

Übungsblatt 4 Stand: 18.09.2014

Bearbeiten Sie alle Aufgaben aller Übungsblätter – soweit nicht anders angegeben – in festen **Zweiergruppen**. Beide Gruppenmitglieder müssen die Aufgabe gemeinsam bearbeiten und die Lösung alleine erläutern können.

Aufgabe 10 (Wide-Column Datenbank)

Zum Abschluss des Datenbankteils des Praktikums eine letzte Aufgabe (bis auf "b") alles wie immer):

- Importieren sie die Daten aus „plz.data“ (siehe EMIL) in HBase. Die ZIP soll den Schlüssel bilden und alle sonstigen Daten in **eine** ColumnFamily eingefügt werden.
- Ergänzen sie eine weitere ColumnFamily "Fussball", in der für alle Städte mit dem Namen "HAMBURG" und "BREMEN" der String "ja" eingefügt wird, während er für alle anderen Städte leer bleiben soll.
- Schreiben sie ein Programm in einer beliebigen Sprache, das auf ihre Datenbank zugreift und für eine beliebige interaktiv eingegebene PLZ den zugehörigen Ort und den Staat ausgibt.
- Erweitern sie ihr Programm: Es soll jetzt die PLZ für eine interaktiv einzugebende Stadt bestimmen. Testen sie das Programm mit den Städten „TUMTUM“ und „HAMBURG“.
- Vergleichen Sie den Aufwand für die Realisierung des Programms (LoC?, Arbeitszeit?) und die Ausführungszeit des Programms (messen?) mit dem entsprechenden Programm für die Key-Value- und die dokumentenorientierte Datenbank.

Zeigen sie mir beim Termin doch bitte ihr Programm, und die Ergebnisse des Vergleichs. Eine GUI ist hier sicherlich sinnvoll, für die PVL aber nicht unbedingt erforderlich.

P.S. Jetzt haben sie Programme für die vier grundverschiedenen Arten von NoSQL-Datenbanksystemen. Lehnen Sie sich doch bitte kurz zurück und betrachten mit etwas Abstand die Eignung der vier Systeme für unterschiedliche Aufgabenszenarien.

Aufgabe 11 (Map-Reduce)

Installieren Sie Hadoop in ihrer virtuellen Maschine (das gilt als nicht ganz einfach). Bringen sie anschließend ein Map-Reduce-Programm zum Wörterzählen zum Ablauf. Das war dann schon der "Pflichtteil" dieser Aufgabe!

Als freiwillige "Kür": Sie können gerne einen Datenbankanschluss (Daten aus HBase lesen) realisieren (TableInputFormat) oder ein wirklich interessantes Map-Reduce-Programm schreiben! Wann, wenn nicht jetzt?