

Big Data Visualisierung

Stephan Lex

8. März 2015

PROBLEMSTELLUNG

Um eine große Menge an Daten korrekt verarbeiten zu können, bietet sich momentan eine Lösung im Rahmen von Apache Hadoop an. Allerdings muss auch hier eine Auswahl an Tools getroffen werden, um die verschiedenen Teilbereiche zu decken.

IMPORTER

Hier bietet sich Apache Sqoop¹ an. Sqoop stellt ein Framework dar, das einen effizienten Ansatz bietet Daten von einem RDBMS in das Hadoop-eigene Dateisystem HDFS zu importieren. Semi-structured-data hingegen kann relativ simpel mit der Hadoop Core API oder top-level Projekte wie Hive importiert werden.

STORAGE

Da der Ansatz mit Hadoop gewählt wurde, wird hier das Hadoop Distributed File System (HDFS)² verwendet. Da HDFS nach dem Master-Slave Prinzip aufgebaut ist, ist die Skalierbarkeit sehr groß. So können Cluster auf benötigte und beinahe beliebige Größe ausgedehnt werden. Das von HDFS in der Regel verwendete OS ist Linux.

¹<http://sqoop.apache.org/>

²<http://hortonworks.com/hadoop/hdfs/>

FILTERING UND ZUGRIFF

Um die zu verwaltenden Datensätze analysieren zu können gibt es eigene Toolsets wie Hive³ und HBase⁴. Diese bieten benutzerfreundlichen Umgang mit der Big Data auf Basis von HDFS und ermöglichen individuelle Abhandlungen der spezifischen Anforderungen an das System, wie z.B. die wichtigen MapReduce Jobs. In diesem Bereich gibt es auch diverse Distributionen verschiedener Firmen, die viele nützliche Tools im Umgang mit Hadoop unter einen Hut bringen, wie etwa Hortonworks⁵.

VISUALISIERUNG

Im Bereich der Visualisierung der gewonnenen Daten, gibt es bereits mehrere Softwarelösungen, die auf dem Markt etabliert sind. Diese können zumeist variabel beschaffene und große Daten importieren sowie verarbeiten und anschließend per benutzerfreundlicher GUI (Drag and Drop) visualisiert werden. Zu den bekanntesten darunter zählen momentan Qlik Sense⁶ und Tableau⁷

³<https://hive.apache.org/>

⁴<http://hbase.apache.org/>

⁵<http://hortonworks.com/>

⁶<http://www.qlik.com/>

⁷<http://www.tableau.com/de-de>