Lessons Learned Gruppe A

- Man sollte sich erst informieren, ob ein Cluster mit der kostenlosen Version einer Datenbank möglich ist. Bei neo4j ist das nicht der Fall.
- Datenmodellierung für Cassandra ist unintuitiv und schwierig, wenn man relationale Datenbanken gewohnt ist. Dafür ist es sehr einfach, ein Cluster hochzuziehen. Auch das Schreiben von Abfragen ist einfach, solange man die Tabellen entsprechend konzipiert.
- Neo4j lässt sich sehr einfach aufsetzen, Datenmodellierung ist relativ einfach, Abfragen und Importieren von Daten auch. Die Performance ist auch sehr gut, selbst das Importieren von mehreren Millionen Beziehungen ist in absehbarer Zeit erledigt. Auch Abfragen über viele Beziehungen brauchen nicht viel Zeit.
- Scheinbar kleine Features wie Likes können sich als ganz schön aufwändig herausstellen und ein Umplanen der Architektur erfordern. Denn bei der Umsetzung von Likes muss bedacht werden, dass jeder Nutzer einen Post liken kann oder auch nicht. Hieraus ergibt sich kein einfacher Zähler, der hochzählt, sondern eine Beziehung zwischen einem Nutzer und einem Post.
- Das Importieren von Daten kann viele Probleme bereithalten:
    - Das Importieren von CSV-Dateien wird komplizierter, wenn ein Feld Zeilenumbrüche, Kommata und/oder Anführungszeichen enthält. Dies kann die gesamte Formatierung zerschießen und zu unsinnigen Einträgen in der Datenbank führen, wenn es sich überhaupt importieren lässt. Nur weil ein Programm wie Excel mit diesen Fällen richtig umgehen kann, bedeutet das nicht, dass eine Datenbank wie zum Beispiel Cassandra diese genauso interpretiert.
    - Der COPY-Befehl zum Import von CSV-Dateien in Cassandra ist ein reiner CQLSH-Befehl, kein CQL-Befehl. Bis zu dieser Erkenntnis kann einige Zeit verstreichen, wenn man den Fehler in dem geschriebenen Befehl sucht und nicht auf die Idee kommt, dass die Umgebung das Problem ist.
    - Gerade Zeilenumbrüche in CSV-Dateien können beim Vorverarbeiten von Daten zu unvorhergesehenen Problemen führen, die nicht direkt offensichtlich sind. Zwar lassen sich diese damit beheben, dass man ein Feld mit Zeilenumbrüchen in Anführungszeichen setzt, jedoch kommen dann Probleme mit bereits existenten Anführungszeichen in dem Feld auf. Diese lassen sich zwar mit einem weiteren Anführungszeichen escapen. Bei einem Feld, das selbst mit einem Anführungszeichen beginnt, funktiniert das jedoch nicht so gut, weil dann am Anfang des Feldes drei Anführungszeichen stehen und statt dass das dritte durch das zweite escapet wird, wird das zweite durch das erste escapet.
- So wie wir die Aufgabe umgesetzt haben, wäre vermutlich die Verwendung von ausschließlich neo4j einfacher und ausreichend gewesen, jedoch wollten wir unbedingt einen funktionierenden Cluster haben, der nicht nur theoretisch umgesetzt ist, sondern auch praktisch funktioniert.