

## Data-Science-Kompetenzzentren als Schlüssel zum Erfolg

*Von Prof. Dr. Erhard Rahm, Universität Leipzig / ScaDS Dresden / Leipzig Competence Center for Scalable Data Services and Solutions*

**Erfolgreiche Data-Science-Lösungen erfordern die kombinierte Expertise zu Datenaufbereitung/Big Data, Data Mining und maschinellem Lernen sowie Anwendungshintergründen. Diese breite Expertise kann in der Regel nur durch ein Team von Spezialisten bereitgestellt werden, die idealerweise an Data-Science-Kompetenzzentren kooperieren. In der Hochschulausbildung sollte Data Science sowohl auf Bachelor- auch als auf Masterebene stark ausgeweitet werden.**

Data Science beinhaltet Methoden zur umfassenden Analyse von Daten zur Gewinnung neuer Erkenntnisse in unterschiedlichsten Anwendungsgebieten. Das Spektrum der Analyseverfahren ist weit gefächert und reicht von einfachen Datenabfragen bis hin zu statistischen Data-Mining-Verfahren und Techniken des maschinellen Lernens. Wesentlich für die Aussagekraft und Qualität der Analyse ist eine umfassende Aufbereitung der Daten (Datenbereinigung und -konsolidierung), die oft aus unterschiedlichen Quellen stammen, teilweise unstrukturiert sind und zahlreiche Qualitätsmängel und Fehler beinhalten können. Die Daten sind zudem teilweise hoch-dynamisch (z.B. Sensordaten) und oft von sehr großem Umfang (Big Data). Diese Aspekte verlangen sehr leistungsfähige, hochparallele Lösungsansätze zur Datenverarbeitung, z.B. mit zahlreichen Prozessoren in lokalen Clustern.

Erfolgreiche Data-Science-Lösungen erfordern somit die kombinierte Expertise zu

- Big-Data-Technologien wie Datenaufbereitung/Datenintegration und skalierbare Ansätze zur parallelen Datenverarbeitung,
- zu Data Mining und maschinellem Lernen
- sowie die Anwendungsexpertise, um für die spezifischen Problemstellungen adäquate Verfahren zur Datenaufbereitung und Datenanalyse zu bestimmen.

Einzelne Personen werden in der Regel nur ein bis zwei dieser drei Bereiche ausreichend vertieft beherrschen, so dass anspruchsvolle Data-Science-Projekte in der Regel ein Team von Experten benötigen. Kompetenzzentren zu Data Science (bzw. Big Data oder Machine Learning) können die unterschiedlichen Kompetenzen bündeln und damit eine schnellere Nutz-

barkeit von Data Science in Wissenschaft und Praxis unterstützen. Dort können auch die entsprechenden Hardwareressourcen, in Kombination mit geeigneter Administrations- und Serviceunterstützung, bereitgehalten werden. Vorreiter solcher Zentren sind die beiden BMBF-geförderten Kompetenzzentren für Big Data, ScaDS Dresden/Leipzig sowie BBDC. Weitere Zentren sind geplant oder bereits an anderen Orten eingerichtet worden, um den hohen Bedarf an Data-Science-Lösungen in Deutschland zu adressieren.

Die Ausbildung zu Data Science sollte an den Hochschulen auf dem Bachelor- und Masterlevel stark ausgebaut werden, um der immensen Nachfrage nach akademisch ausgebildeten Data-Science-Spezialisten Rechnung zu tragen. Dies kann in eigenständigen Studiengängen erfolgen oder durch entsprechende neue Schwerpunkte in existierenden Studiengängen, z.B. der Informatik. Dabei sind in Informatikstudiengängen vor allem die Methoden in den beiden

ersten der oben genannten Gebiete zu berücksichtigen. Die Anwendungsaspekte können im Rahmen eines Ergänzungsfachs sowie durch Praktika und in der Abschlussarbeit berücksichtigt werden.

### Fazit und Ausblick

- 1) Die Einrichtung von Kompetenzzentren zu Data Science (Big Data, Machine Learning) hat sich bewährt, sollte durch Bund und Länder jedoch wesentlich verstärkt werden, jeweils mit standortspezifischen Forschungs- und Anwendungsschwerpunkten.
- 2) Umfassende Mehranstrengungen sind in der Data-Science-Ausbildung erforderlich, insbesondere zur Etablierung neuer bzw. angepasster Bachelor- und Masterstudiengänge an den Hochschulen.
- 3) Um einem Wildwuchs solcher Studiengänge entgegenzutreten, wäre es hilfreich, wenn die GI Empfehlungen zur inhaltlichen Ausgestaltung solcher Studienangebote erstellen könnte.

### Über den Autor

*Prof. Dr. Erhard Rahm ist Professor für Informatik (Datenbanken) an der Universität Leipzig und einer der beiden Direktoren des BMBF-geförderten Big-Data-Kompetenzzentrums ScaDS Dresden/Leipzig. Seine Forschungsgebiete sind Big Data und Datenintegration. Er ist Autor mehrerer Bücher und von über 200 wissenschaftlichen Publikationen. Seine Forschungsergebnisse wurden bereits mehrfach ausgezeichnet, u.a. mit den angesehenen VLDB Ten Year Best Paper Award sowie dem ICDE Influential Paper Award. Prof. Rahm ist gewähltes Mitglied im DFG-*



*Fachkollegium Informatik. In der GI ist er Sprecher des Fachbereichs Datenbanken und Informationssysteme (DBIS).*