

Automatisch uitrol van database systemen en vergelijking van beschikbaarheid

Thomas Uyttendaele

Thesis voorgedragen tot het behalen
van de graad van Master of Science
in de ingenieurswetenschappen:
computerwetenschappen,
hoofdspecialisatie Gedistribueerde
systemen

Promotor:

Prof. dr. ir. Wouter Joosen

Assessor:

Ir.

Begeleider:

Ir. B. Vanbrabant

© Copyright KU Leuven

Zonder voorafgaande schriftelijke toestemming van zowel de promotor als de auteur is overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie i.v.m. het overnemen en/of gebruik en/of realisatie van gedeelten uit deze publicatie, wend u tot het Departement Computerwetenschappen, Celestijnenlaan 200A bus 2402, B-3001 Heverlee, +32-16-327700 of via e-mail info@cs.kuleuven.be.

Voorafgaande schriftelijke toestemming van de promotor is eveneens vereist voor het aanwenden van de in deze masterproef beschreven (originele) methoden, producten, schakelingen en programma's voor industrieel of commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden.

Voorwoord

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Voorwoord
schrijven

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Thomas Uyttendaele

Inhoudsopgave

Voorwoord	i
Samenvatting	iv
Lijst van figuren en tabellen	v
Lijst van afkortingen en symbolen	vi
1 Inleiding	1
2 Overzicht van de technologie	3
2.1 Geschiedenis van de databasemanagementsystemen	3
2.2 Relationale vs NoSQL databases	4
2.3	4
2.4 Vereisten van de systemen	4
2.5 Bestudeerde database systemen	4
2.6 Verschillen in database systemen	5
2.7 Selectie van database systemen	5
3 Methodiek van de testen	7
4 Implementatie	9
4.1 Doelstelling	9
4.2 Gebruik IMP	9
4.3 Gebruik YCSB	9
5 Observaties	11
6 Analyse van de resultaten	13
7 Conclusie	15
Bibliografie	20

Todo list

Voorwoord schrijven	i
Abstract schrijven	iv
Mss ook het toevoegen van IBM hier + (IDS)?	3
Relationeel model basis	3

Samenvatting

Abstract schrijven

In dit **abstract** environment wordt een al dan niet uitgebreide samenvatting van het werk gegeven. De bedoeling is wel dat dit tot 1 bladzijde beperkt blijft. Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Lijst van figuren en tabellen

Lijst van figuren

Lijst van tabellen

Lijst van afkortingen en symbolen

Afkortingen

IMP	Infrastructure Management Platform
DBMS	Databasemanagementsysteem

Symbolen

42	aaa
----	-----

Hoofdstuk 1

Inleiding

Hoofdstuk 2

Overzicht van de technologie

2.1 Geschiedenis van de databasemanagementsystemen

Doorheen de geschiedenis, heeft de mens verschillende manieren gebruikt om data op te slaan en nadien terug te vinden. Doorheen de geschiedenis zijn er verschillende stappen van data management geweest, tot voor het ontstaan van de computer ging dit met pen en papier of met ponskaarten[2]. Met de opkomst van de computer, werden nieuwe methodes gebruikt die doorheen de tijd zijn mee geëvolueerd met de vooruitgang in de technologie en de veranderingen in het gebruik van de data. De hiervoor ontwikkelde software wordt gecategoriseerd onder het **databasemanagementsysteem**(DBMS). De ontwikkeling en opkomst van de DBMS kan in verschillende fasen opgedeeld worden.

De eerste DBMS zijn er gekomen met de introductie van de mainframes zoals UNIVAC1 en de ontwikkeling van specifieke programmeertalen voor het werken met deze data, onder andere conferenties zoals CODASYL hebben de ontwikkeling van COBOL en andere standaarden mee ontwikkeld[2].

De volgende grote verandering in DBMS is er gekomen door het artikel van E. Codd over het relationele model in 1969 [1]. Het relationele model is gebaseerd op theoretische wiskundige principes zoals de set-theorie en eerste-orde predicaten logica. Dit organiseert de data in relaties, een relatie kan gezien worden als een tabel waar een rij een collectie van gerelateerde datawaarden zijn. Een belangrijke eigenschap in dit model is dat de relaties genormaliseerd wordt, database normalisatie is een proces waarbij de redundantie in de data wordt geminimaliseerd. Voorbeelden van populaire DBMS die het relationele model implementeren zijn Oracle, MySQL en PostgreSQL.

Mss ook het toevoegen van IBM hier + (IDS)?

Relationeel model basis

De laatste nieuwe generatie zijn NoSQL databases die sinds 2000 zijn begonnen, NoSQL staat voor '*Not only SQL*'. Deze systemen zijn er gekomen als reactie op

het relationele model voor een meer flexibele database, lagere complexiteit, hogere doorvoer van data, horizontale schaalbaarheid en het draaien op commodity hardware. Verschillende voorbeelden van NoSQL systemen zijn Google BigTable, Amazon Dynamo, HBase, MongoDB, ... [3] Meer informatie en een vergelijking met relationele databases komt aanbod in de volgende sectie.

2.2 Relationele vs NoSQL databases

2.3

Wat en waarom NoSQL en wat zijn de belangrijkste elementen.

2.4 Vereisten van de systemen

- Persistentie
- Distributie
- Replicatie
- Open-source

2.5 Bestudeerde database systemen

- Column database
- Document database
- Graph database
- Key-Value database
- Relationele database

2.5.1 Column database

- Cassandra
- HBase

2.5.2 Document database

- Apache CouchDB
- MongoDB

2.5.3 Graph database

2.5.4 Key-Value database

- Lightcloud (Tokyo)
- MemCache
- Redis
- Riak

2.5.5 Relationale database

- MySQL
- Pgpool-II (PostgreSQL)

2.5.6 Gekozen systemen

HBase

MongoDB

Pgpool-II (PostgreSQL)

2.6 Verschillen in database systemen

2.7 Selectie van database systemen

Hoofdstuk 3

Methodiek van de testen

Hoofdstuk 4

Implementatie

4.1 Doelstelling

4.2 Gebruik IMP

4.3 Gebruik YCSB

Hoofdstuk 5

Observaties

Hoofdstuk 6

Analyse van de resultaten

Hoofdstuk 7

Conclusie

Bijlagen

Bibliografie

- [1] Edgar F Codd. „A relational model of data for large shared data banks”. In: *Communications of the ACM* 13.6 (1970), p. 377–387.
- [2] Jim Gray. „Data Management: Past, Present, and Future”. In: *arXiv preprint cs/0701156* (2007).
- [3] Christof Strauch. *NoSQL Databases*. 2010. URL: <http://www.christof-strauch.de/nosql dbs.pdf>.

Fiche masterproef

Student: Thomas Uyttendaele

Titel: Automatisch uitrol van database systemen en vergelijking van beschikbaarheid

Engelse titel: Automatisch uitrol van database systemen en vergelijking van beschikbaarheid

UDC: 681.3

Korte inhoud:

Hier komt een heel bondig abstract van hooguit 500 woorden. \LaTeX commando's mogen hier gebruikt worden. Blanco lijnen (of het commando `\par`) zijn wel niet toegelaten!

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Thesis voorgedragen tot het behalen van de graad van Master of Science in de ingenieurswetenschappen: computerwetenschappen, hoofdspecialisatie
Gedistribueerde systemen

Promotor: Prof. dr. ir. Wouter Joosen

Assessor: Ir.

Begeleider: Ir. B. Vanbrabant