

# Management Summary

FINISHED ▶ 🔍 📖 ⚙️

Zielgruppe: Top Management eines KMUs | Der Vorteil der nachfolgend genutzten Technologien (z. B. Hadoop Distributed File System (HDFS) oder Spark) ist eine Zeiteinsparung bei der Berechnung der Abfragen durch parallele Verarbeitung sowie eine schnellere Time2Market. Für das bessere Verständnis für die parallele Verarbeitung verwende ich folgendes Beispiel:

Sie müssen die Buchstaben a und e der pro Harry Potter Buch lesen.  
Alleine haben Sie dazu wahrscheinlich einige Tage bis Wochen.  
Sie erhalten CHF 10'000 wenn sie es innerhalb eines Tages schaffen.  
Sie mobilisieren alle ihre Freunde und jeder führt den Task für zugewiesene Seiten durch.  
Die Anzahl e und a tragen die einzelnen Personen in einer zentralen Excel-Liste ein und  
mittels einer einfachen Summe schaffen Sie es innerhalb von 4 Stunden die Aufgabe zu meistern.  
Das war ein guter Stundenlohn ;-)

Ähnlich wie in diesem Beispiel funktioniert das verteilte Rechnen auf Hadoop. Sie erkennen, dass Sie einerseits viel schneller im Erreichen der Aufgaben und dadurch schneller in der Entscheidungsfindung sind. Weitere Vorteile von Hadoop sind: Skalierbarkeit, Flexibilität, Fehlerresistenz.

Die nachfolgende Arbeit zeigt eine End-zu-End-Lösung für Hadoop ausgehend von Fragestellungen im Hausmietpreisbereich in den USA auf. Die Basis dafür legt Systemarchitektur und die eingesetzten Technologien. Die Daten werden entsprechend ins HDFS gelesen und in eine Datenbank abgelegt. Die Analyse erfolgte direkt über die HDFS-Datei und/oder über die angelegte Datenbank. Dank dem verteilten Rechnen sind die Ergebnisse schnell erzielt. Die Bewältigung der vielen Technologien und deren Einsatzzweck im jeweiligen Anwendungsfall stellt eine Herausforderung dar. Nach einer herausfordernden Einarbeitungszeit können Data Engineers und Data Scientists schnell die investierte Zeit durch Effizienzsteigerung einsparen. Lessons Learned: Anwender sollen sich auf die wesentlichen Punkte konzentrieren und sich schnell darin einarbeiten. Es ist leicht möglich sich im Hadoop-Technologien-Dschungel zu verirren.

Als Aufpassfeld sind sicherlich die Gesamtkosten zu berücksichtigen. Gerade KMU können

Took 0 sec. Last updated by admin at February 14 2020, 8:05:34 PM.