# Aufgabenstellung MC2

## Einleitung

Tobias und Yannic sind zwei junge Journalisten und arbeiten zusammen seit einigen Monaten bei der Nord-West Schweizer Zeitung. Eines Morgens checkt Tobias sein anonymes Postfach im Internet und stösst auf 6 csv-Dateien. Nach genauerer Betrachtung der einzelnen Dateien vermutet Tobias, dass es sich dabei um etwas riesiges handeln könnte! Die Daten sind bis auf einige Kleinigkeiten bereits feinsäuberlich aufbereitet, darin enthalten sind in 5 Unterschiedlichen Tabellen Adressen, Branchen sowie Personen- und Firmennamen. Tobias holt rasch seinen Kollegen Yannic und die beiden beschliessen damit in die IT zu gehen. Andreas der IT-Fachmann für Datenbanken erkennt sofort um was es sich dabei handelt. Der mysteriöse Whistleblower hat die Daten feinsäuberlich für eine Graphen Datenbank aufbereitet. «Wieso er wohl ausgerechnet so eine komplizierte Datenbank verwendete?», Fragten sich Tobias und Yannic. «Das will ich euch mal genauer erklären», antwortete Andreas.

Nebst der Tatsache, dass in diesen Daten die ganzen Beziehungen schon feinsäuberlich aufbereitet sind, **eignen sich Graph-Datenbanken hervorragend, um eben diese Beziehungen darzustellen***.* In dem oben beschriebenen Fall können so einfach Muster erkennt werde, welche in einer Relationalen-Datenbank über viele verschiedene Joins erforscht werden müssten. **Zudem ist die Abfragegeschwindigkeit massiv schneller,** somit können wertvolle Minuten und Ressourcen eingespart werden. Ein weiterer Vorteil ist die Möglichkeit Beziehungen zu gewichten. Beispielsweise könnte man die Dauer einer Reise oder aber auch in unserem Beispiel menschlichen Beziehungen mehr oder weniger Wert zuzusprechen.

«Wenn wir schon eine No-SQL Datenbank nehmen, wieso denn keine Dokumenten-Datenbank, da hätten wir den Vorteil von einer Schemalosen Abspeicherung!», antwortete Yannic.

«Möchten wir beispielsweise die erhaltenen Daten in einer Dokumenten-Datenbank wie Mongo DB abfragen, hätten wir die gleichen Probleme wie mit einer Relationalen Datenbank. Zudem können wir den Vorteil der Schemalosen Datenbank nicht ausspielen, da auch eine Graph-Datenbank Schemalos sein kann. Wir würden nur noch mehr Zeit bei den Abfragen verlieren, wie mit einer Relationalen Datenbank.», versuchte Andrea den beiden zu erklären.

«Und was wäre mit einer Wide-Column-Datenbank?», fragte Tobias.

«Auch hier haben wir wieder die ähnlichen Probleme wie vorhin. Die Abfrageperformance ist bei weitem nicht so gut. Ausserdem sind auch hier intuitive Zusammenhänge schwieriger zu erkennen. Ich zeige euch wie man mit einer Graphen-Datenbank arbeitet, glaubt mir der Mehraufwand wird sich für die Erkenntnisse lohnen. Das Datenbank-Management-System Neo-4j ist Open Source und wir können die heiklen Informationen auf unseren eigenen Servern ohne fremde Hilfe speichern.» sagte Andrea überzeugen.

Eventuell vergleich mit Time-Series statt Wide-Column/key-Value

Fragen Konzeptionelles und logisches Datenmodel