

Microsoft Azure bietet eine Vielzahl von Daten-Analyse Lösungen in der Cloud an.

Ihr Nutzen

Dieses Seminar richtet sich an Daten-Analysten, die mit HDInsight Cluster, Spark, Stream Analytics Big-Data Workflows implementieren möchten.

Preis pro Teilnehmer

EUR 2360,- exklusive der gesetzlichen MwSt.

Seminardauer

4 Tag(e)/Day(s)

Seminarinhalte

1. Tag

- * Einführung in HDInsight
 - Was ist Big Data?
 - Einführung in Hadoop
 - Arbeiten mit der MapReduce Funktion
 - Einführung in HDInsight
- * Bereitstellen von HDInsight
 - HDInsight Cluster Types
 - Verwaltung über das Azure Portal und Azure PowerShell
- * Benutzer und Berechtigungen
 - Domain-Joined und Non-Domain Joined Clusters
 - Verwaltung über die Azure PowerShell
 - Das Ambari Management UI
 - Das Ranger Admin UI
 - Hive Policies
- * Daten in HDInsight laden
 - Speicher für HDInsight
 - Azure Blob und Azure Data Lake Storage
 - Verwenden der Data Loading Tools

2. Tag

- * Troubleshooting und Logging in HDInsight
 - HDInsight Logs
 - YARN Logs
 - Heap Dumps
 - OMS
- * Batch Solutions
 - Apache Hive Storage
 - Data Queries mit Hive und Pig
- * Batch ETL Solutions mit Spark
 - Was ist Spark?
 - ETL mit Spark

3. Tag

- * Daten mit Spark SQL analysieren

Voraussetzungen

Kenntnisse des Windows Betriebssystem von Datenanalyse und Statistik und Kenntnisse der Programmiersprache R.

Hinweise

DP203-T00,

Version: N/A

- Interative und interaktive Abfragen
- Verwenden von Zeppelin und Livy
- * Daten mit Hive und Phoenix analysieren
- Interaktive Abfragen mit interactive hive
- Interactive Processing mit Apache Phoenix
- 4. Tag
- * Stream Analytics
 - Verarbeiten von Streaming Daten
 - Verwalten von Stream Analytics Jobs
- * Implementierung von Streaming Solutions
 - Aufbau und Bereitstellung eines Kafka Clusters
 - Storm Cluster, Kafka Producer und Power BI Dashboards
 - Publizieren und Nutzen des Kafka Clusters
 - Verwenden von HBase für Storage und Abfrage
- 5. Tag
- * Real Time Processing Solutions mit Apache Storm
 - Streaming von Daten mit Storm
 - Storm Topologies
- * Spark Streaming Applications erstellen
 - Arbeiten mit Spark Streaming
 - Spark Structured Streaming Applications
 - Erstellen einer Spark Streaming Pipeline
 - Verwenden von DStreams

