



Schulung

## Hadoop Grundlagen

powered by Lufthansa Industry Solutions

Gerade im Kontext von Big Data gewinnt Hadoop an Bedeutung, da damit intensive Rechenprozesse mit großen Datenmengen auf Computerclustern durchgeführt werden können. In der zweitägigen Hadoop Grundlagenschulung werden den Teilnehmern Entwicklungen im Bereich Big Data und diverse Technologien vorgestellt, die unter dem Schlagwort Hadoop zusammengefasst werden. Durch das praxisnahe Konzept sammeln die Teilnehmer erste Erfahrungen mit dem Hadoop Distributed File System, MapReduce, Spark, Hive oder dem Resource Manager Yarn. Dadurch lernen sie, wie die verschiedenen Technologien im Hadoop Ökosystem gemeinsam einen Mehrwert für Unternehmen bieten. Der Überblick reicht von den technologischen Plattformen, auf denen Hadoop Cluster aufsetzen, über virtuelle Maschinen, Netzwerktopologien bis hin zu verschiedenen Speichersystemen. Die wichtigen Themen Sicherheit und Wartung sind ebenso Bestandteil dieser Schulung.

### Kontakt und Buchung

Lufthansa Industry Solutions  
E-Mail: [schulungen@lhind.dlh.de](mailto:schulungen@lhind.dlh.de)  
[www.lhind.de/schulungen](http://www.lhind.de/schulungen)

### Seminarziele

Durch den hohen Praxisanteil lernen die Teilnehmer die Hadoop Technologien kennen, erhalten einen Überblick über die verschiedenen Möglichkeiten und sammeln viele hilfreiche Anwendungserfahrungen.

Nach dem Seminar können sie Daten mit Flume oder Sqoop in ein Hadoop Cluster importieren. Sie wissen, wie sie Daten mit Spark, HBase oder Hive verarbeiten und auch wieder exportieren. Sie können mit dem Hadoop File System, MapReduce, Hive und dem Resource Manager Yarn umgehen. Die einheitliche Verwaltung einer Vielzahl von Applikationen mit Zookeeper ist ihnen geläufig.

Übergeordnetes Ziel dieses Seminars ist, dass sie weitere Schritte in der Big-Data-Welt auch allein bewerkstelligen können.

### Teilnehmer

- IT-Architekten
- Applikations-Entwickler/-Verantwortliche
- Datenbank-Entwickler und Administratoren

### Teilnahmevoraussetzungen

Grundkenntnisse in Linux und SQL



## Hadoop Grundlagen

### Modulübersicht

Modulname	Beschreibung
<b>Tag 1: Hadoop Grundlagen</b>	
<b>Warum Hadoop?</b>	<ul style="list-style-type: none"> <li>typische Anwendungsfälle</li> <li>von Google's MapReduce Framework zu Apache Hadoop</li> <li>Datenlokalität</li> <li>„Design for failure“</li> </ul>
<b>Anatomie eines Hadoop 1.0 Clusters</b>	<ul style="list-style-type: none"> <li>Name Nodes, Data Nodes, Job Tracker, Task Tracker</li> <li>HDFS</li> </ul>
<b>Von Hadoop 1.0 zu Hadoop 2.0</b>	<ul style="list-style-type: none"> <li>Resource Management und Scheduling</li> <li>Hadoop 2.0 mit YARN</li> <li>YARN vs. Mesos</li> </ul>
	<ul style="list-style-type: none"> <li>HA in Hadoop 2.0</li> </ul>
<b>Hadoop Jobs</b>	<ul style="list-style-type: none"> <li>Ausführung eines MapReduce Jobs</li> <li>Ausführung eines Spark Jobs</li> </ul>
<b>Hadoop Ökosystem</b>	<ul style="list-style-type: none"> <li>Zookeeper</li> <li>Hive und Pig</li> <li>Sqoop und Flume</li> <li>Oozie</li> <li>HBase</li> </ul>
<b>Datenimport und Datenexport</b>	<ul style="list-style-type: none"> <li>Sqoop</li> <li>Flume</li> <li>Kafka</li> </ul>
<b>Hands-On-Training (Tag 1)</b>	<ul style="list-style-type: none"> <li>Daten in und aus dem HDFS kopieren</li> <li>Übungen mit Hive</li> <li>Submitting und Monitoring von MR und Spark Job</li> </ul>
<b>Tag 2: Planung, Installation und Betrieb</b>	
<b>Planung</b>	<ul style="list-style-type: none"> <li>typische Big Data Use Cases</li> <li>Maschinen</li> <li>Netzwerk</li> <li>HDs/SSDs und RAID</li> <li>Betriebssystem</li> </ul>
<b>Aufbau und Installation</b>	<ul style="list-style-type: none"> <li>Installation</li> <li>Cluster-Management</li> </ul>
<b>Konfiguration</b>	<ul style="list-style-type: none"> <li>Hadoop Konfigurationsdateien</li> <li>HDFS Konfiguration</li> <li>YARN Konfiguration</li> <li>Konfiguration der Clients</li> </ul>
	<ul style="list-style-type: none"> <li>Hadoop Secure Mode</li> </ul>
<b>Operations und Maintenance</b>	<ul style="list-style-type: none"> <li>Scheduling von Jobs</li> <li>Monitoring</li> <li>Logging</li> <li>Patches und Upgrades einspielen</li> </ul>
<b>Performance-Tuning</b>	<ul style="list-style-type: none"> <li>JVMs</li> <li>Hadoop Parameter</li> <li>Einfluss der Client-Parameter auf Performance</li> </ul>
<b>Hands-On-Training (Tag 2)</b>	<ul style="list-style-type: none"> <li>Installation eines kleinen Hadoop-Systems</li> <li>Simulation von Server-Ausfällen</li> </ul>

#### Methoden

Präsentationen, Anwendungsbeispiele, Diskussionen und praktische Übungen im System

#### Termine, Ort und Dauer

- flexible Termine nach Ihrem Bedarf
- InHouse-Training bei Ihnen vor Ort
- 2 Tage

#### Teilnahmegebühr

auf Anfrage

**Individuelle Änderungen bei Schulungsinhalten und Ablauf sind auf Kundenwunsch möglich.**