

Geomatik - Kolloquium

Do. 20.06. ab 16:15 Uhr GFZ Haus H, Raum VR1

Andre Luckow

Big Data: Anwendungsfelder in der Wissenschaft und Technologien

Die Menge an Daten, die weltweit anfallen und verarbeitet werden, steigt ständig an. Ein Grund ist die allgegenwärtige Verfügbarkeit von Sensoren, z.B. in mobilen Geräten, Maschinen, usw., die das umfangreiche Sammeln von Daten ermöglichen. Nach einer aktuellen Studie des IDC wurden im Jahr 2012 rund 2.2 Zetabyte digitale Daten erstellt, abgerufen oder repliziert; dazu zählen Texte, Bilder, Videos, Aufnahmen von Überwachungskameras, Lokationsdaten, usw. Der größte Teil dieser Daten wird dabei von Maschinen, z.B. mobilen Geräten, Flugzeugen, Fahrzeugen, Industrie-Robotern, erzeugt. Insbesondere in verschiedenen Wissenschaftsdisziplinen werden immense Datenmengen erzeugt: so generieren die Detektoren des Large Hadron Colliders am CERN rund 1 PB Daten am Tag; eine moderne Genome Sequenzierungsmaschine kann >1 TB/Tag Daten pro Tag erzeugen. Das Datenwachstum wird in naher Zukunft weiter zunehmen. Die Bewältigung dieser Datenflut ist mit verschiedenen infrastrukturellen Herausforderungen verbunden: so müssen neben dem Storage, Datentransfers, Datenreplikationen effizient durchgeführt werden; ein gutes Daten/Compute Co-Placement ist die Voraussetzung für die optimale Verarbeitung der Daten.

Cloud-Anbieter, wie Google und Amazon besitzen mittlerweile große, verteilte Infrastrukturen für die Verarbeitung von den durch ihre Diensten generierten Daten. Einen Teil der Infrastruktur wird auch externen Nutzern als Cloud-Dienst zur Verfügung gestellt. Infrastrukturen, wie EGI/Prace in Europa oder XSEDE in den USA, bieten immense Compute und Daten-Ressourcen für wissenschaftliche Anwendungen. Klassisch stand der Aspekt Compute im Mittelpunkt von solchen High Performance Infrastrukturen. Technologien aus dem Cluster und Grid Computing beschäftigten sich primär mit der Ausführung von Anwendungen in Rechnerverbänden, so genannten Clusters, sowie in hoch-verteilten Rechen-Grids. Daten-Handling Aspekte (Dateitransfers, File I/O) wurden in der Regel dem Compute Aspekt untergeordnet. Die Auswertung von Daten spielt aber zunehmend eine wichtigere Rolle - der Anteil an daten-intensiven Anwendungen steigt unaufhaltsam.

Big Data beschreibt den aktuellen Trend und Hype um Technologien die den Umgang mit großen Datenmengen ermöglichen. Ein Beispiel für eine solche Technologie ist Hadoop, ein Framework welches auf dem MapReduce Programmiermodell aufbaut und die effiziente Ausführung von datenintensiven Anwendungen in einem Cluster erlaubt. Im Rahmen dieses Vortrages werden verschiedene Anwendungsfelder sowie Technologien für den Umgang mit Big Data vorgestellt.

Andre Luckow arbeitet in der Automobil-Industrie und beschäftigt sich mit der Anwendung innovativer Technologien in der Industrie. Des Weiteren ist er Forscher an der Rutgers Universität, New Jersey, USA und Lehrbeauftragter an der Hochschule München. Seine aktuellen Forschungsarbeiten beschäftigen sich mit den Herausforderungen rund um Big Data Technologien und der Entwicklung von higher-level Abstraktionen für verteilte, daten-intensive Anwendungen.



Next talk

27.06.2013 Prof. Dr. Ulrike Lucke, Navigationsunterstützung für Blinde und Sehbehinderte
16:15 Uhr **GFZ Haus H, Raum VR1**

Veranstalter

Prof. Dr. Bettina Schnor, Institute of Computer Science, Universität Potsdam
Prof. Dr. Joachim Waechter, German Research Centre for Geosciences