

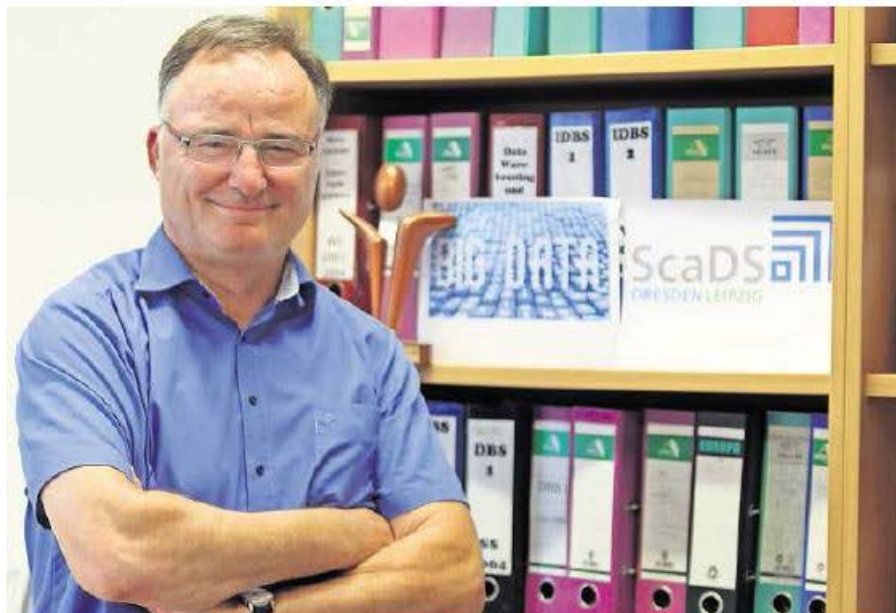
Galaxy und Sirius: Big-Data-Zentrum an der Uni mit mehr Computerpower

1,3 Millionen Euro für neue Server plus Zubehör / Professor Rahm: „Viel Performance durch Parallelverarbeitung“

VON MARIO BECK

Das technische Rückgrat des von der Leipziger Universität zusammen mit der TU Dresden betriebenen Big-Data-Kompetenzzentrums (Scads) wird weiter gestärkt. Zur bisher installierten Hardware kommt jetzt ein als Galaxy-Cluster bezeichneter Komplex mit 90 Servern. 60 davon sind an der Alma mater schon installiert und gehen nach und nach ans Netz. 30 Server sind für Dresden bestimmt, werden aber auch vom hiesigen Uni-Rechenzentrum (URZ) aus gesteuert. Über exakt 1080 Prozessoren verfügt das Galaxy-Gesamtsystem, die Festspeicherkapazität erreicht 2,16 Petabyte – ein Petabyte entspricht einer Million Gigabyte. „Das ist schon eine richtige Hausnummer“, sagt Erhard Rahm, der an der Uni die Professur für Datenbanken innehat und am Scads einer der zwei Koordinatoren ist.

Für dieses Jahr wird am URZ noch die Lieferung des Sirius-Clusters mit zwei Memory-Servern erwartet, die alleine in ihren Hauptspeichern je sechs Terabyte Daten bunkern können. Solche Mengen fallen beispielsweise bei der Erforschung von Genomen oder Krankheitsmarkern an. Beides sind Felder, in die sich die im Scads vereinten Experten unter anderem einbringen. Vor zwei Jahren war das Big-Data-Zentrum als Gemeinschaftsprojekt



Professor Erhard Rahm leitet seitens der Leipziger Uni das gemeinsam mit der TU Dresden betriebene Big-Data-Kompetenzzentrum.

Foto: André Kempner

der Unis in Leipzig und Dresden mit einem Fördervolumen von 5,6 Millionen Euro an den Start gegangen. Galaxy, Sirius und Zubehör schlagen jetzt mit rund 1,3 Millionen Euro zu Buche.

Zum Scads-Konsortium gehören noch das Leibniz-Institut für ökologische Raumentwicklung und das Max-Planck-Institut

für Zellbiologie und Genetik, die in Dresden ansässig sind. Hinzu gesellen sich eine Vielzahl assoziierter Partner aus Wissenschaft und Wirtschaft, darunter die Leipziger Data Virtuality GmbH und das Logistik-Netzwerk Leipzig-Halle.

Die neuen Galaxy-Server, die sowohl für die Grundlagen- als auch die ange-

wandte Forschung eingespannt werden, sind in einer Shared-Nothing-Architektur konfiguriert. Rahm: „Im Verbund liegt ihre Leistungsstärke. Die Performance kommt aus der Parallelverarbeitung.“ Aufgesetzt ist dabei meist eine Software-Technik namens Hadoop, auf deren Basis Rahm mit seinem Team ein Vorhaben mit der Kennung Gradoo vorantreibt. Es zielt auf die schnelle Analyse vernetzter Daten, die – wie bei Facebook – von hoher Beziehungsdynamik geprägt sind. Mit den Themen Datenintegration, Wissensextraktion und Visualisierung hat das Scads drei feste Standbeine und strahlt stark aus: In die Lebens- und Umweltwissenschaften, die Werkstoffkunde oder die digitalen Geisteswissenschaften. Der gute Ruf des Zentrums hat sich schon international herumgesprochen und sorgt für Echo.

Zur jüngsten Big-Data-Sommerschule an der Uni war der Andrang aus aller Welt groß. Rund 120 Teilnehmer kamen – und Rahm ließ wegen seiner Verpflichtungen bei der Summerschool eine Einladung in die USA sausen. Er war zum 25. Geburtstag des Microsoft-Research-Centers (MRC) gebeten worden, um dort den Outstanding Collaborator Award entgegenzunehmen. 32 Auserwählte aus aller Herren Länder, die in der Vergangenheit als Gastforscher am MRC gewirkt und dabei die nachhaltigsten Spuren hinterlassen hatten, bekamen die Ehrung.