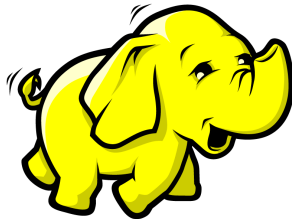


Trainings

Administration von Apache Hadoop Clustern



Das Aufsetzen und Betreiben von Apache Hadoop Clustern ist eine komplexe Aufgabe. Dieses Training vermittelt das Wissen darüber auf sehr praktische Art und Weise.

Dauer: 3 Tage

Zielgruppe: Linux-Administrator:innen, Software-Entwickler:innen (mit grundlegenden Linux-Kenntnissen)

In den letzten Jahren hat sich Hadoop als der Standard für Big Data Systeme durchgesetzt. Aufsetzen und Betreiben von Hadoop Clustern ist jedoch eine wesentlich komplexere Aufgabe als etwa Aufsatz und Betrieb von Webserver-Farmen oder auch klassischen Datenbank-Clustern. Die hohe Vernetzung der Hadoop-Knoten und das „Eventual Consistency“ Paradigma, erfordern tiefgründiges Wissen über Architektur und die wesentlichen Prozesse bei Hadoop Clustern, um einen hochverfügbaren, performanten und sicheren Betrieb zu gewährleisten.

Dieses Training vermittelt eben dieses Wissen auf sehr praktische Art und Weise: Die Teilnehmer:innen bauen während des Trainings einen eigenen Hadoop Cluster auf, mit welchem dann typische Szenarien wie die Anbindung von Drittsystemen, die Erstellung eines Backups oder der Umgang mit einem ausgefallenen Knoten geübt werden.

Agenda:

- Hadoop Grundlagen, Infrastruktur und Architektur
- Einrichtung und Konfiguration eines Hadoop Clusters
- Hadoop Security-Grundlagen
- Hochverfügbarkeit für Hadoop Cluster
- Backup- und Recovery-Strategien für Hadoop Cluster

- Vorstellung des Ökosystems und Integration häufig verwendeter Komponenten
- Monitoring und laufende Optimierung eines Hadoop Clusters
- Hadoop Logfile-Analyse