

## **Praktikum 6 (Freiwilliges Zusatzpraktikum): Datenanalyse mit Pig**

### **Ziel des Praktikums**

Ziel dieses Praktikums ist es, sich mit der Arbeit mit Pig vertraut zu machen und Unterschiede bei der Verwendung von nativen MapReduce vs. Pig zu analysieren.

**Für die erfolgreiche Bearbeitung der Zusatzaufgabe können Sie bis zu 3 Bonuspunkte für die Klausur erhalten.**

### **Aufgabe 1 (Datenbankabfragen mit Pig)**

Informationen zur Arbeit mit Pig finden Sie im Wiki.

Realisieren Sie jetzt die Datenbankabfragen aus dem 4. Praktikum mit Pig auf Hadoop und vergleichen Sie geeignet die Zeiten mit denen der nativen MapReduce-Jobs aus dem letzten Praktikum. Dokumentieren Sie dabei, wie die Zeiten jeweils ermittelt wurden.

Interpretieren Sie die Ihre Messergebnisse. Was könnte der Grund für die unterschiedlichen Zeiten zwischen nativen MapReduce und Pig sein?

Was fällt Ihnen bezüglich der Implementierung der Datenbankabfragen im Vergleich zur nativen MapReduce-Implementierung auf?

Tipp: Nutzen Sie den Operator `describe` für Zwischenschritte, um sich das Schema ihrer jeweiligen Relation ausgeben zu lassen.

### **Praktikumsbericht**

Abgabe des Praktikumsberichts bis zum 5. Juli.