





# WERDEN SIE DATA SCIENTIST!

Um das Potenzial von Big-Data-Technologien nutzen zu können, bedarf es Kompetenzen aus unterschiedlichsten Bereichen. Bislang gibt es nur wenige Fachkräfte und Unternehmen, die das benötigte Wissensspektrum abdecken. Solche Experten werden jedoch dringend gesucht. Aus diesem Bedarf heraus ist das Kompetenzfeld »Data Science« entstanden. Ein »Data Scientist« bündelt das nötige heterogene Fachwissen, um mit Hilfe von Big-Data-Technologien die Produkte, Dienstleistungen und Betriebsabläufe eines Unternehmens zu optimieren.

Unser Schulungsprogramm Data Science widmet sich der Vermittlung der dazu nötigen Kompetenzen. In Lehreinheiten und Workshops lernen Entscheider sowie Fachkräfte praxisnah und gleichzeitig fundiert die wesentlichen Prinzipien und Vorgehensweisen für Data-Science-Anwendungen kennen. Diese können sie anschließend unmittelbar auf das eigene Unternehmen und die dortigen Big-Data-Anwendungsfälle und Technologien übertragen.

## WIR BIETEN FOLGENDE MODULE AN:

### Data Scientist Basic Analytics

Fachkräfte mit Grundkenntnissen in der Datenanalyse und erster Programmiererfahrung erlernen die Grundlagen über:

- Vorverarbeitung und Sampling
- Clustering
- Regression
- Klassifikation
- Verteilte Analyse mit MapReduce
- Einsatz in der Cloud

Für die praktischen Übungen kommen Tools wie R, Rapid Miner und Hadoop zum Einsatz. Nach der Schulung sind Sie in der Lage, erste eigene Analysefragestellungen zu bearbeiten und den Nutzen des Einsatzes von Big-Data-Tools zu bewerten.

### Data Scientist Text Analytics

Die Schulung ist für Experten und Unternehmen geeignet, die Big-Data-Anwendungen auf Basis von unstrukturierten Textdaten umsetzen. Sie lernen unter Anwendung von Open-Source-Werkzeugen wie NLTK (Natural Language Toolkit) und UIMA (Unstructured Information Management Architecture) folgende Methoden kennen:

- Vorverarbeitung von unstrukturierten Textdaten
- Textklassifikation
- Konzeptanalyse und Topic Modelle
- Disambiguierung
- Sequenzmodelle
- Textextraktionsaufgaben

Die Analysetechniken werden anhand offener Datenquellen wie dem Common Crawl, Wikipedia und DBPedia erprobt. Voraussetzung sind erste Python-Kenntnisse. Nach Abschluss des Moduls sind Sie in der Lage, größere Textdatenbestände effizient zu strukturieren, um daraus relevante Inhalte und Muster zu extrahieren.

### Data Scientist Visual Analytics

Sie analysieren in Ihrem Unternehmen bereits Big Data und möchten die Ergebnisse besser kommunizieren? Die Schulung »Visual Analytics« vermittelt Techniken der Visualisierung und explorativen Analyse von:

- Textdaten und Topic Modellen
- Social-Media-Daten, z.B. aus sozialen Netzwerken
- Raumbezogenen Daten und Zeitreihen
- Bewegungsmustern, z.B. aus GPS Daten

Als Werkzeuge werden unter anderem Gephi und Common-GIS eingesetzt. Die exemplarischen Datenquellen umfassen Beispieldatensätze aus Transaktionssystemen, GPS-Daten und dem Internet. Voraussetzungen zur Teilnahme sind Grundkenntnisse der Datenanalyse, Programmierkenntnisse oder Erfahrung mit einem Statistikpaket (R, SPSS, ...) sowie



statistische Grundkenntnisse. Nach der Teilnahme an diesem Modul sind Sie in der Lage, geeignete Visualisierungsformen zur Interpretation und Kommunikation komplexer Sachverhalte auszuwählen und anzuwenden.

### Big Data Architecture

Das Modul »Big Data Architecture« vermittelt IT-Entwicklern die Grundlagen wesentlicher Big-Data-Infrastrukturen:

- ! Speicherung großer Datenmengen in verteilten Dateisystemen
- ! MapReduce-Verfahren
- ! Echtzeitanalyse in Streaming-Daten
- ! Verteilung und Skalierbarkeit
- ! Definition von Datenworkflows
- ! NoSQL-Datenbanken

Als Werkzeuge kommen Hadoop, Cassandra, HDFS, Lucene und Storm zum Einsatz. Im »Living Lab Big Data« experimentieren Sie mit Datenquellen wie dem Common Crawl sowie mit Daten aus eigenen Web Crawlern. Die Schulung ermöglicht es Ihnen, die Einsatzmöglichkeiten von Big-Data-Technologien für verschiedene Einsatzszenarien fundiert zu bewerten und mit eigenen Experimenten zu beginnen.

### EXPERIMENTIEREN IM »LIVING LAB BIG DATA«

Alle Schulungsangebote finden im »Living Lab Big Data« des Fraunhofer IAIS statt, einer Lern- und Experimentierumgebung für Unternehmen. Im Living Lab lassen sich Big-Data-Problemlösungen anhand der exemplarischen Anwendung eines Technologie-Stacks erproben, die verschiedene öffentlich verfügbare Big-Data-Datenbestände vorhält und zum Testen zur Verfügung stellt.

Die dabei genutzte Big Data-Architektur (»Lambda Architektur«) erlaubt es vorwettbewerblich, herstellerneutral

und erweiterbar unterschiedliche Technologien zu testen. Damit steht nicht eine bestimmte Big-Data-Technologie im Vordergrund, sondern es wird eine Architektur genutzt, die aufgrund ihrer Flexibilität verschiedenste Einsatzzwecke in unterschiedlichen Branchen abdecken kann. Die Architektur sieht vor, dass Anwendungen sowohl für Analyseaufgaben in Echtzeit (Streaming) als auch für umfangreiche, aber weniger zeitkritische Analysen (Batch) umsetzbar sind.

### DAUER, PREISE UND ANMELDUNG

#### Dauer der Schulungsmodule

jeweils 2 Tage

#### Veranstaltungsort

Fraunhofer-Institutszentrum Schloss Birlinghoven  
53757 Sankt Augustin

#### Teilnahmegebühr pro Schulung

1.900 Euro (steuerfrei gem. § 4 Nr. 22a UStG)  
Die Gebühr enthält Begleitunterlagen und Verpflegung. Frühbucherrabatt ist möglich, maximale Teilnehmerzahl: 10. Bitte beachten Sie die Storno- und Teilnahmebedingungen unter [www.iais.fraunhofer.de/data-scientist.html](http://www.iais.fraunhofer.de/data-scientist.html)  
Im Anschluss an jedes Schulungsmodul erhalten Sie eine Teilnahmebescheinigung.

#### Organisation und Anmeldung

Frau Renate Henkeler  
Telefon 02241 14-2332  
E-Mail [renate.henkeler@iais.fraunhofer.de](mailto:renate.henkeler@iais.fraunhofer.de)

Gerne können Sie sich auch online anmelden unter [www.iais.fraunhofer.de/data-scientist.html](http://www.iais.fraunhofer.de/data-scientist.html)

**Fraunhofer-Institut  
für Intelligente Analyse- und  
Informationssysteme IAIS**

Schloss Birlinghoven  
53757 Sankt Augustin

**Knowledge Discovery**

Ansprechpartner: Dr. Michael May  
Telefon +49 2241 14-2039  
michael.may@iais.fraunhofer.de

[www.iais.fraunhofer.de/bigdata](http://www.iais.fraunhofer.de/bigdata)