

# PolyGlot-Database Performance

Benchmark Framework - MongoDB vs Neo4J

---

Hyeon Ung Kim, Tim Niehoff

6. August 2018

# Aufgabenstellung

---

- Verschiedene DB Typen mit unterschiedlichen Vorteilen<sup>1</sup>

---

<sup>1</sup>Inhalt der Folie von

[https://dbs.uni-leipzig.de/file/Intro\\_bdprak\\_final.pdf](https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf)

- Verschiedene DB Typen mit unterschiedlichen Vorteilen<sup>1</sup>
  - Relational: Sicherheit, homogene Daten

---

<sup>1</sup>Inhalt der Folie von

[https://dbs.uni-leipzig.de/file/Intro\\_bdprak\\_final.pdf](https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf)

- Verschiedene DB Typen mit unterschiedlichen Vorteilen<sup>1</sup>
  - Relational: Sicherheit, homogene Daten
  - Document: Flexibles Schema, Suchfunktionen

---

<sup>1</sup>Inhalt der Folie von

[https://dbs.uni-leipzig.de/file/Intro\\_bdprak\\_final.pdf](https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf)

- Verschiedene DB Typen mit unterschiedlichen Vorteilen<sup>1</sup>
  - Relational: Sicherheit, homogene Daten
  - Document: Flexibles Schema, Suchfunktionen
  - Graph: Beziehungen, Traversal

---

<sup>1</sup>Inhalt der Folie von

[https://dbs.uni-leipzig.de/file/Intro\\_bdprak\\_final.pdf](https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf)

- Verschiedene DB Typen mit unterschiedlichen Vorteilen<sup>1</sup>
  - Relational: Sicherheit, homogene Daten
  - Document: Flexibles Schema, Suchfunktionen
  - Graph: Beziehungen, Traversal
- Polyglot: Verwendung mehrerer DB-Typen für untersch. Anwendungsfälle

---

<sup>1</sup>Inhalt der Folie von

[https://dbs.uni-leipzig.de/file/Intro\\_bdprak\\_final.pdf](https://dbs.uni-leipzig.de/file/Intro_bdprak_final.pdf)

- Verschiedene DB Typen mit unterschiedlichen Vorteilen<sup>1</sup>
  - 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

---

<sup>1</sup>Inhalt der Folie von

[https://dbs.uni-leipzig.de/file/Intro\\_bdprak\\_final.pdf](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**

---

## **Pipline und Workflow**

- File Struktur (Main, Mongo, Neo4j, Benchmark)

- File Struktur (Main, Mongo, Neo4j, Benchmark)
- Command line Interface

- File Struktur (Main, Mongo, Neo4j, Benchmark)
- Command line Interface
- nur Queryangabe ist pflicht, alles andere optional.



- File Struktur (Main, Mongo, Neo4j, Benchmark)
- Command line Interface
- nur Queryangabe ist pflicht, alles andere optional.
- gibt vorgefertigte Queries für Yelp Datensatz

- File Struktur (Main, Mongo, Neo4j, Benchmark)
- Command line Interface
- nur Queryangabe ist pflicht, alles andere optional.
- gibt vorgefertigte Queries für Yelp Datensatz
- Nutzer kann JSON Datensatz importieren

- File Struktur (Main, Mongo, Neo4j, Benchmark)
- Command line Interface
- nur Queryangabe ist pflicht, alles andere optional.
- gibt vorgefertigte Queries für Yelp Datensatz
- Nutzer kann JSON Datensatz importieren
- Wird in MongoCollections verarbeitet und durch MongoConnector erfolgt danach das Übertragen in Neo4j

- File Struktur (Main, Mongo, Neo4j, Benchmark)
- Command line Interface
- nur Queryangabe ist pflicht, alles andere optional.
- gibt vorgefertigte Queries für Yelp Datensatz
- Nutzer kann JSON Datensatz importieren
- Wird in MongoCollections verarbeitet und durch MongoConnector erfolgt danach das Übertragen in Neo4j
- Queries werden ausgeführt

- File Struktur (Main, Mongo, Neo4j, Benchmark)
- Command line Interface
- nur Queryangabe ist pflicht, alles andere optional.
- gibt vorgefertigte Queries für Yelp Datensatz
- Nutzer kann JSON Datensatz importieren
- Wird in MongoCollections verarbeitet und durch MongoConnector erfolgt danach das Übertragen in Neo4j
- Queries werden ausgeführt
- vor und jedem dieser Schritte zeit gestoppt, geloggt und schließlich ausgegeben

- File Struktur (Main, Mongo, Neo4j, Benchmark)
- Command line Interface
- nur Queryangabe ist pflicht, alles andere optional.
- gibt vorgefertigte Queries für Yelp Datensatz
- Nutzer kann JSON Datensatz importieren
- Wird in MongoCollections verarbeitet und durch MongoConnector erfolgt danach das Übertragen in Neo4j
- Queries werden ausgeführt
- vor und jedem dieser Schritte zeit gestoppt, geloggt und schließlich ausgegeben
- kann mithilfe unserer Javascript Anwendung visualisiert werden.

**Lets run it. With the help of  
YELP**

---

**Lets run it. With the help of  
YELP**

---

**Datensatz + DatenModell**



- YELP = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte

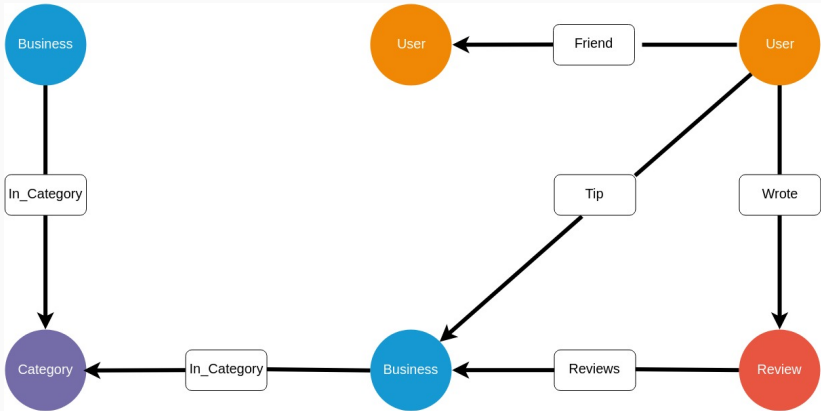
- YELP = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte
- YELP Datensatz X GB groß

- YELP = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte
- YELP Datensatz X GB groß
- besteht aus X .jsons: users, business, review, ...

- YELP = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte
- YELP Datensatz X GB groß
- besteht aus X .jsons: users, business, review, ...
- Beispiele eines JSON Business Objektes:

- YELP = Suchmaschine und Empfehlungsportal für Restaurants und Geschäfte
- YELP Datensatz X GB groß
- besteht aus X .jsons: users, business, review, ...
- Beispiele eines JSON Business Objektes:
  - ...

# Neo4j Datenmodell



**Lets run it. With the help of  
YELP**

---

**Ergebnisse**

# **Zusammenfassung**

---



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

# Zusammenfassung

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

# Zusammenfassung

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

# Zusammenfassung

- Verschiedene DB-Typen haben versch. Vor- und Nachteile
- PolyG-DBP ist ein Framework zum Vergleich einer Dokument-DB mit einer Graphdatenbank
- konkret: MongoDB und Neo4j werden getestet
- Nutzer kann prebuilt queries ausführen lassen oder eigene Queries testen lassen

# Zusammenfassung

- Verschiedene DB-Typen haben versch. Vor- und Nachteile
- PolyG-DBP ist ein Framework zum Vergleich einer Dokument-DB mit einer Graphdatenbank
- konkret: MongoDB und Neo4j werden getestet
- Nutzer kann prebuilt queries ausführen lassen oder eigene Queries testen lassen
- Ausführungszeiten der Queries werden gemessen und verglichen

# Zusammenfassung

- Verschiedene DB-Typen haben versch. Vor- und Nachteile
- PolyG-DBP ist ein Framework zum Vergleich einer Dokument-DB mit einer Graphdatenbank
- konkret: MongoDB und Neo4j werden getestet
- Nutzer kann prebuilt queries ausführen lassen oder eigene Queries testen lassen
- Ausführungszeiten der Queries werden gemessen und verglichen
- Unsere Tests mithilfe von PolyG-DBP und Yelp zeigen: Neo4j und MongoDB siegen bei bestimmten Arten von Queries

# Fragen und Diskussion