

# Big Data Visualisierung

---

Stephan Lex

13. März 2015

## TECHNOLOGIERECHERCHE

Im Rahmen der Evaluierung möglicher Technologien zur Analyse und Visualisierung von Big Data wird im folgenden eine Technologierecherche vorgenommen.

### 1 TECHNISCHE ANFORDERUNGEN

- Verwaltung sehr großer Datenmengen
- Analyse dieser Datenmengen
- Visualisierung der Relationen innerhalb der Menge
- Interpretation und Filterung einzelner relevanten Teile
- Unabhängiges Querying
- Simples und ansprechendes Design für den Endnutzer

### 2 MÖGLICHE TECHNOLOGIEN (WEBRECHERCHE)

Wenn Daten zu groß und kompliziert werden, um sie mit einem einzelnen Rechner zu bearbeiten, ist es unabdingbar von herkömmlichen Systemen abzuweichen und andere Technologien in Betracht zu ziehen. Der momentan dominierende Ansatz zur Verarbeitung solcher Datenmengen etwa bietet der MapReduce Ansatz<sup>1</sup>. Allerdings gibt es heutzutage unzählige Tools und Frameworks die eine Implementierung des MapReduce Ansatzes oder alternativer

---

<sup>1</sup><http://www.pal-blog.de/entwicklung/perl/einfach-erklart-mapreduct-tutorial.html>

Technologien bereitstellen<sup>2</sup>. Im folgenden werden die präsentesten Ansätze evaluiert.

Durch Recherche im Web wurden folgende Ansätze ermittelt:

- Apache Hadoop <sup>3</sup>
- Apache Spark <sup>4</sup>
- Disco Project <sup>5</sup>
- HPCC Systems <sup>6</sup>

### 3 BEWERTUNGSKRITERIEN

Um eine passende Auswahl zu treffen müssen die gewählten Ansätze hinsichtlich einiger Kriterien verglichen werden. Im Fall der Big Data Visualisierung sind das primär:

- Effizienz
- Zukunftssicherheit
- verfügbare Tools
- Verständlichkeit und Einarbeitungsaufwand
- Instandhaltungskosten

---

<sup>2</sup><http://www.bytemining.com/2011/08/hadoop-fatigue-alternatives-to-hadoop/>

<sup>3</sup><http://hadoop.apache.org/>

<sup>4</sup><https://spark.apache.org/>

<sup>5</sup><http://discoproject.org/>

<sup>6</sup><http://hpccsystems.com/>