TU München, Fakultät für Informatik
Lehrstuhl III: Datenbanksysteme
Prof. Dr. Thomas Neumann

Exercises for *Foundations in Data Engineering*, WiSe 23/24

Alexander Beischl, Maximilian Reif (i3fde@in.tum.de)
http://db.in.tum.de/teaching/ws2324/foundationsde

Sheet Nr. 13

Exercise 1 Apache Cassandra is a recently popular no-SQL database system. It is built to be very be very scalable and available, sacrificing consistency if needbe. For the following questions, get familiar with the database system: http://docs.datastax.com/en/cassandra/3.0/

1. Which ACID guarantees does Cassandra provide?
2. Which guarantees of a classic ACID database system does Cassandra not provide? What can be gained by ommiting those?
3. What is Cassandra's scale-out strategy?
4. What is Cassandra's strategy to handle node failures?

Exercise 2 Familiarize yourself with how data is represented and stored in Cassandra: documentation.

Now create a data model for a Twitter like application that has users, a friends relation and a follower relation between users. When designing the tables, make sure that these queries can be answered efficiently:

1. List all friends of one user
2. List all followers of one user

Explain why in your solution the queries can be efficiently processed in Cassandra, even when there are hundreds of thousands of users with each having many friends and followers.

Exercise 3 Cassandra uses tunable consistency among replications to achieve fault tolerance in the cluster. On a per-query basis, the consistency level can be chosen. See here for a list of available levels. (If needbe, refresh your memory of quorum consensus here.)

1. What is the difference between replicas and partitions in Cassandra?
2. Which read write consistency mode combinations produce consistent answers?

Exercise 4 Semantic Web, RDF and Sparql

Complementary to the contents of the lecture, you may want to read the following sections of "Programming the Semantic Web" by Toby Segaran, Colin Evans and Jamie Taylor: I1. Why Semantics?, I2 Expressing Meaning, II4 Just Enough RDF - The RDF Data Model It is available online in full text via the TUM library.

Equiped with this information, you are ready to explore wikidata.org. It is a project which provides structured information to Wikimedia sister projects. Find information about the data model at https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer. Conveniently, it also offers a SPARQL interface for your explorations at query.wikidata.org.

Write SPARQL queries to answer the following questions:

1. List everything that uses Munich as object. Wikidata gave the URI http://www.wikidata.org/entity/Q1726 to munich. Therefore, using a prefix definition, you may refer to Munich by using wd:Q1726.
2. Which predicate is used most?
3. Which of the cities in the database has the earliest written record?
4. List the transitive subclasses of sport (Q349).
5. List the subclasses of sport (Q349) and their labels if there exists one.