

## Editorial

Andreas Thor · Stefanie Scherzinger · Günther Specht

Online publiziert: 12. Juni 2014  
© Springer-Verlag Berlin Heidelberg 2014

### Schwerpunktthema: Datenmanagement in der Cloud

Anwendungen, Institutionen und Unternehmen müssen mit immer größeren Datenmengen hantieren. Sowohl wissenschaftliche Experimente (z. B. Genomsequenzierung, Teilchenbeschleuniger) als auch die Aufzeichnung menschlicher Aktivitäten (z. B. soziale Netzwerke, Online-Shopping) erzeugen Datenmengen, die oft nur noch unter Einsatz von Cloud-Infrastrukturen beherrschbar sind. Die Kopplung hunderter bis tausender Rechner ermöglicht dabei eine ökonomische Verarbeitung großer Datenmengen, da Ressourcen (u. a. Rechenkapazität und Speicherplatz) flexibel „on-demand“ an den Bedarf von Nutzern und Anwendungen angepasst werden können. Gleichzeitig ergeben sich jedoch eine Vielzahl wissenschaftlicher Fragestellungen hinsichtlich der Speicherung, Verarbeitung und Analyse großer Datenmengen.

Dieses Themenheft soll einen Überblick über aktuelle Entwicklungen und Herausforderungen im Bereich Cloud-Datenmanagement geben und mögliche Richtungen für Forschung, Lehre und Entwicklung aufzeigen. Es wird vom Arbeitskreis *Datenmanagement in der Cloud* (AK DMC) herausgegeben. Er ist der jüngste AK im Fachbereich Datenbanken und Informationssysteme der GI und wurde im

Herbst 2012 gegründet. Heute stellt er sich mit den ersten fünf Beiträgen in diesem Heft des Datenbank-Spektrums vor. Aber auch sonst ist in den ersten knapp zwei Jahren bereits viel passiert: Ein erster Workshop fand im Frühjahr 2013 auf der BTW in Magdeburg statt; der nächste ist für die Jahrestagung der GI in Stuttgart im September 2014 angekündigt und fungiert gleichzeitig als Herbsttreffen der GI-Fachgruppe Datenbanksysteme.

Inhaltlich hat sich in den letzten Jahren in der Cloud-Diskussion einiges getan. Man braucht dazu nur die Buzzwords Revue passieren lassen: Wurden am Anfang noch die unterschiedlichen Servicemodelle zwischen SaaS (Software as a Service), PaaS (Platform as a Service) und IaaS (Infrastructure as a Service) hervorgehoben, so standen bald MapReduce (und Hadoop) als Möglichkeit der parallelen Datenverarbeitung sehr großer Datenmengen in der Cloud und die skalierbare und von überall zugängliche Speicherung der Daten in der Cloud im Vordergrund. Das Potenzial der Auswertung und Nutzung von *Big Data* führte spätestens mit dem NSA-Skandal zur Diskussion um Datenschutz und Datensicherheit in der Cloud.

Dennoch ist *die Cloud* ein voller Erfolg. Die Virtualisierung der Speicherung und Berechnungen in der Cloud ermöglicht einerseits Kostenvorteile für kleinere und mittlere Unternehmen, gerade bei sehr variablen Kapazitätsanforderungen. Andererseits ermöglicht die Übertragung der Auswertung der Daten in die Cloud gerade im wissenschaftlichen Umfeld überhaupt erst eine zeitnahe Auswertung. Letzteres wird nicht zuletzt in der Genomanalyse augenscheinlich, wo Performanzgewinne bei der Sequenzierung eines ganzen Genoms von Tagen auf Minuten erzielt werden können. Bestehen bleibt die Problematik der Datensicherheit beim Cloud Computing. Leichte Benutzbarkeit (auch für Endanwender) und *State of the Art*-Kryptographie oder -Anonymisierung

---

A. Thor (✉)  
Deutsche Telekom, Hochschule für Telekommunikation Leipzig,  
Leipzig, Deutschland  
E-Mail: thor@hft-leipzig.de

S. Scherzinger  
OTH Regensburg, Regensburg, Deutschland  
E-Mail: stefanie.scherzinger@hs-regensburg.de

G. Specht  
Universität Innsbruck, Innsbruck, Deutschland  
E-Mail: guenther.specht@uibk.ac.at

stehen sich bisweilen gegenüber; erst recht bei freigiebigem Umgang in Social-Media-Plattformen, seien sie offen oder geschlossen.

Das Gebiet des AK *Datenmanagement in der Cloud* ist also sehr weit gespannt, wengleich naturgemäß der Datenbankaspekt, also die Speicherung und Verarbeitung großer Datenmengen in der Cloud, im Vordergrund steht. Diese Breite zeigt sich auch in den Artikeln dieses Heftes. Der *Call for Papers* stieß wegen der Aktualität und Dynamik dieses Themas auf große Resonanz, sodass nur die Hälfte der eingereichten Beiträge zur Veröffentlichung in diesem Heft angenommen werden konnten.

Der erste Beitrag *Database Backend as a Service – Automatic Generation, Deployment, and Management of Database Backends for Mobile Applications* von Francis Gropengießer und Kai-Uwe Sattler (TU Ilmenau) geht von den stark fluktuierenden Lastanforderungen bei Events, Festivals oder Konferenzen durch mobile Anwendungen aus. Die dynamische Skalierung ist dabei die Hauptherausforderung. Dafür verlegen sie die Datenbank als Service in die Cloud und stellen ein Framework zur automatischen Generierung, zum Überwachen und zur Skalierung vor.

Der zweite Beitrag *A Real-time Materialized View Approach for Analytic Flows in Hybrid Cloud Environments* von Weiping Qu und Stefan Deßloch (TU Kaiserslautern) schlägt einen Ansatz für Echtzeit-Materialisierung von Views in stark datenflussorientierten Anwendungen wie Business Intelligenz, ETL oder Hadoop vor. In traditionellen Datenbanken werden materialisierte Views zur Performanz-Optimierung eingesetzt. In diesem Beitrag wird untersucht, wie diese Idee auf verteilte Plattformen in der Cloud mit großen Datentransformationsprozessen übertragen werden kann und welche Erweiterungen für stark datenflussorientierte Auswertungen nötig sind.

Im dritten Beitrag des Heftes *Iterative Computation of Connected Graph Componentens with MapReduce* von Lars Kolb, Ziad Sehili und Erhard Rahm (Universität Leipzig) werden drei Erweiterungen vorgeschlagen, um sowohl die Menge der Zwischenergebnisse als auch die Anzahl der Iterationen zu reduzieren, die zur Berechnung zusammenhängender Komponenten in großen Graphen benötigt werden. Dabei werden sowohl die Map- als auch die Reduce-Phase genauer analysiert. Insbesondere die frühe Separierung stabiler Komponenten aus der weiteren Berechnung bringt hier einen signifikanten Performance-Gewinn.

Der vierte Beitrag *Datenbanken ohne Schema* von Meike Klettke (Universität Rostock), Stefanie Scherzinger (OTH Regensburg) und Uta Störl (Hochschule Darmstadt) beschäftigt sich mit der agilen Anwendungsentwicklung mit schema-flexiblen NoSQL-Datenbanksystemen. Bei der Speicherung und Verarbeitung großer Datenmengen in der Cloud unter Verwendung eines agilen Ansatzes in der Softwareentwicklung treten Schemaänderungen gerade am

Anfang laufend auf. Verwendet man relationale Datenbanksysteme, geht man immer von festen, in die Cloud zu exportierenden Schemata aus. Schemaänderungen sind dann aufwändig. Demgegenüber werden in diesem Beitrag schemalose NoSQL-Datenbanken als Lösung untersucht.

Schließlich rundet ein Community-Beitrag zu *Cloud-Technologien in der Hochschullehre – Pflicht oder Kür?* von Stefanie Scherzinger (OTH Regensburg) und Andreas Thor (Hochschule für Telekommunikation Leipzig) das Schwerpunktthema dieses Heftes ab. Sie führten dazu eine Umfrage unter zwanzig Hochschulen durch und werteten diese aus. Es zeigte sich deutlich, dass sich das Thema Cloud in der Lehre zunehmend etabliert. Die Autoren zeigen auch interessante Vorschläge in Richtung praktische Übungen oder Praktika in diesem Bereich auf.

Die fünf Beiträge zum Schwerpunktthema dieses Heftes werden durch einen Fachbeitrag *Unleashing XQuery for Data-independent Programming* von Sebastian Bächle (SAP Deutschland) und Caetano Sauer (TU Kaiserslautern) ergänzt. Die Sprache XQuery wurde ursprünglich als SQL-Äquivalent für XML-Daten entwickelt. Ihre Wurzeln in der funktionalen Programmierung erlauben jedoch eine einfache Verarbeitung fast aller Arten von strukturierten und semi-strukturierten Daten, wobei die vollständige Sprache insbesondere bei erweiterten Sprachkonzepten für große Datenvolumina kaum in effizienter Weise zu implementieren ist. Daher schlägt diese Arbeit eine neuartige Compilierungsstrategie vor, die auf Erweiterbarkeit und effiziente Verarbeitung von XQuery abzielt. Zur Vereinfachung von Optimierungs- und Übersetzungsprozessen strebt sie eine strikte Trennung von logischen, mengenorientierten Konzepten und physischen Aspekten an. Dadurch lassen sich teure iterative Abschnitte bei einer Anfragebearbeitung in eine Pipeline von relationenartigen Operatoren überführen, die dann für die Optimierung von Verbundverarbeitung, Indexnutzung und Parallelisierung zugänglich ist. Der hybride Compilationsansatz führt auf eine effiziente und hochgradig erweiterbare Query Engine, die eine Berechnung von einfachen XML-Transformationen bis zu komplexen Datenanalysen, auch auf Nicht-XML-Daten, erlaubt. Vielfältige Experimente und Vergleiche mit weltweit verfügbaren Systemen von Wettbewerbern bestätigen die Allgemeingültigkeit und Effizienz des gewählten Ansatzes.

Weiterhin finden Sie unter der Rubrik „Datenbankgruppen vorgestellt“ einen Beitrag von Klaus Meyer-Wegener und Richard Lenz über die *Datenbank-Forschung am Lehrstuhl für Informatik 6 (Datenmanagement) der Friedrich-Alexander-Universität Erlangen-Nürnberg*, der nach einem Blick auf die geschichtliche Entwicklung des Lehrstuhls einen Überblick über die aktuellen Forschungsthemen und die Aufgaben in der Lehre gibt.

In diesem Heft bietet die Rubrik „Dissertationen“ sechs Kurzfassungen von Dissertationen aus der deutschen DBIS-Community.

Die Rubrik „Community“ enthält unter *News* den *Call for Papers* für die BTW 2015 in Hamburg und weitere aktuelle Informationen.

## Künftige Schwerpunktthemen

### Data Management on New Hardware

Continued progress and evolution of computer hardware and infrastructure provide new optimization opportunities for data management concerning performance, scalability, availability, energy efficiency, and more. The traditional DB focus on I/O optimization is not enough to fully utilize this potential. New aspects include processors (multi-core CPUs, GPUs, FGPAs, etc.), multi-level memory hierarchies, storage systems (storage-class memories such as flash (SSDs) and phase-change memory), or low-power hardware. They impose new challenges and bottlenecks to DB performance, but, in turn, provide opportunities to improve and accelerate data-intensive systems for differing kinds of workloads (e.g., OLTP, DW/OLAP, ETL, Streaming/Realtime, or XML/RDF processing). An important goal is the automatic and transparent adaptation to the new hardware characteristics, thereby achieving maximum gain (e.g., performance, energy efficiency, etc.) for the applications.

A special issue of *Datenbank-Spektrum* wants to publish research contributions providing an overview over ongoing work in the data management area. Submissions covering topics from the following non-exclusive list are encouraged:

- Database algorithms (e.g., caching) and data structures (e.g., indexes) on modern hardware
- Hardware or memory support for query processing
- Cost models and query optimization for novel hierarchical memory systems
- Database architectures on multi-threaded and chip multi-processors
- Databases and transactional memory systems
- Database systems supporting energy efficiency and energy proportionality
- New benchmarks for or performance analysis of database workloads on modern hardware

Paper format: 8–10 pages, double column

Guest editors:

Theo Härder, University of Kaiserslautern,  
haerder@cs.uni-kl.de

Jens Teubner, TU Dortmund,  
jens.teubner@cs.tu-dortmund.de

### Informationsmanagement für Digital Humanities

In den Geisteswissenschaften fallen in immer größerer Menge digitale Forschungsdaten an. Dabei ergeben sich durch die spezifischen Rahmenbedingungen zahlreiche Herausforderungen für Datenbanken und IR-Systeme: Die Daten und Dokumente sind heterogen in Sprache, Struktur und Qualität. Es gibt zwar eine Vielzahl von Standards und Methoden, eine übergreifende Sicht existiert aber kaum. Relevante Kollektionen mit elektronischen Texten, Metadaten, Bildern und anderen multimedialen Ressourcen liegen in verschiedenen Disziplinen und Institutionen vor und bilden eine hochgradig verteilte und heterogene Informationslandschaft, deren Verarbeitung oft im Rahmen spezifischer, geisteswissenschaftlicher Forschungsfragen erfolgt. Von besonderer Bedeutung sind die Erschließung, Veröffentlichung und Verwaltung digitaler Ressourcen im Rahmen spezifischer Anwendungen z. B. in der Archäologie, den Geschichts-, Sprach- oder Religionswissenschaften, aber insbesondere auch im Kontext interdisziplinärer Forschung. Im Themenheft sollen einführende und überblicksartige Artikel sowie aktuelle Forschungsergebnisse zu ausgewählten Themen ein breites Bild zum aktuellen Stand des Informationsmanagements für Digital Humanities geben.

Mögliche Themen aus diesem Bereich könnten z. B. sein:

- Integrierte Analyse, Verarbeitung und Visualisierung verteilter bzw. heterogener Kollektionen
- Nutzung, Entwicklung und Auswertung von Vokabularen, Thesauri und Ontologien
- Langzeitarchivierung und Datenprovenienz
- Katalogisierung, Annotation und Dokumentation von Ressourcen (Data Curation)
- Erkennung, Analyse und Visualisierung kollektionsinterner oder -übergreifender Zusammenhänge z. B. durch Analyse von Ort und Zeit, Themen, Named Entities
- Aspekte der Usability im Umgang mit verteilten und heterogenen Ressourcen
- Anwendungen zum Datenmanagement, zur Suche und zur Analyse in speziellen Anwendungsfeldern aus den Geisteswissenschaften
- *Big Data*-Technologien für die Digital Humanities
- Forschungsinfrastrukturen für die Digital Humanities

Beitragsformat: 8–10 Seiten, zweispaltig

Ankündigung einer Beitragseinreichung bis zum 1. August 2014

Gastherausgeber:

Andreas Henrich, Otto-Friedrich-Universität Bamberg  
andreas.henrich@uni-bamberg.de

Gerhard Heyer, Universität Leipzig  
heyer@informatik.uni-leipzig.de

Christoph Schlieder, Otto-Friedrich-Universität Bamberg  
christoph.schlieder@uni-bamberg.de

Einreichung der Beiträge bis zum 1. Oktober 2014

Best Workshop Papers of BTW 2015

This special issue of the “Datenbank-Spektrum” is dedicated to the Best Papers of the Workshops running at the BTW 2015 at the University of Hamburg. The selected Workshop contributions should be extended to match the format of regular DASP papers.

Paper format: 8–10 pages, double column

Selection of the Best Papers by the Workshop chairs and the guest editor: April 15th, 2015

Guest editor:  
Theo Härder, University of Kaiserslautern,  
haerder@cs.uni-kl.de

Deadline for submissions: June 1st, 2015