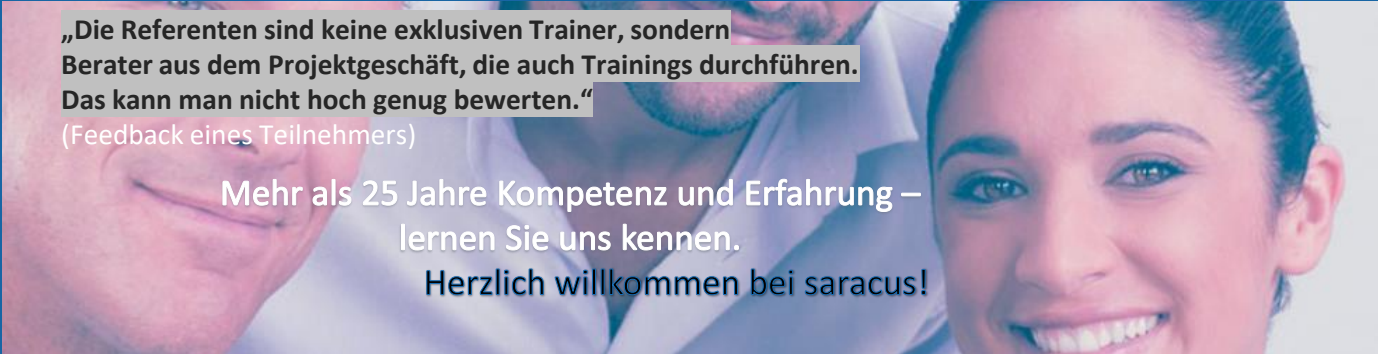


# Hadoop Administrator Training

## Viertägiges Intensivseminar



„Die Referenten sind keine exklusiven Trainer, sondern Berater aus dem Projektgeschäft, die auch Trainings durchführen. Das kann man nicht hoch genug bewerten.“

(Feedback eines Teilnehmers)

Mehr als 25 Jahre Kompetenz und Erfahrung –  
lernen Sie uns kennen.  
Herzlich willkommen bei saracus!

Das Ziel dieses viertägigen Seminars ist die Vermittlung von praktischen und theoretischem Wissen, welches für den Einsatz eines Hadoop Clusters benötigt wird. Es werden unter anderem Installationsschritte behandelt sowie die Anpassung von einer Vielzahl von Konfigurationsparametern. Die Teilnehmer lernen, wie man ein Hadoop Cluster wartet und überwacht, auftretende Fehlermeldungen bearbeitet und praktische Lösungen dafür findet. Der Schwerpunkt liegt hierbei insbesondere auf den praktischen Übungen, die das Ziel haben, dass die Teilnehmer nach dem Seminar als Hadoop Administratoren arbeiten können.

Während des Seminars werden die Teilnehmer in die Rolle des Hadoop Administrators versetzt, der für 6 Maschinen, die in einer Cloud gehalten werden, zuständig ist. Das Ziel ist es, ein Hadoop Cluster mit mehreren Server-Knoten (Multi-Node) zu installieren und entsprechend zu konfigurieren. Das Cluster muss voll funktionsfähig sein und so konfiguriert sein, dass es unterschiedlichen Ausfällen standhält. Neben den Tätigkeiten wie auf Alerts zu antworten und die Konfiguration zu optimieren, sind die Teilnehmer für typische Wartungsaufgaben wie Hinzufügen von neuen Diensten, Verschieben von Diensten über die Knoten, Hinzufügen von neuen Knoten, Tuning des YARN-Schedulers sowie die Visualisierung von Metriken etc verantwortlich.

### An wen richtet sich das Seminar?

Das Seminar richtet sich an IT-Mitarbeiter, die für die Installation, Konfiguration und Wartung eines Hadoop Clusters zuständig sind.

### Voraussetzungen

Wünschenswert ist Basiswissen von Linux Systemen. Erfahrung im Bereich Hadoop ist nicht notwendig.

# Agenda - Tag 1 (Basiswissen Hadoop, Wartung, Betrieb)

**09.00 Uhr**

**Begrüßung durch die Seminarleitung**

**09.15 Uhr – 17.00 Uhr**

## **Einführung**

### **Überblick: HDFS**

- Grundlegende Schritte wie z.B. Schreiben/Lesen von Files, Replikation, Metadaten und Datenblöcke
- Daemons und Cluster Infrastruktur
- Grundlegende Eigenschaften und Use Case

### **Überblick: YARN**

- Einführung
- Daemons und Cluster Infrastruktur
  
- Übung: Installation der zentralen Komponenten eines Hadoop Clusters
- Übung: Überprüfung einer HDFS Installation
- Übung: Überprüfung einer YARN Installation
- Übung: Erweitern des Clusters
- Übung: Ausbalancierung von HDFS
- Übung: Konfigurationsänderungen

# Agenda - Tag 2 (Hadoop für Fortgeschrittene)

## **Administrative Aspekte von HDFS**

- NameNode Internia wie z.B. Metadaten-Management, Startup Prozedur, Checkpointing
- Wichtige Konfigurationseinstellungen
- Übung: Wechsel der Java-Heap-Größe, Restart eines NameNode, Überprüfung des Checkpointing Status, Restart eines DataNodes

## **Administrative Aspekte von YARN**

- Cluster Ressourcen wie z.B. Container Sizing, Limitationen und Management
- Wichtige Konfigurationseinstellungen
- Übung: Überprüfung und Tuning von ressourcenbezogene Einstellungen wie vcores und RAM.

## **Überblick auf Projekte eines Hadoop Ecosystems**

- Transfer von Daten zum HDFS mit Sqoop
- Verarbeitung von Daten in einem Hadoop Cluster mit Hive
- Analyse mit Spark

**Kaffee- und Teepausen finden nach Bedarf statt!**

# Agenda - Tag 3 (Sicherheit und hohe Verfügbarkeit)

## Hadoop Sicherheit

- Authentifizierung mit Kerberos
- Autorisierung für Hadoop (z.B. Sentry, Ranger)
- Sicherheits-bezogene Features

## Backup und Wiederherstellung im Notfall

### Fortgeschrittene Konfigurationseinstellungen für die Cluster-Verfügbarkeit & Performance Hohe Verfügbarkeit für Hadoop Komponenten wie z.B. HDFS, YARN.

- Übungen: Aktivierung eines NameNode HA und Überprüfung der Korrektheit
- Übungen: Migration eines NameNode auf einen anderen Host
- Bonus-Übungen: Aktivierung und Überprüfung von ResourceManager HA

# Agenda - Tag 4 (Wartung, Betrieb und Tuning)

## Hardware- und Software-Auswahl für Hadoop Cluster

### YARN Scheduler

- Übung: Konfiguration von multi-tenant Queues und ACLs in dem Scheduler

### Monitoring und Alerting

- Übung: Erstellen von benutzerdefinierten Diagrammen, Dashboards und Alerts

### Wartungstätigkeiten des Clusters

- Übungen: Erweitern des Clusters, Ausbalancierung von HDFS, Konfigurationsänderungen, Troubleshooting von einer Spark Applikation

## Backup und Wiederherstellung im Notfall

# Anpassungsmöglichkeiten im Rahmen des Seminars

Aufgrund der praktischen Erfahrung mit Cloudera oder Hortonworks Distributionen ist es möglich, ein auf die Teilnehmer zugeschnittenes Seminar anzubieten, in dem auch die Agenda entsprechend den Kunden-Anforderungen an das eigene Hadoop Cluster angepasst werden kann.

### Möglich sind die folgenden Anpassungen:

- HDP (Apache Ambari) oder CDH (Cloudera Manager)
- Hinzufügen einiger Komponenten: Cloudera Impala, Apache Tez, Facebook Presto, Apache Flume, Apache Kafka, Apache Sentry, Apache Ranger, Search (Apache Solr)
- Übungen für den Capacity Scheduler oder den Fair Scheduler

# Anmeldung zum Seminar



- Ja**, ich nehme an dem Seminar  
„Hadoop Administrator Training (viertägig)“ teil.

Firma: \_\_\_\_\_

Name: \_\_\_\_\_ Vorname: \_\_\_\_\_

Funktion: \_\_\_\_\_

Straße: \_\_\_\_\_

PLZ/Ort: \_\_\_\_\_

Tel.-Nr.: \_\_\_\_\_ Fax-Nr.: \_\_\_\_\_

Email: \_\_\_\_\_

Datum: \_\_\_\_\_ Unterschrift: \_\_\_\_\_

**Die Seminargebühr für dieses viertägige Seminar beträgt 3.200 Euro in Deutschland / 3.800 CHF in der Schweiz pro Teilnehmer.**

**Gebuchter Seminartermin (bitte angeben):** \_\_\_\_\_

Ihre Anmeldung sollte spätestens 14 Tage vor dem Seminartermin erfolgen.

**Gebuchter Seminarort (bitte kreuzen Sie an):**

- Baden-Dättwil(CH)  Münster (D)  
 Ich benötige ein Hotel!

Nach Ablauf der Anmeldefrist erhalten Sie von uns eine Anmeldebestätigung und eine Rechnung sowie eine Anfahrtsskizze. Bitte überweisen Sie den Rechnungsbetrag vor Veranstaltungsbeginn.

**Bitte senden Sie diese Seite per Fax an +49 (0251) 98721-26  
oder per Post an: saracus consulting GmbH  
„Hadoop Administrator Training (viertägig)“  
Hafenweg 46, D- 48155 Münster**

**Voraussetzung:**

Berücksichtigen Sie bitte, dass eine Mindestteilnehmerzahl von 4 Personen vorgesehen ist. Ansonsten kann das Seminar nicht stattfinden.

**Anmeldung:**

Bitte nehmen Sie Ihre Abmeldung schriftlich vor. Selbstverständlich ist eine Vertretung des angemeldeten Teilnehmers möglich. Alle genannten Preise verstehen sich zzgl. der gesetzl. MwSt.

**Stornierung:**

Bei Stornierung bis zu 14 Tage vor Veranstaltungsbeginn erheben wir eine Bearbeitungsgebühr von 50% der Gebühr. Bei späteren Absagen wird der gesamte Beitrag fällig.

**Seminarabsage:**

Sollten wir aus wichtigen Grund (z. B. Erkrankung des/der Referenten) gezwungen sein, den Kurs abzusagen, so teilen wir Ihnen dieses umgehend mit. Wir werden Ihnen in diesem Fall einen Ersatztermin anbieten. Passt dieser nicht zu Ihrer Terminplanung, erhalten Sie die bereits gezahlte Seminargebühr in voller Höhe zurück. Darüber hinausgehende Ansprüche bestehen nicht.



**Für Rückfragen steht Ihnen gerne das Team der Data & Analytics academy der saracus consulting unter +49 (0251) 98721-0 zur Verfügung. Email: [saracus@saracus.com](mailto:saracus@saracus.com) Internet: <https://www.saracus.com/know-how/data-and-analytics-academy/>**