

Analysis of efficient data storage schemes for a structured social notification graph

Research thesis proposal

Florian Dejonckheere

Abstract

Ever since the dawn of dynamic web applications and social networks, there has been an increasing demand for performant storage systems to store the vast deluge of data that is generated by the social activity of users. This data is characterised by its size and inherently structured nature most types of social notification feeds can be reduced to a representation of an acyclic graph. In this paper, the current state of affairs pertaining to the storage of data in small to medium-size web applications will be summarized. An analysis of the structured data within the context of the Open Weblides project will be made, and an attempt will be made to match this analysis to a stable, scalable and maintained storage system. A proposed implementation for a normalized, efficient data schema will also be presented.

Keywords

Webapplications. Graph — Social feed — Data storage — Cloud — NoSQL

Contact: florian@floriandejonckheere.be

Contents

1	Introduction	1
2	State-of-the-art	1
3	Methodology	1
4	Expected results	1
5	Expected conclusions	2

1. Introduction

The Open Weblides project provides a user-friendly platform to collaborate on HTML weblides. One of the core features this application provides is *co-creation*. The co-creation aspect manifests itself in several forms within the application; annotations on slides and a change suggesting system resembling GitHub's pull request feature are the main mechanisms. Because of the inherent social nature of co-creation, a basic notifications feed was also implemented. This feed is tailored to the user, and reflects the most recent changes relevant to the slidedecks the user is interested in.

However, the functionality implemented in the system contains only the bare necessities at the moment. The module will be expanded in the future, and doing so requires a structural rethinking of how the notifications are generated, stored and queried. This paper has two concrete goals: First, it aims to analyse and summarize the existing frameworks and software packages used in the industry to store structured non-relational graph data.

Second, the structure and data provided by the Open Weblides' social notification feed will be interpreted in the context of the aforementioned analysis. Finally, a recommendation will be made for a concrete implementation of the described data storage schema.

2. State-of-the-art

Hier beschrijf je de *state-of-the-art* rondom je gekozen onderzoeksdomein. Dit kan bijvoorbeeld een literatuurstudie zijn. Je mag de titel van deze sectie ook aanpassen (literatuurstudie, stand van zaken, enz.). Zijn er al gelijkaardige onderzoeken gevoerd? Wat concluderen ze? Wat is het verschil met jouw onderzoek? Wat is de relevantie met jouw onderzoek?

Verwijs bij elke introductie van een term of bewering over het domein naar de vakliteratuur, bijvoorbeeld (Doll1954)! Denk zeker goed na welke werken je refereert en waarom.

Je mag gerust gebruik maken van subsecties in dit onderdeel.

3. Methodology

Hier beschrijf je hoe je van plan bent het onderzoek te voeren. Welke onderzoekstechniek ga je toepassen om elk van je onderzoeksvragen te beantwoorden? Gebruik je hiervoor experimenten, vragenlijsten, simulaties? Je beschrijft ook al welke tools je denkt hiervoor te gebruiken of te ontwikkelen.

4. Expected results

Hier beschrijf je welke resultaten je verwacht. Als je metingen en simulaties uitvoert, kan je hier al mock-ups maken van de grafieken samen met de verwachte conclusies. Benoem zeker al je assen en de stukken van de grafiek die je gaat gebruiken. Dit zorgt ervoor dat je concreet weet hoe je je data gaat moeten structureren.

5. Expected conclusions

Hier beschrijf je wat je verwacht uit je onderzoek, met de motivatie waarom. Het is **niet** erg indien uit je onderzoek andere resultaten en conclusies vloeien dan dat je hier beschrijft: het is dan juist interessant om te onderzoeken waarom jouw hypothesen niet overeenkomen met de resultaten.