




Kursstart alle 4 Wochen


# Big Data Engineer mit Statistik


Der Kurs behandelt den Bereich Data Engineering und das Verarbeiten und Berechnen großer Datenmengen. Zudem wird der Umgang mit einem statistischen Programm erläutert, um Zusammenhänge zu erkennen und Beobachtungen zu verifizieren. Auch der Einsatz von Künstlicher Intelligenz (KI) in diesem Bereich wird erläutert.

 **Abschlussart**  
Zertifikat „Big Data Engineer“  
Zertifikat „Statistik“

 **Abschlussprüfung**  
Praxisbezogene Projektarbeiten mit Abschlusspräsentationen

 **Dauer**  
12 Wochen

 **Unterrichtszeiten**  
Montag bis Freitag von 8:30 bis 15:35 Uhr  
(in Wochen mit Feiertagen von 8:30 bis 17:10 Uhr)

 **Nächste Kursstarts**  
14.10.2024  
11.11.2024  
09.12.2024

## LEHRGANGSZIEL

Du beherrschst die Prozesse rund um die Zusammenführung, Aufbereitung, Anreicherung und Weitergabe von Daten. Außerdem kannst du große, unstrukturierte Datenmengen mit Hilfe von branchenspezifischer Software verarbeiten. Du verfügst über Kenntnisse im Framework Apache und weißt, wie Daten ansprechend visualisiert werden.

Des Weiteren verstehst du die Grundlagen der Statistik, kannst Daten aufbereiten, auswerten sowie statistische Datenanalysen und Ergebnisse mit Grafiken darstellen, erläutern und interpretieren.

## ZIELGRUPPE

Der Lehrgang richtet sich an Personen mit abgeschlossenem Studium in der Informatik, Wirtschaftsinformatik, BWL, Mathematik oder vergleichbarer Qualifikation.

## BERUFSAUSSICHTEN

Big Data wird in Unternehmen zur interdisziplinären Analyse und Konzeption von IT-Lösungen in Zusammenarbeit mit Entwicklungs- und Betriebsteams eingesetzt. Big Data Engineers sind sowohl bei großen als auch mittelständischen Unternehmen in Industrie, Handel, Dienstleistungs- und Finanzwesen nachgefragt.

Fundierte Kenntnisse in der Statistik stellen eine wertvolle Zusatzqualifikation dar, die in der industriellen Forschung und Entwicklung, in der Arzneimittelentwicklung, in der Betreuung medizinischer Studien, im Bereich Finanz- und Versicherungswesen, in der Informationstechnologie oder in der öffentlichen Verwaltung sehr gefragt. Dein aussagekräftiges Zertifikat gibt detaillierten Einblick in deine erworbenen Qualifikationen und verbessert deine beruflichen Chancen.

## VORAUSSETZUNGEN

Für diesen Lehrgang werden Programmierkenntnisse (idealerweise Python) und Erfahrungen mit Datenbanken (SQL) vorausgesetzt.

## LEHRGANGSINHALTE

### DATA ENGINEER

#### Grundlagen Business Intelligence (ca. 2 Tage)

Anwendungsfelder, Dimensionen einer BI Architektur  
Grundlagen Business Intelligence, OLAP, OLTP, Aufgaben der Data Engineers  
Data Warehousing (DWH): Umgang und Verarbeitung von strukturierten, semi-strukturierten und unstrukturierten Daten

#### Anforderungsmanagement (ca. 2 Tage)

Aufgaben, Ziele und Vorgehensweise in der Anforderungsanalyse  
Datenmodellierung, Einführung/Modellierung mit ERM  
Einführung/Modellierung in der UML  
· Klassendiagramme  
· Use-Case Analyse  
· Aktivitätsdiagramme

#### Künstliche Intelligenz (KI) im Arbeitsprozess

Vorstellung von konkreten KI-Technologien im beruflichen Umfeld  
Anwendungsmöglichkeiten und Praxis-Übungen

#### Datenbanken (ca. 3 Tage)

Grundlagen von Datenbanksystemen  
Architektur von Datenbankmanagementsystemen  
Anwendung RDBMS  
Umsetzung Datenmodell in RDBMS, Normalformen  
Praktische und theoretische Einführung in SQL  
Grenzen von Relationalen Datenbanken, csv, json

## Data Warehouse (ca. 4 Tage)

Star Schema  
Datenmodellierung  
Erstellung Star Schema in RDBMS  
Snowflake Schema, Grundlagen, Datenmodellierung  
Erstellung Snowflake Schema in RDBMS  
Galaxy Schema: Grundlagen, Datenmodellierung  
Slowly Changing Dimension Tables Typ 1 bis 5 – Restating, Stacking, Reorganizing, mini Dimension und Typ 5  
Einführung in normal, causal, mini und monster, heterogeneous und sub Dimensions  
Vergleich von state und transaction oriented Faktentabellen, Density und Storage vom DWH

## ETL (ca. 4 Tage)

Data Cleansing  
· Null Values  
· Aufbereitung von Daten  
· Harmonisierung von Daten  
· Anwendung von Regular Expressions  
Data Understanding  
· Datenvalidierung  
· Statistische Datenanalyse  
Datenschutz, Datensicherheit  
Praktischer Aufbau von ETL-Strecken  
Data Vault 2.0, Grundlagen, Hubs, Links, Satellites, Hash Key, Hash Diff.  
Data Vault Datenmodellierung  
Praktischer Aufbau eines Data Vault Modells – Raw Vault, Praktische Umsetzung von Hash-Verfahren

## Projektarbeit (ca. 5 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

---

## BIG DATA SPECIALIST

### Was ist Big Data? (ca. 1 Tag)

Volume, Velocity, Variety, Value, Veracity  
Chancen und Risiken großer Datenmengen  
Abgrenzung: Business Intelligence, Data Analytics, Data Science  
Was ist Data Mining?

### Einführung in Apache Frameworks (ca. 2 Tage)

Big-Data-Lösungen in der Cloud  
Datenzugriffsmuster  
Datenspeicherung

### MapReduce (ca. 3 Tage)

MapReduce Philosophie  
Hadoop Cluster  
Verkettung von MapReduce Jobs

### Künstliche Intelligenz (KI) im Arbeitsprozess

Vorstellung von konkreten KI-Technologien im beruflichen Umfeld  
Anwendungsmöglichkeiten und Praxis-Übungen

### Komponenten (ca. 3 Tage)

Kurzvorstellung von verschiedenen Tools  
Datenübertragung  
YARN-Anwendungen  
Hadoop JAVA-API  
Apache Spark

### NoSQL und HBase (ca. 3 Tage)

CAP-Theorem  
ACID und BASE  
Typen von Datenbanken  
HBase

## Big Data Visualisierung (ca. 3 Tage)

Theorien der Visualisierung  
Diagrammauswahl  
Neue Diagrammarten  
Werkzeuge zur Datenvisualisierung

## Projektarbeit (ca. 5 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

---

## STATISTIK

### Statistische Grundlagen (ca. 6 Tage)

Messtheoretische Grundlagen (Grundgesamtheit und Stichprobe, Stichprobenarten, Messung und Skalenniveaus)  
Univariate Deskriptivstatistik (Häufigkeitsverteilungen, Zentralmaße, Streuungsmaße, Standardwert, Histogramme, Balkendiagramme, Kreisdiagramme, Liniendiagramme und Boxplots)  
Bivariate Deskriptivstatistik (Zusammenhangsmaße, Korrelationskoeffizienten, Kreuztabellen, Streudiagramme und gruppierte Balkendiagramme)  
Grundlagen der induktiven Inferenzstatistik (Wahrscheinlichkeitsverteilung, Normalverteilung, Mittelwertverteilung, Signifikanztest, Nullhypothesentest nach Fisher, Effektgröße, Parameterschätzung, Konfidenzintervalle, Fehlerbalkendiagramme, Poweranalysen und Ermittlung des optimalen Stichprobenumfangs)

### Künstliche Intelligenz (KI) im Arbeitsprozess

Vorstellung von konkreten KI-Technologien im beruflichen Umfeld  
Anwendungsmöglichkeiten und Praxis-Übungen

### Methoden zum Vergleich von zwei Gruppen (ca. 5 Tage)

z- und t-Test für eine Stichprobe (Abweichung von einem vorgegebenen Wert)  
t-Test für den Mittelwertsunterschied von zwei unabhängigen/verbundenen Stichproben  
Prüfung der Wirksamkeit von Aktionen, Maßnahmen, Interventionen und anderen Veränderungen mit t-Tests (Pretest-Posttest-Designs mit zwei Gruppen)  
Unterstützende Signifikanztests (Anderson-Darling-Test, Ryan-Joiner-Test, Levene-Test, Bonnet-Test, Signifikanztest für Korrelationen)  
Nonparametrische Verfahren (Wilcoxon-Test, Vorzeichentest, Mann-Whitney-Test)  
Kontingenzanalysen (Binomialtest, Exakter Test nach Fisher, Chi-Quadrat-Test, Kreuztabellen mit Assoziationsmaße)

### Methoden zum Mittelwertvergleich von mehreren Gruppen (ca. 5 Tage)

Ein- und zweifaktorielle Varianzanalyse (einfache und balancierte ANOVA)  
Mehrfaktorielle Varianzanalyse (Allgemeines lineares Modell)  
Feste, zufällige, gekreuzte und geschachtelte Faktoren  
Mehrfachvergleichsverfahren (Tukey-HSD, Dunnett, Hsu-MCB, Games-Howell)  
Interaktionsanalyse (Analyse von Wechselwirkungseffekten)  
Trennschärfe und Poweranalyse bei Varianzanalysen

### Einführung in die Versuchsplanung (DoE, Design of Experiments) (ca. 1 Tag)

Voll- und teilfaktorielle Versuchspläne

### Projektarbeit (ca. 3 Tage)

Zur Vertiefung der gelernten Inhalte  
Präsentation der Projektergebnisse

## UNTERRICHTSKONZEPT

### Didaktisches Konzept

Deine Dozierenden sind sowohl fachlich als auch didaktisch hoch qualifiziert und werden dich vom ersten bis zum letzten Tag unterrichten (kein Selbstlernsystem).

Du lernst in effektiven Kleingruppen. Die Kurse bestehen in der Regel aus 6 bis 25 Teilnehmenden. Der allgemeine Unterricht wird in allen Kursmodulen durch zahlreiche praxisbezogene Übungen ergänzt. Die Übungsphase ist ein wichtiger Bestandteil des Unterrichts, denn in dieser Zeit verarbeitest du das neu Erlernte und erlangst Sicherheit und Routine in der Anwendung. Im letzten Abschnitt des Lehrgangs findet eine Projektarbeit, eine Fallstudie oder eine Abschlussprüfung statt.

### Virtueller Klassenraum alfaview®

Der Unterricht findet über die moderne Videotechnik alfaview® statt - entweder bequem von zu Hause oder bei uns im Bildungszentrum. Über alfaview® kann sich der gesamte Kurs face-to-face sehen, in lippensynchroner Sprachqualität miteinander kommunizieren und an gemeinsamen Projekten arbeiten. Du kannst selbstverständlich auch deine zugeschalteten Trainer:innen jederzeit live sehen, mit diesen sprechen und du wirst während der gesamten Kursdauer von deinen Dozierenden in Echtzeit unterrichtet. Der Unterricht ist kein E-Learning, sondern echter Live-Präsenzunterricht über Videotechnik.

## FÖRDERMÖGLICHKEITEN

Alle Lehrgänge werden von der Agentur für Arbeit gefördert und sind nach der Zulassungsverordnung AZAV zertifiziert. Bei der Einreichung eines Bildungsgutscheines oder eines Aktivierungs- und Vermittlungsgutscheines werden in der Regel die gesamten Lehrgangskosten von Ihrer Förderstelle übernommen.

Eine Förderung ist auch über den Europäischen Sozialfonds (ESF), die Deutsche Rentenversicherung (DRV) oder über regionale Förderprogramme möglich. Als Zeitsoldat:in besteht die Möglichkeit, Weiterbildungen über den Berufsförderungsdienst (BFD) zu besuchen. Auch Firmen können ihre Mitarbeiter:innen über eine Förderung der Agentur für Arbeit (Qualifizierungschancengesetz) qualifizieren lassen.

- ① Änderungen möglich. Die Lehrgangsinhalte werden regelmäßig aktualisiert. Die aktuellen Lehrgangsinhalte findest Du immer unter [smartbuilding.alfatraining.de](https://smartbuilding.alfatraining.de).