

# Projektskizze

von

Steffen Hafner & Daniel Landler-Gärtner

## Projekttitel

„Analyse von Log-Files eines Web-Servers mithilfe/innerhalb eines Hadoop Clusters“

## Einleitung

Die Verfügbarkeit von Daten hat sich in den vergangenen zehn Jahren drastisch verändert. Die Anzahl verschiedener Datenquellen steigt stetig durch die zunehmende Verbreitung mobiler und internetfähiger Geräte. Dadurch sehen sich Unternehmen heute mit sehr viel größeren Datenmengen konfrontiert. Diese gilt es zu erfassen, zu speichern und auszuwerten. Dabei ist es nicht nur die Datenmenge selbst, die den Unternehmen Probleme bereitet, sondern darüber hinaus auch die Struktur und die Art der Daten, sowie die Geschwindigkeit, mit der sie anfallen.

Der Begriff Big-Data ist in den letzten Jahren vom bloßen Buzz-Word hin zu einem greifbaren technischen Begriff gereift. Hadoop und NoSQL-Technologien haben maßgeblich zu dieser Evolution beigetragen.

## Problem

Bereits vor dem Aufkommen von Big-Data haben Unternehmen Log-Files zur Gewinnung von Einblicken genutzt. Jedoch ist das Problem, dass durch das exponentielle Wachstum aller Datenquellen die Verwaltung und Analyse der Log-Files zu einer immer größeren Herausforderung wird. Log-Files enthalten eine große Menge an Informationen, wobei nicht alle für den jeweiligen Betreiber von gleicher Bedeutung sind. Zudem liegen die Informationen innerhalb der Log-Files in keinem gut leserlichen Format vor, weshalb eine Analyse von relevanten Informationen und Korrelationen sehr aufwendig ist.

Für die Veranschaulichung des genannten Problems verwenden wir die Log-Files des Webserver der Hochschule Ravensburg-Weingarten.

## Anforderungsanalyse

