UNIVERSITY OF TRENTO - Italy

Department of Information Engineering and Computer Science

Bachelor's Degree in Computer Science

Final Dissertation

Evaluating MongoDb Performace:

How and where NoSQL databases are getting over Relational Databases

Supervisor
Alberto Montresor

Student
Michele Romani

Academic year 2015/2016

Acknowledgements

In my experience, everything we do in our lives has some kind of contribution from the people that live around us at home, at school, at work, everywhere. Unless you are a hermit, of course. And there are different kinds of support people can give you, depending on the relation you have with them. On a more affective side, I'd like to thank my family for always believing in me and in what I do, and supporting me in my choices. My girlfriend and my friends, for their support and the great time I always spend with them and that helps me recovering from the overwhelming amount of commitments of every day life. And my course mates, for all the time we studied together sharing knowledge and useful suggestions. On the professional side, first I'd like to express my deepest gratitude to my Supervisor and Professor Alberto Montresor for the passion he transmitted me while teaching Algorithms on the second year, for involving me in interesting projects and for his supervision in this last work of my bachelor degree. I'd also like to thank my Erasmus Professor Erki Eassar for his skill in making me appreciate the course of Databases and for sharing me useful suggestions and materials even months after being his student in Tallinn University of Technology. At last, I'd like to acknowledge my internship tutors in Tai, Andrea Carpineti and David Votino, for their technical and motivational suggestions about the project from which this dissertation has origin, and my colleague Bruno Graziano for his precious help in configuring and maintaining the system of virtual machines that hosted my clients and my Mongo nodes for the main tests. Have a good reading.

Contents

Summary

The IT world is evolving faster and faster every year, with new breaking technologies coming to our lives, even changing our way to communicate and live in our society. With the advent of social networks, cloud storage and computing, a new definition for the amount of data they involve has been coined: Big Data[1]. The challenge of Big Data involves both developing better retrieving solutions using advanced data mining techniques and functional storage solutions. Several companies are switching their old systems and technologies to more scalable and reliable solutions to optimize their costs in terms of time and money, improving their profits. The company in which I am actually working entrusted me to develop a software in order to evaluate MongoDB [2], a new non-relational database technology in anticipation of a new contract from a customer that needs to support an application with several hundred thousand of users and millions of records. The challenge is to obtain acceptable results from Mongo in stressing conditions like a production software: retrieving data in less than 2 seconds, preventing loss of data and most importantly, preventing a system crash of the database. I entirely developed a Java software based on the Spring framework, following my project leader and my tutor directives, capable of launching specific benchmark tests aimed at stress-testing and maybe even crashing a virtual machine running a MongoDB instance. For my architecture used the technique of Microservices [3], that consists in building a modular application, with each module dedicated to a specific service. It is an advanced development technique that is getting more and more successful, also thanks to famous use cases such as Netflix, with the strength of easy reusability and maintainability of the software. My choice of this technique is due to a possible future experimentation of other storage technologies, even relational, as the modularity of the applications allow to quickly develop and connect a new module with drivers for other Database Management Systems. The choice of MongoDB was made by both our manager and our customer because of its ease of configuration and its availability as an open-source software. This research aims to explain many reasons why NoSQL technologies are taking over the well-known relational databases in new enterprise applications, focusing on selected use cases. In particular, I have been committed to develop a software that could perform a stress-test on MongoDB to verify if it could stand the customer requirements. The team involved 3 persons:

- me as Software Developer.
- an internal System Engineer that helped me configuring MongoDB instances on different nodes (depending on the test requirements) and configuring the virtual machines that have been used through the evaluation.
- an internal Software Engineer, my stage tutor, that helped me define the architecture of the application and choose the right frameworks for both backend and frontend. He also contributed in defining test cases and testing the functionalities of the software.

To clarify, the research does not demonstrate that NoSQL technologies are a better choice than Relational DBMS in any case, nor that the relational databases will get outdated and out of use. In fact, both of them have strenghts and weaknesses based on the situation in which they are used. The future of databases will likely involve the parallel use of different technologies or maybe a "fusion" like, for example, NewSQL databases that are currently under experimental development. But even tough relational databases are not going to disappear soon, we will explain why NoSQL are really taking over them in the highly demanding requests of the new market of Big Data challenging applications in terms

[1] https://datascience.berkeley.edu/what-is-big-data/
[2] https://www.mongodb.com/what-is-mongodb
[3] http://microservices.io/patterns/microservices.html

of high scalability, usability and performance. This will likely lead to a relegation of Relational DBMS to specific roles in a system or to specific use cases in which they still have better reliability or even better performance than NoSQL regardless their higher cost of configuration and maintainability and their restrictions as explained by the *CAP theorem* [4]. MongoDb Performcance was developed entirely in Java on the backend side, while the frontend side was developed in Html 5, Css and Javascript. It depends on several frameworks and libraries, among which the most relevant are:

- *Java Spring* [5] - The future of Java Development, based on REST calls and Annotations.
- *AngularJS* [6] - An essential web framework to build single page applications with dynamic loading of contents, used in combination with *Twitter Bootstrap* [7].
- *MetricsGraphics.js*[8] - A versatile Javascript framework based on D3, used to plot data.

The code of the project can be visualized on GitHub [9] only after authorization as its property rights are owned by the company.

[4] Also named Brewer's theorem, will be explained in chapter 2
[5] https://spring.io
[6] https://angularjs.org
[7] http://getbootstrap.com
[8] http://www.metricsgraphicsjs.org
[9] https://github.com/BRomans/mongodb-performance-app

1 Introduction

In the first two parts of this chapter a brief overview will explain the most known databases technologies while in the last part NoSQL databases will be introduced through the descrption of the most famous implementations from which many others derive.

1.1 Discovery of NoSQL technologies

Commonly students have their first encounters with database technologies during their studies in high school or bachelor degree and, to better understand all the fundamentals concepts, they are taught the basic principles of relational databases. It i's the most common choice of every school teaching the very foundation of Databases to make students understand the the meaning of *CRUD operations, relations, consistency, redundancy*, etc... and how to correctly set up the entities of their systems following proven patterns and constraints. Detaching from well-known developing habits is not always so simple, but it is necessary to understand why big companies such as Facebook decided to invest money in developing their own database solution, Cassandra, instead of using an existing relational database. It is important to know that there are many different ways to build a database, some are better than others in certain use cases. Nowadays, an huge amount of data are spread around the world everyday through the Web and it needs to be stored and retriewed quickly to save companies' money and give the users a perfect feeling of resposivity. But let's start from the beginning to get an overview of a technology we rely on every day, even without being aware of its presence in every single application we use.

[1] [2]

1.2 Databases

A Database is an organized collection of data even though we often use the term to refer to the entire database system. The Database Management System, or DBMS, is the name of the entire system that handles data, transactions, relations and eventually problems.

1.2.1 Relational Databases

Probably the most popular and for many years also most used model, a relational database is composed by tables representing entities (users, customers, courses...) where each column represents a field or attribute and each row represents a record. Tables can have relations each other with the use of foreign keys, and each table has a primary key that is unique on each record.

2 Proin rhoncus a sapien in.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

2.1 Cras in aliquam quam, et

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

2.1.1 Sed pulvinar placerat enim, a

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

2.2 Vivamus hendrerit imperdiet ex. Vivamus

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis.

Phasellus gravida tellus velit, non eleifend justo lobortis eget.

3 Conclusioni

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

Bibliography

[1] Ict business. http://www.ictbusiness.it/. ultimo accesso 15/06/2015.

[2] Donoho D. L. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

Allegato A Titolo primo allegato

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

A.1 Titolo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

A.1.1 Sottotitolo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

Allegato B Titolo secondo allegato

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

B.1 Titolo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

B.1.1 Sottotitolo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.