Dissertation Project

A Comparison of NoSQL and Indexing

Solutions for Big Data

Author: Callum George William Guthrie

Supervisor: Dr. Albert Burger

Heriot Watt University

Edinburgh, Scotland

A dissertation submitted in partial fulfilment of the requirements for the degree of

Bachelor of Science.

May 2016

Declaration

I, Callum George William Guthrie confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are properly acknowledged at any point of their use. A list of the references employed is included.

Signed:

Date:

Abstract

The era of Big Data is upon us bringing with it a range of new challenges, and encouraging the formulation of new approaches for cleaning, processing and using these enormous amounts of data. These new methods have led to the development of a range of technologies designed to meet the needs of Big Data.

This project focuses on a subset of the new technologies, in particular those products designed to deliver high performance querying of large data sets. It shall compare leading NOSQL solutions (e.g., MongoDB [1], and Neo4j [2]) against modern search and analytics engines (e.g., ElasticSearch [3], and SOLR [4]). The overall goal is to compare and contrast the functionally, performance and analytical capabilities of the different solutions. With the ultimate aim of gaining an understanding of, and insight into, these technologies and their application to Big Data.

The student will produce several versions of a prototype application; with each version employing a different technology (or approach). This work will be undertaken using a real world dataset from the biological environment.

[1] http://www.mongodb.org/about/introduction/

[2] http://www.neo4j.org/

[3] http://www.elasticsearch.org/webinars/introduction-elk-stack/

[4] http://lucene.apache.org/solr/

Contents

Chapter 1

The title of chapter one

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec non ante sem. Aliquam volutpat nisi erat, quis pharetra nunc lobortis eget. Nam elementum urna mattis, tincidunt eros porttitor, vehicula dui. Integer nisi erat, mattis eget pretium nec, luctus ac neque. Quisque volutpat, dolor nec facilisis pharetra, diam est efficitur ligula, sit amet tincidunt nunc ante at mauris. Integer mattis ultricies dolor, sit amet sagittis nulla facilisis ut. Duis imperdiet ultrices metus, eget fringilla tortor tincidunt non. Nunc condimentum justo at neque consectetur aliquam. Aliquam eu enim sapien. Vivamus sodales nunc ligula, sed sagittis mauris cursus vitae. Etiam sit amet leo sollicitudin, iaculis enim dignissim, aliquet lorem. Aliquam varius euismod risus, sit amet posuere mi pharetra eu. Sed tempor, velit id tristique gravida, elit nunc ultricies tellus, a elementum arcu ante vitae dui. Suspendisse fringilla leo id nisi ultricies, id imperdiet purus tristique. Maecenas non tincidunt risus $x = 1/\alpha$ magna rhoncus neque, id pulvinar odio lorem non turpis [1, 3]. Nullam sit amet enim. Suspendisse id velit vitae ligula volutpat condimentum. Aliquam erat volutpat. Sed quis velit. Nulla facilisi. Nulla libero. Vivamus pharetra posuere sapien. Nam consectetuer. Sed aliquam, nunc eget euismod ullamcorper, lectus nunc ullamcorper orci, fermentum bibendum enim nibh eget ipsum. Donec porttitor ligula eu dolor. Maecenas vitae nulla consequat libero cursus venenatis. Nam magna enim, accumsan eu, blandit sed, blandit a, eros [2].

1.1 Section 1

Quisque facilisis erat a dui. Nam malesuada ornare dolor. Cras gravida, diam sit amet rhoncus ornare, erat elit consectetuer erat, id egestas pede nibh eget odio. Proin tincidunt, velit vel porta elementum, magna diam molestie sapien, non aliquet massa pede eu diam. Aliquam iaculis. Fusce et ipsum et nulla tristique facilisis. Donec eget sem sit amet ligula viverra gravida. Etiam vehicula urna vel turpis. Suspendisse sagittis ante a urna. Morbi a est quis orci consequat rutrum. Nullam egestas feugiat felis.

The epipolar lines l can be computed as follows:

Bibliography

[1] Sergey Brin and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833, 2012.

[2] Google Inc. Google home page, 2015.

[3] Stuart Russell and Peter Norvig. Artificial intelligence: a modern approach. 1995.