

Übungsblatt 6

Aufgabenblock: Big Data Analytics

Aufgabe 1

- A. Erläutern Sie das Funktionsprinzip eines künstlichen neuronalen Netzes (KNN) an einem Anwendungsbeispiel Ihrer Wahl. Gehen Sie dabei auf das Perzeptron-Modell ein.
- B. Erläutern Sie die sich durch Big Data verändernden Anforderungen an Datenbanksysteme.
- C. Erläutern Sie, welche Schwächen relationale Datenbanksysteme in Hinblick auf die neuen Anforderungen in Big Data haben.
- D. Stellen Sie das ACID-Kriterium relationaler Datenbanken und das BASE-Kriterium nicht-relationaler Datenbanken gegenüber.

Aufgabe 2

- A. Erläutern Sie das Funktionsprinzip von Key-Value Stores.
- B. Beschreiben Sie die Architektur von Apache Hadoop und gehen Sie dabei kurz auf die Funktionen der einzelnen Systembestandteile ein.
- C. Worin unterscheidet sich Apache Spark von MapReduce?
- D. Eine typische Dateisystemblockgröße beträgt 4096 Bytes. Ein Block in HDFS ist hingegen z.B. 64 oder 128 Megabytes groß. Warum ist dies so? Nennen Sie zwei Vorteile größerer Blöcke.
- E. Wie wachsen die Hardware-Kosten als Funktion der Datenmenge, wenn in einem verteilten Dateisystem wie HDFS gespeichert wird? Warum?
- F. Welche Komponente ist der single point of failure in Hadoop?
- G. Erläutern Sie, wie das Dateisystem HDFS von Hadoop die folgenden Anforderungen erfüllt:
 - 1. Skalierbarkeit
 - 2. Haltbarkeit
 - 3. Hohe sequentielle Lese- / Schreibleistung
- H. Vergleichen Sie Objektspeicher (OO-DB), Blockspeicher (rel. DB) und Key-Value-Speicher. Geben Sie für jeden der folgenden Anwendungsfälle an, welche Technologie besser zu den Anforderungen passt.
 - 1. Speichern von Netflix-Filmdateien, sodass sie von vielen Client-Anwendungen aus gleichzeitig zugänglich sind.
 - 2. Speichern der Buchungen für innerdeutsche Flüge.
 - 3. Speichern einer Datenbank für Browser-Cookies.
 - 4. Speichern der Auto-Backups von iPhone / Android-Geräten.