

BIG DATA - DATA ENGINEERING GRUNDLAGEN

Big Data-Architektur, -Aufbau und die Komponenten

Dauer: 3 Tage

Nr.: 54010

Preis: 1.890€ netto / 2.249,10 € inkl. 19 % MwSt.

Durchführungsart: Präsenztraining

Zielgruppe: Das Seminar richtet sich an Datenbank Experten, BigData Experten, Sicherheitsexperten, Systemadministratoren, Netzwerkadministratoren, Data Science Mitarbeiter, Entwickler sowie alle Interessierten, die dem Thema BigData tiefer unter die Haube schauen möchten.

Schulungsmethode: Die Inhalte werden durch Vorträge, Diskussionen, Demonstrationen und Praktikum am System vermittelt.

Voraussetzungen: Neben generellen IT Kenntnissen in SQL und Programmierung sind keine besonderen Vorkenntnisse erforderlich. Der vorherige Besuch unseres Grundlagenseminars 4704 erleichtert das Verständnis.

Das Thema Big Data ist aktueller denn je, pro Tag werde viele Petabyte an Daten generiert und warten darauf analysiert und verarbeitet zu werden.

Mit dem Apache Hadoop Ökosystem wurde eine Möglichkeit geschaffen, große Datenmengen in einem verteilten System zu persistieren und Analysen und Vorhersagen auszuführen.

Die damit verbundenen Begriffe rund um Data Science und Data Engineering sind allgegenwärtig. Nur was verbirgt sich hinter all den Trendbegriffen?

Die Schulung gibt einen ersten Ein- und Überblick über die Komponenten, Methoden und Begrifflichkeiten zu diesen Themen. Dieser beschränkt sich nicht auf reine Theorie, sondern ermöglicht Ihnen erste praktische Erfahrungen anhand eines fiktiven Projekts zu sammeln.

Nach einer Einführung in die Paradigmen von Big Data lernen Sie anhand eines realen Twitter Datensets, wie einfach Sie Daten in den Hadoop Cluster streamen können. Danach nehmen Sie eigenhändig Daten-Transformationen vor und orientieren sich dabei immer am besprochenen Data Lake Konzept. Am Ende des Data Engineering Parts haben Sie die Daten soweit verdichtet und aufbereitet, dass Sie im nächsten Schritt mit Data Science Methoden Analysen darauf ausführen.

Programm

Data Engineering

Einführung Big Data:

- Was ist Big Data?
- Die drei Vs
- Abgrenzung zu relationalen Datenbanken
- Data Lake und hybride Architektur Konzept

IT Grundlagen:

- Basiswissen Netzwerke
- Basiswissen verteilte Systeme

Ökosystem Hadoop:

- Hadoop Cluster Architektur
- Überblick über die Hadoop Distributionen
- HBase
- HDFS
- Hive
- Kafka
- NiFi
- Pig
- Spark / Spark Streaming
- Sqoop
- Ambari
- Zookeeper
- Überblick weiterer Komponenten

Security in Hadoop:

- Ranger
- Kerberos

Hands on Hadoop - Praxisteil:

- HDFS Puzzle
- Administration Grundlagen
- Ambari Monitoring M/R
- HDFS Strukturen anlegen
- Zookeeper Nodes
- HBase
- HDFS
- Hive
- Kafka
- NiFi
- Pig
- Spark / Spark Streaming
- Sqoop

Vertiefung der Hadoop Komponenten - Praxisteil:

- Streaming
- Buffering
- Archivierung
- Transformation
- Analyse

Termine und Orte - Nr.: 54010

Düsseldorf

14 Aug - 16 Aug 2019

18 Nov - 20 Nov 2019

Frankfurt

18 Sep - 20 Sep 2019

München

16 Oct - 18 Oct 2019

Stuttgart

09 Dec - 11 Dec 2019

Online Anmeldung:

Kundenservice | Tel. 0711 62010 100 | Fax: 0711 62010 267 | seminaranmeldung@integrata.de

<https://www.integrata.de/54010>

19/06/2019