

PROFIreferenz

Die Success-Storys der PROFI AG

Technische Universität Berlin

Spitzenforschung im Bereich Big Data



Große Datenmengen analysieren, Erkenntnisse für Theorie und Praxis gewinnen

© TU Berlin/Dahl

TU Berlin

Die Technische Universität Berlin ist eine international renommierte Universität. Herausragende Leistungen in Forschung und Lehre, die Qualifikation von sehr guten Absolventinnen und Absolventen sowie eine starke regionale, nationale und internationale Vernetzung mit Wissenschaft und Wirtschaft stehen im Fokus. Das Fachgebiet für Datenbanksysteme und Informationsmanagement (DIMA) unter der Leitung von Prof. Dr. Volker Markl forscht momentan im Rahmen des Projekts Stratosphere daran, die Analyse von sehr großen Datenmengen, wie sie in Unternehmen und in der Forschung anfallen, zu vereinfachen. Dafür untersuchen er und sein Team die massiv-parallele Datenverarbeitung in einer Cloud aus virtualisierten Rechnern. Diese Arbeit ergänzt einige Projekte im deutschen IBM Forschungs- und Entwicklungszentrum bei Stuttgart, wo Produkte und Lösungen zu den Themen Big Data und Analytics für den Weltmarkt entwickelt werden.

Die Ausgangssituation

Für die Grundlagenforschung zum Thema Big Data standen dem Forscherteam eigene Server-Systeme zur Verfügung. Es handelte sich jedoch um kleinere Einzelsysteme, die bei großen Rechenaufgaben im Giga- und Terabyte-Bereich schnell an ihre Leistungsgrenze stießen.

Die Ziele: Mehr Leistung für Big Data

Um auch künftig exzellente Forschungsergebnisse im Bereich Big Data zu erzielen, benötigten die Forscher eine leistungsfähige, erweiterbare IT-Infrastruktur aus einem Guss. Außerdem würde eine einheitliche IT-Umgebung dazu beitragen, die Administration einfacher und effizienter zu gestalten. Zudem wären die Systeme besser ausgelastet. Big Data und Analytics sind ein wichtiger Bereich der IBM-Produktentwicklung. So entschloss sich das Unternehmen, die Forschungsarbeit von Professor Markl und seinem Team zu fördern. IBM zeichnete Professor Markl mit dem IBM Shared University Research Award (SUR) aus und stiftete in diesem Rahmen einen Großteil des IBM PowerLinux Big Data Clusters für Forschungszwecke. Das gemeinsame Vorhaben von TU Berlin, IBM und PROFI als IBM Premier Business Partner sollte den Austausch zwischen Wissenschaft und Praxis fördern und einen Erkenntnisgewinn für alle Seiten bringen.

Die PROFI-Leistung

Der Cluster besteht aus 120 POWER7-Prozessoren, 640 GB Hauptspeicher und rund 40 Terabyte Plattenspeicher. Durch die Konfiguration mit IBM Infosphere Big Insights Software kann das System große Datenmengen in Echtzeit verarbeiten und ana-

“Der Preis von IBM ermöglicht es meinem Fachgebiet sowie meinen Kollegen am Data Analytics Lab der TU Berlin, anhand von realen Problemen mit riesigen Datenmengen Spitzenforschung und praxisnahe Lehre durchzuführen.“

Prof. Dr. Volker Markl, TU Berlin



lysieren. PROFI lieferte hierfür alle notwendigen Komponenten und kümmerte sich um die komplette Abwicklung sowie die individuelle Logistik. Im Rechenzentrum der TU Berlin wurde das neue System in die bestehende Infrastruktur integriert und an die Anforderungen des Forscherteams angepasst. Wenn das Forschungsteam die neue Plattform erweitern möchte, übernimmt PROFI die nötigen Aufgaben.

Die Vorteile: Forschung und Praxis im Dialog

Das gemeinsame Projekt fördert den Austausch zwischen Forschung und Praxis und schafft Vorteile für beide Seiten: Anhand realer Probleme mit enormen Datenmengen kann die TU Berlin Spitzenforschung betreiben und in der Lehre einen größeren Praxisbezug herstellen. Organisationen aus Wirtschaft und Wissenschaft können in verschiedenen Bereichen von den Forschungsergebnissen profitieren: Beispiele sind Datenbanksysteme und ihre Ablagestruktur, Marktforschung und Klimadatenanalyse.

Dank der exzellenten Zusammenarbeit mit den TU-Mitarbeitern konnten wir das komplexe Projekt nahezu reibungslos abwickeln und eine Big Data-Umgebung schaffen, die den Bedürfnissen der Forscher genau entspricht.“

Jonas Splettstößer, Vertriebsbeauftragter der PROFI AG

IBM kann die Forschungsergebnisse für die Produktentwicklung verwenden, sodass Kunden später wiederum von optimierten Produkten profitieren. Der leistungsstarke, bedarfsgerecht erweiterbare PowerLinux Cluster unterstützt die Wissenschaftler dabei, auch in Zukunft exzellente Forschungsergebnisse zu erzielen – denn die Server-Systeme können auch weiter steigende Datenmengen zuverlässig verarbeiten. Heute ist der Faktor Zeit oftmals wichtig für den Geschäftserfolg. Das leistungsfähige IT-Umfeld unterstützt das Forscherteam bei der Zielsetzung, die Analyse riesiger Datenmengen zu beschleunigen und Organisationen zu befähigen, rechtzeitig Entscheidungen aus den Daten abzuleiten. Die einheitliche IT-Infrastruktur sorgt für eine einfachere Administration und bessere Auslastung der Systeme.

Fazit

Professor Markl und sein Team arbeiten daran, die Analyse sehr großer Datenmengen zu vereinfachen. IBM fördert die Forschungsarbeit von Prof. Dr. Volker Markl und seinem Team im Bereich Big Data an der TU Berlin mit der Stiftung des europaweit leistungsstärksten PowerLinux Big Data Clusters. Innerhalb des Projekts ist PROFI dafür verantwortlich, das System an die Anforderungen des Kunden anzupassen, bei Bedarf zu erweitern, die entsprechenden Teile bereit zu stellen und als lokaler Technologiepartner den reibungslosen Betrieb zu sichern. Auch weiterhin werden die Projektpartner den PowerLinux Cluster an

Technische Details

- » 1 x IBM Hardware Management Console
- » 1 x IBM 42U Deep Dynamic Rack
- » 1 x IBM G8124ER 10GbE Rack Switch
- » 1 x IBM 24 Port 1Gb EX220 (Management)

- » 10 x IBM PowerLinux Server

- » PowerVM Enterprise Edition

- » 12 Core mit 3,7 GHz per Node
- » 20 x 600 GB 10k (12 TB Storage)
- » 20 x 73,4 GB 15k (1,46 TB Storage)
- » 8 x 8 x 10 (640 GB Hauptspeicher)

die Erfordernisse der Wissenschaftler anpassen. So kann die TU Berlin auf lange Sicht eine leistungsfähige, zukunftsorientierte IT-Plattform für ihre Forschung im Bereich Big Data nutzen. Das Projekt fördert den Austausch zwischen Wissenschaft und Praxis und bringt neue Erkenntnisse für beide Seiten.

Kundennutzen

- » **Zukunftsfähige Forschungsarbeit durch leistungsstarken, bedarfsgerecht erweiterbaren PowerLinux Cluster**
- » **Förderung der Spitzenforschung und praxisnahen Lehre an der TU Berlin durch Untersuchungen anhand realer Probleme mit riesigen Datenmengen**
- » **Nutzbringende Verwendung der Forschungsergebnisse in verschiedenen Praxisbereichen wie Produktentwicklung, Marktforschung oder Klimadatenanalyse**
- » **Fähigkeit, rechtzeitig Entscheidungen aus enormen Datenmengen abzuleiten wird durch leistungsstarkes Rechnersystem vorangebracht**
- » **Einfachere Administration und bessere Auslastung der Systeme durch einheitliche IT-Landschaft**

Riesige Datenmengen in jedem Netz

- » Heutzutage verwalten moderne Datenbanksysteme in globalisierten Unternehmen nahezu 400 Millionen Konten mit 10.000 Transaktionen pro Sekunde.
- » Data-Warehouses speichern über eine Milliarde strukturierter Datensätze in einem Umfang von mehreren Petabytes.
- » Unternehmensnetze organisieren Daten in Millionen von Seiten.
- » Im Internet befinden sich ca. 200 Millionen Webauftritte mit über 50 Milliarden Webseiten. Dynamisch generierte Inhalte und Mashup im Web 2.0 erhöhen die Anzahl der vorhandenen Informationen exponentiell.

Das Projekt Stratosphere

Stratosphere ist eine von der Deutschen Forschungsgemeinschaft (DFG) geförderte Forschergruppe. Hier forschen Prof. Dr. Volker Markl und sein Team daran, die Analyse von sehr großen Datenmengen, wie sie in Unternehmen und der Forschung anfallen, zu vereinfachen. Dafür untersuchen er und sein Team die massiv-parallele Datenverarbeitung in der Cloud aus virtualisierten Rechnern. Die Forschungsarbeiten sollen Organisationen dabei unterstützen, rechtzeitige Entscheidungen auch bei der Auswertung von riesigen Datenmengen zu treffen.

Unsere IT-Lösungen für Ihren Erfolg

- » System-Infrastruktur
- » Software-Lösungen
- » SAP-Services
- » Projekt-Services
- » Managed Services

Unsere Partner

Gemeinsam mit unseren zertifizierten Partnern entwickeln wir für Sie die optimale Lösung.

