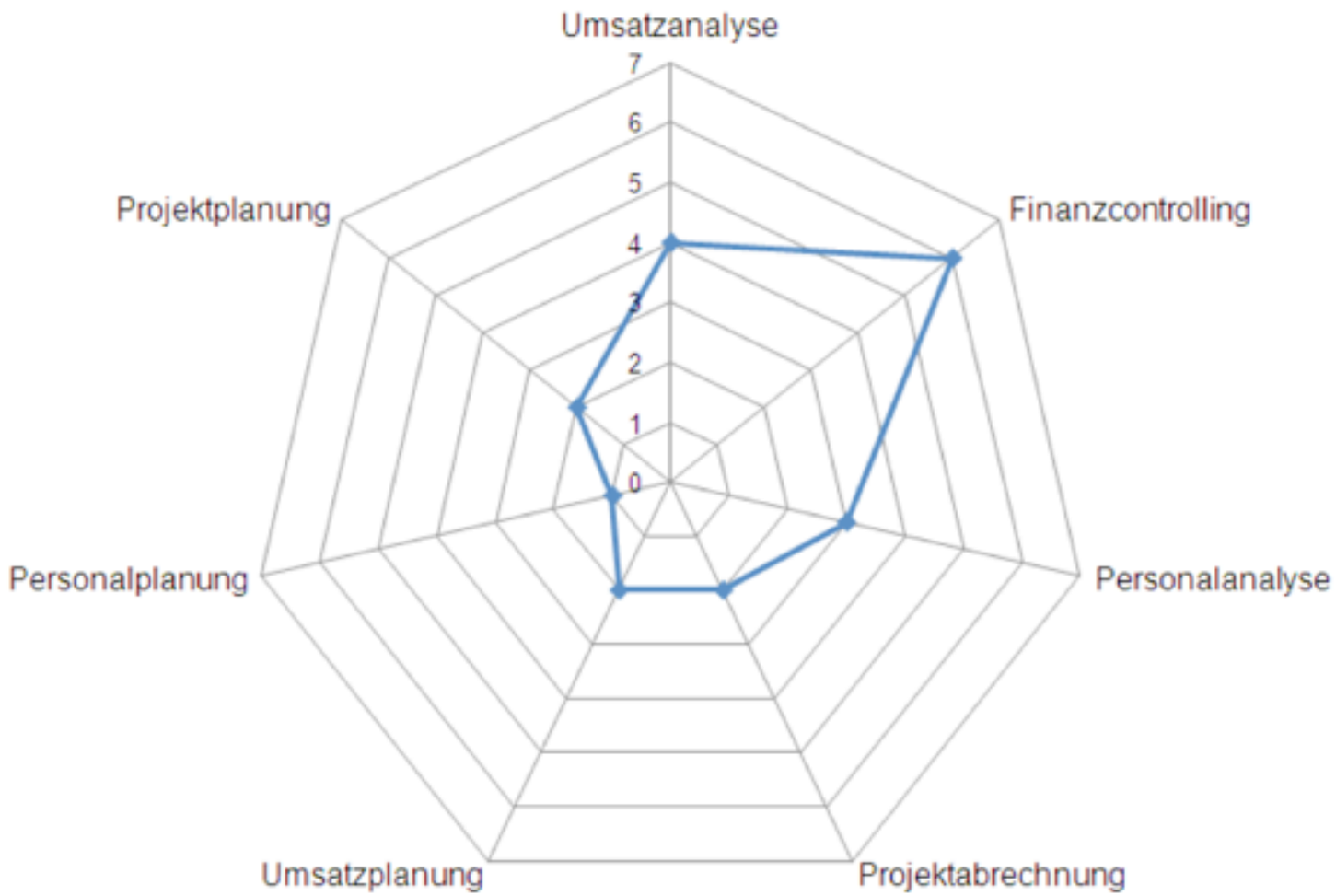
Auswahl der OLAP-Server

- Kriterien
  - Nutzung der In-Memory-Technologie
  - Planungsfunktion bereits enthalten („Splashing“)
- Testkandidaten
  - Jedox Palo (Entwicklungsversion zum Release 3.2 vom 24.08.10)
  - Infor PM 10
  - IBM Cognos TM1 (Version 8)

Hardware

- CPU AMD Phenom II X4 920 (4 x 2,80 GHz)
- 16 GB Arbeitsspeicher DDR-2
- SATA-Festplatte 250 GB für Betriebssystem und OLAP-Server
- SATA-Festplatte 500 GB für Quelldaten
- 4 GPUs Nvidia Tesla C1060 (jeweils 240 Grafikkerne und 4 GB Grafikspeicher)

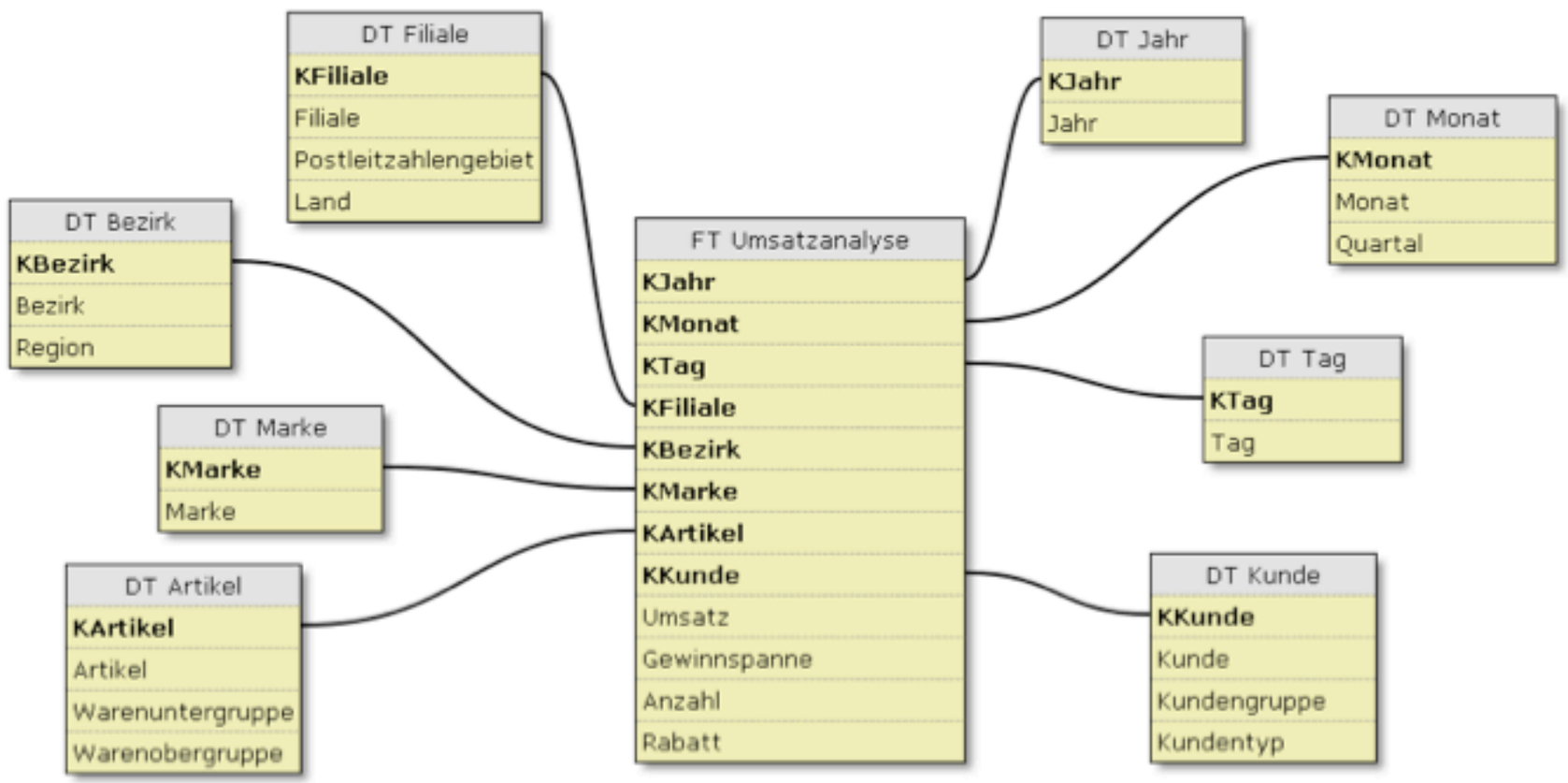
Analyse von Kundenanwendungen



- Umsatzanalyse weit verbreitet
- Planungsanwendung werden öfter

Teildisziplin Analyse - Datenmodell

- Umsatzanalyse
  - „Klassisches Beispiel“ für Business Intelligence
  - Datenmenge auch in der Praxis schnell sehr groß

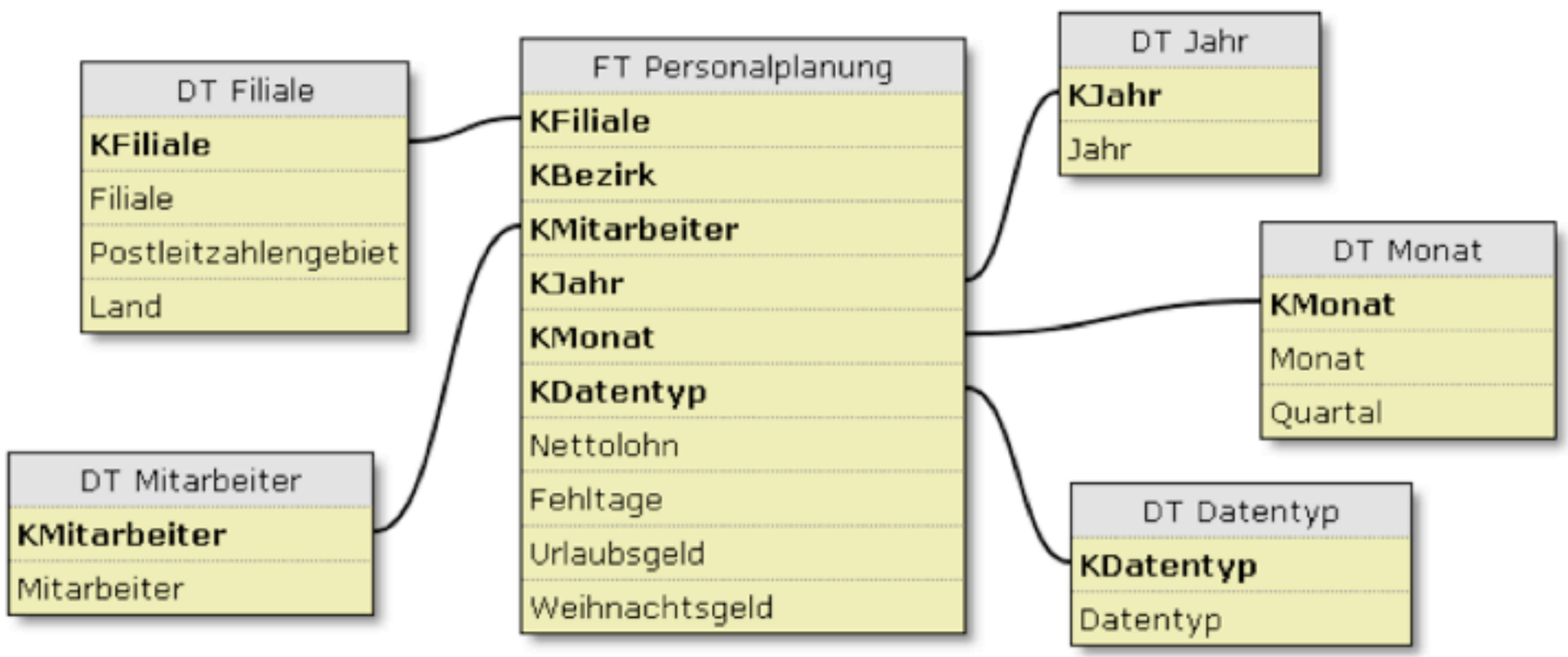


Teildisziplin Analyse - Abfrage

- Gesamtumsatz ohne jegliche Einschränkung
- Anzahl der in 1 Jahr verkauften Artikel
- Reingewinn einer Marke in einem Land
- Gesamtumsatz einer Marke in 1 Jahr in einem Land
- Summe der innerhalb eines Quartals an alle Geschäftskunden eines Landes gewährten Rabatte

Teildisziplin Planung - Datenmodell

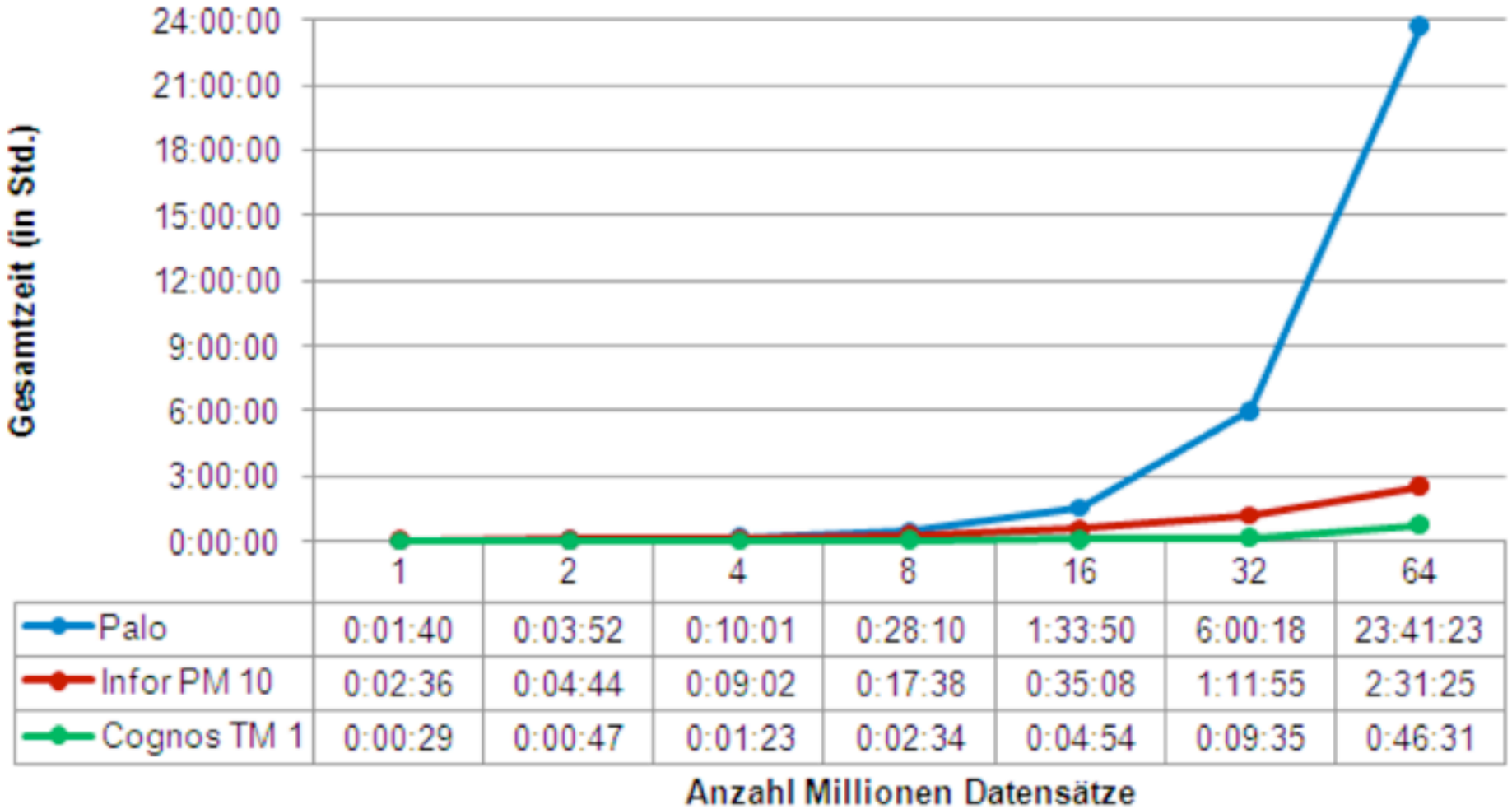
- Personalplanung
  - in jeder Firma nötig, Aufbau ähnlich
  - Datenmodell meist relativ einfach



Teildisziplin Planung - Operationen

- Nettogehälter (Ist) von 2009 als Planzahlen für 2010 kopieren
- Nettogehälter (Plan) um 10 % erhöhen
- Urlaubsgeld 2010 (Plan) mit Auszahlung im Juli 2010 beträgt pro Mitarbeiter und Filiale 100 €
- Weihnachtsgeld 2009 (Ist) mit Auszahlung im November 2009 lag pro Mitarbeiter um 100 € höher
- Fehltage 2010 (Plan) sollen 20 % unter Niveau von 2007 liegen, Gleichverteilung auf alle Mitarbeiter
- Personalkosten 2011 auf 75 % des Jahres 2009 planen

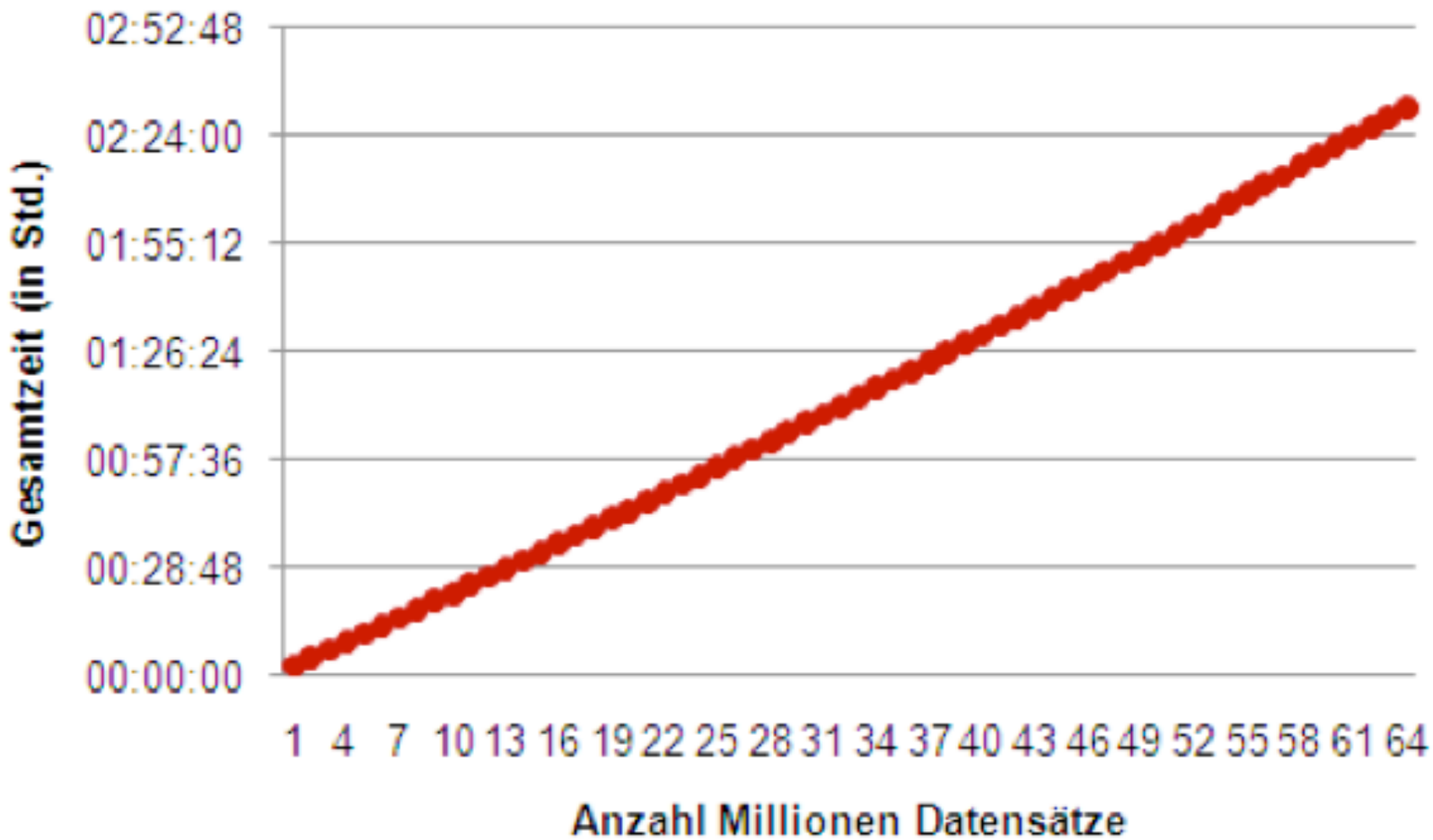
Benchmark der Datenintegration (1)



	1	2	4	8	16	32	64
Palo	0:01:40	0:03:52	0:10:01	0:28:10	1:33:50	6:00:18	23:41:23
Infor PM 10	0:02:36	0:04:44	0:09:02	0:17:38	0:35:08	1:11:55	2:31:25
Cognos TM 1	0:00:29	0:00:47	0:01:23	0:02:34	0:04:54	0:09:35	0:46:31

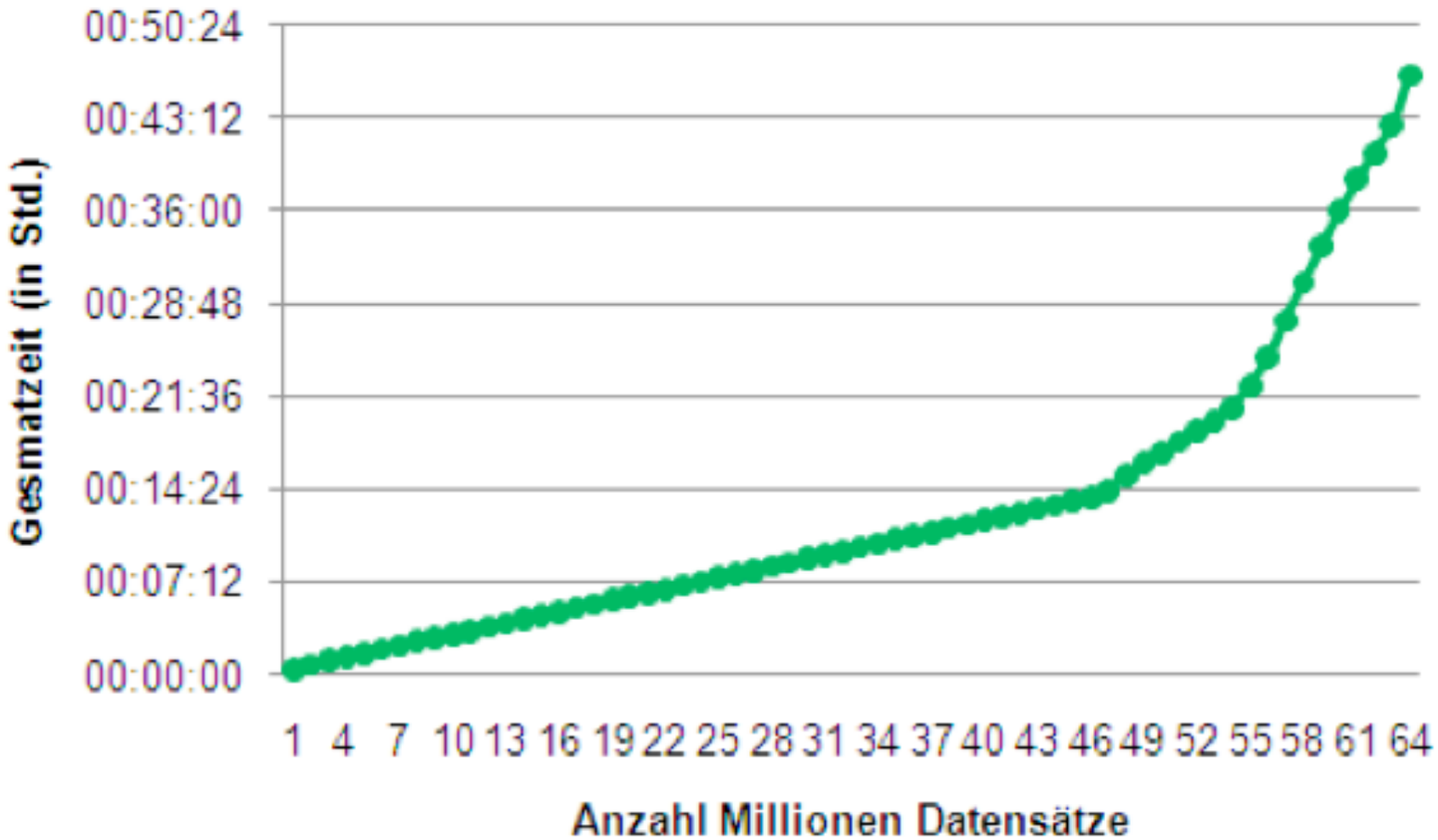
- Performance von Palo ab 8 Mio. Datensätzen deutlich schlechter
- Infor um Faktor 10 schneller, Cognos nochmals besser

Benchmark der Datenintegration (2)



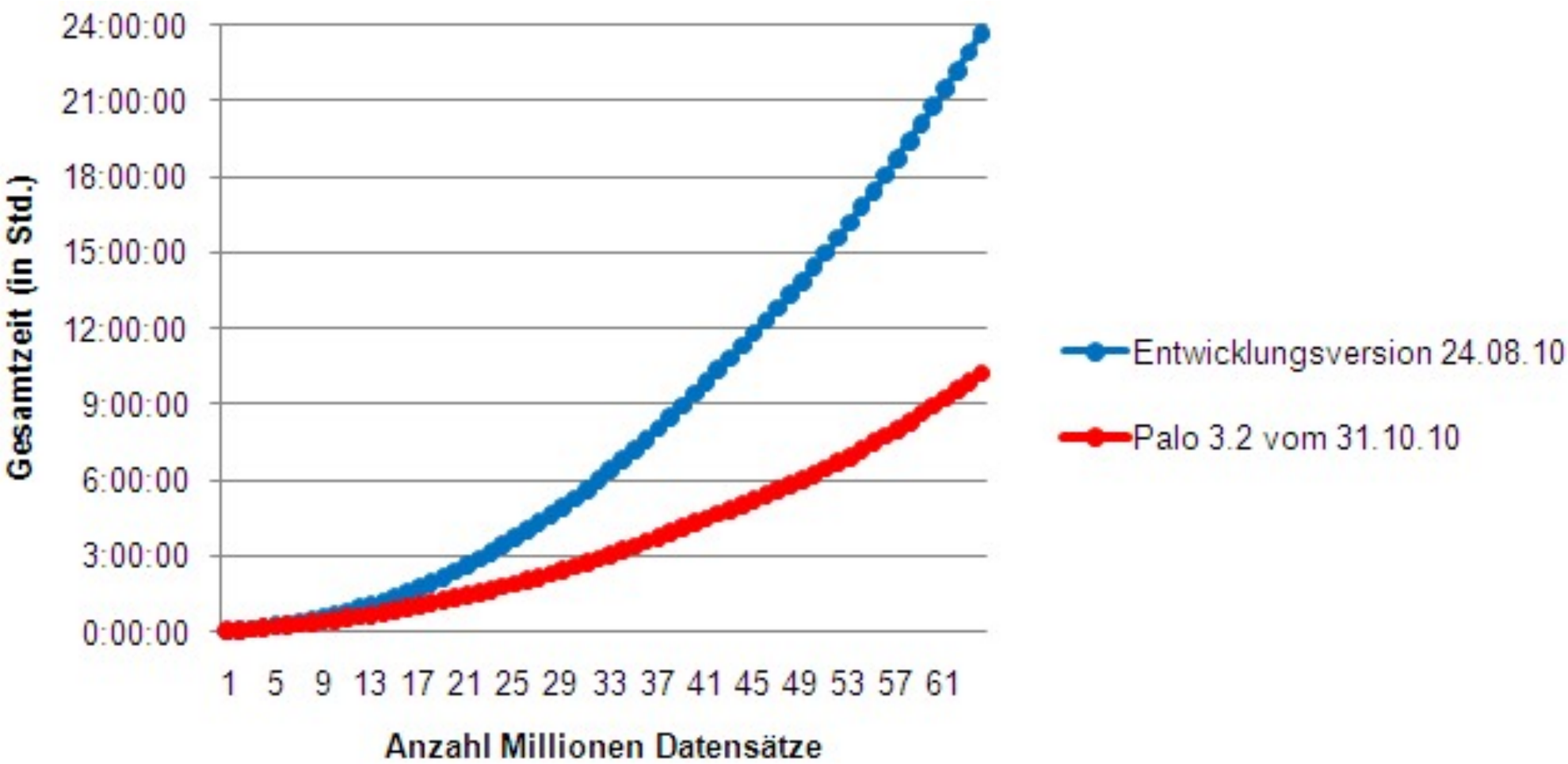
- Infor PM 10
  - lineare, berechenbare Entwicklung bis zum Ende

Benchmark der Datenintegration (3)



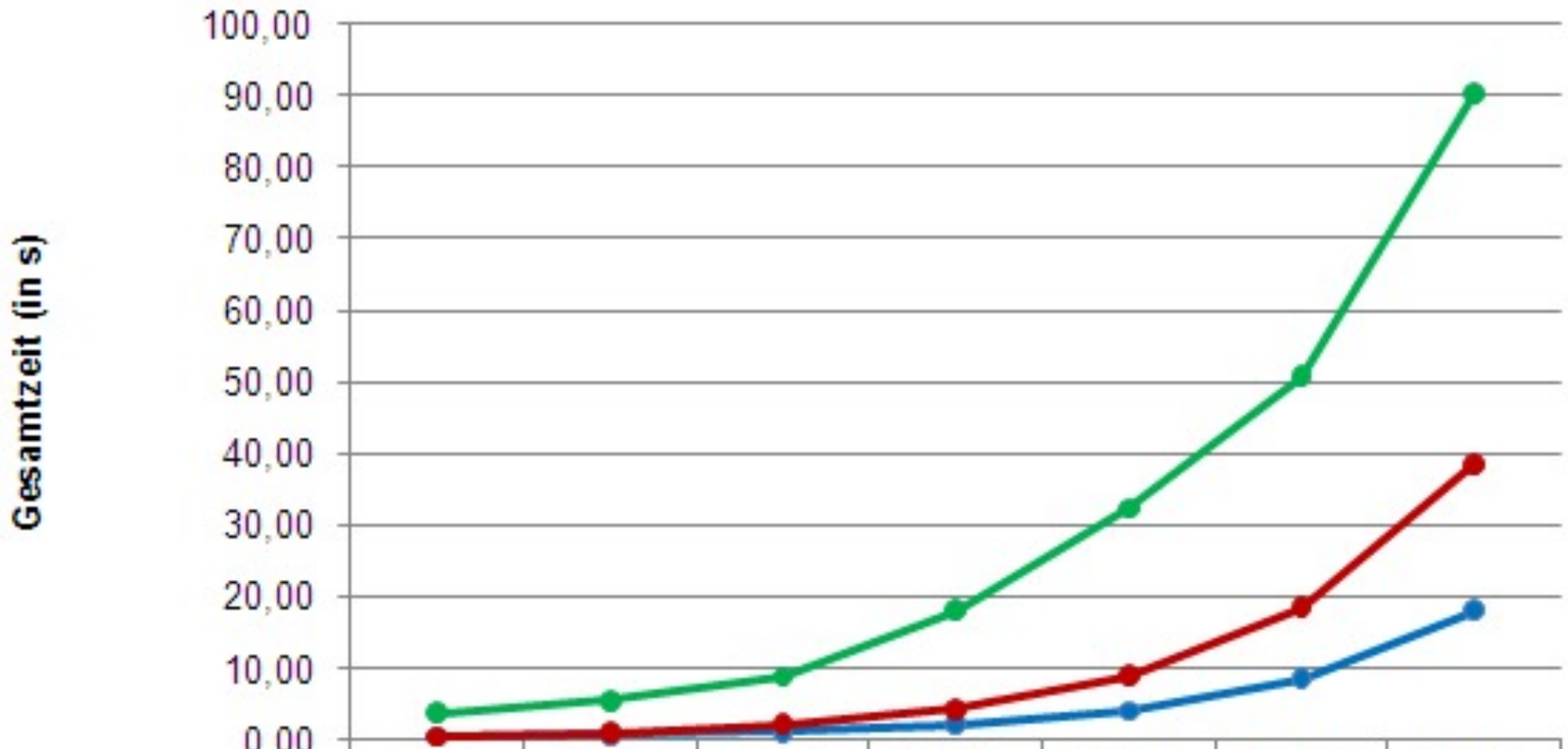
- IBM Cognos TM1
  - enormer Speicherbedarf
  - langwierige Optimierung nach Abschluss des Ladens

Benchmark der Datenintegration (4)



- Nachtest der finalen Version 3.2 am 05.11.2011
  - deutliche Performance-Verbesserung

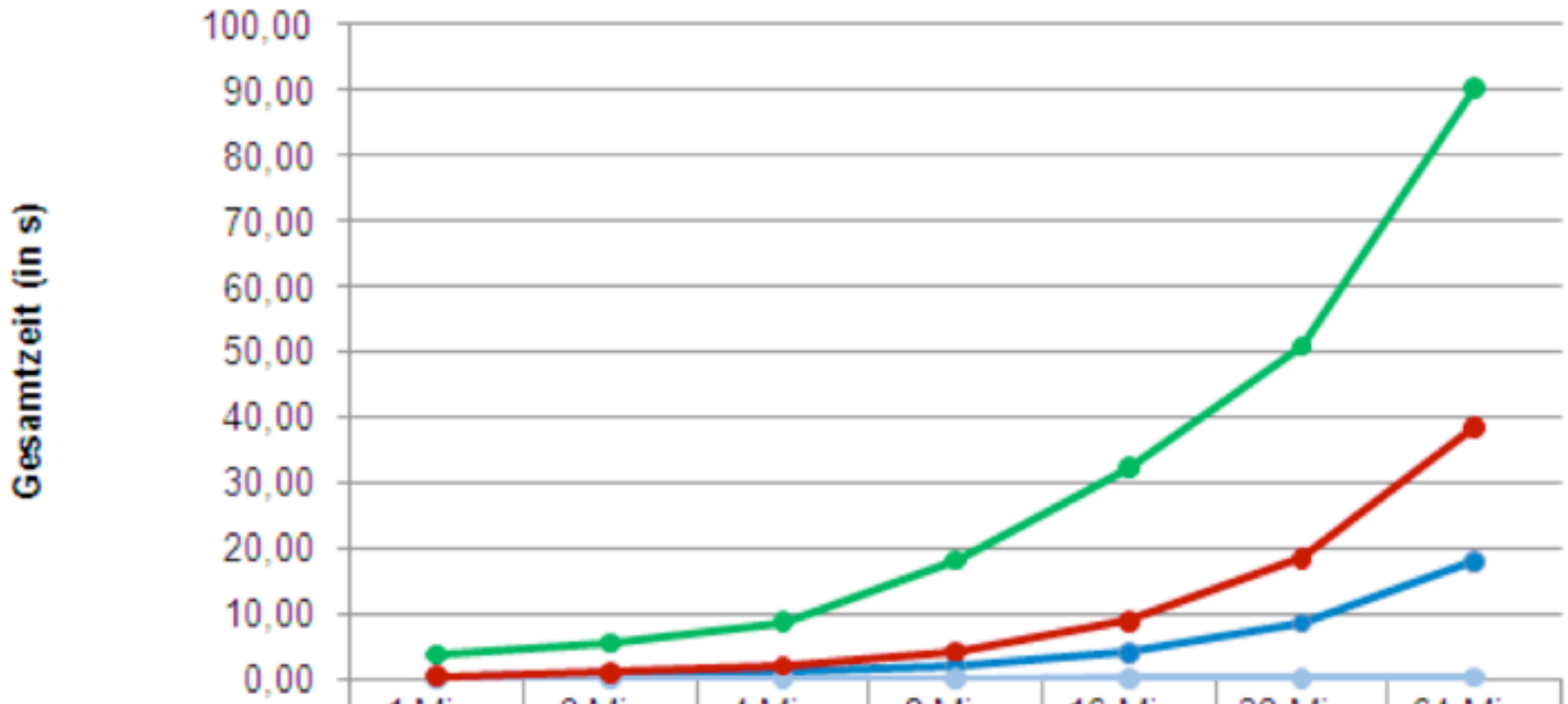
Benchmark der Analyse (1)



	1 Mio.	2 Mio.	4 Mio.	8 Mio.	16 Mio.	32 Mio.	64 Mio.
Palo	0,43	0,72	1,10	2,14	4,07	8,60	18,13
Infor PM 10	0,51	1,03	2,13	4,33	8,91	18,51	38,54
Cognos TM 1	3,82	5,55	8,83	18,17	32,27	50,80	90,34

- Palo bester OLAP Server
- Infor PM 10 um Faktor 2, IBM Cognos TM1 um Faktor 5 langsamer

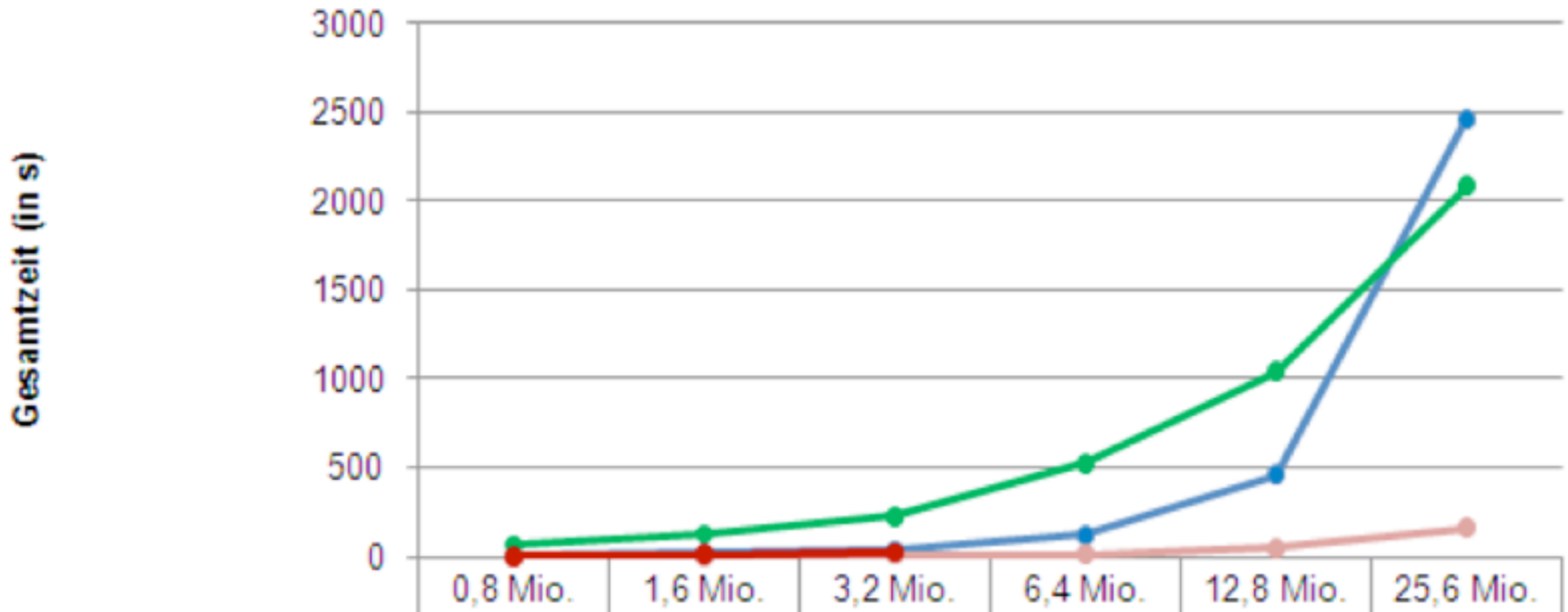
Benchmark der Analyse (2)



	1 Mio.	2 Mio.	4 Mio.	8 Mio.	16 Mio.	32 Mio.	64 Mio.
Palo	0,43	0,72	1,10	2,14	4,07	8,60	18,13
Palo GPU	0,10	0,12	0,13	0,14	0,17	0,22	0,36
Infor PM 10	0,51	1,03	2,13	4,33	8,91	18,51	38,54
Cognos TM 1	3,82	5,55	8,83	18,17	32,27	50,80	90,34

- Palo mit Unterstützung der 4 GPUs bis zu 50 Mal schneller als ohne GPUs

Benchmark der Planung



	0,8 Mio.	1,6 Mio.	3,2 Mio.	6,4 Mio.	12,8 Mio.	25,6 Mio.
Palo	7	15	39	122	461	2462
Infor PM 10	4	10	23			
- ohne Operation 5	1	2	5	14	47	162
IBM Cognos TM 1	65	127	224	524	1039	2085

- Palo und IBM Cognos TM1 tauschen Positionen bei 25,6 Mio. Datensätzen
- Infor PM 10 versagt bei Planungsoperation 5

3 Disziplinen, 3 Sieger

- Teildisziplinen nicht miteinander vergleichbar
- Anforderungen müssen klar definiert sein, um die richtige Entscheidung treffen zu können

Fazit (1)

- Benchmark-Ergenisse mit leichten Einschränkungen zu betrachten
  - Systeme zu unterschiedlich in Architektur und Anforderungen
  - zahlreiche Workarounds nötig
- Benchmark liefert erste Anhaltspunkte, ersetzt aber keinen Proof-of-Concept

Fazit (2)

- In-Memory OLAP-Systeme erlauben sehr schnelle Auswertungen bei überschaubarer Datenmenge
- GPU-Technik könnte neuen Evolutionsprung in der Business Intelligence bedeuten

Fazit (3)

- Open-Source-Systeme brauchen sich als auch in Sachen Performance nicht vor der kommerziellen Konkurrent zu verstecken.

Literatur

- BG09: Andreas Bauer / Holger Günzel. Data-Warehouse-Systeme – Architektur, Entwicklung, Anwendung. dpunkt.verlag, Heidelberg, 3. überarbeitete und aktualisierte Auflage, 2009.

- computerwoche.de. Fragwürdiger Benchmark-Test für Olap-Server. http://www.computerwoche.de/heftarchiv/1996/19/1106318/, 1996. Letzte Abfrage: 21. April 2010.

- NVIDIA Corporation. CUDA in Aktion. http://www.nvidia.de/object/cuda_in_action_de.html, 2010. Letzte Abfrage: 27. Oktober 2010.

- OLAP Council. OLAP Council APB-1 OLAP Benchmark Release II. http://www.olapcouncil.org/research/APB1R2_spec.pdf, 1998. Letz- te Abfrage: 21. April 2010.

- Transaction Processing Performance Council. TPC BENCHMARKTMH (Decision Support) Standard Specification Revision 2.11.0. http://www. tpc.org/tpch/spec/tpch2.11.0.pdf, 2010. Letzte Abfrage: 5. Mai 2010.

- Barney Finucane / Melanie Mack / Carsten Bange / Patrick Keller / Steffen Bierkorn. The BI Survey 9. Business Application Research Center, Würzburg, 2010.

Fragen?

Any questions?

Domandas?

¿Preguntas?

问题?

... für die Aufmerksamkeit

OSBI Blog
Open Source und Business Intelligence

mail@christian-kolodziej.de
www.osbi-blog.de