Voorstel onderwerp bachelorproef

|  |  |
| --- | --- |
| **Student** | LORENZ VERSCHINGEL  lorenz.verschingel.r1708@student.hogent.be |
| **Promotor** |  |
| **Co-promotor** | DE CLERCQ Jean-Jacques  [Jeanjacques.declercq@Ugent.be](mailto:Jeanjacques.declercq@Ugent.be)  Universiteit Gent, Faculteit Economie, Vakgroep CRM |
| **Werktitel** | NoSQL: Apache Cassandra |
| **Domein** | Databanken |
| **Vrije trefwoorden** | NoSQL, Cassandra |

1. Probleemstelling en context

Met de opkomst van NoSQL databanken zijn er 4 grote soorten NoSQL databanken naar voor gekomen, namelijk key-value stores (Riak, Redis...), document stores (MongoDB, CouchDB...), Column Family Stores (Cassandra, HBase...) en Graph databases (Neo4J, Infinite Graph...). Deze NoSQL databanken kwamen er in reactie op de problemen die men ondervond met relationele databanken. Een van de grootste frustraties van softwareontwikkelaars in de "impedance mismatch" tussen de relationele datastructuren en de in-memory datastructuren van de applicaties. Een ander voorbeeld is dat relationele databases niet geoptimaliseerd zijn voor het runnen op clusters.

In het kader van mijn bachelorproef wil ik mij graag verder verdiepen in de Column based NoSQL databank Cassandra met de doelstelling om deze verder toe te lichten. Een van de problemen met relationele databanken, zoals eerder vermeld is de schaalbaarheid en de installatie op clusters. Hierdoor wil ik me focussen hoe Cassandra met deze problemen omgaat. Cassandra garandeert de betrouwbaarheid via meervoudig wegschrijven. Hierdoor zouden backup’s overbodig worden. Dit zou ook tot het onderzoek behoren.

1. Doelstelling en onderzoeksvraag/-vragen

Hoe werkt Cassandra:

* Theoretisch
  + Hoe/waar wordt de data opgeslaan?
    - Meervoudig opslaan → Betrouwbaarheid
    - Belang van backups?
  + Belangrijke verschillen in het paradigma als je migreert van relationeel naar een NoSQL model. (Focus op Cassandra)
* Praktisch:
  + Installatie van Cassandra op een cluster
  + Schaalbaarheid van Cassandra nagaan
  + Betrouwbaarheid van Cassandra nagaan
* Voor welke problemen biedt Cassandra een goede oplossing

1. Plan van aanpak

* Literatuurstudie
* Installatie van Cassandra op een cluster (d.m.v. virtuele machines).
* Schaalbaarheid onderzoeken
* Betrouwbaarheid nagaan
  + Deel van de cluster uitschakelen en nagaan hoe er gereageerd wordt

1. Referentielijst

Sadalage, P. (2014). NoSQL Databases: An Overview. Geraadpleegd op 11/01/2016 via https://www.thoughtworks.com/insights/blog/nosql-databases-overview

[GOTO Conferences]. (2013) Introduction to NoSQL - Martin Fowler. [Video File]. via https://www.youtube.com/watch?v=qI\_g07C\_Q5I.

Schumacher, R. (2015). A Brief Introduction to Apache Cassandra. Geraadpleegd op 11/01/2016 via https://academy.datastax.com/demos/brief-introduction-apache-cassandra