

Der „Homo Connectus“ – verloren im Datenmüll?

Unaufhaltsam wachsende Datenmengen
erfordern neue Technologien und
bieten neue Geschäftschancen



Jörg Bienert ist zusammen mit Michael Hummel Gründer/Geschäftsführer der empulse GmbH. Er verfügt über 20 Jahre Berufserfahrung in verschiedenen Führungspositionen und Branchen. Empulse bietet mit ParStream eine innovative Datenbanktechnologie zur Analyse von großen Datenmengen an.



Ekkehart Gerlach ist Geschäftsführer der deutschen medienakademie in Köln, die ab 2000 auf Anregung des Landes NRW von der Bertelsmann Stiftung aufgebaut wurde und als Fachakademie für Führungskräfte in den Bereichen Kommunikation und Medien, insbesondere Neue Medien, TK und IT tätig ist. Zuvor war er als Experte für Unternehmensstrategie und -entwicklung lange Jahre in den drei Bereichen tätig, die heute die Konvergenz zu Multimedia ausmachen: Informationstechnologie, Telekommunikation und Content.

Die Menge der weltweit erzeugten und gespeicherten Daten steigt exponentiell.

Laut Marktanalyst IDC verdoppelt sich das Datenvolumen derzeit alle zwei Jahre und wird noch im Jahr 2011 die 1,8 Zetabyte-Marke erreichen. Um diese „Big Data“ zu verarbeiten, braucht es neue Technologie-Ansätze. Von Jörg Bienert und Ekkehart Gerlach

Eric Schmidt, Ex-CEO von Google, hat errechnet, dass die Datenmenge, die heute innerhalb von zwei Tagen weltweit gespeichert wird, so groß ist wie alle gespeicherten Daten vom Beginn der Zivilisation bis zum Jahr 2003. Fachleute sind sich einig, dass die über uns hereinbrechenden Datenfluten – die so genannten „Big Data“ – zu einer wesentlichen Herausforderung unserer Gesellschaft werden, aber auch Chancen für neue Geschäftsfelder eröffnen. Der Begriff Big Data steht für die explodierende Datenmenge, ihre Analyse und neue Lösungsansätze. Was nach Unternehmen und deren IT klingt, reicht heute weit darüber hinaus bis hinein in den privaten Bereich. Große Datenmengen fallen in vielen Bereichen des Geschäfts- und Privatlebens an – längst nicht mehr nur bei traditionellen Datenlieferanten. So werden beispielsweise die Informationskonsumenten durch interaktive und soziale Medien immer mehr zu Datenproduzenten, millionenfach und jeden Tag. Mobile Endgeräte, Sensor-Netzwerke, Machine-to-Machine Kommunikation, RFID-Leser, Software Logs, Kameras und Mikrophone generieren weltweit einen kontinuierlichen Datenstrom von mehreren Terabyte pro Tag.

Nase vorn im Wettbewerb

Laut Roger Magoulas vom us-amerikanischen Fachverlag „O'Reilly“ wird die Fähigkeit, die grossen Datenmengen zu analysieren immer mehr zu einer Kernkompetenz im Informationszeitalter. Unternehmen, die dies beherrschen, werden bessere Voraussagen für die Markt- und Geschäftsentwicklung erstellen, ihre Wettbewerbsposition verbessern, die Produktivität steigern und neue Geschäftsmodelle erschließen können. In einer Studie berechnet die Management-Beratung McKinsey beispielhaft, dass in Europas öffentlichem Sektor durch Big Data-Analytics Mehrwerte von 250 Milliarden Euro geschaffen werden könnten. Auch könnten die Margen im Einzelhandel um 60 Prozent steigen, wenn die anfallenden Kundendaten schneller und besser als bei der Konkurrenz ausgewertet würden.

Neue Technologie-Ansätze sind gefragt

IT-Abteilungen versuchen häufig, der steigenden Datenflut mit traditionellen Mitteln zu begegnen, nämlich dem Ausbau von Infrastruktur und Hardware bei ansonsten unveränderter Software. Klassische Data Warehouses und Business Intelligence- (BI) Systeme sind aber immer weniger in der Lage, die Daten aktuell und mit kurzen Antwortzeiten zu vertretbaren Kosten zu verarbeiten. Die herkömmlichen IT-Technologien und -Prozesse sind nur unzureichend geeignet, sehr große Datenmengen zu analysieren – das besagt auch eine aktuelle Studie des Marktforschungsunternehmens Gartner. Da die Datenmengen schneller wachsen als die Verarbeitungsmöglichkeiten durch zusätzliche Server oder bessere Prozessoren, sind neue Technologie-Ansätze gefragt. „Beyond-BI“-Technologien beispielsweise führen zu einem Paradigmenwechsel in Strukturen, Programmierung und Prozessen sowie dem erforderlichem Mitarbeiter-Know-how. Vor diesem Hintergrund entstanden in den vergangenen Jahren neue Ansätze und Technologien zur Datenanalyse, die sich grob in zwei Kategorien aufteilen lassen: NoSQL-Daten-speicher und analytische Datenbanken.

Der populärste Vertreter im Bereich NoSQL („Not only SQL“) ist „Hadoop“. Hadoop basiert auf dem von Google entwickelten so genannten „Map-Reduce“- Ansatz und wird von der „Apache Foundation“ als „Open Source“ bereitgestellt. Hier werden zu verarbeitende Daten auf verschiedene Prozesse und Rechner aufgeteilt und verarbeitet („Map“), die Ergebnisse werden anschließend zusammengeführt („Reduce“). Hadoop wird von einer breiten Entwicklergemeinschaft unterstützt. Den Vorteilen eines einfachen transparenten und kostenlosen Parallelisierungs-Frameworks stehen eine Reihe von Nachteilen gegenüber, insbesondere, dass Auswertungen in Echtzeit nicht möglich und die Infrastrukturanforderungen hoch sind.

Im Bereich der analytischen Datenbanken sind einige neue Anbieter auf den Markt gekommen. „Netezza“ bietet eine hardware-

basierte Lösung, wobei spezielle Festplatten-controller mit so genannten „FPGAs“ („Field Programmable Gate Array“, zum Einsatz kommen. „Asterdata“ und „Greenplum“ haben Datenbankfunktionalitäten um einen „Map-Reduce“-Kern implementiert. „Vertica“ ist eine spaltenorientierte Datenbank mit einer „Shared-Nothing“-Architektur, „Exasol“ erzielt die Performance dadurch, dass die Daten komplett in den Hauptspeicher geladen werden (so genannte „in-memory“-Technologie) und „Parstream“ ist eine auf Big Data-Analytics spezialisierte Lösung, die aktuellste Datenbanktechnologien (wie z.B. Column-Store, in-memory, Shared-Nothing-Architektur) mit einem für die Echtzeit-Analyse von Massendaten entwickelten „High Performance Index“ vereint.

Big Data – in Zukunft immer wichtiger

Einige dieser Ansätze sind sicherlich noch Zukunftsmusik. Doch aufgrund der unaufhaltsam steigenden Datenflut ist der Zwang übergroß, neue Instrumente zu ihrer Beherrschung einzusetzen. Das führt möglicherweise sogar dazu, dass Big Data-Ansätze in den nächsten Jahren eine ähnlich zentrale Bedeutung für die IT erlangen, wie es gegenwärtig bei Cloud Computing der Fall ist. ■

DAS SEMINAR

01

Future Internet – Clean Slate auf grünen Wiesen oder Renovierung an denkmalgeschützten Gebäuden?

INHALTE

Mit dem Feldafinger Kreis hatten BITKOM, Bitkom Akademie und deutsche medienakademie im Januar 2011 über die Entwicklungen für das Internet der Zukunft konferiert. Aufgrund zahlreicher Anregungen präsentieren Bitkom Akademie, deutsche medienakademie und das Forschungsprojekt-Plattform G-Lab einen vertiefenden Workshop, der relevante Forschungsprojekte beschreibt. Ausgehend von möglichen zukunftsorientierten Architekturen werden dabei internationale Projekte mit einbezogen. Der Workshop will aufzeigen, wo welche Optionen und Chancen bestehen und welche Hemmnisse dem neuen Internet in die Quere kommen könnten.

ORT

Köln

TERMIN

29. Februar 2012, 13.00 – 19.00 Uhr

PREIS (zzgl. USt.)

290,- EUR / 190,- EUR (BITKOM-Mitglieder)