

Übungen zur Vorlesung
Spezielle Gebiete zum Software Engineering
Sommersemester 2015

Aufgabenblatt 4: MongoDB

Aufgabe 4.1: JSON-Import und Query-Konvertierung von SQL nach MongoDB (5 Punkte)

Im ILIAS findet sich eine Text-Datei, in der eine Menge von Städten mit weiteren Metadaten im JSON-Format gespeichert ist. Importieren Sie diese Datenmenge in eine Collection mit dem Namen „zips“.

Weiter unten sehen Sie eine Menge von SQL-Queries, die Abfragen auf diese Datenmenge deklarieren. Formulieren Sie diese SQL-Queries in der Query-Syntax von MongoDB und zeigen Sie, dass die umformulierten Queries in der Mongo-Shell lauffähig sind und Resultate liefern:

- `INSERT INTO zips VALUES ('Minden', 25000)`
- `SELECT * FROM zips`
- `SELECT city, pop FROM zips`
- `SELECT city, pop FROM zips WHERE pop = 33`
- `SELECT * FROM zips WHERE pop > 33`
- `SELECT * FROM zips WHERE pop <= 33`
- `SELECT * FROM zips WHERE pop > 33 AND pop < 40`
- `SELECT * FROM zips WHERE pop = 32 AND city = 'Bob'`
- `SELECT * FROM zips WHERE pop = 33 OR city = 'Bob'`
- `SELECT * FROM zips WHERE pop = 33 ORDER BY city ASC`
- `SELECT * FROM zips ORDER BY city DESC`
- `SELECT * FROM zips WHERE city LIKE '%inde%'`
- `SELECT * FROM zips WHERE city LIKE 'inde%'`
- `SELECT * FROM zips LIMIT 10 SKIP 20`
- `SELECT * FROM zips LIMIT 1`
- `SELECT DISTINCT city FROM zips`
- `SELECT COUNT(*) FROM zips`
- `SELECT COUNT(*) FROM zips WHERE pop > 30`
- `SELECT COUNT(pop) FROM zips`
- `UPDATE zips SET pop = 33 WHERE city = 'Minden'`
- `UPDATE zips SET pop = pop + 2 WHERE city = 'Bob'`
- `DELETE FROM zips WHERE city = 'Minden'`
- `CREATE INDEX ON zips (city ASC)`
- `CREATE INDEX ON zips (city ASC, pop DESC)`
- `EXPLAIN SELECT * FROM zips WHERE pop = 32`

Aufgabe 4.2: Formulierung von Queries für E-Mail-Collection (2 Punkte)

Im ILIAS findet sich eine Text-Datei, in der eine Menge von E-Mails im JSON-Format gespeichert ist. Importieren Sie diese Nachrichten in eine MongoDB-Collection. Schreiben Sie dann Queries, mit denen Sie folgende Anfragen beantworten können:

1. Wie lauten die eindeutigen Namen der „Folder“ der E-Mails?
2. Wieviel E-Mails wurden mit der Adresse „rosalee.fleming@enron.com“ gesendet?
3. Was ist die E-Mail mit der höchsten Anzahl an Rezipienten und was ist der Nachrichtentext der E-Mail?
4. Wieviel E-Mails enthalten genau einen Sender und genau einen Rezipienten?
5. Wieviel E-Mails wurde im Jahr 2001 versendet?

Aufgabe 4.3: Aggregation mit MapReduce (3 Punkte)

Jedes Dokument der E-Mail-Collection enthält ein Feld für den Sender und die Rezipienten. Ziel dieser Aufgabe ist es, mit dem MapReduce-Verfahren die Menge der eindeutigen Sender/Rezipienten-Paare zu ermitteln (Map) und zu zählen, wie oft diese Paare vorhanden sind (Reduce). Geben Sie die Liste geordnet aus, so dass das Paar, das am häufigsten auftaucht, zuerst ausgegeben wird.