

“Lessons Learned” NoSQL Social Network

Für die Abgabe entschieden wir uns für ArangoDB, ein verteiltes Multimodel DBMS, über das Datenhaltung über eine Mischung aus dokumentenbasiert und Graph abgebildet wird. Das Aufsetzen von mehreren ArangoDB über Docker war nachvollziehbar über die Arango Docs (<https://www.arangodb.com/docs/stable/tutorials-starter.html>). Eine ArangoDB Cluster besteht dabei mindestens aus jeweils aus einem Agent, einem Coordinator und einem DB-Server (in unserem Fall zwei).

- Exposen der Ports von Docker-Containern
- große Datenmengen (101M Einträge) führte oft zum Überlaufen des Speicherbereich. Die Erstellung und Speicherung der Daten musste Sequenzweise erstellt werden. Nutzen von Asynchronen Funktionen führte meist dazu, dass die Daten schneller erstellt wurden, als sie von der DB gespeichert wurden => Arbeitsspeicher wurde immer voller)
- Queries von ArangoDB waren mit AQL ziemlich einfach zu verstehen
- Edges machen Verlinkungen sehr einfach
- Redundantes Abspeichern für schnellere Abfragen notwendig (z.B. über das Attribut Likes vom Tweets suchen anstatt über die Likes-Edges direkt)
- Datentypen mussten beachtet werden (Likes waren mal als String und mal als Number gespeichert worden, was die Suche nach dem beliebtesten Tweet fehlerhaft machte)

Im Endeffekt waren wir mit unserer Wahl des DBMS sehr zufrieden.