

PolyGlot-Database Performance

Benchmark Framework - MongoDB vs Neo4J

Hyeon Ung Kim, Tim Niehoff

6. August 2018

Aufgabenstellung

- Verschiedene DB Typen mit unterschiedlichen Vorteilen¹

¹Inhalt der Folie von

https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf

- Verschiedene DB Typen mit unterschiedlichen Vorteilen¹
 - Relational: Sicherheit, homogene Daten

¹Inhalt der Folie von

https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf

- Verschiedene DB Typen mit unterschiedlichen Vorteilen¹
 - Relational: Sicherheit, homogene Daten
 - Document: Flexibles Schema, Suchfunktionen

¹Inhalt der Folie von

https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf

- Verschiedene DB Typen mit unterschiedlichen Vorteilen¹
 - Relational: Sicherheit, homogene Daten
 - Document: Flexibles Schema, Suchfunktionen
 - Graph: Beziehungen, Traversal

¹Inhalt der Folie von

https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf

- Verschiedene DB Typen mit unterschiedlichen Vorteilen¹
 - Relational: Sicherheit, homogene Daten
 - Document: Flexibles Schema, Suchfunktionen
 - Graph: Beziehungen, Traversal
- Polyglot: Verwendung mehrerer DB-Typen für untersch. Anwendungsfälle

¹Inhalt der Folie von

https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf

- Verschiedene DB Typen mit unterschiedlichen Vorteilen¹
 - Relational: Sicherheit, homogene Daten
 - Document: Flexibles Schema, Suchfunktionen
 - Graph: Beziehungen, Traversal
- Polyglot: Verwendung mehrerer DB-Typen für untersch. Anwendungsfälle
- Aufgabe: Vergleich einer Graphdatenbank mit einer Dokumenten-Datenbank

¹Inhalt der Folie von

https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf

gegebene Werkzeuge



**Herausforderung: Gleicher
Datensatz auf beiden DB**

Herausforderung: Gleicher Datensatz auf beiden DB

**Mongo Connector + Neo4j Doc
Manager**

Mongo Connector + Neo4j Doc Manager

- Software jeweils von MongoDB und Neo4j

Mongo Connector + Neo4j Doc Manager

- Software jeweils von MongoDB und Neo4j
- Dauerhafte Synchronisation zwischen Mongo und Neo4j möglich

Mongo Connector + Neo4j Doc Manager

- Software jeweils von MongoDB und Neo4j
- Dauerhafte Synchronisation zwischen Mongo und Neo4j möglich
- Generisches Erstellen des Neo4j Datenmodells

Herausforderung: Gleicher Datensatz auf beiden DB

Apoc

- Software von Neo4j zum Import von JSONs in Neo4j

- Software von Neo4j zum Import von JSONs in Neo4j
- keine automatische Synchronisierung zwischen Mongo und Neo4j

- Software von Neo4j zum Import von JSONs in Neo4j
- keine automatische Synchronisierung zwischen Mongo und Neo4j
- kein generisches Datenmodell

- Software von Neo4j zum Import von JSONs in Neo4j
- keine automatische Synchronisierung zwischen Mongo und Neo4j
- kein generisches Datenmodell
- Vorteil des nutzerspezifischen Datenmodells: Ausnutzen graphdatenbankspezifischer Performanzvorteile

Resultat: PolyGDBP

Resultat: PolyGDBP

Features

- Command line Interface mit Parametereingabe (Query, Serveradressen, Datensatz, ...)

- Command line Interface mit Parametereingabe (Query, Serveradressen, Datensatz, ...)
- Obligatorisch nur Queryangabe

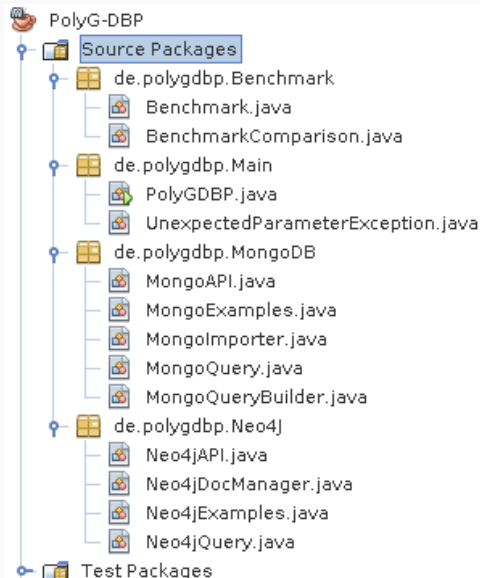
- Command line Interface mit Parametereingabe (Query, Serveradressen, Datensatz, ...)
- Obligatorisch nur Queryangabe
- Vorgefertigte Queries für Yelp Datensatz, ansonsten Angabe benutzerspezifischer Queries

- Command line Interface mit Parametereingabe (Query, Serveradressen, Datensatz, ...)
- Obligatorisch nur Queryangabe
- Vorgefertigte Queries für Yelp Datensatz, ansonsten Angabe benutzerspezifischer Queries
- Reduzieren und Einlesen des Datensatzes möglich (via MongoAPI und MongoConnector)

- Command line Interface mit Parametereingabe (Query, Serveradressen, Datensatz, ...)
- Obligatorisch nur Queryangabe
- Vorgefertigte Queries für Yelp Datensatz, ansonsten Angabe benutzerspezifischer Queries
- Reduzieren und Einlesen des Datensatzes möglich (via MongoAPI und MongoConnector)
- Vor und nach jedem Schritt Stoppen der Zeit, Logging, Filewriting

- Command line Interface mit Parametereingabe (Query, Serveradressen, Datensatz, ...)
- Obligatorisch nur Queryangabe
- Vorgefertigte Queries für Yelp Datensatz, ansonsten Angabe benutzerspezifischer Queries
- Reduzieren und Einlesen des Datensatzes möglich (via MongoAPI und MongoConnector)
- Vor und nach jedem Schritt Stoppen der Zeit, Logging, Filewriting
- Visualisierung mittels Javascript Anwendung möglich

Programmstruktur



Tests mit dem Yelp Dataset

Tests mit dem Yelp Dataset

Datensatz + DatenModell

- Yelp = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte

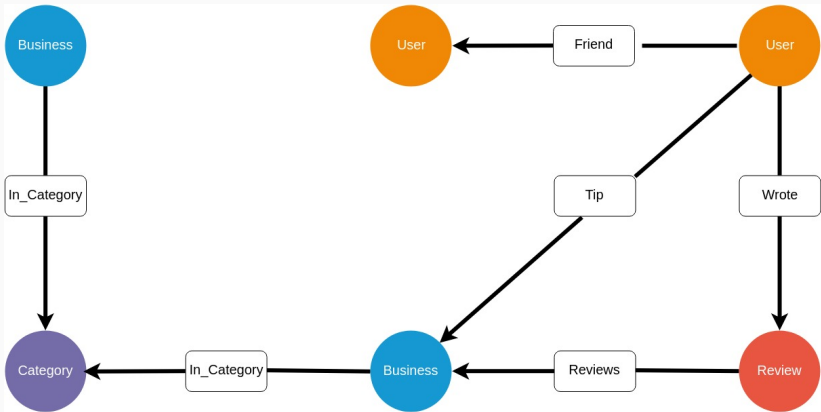
- Yelp = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte
- Datensatz $\geq 6,5$ GB

- Yelp = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte
- Datensatz $\geq 6,5$ GB
- besteht aus 6 *.jsons: business, checkin, photos, review, tip, user

- Yelp = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte
- Datensatz $\geq 6,5$ GB
- besteht aus 6 *.jsons: business, checkin, photos, review, tip, user

```
{
  "business_id": "FYWN1wneV18bWNgQjJ2GNg",
  "name": "Dental by Design",
  "neighborhood": "",
  "address": "4855 E Warner Rd, Ste B9",
  "city": "Ahwatukee",
  "state": "AZ",
  "postal_code": "85044",
  "latitude": 33.3306902,
  "longitude": -111.9785992,
  "stars": 4.0,
  "review_count": 22,
  "is_open": 1,
  "attributes": {
    "AcceptsInsurance": true,
    "ByAppointmentOnly": true,
    "BusinessAcceptsCreditCards": true
  },
  "categories": [
    "Dentists",
    "General Dentistry",
    "Health & Medical",
    "Oral Surgeons",
    "Cosmetic Dentists",
    "Orthodontists"
  ],
  "hours": {
    "Friday": "7:30-17:00",
    "Tuesday": "7:30-17:00",
    "Thursday": "7:30-17:00",
    "Wednesday": "7:30-17:00",
    "Monday": "7:30-17:00"
  }
}
```

Neo4j Datenmodell



Tests mit dem Yelp Dataset

Ergebnisse

Zusammenfassung

- Verschiedene DB-Typen mit versch. Vor- und Nachteile

Zusammenfassung

- Verschiedene DB-Typen mit versch. Vor- und Nachteile
- PolyG-DBP = Framework zum Vergleich einer Dokument-DB mit einer Graphdatenbank

Zusammenfassung

- Verschiedene DB-Typen mit versch. Vor- und Nachteile
- PolyG-DBP = Framework zum Vergleich einer Dokument-DB mit einer Graphdatenbank
- konkret: MongoDB und Neo4j werden getestet

Zusammenfassung

- Verschiedene DB-Typen mit versch. Vor- und Nachteile
- PolyG-DBP = Framework zum Vergleich einer Dokument-DB mit einer Graphdatenbank
- konkret: MongoDB und Neo4j werden getestet
- Ausführung von prebuilt Queries oder nutzerspezifischen Queries

Zusammenfassung

- Verschiedene DB-Typen mit versch. Vor- und Nachteile
- PolyG-DBP = Framework zum Vergleich einer Dokument-DB mit einer Graphdatenbank
- konkret: MongoDB und Neo4j werden getestet
- Ausführung von prebuilt Queries oder nutzerspezifischen Queries
- Messen und Vergleich der Query-Ausführungszeiten

Zusammenfassung

- Verschiedene DB-Typen mit versch. Vor- und Nachteile
- PolyG-DBP = Framework zum Vergleich einer Dokument-DB mit einer Graphdatenbank
- konkret: MongoDB und Neo4j werden getestet
- Ausführung von prebuilt Queries oder nutzerspezifischen Queries
- Messen und Vergleich der Query-Ausführungszeiten
- Vergleich Neo4j und MongoDB mit Yelp Datensatz: MongoDB performanter bei einfachen Queries, Neo4j siegt bei Queries mit mehreren Relationen/Collections

Fragen und Diskussion