

# In-Memory – Einfluss auf System- und Enterprise- Architektur

Hamburg, 10.7.2015, Tim Lüecke



# Eine Minute im Internet...



<http://pennystocks.la/internet-in-real-time/>

# Über mich



**Tim Luecke**

Senior Solution Architect  
Lübecker Straße 128, Hamburg

Phone: +49 40 254491 314

E-Mail: [tim.luecke@capgemini.com](mailto:tim.luecke@capgemini.com)

# Agenda

- Motivation
- In-Memory im Überblick
- Paradigmen im Wandel?



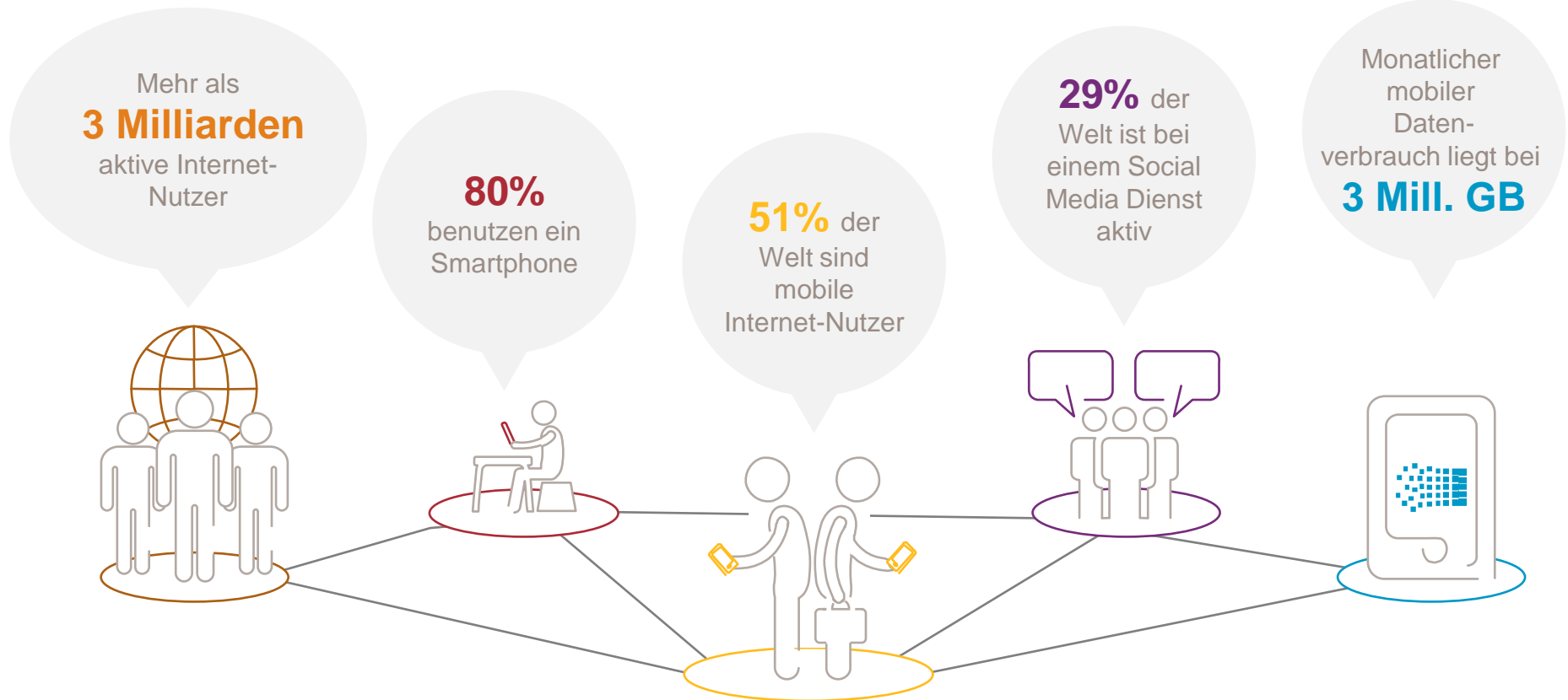
# Agenda

## ■ Motivation

- In-Memory im Überblick
- Paradigmen im Wandel?

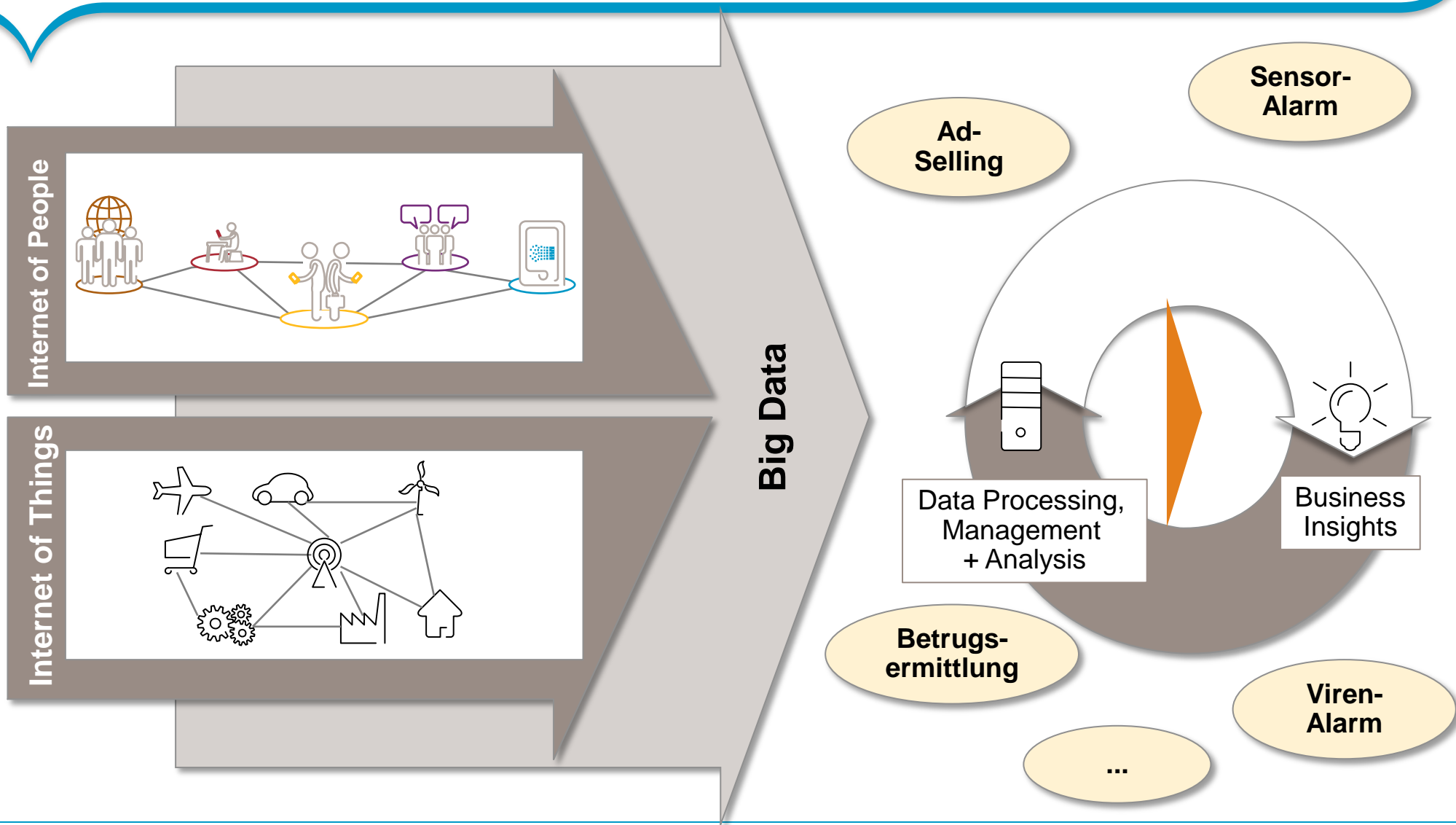


# Digitalisierung – einige Fakten



<http://wearesocial.sg/blog/2015/01/digital-social-mobile-2015/>  
<https://www.globalwebindex.net/>

# Digitalisierung und Big Data





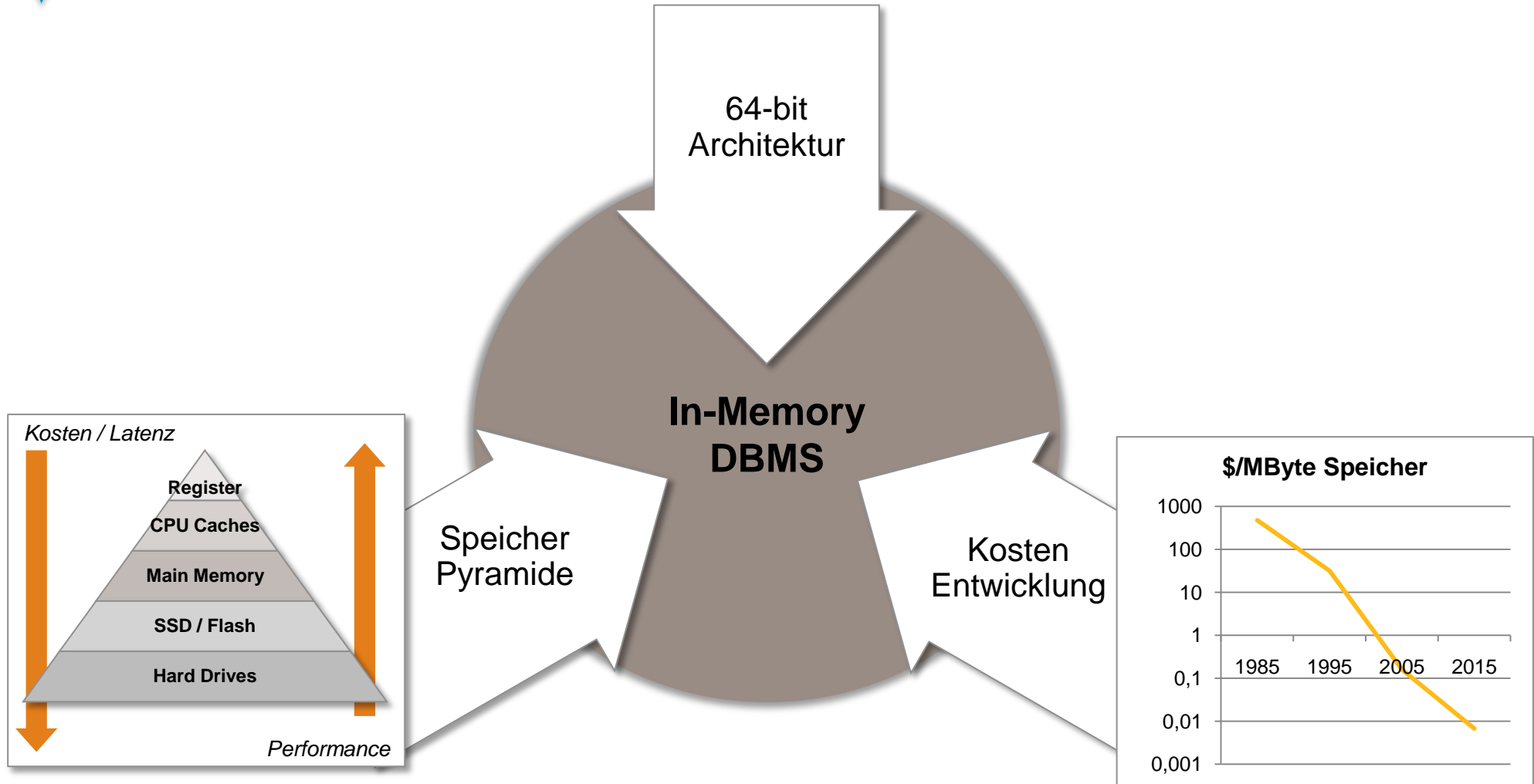
# Agenda

- Motivation
- **In-Memory im Überblick**
- Paradigmen im Wandel?

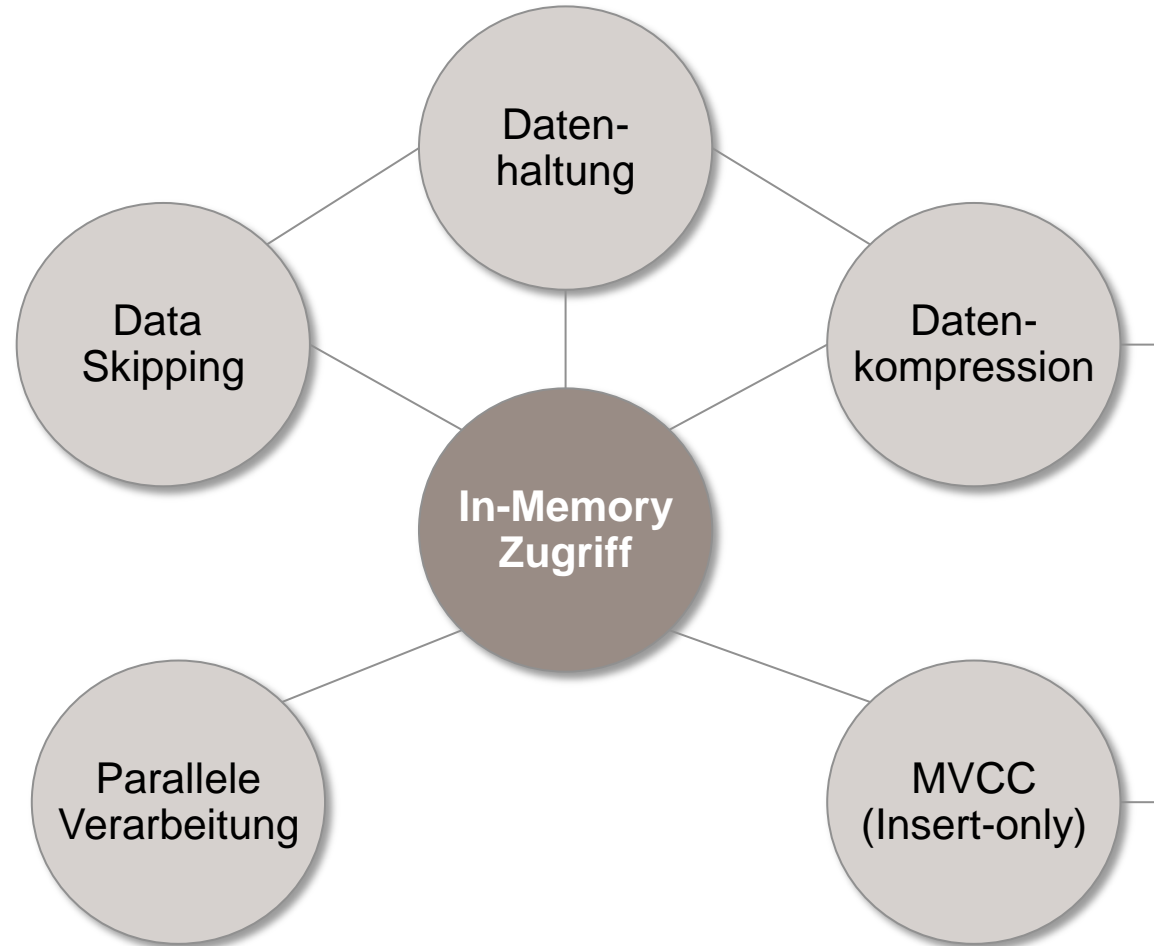




# In-Memory DBMS als mögliche Lösung



# In-Memory Techniken



# In-Memory Angebote der „Big Player“

	Microsoft SQL Server 2014	IBM DB2 BLU Acceleration	Oracle 12c w/ In-Memory
<b>Fokussierung</b>	OLTP (OLAP)	OLAP	OLAP (OLTP)
<b>Datenhaltung</b>	Zeilen (In-Memory Organized Tables) Spalten (Columnstore indices)	Spalten (Column-organized tables)	Beides simultan (Buffer Cache für Zeilen)
<b>Daten Kompression</b>	<ul style="list-style-type: none"> <li>• Dictionary encoding</li> <li>• Value encoding</li> <li>• Run-length encoding</li> <li>• Bit-packing</li> </ul>	<ul style="list-style-type: none"> <li>• Dictionary encoding</li> <li>• Frequency compression</li> <li>• Page compression</li> </ul>	<ul style="list-style-type: none"> <li>• Dictionary encoding</li> <li>• Run-length encoding</li> <li>• Bit-packing</li> </ul>
<b>Data Skipping</b>		Ja	Ja
<b>MVCC</b>	Ja	Ja (Optimiert für Batch-Insert)	
<b>Parallele Verarbeitung</b>		<ul style="list-style-type: none"> <li>• SIMD processing</li> <li>• Multi-core processing</li> </ul>	<ul style="list-style-type: none"> <li>• SIMD Processing</li> <li>• RAC (Shared nothing)</li> </ul>
<b>Besonderheiten</b>	<ul style="list-style-type: none"> <li>• DLL für DML Anfragen (kein ALTER TABLE)</li> <li>• Flexible Ablage auf “heap”</li> <li>• Compiled Stored Procedures</li> </ul>	<ul style="list-style-type: none"> <li>• Shadow Tables für OLTP Systeme (über Data Replication)</li> </ul>	<ul style="list-style-type: none"> <li>• Disk bleibt führendes Speichermedium</li> </ul>

# „NewSQL“-Anbieter (eine Auswahl...)

	VoltDB	MemSQL	SAP HANA
<b>Fokussierung</b>	OLTP mit hohem Durchsatz inkl. Kontext-Evaluierung	OLTP mit hohem Durchsatz bei gleichzeitigem Zugriff	OLTP + OLAP in einem System
<b>Datenhaltung</b>	Zeilen (?)	Zeilen	Beides (Spalten präferiert)
<b>Daten Kompression</b>	N/A	N/A	<ul style="list-style-type: none"> <li>• Dictionary encoding</li> <li>• Run-Length encoding</li> <li>• Cluster encoding</li> </ul>
<b>Data Skipping</b>	N/A	N/A	
<b>MVCC</b>	Ja	Ja	Ja
<b>Parallele Verarbeitung</b>	<ul style="list-style-type: none"> <li>• Single Threaded Transactions</li> <li>• Parallelverarbeitung im Cluster (Shared Nothing)</li> </ul>	Parallelverarbeitung im Cluster (Shared nothing)	Parallelverarbeitung im Cluster (Shared nothing)
<b>Besonderheiten</b>	<ul style="list-style-type: none"> <li>• Zugriff über kompilierte Stored Procedures</li> <li>• DB komplett In-Memory</li> </ul>	<ul style="list-style-type: none"> <li>• SQL Runtime Code Generation (C++)</li> <li>• Keine Locks auf Datenstrukturen (z.B. skip lists)</li> </ul>	<ul style="list-style-type: none"> <li>• DB komplett In-Memory</li> <li>• Plattform für Anwendungs-Entwicklung</li> </ul>

# SAP HANA Proof-of-Concept

## Ziele

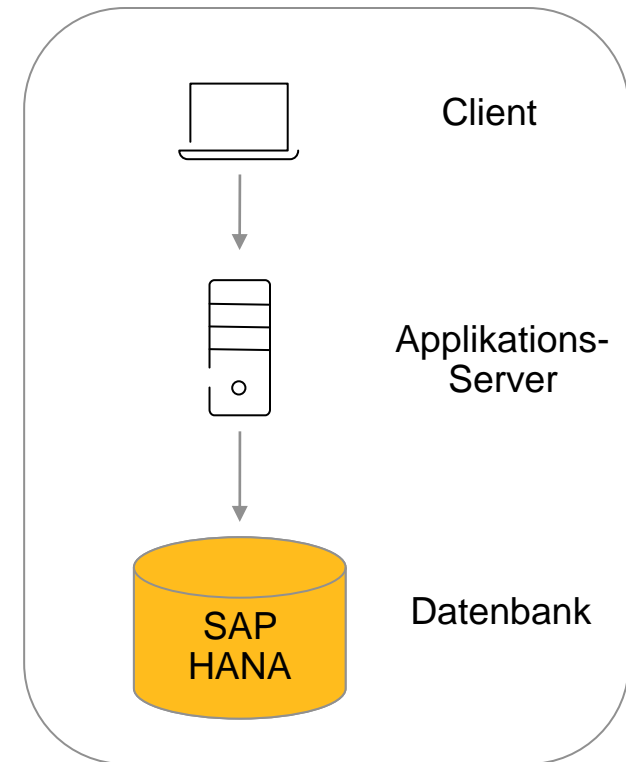
- Technische Machbarkeit
- Performance-Vergleich

## Vorgehen

- Migration der produktiven Daten nach HANA
- Daten werden komplett in Spalten abgelegt
- Keinerlei weitere Performance-Optimierung (!)
- Anwendung bleibt komplett unberührt

### Benchmark:

- Erstelle SQL Trace aus produktivem Betrieb
- Trace auf SAP HANA erneut ablaufen lassen



<https://www.de.capgemini.com/customer-experience-management/realtime-applications>

# Bisherige Ergebnisse

Projektbeispiele

Daten

Kompabilität

Performance

## Ergebnisse

- Enterprise Anwendung mit komplexen **Ad-hoc Abfragen** für Super-User
- Medienarchiv mit integrierter **Volltext-Suche**
  
- Erfolgreicher, automatischer Import der Daten **ohne Verlust**
- Datenbank-Größe um bis zu 65% **reduziert**
  
- **Großteil** der Queries (fast) **kompatibel**
- DB-spezifische Queries benötigen Bearbeitung
  
- Generell: **schneller** im Vergleich zur Produktion
- Für Komplexe Ad-hoc Queries **signifikant schneller**

# Agenda

- Motivation
- In-Memory im Überblick
- **Paradigmen im Wandel?**





# Paradigmen im Wandel?

## System-Architektur

### Paradigma

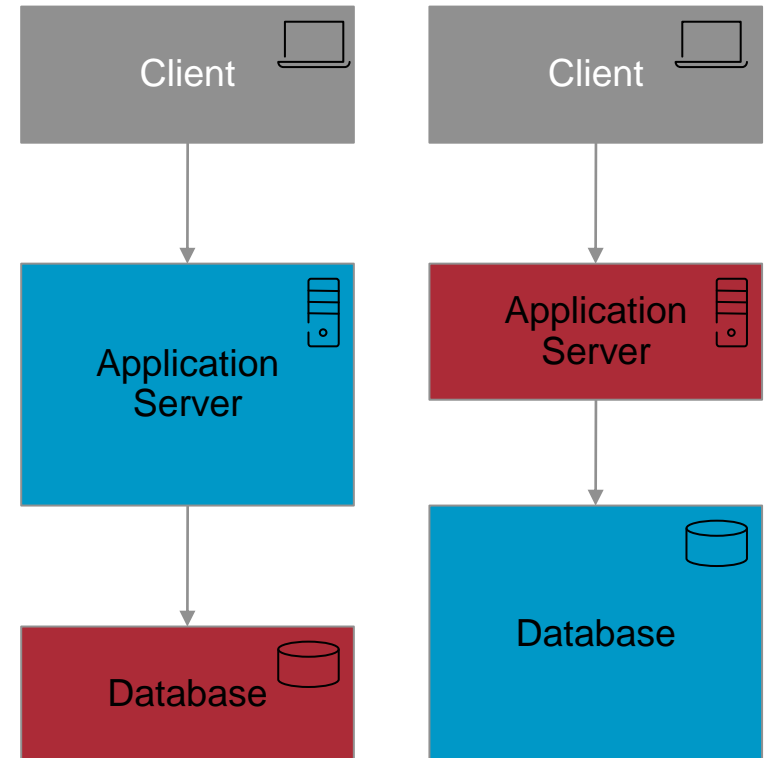
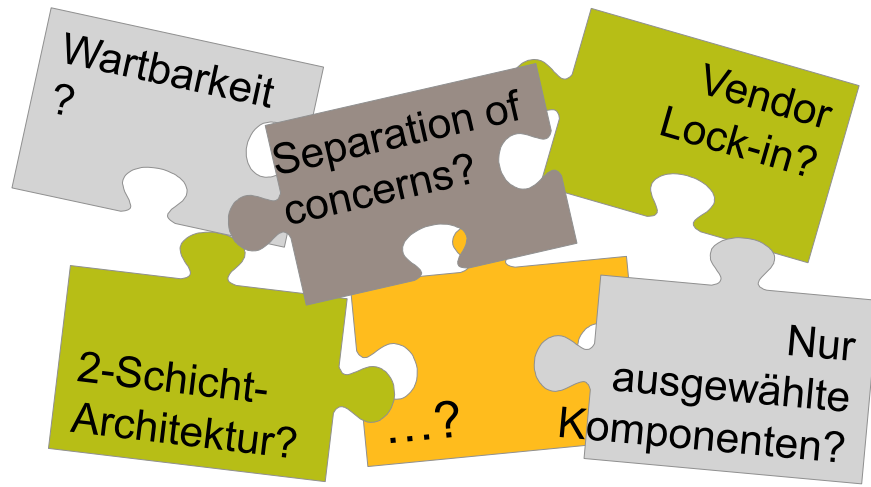
---

- Code-to-data
- 3-Schichten-Architektur

## Enterprise-Architektur

- Trennung von OLTP und OLAP
- vs. Fragmentierung der Data Management Systeme

# Code-to-data vs. 3-Schichten-Architektur



# Trennung zwischen OLTP und OLAP Systemen

## OLAP vs. OLTP



- Trennung bzgl. Arbeitslast
- Optimiert für die jeweilige Arbeitslast
- ETL Prozesse benötigt, um Daten zu portieren

## Nachteile

- Real-time Analyse stark erschwert
- OLAP meist eingeschränkt bzgl. Fragestellungen
- ETL Kosten
- Hohe Redundanz

## Mögliche Vision

- Einheitliche Datenhaltung sowohl für OLTP als auch für OLAP
- Abfragen direkt auf Produktionsdaten
- „Googling for Data Insights“

# OLTP / OLAP Charakteristiken neu bewertet

Charakteristik	OLTP	OLAP	Bewertung
<b>Performance</b>	Kritisch	Entspannter	<ul style="list-style-type: none"> <li>• <i>Trennung eher gezwungene Notwendigkeit</i></li> <li>• <i>keine inhärente Anforderung</i></li> </ul>
<b>Zugriffsmenge</b>	Klein	Groß	<ul style="list-style-type: none"> <li>• <i>OLTP Untermenge von OLAP</i></li> <li>• <i>Falls effizient, kein Argument für eine Trennung</i></li> </ul>
<b>Zugriffsart</b>	CRUD	Read-only Batch Update	<i>OLAP Untermenge von OLTP</i>
<b>Redundanz</b>	Nein	Ja	<i>Wird nicht mehr benötigt, Aggregation im Speicher</i>
<b>Datenmenge</b>	Klein	Groß	<i>Wenn parallelisierbar können auch große Datenmengen im Hauptspeicher eines Clusters gehalten werden</i>
<b>Anzahl User</b>	Groß	Klein	<i>Ebenfalls eine Frage der Parallelisierbarkeit</i>
<b>Verwendung</b>	Produktion	Analyse	<i>Letztlich nur die Frage welcher Client das System nutzt</i>
<b>Orientierung</b>	Applikation	Domäne	<ul style="list-style-type: none"> <li>• <i>Datenintegration/-Konsolidierung weiterhin benötigt</i></li> <li>• <i>Valides Argument für zentralisierte Data Warehouses</i></li> </ul>
<b>Datenhistorie</b>	Nein	Ja	<ul style="list-style-type: none"> <li>• <i>Valider Grund um OLTP nicht aufzublähen</i></li> <li>• <i>Insert-Only könnte Alternative darstellen</i></li> </ul>
<b>Mission Critical</b>	Ja	Nein	<i>Valider Grund zur Risikominimierung</i>

# Zusammenfassung

- Digitalisierung und Big Data benötigen neue technologische Antworten
- In-Memory ist eine mögliche Antwort und geht einher mit einer Sammlung von dazugehörigen Techniken
- Alle namhaften und zahlreichen neue Anbieter drängen auf den Markt
- Alte Paradigmen sollten im Zeitalter von In-Memory zumindest hinterfragt werden

People matter, results count.

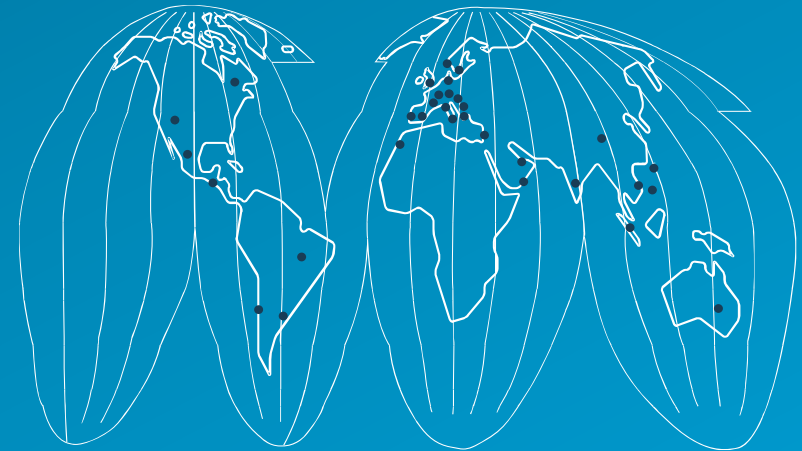


## Über Capgemini

Mit mehr als 140.000 Mitarbeitern in über 40 Ländern ist Capgemini einer der weltweit führenden Anbieter von Management- und IT-Beratung, Technologie-Services sowie Outsourcing-Dienstleistungen. Im Jahr 2013 betrug der Umsatz der Capgemini-Gruppe 10,1 Milliarden Euro.

Gemeinsam mit seinen Kunden erstellt Capgemini Geschäfts- wie auch Technologielösungen, die passgenau auf die individuellen Anforderungen zugeschnitten sind. Auf der Grundlage seines weltweiten Liefermodells Rightshore® zeichnet sich Capgemini als multinationale Organisation durch seine besondere Art der Zusammenarbeit aus – die Collaborative Business Experience™.

*Rightshore® ist eine eingetragene Marke von Capgemini*



[www.de.capgemini.com](http://www.de.capgemini.com)

