Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

Localised Internet Scanning Infrastructure

Jordan Myers
15323206

B.A. (Mod) Integrated Computer Science
Final Year Project April 2019
Supervisor: Dr. Stephen Farrell

Declaration

I hereby declare that this project is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

Signed: _____ Date: _____

Abstract

A short summary of the problem investigated, the approach taken and the key findings. This should be around 400 words, or less.

This should be on a separate page.

Acknowledgements

You should acknowledge any help that you have received (for example from technical staff), or input provided by, for example, a company.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Goal

The goal of this project is to design and implement a data store and data visualisation tool for a localised internet scanning infrastructure which supports Ireland-sized internet scans.

1.2 Motivation

While large-scale, internet-wide scanning is a well-explored topic, more local scans could produce more actionable results. This project focuses on Ireland-wide scans, but it could be used for other similarly sized scans.

1.3 Report Structure

Chapter 2 outlines the background of the project and related work. It explains what internet scanning is, some of the reasons for it and some of the technologies used.

Chapter 3 gives an overview of the data, examines some of the design considerations and introduces the Elastic Stack.

Chapter 4 explains the implementation of Elasticsearch as a data store and some of the challenges associated with it.

Chapter 5 describes the use of Kibana as a data visualisation tool and looks at some high-level visualisations and the insights provided by them.

Chapter 6 attempts to determine the success of the project by comparing implementations of the same experiment with and without the infrastructure implemented in this project both in terms of code (quantity and readability) and execution time. It also highlights some of the advantages and disadvantages of the new solution.

Chapter 7 provides some final remarks on the project.

Chapter 8 highlights some work that could be done following on from this project.

2 Background

2.1 Internet Scanning

2.2 ZMap/ZGrab

2.3 Censys.io

3 Design

3.1 Data Overview

3.2 Considerations

3.2.1 Data Sensitivity

3.2.2 Network Architecture & Hardware

3.2.3 Open Source

3.3 Elastic Stack

4 Data Store

4.1 Elasticsearch

4.2 Inserting Data

4.3 Data Types and Mappings

4.4 Challenges

Java OOM error on bulk insert

5 Data Analysis

5.1 Kibana

5.1.1 Web Server and TLS Certificate

5.2 Dashboards

5.2.1 Ireland Data

5.2.2 Aviation Data Subset

5.3 Results/Insights

6 Evaluation

6.1 Same Keys

6.1.1 Lines of Code

6.1.2 Run Time

6.2 Scalability

6.3 Usability

7 Conclusion

8 Future Work

Bibliography

A1 Appendix