

Log Files for proactive monitoring of Big Data.

Linda Quézet  
MCIS

2017

# Log Files for proactive monitoring of Big Data.

Linda Quézet

A thesis submitted to  
Auckland University of Technology in partial fulfilment of the requirements for the degree  
of  
Master of Computer and Information Sciences (MCIS)

2017

School of Computing and Mathematical Sciences  
Auckland University of Technology

## Abstract

This thesis reports on a qualitative study of Big Data and Analytics, with an emphasis on tools that can be used to monitor systems within organizations. An overview of the literature available has been given as to what Big Data means, as well as how this data could be used to provide information to an organization.

Big Data is stockpiled into log files by the various systems that have been implemented within an organization. In order to interrogate and find information, as well as deal with issues that may arise, an organization requires evidence to be gathered from log data. This could be used to assist with the decision making process within organizations. Three tools, namely N-Able, AWS CloudWatch and Sumo Logic were implemented in order to gain an understanding of how they could be used to provide information from the data contained in these log files. During the analysis of these tools the focus was on what data these tools provided, and if and how these tools could provide an in-depth analysis by utilizing Big Data and Big Data Analytics. The use of the tools, allows for an organization to monitor and alert on their infrastructure and other environments. In all cases the tools reviewed were able to provide this as a basic foundation. Sumo Logic stood out as the most productive tool, as it had a similar basic foundation as the other two tools, but in addition had the Big Data Analytical capability inbuilt.

From the review of tools and literature during this research it came to the fore that there is a requirement for Big Data Analytical Tools. In order for information to be collected and assimilated into something useful – Big Data Analytical tools provide the means and methods to assist in providing information that is both useful and timeous.

# TABLE OF CONTENTS

---

<b>ABSTRACT .....</b>	<b>I</b>
<b>LIST OF FIGURES .....</b>	<b>IV</b>
<b>LIST OF TABLES .....</b>	<b>VI</b>
<b>ATTESTATION OF AUTHORSHIP .....</b>	<b>VII</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>VIII</b>
<b>CHAPTER 1 – INTRODUCTION .....</b>	<b>1</b>
<b>CHAPTER 2 LITERATURE REVIEW .....</b>	<b>4</b>
2.1 INTRODUCTION.....	4
2.2 WHAT IS BIG DATA .....	4
2.3 BIG DATA ANALYTICS.....	8
2.4 LOG FILES .....	9
2.5 MONITORING TOOLS .....	11
2.6 BIG DATA IN ORGANIZATIONS .....	18
<b>CHAPTER 3: RESEARCH METHODOLOGY.....</b>	<b>20</b>
3.1 INTRODUCTION.....	20
3.2 RESEARCH QUESTIONS .....	20
3.3 N-ABLE.....	23
3.4 AMAZON CLOUDWATCH.....	23
3.5 SUMO LOGIC.....	23
3.6 EVALUATION OF TOOLS .....	24
<b>CHAPTER 4 IMPLEMENTATION AND FINDINGS.....</b>	<b>25</b>
4.1 INTRODUCTION.....	25
4.2 CRITERIA USED .....	25
4.3 ANALYSIS OF CHOSEN TOOLS.....	26
<b>CHAPTER 5 ANALYSIS .....</b>	<b>37</b>
5.1 INTRODUCTION.....	37
5.2 EVALUATION OF MONITORING PLATFORMS .....	37
<b>CHAPTER 6 OBSERVATIONS AND CONCLUSION .....</b>	<b>84</b>
6.1 INTRODUCTION.....	84
6.2 OBSERVATIONS.....	84
6.3 DISCUSSION .....	89

6.4	CONCLUSION .....	92
<b>CHAPTER 7 LIMITATIONS AND FUTURE RESEARCH .....</b>		<b>93</b>
7.1	INTRODUCTION.....	93
7.2	LIMITATIONS .....	93
7.3	FUTURE RESEARCH .....	94
<b>REFERENCES.....</b>		<b>107</b>
<b>GLOSSARY .....</b>		<b>117</b>

## List of Figures

Figure 1 - NAT-based N-Central configuration (N-Able) .....	38
Figure 2 - N-Able Service Templates .....	39
Figure 3 - N-Able Dell Server Template .....	39
Figure 4 - N-Able Agent Information .....	41
Figure 5 - N-Able Drill Down View .....	42
Figure 6 - N-Able configuration Options .....	43
Figure 7 - N-Able Monitoring Options .....	43
Figure 8 - N-Able Downtime Options .....	44
Figure 9 - N-Able Status Tab .....	44
Figure 10 - N-Able Object Status Tab – CPU Utilization .....	45
Figure 11 - N-Able Object Service Details Tab – CPU Utilization .....	46
Figure 12 - N-Able Object Thresholds Tab - CPU Utilization .....	46
Figure 13 - N-Able Object Self-Healing Tab - CPU Utilization .....	47
Figure 14 - N-Able Asset Summary - Asset Tab .....	48
Figure 15 - N-Able Asset Summary - Asset Details .....	49
Figure 16 - N-Able Asset Summary - Configurable Asset Details .....	50
Figure 17 - N-Able Maintenance Windows .....	50
Figure 18 - N-Able System Change .....	51
Figure 19 - N-Able Remote Control .....	51
Figure 20 - N-Able Reports Tab .....	52
Figure 21 - N-Able Reports - Remote Support Manager .....	53
Figure 22 - N-Able Reports - View of Report .....	53
Figure 23 - N-Able Reports - Asset Reporting .....	54
Figure 24 - N-Able Reports - Availability Reporting .....	57
Figure 25 - N-Able Reports - Availability Aggregated for Device .....	58
Figure 26 - N-Able Reports - Availability Aggregated for One Service on One Device .....	59
Figure 27 - N-Able Device Report Drill Down .....	60
Figure 28 - N-Able Object Reporting Tab - CPU Utilization .....	61
Figure 29 - EC2 Dashboard .....	65
Figure 30 - CloudWatch Main Landing Page .....	65
Figure 31- CloudWatch Dashboards Page .....	66
Figure 32- CloudWatch - CPU Utilization of Server "Prod Test Environment" over the last 7 days .....	67
Figure 33 - CloudWatch - DiskWrieOps of Server "Prod Test Environment" over the last 7 days .....	67
Figure 34 - CloudWatch - NetworkIN of Server "Prod Test Environment" over the last 7 days .....	68
Figure 35 - CloudWatch - NetworkOUT of Server "Prod Test Environment" over the last 7 days .....	68
Figure 36 - CloudWatch - Alarm - High CPU for Server "Prod Test Environment" .....	69

Figure 37 - CPU Utilization Graphic (AWS, 2015).....	70
Figure 38 - Metrics available for CPU Utilization (AWS, 2015) .....	71
Figure 39 - CloudWatch command line interface .....	71
Figure 40 - Sumo Logic Single Collector (Sumo Logic, 2015) .....	73
Figure 41 - Sumo Logic Multiple Collectors (Sumo Logic, 2015).....	74
Figure 42 - Sumo Logic Hosted Collectors.....	75
Figure 43 Sumo Logic Alert and Notify (Sumo Logic, 2015).....	76
Figure 44 - Sumo Logic Collect and Centralize .....	78
Figure 45 – Sumo Logic Search and Analyze (Sumo Logic, 2015) .....	79
Figure 46 - Sumo Logic Detect and Predict (Sumo Logic, 2015) .....	80
Figure 47 Sumo Logic Monitor and Visualize (Sumo Logic, 2015) .....	81

# List of Tables

<i>Table 1 - N-Able Costs .....</i>	<i>62</i>
<i>Table 2 N-Able Licensing Costs.....</i>	<i>63</i>
<i>Table 3 N-Able Service Offerings and Costings .....</i>	<i>63</i>
<i>Table 4 - AWS CloudWatch Options and Pricing.....</i>	<i>71</i>
<i>Table 5 - Sumo Logic Costs .....</i>	<i>82</i>



## Attestation of Authorship

“I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgments), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.”

Yours sincerely,

Linda Quézet

## Acknowledgements

This thesis is dedicated in loving memory of my parents, Jette and Manfred Gudenberg. Your encouragement and support of me over the years showed me that nothing is impossible.

To my husband, Gary, thank you for your eternal patience, and for encouraging me in all that I do. My children, Samantha and Calvin, thank you for supporting and standing by me. This accomplishment would not have been possible without all of you.

Thank you to my supervisor, Shoba Tegginmath, for all her support and assistance.

And thank you God, with whom all things are possible.

## Chapter 1 – Introduction

Organizations are relying more and more on IT systems to run their business. By implementing monitoring tools or platforms to monitor and alert on the systems is a first step in being able to determine availability and performance (Hernantes, Gallardo, & Serrano, IT Infrastructure-Monitoring Tools, 2015).

This thesis considers the tools that are available in the marketplace that can be used to monitor an organization's systems. In addition, the monitoring tools or platforms that can provide the means to get information from the data which is collected during the monitoring process were also investigated. It has always been a challenge to find the right tools to perform the analyses that are required to gain more insight into the log data that systems generate. Through the implementation and use of monitoring and alerting tools on systems within an organization, some proactive measures to alleviate system inefficiencies and downtime are possible. In addition, monitoring can be taken further to provide insights and information into the way that an organization functions which in turn helps the organization function better in the future.

When an organization monitors its infrastructure, the monitoring has the potential to shed light on the systems used and their health. Some of the available monitoring tools provide basic monitoring of standard information on the various information technologies in use in an organization while other monitoring tools are more intricate and customizable. It is not always easy to choose the right tool. The options available should be weighed according to business requirements and the level of competency within the organization's IT department.

Coming to how the monitoring of an organization's infrastructure fits in with Big Data—monitoring of systems produces vast amounts of data, which are not always easy to manage.

How an organization uses this Big Data depends largely upon the tools used to mine it.

Proactive monitoring of systems requires fore knowledge of what the system is doing and how it is performing, what errors are being generated and why.

Many organizations are moving to Cloud-based infrastructure, as well as other services provided by Cloud options like SaaS and PaaS. Some organizations have a hybrid approach in that they have some infrastructure on premise and use Cloud provisions for various other systems/services

that they provide. This makes monitoring of the infrastructure more complex in that there needs to be a view of all systems in operation, no matter where situated: in the Cloud or in a regular Data Centre, or within the organizations offices. When investigating the tools, it is good to understand what capabilities are required from the tool, basic or in-depth. Many organizations are mandated to increase performance of their IT systems, as well as provide services that are available 24/7 (Pandian & Chinnathurai, 2016). It is thus important to look at monitoring tools that can assist with providing early warning of capacity and other issues to be able to troubleshoot and correct these issues (Heath, 2011). Some monitoring tools can provide monitoring across all infrastructures, local or remote, including infrastructure in the cloud. The tools will however become more complex in order to cover all options. The tool chosen should be able to monitor for example various operating systems, different hardware, different network equipment and web applications, as required by the organization. Finding a tool that fits all requirements can be an ominous task for an organization, as well as an expensive one.

The objective of this research was to gain insight into how best to approach finding and implementing tools that can assist organizations to better understand and utilize the information that they have in their log data. To this end, 2 research questions investigated in this thesis are: What capabilities of a monitoring platform should an organization consider while choosing one to implement; and how can log files be exploited by Big Data Analytics to gain benefits for the organization.

In order to answer these questions, an investigation of three Monitoring Toolsets was undertaken. The three tools chosen for investigation have varying degrees of usability, interoperability and scope. The first of these tools has the ability to monitor across a hybrid environment with relative ease and provide information on devices, systems, and applications. The second tool is for use in the AWS environment and can provide logging and monitoring capabilities within this environment. The third one can effectively monitor hybrid environments as well as provide functionality around log aggregation and analytics.

In Chapter 2, the literature review, gives a summary of what Big Data is purported to be. There are many types of Big Data found in many different forms. Log files contain large quantities of data in different formats which can give insight into various applications and business drivers. The history of Big Data is included in this chapter to give an overview of how this has developed

over the years. There is also a section on what Big Data Analytics is about and what this can mean when looking at data contained in log files.

Chapter 3 gives an outline of the research methodology used in this research as well as the motivations thereof. Chapter 4 details the findings of the research, including the additional value the use of Tools can provide.

Chapter 5 gives an analysis of Big Data in Log Files, monitoring of infrastructure and the value that can be gained from log files. This chapter also details the review of the implementation of the three tools, and gives the reader a holistic view of the review undertaken on the chosen tools.

Chapter 6 discusses the three monitoring tools reviewed and discusses the outcome from the implementations and analysis. Chapter 7 looks at the limitations noted during the course of this thesis, as well as what further research could be undertaken.

## Chapter 2 Literature Review

### 2.1 Introduction

This chapter will undertake to lay the foundation for the thesis, by undertaking a review of the literature of the topics covered, as well as trying to identify gaps in previous research. An initial outline of some of the terminology is given so as to give an understanding of each. Various sources within the literature review looked at what Big Data Analytics means. From these sources there seems to be consensus on what Big Data Analytics can provide. The value from analyzing large pools of data, when done with the organizational environment in mind, is great. The information that is produced can give more knowledge and insight into aspects previously unknown or unanalyzed.

During the course of this thesis both commercial and academic resources were utilized to provide a holistic view of the current state of monitoring, alerting and analyzing within the Big Data and Big Data Analytics fields.

### 2.2 What is Big Data

#### 2.2.1 Big Data Definitions

There does not seem to be one agreed definition of what Big Data is. There are many different definitions used, however all point towards Big Data meaning large volumes of data generated by operating systems and applications.

“Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information” (Rouse, 2014, p. 1).

When defining Big Data, the 3Vs come into play. The first is Volume; this depicts the very large volumes of data that is generated. Secondly the Variety of data that is amassed, there are many types of data, structured, unstructured that are collected. Lastly is Velocity, this is the speed at which this data must be processed, in order provide information that is timely.

In an article in IGI Global the authors define Big Data as the process of extracting information from large amounts of data.

“It (...Big Data) refers roughly to extraction of actionable intelligence from a large amount of data, including social Web data, and applying it to some important needs of an organization. The data may be stored in the proprietary databases of an organization or purchased from third-party data providers or may be gathered from the Internet” (Kim, Jeong, & Kim, 2014, p. 1).

Over the years, the amount of data generated has grown exponentially. There has also been an increasing awareness of this growth of data, as well as to use this data in work and personal environments (George, Hass, & Pentland, 2014). Big data analytics developed to deal with how to gain more information from this data. Big Data generated from a multitude of sources, including user clicks on the internet, through to sensor driven machinery. Being big or having large volumes is not the only defining factor of Big Data; it is also the nature of the data, how fine-grained it is, so that there is a shift from size to granularity contained therein.

The evolution of the term Big Data has happened rapidly and often caused much confusion. A survey in 2012 to gain an understanding of what Executives thought Big Data was, showed that everyone seemed to differ in his or her understanding of what it actually meant (Gandomi & Haider, 2015). One of the first things that comes to mind when mentioning Big Data is size. However, Big Data can be defined using many different characteristics, such as the 3 V's (Volume, Variety and Velocity). Another definition from TechAmerica's Foundation Federal Big Data Commission in 2012 follows:

“Big data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information” (Gandomi & Haider, 2015, p. 2).

Big Data is a technology developed in order to deal with vast amounts of data generated by machines, programs and other sources. Edd Dumbill in 2013 attempted a definition in an article called “Making Sense of Big Data” wherein he stated,

“Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures. To gain value from this data, you must choose an alternative way to process it” (Dumbill, 2013, p. 1).

Social media plays an ever more important role in many organizations and the information obtained from these sources can be invaluable.

“Big Data can yield profound insights into what consumers really want and need. Specifically, what you’re hunting for are actionable insights that you can fulfill better, more easily, or more inexpensively than your competitors can” (Stringfellow, 2015, p. 3).

All organizations have data, and this data forms an invaluable asset if utilized to its full potential.

“...information has far lesser impact when presented as raw data. In order to maximize the value of information, it must be captured, analyzed, quantified, compiled, manipulated, made accessible, and shared“ (Moga, 2014, p. 1).

The main argument or point from this definition is that organizations have vast amounts of data, but in order to gain valuable information it needs to be processed.

### **2.2.2 History of Big Data**

Big Data predates the use of computers. Many countries gathered data on population, taxes, diseases, etc. and transformed the data into information to allow the political party of the time to make decisions. Granted it took longer without computers to gather the data, makes sense of it and output it into useful information. Even so, it constituted Big Data, it was massive amounts of data that needed to be analyzed and sorted; information from this was used for decision-making. One of the earliest records of data being gathered and used for tracking in a business, dates back to around 1663, when John Aunt recorded and examined information about mortality rates in London (van Rijmenam, n.d.) .

In 1937, Franklin D. Roosevelt’s administration required that due to the Social Security Act law, the government required to keep track of the contributions made by employees and employers. IBM granted the contract, developed a punch card-reading machine used to capture this data.

In 1943, a data-processing machine developed by the British and used to decipher Nazi codes in World War II. This machine called the “Colossus” used to look for patterns in messages that had been intercepted, and it did this at a speed of around 5000 characters per second. This assisted in reducing the interpretation tasks from weeks to hours.



The NSA, created in 1952, proceeded to employ around 12000 cryptologists, who were to deal with all the data generated during the Cold War. The cryptologists collected and processed intelligence signals automatically.

1965 saw the first data center built by the United States government, to store the over 742 million tax returns and around 175 million sets of fingerprints. Records were stored on magnetic tape at this location; the project however, discontinued due to privacy reasons.

British scientist Tim Barnes-Lee, in 1989, invented the World Wide Web, as a means to transfer information via system using hypertext. The impact of this creation was massive; it led to even greater collections of data, and this made information more accessible to the world as a whole.

During the 1990's, data creation was spurred on by more and more devices being connected to the internet. The first super computer built in 1995 was able to do as much work in one second as a calculator operated by one person could in 30000 years.

### **2.2.3 Types of Big Data**

Data can exist within internal and external to an organization. In an article on the 7 Important Types of Big Data Michele Nemschoff gives an outline of what is considered structured and unstructured data created outside of an organization (Nemschoff M. , 2014). Firstly, she discusses the types of structured data created outside of an organization. Structured data can be created, provoked, transacted, compiled or be experimental. She then moves on to discuss the types of unstructured data, that is, either captured or user-generated. Created data is data that a business has purposefully input into a system, for use in various business systems, including for marketing and research purposes. Provoked data is data related to opinions and ratings given by people for products and services that they purchase from an organization. Transacted data is general business data captured during the undertaking of business transactions, when customers purchase products or services. Information relating to the transaction is captured in the business systems for storage and future use. Compiled data is the information collected by organizations for the purposes of gaining information in regards to where purchases were made, for example, which is often used in marketing campaigns. Experimental data is data obtained by an organization trying out different products or services within the market. Unstructured data as stated can be of two types, namely captured or user-generated. Captured data is data obtained

passively by actions undertaken by people in their daily lives. User-generated data is information shared by a user on various forums and social and business sites. This information can provide material for organizations to use in their marketing campaigns in order for them to target certain market segments.

Lisa Arthur wrote in an article on the Forbes.com website that data falls into two categories, unstructured data and multi-structured data (Arthur, 2013). Unstructured data is data not in any particular format and can generally not be utilized or organized by many of the traditional databases or models that are utilized for structured data. Multi-Structured data is an assortment of text, data types and formats, images and even log files from various services or machinery.

The above descriptions of types of Big Data, show that log files are considered to be Big Data. Some log files are more structured than others, but they are mostly of the unstructured or multi-structured nature.

### **2.3 Big Data Analytics**

Predictive analytical tools can provide information if the focus is towards the quality of the data that is being generated and collected, rather than just the volume thereof. By improving the quality an organization's IT department is better equipped to generate real-time insights from the vast amount of data generated (Subramanian, 2015).

With Data Analytics, information is provided in many organizational sectors, to the right people at the right time for the decision makers to have everything they need to hand to grow their organization. Using the vast amounts of data that an organization collects in many forms, including in log files for infrastructure, and application, can provide valuable information if analyzed correctly (Zulkernine, et al., 2013).

Gathering the information from the vast amounts of data that an organization captures requires the use of intelligent and scalable analytics services, tools and applications (Talía, 2013). Big Data Analytics utilizes:

“...compute-intensive data mining algorithms that require high-performance processors to produce timely results” (Talía, 2013, p. 98).

Data is plentiful and everything that organizations do on systems creates more and more data. Finding a way to analyze this data and gain information from it is the key (Power, 2014). Having large datasets means that large storage areas are required in order to house the data. Wanting to use this data and turn it into information that is valuable to the organization requires toolsets and human resources. In addition, most of these resources, hardware and human, are expensive. One of the challenges that organizations face is finding the right mix of decision support and information technology to help identify the use cases to extract this large volume of data and transform it into the information that the organization needs to run, as well as to gain competitive advantage.

Big Data Analytics requires vast amounts of processing power, as well as a large amount of storage space. Parallel and distributed systems lend themselves well towards solving for the needs of Big Data Analytics (Kambatla, Kollias, Kumar, & Grama, 2014). Data collected by an organization is dispersed across different platforms, applications and systems.

## **2.4 Log files**

Log files contain large quantities of data in different formats, which can give insight into various applications and business drivers. Organizations of different types have data that is stored in many different ways, as well as this, data stored can be specific to that industry in which the organization functions.

Log files contain:

“...a gold mine of analytic information, they can help you find problems, and they can tell you about your IT infrastructure, the behavior of your users, and identify potential attackers” (Splunk, 2016, p. 1).

Log files are often set up across many systems within an organization, and this can be leveraged in order to provide value. Logs contain various types of data; such as what activity is being undertaken on a system(s) as well as information on the systems themselves. This data needs to be transformed into information to provide value that can be leveraged by an organization.

Log files can be structured, semi-structured and unstructured.

“Unstructured data refers to information that either does not have a pre-defined data model and/or is not organized in a predefined manner” (Nemschoff M. , 2014, p. 1).

Unstructured data is more difficult to sift through, even though it may reveal patterns that are useful and can give rise to important information. Semi-structured data does not usually fit into relational databases; it does however have some structural properties that allow for easier analysis (Ronk, 2014).

Log data that is kept in files generated by operating systems and applications, including network infrastructure hardware and software, can be a very rich source of information and operational data. It may also be termed the very first instance of what is today termed Big Data (Kusnetzky, 2012).

Log files are considered to be semi structured, and therefore more difficult to analyze and get information from (Di Martino, Aversa, Cretella, Esposito, & Kolodzeij, 2014). Log files for certain applications can also be unstructured. These types of log files often do not follow a serialized standard and dividing them into big data flows can be difficult, and often almost impossible without the right tools.

Log files can be used to gather information about user behavior from a system with which they interact (Agosti, Ferro, Peters, de Rijke, & Smeaton, 2010). How users interact with a system can be determined in part from the information contained in log files, which are generated by the system that the user is interacting with. Useful trends and activities that are commonly undertaken by users can be gleaned from analyzing these log files. Using log files to evaluate the data can reveal insights into the way users interact with the system, and the way that they interpret the flow of the system. These valuable insights can be used to direct studies on usability, and various other informational studies.

Log files create huge amounts of data, by logging or writing the status of each component that reside on hardware and form part of a system. Included in these are the statuses, such as stop or start of the component, including errors or successes thereof (Peng , Li, & Ma, 2005). Logs are written in different formats by diverse vendors and are dissimilar for different types of devices or systems. This makes gaining information from log files cumbersome. In order to automate

analysis of this data the messages in the log files need to be categorized and context needs to be taken into account.

Overall log files contain large amounts of data as is stated above, but they are often not easily readable. A lot of information can be gathered from log files, and this requires the use of tools that can assimilate the data from these.

Monitoring tools can provide visibility via dashboards or reports and thus help with troubleshooting and correction of faults that may occur. They also can provide the basis for reporting on uptime and what is going well within the infrastructure, as well as assist with capacity planning.

Cause and effect relationships are not always straightforward and are often impossible to identify (Bigelow, Log analysis tools join data center management arsenal, 2015). Some logs contain information tracking what users are doing in the systems, and showing if any errors have crept into the code. It can be a fulltime job for someone to look through the log files to ascertain where and when something went wrong. It can be just as time consuming for someone to look at the log files to determine what was going well and how to improve upon that. Tools for monitoring systems enable views of these log files to be visualized and alerted upon to save enormous amounts of time.

## **2.5 Monitoring Tools**

Monitoring tools are imperative in most environments, this is really due to the volume of logs that can get written every day, especially when an issue occurs. Mostly the logs are unstructured and come from many different places (Shoor , 2015). Log management tools allow for these logs to be centralized and kept in a more productive format to provide information. Monitoring tools come in all shapes and sizes, from the very basic to the more complex and integrated.

There are also many offerings of each type of monitoring tool. Andy Lurie (Lurie, 2014) provided a brief summation of 47 log management tools that are available in the market. This goes to show that there is a huge offering in the market and an organization will need to review their needs and choose wisely when looking at what tool or tools to implement. Business requirements will need to be looked at first to ensure that the right tool for the requirements is chosen. It can be a mammoth task, especially since some of the tools are specific to a certain

criterion, like for example, SIEM (Security Information and Event Management) – so if that is the main criterion for an organization the choice may be a little easier. The choice may also include more than one tool in order to cover all aspects of the requirements – however it would be more efficient to look for something that covers most of the requirements without much customization. Tools provide visibility on what is happening within an organization's infrastructure implementations; tools that can be set up to monitor and alert at certain specified thresholds, so that the organization is aware of any issues, are required.

With the plethora of tools available there is a need for a full study and detailed analysis of tools in order to ensure that an organization is able to knowledgeably select one or even two tools that can fulfill their requirements. The following sub sections provide a summary of the capabilities of 8 such tools.

Moving towards more proactive monitoring can take a fair amount of knowledge, investigation and effort especially knowledge about the systems that are being monitored. This in itself allows for some automated tasks to be performed which gives a more proactive approach to standard monitoring (Halcyon, 2016). When looking for monitoring software that would provide this level of monitoring it is important to look at an offering which has standardized templates which can be used. This will give the organization some significant gains, in that they do not have to develop customized monitoring which can take up a lot of time, and requires development and constant revision. Doing some customization can provide additional benefits over time as well as help by gaining more understanding of the tool in use. Monitoring tools when set up correctly can provide very valuable proactive capabilities. It can be the difference between knowing something is wrong or being told something is wrong. Knowing that something is wrong before others know gives an opportunity for the issue to be investigated and fixed so as to lessen the impact on the users of the system (Sharp, 2015).

Monitoring tools look for and examine alerts and logs to find performance and other issues. This is defined as being reactive and is usually all about diagnostics. Finding the problem after it has occurred. Digging back into huge amounts of log files is like trying to find the proverbial needle in the haystack. Investing in proactive monitoring up front can save time and money.

Diagnostics is also an important part of monitoring, especially for unexpected issues, and it is also used when monitoring points out potential issues. Balancing these two functions, may mean

the requirement for different tools that are able to perform these functions, however they should be interoperable in order to provide the most benefit.

Most IT architectures consist of complex environments interacting with each other. Performance and uptime is of utmost importance for most systems in an organization (Conley, 2016). With this in mind, monitoring of systems has become more and more important, specifically proactive monitoring. An organization should ensure that it has the right tools in place in order to perform proactive monitoring. The tools can provide insight into problems before they become critical issues that stop the system and therefore the business. Proactive monitoring often entails constant data collection and data retention; this allows for deep analysis of what is happening within the systems. It gives insight into issues or problems that may negatively impact the systems and allows for proactive steps to be taken before something has a major impact.

Different monitoring tools can be effective in providing data at the right time to an organization so that they can ensure uptime and also have a view on who is using their systems, and how they are being used. By using tools to monitor users' interactions with systems in real-time can provide invaluable information to an organization (Croll & Power, 2009). Having something in place to capture this data, can show up a number of potential issues and assist with improvements to the systems that are being used.

When systems are being developed it is important to ensure that the correct data is being captured in log files. Using log files to show application data like errors needs to be done in a systematic way, so as to ensure that the log files show only pertinent data that can assist with the troubleshooting of issues (Eberhardt, 2014). Developers need to know what is important and needs to be logged so as to ensure that there is not too much information contained in the log files as this will hide the true identity of the problem.

By having logging tools in place helps with the analysis of incidents of varying kinds. Being able to have a tool as opposed to delving manually into log files is considered to be adding value. This also allows the administrators of the systems more time to develop and enhance the system instead of spending time analyzing millions of entries in log files to determine issues.

Monitoring tools can work across platforms giving a better view of the environment as a whole and thus providing data that can be more useful than if it were just from one system. The issue may not be noted in a system log file, but may be noted in a network device log file – thus

allowing for the quicker identification of issues and where they may be occurring (Solar Winds N-Able, 2013).

### **2.5.1 Splunk**

Splunk is considered to be the industry leading platform that allows for centralization of all types of logging data and automatically indexes this data. The aim of this management tool is to provide insight into real-time data and allows for scalability and better searching facilities. Processes can be automated and allow for quick searches to be undertaken. Splunk is able to draw data from all types of devices, creating a full organizational picture of the logging data that is captured. Splunk is an enterprise-focused tool and is flexible in deployment; it can be deployed on premise, in hybrid clouds, public or private clouds. However, costs can become quite high as more and more data is centralized (Lurie, 2014). The Cloud option is generally for small medium business (SMB) type organizations, whereas the on premise solution is usually for larger organizations. Splunk offers many features and provides all-encompassing search tools as well as charting tools to visualize the data. It provides collection of data through the use of APIs and its UI, which gives the organization flexibility.

## **2.6 Loggly**

Loggly, a cloud-based only platform, makes the management of log data less burdensome by having a simple set-up and intuitive tools. Loggly has a number of customizable alerts with triggers, as well as easy to use and customizable dashboards for visualization of the data. Costs for the platform are also based, like Splunk, on the amount of data stored and uploaded. This means that it can become prohibitive unless ensuring that retention is kept to a reasonable level (Lurie, 2014).

Loggly as a tool is very much geared towards the operational side of things, and provides devops with information that can be used to find as well as fix operational problems (Weiss, 2014). This tool is focused towards developers and has an easy to use dashboard that can be utilized for monitoring. It is however not a complete infrastructure, security or analytics solution, as it is mainly intended for devops engineers to parse data from application servers.

### **2.6.1 Sumo Logic**

Sumo Logic is a log management tool that is purely provisioned as Software as a Service (SaaS). It is a cloud based service that can leverage Big Data analytics in order to give insights into IT



systems that are configured within an organization. The tool has many features, and has developed into an enterprise class management tool (Weiss, 2014). Set up and operation of the tool is comparatively less costly and easier due to not having to have onsite infrastructure to support it. It can be set up to monitor hybrid environments, as well as providing log aggregation to allow for deeper analysis of logs.

However, moving lots of data up to the Cloud provisioning can be a costly affair. Another possible drawback is the lag in the time it takes to get the data to instance, and being able to use the information contained therein.

As a cloud-based management tool, Sumo as stated has a number of components (applications) which provide specific services. These services operate as lower level modules, making use of common functionality, which in the end results in a unified solution that is enabled for scalability (Lurie, 2014).

Starting out as a SaaS version of Splunk, Sumo Logic has many similar features to Splunk (Shoor, 2015). It has many features that enable reduction, search and charting of masses of data. It has the ability to baseline and uses notifications actively from metrics that have been setup. It also has the ability to use real-time analysis and make use of machine-generated Big Data in order to provide more in depth understandings into systems and their states. Sumo Logic offers many applications especially for the larger tools, but may not have an application for things that are more obscure or older.

### **2.6.2 Logstash**

Another log management tool is Logstash. This is an open source tool, and includes Elasticsearch for indexing as well as Kibana, which allows for the charting and visualization of the data (Weiss, 2014). It is an easy to install tool and allows for ease of setup. It is a tool that has been built on three mature and powerful components which are maintained, and thus made it a robust and very extensible product.

Logstash runs on a Linux platform which allows for real-time pipelining (Killi, 2015). It was originally designed only for data collection, however it has matured and the newer versions have integrated many other capabilities including being able to utilize a wide range of inputs, filtering and output plugins and formats. Data can be unified across source systems and this data can be

normalized to cater for a System Administrators choice of view. The tool also allows a System Administrator the capability of cleansing, comparing and standardizing all logging data, for distinct analytics and visualization.

As an open-source tool, Logstash allows for simple ways to extend for custom logging formats and adding of plugins (Shoor , 2015). The tool however does require the addition of Elasticsearch and Kibana, to give it more substance as on its own it does not provide a proper front or back end. Logstash does not provide any dashboards, but allows for flexibility in where and how to store data. When Logstash is part of a ElasticSearch and Kibana implementation, Kibana is often used as the front end for reporting and visualization. However, other reporting and visualization tools like Graphite, DataDog, or Librato can also be used, due to the extensions within Logstash.

### **2.6.3 AlertLogic**

Monitoring for security related incidents is often important to an organization, especially if the data that they house has value to others. AlertLogic is a Security as a Service provisioning that has the ability to work across various platforms including looking at hybrid implementations (i.e. on-premises and cloud); and provide information from the logs on the security aspects of the application (Kepes, 2015). The tool also utilizes information gained from the analysis of logs from all its clients to provide information on security issues that can affect all their customers. Due to the nature of applications that are currently being developed, it is becoming more challenging to protect and secure the applications. It is therefore important to have the right tools to protect, monitor, analyze and report on possible security related incidents.

AlertLogic offers security as a service as well as a managed service which looks at incident monitoring (Cser, 2015). This particular tool is geared towards security by offering managed services including network-based IDS; SIEM (Security Information and Event Management); firewalling for web applications; Big Data analytics including a correlation engine and of course 24x7 monitoring of incidents.

### **2.6.4 WhatsUpGold**

WhatsUpGold is a tool that provides the A to Z of log collection, monitoring, alerting amongst many other features (Lurie, 2014). The product is able to do real-time monitoring as well as log

based monitoring. It allows for filtering and analyzing of logs to provide clearer insight into issues and alerting. The tool is able to collect logs from a variety of devices and applications and also has the facility for intrusion detection alerting.

The tool WhatsUpGold is supplied and developed by Ipswitch (Rashid, 2016). The tool has the capability to investigate the network and come up with alerting that allows for pro-active monitoring by identifying problems before they become serious. It has the capability of giving an overall view of the entire network via a dashboard which has tools to assist in troubleshooting. The installation is a one-stop implementation including IIS and SQL Server, and there is no need for installation of end-points on each device that needs to be monitored.

### **2.6.5 N-Able**

N-Able (Solar Winds N-Able, 2013) is a Solar Winds product which has a central console which allows for a “single pane of glass” view of an entire environment, whether the infrastructure is in the Cloud or On-Premise, or even in traditional Data Centers. Installing of the agents within the environment can be done automatically and the templates that are used for monitoring allow for quick setup of standard monitoring.

### **2.6.6 CloudWatch**

CloudWatch (AWS, 2015) is a tool provided by Amazon Web Services (AWS). AWS provides this tool to assist with the monitoring of the environment that has been set up by customers. It is generally not automatically set up and needs to be configured in order to start monitoring the various aspects of the customer’s environment. Set up of monitoring and alerting is configurable and relatively easy, giving visibility of environments that have been set up to the customer by using the dashboard.

As can be seen from the tools discussed above, the number of monitoring tools that are available is staggering, and each tool comes with its own advantages and disadvantages. In this research three tools were chosen for further investigation, N-Able, CloudWatch and Sumo Logic. N-Able and Sumo Logic can both operate in hybrid environments i.e. Cloud and On-Premise whereas CloudWatch monitors AWS specific environments. These tools and the reasons why they were chosen are discussed further in the next chapter.

## 2.7 Big Data in Organizations

Many different types of organizations can benefit from the use of Big Data and Big Data Analytics. Tools can assist in quantifying the data from log files by aggregating and analyzing them to provide the organization with the information that they need to make sound business decisions (Kalakota, 2015). Just having data stored in log files that is used by monitoring tools to alert and provide visualizations is not enough, more in depth analysis is needed to provide value information that may help the organization in its quest to be strategically ahead of the pack.

All organizations have systems of some description in place to assist them with their day-to-day business undertakings. Organizations operating in all fields, such as Telecommunications, Healthcare, Retail, Insurance, Banking, and many more, can add value to the data that they collect by utilizing tools to aggregate this data into information. There is so much data available in an organization and the important factor is to consider how best to use the data and gain insight and information that can benefit the organization.

The impact of a system failure on an organization can be large, as well as the process of recovering from the failure can place undue pressure on time and resources (Fronza I. , Sillitti, Succi, Terho, & Vlasenko, 2013). Being able to predict failures and therefore attempt to mitigate the circumstances before they occur can limit the impact of failures. It is however challenging to provide the predictions, as accurately as required, in order to predict instances of failure. Log files are able to assist in this process as they contain a vast amount of data, on the current and past system state. It is also challenging to find the underlying cause of a failure or incident and can take a large amount of resources, and recovery may or may not be a very lengthy process. Having the information to predict potential failures can help in the mitigation of these types of events (Fronza I. , Sillitti, Succi, Terho, & Vlasenko, 2013). Log files can be used to provide the necessary information for these types of circumstances and issues.

Up to date information can make a big difference in an organization, especially when it comes to system management, including troubleshooting and investigating security issues (Mowlem, 2015). The size of an IT Team does not matter; what matters is how log files are analyzed. Log files capture what is happening on the systems that an organization runs and proper analysis of them will provide the correct information at the right time. Thus leveraging the information

gathered from the data contained in the log files can create an advantage for a business, as well as fostering a proactive approach to system.

## Chapter 3: Research Methodology

### 3.1 Introduction

A tool to monitor and alert on the basic events within an organization, such as system uptime and capacity, can enable an organization to better monitor their information systems. Additionally, interrogating the data captured by the tools provides additional benefits such as more in depth information on issues and possible causes as well as capacity planning.

The objective of this research was to gain an insight into how best to find and implement tools in order to enable organization to better understand and utilize the information that they have in their log data.

### 3.2 Research questions

The objective of this research led to two research questions that were formulated and investigated:

Q1: What capabilities of a monitoring platform should an organization consider while choosing one to implement.

Q2: How can log files be exploited by Big Data Analytics to gain benefits for the organization.

#### 3.2.1 Methodology

The methodology used in the research was of a qualitative nature. Qualitative research tends to aim towards gaining meanings which are more subjective in nature. Qualitative methodology is aimed more at exploring and discovering as opposed to testing a specific hypothesis (Hesse-Biber, 2015).

Qualitative research allows for a more exploratory format, opening up opportunities for discovery of emerging information (Campbell, 2014). Generally quantitative research will focus on data collection with precise measurements and analysis, whereas in contrast qualitative allows for open-ended methods.

Qualitative research is often based on more than one methodology and is underpinned by philosophical ideas supporting quality as opposed to quantity (Yilmaz, 2013). There are many

different ways in which qualitative methods can be employed to address the research that is being undertaken.

The main emphasis in this thesis was on looking at which monitoring platforms are available and then reviewing three of these in order to understand their capabilities and gauge their usefulness within an organizational environment. The words monitoring platforms and monitoring tools are used interchangeably throughout this thesis.

### **3.2.2 Criteria Used**

A number of criteria will be reviewed when implementing the tools. The criteria chosen to review the tools against were chosen because they would assist an organization in making a decision on which to implement. It is important to choose a tool that fulfils the needs of the business, is able to be managed and implemented by IT staff and the costs incurred are important (Hernantes, Gallardo, & Serrano, IT Infrastructure-Monitoring Tools, 2015). The criteria chosen are: ease of installation; ease of implementation and setup; log aggregation and reporting; value of use and cost.

All criteria used were measured by means of a scale of 1 and 5, 1 being the lowest score and 5 being the highest score.

Ease of installation will be measured by either installing the tool following the instructions from the vendor, or reviewing the already installed product. A certain amount of technical knowledge is assumed in this review.

Ease of implementation and setup will be measured by looking at the parameters that can be set up for monitoring and alerting. Review of how these are implemented will be undertaken. Again a certain amount of technical knowledge in this field is assumed.

By reviewing these tools, it is hoped to see what type of log aggregation, if any, is offered by each. Log aggregation provides longer term data that can be used to provide additional information. Reporting on various aspects that the platform monitors is important in a number of respects. Setting KPI's and reporting on these for management and capacity planning purposes is one aspect that should at a minimum be available.

The value of use of the platform will be indicated by what is achieved from the monitoring of the infrastructure. If monitoring is more on a reactive basis than a proactive basis it is possibly of lesser value as a tool. Another aspect of value is in the amount of administration required, if the platform is intense in administrative needs, it may not be of great value to an organization who does not have the resources to do this.

Cost is an important factor for any organization. Depending on the size of the infrastructure to be monitored the costs will vary greatly. Value for money is a good objective to have and the review of the platforms will give an overview of the costs involved for each. These costs however do not take into account any additional resourcing required to maintain and administer the platforms.

Some of these criteria can be quite subjective in that, what is easy for some may not be easy for others. However, the audience of this research are those who would have a relatively good background within the Information Technology industry and therefore my evaluation of the criteria should be comparable to the audience's.

#### *3.2.2.1 Selection of tools to investigate*

Monitoring tools provide the basis from which to monitor and alert on events and incidents that may occur within an organization's information systems. They provide data in log files about how an organization's infrastructure, systems and processes are working. Log files come in many different formats; they can be anything from unstructured to structured, or something in between. With monitoring and alerting tools, an organization can have an overview of the health of their systems as well as have the tool provide a centralized mechanism for troubleshooting alerts and for proactive monitoring.

From the overview of the tools in Chapter 2 it was decided to look at the three tools. The three tools are N-Able, CloudWatch and Sumo Logic. The first two, N-Able and CloudWatch, were chosen as these were either fully or partially implemented within the organization. I therefore had easy access to these to review and report on them. The third tool, Sumo Logic, was chosen for its capabilities in the log aggregation field, and that it had a number of features allowing for more in depth analysis of logs. The tools are implemented within an organization which is Healthcare related. The environments that are utilized for the implementation will range from live



production environments to a standalone test environment. The test environment will not be a true reflection of the production environment.

The following paragraphs give an overview of each of the tools reviewed in this research.

### **3.3 N-Able**

N-Able Technologies acquired by Solar Winds in May 2013, to enhance Solar Winds' remote monitoring and management offerings in the market. This acquisition also added MSP (Managed Service Provider) service automation to the range of IT management challenges that company was working to address for the IT community (Solar Winds N-Able, 2013).

N-Able (N-Central) is sold as a complete service delivery platform which can be utilized by small IT businesses as well as in large managed service providers (Solar Winds, 2015). The product can be used to automate monitoring and management of systems and logs. It is platform agnostic and operates in all types of environments.

### **3.4 Amazon CloudWatch**

The implementation of CloudWatch was part of a drive to move infrastructure into the AWS environment. CloudWatch is native to AWS and allows setting up of monitoring and alerting according to AWS practices. If an organization has a hybrid environment, CloudWatch could not monitor anything external to the AWS environment.

Amazon CloudWatch is an Amazon Web Services (AWS) product developed to monitor resources that are set up in AWS (AWS, 2015). The product allows for the collection and tracking of metrics as well as for the collection and monitoring of log files. It has the capability to set alerts and has a centralized monitoring console. Amazon CloudWatch allows for the monitoring of any resource on AWS including EC2 (Elastic Compute) instances, as well as applications and services, with resource monitoring to gain an overall view of resource utilization, performance of an application(s) and the overall operational health of the systems.

### **3.5 Sumo Logic**

Sumo Logic started in 2010, with a group of experts in the field of log management, scalable systems, Big Data as well as security. With the idea of monitoring very large sets of machine data, and using machine-learning algorithms to make sense of the masses of data, they developed the product called Sumo Logic (Sumo Logic, 2015).

Using Sumo Logic organizations can build analytical queries to assist them in transforming daily operations, which give better information for making decisions (Sumo Logic, 2015).

### **3.6 Evaluation of Tools**

Each tool of the platforms implemented were evaluated under each of the criteria, namely ease of installation; ease of setup and implementation; availability of log aggregation and reporting; value of use and the costs involved. For each of the criteria visuals of the tools are given in Chapter 5 so as to show the reader what this involved in the review. A score was given based on a scale of 1 to 5 for each of evaluated criteria, in order to show what each tool provides. The outcomes of this will be discussed in Chapter 6.

## Chapter 4 Implementation and findings

### 4.1 Introduction

This chapter will outline the implementation of the three tools, as well as their setup and configuration. An overview of installation and implementation are given. Each tool is compared against the criteria chosen. The findings compare the tools and give an overview of which can be used most beneficially depending on requirements.

### 4.2 Criteria Used

#### 4.2.1 Ease of installation

When installing a product it should be easy to do so, and little support should be required from a third party or the supplier. Instructions should be easy to follow and no major difficulties should be encountered.

#### 4.2.2 Ease of implementation and setup

If a product is easy to setup, manage and maintain, the costs associated therewith will be less, in that an internal resource (who the organization is already paying for) can be used. It also allows for greater flexibility in setting up the requirements of the organization without too much consultation and time wasting.

#### 4.2.3 Log aggregation and reporting

Reporting is a key component for management in an organization. Being able to report on KPI's that are set will give the management an overview of the environment and the ability to measure the success of the tool implemented. It also provides invaluable information with regards to capacity planning, which in turn allows for scoping and budgetary requirements to be met.

From log aggregation more of the data can be made use of by employing either additional tools, or tools inherent to the implementation. This allows for more in depth analysis of the data to be done, providing information to the organization.

#### 4.2.4 Value of use

When choosing a tool to implement and organization needs to look at the value the tool will bring. By implementing monitoring tools for its information technology environment, a gain

needs to be made. The organization will get value out of the tool if it provides the necessary information that they expect it to. Monitoring and altering will give a view of the health of the environment and give advance notice of issues that occur. These can be set proactively or reactively. Proactively would mean that there would be more value gained from the tool in that potential issues are dealt with prior to them becoming a priority incident.

#### **4.2.5 Cost**

Costs hit the bottom line in any organization, and these should be kept within the bounds of the budget set. Mostly the costs for monitoring tools will be an ongoing cost. Depending on the tool chosen a capital outlay may be required to purchase the hardware needed for the implementation. Licensing costs for the tool need to be taken into account as well as the model by which the licensing is sold, that is perpetual or annual. Perpetual licensing means the organization will own the license outright, and may or may not choose to pay software assurance in order to keep the license current. An annual license is one where the organization will never own the license but will use it whilst it pays for it. It is important to choose the right model, and this is usually dependent on the organizations financial standing and how it budgets for these types of costs. All in all if a tool is to provide value, the costs cannot outweigh the benefits received from its implementation.

### **4.3 Analysis of Chosen Tools**

Three tools were considered for inclusion in this study; two due to their ability to monitor and alert and the third based on its provisos to monitor and provide in depth analysis on data captured. The analysis presented below is on all three of the tools chosen.

#### **4.3.1 N-Able**

##### *4.3.1.1 Ease of Installation*

N-Able was the tool chosen to monitor and alert on the infrastructure by a Health Care related organization. N-Central is the software management console of N-Able which is installed on a centralized server within the infrastructure environment. This can be installed on a physical server or on a virtualized server. Solar Winds do however recommend that the installation be undertaken on a physical server in order to ensure performance metrics are met. The system requirements are linked to the number of devices that are to be monitored.

Once installed N-Central can be used to install N-Able agents onto devices to be monitored. The agents can be set to automatically deploy to devices that are set up within the organization's domain.

The down side of this is that if the device is not on the domain (for whatever reason) the agent needs to be added manually onto the device and the setup of the monitoring and alerting needs to be done from the central console.

As noted previously the installation had already been undertaken. The criteria for ease of implementation was measured against the guide from the Solar Winds. According to this documentation the installation steps looked to be straightforward. Feedback was also sought from the vendor who had installed the product within the environment. The general consensus from this vendor was that implementation was straightforward and the documentation provided by Solar Winds was easy to follow.

#### *4.3.1.2 Ease of Implementation and Setup*

Having agents automatically deployed is a great time saver, and when implementing new devices into an environment very little thinking needs to take place from a monitoring and alerting perspective. However, it is important to note that the initial installation of the agent on a new device sets the agent to monitor at the essentials level only. This means that if full monitoring and alerting as per other devices or infrastructure objects needs to be undertaken a little set up is required from the main console.

Setting up alerting makes use of standard templates. More intricate monitoring or alerting requires deeper knowledge of the product. This can be labor intensive and requires training on the product to ensure that the administrator of the tool is capable and knowledgeable.

It is also important to remember that not all devices need monitoring at the same level. That is, monitoring and alerting on production devices would generally have different requirements to those in a development or test environment. This depends on the organization's function. For example, does the organization care whether or not the development environment runs slow, or is that more important in the production environment? Generally, the rule of thumb is that production devices are more important and need more monitoring in place than a development or

other type of environment. The IT Team needs to liaise with the organization to get the right level of monitoring and alerting in place.

N-Able provides a centralized console – N-Central - to setup and view alerting and monitoring. From an administrative perspective, this provides value and saves time.

#### *4.3.1.3 Log aggregation and reporting*

The reporting functionality on N-Able is not as simple as it initially seems. A lot of thought needs to go into what is required from the reporting functionality. Writing reports requires knowledge of the tool and careful thinking about what metrics need to be included or not. The dashboards and graphs however do provide a good input to management reporting.

N-Able is able to store logs (if set up to do so) for a period of up to 7 years, this may allow for in depth analysis of data to provide more information. However, this does require a secondary database to be set up to store this data. It will also require someone who is trained and knowledgeable on Business Intelligence or Analytics to make use of this data properly. This does mean that an additional costly resource is required to assist with the analysis and output of required information.

#### *4.3.1.4 Value of Use*

N-Able as a monitoring and alerting tool is easy to install, implement and use. The future development that is planned for the tool may bring it more into the realm of being able to use Big Data Analytics in a more efficient and accessible interface. However, it would seem that not everyone could utilize the tool without some training and insight. The tool requires an IT user with training, and understanding of the Information Technology environment.

#### *4.3.1.5 Cost*

The cost of licensing for this software depends on the size of the implementation. The pricing is set a per device cost, which can add up to quite a substantial amount if the implementation is sizable. Depending on the options chosen when implementing the tool, the additional features that are provided can also add to the cost.

In this implementation no additional services other than monitoring and alerting were set up. The table below gives an indication of the cost for this implementation.

<b>Device Type</b>	<b>Cost per month</b>
Server x 240 @ USD9 per server	USD2,160
Reporting License	USD150
User access concurrent usage x 5 @USD50	USD250
Total	USD2,560

The cost per month for this implementation was relatively palatable considering the benefits of having monitoring and alerting in place.

### **4.3.2 CloudWatch**

#### *4.3.2.1 Ease of Installation*

To install CloudWatch you first need to set up and configure an IAM role or user for CloudWatch Logs. This needs to be set up in the AWS Console and within the region that you wish to implement monitoring on. In order to allow this user access to the logs, the role also needs to be set up. This can be done by following the steps on the installation guide of CloudWatch. The role can be set up using the script provided in the guide. After this has been done, you can connect to your EC2 instance and follow the guide for installing the agent on the particular operating system that your instance is running on. Depending on the operating system, the number of steps vary. After you have installed the agent you need to start the service on that instance. CloudWatch is now installed and running on your instance. By following the guide for installation the steps are well laid out and easy to follow. There are guides for each type of operating system that can be set up on AWS (AWS, n.d.).

#### *4.3.2.2 Ease of implementation and setup*

Setting up of alerting on this platform seems relatively uncomplicated, and there are many templates for filters and options. A more in depth knowledge of the tool, and infrastructure is required if additional non-standard monitoring and alerting is required. Alerting can be set up to automatically deploy on any new EC2 instance created. This is done via scripting a configuration file which is stored within a S3 bucket.

Implementation and setup can be done in a couple of ways. Via the CLI (Command line interface) or via the CloudWatch console. There are many metrics that are preconfigured that can be chosen and used to monitor the instance. After choosing the metrics that you wish to monitor you need to set the thresholds and alarms so that you are kept advised of any issues that may occur. Via the console you can create dashboards and graphs which will show the metrics that you have chosen to monitor. These provide a visual view of the metrics and monitoring that has been setup. If monitoring that is not template is required, you have the option of implementing a tool from the market place, or scripting requirements using an API or the CLI.

#### *4.3.2.3 Log aggregation and reporting*

Reporting is not exceptionally complex; however, it is not overly extensive. Additional tools are required in order to get more detailed reporting from the data that is stored by the tool. Setting up a good retention period means that the data can be stored for longer periods. Utilizing additional tools will give the option of being able to use the data to gain information long term and short term. Doing this will mean additional costs for the extra tools, as well as having a resource knowledgeable in the technology chosen.

The dashboards that are created provide reporting from a graphical and deep dive perspective. You can drill down into the metrics from the dashboard, as well as set up graphs which provide information on the environment being monitored. If detailed reports in formats that are easily printable or shareable are required, there are a number of tools on the market place within AWS that can be downloaded.

Log aggregation is done using another AWS centric tool called Amazon Kinesis Streams. This product allows you to aggregate the logs from your applications and can provide a real time response for data processing (AWS, n.d.). Logs from CloudWatch can be stored or archived into storage on AWS so as to provide long term data metrics.

#### *4.3.2.4 Value of use*

The tool itself can provide valuable insights into the infrastructure housed within an AWS environment, but has the limitation of not having visibility of other infrastructure within an organization. AWS are however constantly making improvements in their tools for customers, so this may change in the future.



Being a proprietary tool to AWS means that setting up and monitoring of AWS infrastructure is fully catered for by this tool. It allows for monitoring of all instances within AWS natively and as such should provide good value in this environment.

#### *4.3.2.5 Cost*

The costs for implementing CloudWatch vary according to the size of the environment to be monitored. When subscribing to CloudWatch basic monitoring of up to 10 metrics, alarms will be free. Up to 5GB of data ingestion from this monitoring is also free. When moving to more detailed monitoring the costs increase – it is also important to note that the costs differ in each region. By choosing detailed monitoring specifically within the Sydney region expect to pay around \$3.50 per month per EC2 instance monitored. Setting up alarms on the monitoring will incur an additional 10c per alarm per month.

The table below shows the approximate cost using the same number of servers as was implemented with N-Able to give a good comparison.

Device Type	Cost per month
EC2 instance x 240 @ \$3.50 per instance	\$840
Alarms* on instances 240 x \$0.10 per instance	\$24
Total	\$864

\*Based on only one alarm set per instance.

Costs for the use of CloudWatch seem comparatively low – however the more alerting configured the higher the costs. The storage of the data is free for up to 5GB, and thereafter the costs increase at a rate of approximately \$0.67 per GB ingested.

### **4.3.3 Sumo Logic**

#### *4.3.3.1 Ease of Installation*

The installation of Sumo Logic was in a test environment in AWS, utilizing a trial license. Gaining access to the download for the tool as well as the trial license was simple, with no major hiccups noted. The first step in implementing Sumo Logic is to obtain an account with Sumo

Logic. There are different types of accounts, and these are dependent on the size of the environment you are going to monitor. Once the account is up and running the collectors can be installed on the various devices that need to be monitored. The collectors can be implemented as hosted or installed locally. Hosted collectors are hosted in AWS. The locally installed collectors are deployed within the environment on a local machine. In this implementation both locally installed and hosted collectors were set up. The locally installed collectors were a little more difficult to implement than the hosted ones. However the documentation and troubleshooting guides provided by Sumo Logic are a good source for assistance. Ease of implementation and setup

Once collectors are installed monitoring and alerting metrics can be set up. There are many templates which can be used to assist with this process. In this installation the standard templates were used. It was noted that if logging levels were set high a vast amount of data would be sent to the Sumo Logic Cloud, and this could impact network traffic. However, this data could then provide more in depth information on the workings of the devices. No complex monitoring was setup, mainly due to the time that would be required to implement same, and that the trial was only for a month. Sumo Logic provides extensive documentation on their product available on the vendor website and many other options available for support if required.

#### *4.3.3.2 Log aggregation and reporting*

Sumo Logic provides a report generator which allows for data to be gathered and reported on either in a single or multiple report format. Reporting is based on the JSON and Sumo Logic provide sample configuration files which can be used to assist in setting this up. The reports are generated in an .xlsx (Excel) format. This allows for manipulation of data so as to gain information that is pertinent.

Sumo Logic was designed with log aggregation in mind. Data from the collectors is stored centrally and aggregated for further analysis. Sumo Logic provides the interface for viewing the aggregated information.

#### *4.3.3.3 Value of Use*

Sumo Logic collects data from multiple sources, provides a centralized interface for management purposes, and allows for pre-parsing of queries, as well as parsing of the data as it is gathered.

The vendor provides a number of collectors and API's which allow administrators ease of developing, and integrating data sources into the product. Sumo Logic provides an overall monitoring and alerting platform as well as a Big Data platform with Big Data Analytics as an integral part of the product.

#### *4.3.3.4 Cost*

The costs for Sumo Logic are dependent on the number of users that will be using the tool, as well as the amount of data that will be ingested. Options vary from the free tier to the enterprise tier. The professional tier which is in the middle offers up to 20 users and varying costs for the amount of data. Including in the costs are the implementation, support, alerts, extended retention to name a few.

The pricing model does not lend itself to the same comparative as done for the other tools investigated, that is a per device model. However looking at an approximation of around 5 users the costs would fall in the professional tier. See the table on page 81 for more details. Costs for the product increase as the size of the implementation increases.

#### **4.3.4 Parameters monitored**

A number of sample parameters were set up to be monitored by the tools during the setup phase. These are listed below:

- CPU utilization – monitoring of CPU usage can show how a system is performing as well as identify if there are any issues through alerting (Gill & Hevary, 2016).
- Memory Utilization – by monitoring memory utilization it can be seen whether or not the infrastructure is working optimally. If the utilization is high, additional memory resources may be required, or it may show that there is an issue with the application over-utilizing the memory allocated to it (Nemati, 2016).
- Disk Utilization – having information on disk utilization will give an indication of data growth within an organization. Having this information to hand can assist with capacity planning (Nemati, 2016).
- Uptime – system uptime is important, as it means that the system is available for use. The need to monitor uptime (and subsequently availability) is derived from the need for

the system to be available according to the service levels agreed upon (Treyner, Dhalin, Rau, & Beyer, 2017).

#### **4.3.5 Implementation and Results**

The tools were set up to monitor basic information on devices, like CPU, memory and disk utilization, as well as network statistics. More in depth monitoring and alerting was not set up, mainly due to constraints within the environments. In-depth analysis of log information on user behavior would not have been appropriate due to the privacy and data sovereignty required in the Health Care Industry. Additionally, delving into the data over a period of time was not possible due to time constraints. As such there may not have been enough data gathered to provide capacity planning information.

Monitoring tools within an IT infrastructure that provide overall log monitoring and alerting deliver a level of visibility to an organization of their systems and processes. These tools can save valuable time and effort for the organization in providing data that can be used to assist in early detection of issues. Monitoring tools provide a one stop overview of the environment thus allowing the IT team to be able to concentrate on more strategic and valuable implementations (Hernantes, Gallardo, & Serrano, IT Infrastructure-Monitoring Tools, 2015).

#### **4.3.6 Measuring the Value Added by Big Data Analytics**

The second question referred to finding out how log data could be used by applying Big Data principles to gain valuable information for the organization. If an organization wishes to have more visibility and understanding, it will need to look at investing in Big Data Analytics Tools.

In order to set up an environment that is conducive to analyzing Big Data, may mean a capital outlay for hardware and software, as well as the human resources that are knowledge in these tools. However, this is not the only option available. Organizations do not need to invest in hardware on premise; they can make use of the various provisions from Cloud Providers. Using the capacity provided including scalability and redundancy makes using a Cloud option a very palatable option. The vast amounts of data gathered during logging on operating systems, hardware platforms, applications and social media would normally require a lot of storage capacity. The decision to use Cloud or on-premises hardware and software combinations would depend on the organizations Capex/Opex models. It is also important to note that expertise,

whether in-house or external, will be required to setup and operate the environment in order to gain the advantages expected from the system.

N-Able as a monitoring tool provides data on the systems within the organization, however in order to delve into this data, it needs to be stored and in order to analyze the data, and additional tools are required.

Amazon CloudWatch allows data to be stored, within the Cloud Platform and use of additional tools can assist with correlation and analysis of the data captured within the log files. There is an additional cost associated, not only for additional storage and retention, but also for tools that can do this in depth analysis.

Sumo Logic is able to do log aggregation and analysis as a standard out of the box feature. The tool revolves around monitoring and alerting, but at the same time doing in depth analysis, in addition to the use of algorithms to detect early warning, and other metrics that turn the data into information for an organization.

In order to be able to gain a more in depth analysis and understanding of the data stored in log files, Big Data Analytical tools should be employed. These tools can take the Big Data that is stored in log files over time, and provide information to an organization that can be invaluable. Utilizing Big Data Analytical tools allows for far more proactive monitoring to be achieved than what can be by standard monitoring and alerting tools alone.

#### **4.3.7 Summary**

The analysis chapter (Chapter 5) will show that there is much worth to be gained by having monitoring and alerting tools implemented. These tools can provide invaluable information from the data in the log files that they have a view of. When in troubleshooting mode these tools can provide time saving benefits, so that an IT engineer would not have to troll through thousands of lines of logs. With all the tools reviewed there is a central dashboard with many overviews and reporting to give guidance during troubleshooting.

The table below summarizes the findings according the criteria that was reviewed.

	<b>Ease of Installation</b>	<b>Ease of implementation and setup</b>	<b>Log aggregation and reporting</b>	<b>Value of use</b>	<b>Cost</b>
N-Able	3/5	3/5	3/5	4/5	3/5
CloudWatch	4/5	3/5	3/5	4/5	4/5
Sumo Logic	4/5	3/5	5/5	4/5	4/5

All of the tools reviewed provide a centralized console from which to manage and view events. This made the administration and viewing much easier than having to jump from screen to screen.

The tools reviewed can be used to provide proactive monitoring and provide essential information with regards to capacity planning and how the systems are being used. By using the information contained in the log files, the tools allow the organization to look at system capacity used over time and draw up capacity planning strategies with relative ease. However in order to look at the data from the log files more in depth, either additional tools are required, or by implementing a tool like Sumo Logic which inherently has this as standard would be the best option.

## Chapter 5 Analysis

### 5.1 Introduction

The three monitoring tools that are reviewed during the course of this thesis are N-Able, Amazon CloudWatch and Sumo Logic. When looking at which of the monitoring platforms will provide the most benefit to an organization there are many points that need to be investigated. The criteria chosen in this thesis, as being representative of what organizations generally look for when evaluating a new technology are: Ease of installation, Ease of implementation and setup, Log aggregation and reporting capabilities, Value of use, and Cost.

This chapter discusses how the criteria were evaluated for each of the three chosen monitoring platforms.

### 5.2 Evaluation of Monitoring Platforms

This section will evaluate each of the monitoring tools reviewed using the above criteria. The tools were implemented in a healthcare-related organization, and as such some details had to be obscured in the screenshots below for purposes of confidentiality and privacy. Table 6 at the end of the chapter summarizes the evaluation.

#### 5.2.1 N-Able

As discussed in Chapter 2, N-Able consists of a suite of products including N-Central for monitoring, which provide organizations with MSP service and technology platforms (Solar Winds, 2015).

##### 5.2.1.1 *Ease of Installation*

The Installation of N-Central which is the automated monitoring and management tool provided by N-able, would seem to be simple and easy. The guide gives steps for different types of environments and servers. Installation is usually on a central server, whether it be on a virtual environment or a physical platform. (N-Able) N-central needs to have connectivity from Internet in order to cover the entire network, it should however be placed behind a firewall on a private IP address with forwarding enabled.



Figure 1 - NAT-based N-Central configuration (N-Able)

Steps for installation include setting up the network settings, so that they use the Fully Qualified Domain Name (FQDN), this allows for communication across the organizations domain. The time zone settings need to be configured and synchronized across the network. The server licensing needs to be activated as part of the installation, and this can take up to 24 hours.

The server needs to have backup configuration set, and these should preferably be held in another location. The default is the C drive of the server. At this point setting up access for system administrators and other staff can be undertaken. It is best that each have their own login for tracking and auditing purposes.

#### 5.2.1.2 Ease of implementation and setup

N-Central comes with a number of pre-configured service templates which help with the set-up of standard monitoring. However additional templates can be created making use of the specialized knowledge within the organization of what needs to be monitored. The information from these service templates can be displayed on Dashboards which are totally customizable. N-Able recommends the use of profiles when setting up notifications for various alerts from the monitoring templates. This allows the organization to tailor the various types of alerts from critical to informational, and set up notifications to different groups or individuals.

Service templates can be associated to the device being monitored and there are many pre-determined templates and new ones can be added and configured to suit the device. Figure 2 lists some of the templates. New templates are being developed to cover additional devices by the Solarwinds N-Able team (Zenz, 2015). The new templates for monitoring include services or devices like Lync Server 2013, SharePoint 2013, etc.



**Associated Service Templates**

Service Template	Services
<b>Dell Servers</b>	Fan (Dell), Logical Drive (Dell OM 2.2), Memory Status (Dell), Physical Drive (Dell), Power Supply (Dell), RAID Status (Dell), Server Temperature (Dell)
<b>Fujitsu Servers - Windows</b>	CPU Status (Fujitsu), Fan Status (Fujitsu), Logical Drive (Fujitsu), Memory Status (Fujitsu), Physical Drive (Fujitsu), Power Supply (Fujitsu), RAID Controller Status (Fujitsu), Temperature Status (Fujitsu)
<b>IBM Servers</b>	Fan Status (IBM), Logical Drive (Adaptec), Memory Status (IBM), Physical Drive (Adaptec), RAID Status (Adaptec), Server Temp (IBM)
<b>IIS 6.0</b>	HTTP, HTTPS, IIS, IIS Application Pool, IIS Website Metrics, Process (WMI), SMTP, Windows Event Log
<b>Intel Servers</b>	Fan Status (Intel), Logical Drive (Intel), Physical Drive (Intel), Power Supply (Intel), Server Temp (Intel)
<b>Patch Status Windows Servers - Remove</b>	Patch Status
<b>Servers - Uptime</b>	Uptime
<b>Servers - Windows Base Monitoring</b>	Agent Status, CPU (WMI), Connectivity, Disk (WMI), Disk I/O, Memory (WMI), Patch Status, Uptime
<b>Terminal Service Monitoring</b>	Windows Service

Adds or Modifies Services Removes Services

Figure 2 - N-Able Service Templates

If the device to be monitored is a Dell server, the Dell Server's template can be chosen. Within this template there are many options of what can be monitored on the Dell server, as shown in figure 3.

Name:

Description:

Details Associated Devices Associated Rules

Device Class: Servers - Windows

**Service Template Services**

Service:

<input type="checkbox"/> Name	Service	Action	Automatically add new instances
<input type="checkbox"/> Fan (Dell) (ESM MR Fan1 RPM)	Fan (Dell)	Add or Modify	No
<input type="checkbox"/> Logical Drive (Dell OM 2.2)	Logical Drive (Dell OM 2.2) (Asset Information)	Add or Modify	No
<input type="checkbox"/> Memory Status (Dell)	Memory Status (Dell) (Asset Information)	Add or Modify	No
<input type="checkbox"/> Physical Drive (Dell) (Physical Disk 0:0)	Physical Drive (Dell) (Asset Information)	Add or Modify	No
<input type="checkbox"/> Power Supply (Dell) (Power Supply 1)	Power Supply (Dell)	Add or Modify	No
<input type="checkbox"/> RAID Status (Dell) (PERC 3/D)	RAID Status (Dell) (Asset Information)	Add or Modify	No
<input type="checkbox"/> Server Temperature (Dell) (ESM Frit IO Temp)	Server Temperature (Dell)	Add or Modify	No

Figure 3 - N-Able Dell Server Template

Within the templates tab, associations to other devices can be made, including setting up of standard or configured rules that would be connected with the device and its associations. This allows for showing connectivity between devices, as well as which devices are impacted should this device be unavailable. This type of understanding is optimal for any organization as it gives a view of the connectivity and inter-operability between devices allowing for impact analysis during changes as well as assessment of risk and impact should a device fail. Associated Rules allows for rules to be set up to monitor that particular device.

To allow the N-Able agents and probes to gain access N-Able server ports 80, 443 and 22 need to be opened. WMI and SNMP will also need to be enabled to allow for the monitoring and reporting of the various devices within the organization. The probe can be implemented on different levels, namely domain, workgroup or local. In most organizations this will be domain, or a combination of the classifications. This means that all devices on the domain, within the organization can be monitored by N-Central, as well as other devices that may not be on the domain.

The agents can be pushed out to the devices to be monitored automatically, and this is the recommended implementation model. However, this may mean that devices that are not required to be monitored will have the agent automatically installed on them; these can be removed if not needed. All production instances should be monitored, some organizations may well want to monitor certain development and test environments, but often these environments are changing and may cause false alerting and gathering of “noise”. Some of these systems could only need to be available at certain times of the day and N-Able can be set to monitor/alert as required for these types of devices.

The agent tab has information on the configuration of the agent, see figure 4. This information includes installation and upgrade details, which allows for a quick view to ensure that the agent on the device is the latest version. Asset scan details are also shown here, and can be modified from this tab. Data persistence is also shown under Agent, detailing how long data is kept for, as well as the minimum and maximum size on disk that is allowed to be utilized for monitoring purposes.

## Configuration Details

Activation Key:	aHR0cHM6Ly9nc25jLmxleGVsLmNvLm56OjQ0MyxodHRwciovLzYwLjIzNC40My4xMjA6NDQzfDEyMTUwMjQzNzV8MXww	?
Agent Installer:	<a href="#">WindowsAgentSetup.exe</a>	
* Endpoint Network Address:	[REDACTED]	?
Endpoint Protocol:	HTTPS	?
* Endpoint Port:	443	?
Proxy String:		?
Log Level:	3 - Normal	?
Appliance ID:	1215024375	?
Update Monitored Address:	<input checked="" type="checkbox"/>	?
Last Login:	2015-Nov-08 12:55	

## Installation & Upgrade Details

### Agent

Last Action:	Install
Installed Agent Version:	9.5.1.243 (Software is up to date)
Installing Probe:	[REDACTED] Windows
Update New Version:	Always
Status:	Completed
Detail Status:	.NET4 already installed. Agent installation succeeded.
Log:	<a href="#">Log</a>

## Asset Scan

Perform Asset Scan:	<input checked="" type="checkbox"/>	?
Start Time:	10 : 42	
Days of the Week:	<input checked="" type="checkbox"/> Sun <input checked="" type="checkbox"/> Mon <input checked="" type="checkbox"/> Tue <input checked="" type="checkbox"/> Wed <input checked="" type="checkbox"/> Thu <input checked="" type="checkbox"/> Fri <input checked="" type="checkbox"/> Sat	
Monthly On Days:	<input type="text"/> (Ex. 1,15 or 1-3)	

## Data Persistence

* Drop Policy:	<input checked="" type="radio"/> Oldest <input type="radio"/> Newest	?
* Data Window:	48 Hours 0 Minutes	?
* Minimum Disk Space:	10 MB	?
* Maximum Disk Space:	20 MB	?

Figure 4 - N-Able Agent Information

N-Able can be configured with out of the box templates for a variety of monitoring purposes. These include server monitoring – basics like CPU, Disk and Memory thresholds. Additional monitoring can be put in place and customized in order to monitor critical events or services. It can also be tailored to restart services should they fail; this is called self-healing in N-Able

terminology. This can be further customized to attempt to restart a number of times and only alert if the restart does not occur due to an error or a certain amount of retries.

Monitoring is centralized and can be viewed on a single dashboard. The functionality of the dashboard allows for drill down into various components giving a view of what has been configured.

## Properties Tab

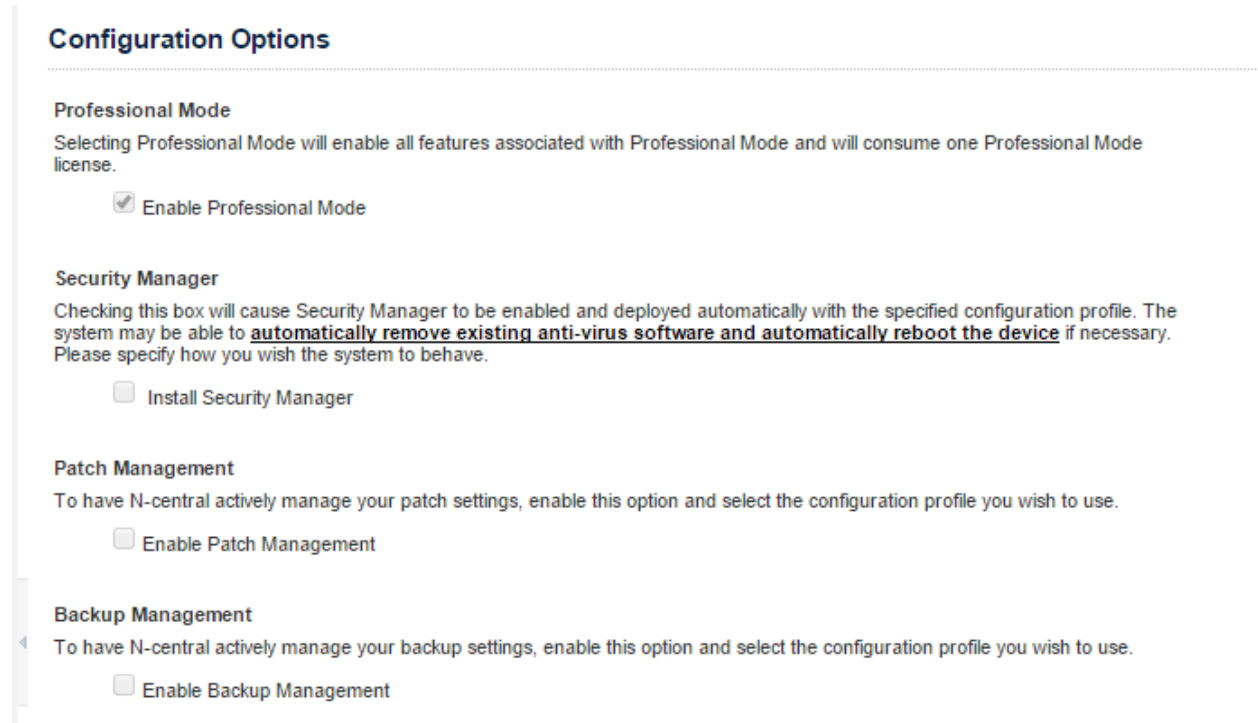
The screenshot shows the 'Properties' tab of a server configuration interface. At the top, there is a header bar with a green 'd' icon, a computer icon, and the IP address '10.10.61.8'. Below this, a note states: '\* Required Field' and 'Note: You have logged in using an User account, which has read-only privileges.' To the right of the note is a 'Remote Control' button with a blue icon and the text 'Click to connect'. Below the header is a horizontal menu with tabs: 'Properties', 'Status', 'Asset', 'Maintenance Windows', 'Service Templates', 'Associations', 'Custom Details', 'Notes', 'Agent', 'System Change', and 'Remote Control'. The 'Properties' tab is selected. Below the menu is the 'Device Settings' section, which contains the following fields: 'Given Name' (text input with value 'AHCALDCVHOST01'), 'Use Discovered Name' (checkbox), 'Discovered Name' (text input with value 'AHCALDCVHOST01'), 'Class' (dropdown menu with value 'Servers - Windows'), 'Network Address' (text input with value '10.10.61.8'), 'Remote Access URI' (text input), 'Operating System' (dropdown menu with value 'Microsoft Windows Server 2012 Datacenter x64 Edition'), 'User Name' (text input with value 'Administrator'), 'Password' (text input with value '(unchanged)'), and 'Show Password' (checkbox).

Figure 5 - N-able Drill Down View

As can be seen from figure 5 which gives information on a particular server, there are many options that can allow for more information to be gathered about the device. Please note that the server name and username has been obscured for privacy reasons.

When clicking into an object on the dashboard the user is present with the options above. The first tab gives details on the properties of the object including Given Name, Discovered Name, Class, Network Address, Remote Access URL, Operating System, User Name and Password. Below the above screen there are additional options for configuration showing how the monitoring has been set up and allowing for modifications to be made.

The Configuration options for this object are shown below:



**Configuration Options**

---

**Professional Mode**  
Selecting Professional Mode will enable all features associated with Professional Mode and will consume one Professional Mode license.

☒ Enable Professional Mode

**Security Manager**  
Checking this box will cause Security Manager to be enabled and deployed automatically with the specified configuration profile. The system may be able to automatically remove existing anti-virus software and automatically reboot the device if necessary. Please specify how you wish the system to behave.

☐ Install Security Manager

**Patch Management**  
To have N-central actively manage your patch settings, enable this option and select the configuration profile you wish to use.

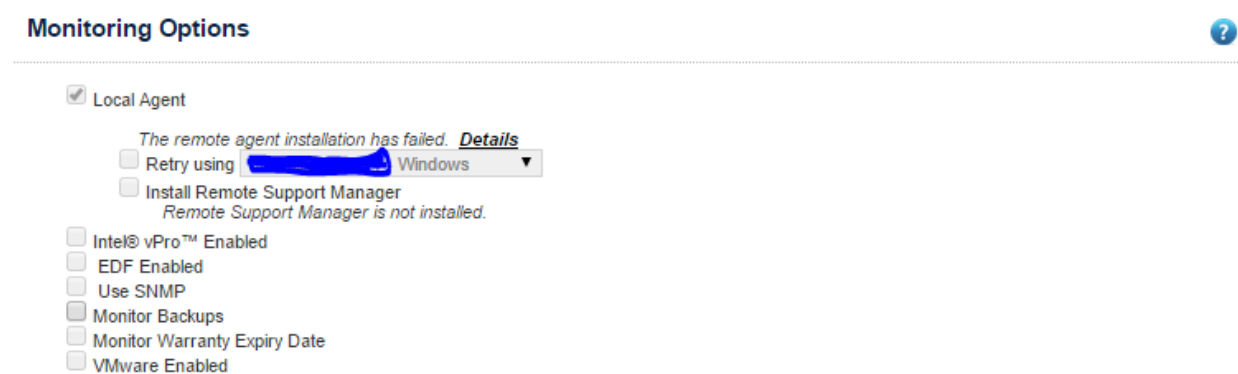
☐ Enable Patch Management

**Backup Management**  
To have N-central actively manage your backup settings, enable this option and select the configuration profile you wish to use.

☐ Enable Backup Management

Figure 6 - N-Able configuration Options

Monitoring options for this object:



**Monitoring Options** ?

---

☒ Local Agent

The remote agent installation has failed. [Details](#)

☐ Retry using Windows

☐ Install Remote Support Manager  
Remote Support Manager is not installed.

☐ Intel® vPro™ Enabled

☐ EDF Enabled

☐ Use SNMP

☐ Monitor Backups

☐ Monitor Warranty Expiry Date

☐ VMware Enabled

Figure 7 - N-Able Monitoring Options

N-Able also has options that can be set for Downtime for planned/unplanned outages:

**Downtime**

☒ Off

☐ One Time Only

Start Time: October 11 2015 11 : 54

End Time: October 12 2015 11 : 54

☐ Recurring

Start Time: 00 : 00

End Time: 00 : 00

Days of the Week: Sun Mon Tue Wed Thu Fri Sat All Clear

Monthly On Days:

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	All	Clear		

☐ Unscheduled Downtime: Change all services on this device to a Disconnected state ( ) when the Agent Status service transitions to a Failed state. The Agent Status service will then also be changed to a Disconnected state.

Figure 8 - N-Able Downtime Options

## Status Tab

Clicking on through from the object page to the Status page shows details of various properties that exist for the object, including what is being monitored, its status, including details of the last time the scan was initiated on the object for the particular service.

**Services**

Service	Status	Transition	Probe/Agent	Last Scan Time
<a href="#">Agent Status</a>		2015-Oct-05 21:12	Central Server Asset	2015-Oct-11 12:08
<a href="#">Connectivity</a>		2015-Oct-08 14:09	Windows	2015-Oct-11 12:07
<a href="#">CPU (WMI)</a>		2015-Oct-11 07:32	Windows	2015-Oct-11 12:02
<a href="#">Disk (WMI) - C:</a>		2015-Oct-08 14:04	Windows	2015-Oct-11 12:08
<a href="#">Disk (WMI) - D:</a>		2015-Oct-08 14:04	Windows	2015-Oct-11 12:08
<a href="#">Disk I/O - Total</a>		2015-Oct-08 14:04	Windows	2015-Oct-11 12:08
<a href="#">LocalIP</a>		2014-Sep-30 10:50	Local Agent	--
<a href="#">Memory (WMI)</a>		2015-Oct-08 14:04	Windows	2015-Oct-11 12:08
<a href="#">Uptime</a>		2015-Oct-08 14:04	Windows	2015-Oct-11 12:09
<a href="#">Windows Service - Cluster Service</a>		2015-Oct-08 14:04	Windows	2015-Oct-11 12:09
<a href="#">Windows Service - Hyper-V Virtual Machine Management</a>		2015-Oct-08 14:04	Windows	2015-Oct-11 12:09

Figure 9 - N-Able Status Tab

Additional information on each of the services above can be seen by clicking on the service. For example, it may be necessary to view more detail on CPU Utilization to gain an understanding of possible performance issues.

Clicking on the CPU service item will bring up another screen, as seen in figure 10, which gives more detail about the CPU usage on the server in question. There are additional tabs here again to allow for more in depth investigation. The first tab is the Status tab, which gives information on the current status of the service. There is also an associated graph which gives information on the CPU utilization by the Top 5 Processes on the server. The graph also presents an historical view so as to enable the viewer an overall picture of the usage on the CPU over the past number of days. This information is configurable and more or less data can be viewed according to requirements.

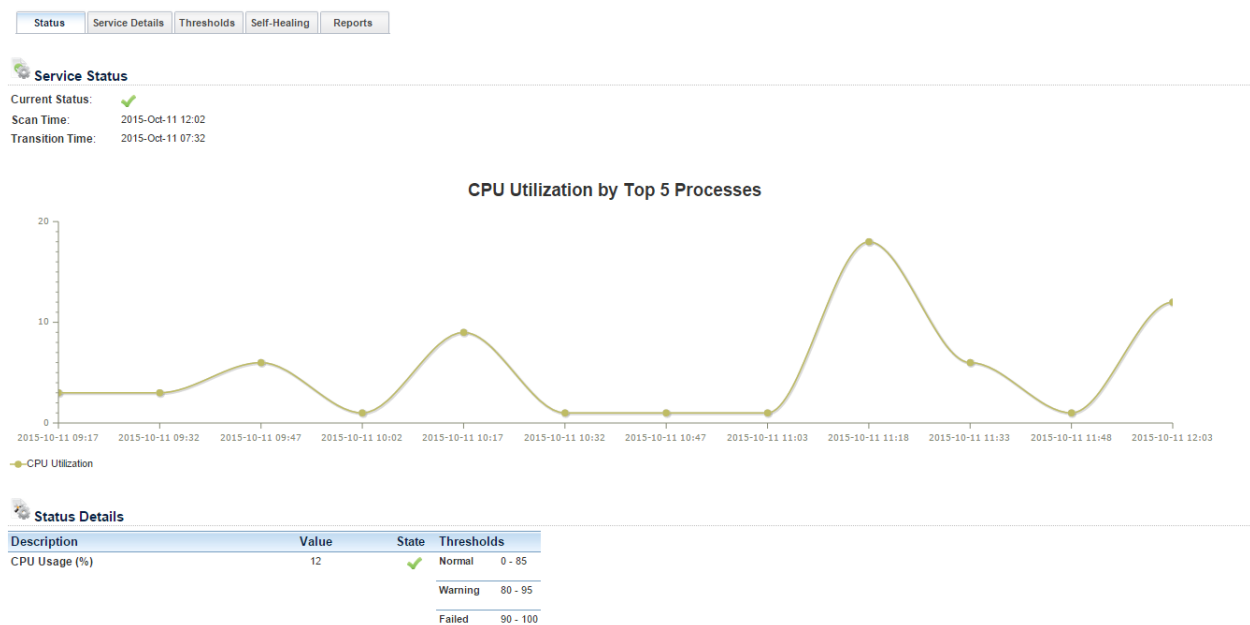


Figure 10 - N-able Object Status Tab – CPU Utilization

Service Details give more information on the monitoring of the service.

StatusService DetailsThresholdsSelf-HealingReports

Details

\* Monitoring:Enabled

\* Time To Stale:45

Service Description:Added by service template [ Servers - Windows Base Monitoring ]

\* Scan Interval:15

\* Processor Name:CPU0

Figure 11 - N-Able Object Service Details Tab – CPU Utilization

The Thresholds tab shows what values have been setup for the service alerting.

StatusService DetailsThresholdsSelf-HealingReports

Thresholds

If you set the Monitoring field for a threshold to Off, the threshold is no longer used to determine the status of the service. If the Monitoring field for all of the thresholds is set to Off, the service status always displays as Normal, regardless of its true state (e.g. Failed or Warning).

CPU Usage (%):

Monitoring:Normal

Range:0 - 65.535

Normal:0 - 85

Warning:80 - 95

Failed:90 - 100

Cancel

Figure 12 - N-Able Object Thresholds Tab - CPU Utilization



The Self-Healing tab gives information on what has been set up to alleviate the problem should it occur.

Status	Service Details	Thresholds	<b>Self-Healing</b>	Reports
--------	-----------------	------------	---------------------	---------

<b>Trigger</b>	Action	Notification	Results
----------------	--------	--------------	---------

---

### When the status changes ?

---

From

Failed  
Warning  
Normal  
Misconfigured  
Stale

>>  
>  
<  
<<

To

---

### Before Self-Healing Action

---

Retry Monitoring: ☐

Scan the service:  times

Wait between each scan:  seconds

---

### Self-Healing

---

Execute Self-Healing action: ☐

---

### After Self-Healing Action

---

After executing Self-Healing wait:  Minutes

Scan the service:  times

Wait between each scan:  seconds

Figure 13 - N-Able Object Self-Healing Tab - CPU Utilization

The asset tab in figure 14 gives information pertaining to the asset. This allows for a view of what is currently setup and configured on a device, including software, patching and provides reporting information on the asset. Additional fields allow for additional information to be entered and reported on, for instance Warranty expiry date, etc. and are configurable to the needs of the organization. N-Able can as such be utilized as a CMDB (Configuration Management Database), allowing an organization to keep track of assets, software, warranties and licensing information and renewals.

#### Asset Summary

---

Make and Model:	Microsoft Corporation ( Virtual Machine )
Serial Number:	8704-1994-3587-4360-3125-5177-96
Operating System:	Microsoft Windows Server 2012 Datacenter
CPU speed:	Intel(R) Xeon(R) CPU E5-2660 0 @ 2.20GHz (2.20GHz)
System Memory:	4.00GB
Hard Drive Capacity:	C: 59.66 GB
Device Created On:	2014-09-19 09:59
Last System Warranty Discovery:	2015-10-17 10:47
Last Discovery:	2015-10-17 10:47
Logged in User:	

Figure 14 - N-Able Asset Summary - Asset Tab

The figure below shows the Asset Details – details can be expanded individually or in their entirety to give a view of what components are parts of the server or object. In this screen shot the object is a server and the fields that are shown include Computer System, Processor (expanded to show more details), Memory, Motherboard, BIOS, Network Adapters, Logical Drives, Printers, Ports (expanded to show more details), Operating System, Services, Patches, Applications, OS Features, Roles/Features and MSSQL (i.e. present or not).

**Asset Details** ?

☐ Expand all

Component	Count
<input type="checkbox"/> Computer System	1
<input type="checkbox"/> Processor CPU ID: CPU0 Name: Intel(R) Xeon(R) CPU E5-2660 0 @ 2.20GHz Description: Intel64 Family 6 Model 45 Stepping 7 Max clock speed: 2.20GHz	1
<input type="checkbox"/> Memory	2
<input type="checkbox"/> Motherboard	1
<input type="checkbox"/> BIOS	1
<input type="checkbox"/> Network Adapters	1
<input type="checkbox"/> Logical Drives	3
<input type="checkbox"/> Physical Drives	2
<input type="checkbox"/> Printers	1
<input type="checkbox"/> Ports Generic SQL Server: 1433 HTTP: 80 HTTPS: 443 ping: 0 Windows Terminal Server: 3389	5
<input type="checkbox"/> Operating System	1
<input type="checkbox"/> Services	139
<input type="checkbox"/> Patches	22
<input type="checkbox"/> Applications	48
<input type="checkbox"/> OS Features	2
<input type="checkbox"/> Roles/Features	13
<input type="checkbox"/> MSSQL	1

Figure 15 - N-Able Asset Summary - Asset Details

The configurable asset details for this particular server have not been configured, however the screenshot below shows what has been set up for capture, and if needed can be activated.

Configurable Asset Details

Warranty Expiry Date:

Lease Expiry Date:

Expected Replacement Date:

Purchase Date:

Cost:

\$

Location:

Asset Tag:

Description:

Network device discovered using Asset Discovery - 2041331023

Figure 16 - N-Able Asset Summary - Configurable Asset Details

Maintenance Window Tab

Scheduled maintenance can be viewed on this tab for the object, see figure 17. There are two types of maintenance of that can be configured automatically. The first is Schedule Maintenance, predefined actions can be set in motion by N-Able in order to undertake scheduled maintenance, and at the same time the object will be put into scheduled maintenance mode so that alerting can be halted during this period.

The second type of maintenance window is for scheduled reboots on the object. This can be set up for a scheduled time and day, and the reboot will only be allowed to occur during the specified window.

Maintenance Windows

The N-central Agent can be configured to automatically complete tasks for various N-central features according to a defined schedule. The two types of windows that can be created are:

Scheduled Maintenance: The device will be allowed to initiate the selected actions for the selected feature.

Scheduled Reboot: The device will reboot according to the schedule. The reboot will only be allowed to occur during the window specified.

Name	Associating Rule	Last Modified By	Last Modified Time	Type	Schedule	Status
<div><div></div>There are currently no Maintenance Windows</div>						

Figure 17 - N-Able Maintenance Windows

## System Change

The system change tab shows information related to recent changes made with regards to the device being monitored. Each device may have different properties which are monitored for change. In this instance it is a server device, and information pertaining to Network, Processor, Physical Drives, Computer system, Operating system and Motherboard is shown. Below is a screen capture showing details for the Motherboard.

Property: Motherboard				
Baseline:				
manufacturer	product	version	serialNumber	biosVersion
Microsoft Corporation	Virtual Machine	7.0	9074-1267-5183-9824-5858-4002-25	-
Current:				
manufacturer	product	version	serialNumber	biosVersion
Microsoft Corporation	Virtual Machine	7.0	9074-1267-5183-9824-5858-4002-25	-

Figure 18 - N-Able System Change

## Remote Control

The remote control tab allows for remote connectivity to the device. There are a few options under the connection type – and what is used would vary according the device. The default is Remote Desktop, but other options like ssh, telnet and web page can be chosen.

### Remote Control Settings

Allow Remote Control: ☒ ?

\* Time Out in Minutes:  ?

Connection Type:  ?

Connect Using:  ?

\* Target Port:  ?

\* Application to Run:   ?

Figure 19 - N-Able Remote Control

### 5.2.1.3 Log aggregation and reporting

Data gathered by N-Able can be stored for up to 7 years. Reports for executive or technical levels can be pulled from this information (n-able, 2015). This data can give insight into many areas of the business that need attention. Reporting at an executive level gives insight into various aspects of the organization's information technology infrastructure. Using this data to report on aspects such as network and security can show where issues are prevalent and this can then be used proactively to rectify and improve the services that are provided.

In order to get more value from the data that is collected and stored, that is aggregation of logs, it is suggested that the data be moved to data warehouse storage and analytical tools be utilized to provide the in-depth analysis required.

The reporting tool within N-Able gives access to a number of options of reporting as shown in figure 20 below.

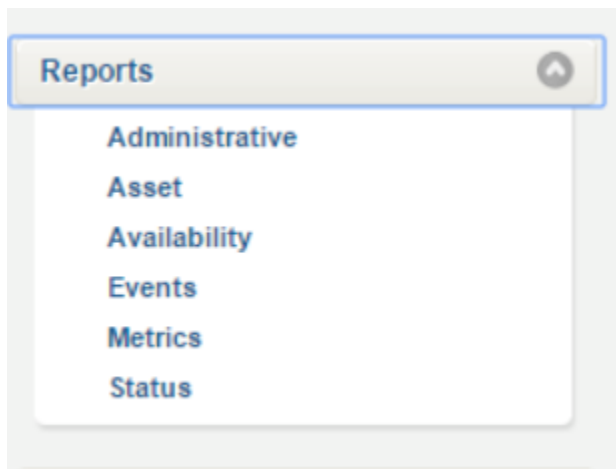
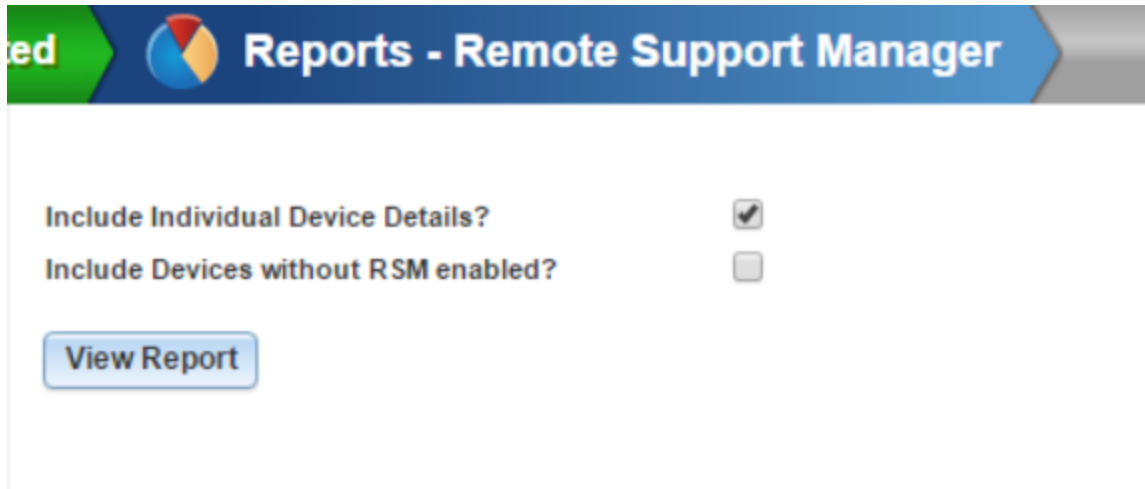



Figure 20 - N-Able Reports Tab

#### **Administrative**

Administrative reporting gives an overview of licensing information as well as which devices have the tool deployed on them. The report can be run with a couple of options, i.e. including device details and show devices without RSM (Remote Support Manager) enabled, as in figure 21.



ed  **Reports - Remote Support Manager**

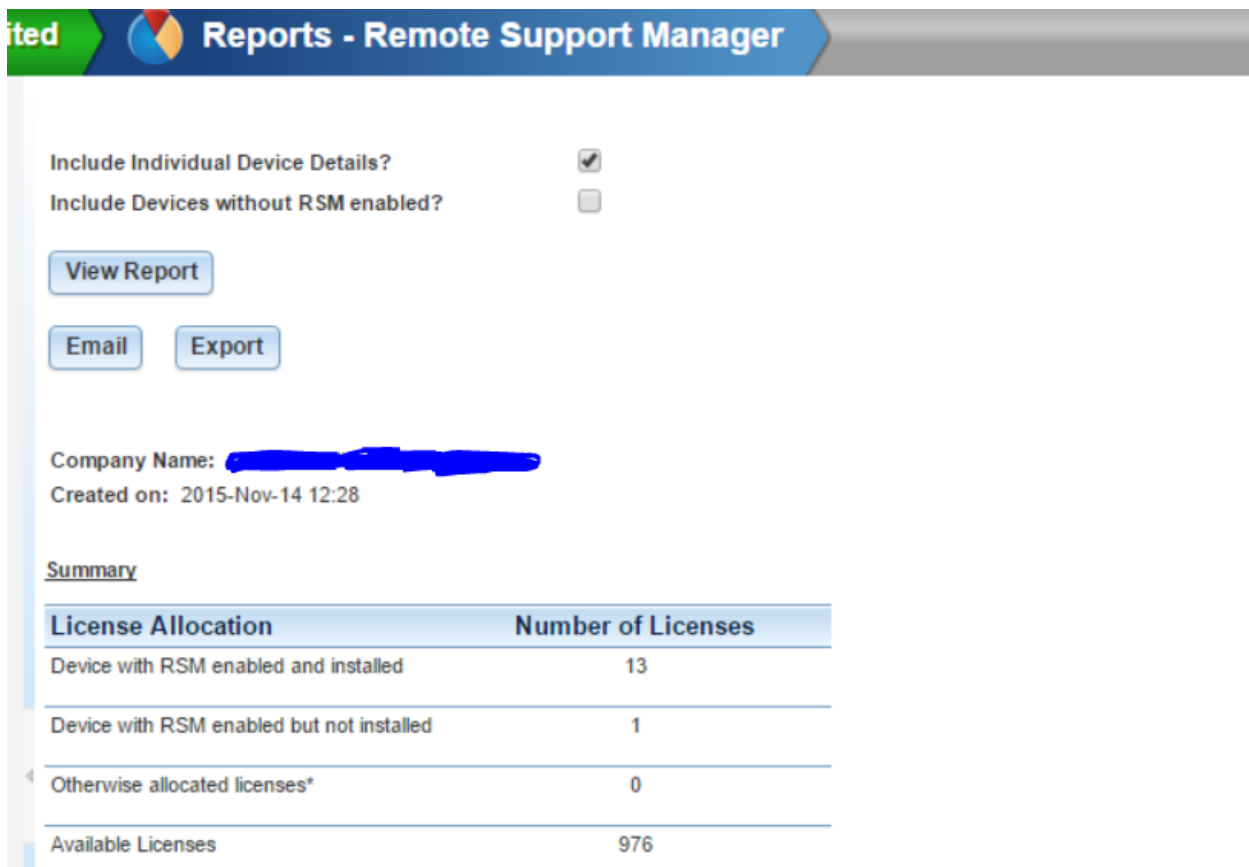
Include Individual Device Details? ☒


Include Devices without RSM enabled? ☐

**View Report**

Figure 21 - N-Able Reports - Remote Support Manager

Clicking on “View Report” will bring the report up on screen. Options to email or export the report are presented on this screen.



ted  **Reports - Remote Support Manager**

Include Individual Device Details? ☒

Include Devices without RSM enabled? ☐

**View Report**

**Email** **Export**

Company Name: [REDACTED]

Created on: 2015-Nov-14 12:28

Summary

License Allocation	Number of Licenses
Device with RSM enabled and installed	13
Device with RSM enabled but not installed	1
Otherwise allocated licenses*	0
Available Licenses	976

Figure 22 - N-Able Reports - View of Report

The report can be exported to PDF or CSV formats. Export to PDF would give a non-editable format, and CSV will allow editing and manipulation of the information presented in the report.

The email option allows for the report to be emailed in PDF or CSV format to an email recipient or multiple email recipients.

## Asset Reporting

This tab, in figure 23, gives options to run various reports including more detailed information on the asset, compliance issues detected on the asset amongst others.

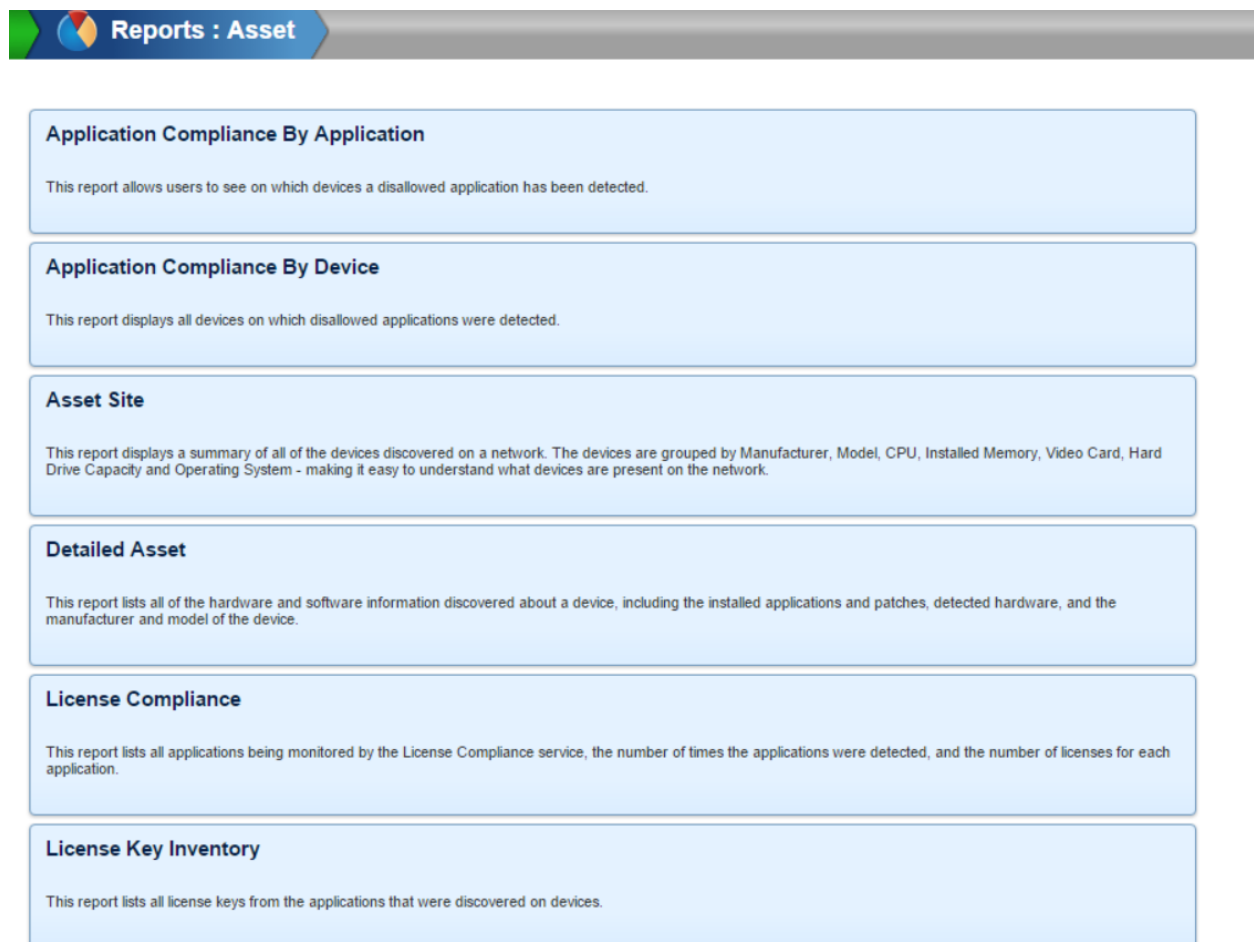


Figure 23 - N-Able Reports - Asset Reporting

These reports also have options to be emailed or exported, either in PDF or CSV format.

The first option “Application Compliance by Application” requires some prior setup in order to be value. It will show applications that have been installed on devices that are not in the



organizations approved list. This is valuable in that certain compliance and licensing issues can be avoided if installation of software that is not authorized is detected sooner rather than later.

The “Application Compliance by Device” will show which devices have unauthorized applications loaded on them. As above this report will give details of which devices have had applications installed on them that are not in the compliant listing. This will allow the organization the opportunity to remove any unauthorized applications before issues arise.

“Asset Site” gives some very valuable information with regard to the devices deployed across the organization. Assets are grouped according to manufacturer and various other components that are present. This gives the organization an overview of the devices on the network. This information can be exported into CSV format and uploaded into the organization’s CMDB (Configuration Management Database); if this is done regularly the CMDB will be fairly up to date. Using a CMDB is important as it gives an overall or holistic view of all the IT assets that are deployed throughout an organization (Guddemi, 2014). A CMDB should be used to evaluate the impact of changes that occur in the environment. As such a CMDB is used to serve two purposes, that of change control/change management and disaster recovery. It is also imperative to ensure that the CMDB includes all Cloud deployments (Sheppard, 2015). In this article Sheppard states: “An accurate, well-maintained Configuration Management Database (CMDB) can help to reduce the risks of the cloud transition and support day-to-day operations and maintenance processes.”

Alternatively, the N-Able software can be used as the CMDB if the organization does not have an external source for this information to be held in.

The “Detailed Asset” report gives additional information on the devices deployed, including software that has been deployed on the device, what patch levels the devices are on, as well as the manufacturer and model information. This information run regularly could also be an input into the organization’s CMDB.

Audits of licensing should be undertaken at least quarterly to ensure that the organization is in fact licensed for all the applications or software that is installed and in use. The “License Compliance” report gives details of the software/applications installed on devices as well as licensing information. Each piece of software loaded on a device unless freeware must be

licensed. All software vendors are by law able to audit an organization in order to ensure that the number of licenses purchased for their software tallies up to the number of devices the software is deployed on (Prescott, 2015). If an organization is under licensed the software vendor has the right to claim the licensing costs that are outstanding, and back date them. Some vendors even add additional fines onto this, and as such it can be very costly if you are not licensed properly, often even more expensive than being properly licensed.

Having information centrally of all license keys that are deployed on devices is important. Sometimes a device crashes and needs to be reinstalled; having information on license keys in a centralized repository is very helpful in these situations. It is also useful for auditing purposes. Running the “Licensing Key Inventory” report will give details of the license keys for the software loaded on the various devices. However, it would probably be good to export these to CSV and store them in a secure location in case the device is lost, stolen or destroyed. N-Able relies on the device being online and on the relevant network to probe for information. As such having an external copy of this information – regularly updated – would be very useful.

## Availability

A variety of availability reporting is provided under this section.

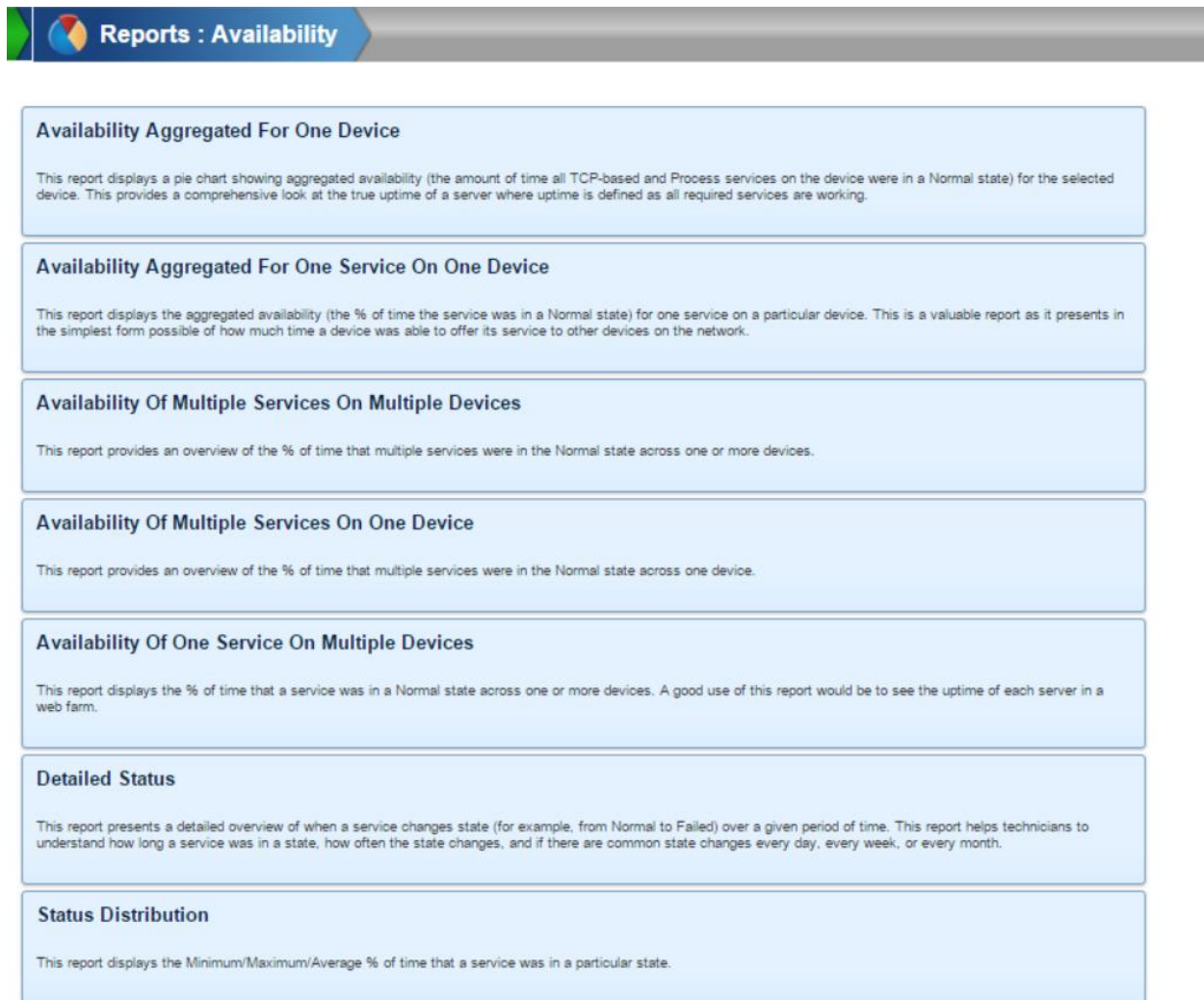


Figure 24 - N-able Reports - Availability Reporting

“Availability Aggregated for one Device” will provide in depth availability information for a specific device for a specified period of time, from a starting point. The information is presented graphically in the form of a pie chart



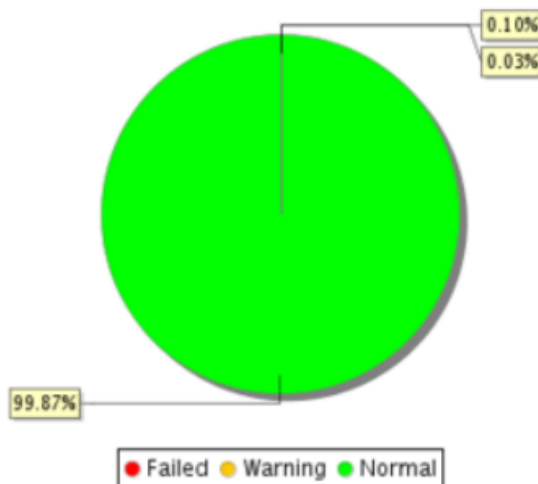
Figure 25 - N-Able Reports - Availability Aggregated for Device

Graphs and charts are always good ways to give a visual view of a networked environment especially for reporting to senior executives.

The “Availability Aggregated For One Service On One Device” Report can provide additional information about the availability of the device, by delving into each service. See below – the report was run for the same time period, same device but for the service of Connectivity.

Company Name: [REDACTED]  
 Report: Availability Aggregated For One Service On One Device  
 Period: 1 Month  
 Start Date and Time: 2015-Oct-14 14:00  
 Device: 10.40.10.1-10.40.10.1  
 Service: Connectivity [REDACTED] Windows

### Connectivity - [REDACTED] - Windows



State	Time	%
Misconfigured	0 Seconds	0.00%
Stale	0 Seconds	0.00%
No Data	0 Seconds	0.00%
Disabled	0 Seconds	0.00%
Disconnected	0 Seconds	0.00%
Failed	44 Minutes 21 Seconds	0.10%
Warning	15 Minutes 4 Seconds	0.03%
Normal	743 Hours 0 Minutes 33 Seconds	99.87%

Figure 26 -N-Able Reports - Availability Aggregated for One Service on One Device

The reports screen under the device shows details about the device and the monitoring that has been included. Drill options include seeing monitoring statistics over a time period, quick view of what is being monitored, graphical interface showing monitoring statistics and a table of uptime statistics for the device. As can be seen from the below uptime monitoring of a device can provide very useful data for reporting within the organization.

This particular uptime table shows the services being monitored, as well as the Uptime of the Service by Minimum, Maximum and Average. All services except for the Agent Status service have had 100% uptime. This report should be used to investigate why the Agent Status has had some downtime.

Service	Uptime of Service (%)		
	Minimum	Maximum	Average
Agent Status - Central Server Asset	92.80	100.00	98.97
Connectivity - [REDACTED] Windows	100.00	100.00	100.00
CPU () - Local Agent	100.00	100.00	100.00
Disk (C:) - Local Agent	100.00	100.00	100.00
Disk (E:) - Local Agent	100.00	100.00	100.00
Disk I/O (_Total) - Local Agent	100.00	100.00	100.00
HTTP () - [REDACTED] Windows	100.00	100.00	100.00
IIS (_Total) - Local Agent	100.00	100.00	100.00
IIS Website Metrics (_Total) - Local Agent	100.00	100.00	100.00
IIS Website Metrics (Default Web Site) - Local Agent	100.00	100.00	100.00
IIS Website Metrics (Microsoft Dynamics CRM) - Local Agent	100.00	100.00	100.00

Figure 27 - N-Able Device Report Drill Down

The Reports Tab under the each object allows for additional reporting on the object and the particular service. It gives the flexibility of viewing the report, exporting it or emailing it. The period for which the reporting is required can be chosen here.

Status	Service Details	Thresholds	Self-Healing	Reports
--------	-----------------	------------	--------------	---------

Period : 1 Month ▼

Start Date and Time : July ▼ 11 ▼ 2015 ▼ At: 13:00 ▼ Hours

\* Scan Details:

CPU Usage

☒ Overlay Thresholds

View Report

Email

Export

Company Name:   
 Report: null  
 Period: 12 Hours  
 Start Date and Time: 2015-Oct-11 00:00  
 Service: CPU (WMI)  
 Devices:

### CPU Usage (%) Vs. Time

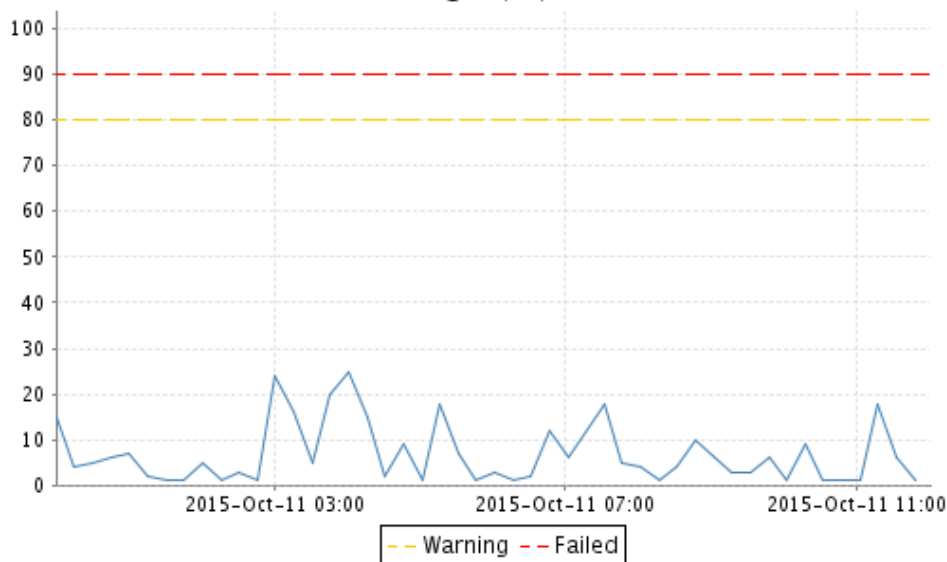


Figure 28 - N-Able Object Reporting Tab - CPU Utilization

#### 5.2.1.4 Costs

N-able offer Cloud or on-premises implementations of their software.

*Table 1 - N-Able Costs*

<b>Tool</b>	<b>Average Per Month Cost (in USD)</b>
Attended Remote Control	\$50 per concurrent session
Help Desk	\$23 per technician
Reporting	\$150



The cost for monitoring devices is set as a per device license fee as shown in Table 2:

*Table 2 N-Able Licensing Costs*

Service	Average Per Device Per Month Cost (in USD)
<b>Device Management</b>	
Managed Server	\$9.00
Managed Network	\$4.50
Managed Workstation	\$2.10
<b>A-La-Carte Services</b>	
Managed Security	\$1.00
Managed Patch – Applications	\$0.25
Managed Path – Windows	\$1.00
Managed Mobile	\$1.00
Managed Backup	Priced per GB or Server (provided upon request)
Managed Compliance	\$1.00 per IP

N-Able also offers different types of services, see Table 3, which can then be used to provide an organization with either a robust or basic offering.

*Table 3 N-Able Service Offerings and Costings*

Service	Average Monthly Cost Per Device (in USD)
Light Monitoring <ul style="list-style-type: none"> <li>• Light Monitoring</li> <li>• Asset/Software/Hardware reporting</li> <li>• Branded sys-try icon</li> <li>• Remote Control</li> </ul>	Free
Managed Security <ul style="list-style-type: none"> <li>• Industry Leading AV protection</li> <li>• AV Monitoring and Updates</li> <li>• AV Threat/Status Reporting</li> </ul>	\$1 per device
Managed Patch <ul style="list-style-type: none"> <li>• Windows Patch Management</li> <li>• Patch Status Reporting</li> </ul>	\$1 per device
Managed Mobile <ul style="list-style-type: none"> <li>• Mobile management and support</li> <li>• Mobile reporting</li> </ul>	\$1 per device
Managed Backup <ul style="list-style-type: none"> <li>• Offsite storage to secure datacenter</li> <li>• Backup status reporting</li> </ul>	Provided on request only
Managed Compliance	\$1.00 per IP

Service	Average Monthly Cost Per Device (in USD)
<ul style="list-style-type: none"> <li>Monthly vulnerability and compliance reporting</li> </ul>	

### 5.2.2 Amazon CloudWatch

Amazon Cloudwatch is a product used for monitoring AWS cloud resources and applications that are run within AWS (AWS, 2015). It can be used to collect and track metrics as well collect and monitor log files. It also has the facility of setting alarms up for thresholds that are defined. The product is available to monitor AWS resources including, EC2 (Elastic Compute); DynamoDB tables; RDS DB instances and can also be customized for an organization's applications and services. Amazon CloudWatch gives an overall view, system-wide, into the utilization of resources, performance of applications, as well as the overall operational health of the systems.

#### 5.2.2.1 Ease of Implementation

Setting up Amazon CloudWatch is relatively simple and there are a number of steps set out in the instructions from AWS that need to be followed.

The first step is to install and configure the Amazon CloudWatch Log Agents to send logs to Amazon CloudWatch. The agent can be installed on a new EC2 instance or one that is already running, it can also be undertaken from Amazon CloudFormation or Chef (AWS, 2015).

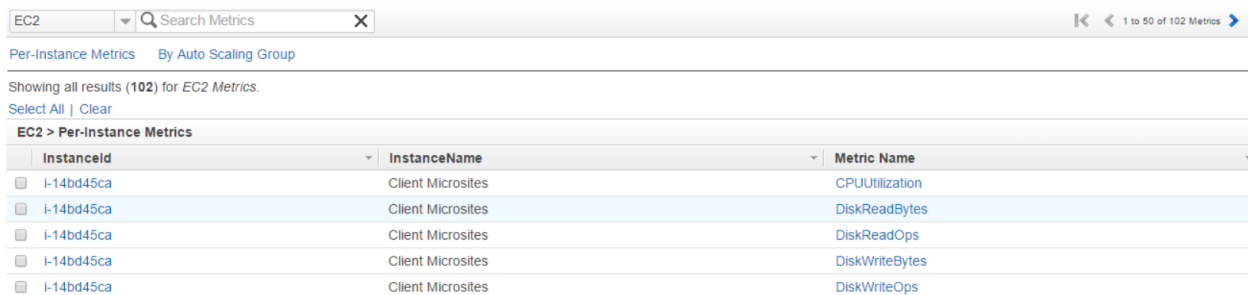
If the organization has Amazon EC2 instance(s) they are automatically registered for Amazon CloudWatch, which means that the EC2 instances use the basic monitoring. If additional monitoring is required, the organization can enable detailed monitoring when launching the instance. It can also be enabled at a later stage by right clicking on the instance and selecting to enable this option (AWS, 2015).

#### 5.2.2.2 Ease of implementation and setup

There are many resources that have free and automatic monitoring in place on AWS. Elastic Load Balancers set up in AWS are automatically monitored with Amazon CloudWatch for metrics such as latency and request count. Amazon RDS DB instances are also automatically monitored; the metrics include free memory and available storage space (AWS, 2015).

An organization can set up their own metrics that need to be monitored for their custom applications, etc. This is achieved by using API calls, which are simple and easy to set up (AWS, 2015).

Amazon CloudWatch provides a number of metrics that are available out of the box to be used to monitor various objects in AWS. Metrics for EC2 instances for example, include CPUUtilization, DiskReadBytes, DiskReadOps, and many more. The metrics that are available are also reliant on the area where the AWS instances are serviced from (Amazon, 2015).



InstanceId	InstanceName	Metric Name
i-14bd45ca	Client Microsites	CPUUtilization
i-14bd45ca	Client Microsites	DiskReadBytes
i-14bd45ca	Client Microsites	DiskReadOps
i-14bd45ca	Client Microsites	DiskWriteBytes
i-14bd45ca	Client Microsites	DiskWriteOps

Figure 29 - EC2 Dashboard

## Monitoring

From within CloudWatch, there are options to set up Dashboards, Alarms, Logs, and Metrics. These can all be tailored to requirements.

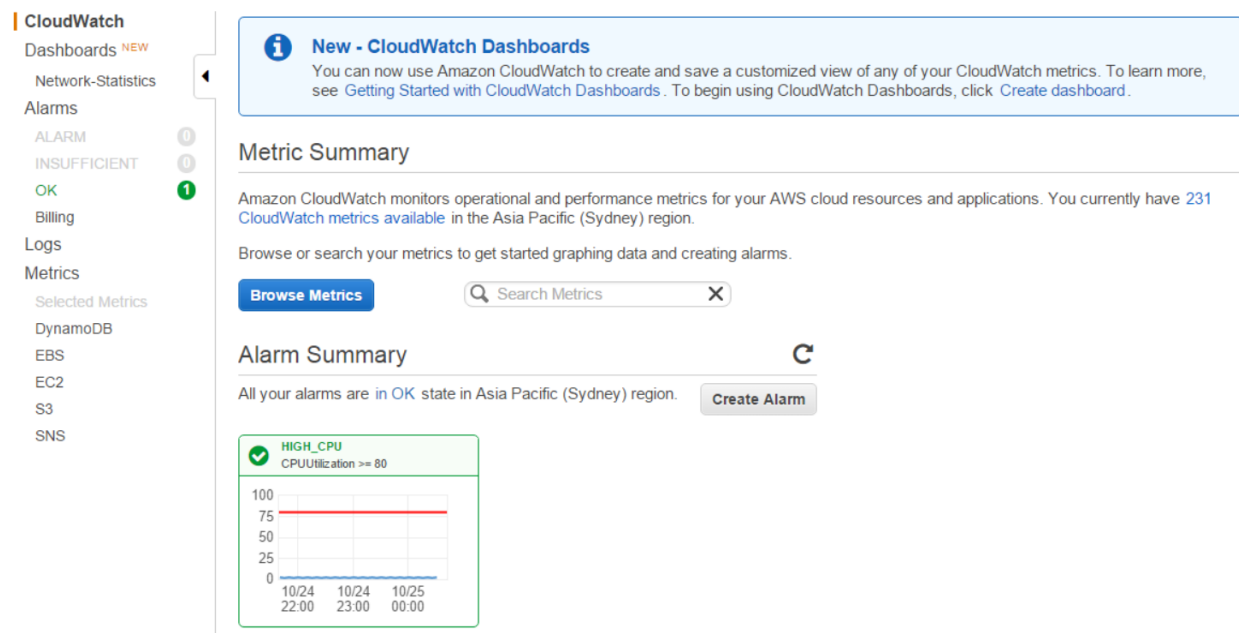


Figure 30 - CloudWatch Main Landing Page

The above Figure shows the details of the status of any alarms that have been configured and gives options to set up dashboards and so on.

In this particular environment (host in Sydney) three Dashboards have been configured. There is one Dashboard for CPU Utilization, one for Disk Writes and another showing Network Statistics (incoming and outgoing).



Name	Last updated (UTC)
<a href="#">CPU-Utilization</a>	2015-10-24 00:03
<a href="#">Disk-Writes</a>	2015-10-24 00:58
<a href="#">Network-IN</a>	2015-10-25 00:51
<a href="#">Network-OUT</a>	2015-10-25 00:52

Figure 31- CloudWatch Dashboards Page

Click on the CPU-Utilization Dashboard a quick view of the state of this metric is given in graphic format. The time range of the metric can be adjusted to view more or less in the graph. In the figure below the graph shows 7 days' worth of data collected on the CPU Utilization of a server.

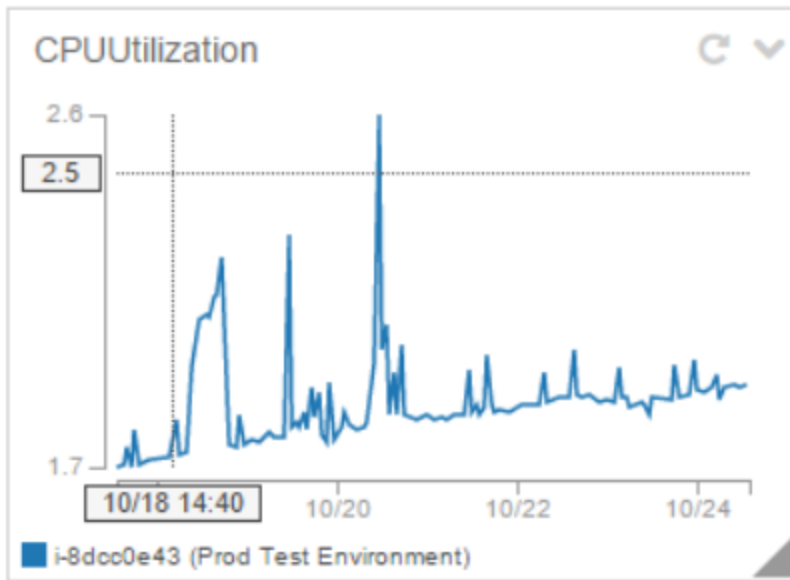


Figure 32- CloudWatch - CPU Utilization of Server "Prod Test Environment" over the last 7 days

Looking at the DiskWriteOps Dashboard for the same period shows that this server is not performing high disk write operations.

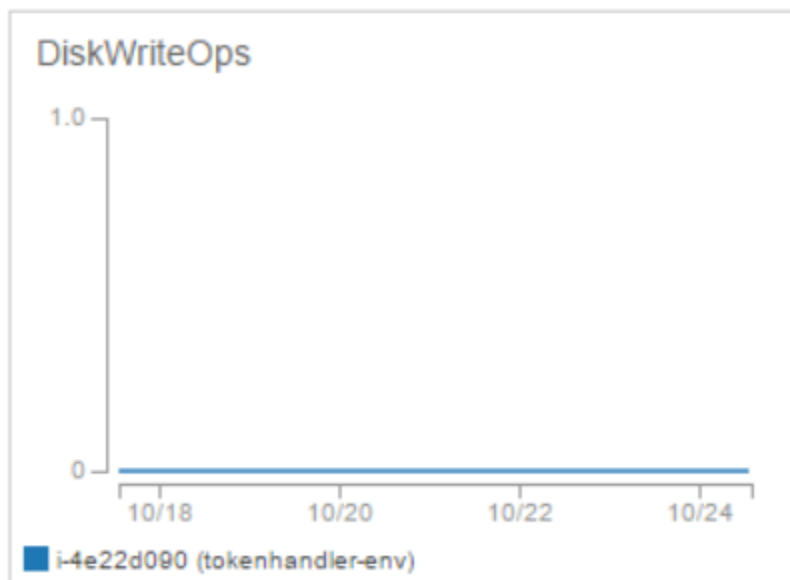


Figure 33 - CloudWatch - DiskWriteOps of Server "Prod Test Environment" over the last 7 days

The Dashboard for Network-IN shows that for the same period most of the incoming network activity occurred on the 20<sup>th</sup> October 2015 at approximately 22:06.

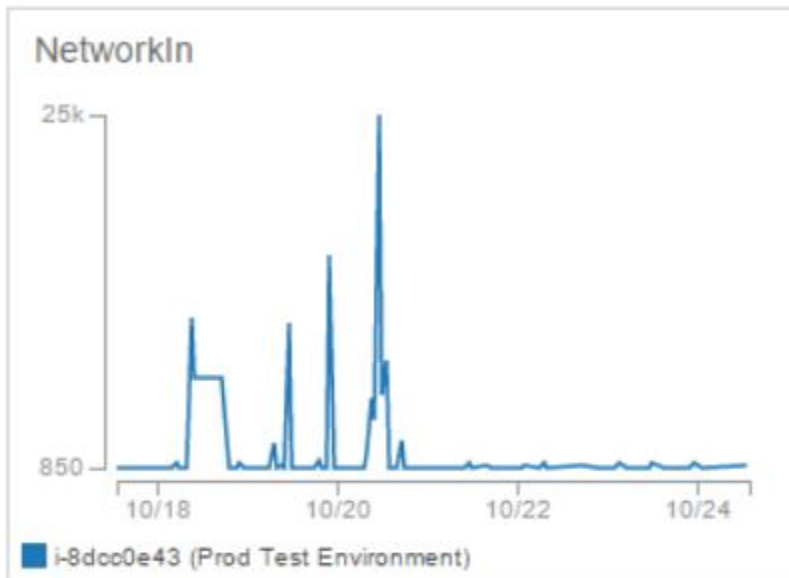


Figure 34 - CloudWatch - NetworkIN of Server "Prod Test Environment" over the last 7 days

The NetworkOUT dashboard in figure 35 shows similar results – i.e. that the most outgoing traffic was on the 22<sup>nd</sup> October 2015; however the peak was at 08:06 of that day.

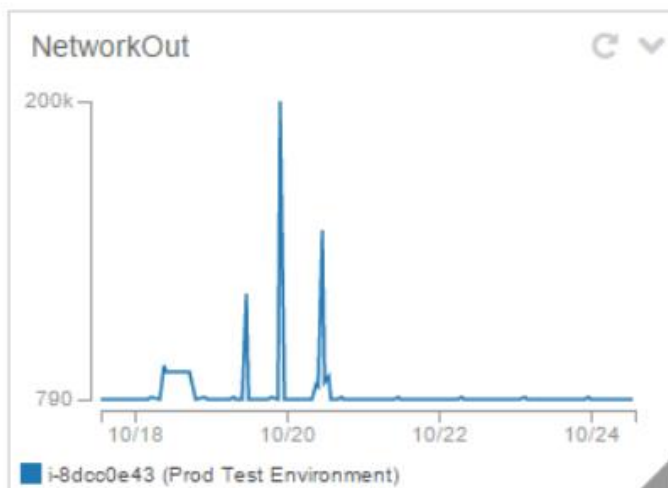


Figure 35 - CloudWatch - NetworkOUT of Server "Prod Test Environment" over the last 7 days

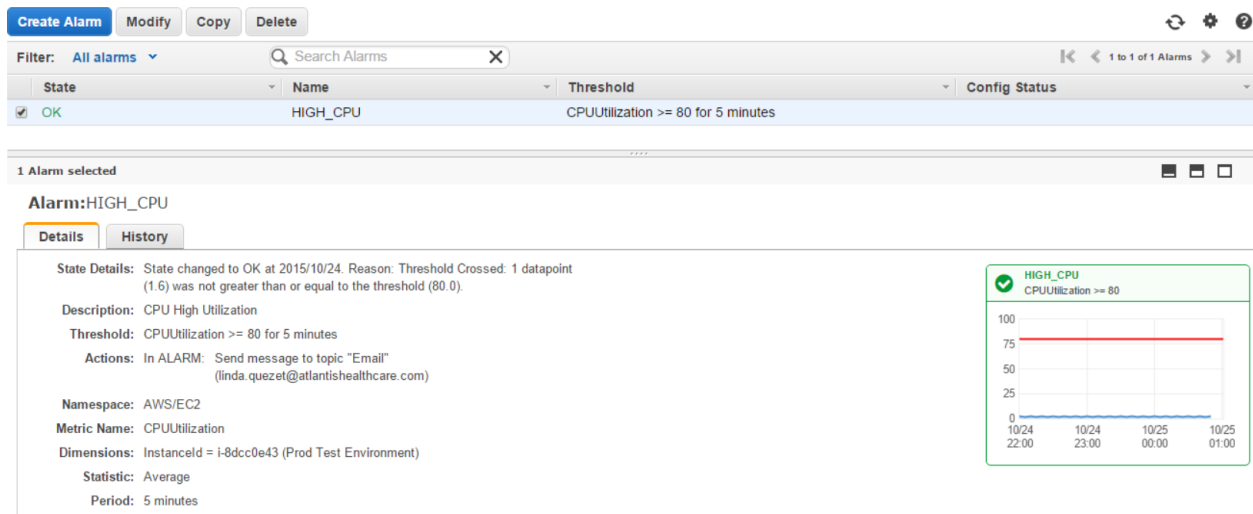


Figure 36 - CloudWatch - Alarm - High CPU for Server "Prod Test Environment"

The above is a depiction of CPU utilization that is set to a threshold of 80%. The alarm will only be activated if this threshold is maintained for 5 or more minutes.

### 5.2.2.3 Log aggregation and reporting

Additional tools are required to make full use of the metrics provided under CloudWatch. There are a number of plug-ins that are available on AWS, mostly third party developed tools, which give access to more than just standard log and event information. CloudWatch does not natively provide this level of detail.

One such tool that AWS promotes is Logentries, which partnered with AWS to provide centralized and real-time analytics integration (Logentries, 2014). Logentries collects and centralizes information gathered from CloudWatch and other AWS native tools such as CloudTrail. CloudTrail is a:

“Web service that records AWS API calls for your account and delivers log files to you. The recorded information includes the identity of the API caller, the time of the API call, the source IP address of the API caller, the request parameters, and the response elements returned by the AWS service” (AWS, 2015, p. 1).

Logentries was developed specifically for the cloud and is built using open architectures so that it can support connectivity and integration to AWS. The tool also supports the real-time aggregation and correlation of logging information that is generated within AWS (Logentries, 2014).

CloudWatch reporting is pretty basic in that once setup viewing metrics for the monitored events and services is provided via the dashboard. CLI (Command Line Interface) code can be used to specify utilization or other metrics for a particular instance or service that is being monitored by CloudWatch. This can be undertaken for various time periods, and setup to show average, minimum, maximum, sum or data samples of what is being monitored (AWS, 2015). The reports are graphically represented and show utilization or metric being monitored. Below is an example of the dashboard, and graphical representation of CPUUtilization.

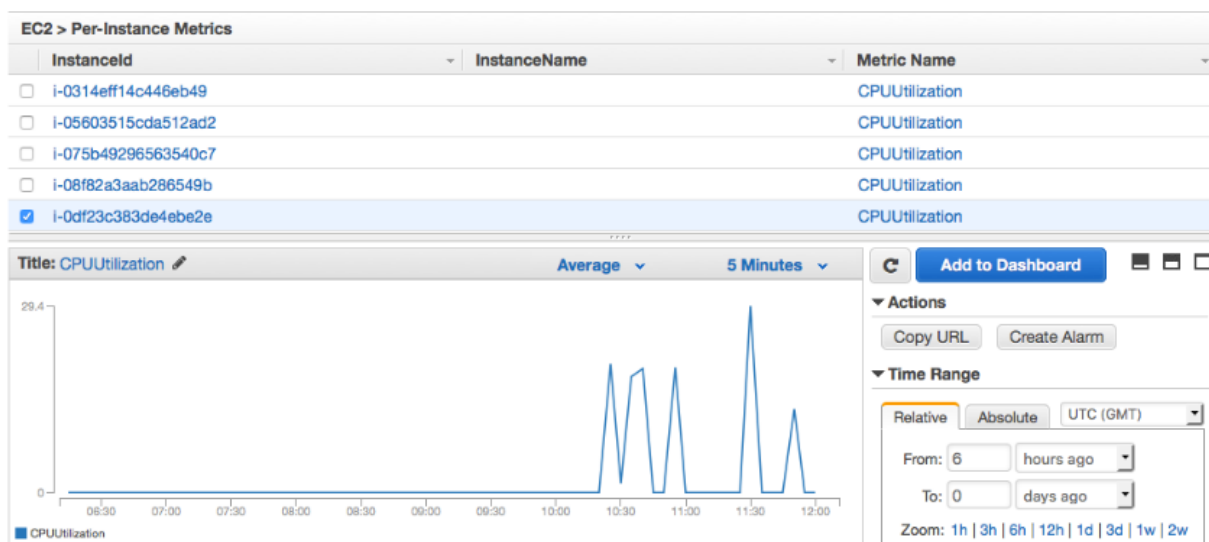


Figure 37 - CPU Utilization Graphic (AWS, 2015)



There are a couple of options that can be utilized to show more details as shown below.

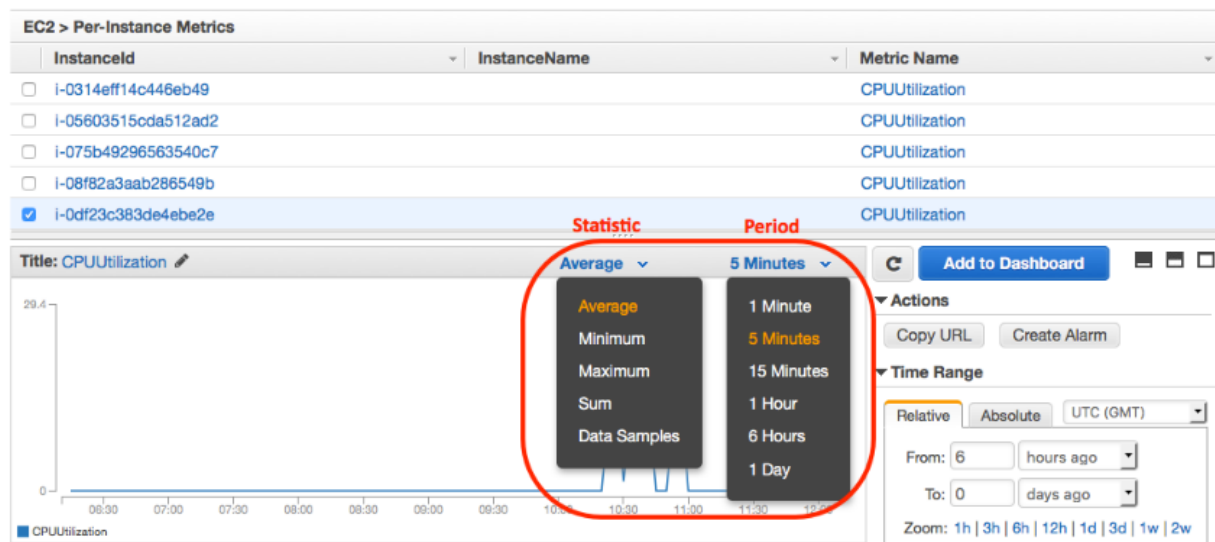


Figure 38 - Metrics available for CPU Utilization (AWS, 2015)

The CLI (Command Line Interface) allows for use of specific strings to provide information and reporting on the various metrics and alerting set. The CLI allows for more flexibility than the standard dashboard interface. An example of accessing information on CPUUtilization using the CLI is set out below.

```
$ aws cloudwatch get-metric-statistics --namespace AWS/EC2 --metric-name
CPUUtilization --period 3600 --statistics Maximum --dimensions
Name=InstanceId,Value=i-1234567890abcdef0 --start-time 2015-02-18T23:18:00 --
end-time 201502-19T23:18:00
```

Figure 39 - CloudWatch command line interface

#### 5.2.2.4 Cost

### Pricing Models

Amazon CloudWatch pricing is based on the region where the AWS instance is implemented and this can and does vary (AWS, 2015).

Table 4 - AWS CloudWatch Options and Pricing

Option Description	Pricing
Basic monitoring of Amazon EC2 instances (metric of five minute intervals) as well as metrics for EBS (Elastic Bean	Free

Option Description	Pricing
Stalk) volumes, Elastic load balancers and RDS DB instances. Organizations making use of this Tier, receive 10 metrics, 10 alarms and 1 million API requests each month, with no charge. Up to 5GB of data ingestion and storage is also at no charge.	
Detailed Monitoring for Amazon EC2 Instances	The costs for this particular option in Sydney are \$3.50 per instance per month, using 1-minute frequency intervals. (AWS, 2015)
Amazon CloudWatch Custom Metrics	Charged at \$0.50 per metric per month in the Sydney region. (AWS, 2015)
Amazon CloudWatch Alarms	\$0.10 per alarm per month in the Sydney region. (AWS, 2015)
Amazon CloudWatch API Requests	\$0.01 per 1,000 GetMetricStatistics, ListMetrics, or PutMetricData requests, in the Sydney region. (AWS, 2015)
Amazon CloudWatch Logs	\$0.67 per GB ingested** \$0.033 per GB archived per month*** Data Transfer OUT from Amazon CloudWatch Logs is priced equivalent to the “Data Transfer OUT from Amazon EC2 To” and “Data Transfer OUT from Amazon EC2 to Internet” tables on the EC2 Pricing Page. This is also the Sydney region. (AWS, 2015)

### 5.2.3 Sumo Logic

Sumo Logic is a purpose built cloud service, whose mission it is to give organizations the ability to:

“...harness the power of machine data to improve their operations and deliver outstanding customer experience at Enterprise scale” (Sumo Logic, 2015, p. 1).

#### 5.2.3.1 Ease of installation

Sumo Logic provides a number of options on where to install collectors. An important factor in deciding where to install the collectors is the network topology, as well as the bandwidth that is available on the network and the various domains and user groups that exist on the network (Sumo Logic, 2015).

There are also different types of collectors – hosted or installed. Hosted collectors do not need to be installed or activated, and they have no physical requirements as these are stored in Amazon Web Services (AWS) by Sumo Logic.

#### Installed Collectors

An installed collector can be installed on any server within the network that is used for aggregation of logs or other services on the network. This is ideal if the organization has a central network location for access to data (Sumo Logic, 2015).

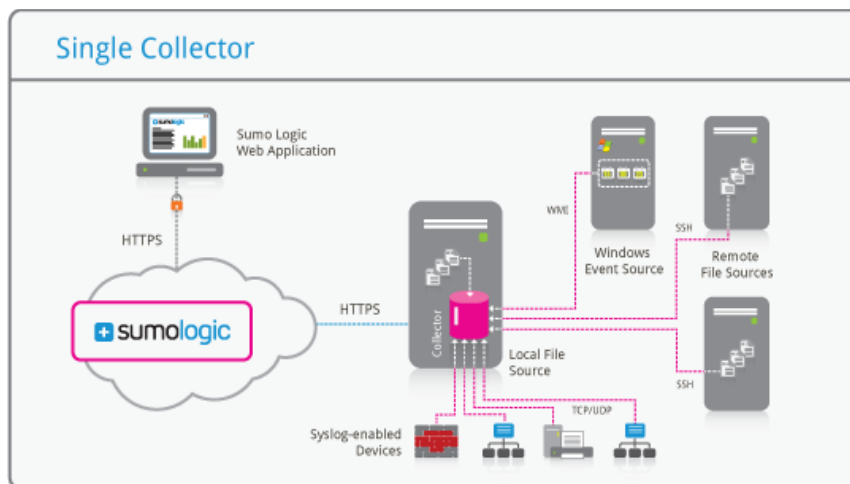


Figure 40 - Sumo Logic Single Collector (Sumo Logic, 2015)

If the organization has a distributed topology it may be best to install the Sumo Logic collectors on multiple machines throughout the organization allowing for a combination of sources to collect the data.

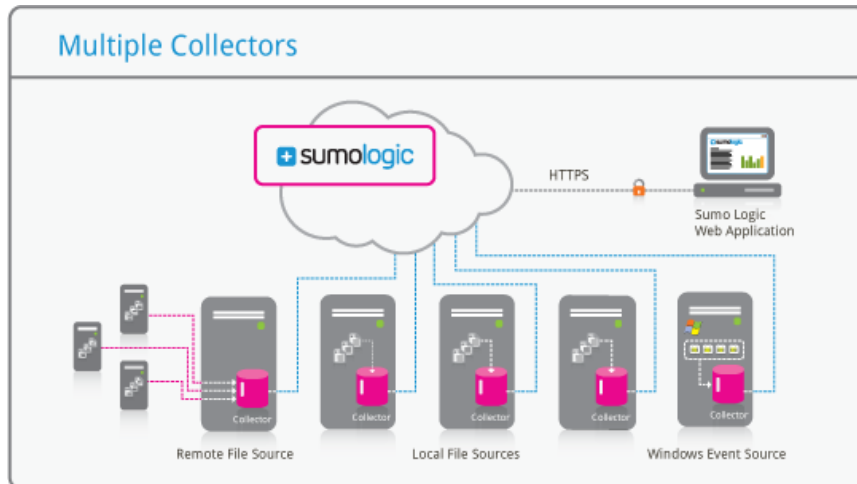


Figure 41 - Sumo Logic Multiple Collectors (Sumo Logic, 2015)

Installed collectors can also be deployed within cloud deployments or in a data center configuration. The collectors that are set up on each of the machines independently send unique log information to Sumo Logic so that queries can be run against any machine or server.

## Hosted Collectors

These collectors are not installed on the organization's local systems, they (and the collectors' sources) are instead hosted by Sumo Logic within AWS. A hosted collector allows for configuration of Amazon S3 sources which allow for data to be moved from the organization's S3 buckets into Sumo Logic directly, or HTTPs sources. One hosted collector can be configured with many S3 or Https sources (Sumo Logic, 2015).

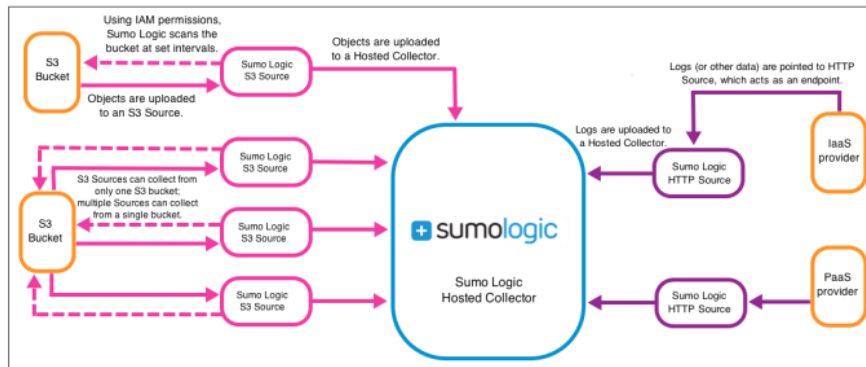


Figure 42 - Sumo Logic Hosted Collectors

### 5.2.3.1.1 Setup of Monitoring and Alerting

#### Alert and Notify

A monitoring system would not be one without alerts and notifications. Sumo Logic allows for alerts to be customized to allow for proactive notification of specific events and possible outliers that occur in the systems, see figure 43. A patent pending Push Analytics technology uses LogReduce to establish baselines. Notifications are set up to be generated when the data deviates from these baselines, or exceeds certain thresholds that have been set (Sumo Logic, 2015).

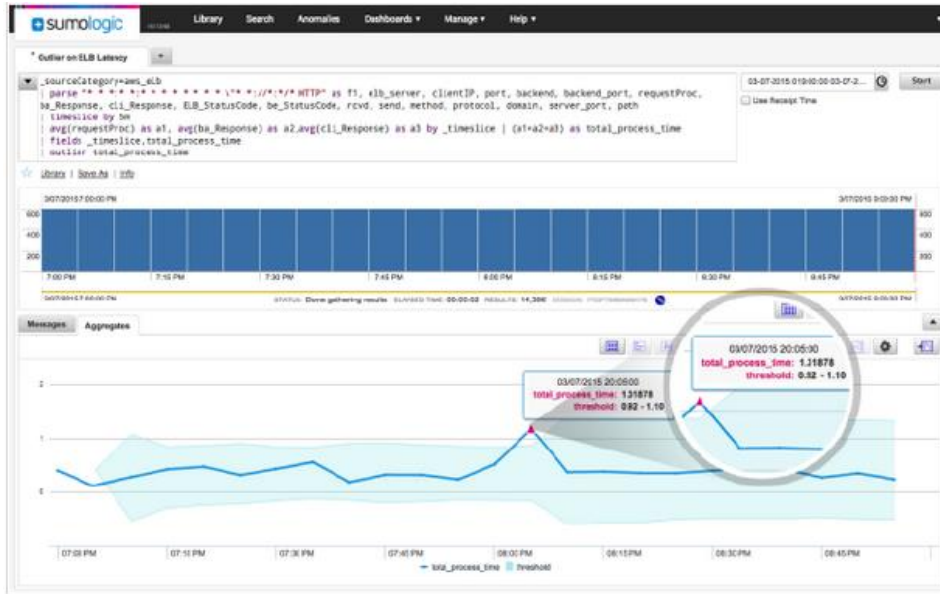


Figure 43 Sumo Logic Alert and Notify (Sumo Logic, 2015)

### 5.2.3.2 Log aggregation and reporting

Sumo Logic looks at analytics from two angles; namely how to address the quantity of data and secondly how to support the two styles of analytics (Sumo Logic, 2016). The two styles of analytics that Sumo Logic considers are firstly the known data that an organization has; that is analyzing questions that an organization knows that it needs to ask about its infrastructure, and secondly looking at the unknowns in order to gain insight into questions that the organization does not know how to ask or answer.

Sumo Logic meets Big Data requirements by implementing universal collection. Data is collected and analyzed from many different sources, without worrying about the volume, format or location of the data that this being collected. Sources can be located in many different places, including on premise and cloud. Data is collected by utilizing either the local Sumo Logic collectors or via hosted collection using https or directly from Amazon S3 buckets (Sumo Logic, 2016).

The data that is gathered is compressed at a ratio of 10x and encrypted for security before it is transmitted. Data is collected in its raw format and all processing of this data is undertaken in the cloud. Sumo Logic process data in this way so as to minimize performance issues associated with parsing and processing data locally (Sumo Logic, 2016).

Sumo Logic offers a centralized and single repository for machine data. This is important because the number of disparate systems and locations makes it difficult to consolidate and centralize data. If looking at a particular transaction it could potentially traverse multiple systems in order to complete, and by having the data from log files about this single transaction centrally available, it allows for quicker and easier troubleshooting and problem isolation (Sumo Logic, 2016).

As stated above Sumo Logic utilizes two types of Analytics. The first is the “Known Unknowns”, which is based on setting up queries for things that an organization knows that it needs to look for, such as error conditions. Sumo Logic utilizes a powerful query language which can be used to query, as well as “slice and dice data” (Sumo Logic, 2016) and at the same time allows the organization to ask the questions for which it needs answers. Queries can be set up to be parsed immediately with information being posted to a web-based UI (User Interface). Fields and data can also be included in statistical analysis which is enabled by Sumo Logic’s support of mathematical libraries, thus allowing for data to be further analyzed. The query options provided gives the organization’s IT teams access to look at large amounts of data quickly and efficiently, in order to find anomalies and patterns that could pin point sources of issues that are being experienced. Dashboards are formulated so that environments can be monitored in real-time and these also provide a good visualization platform.

The second type of analytics that Sumo Logic undertakes is the “Unknown Unknowns” (Sumo Logic, 2016) Sumo Logic have developed a patent pending LogReduce technology that uses machine learning to assist in proactively identifying patterns and provide insights into data, when users are not certain about what to look for.

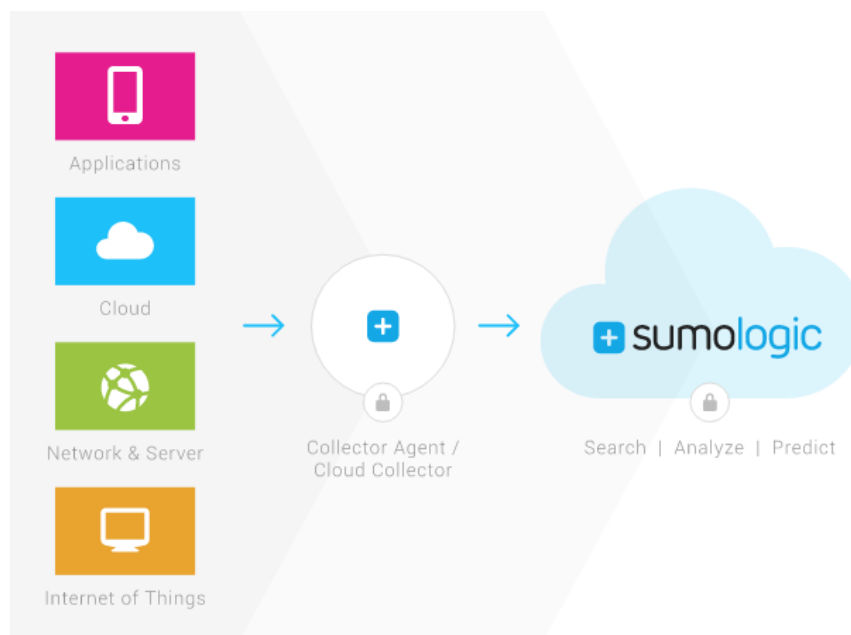
“LogReduce takes millions of lines of log results and distills them into a discernible set of underlying patterns, all without users ever writing a specific query. In other words, LogReduce reduces the prevalence of “unknown unknowns” and turns them into “known knowns”” (Sumo Logic, 2016, p. 1)

Another advantage of this is that LogReduce is capable of learning and improving over time, giving users the ability to be able to look at refining queries, and personalizing results. The advantage of being able to personalize with LogReduce, means that users are able to more easily uncover things that are considered important to the organization. The anomaly detection

extends LogReduce by showing changes in baselines as well as differences in system behaviors (Sumo Logic, 2016).

### **Collect and Centralize**

Sumo Logic is able to collect very large amounts of data from any type of application, device, cloud implementation, and many more (Sumo Logic, 2015). This is implemented by using lightweight collectors which are used to collect, compress, cache and encrypt data so that it can be securely transferred to a central location. This centralization eliminates the requirements for additional backup or archiving of logs. Sumo Logic allows for the data to be pre-parsed and it can be partitioned when being ingested. There are many collectors as well as API's which allow administrators to easily integrate or develop data sources with Sumo Logic services.



*Figure 44 - Sumo Logic Collect and Centralize*

### **Search and Analyze**

Sumo Logic features searches that can be run in real-time by administrators. It also allows for event correlation in real-time using engine-like syntax that is easy to use. This LogReduce technology (patent pending) groups the hundreds of thousand log events into groups of patterns. This gives easy filtering which then assists with the identification of issues, by



filtering out the noise. This noise often makes it very difficult to find and identify issues timeously (Sumo Logic, 2015).

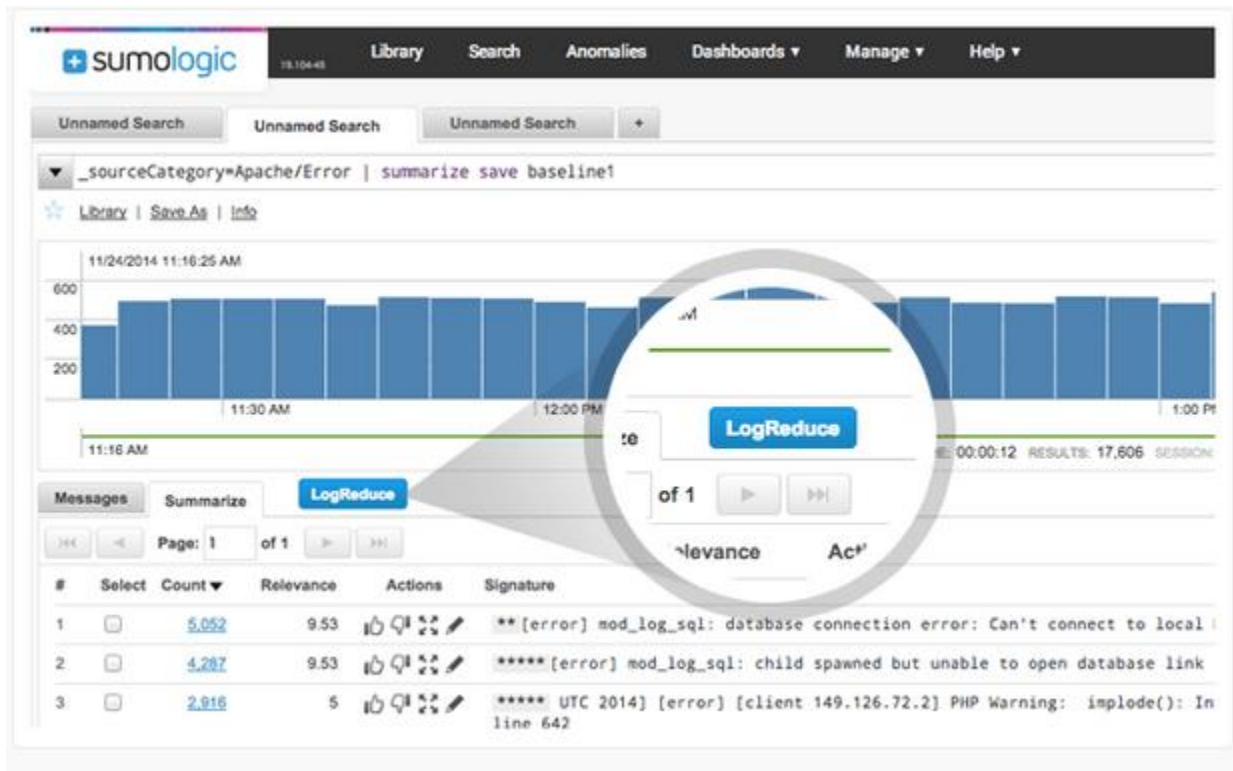


Figure 45 – Sumo Logic Search and Analyze (Sumo Logic, 2015)

## Detect and Predict

There are times when the rules that have been set up are not enough to give the full picture of what is happening on the systems, it is then that the Anomaly Detection technology kicks in.

This is driven by machine-learning algorithms which detect deviations in order to uncover the unknowns in the data. To give the administrators of the systems better visibility of Key Performance Indicators (KPIs) the purpose built visualization can highlight the abnormalities and give access to information for troubleshooting purposes and remedial action. This is customizable in that the users can annotate and add additional information to train the algorithm and thus help to reduce the time it takes to implement a fix.

The functionality is further extended by predictive analytics which complements the Anomaly and Outlier Detection by forecasting future KPI violations and abnormal behaviors by using a linear projection model. The fact that possible future errors can be detected or predicted allows for the addressing of problems before they actually impact the organization (Sumo Logic, 2015).

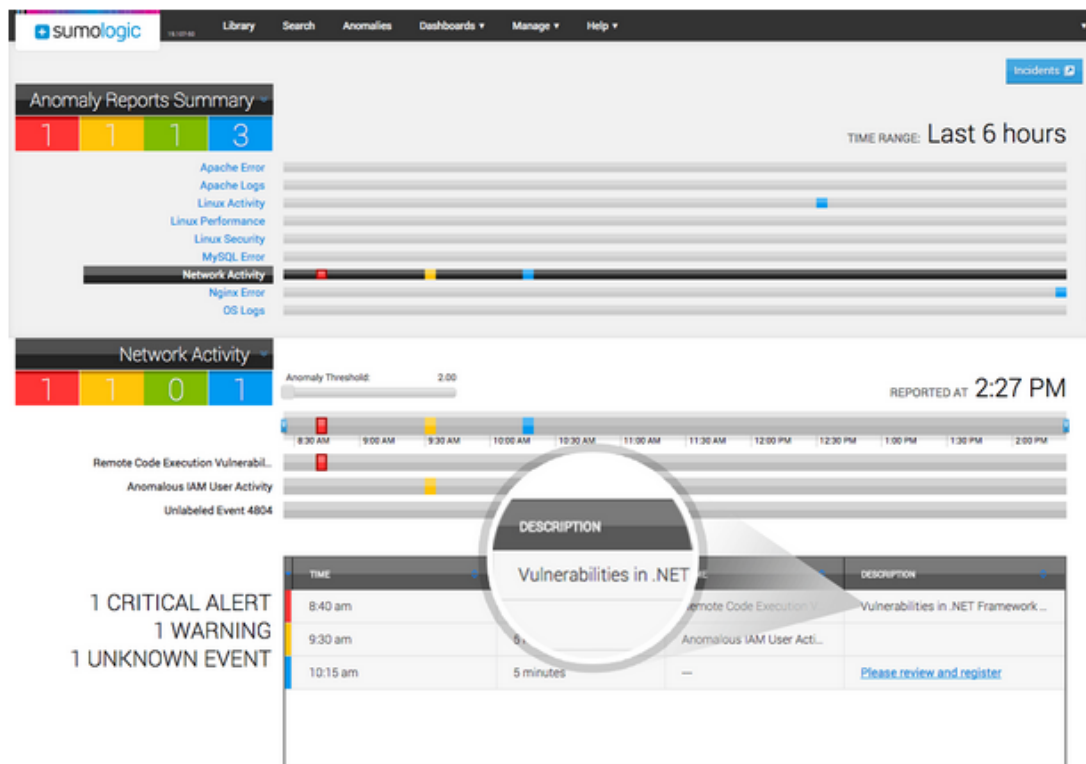


Figure 46 - Sumo Logic Detect and Predict (Sumo Logic, 2015)

## Monitor and Visualize

Sumo Logic allows for the customization of dashboards giving the user an easy way to monitor data real-time. There are also charting capabilities for various types of charts like pie, bar, etc. which can assist with reporting and visualization of the data (Sumo Logic, 2015).

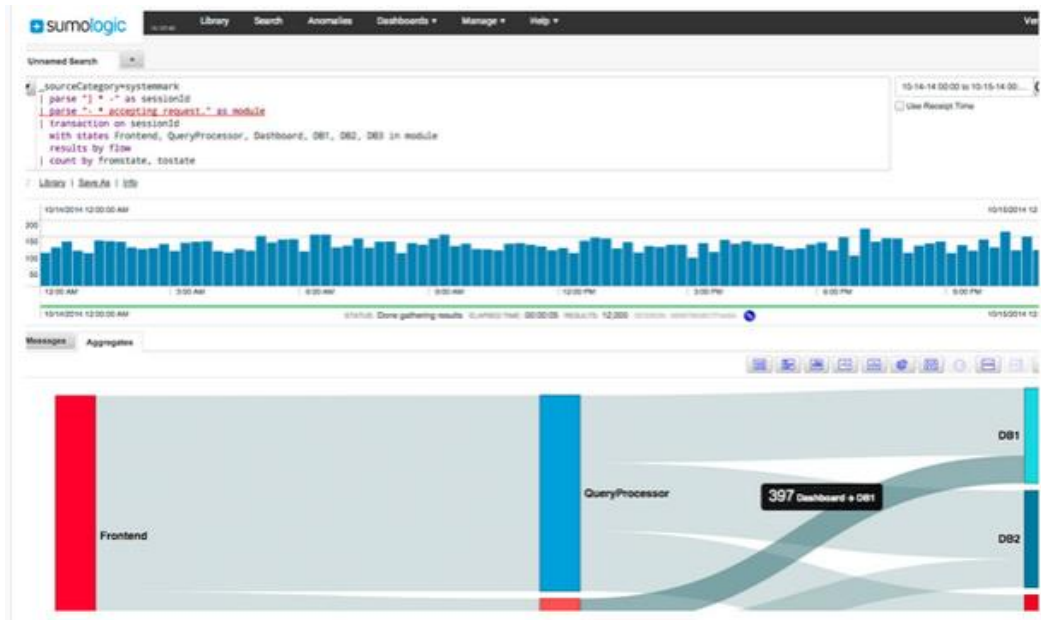


Figure 47 Sumo Logic Monitor and Visualize (Sumo Logic, 2015)

### 5.2.3.3 Costs

Sumo Logic offers various pricing models as shown in Table 5. Initially a trial can be run to evaluate the software, at no charge for data volumes of up to 500MB per day. Their next offering is called Professional and which offers up to 1GB of data per day at a cost of \$90 (USD) per month. Sumo Logic also offers Enterprise licensing which can be customized to meet the organization's needs (Sumo Logic, 2015).

*Table 5 - Sumo Logic Costs*

Product Tier	Number of Users	Features/Support	Support	Cost
Sumo Logic Free	1–3 users	<ul style="list-style-type: none"> <li>• LogReduce Analytics</li> <li>• Data Collection (any source)</li> <li>• Outlier Detection</li> <li>• Predictive Analytics</li> <li>• Live Streaming Dashboards</li> <li>• Powerful Search</li> <li>• PCI, SOC 2 Type 2, HIPAA Certifications</li> <li>• 7 Days Data Retention</li> </ul>	Community Forum Tech Support	Free
Professional	3–20 Users	<ul style="list-style-type: none"> <li>• All Sumo Logic Free Features</li> <li>• Real Time Alerts</li> <li>• Advanced Search Performance</li> <li>• Collector Management API</li> <li>• Data forwarding</li> <li>• Extended Retention to 30 days</li> </ul>	Business Hour Tech Support	1-3GB per Day = \$90 per month 5-10Gb per Day = \$450 per month 20GB per Day = \$1800 per month
Enterprise	20+ Users	<ul style="list-style-type: none"> <li>• All Professional Features</li> <li>• Anomaly Detection</li> <li>• Analytics API</li> <li>• Enterprise Integrations</li> <li>• PCI Compliance App</li> <li>• Single Sign-on Integration</li> <li>• Multi-Year Extended Retention</li> </ul>	Proof Concepts & RFP Support Technical Account Manager Professional Services Trainer Optional 24x7 Tech Support	On request

#### 5.2.4 Summary

The above overview of the three tools that were reviewed shows that there are differences in the way things can be set up, monitored and that there are many configurable options in each of the tools. The tools each come with standard templates that can be set up to be used for monitoring and alerting. These are base templates and monitor events like high CPU for instance, according to a standard threshold level that is considered to be an industry norm. If this is not the right level of monitoring, custom monitoring and alerting should be set up. This will require a more in

depth knowledge and understanding of the tool, which invariably means a resource will need to be trained and have dedicated time to setting up, monitoring and maintaining the tool.

It is important to come up with the right balance for the organization, in order to gain the most benefit from the use of any tool. For some organizations it may be more important to monitor high memory usage, which for a software development company, may show that there is a memory leak or issue with their software. Having this view of memory usage may assist the developers with isolating the issue, and being able to fix the problem. If no monitoring was in place it would be more difficult to understand when the problem occurred, and what was happening on the system at the time the issue occurred. Monitoring and alerting tools can provide valuable information to an organization, not only by reactionary means but also by storing data that can provide information to be used for proactive and strategic planning purposes.

Understanding the environment to be monitored will assist greatly with what is required from the tools that are available. The monitoring and/or alerting needs of an organization will generally be different for an organization that develops software to one that runs the medical machinery in a hospital. Different levels of criticality will be involved in each of these environments, and what is important for one may not be at all important for the other. Thus understanding what needs to be monitored, from having a base understanding of the environment in which the organization operates, will provide a basis from which to look at the requirements needed in the tools.

There are many options in the market place, from basic monitoring and alerting to more advanced and customizable tools. Doing a comparative analysis of what is available in the market place, and looking at the pros and cons and ease of use and installation, will help ensure that the right tool for the right environment is chosen.

The next Chapter, Chapter 6, is centralized on a discussion with regard to these particular tools. N-Able is discussed in point 6.2.1 and gives an overview of the implementation and functions that the tool provides. 6.2.2 gives an outline of AWS CloudWatch and what can be gained by implementing this AWS tool. Lastly point 6.2.3 reflects on Sumo Logic and its capabilities.

## Chapter 6 Observations and Conclusion

### 6.1 Introduction

All of the tools that were analyzed had good monitoring and alerting capabilities, allowing for more in-depth settings to be utilized as well as customizable alerting and reporting. Initial setup and implementation could be done on all the tools with relative ease, however if more sophisticated information was required, the set up become more complicated. Documentation and support for each was good and easy to find, either on the vendors own websites or on forums which discussed various implementation types and troubleshooting measures.

All of the tools implemented during this research were implemented within an organization which was Healthcare related. The environments utilized for the implementations ranged from live production environments to a standalone test environment. The test environment used was not a true reflection of the production environment.

### 6.2 Observations

Looking at an AWS (Amazon Web Services) implementation, CloudWatch provides all the necessary monitoring and alerting. Similarly, other tools in other environments specifically tailored to Cloud monitoring and alerting can be implemented, for example in Microsoft's Azure Platform. N-Able as a log monitoring and alerting tool can provide good visibility of an organizations implementations. Implementing Sumo Logic presents an organization with more detailed analysis of log data that is captured, in one tool. In turn this can provide insights into various aspects that the organization may be interested in, such as consumer behavior, analysis of uptime and related issues.

Whilst all the tools analyzed in this research could be used to gain further insight into the Big Data that is stored within log files, only one was an all-in-one tool, namely Sumo Logic. The other two tools, N-Able and CloudWatch, had the capability of storing this Big Data from log files in other locations, and you could add on Big Data Analytical tools to cater for these requirements. Even though this increases the complexity and the knowledge required to implement the tools, they still have the capability of providing more in-depth information.

Using a tool that can do Big Data Analytics to gain information on correlations and causes of issues, as well as provide insights into organizational growth areas, will give an organization more of an edge over their competitors. By using tools that are Big Data centric, such as Big Data Analytics and Data Mining tools, more information can be obtained from the data that is stored within log files. These log files contain data from applications, infrastructure and other sources like social media that can add value to an organization.

### **6.2.1 N-Able**

This product, installed within the environment, was set up to monitor and alert on various metrics on devices within the organizations network. Many of the devices that were monitored and alerted on were set up on-premises, and some were in AWS in various regions throughout the world. The agents were automatically installed on any device that was set up in the domain. However, it is important to note that the agent is initially set to essentials mode. Additional configuration is therefore required in order to ensure that the correct level of monitoring and alerting for that particular device was set.

During the analysis of this tool, a number of devices had been created in the domain, and the agents automatically installed. It was imperative to remember to go into the N-Able console in order to ensure that the correct level of monitoring and alerting was set up. This potentially could cause a false sense of security in that it is not evident that the level of monitoring and alerting is at a basic level when the agent is initially automatically installed. Knowledge of the product is definitely required. Setting the levels of monitoring and alerting to ensure that issues are caught before they become incidents is important.

In order to set up more complex monitoring and self-healing with this tool, a more in depth knowledge of the tool is required. It is not enough to just have a basic understanding. Setting up basic monitoring and alerting using the templates and putting various devices into categories, is pretty simple. There are many additional things that can be done, by writing scripts (using PowerShell for the Microsoft environment as an example), that can provide more in depth monitoring and alerting. The tool also has additional optional (with extra costs!) add-ins or plug-ins that allow for more specific monitoring of certain types of software/hardware. For instance, SQL Server or Oracle Database options will allow for better and more in depth analysis,

monitoring and alerting on these databases. An organization needs to take into account the additional costs of these options, and decide on the return on investment when looking at setting these up.

It is important to remember to use the maintenance schedules, when taking devices out of service for maintenance. This is especially important when reporting on SLA's, and also to ensure that false alerts are not received by IT engineers. Also, remembering to remove devices that are being decommissioned, from being monitored. This is not automatically done, and can cause disruption by alerts being received from a device that is no longer running.

The reporting functionality within N-Able is very broad and allows for a number of different types of out of the box reports to be run and set up. These can be automated or run on demand. It was noted that the format of some of the reports, for example a report showing all devices with the agent installed, was not really conducive to additional manipulation – even when exported to excel. The format did not lend itself to ease of sorting and grouping and as such made it quite difficult to view. However, the reporting functionality itself did lend itself to being able to produce reports that could be required for management purposes, including graphs which provide a good visualization of the environment.

In order to gain more in depth analysis from the data that has been gathered, the use of a data store in a separate database is required. This data can then be interrogated and manipulated to provide additional information that may be of value to the organization. Being able to keep the data for long periods of time is a standard out of the box feature, using a separate database. You still need to invest in Big Data Analytical tools in order to get information from the stored log data. Unfortunately due to time constraints this was not undertaken.

### **6.2.2 AWS CloudWatch**

AWS CloudWatch is proprietary to AWS and allows for setting up of monitoring and alerting of AWS implemented instances. Implementing AWS CloudWatch within an AWS environment is a matter of subscribing either to the free levels of monitoring or paying for something a little more in depth. What will be required will be dependent on the organizations needs and requirements with regard to monitoring and alerting of their infrastructure. The free monitoring



and alerting can provide some information about systems in AWS, and would be feasible for organizations that did had small instances and did not want to invest in more complex infrastructure.

In the test environment it was decided that it was not necessary to monitor all devices all of the time. Specific monitoring was set up during working hours for monitoring as it was not necessary to monitor these 24/7. The production environment was heavily used, mostly for web applications, and it was important to ensure that these instances were always running effectively. As such the production environment was monitored and alerted on 24/7 with thresholds to ensure that no issue impacted customers. Some of the other development and QA environments, like the test environment, were only monitored during set working hours. The tool has the flexibility to set thresholds and limits for different types of environments.

Setting up alerting according to well thought out thresholds will ensure that alerts are only received for valid criteria, thus giving the IT engineer enough time to rectify issues that have occurred before they become all encompassing. It is also important to set up maintenance schedules when doing maintenance on devices, so as not to be inundated with alerts that are not valid. The same goes for removing devices from monitoring when they are decommissioned.

CloudWatch provides good monitoring and alerting capabilities, as well as reporting on top of this to give an overview of the environment and its overall health and status. The reporting functionality can provide some valuable information that allows for reports to be generated automatically or on demand. This is especially helpful when having to report to management on the overall status and health of the systems.

Storing of log files is for up to 15 months free of charge. If more time is required, additional storage of these logs is required, costs will be incurred. Keeping the logs for a longer period will give more value over time, as the trends will become more apparent. However, in order to delve more deeply into the data that is held within the log files, additional tools would need to be utilized. AWS does have a range of additional Big Data and Big Data Analytical tools available at additional cost that can assist. These tools integrate into the environment well as they have all been coded and set up to work together and interface with each other.

### **6.2.3 Sumo Logic**

The implementation of Sumo Logic was on a trial license within the same Healthcare related environment. The trial allowed for many devices to be set up for monitoring. The only downfall was the trial was for a period of 1 month. This was a little limiting, but the time was utilized to set up monitoring and alerting in test environments.

The installation of the centralized console was easy enough. Installing the collectors on the devices to be monitored was a little more intricate than initially thought, but the documentation from the vendor was readily available and of a good standard. With regards to setting the monitoring and alerting up, the product had some relatively simple templates which could be utilized. More in depth monitoring and alerting, including remedial actions, was not undertaken due to the time limitations of the trial license. These are however well documented with easy to follow guidelines. An installed collector was utilized for this trial implementation. This was chosen so as to have the results housed in a centralized place, allowing for results of monitoring and alerting to be easily accessible for this period.

Sumo Logic allows for a number of options with regard to setting up alerting, from basic to more in depth. Alerting was set up on the same basis as the other tools utilized in the environment so as to have a relatively comparable base to do an overview of the tools.

The basis of the technology provided by Sumo Logic allows for a more in depth analysis of the data that is collected in log files. It allows for real-time monitoring and alerting to take place, by affording the user an interface from which to gather data. The technology used in correlating and gathering data is done via LogReduce technology which allows for the grouping and identification of issues. Filtering can also be undertaken to ensure that non-impacting issues do not cloud the actual issues that need to be investigated. Sumo Logic also caters for anomaly detection which can give a broader insight into the data that is being collected.

Reporting within Sumo Logic has been well designed. It allows for Key Performance Indicators to be set up and reported on, which in turn makes reporting to management a lot easier. The format of the reports and graphic views give the report reader a good overall understanding of the health of the environment. The user of the reporting function is able to add additional information into the system, so that it can be used in further predictive analysis of the environment.

### 6.3 Discussion

Big Data is a buzzword that is bandied about quite a bit in the business sector. Big Data Analytical tools are also not something quick and easy to understand and implement, and can also be very costly. In order for data to make sense and be used in a way that benefits an organization, some organizational knowledge and technical knowledge is required. That is, the organization has to have the knowhow and understanding of their business, and what the data means to them; this needs to be combined with technology and skills in the technology arena to be able to gain the most from the data. Finding the right tool that will work to an organization's advantage is vital. Tools that can be used for monitoring and alerting can give a good centralized overview of an organizations infrastructure and systems. This in turn can save a lot of time and resource when investigating issues and troubleshooting as well as assist with planning for future growth.

There are a variety of tools available in the market, some are proprietary that come with a system or hardware. Other tools are multi-platform and allow for the monitoring and alerting of all types of devices within the infrastructure environment. Yet more advanced tools or functions within tools can be used to monitor an application or system following the path from beginning to end, that is, from the first key stroke to the end of the function that the application is to perform. And others can provide more in depth information on the usage of the system and how the users' preferences impact their usage of the systems.

With this in mind, as part of undertaking the analysis of the tools in this thesis, the pre-configured templates that came with each was looked at to ensure that the right levels were being utilized. Mostly, the levels were pretty generic, and needed some adjustments depending on what was being monitored.

In this thesis the tools that were analyzed were based on monitoring of infrastructure, namely the devices within a network. This included servers, network appliances and other devices connected to the network. These tools were set up to monitor connectivity to the network, as well as to look at certain performance areas of the devices. Alerting thresholds were set at values to ensure that IT engineers were kept informed, and were able to investigate and fix issues before they became a major outage. For example, setting the disk utilization threshold on a server to warn at 80% of utilization, would generally give an IT engineer enough time to alleviate the

issue by removing unneeded data, or by adding additional space to the device. There is no point in monitoring a server for disk utilization when the threshold is set to alert at 99% or a 100%. It would be too late to avoid an outage. Having an understanding that setting up alerting on a disk for example that is very large, say 2Tb, and the threshold is set to 80% is pretty ineffective. Generally, if this is the threshold there is still loads of space available, and the IT engineer would be alerted for something that would not require him/her to take any action at this stage. Thus, setting of thresholds needs to be carefully thought out.

The monitoring and alerting tools analyzed during this research ranged from the proprietary to the multi-platform. Most proprietary tools are generally specific to a particular vendor and their hardware or software implementations. Other tools which may be from specific vendors, or open source, can monitor and alert across platforms using standard protocols, and often plugins that have been developed for vendor specific integrations.

There are also certain devices that do not need to have monitoring and alerts set up or need the thresholds adjusted to suit the environment. Knowledge of the environment being implemented, and subsequently monitored and alerted on, is important. The IT engineer implementing the tool needs to know what the organization considers to be important in its day to day operations. This should be set according to the SLA's that have been set with the organization.

Alerts should be set up on devices to allow for proactive monitoring, as well as for meeting the required SLA's that are set within an organization. This could be anything from standard uptime to certain performance requirements. It is not good practice to be getting alerts on things that are not important, as this can create a fire storm of alerts which in turn may mean that they are either ignored or something that is vital is missed.

Being able to utilize the information from the log files, requires some additional manipulation of the data contained therein. Tools have been written specifically to get information from log files, and these can be tweaked and adjusted to obtain the right information at the right time. These tools make use of Big Data contained in log files, and by implementing Big Data Analytical tools the data is shaped and formed into information that is more useful than the raw data.

The three tools that were reviewed shows that there are differences in the way things can be set up, monitored and that there are many configurable options in each of the tools. The tools each

come with standard templates that can be set up to be used for monitoring and alerting. These are base templates and monitor events like high CPU for instance, according to a standard threshold level that is considered to be an industry norm. If this is not the right level of monitoring, custom monitoring and alerting should be set up.

It is important to come up with the right balance for the organization, in order to gain the most benefit from the use of any tool. For some organizations it may be more important to monitor high memory usage, and others be more interested in storage. Monitoring and alerting tools can provide valuable information to an organization, not only by reactionary means but also by storing data that can provide information to be used for proactive and strategic planning purposes.

One of the major differences between N-Able, Cloudwatch and Sumo Logic, is that Sumo Logic has an inbuilt predictive analytics system that can be used. This is a major plus for the product in that it is an all in one tool that can be used for monitoring, alerting and full analysis. It is a Big Data tool with Big Data Analytics built in. This allows for greater flexibility in providing information on the environment to an organization, for capacity planning and other functions that can allow the organization an advantage.

The below table is a summary of each tool with measures against the criteria.

*Table 7 – Comparative Evaluation of Tools*

<b>Product</b>	<b>Ease of Installation</b>	<b>Ease of Implementation and Setup</b>	<b>Log Aggregation and Reporting</b>	<b>Value of use</b>	<b>Cost</b>	<b>Totals</b>
N-Able	3/5	3/5	3/5	4/5	3/5	16/25
CloudWatch	4/5	3/5	3/5	4/5	4/5	18/25
Sumo Logic	4/5	3/5	5/5	4/5	4/5	20/25

Each of the installations were relatively easy, although N-Able looks like it requires a little more work in order to get the central console setup.

All products were similar in their implementation and setup. Each had templates that could be used for basic monitoring, but all required additional product knowledge to set up more intricate monitoring and alerting.

Reporting was over all good for all products, with N-Able having a few things that could be improved upon with regard to layout. The only product that had log aggregation built in as part of the tool was Sumo Logic. The other two tools needed additional tools and configuration in order to achieve the same objective.

Each tool has value in its use. It just depends on what the organization requires. If stock standard monitoring and alerting is required, N-Able would be a good tool. It also can monitor on premise and cloud environments which gives it an edge over CloudWatch. CloudWatch monitors only within the AWS environment, and if the organization is predominantly in AWS then it would be a good tool to use. Sumo Logic is able to monitor, alert and aggregate logs in both an on premise and cloud environment, and therefore provides more flexibility and in depth information gathering. All the tools had good user interfaces, making it easy to navigate and use from the central consoles.

Costs were relatively low for each, although N-Able was slightly higher in costs. The costs however would increase for N-Able and CloudWatch specifically if logs were to be kept for longer periods. In order to interrogate this data additional tools would be required for these products and this would increase costs. Sumo Logic costs include these options as standard.

## **6.4 Conclusion**

This chapter looked at/discussed a review of the tools that were reviewed. The tools that were chosen have the ability to provide the information needed, in a format that is easily understood and accepted. Being able to use the system effectively means that users need to be able to define requirements easily. Additionally by being able to mine the data that is logged, an organization can find advantages in the market, as well as being proactive in monitoring of systems, and being aware of what is happening within and on the systems that they utilize.

## Chapter 7 Limitations and Future Research

### 7.1 Introduction

Big Data is not a new concept and has been around for many years, although not necessarily named as such. Log files have always held data about systems, processes and errors. Finding a way to make use of this data, in a concise and economical way is the way forward. By using Big Data Analytical tools to assimilate the data, information can be gained that can assist organizations to better strategize and avoid unnecessary issues and outages.

This chapter looks at the limitations experienced during the writing of this thesis, as well as the future research that could be undertaken in this field.

### 7.2 Limitations

As the implementation of N-Able and CloudWatch had been completed before this research began, there was the element of requiring some knowledge on how the tools were set up initially in order to be able to utilize them to the full advantage. It was a little disconcerting not having installed the product myself, as there was never surety that they had been installed with the correct parameters required in the first instance. By reviewing the implementation documentation on each, and then using the console of each product to check how the initial implementation was undertaken, gave clarity on the implementation. From this review it was obvious that the installations were done according to the requirements set out in the documentation provided by each vendor.

The implementation of Sumo Logic was then done within a test environment and not a live environment, which meant that not all data captured would be of the same nature as that within a production environment. Utilizing a live environment would have been more beneficial for gaining insights. This is mainly due to the fact that a live (production) environment will generally have larger workloads than a test environment. Even though this was a limiting factor, the implementation and use of the tool provided insight into what it could provide. Another limitation was that I was only able to obtain a trial license for a month. This could have been extended by the Vendor but the test environment was due to be shut down and as such this would not have been of much value.

All alerting and monitoring that was set up in the three tools for use during this thesis was similar in nature. This gave a good basis for comparison of the performance of the tools. The basic monitoring and alerting across the tools was set for CPU, Memory, Disk Utilization and Network Connectivity. The thresholds for CPU, Memory and Disk Utilization were set on the same basis, that is warning levels were set at 80%, and the critical level was set to 95%. Network Connectivity was set to report on up and down statuses. The limitations experienced were that there was not enough time to set up more in-depth monitoring, which meant that the customization of the tools reviewed was not tested fully. The differences in the production and test environments posed another possible discrepancy in that the production environment collected data 24/7 and the test environment only during working hours (8am to 5pm NZST).

Some tools that are used to monitor an environment can impact the performance of the devices being monitored. Therefore, another important factor to consider when implementing monitoring and alerting tools, is to ensure that they do not impact performance detrimentally on the devices that they are installed on. Having to keep the agent software up to date can also be a drawback (Bigelow, TechTarget, 2013). Luckily the tools under investigation during this research did not have a major impact on the operation of the devices within the environment. However they would still require to be updated when updates are released. This does mean additional administration is required.

For an organization to implement solutions that will drive their insight into what benefits they can realize will mean that they need to invest in tools that can provide this. The costs of the tools can vary and understanding what the goal of the organizations requirements are, will assist in choosing the right tool at the right cost. Managing data effectively can provide the insight required for investigation, debugging and gaining additional understandings into the business environment (Google, 2015).

### **7.3 Future Research**

This thesis has not covered all aspects of Big Data Analytics with regards to information contained in log files, and how best to utilize that data. Further studies and research into the various aspects of Big Data and Big Data Analytics would provide valuable information on this topic. The way data is collected and stored as well as processed means having to look at new



ways to set up platforms and tools. Planning the architecture to be able to implement and use data gathered is important, and it is not necessarily based on traditional infrastructure architecture (Kumar & Bhatnagar, 2015).

### **7.3.1 Skills for Big Data Analytics**

The future of Big Data Analytics also requires a different skillset. The skills required mean that requirements for more analytical thinking and training is required. Not only will this need to come specifically from people involved in data science but will need to extend into organizations. People within the organization will be required to develop analytical and data modelling skills in order to make the most of this field (Wamba, 2015).

### **7.3.2 E-Commerce and Big Data Analytics**

Big Data Analytics can assist organization within the E-Commerce sector. Some organizations already do make use of the Big Data that they are gathering from online shoppers, and are able to tailor suggestions for these shoppers. Information from log files on e-commerce websites can help an organization to drive sales up. Amazon is a good example of this, in that when a person is shopping for something, a list of other things that may interest the shopper is shown. This information is taken from the stored data transformed into information by Big Data Analytics (Buttolph, 2016). Amazon uses log files to assist in driving sales by predicting what customers want to purchase by looking at their history. Amazon analyze log data from their customers' accounts, and use this Big Data to help them drive additional sales and predict additional goods, and services that their customers will have an interest in. Amazon use a technique called "item-to-item collaborative filtering" which uses structured and unstructured data to tailor a customers' shopping experience. There is a great deal more that can be researched about this particular topic, which was not investigated within the realm of this thesis.

Nordstrom, another retail venture, using on-line and in-store techniques also makes use of Big Data to achieve their goals of being able to predict what products should be promoted to customers, using channels that are specific (SQream Technologies, 2013). One method that Nordstrom uses to predict customer tastes is to use log information from their websites and in-store sales, as well as utilizing social media networks. These logs are used in an innovation lab utilizing Big Data to assist in gaining insights and information from these logs, to assist in

providing predictable behaviors and purchasing trends of customers, in order to tailor their marketing of additional goods to these consumers.

Leveraging Big Data has the potential to unveil information that can benefit an organization and improve their business. Big Data reveals trends as well as patterns, which if utilized correctly by an organization, can lead to providing inputs into strategies especially marketing strategy (Stringfellow, 2015). By utilizing data from an organization's internal systems, and combining this with the data obtained from social media, a broad insight into the needs and wants of consumers can be gleaned. Such insights can be invaluable in marketing an organization's goods and services effectively.

### **7.3.3 Security and Privacy**

Security and privacy are other fields of research that have not been covered in this thesis. These topics are important to all organizations and additional research into these respective topics with regard to Big Data Analytics would be beneficial. There have been, and continue to be, advances within the field of security related to encryption and privacy of data. Big Data is often shared amongst organizations which leads to a number of concerns in relation to privacy and security. In order to deal with these concerns new technologies are being developed and new techniques are being employed (Li & Gao, 2016).

### **7.3.4 Traffic Monitoring**

In the past, conventional methods were used to monitor traffic flow patterns and monitor congestion; these included "floating car surveys" and "traffic count data" (Iteris, n.d.) Advancements in technology such as infrastructure sensors and probes that are installed in vehicles, has meant that collection of this data, and the use of Big Data Analytics, provides a new way of system monitoring. From these advances in technology, planning for future infrastructure upgrades, as well as providing public transportation, will be a good deal easier for city planners.

The Japanese application Zenryoku Annai! which helps drivers get to where they are going more efficiently and quickly utilizes in-memory computing and Big Data (Mullich, 2013). This application combines information from various sources, like satellite navigations systems, as well as traffic data gathered by statistical analysis from various sources. The data collected is

pipelined into the Zenryoku Annai! application and allows for better analysis of road conditions as well as allowing for better route planning by drivers. By using Big Data and NRI's research and traffic analysis techniques, subscribers to the application are able to plot their routes, avoid major traffic congestion and are given the ability to time their arrivals at destinations in a much more efficient manner.

Real-time data provided to motorists via their GPS and data connected Tom Tom navigation device is another example of utilizing Big Data to improve services provided (van der Lande, 2013). The real-time data allows drivers to avoid traffic jams and plan alternative routes. Mobile data networks provide connectivity to the device and provide up to date traffic information, which integrates with the Tom Tom navigation device. There is also the provision of High Definition Traffic information, using anonymous data from the speed and direction in which mobile devices are travelling in, available to motorists.

### **7.3.5 Organizations and Big Data**

Different business sectors can make use of data – each organization has some valuable data that it can transform into information. The following is an overview of the sectors and possibilities of what data and information can be provided and used.

#### *7.3.5.1 Insurance Sector*

Insurance companies are increasingly becoming the targets of fraudulent practices. This is in part because organized crime rings are becoming more knowledgeable about the various regulatory loopholes, as well as being able to utilize the information highway, which provides easier access to information than in the past (IBM, 2013). It is often a real burden on the insurance company to investigate, and monitor for fraudulent activities. Moreover, not only do organized criminal organizations defraud insurance companies; there are also individuals who undertake fraudulent activities. IBM has developed a claims fraud solution that uses a range of analytical technologies, and tools, which allow an organization to, anticipate, prevent, predict, identify, discover and investigate fraudulent activities. IBM leverage Big Data, and use analysis thereof to implement their fraud detection system. The software is able to analyze huge amounts of data, and process these to look for emerging patterns or even monitor changes in the claims lifecycle. Having a tool to assist with the issue of fraudulent activity, is a great help to the insurance industry, as it provides an effective and relatively cost effective model for fraud

prevention and investigation. Previously this would have been an extensive manual task, considering many resources, including valuable human resources.

#### *7.3.5.2 Banking Sector*

Banks invest heavily in data, analytics and technology to assist them in fraud detection and prevention methods (Smith & Gasan, 2015). The banking industry relies on internal and external IT resources to gather store and convert Big Data into business intelligence used to prevent and identify fraudulent activities. It is important that the right tools be used to analyze the data that is gathered, otherwise there is no particular value added. Looking for the right patterns and understanding how fraud is being committed is tantamount for the banking industry. Combined with the ever-changing way in which fraud is being committed, having a tool that is able to learn from previous patterns and has the capability of learning and finding new patterns is imperative.

Big Data is more frequently being utilized by the banking sector to look for assistance in risk management, fraud detection as well as looking at consumer behavior (Kiryakov, 2016). By using Big Data, analytics to compare data that comes from many different sources and to look for any inconsistencies allows for the flagging of things that need immediate action. Thus by utilizing predictive models banks are more easily able to detect fraudulent behavior. However, detection of fraud is not the only thing that Big Data analytics is good for, it can also assist the banking sector with regulatory compliance. By not complying with regulatory compliance may cost a bank a lot of money. In order to ensure compliance, Big Data can provide the information needed to circumvent any issues before they appear. Big Data also has the ability to look at consumer behavior with a view to increasing revenue and enhancing customers' experiences. Banks can leverage Big Data analytics to gain further insights into what customers are doing and how to improve the experience that they have with the banks. It is important to drive customer loyalty as well as gain new customers, and by knowing what customers want and need, marketing efforts can be funneled to achieving this objective.

#### *7.3.5.3 Healthcare Sector*

Healthcare organizations are moving more towards digitizing the information that they used to store on paper. The data that is now available in the form of digitized records holds a very wide range of information, which can utilized to support many health care functions including disease surveillance and population health management (Raghupathi & Raghupathi, 2014). Through the

right use of this data, there is huge potential in the medical arena. This can be through discovery of patterns and trends and gaining understanding of these. Overall, these breakthroughs could lead to improved patient care, saving of lives and lower medical costs. From this analysis, it would be probable that greater insight into patient history will help with diagnosis, and treatment.

Rocana, a Canadian organization, is looking at leveraging big data to improve organizations IT operational environments in monitor IT systems by providing alerts with more perspective to them (Hilson, 2016). They have developed an advanced analytics module, which gives better functionality to their existing monitoring capabilities, which includes anomaly detection. They have called this module “WARN (Weighted Analytic Risk Notifications)” and it has been developed as a second order analytic that will investigate the history of any previous anomalies that have occurred. It does this by looking at each object, whether it be a service, host or other object, and does computations on a score for that object which will give insight to IT teams so that they can understand the pervasiveness or criticality of the anomalies.

“By applying historical context in creating a risk score, Rocana gives IT administrators the ability to identify which entities in their IT environment are having the most unusual behavior, and drill down to investigate or take further action as needed” (Hilson, 2016, p. 1).

Medical and healthcare is another area where real-time monitoring has been shown to highly effective, even at times providing life-saving information. When patients are in an ICU or other wards, and are under Real-time data sensing, it is critical for information on patient deterioration to be received early so that remedial action can be taken (Mao, et al., 2012). In this particular paper, the authors describe how they use data mining as an integrated approach to find general and sudden deterioration warnings. Monitoring patients using RDS and then using data mining on the collected information could save many patients’ lives.

As stated, the Healthcare industry looks to benefit from Big Data (Vyas, 2016). Three technology trends have come to the fore in recent years; physiological sensor technology, smart applications on mobile phones and big data analytics. Healthcare professionals now have digitized healthcare data at the tips of their fingers. These records of patients, pharmaceutical companies as well as information captured by healthcare insurers and providers, is all aggregated

into a healthcare database. The amount of data collected in these databases is extremely large, and in order to process or make sense of it Big Data Analytics has come to the rescue. Big Data analytics can look at a number of things, from the reactive that is analyzing characteristics of patients to the proactive side of things, by looking at identifying signals and by aiding in disease identification and prevention. Sensors can enable this proactive monitoring by gathering information about the patient wearing the device.

“These sensors coupled with external processing units like computers or mobile devices for converting this data into classified events using pattern recognition and machine learning algorithms and in turn alerting the patient or caregiver will reduce adverse incidents” (Vyas, 2016, p. 1).

In another medical related study, Pfizer and IBM have entered into a partnership:

“...to develop a state of the art monitoring platform or IoT (internet of things), for the monitoring of Parkinson’s patients. This will give physicians remote access to patients, and they will also be able to monitor the effects of dosage routines” (Farooq, 2016, p. 1).

The objective is to be able to monitor Parkinson’s patients by using sensors, mobile devices and other equipment, which can provide real-time information to doctors and researchers. By gathering this information, an overview of a patient and their progress can be obtained as well as a view on their response to various medications. This can assist by providing input into development of new drugs as well as the clinical trial design around these. Previously patients with Parkinson’s were requested to keep a diary with this type of information, which could be very inaccurate, thus the move towards using an automated technology to support the gathering of this information.

Data within Healthcare in general is increasing at an overwhelming rate and organizations in these fields have generally not deployed any level of management of data with a view to the analysis thereof (Wills, 2014). In this regard, Big Data Analytics can offer a lot, especially when looking to establish trends and predictions, and can offer solutions that can improve the quality of care given, whilst maintaining or containing costs and simplifying operational tasks. In general, most Healthcare organizations have not invested heavily in information technology infrastructure and systems, especially as they often consider the data local. With government

intervention however, this needs to change. This is evident especially when looking at the United States, where various information acts are spurring trends towards data sharing and analysis on. In order for a healthcare organization to take the first steps towards analyzing data, the author of this article suggests that these organizations look at the small data aspects as a starting point. Mostly small data analytics will not require huge capital outlays for organizations, but still give them some relevant information that can help transform the way they function. In looking towards the smaller subsets of data, it can be mined in real-time, relatively quickly in order to provide insights into patient care and costs.

Hospitals make use of many different medical devices to monitor patients. In a particular instance a Children's Hospital in Toronto, Canada, use medical devices to monitor premature babies. These devices log at a very high rate of around a 1000 readings per second, and the staff monitoring these machines are only able to capture a reading either hourly or half-hourly (IBM, 2010). This means that there is a lot data missed that could potentially provide lifesaving information. In order to be able to access all this data, Dr McGregor assisted with the development of a system called Artemis, which was supported by IBM infrastructure and technology. This system, by utilizing IBM as a platform, allows the data to be streamed to where it can be streamed in order to find patterns that can produce information for medical professionals.

#### *7.3.5.4 Services Sector*

Depending on the organizations function, real-time monitoring may be of extreme importance. Contact Centre's are one such type of organization where real-time monitoring can provide invaluable information.

“Knowing exactly what is going on in your contact center, what transpires between your customers and agents during the call, and how well your agents adhere to compliance and quality guidelines is critical to long-term success” (CallMiner, 2013, p. 1)

Being able to monitor calls in real-time and get information with regards to issues, is an important factor in an organization that has real-time processes and especially customer interaction.

#### *7.3.5.5 Food Industry Sector*

Innovations in the food industry have led to using technology to focus on food safety. In the city of Cambridge, UK, an interactive digital checklist and automated cloud-based system was implemented to replace paper-based systems that are widely used by restaurants and food chains (Millman, 2016). The system is based on automated monitoring from Checkit (checkit, 2016), who have developed automated monitoring in part for the food industry. The Cambridge implementation provides staff with hand held devices that give detail of the tasks that they need to undertake. They are able to check off each task as it is completed. Management is able to view that the staff have completed the steps, by centralized management of the logs, thus ensuring compliance with business processes and procedures.

“The FDA Food Safety Modernization Act (FSMA), the most sweeping reform of our food safety laws in more than 70 years, was signed into law by President Obama on January 4, 2011. It aims to ensure the U.S. food supply is safe by shifting the focus from responding to contamination to preventing it” (IOT, 2016, p. 1).

With this act, manufacturers of food, beverages, as well as producers of produce and retailers are starting to employ more digital solutions in order to manage food compliance and safety needs and regulations (IOT, 2016). Bosch Connected Devices and Solutions are collaborating with Zebra Technologies Corporation

“...to offer a complete reference design for the development of food safety solutions” (IOT, 2016, p. 1)

The system has a number of features, including that it is easy to deploy and uses cloud-based temperature monitoring technologies, including wireless sensors, mobile computing, etc. There are real-time alerts and monitoring of systems, which allow pro-active measures to be taken to alleviate deviations from safety standards. The hope is that this implementation will assist the users of the system with being able to remain mobile and improve their workflow and therefore productivity.

#### *7.3.5.6 Marketing Sector*

The analysis of logs, real-time or log based is a very important part of monitoring. In the book “Complete Web Monitoring” this is pointed out in Chapter 10, where a company initiated a



marketing campaign to drive business to their site (Croll & Power, 2009). The campaign did not go as well as expected even though a lot of monitoring had been undertaken prior to the campaign to ensure that they had the right information to hand. It was however noted by a Web Operator that the problem was not the website itself but its payment portal. Customers were being dropped off or waiting too long to do a payment and therefore never completed the transaction. This part of the web site has not been incorporated in the testing and monitoring. However once this issue was noted and fixed, the company went on to triple its sales. This shows that monitoring not only needs to be set up but also, the right analysis needs to be done on the data that is collected in order to provide the value for the organization.

By leveraging Big Data, organizations that sell groceries can gain a better understanding of the way in which their customers behave, and what they purchase (Teradata, 2015). Retailers, including grocers are making use of other channels like websites, in order to market and sell their products. This can provide opportunities to leverage log data from activities that customers undertake on these websites, including understanding preferences and interests that customers have. With the log file data, these preferences and interests, as well as other interesting facts, can be obtained. This can then be leveraged to provide the customer with a personalized selection of goods, and complimentary goods, which have been selected according to their preferences.

#### *7.3.5.7 Learning Sector*

Learning behaviors of students can be tracked by investigating the data that is accumulated in the log files of the systems used by the students. This is particularly true of E-Learning systems. Performance of students can be looked at to see what characteristics of the environment are pertinent to the performance between groups of e-learners (Jeske, Blackhaus, & Stamov Robnagel, 2013). The navigational patterns found in the log files of the e-learning system can provide information and assist in deducing information regarding self-regulation. Having this information at hand can assist with the development of the e-learning system, as well as give guidance to assist students using these systems.

#### *7.3.5.8 Telecommunications Sector*

Telecommunications organizations collect vast amounts of data in the form of logs pertaining to network stability, usage, customer information, and so on (Shiomoto, 2013). Many kinds of data

are used to assist in management of traffic, data, network configurations and flows, which often result from failure alarms. Using this data to obtain information will help the organization to plan for various network implementations, prioritizations, predicting failures, customer preferences and usage. Information obtained from logs showing traffic characteristics will assist the organization with network planning and engineering activities. Various logs can also be analyzed to assist in predictions of failures, and thus will give the organization the “heads up” so that they can deal with these before they occur.

Telecom organizations have a competitive advantage with the data that they have available from the logs and statistics on their network platform (Neos, 2015). Previously telecom organizations were mainly there to connect one user to another, but things have evolved significantly with the advent of mobile devices. The volume of data traffic has increased significantly over the years, and this in turn means finding new and faster ways to move this traffic to and from the users. Telecommunication organizations need real-time logging systems to monitor this traffic and of course bill for the usage thereof. However, this data can also be utilized to provide other information to the Telco and this is where Big Data analytics comes in. By having a tool that can probe the massive amounts of raw data that is generated, can be exceptionally invaluable. Information that can be gathered from this raw data includes, network traffic analysis, real-time surveillance for fraud detection, tracking customers’ behaviors and at the same time gaining insight into their usage. All of this information can be used to plan for the growth of the business and identify new opportunities.

One the most important assets for a Communications Service Provider (CSP) is data. With the continued and increased use of smart phones and mobile devices, a CSP has access to enormous amounts of data such as profiles of customers, location information, network usage, etc. This data can be considered a “gold mine” for the CSP (Cloudera, 2013). More and more frequently, CSP’s are making use of Hadoop and Big Data Analytics to gain strategic information from this data. There are a number of cases in which this information gained from the data collected can be utilized. Understanding and improving customer experience by using targeted marketing and personalization of the service for the user is one way in which this information can be utilized. In order to optimize and grow the mobile network, using this data can provide valuable information to the CSP. They can gather statistics of heavily utilized network stations and from

that can plan for growth. They are better able to forecast usage from these statistics, which allows for planning for potential outages should equipment becomes overloaded.

An early adopter of Big Data and related tools like Hadoop, is the Telecommunications sector. Many global Telecom organizations have shown what type of value can be gained by using the insights provided by Big Data (Cloudera, 2016). The main area in which this additional value has been achieved is with regard to customer insights, network optimization, operational analytics and data monetization. Having data to hand about how to improve customer experience is invaluable. Many telecommunication providers are looking at the data from customer utilization on their network and using this to provide better service and introduce new and improved services that customers want. Utilizing the data collected about network usage will give the telecommunication organization the advantage of being to plan ahead for growth. In order to ensure that the organization is running effectively, Big Data can provide the necessary information to drive efficiency and economies of scale. The information gathered will provide insight into security and fraud issues that may arise, and give the advantage of pro-active responses.

#### *7.3.5.9 Hospitality Sector*

Organizations in the Hospitality Industry can make use of the vast amounts of Social Media and consumer driven and generated data that is available to help them understand and solve problems and point them in the direction of opportunities (Xiang, Schwartz, Gerdes, & Muzaffer, 2015). By utilizing Big Data Analytics approaches to evaluate this data, a lot of information can be gathered and used to further an organization in the Hospitality Industry. A growth area is that of business intelligence used to provide information on understanding customers, competitors, market characteristics, etc. Mining of this data is attracting a lot of attention for the potential value that it has in relation to public and community data. Marketing tools such as product recommenders can be established by looking at and mining the data that is generated by consumers and other data sources. This information if used correctly can be of absolute value to organizations within the Hospitality Industry.

#### *7.3.5.10 Tourism Sector*

In the tourism industry log files can be utilized to gain more information on the likes and dislikes of tourists. Due to the use of the World Wide Web by tourists for bookings, research into

destinations and the so on, the electronic traces in log files, can provide insightful information for the tourism industry (Fuchs, Hopken, & Lexhagen, 2014). Sources of information can be explicit in nature or implicit. An example of explicit information is the feedback received from a tourist in regards to a questionnaire for feedback or a survey to gain more understanding of the tourists' opinion. Implicit sources could be numerous, including web-navigation data, GPS data, etc. Data mining methods can be used to drill down and find interesting patterns and relationships in the data, which in turn can be used to propose new and improved tourism opportunities.

## References

- Akter, S., & Wamba, S. F. (2016, March 16). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets*, 26(2), 173-194. Retrieved August 13, 2018, from <https://link.springer.com/article/10.1007/s12525-016-0219-0>
- Amazon. (2015, July 26). *Working with metrics*. Retrieved August 16, 2015, from docs.aws.amazon.com: [http://docs.aws.amazon.com/AmazonCloudWatch/latest/DeveloperGuide/working\\_with\\_metrics.html](http://docs.aws.amazon.com/AmazonCloudWatch/latest/DeveloperGuide/working_with_metrics.html)
- Arthur, L. (2013, May 15). *What is Big Data*. Retrieved November 9, 2014, from Forbes.com: <http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/>
- AWS. (2015). *Amazon CloudWatch*. Retrieved June 21, 2015, from aws.amazon.com/cloudwatch: <http://aws.amazon.com/cloudwatch/>
- AWS. (2015). *Amazon CloudWatch Pricing*. Retrieved August 1, 2015, from aws.amazon.com: <http://aws.amazon.com/cloudwatch/pricing/>
- AWS. (2015). *AWS CloudTrail*. Retrieved June 26, 2016, from aws.amazon.com: <https://aws.amazon.com/cloudtrail/>
- AWS. (2015). *Get Statistics for a Specific Instance*. Retrieved June 26, 2016, from docs.aws.com: [http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/US\\_SingleMetricPerInstance.html](http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/US_SingleMetricPerInstance.html)
- AWS. (2015). *Getting Started with Amazon CloudWatch*. Retrieved August 2, 2015, from aws.amazon.com: <http://aws.amazon.com/cloudwatch/getting-started/>
- AWS. (n.d.). *Getting Started with CloudWatch*. Retrieved August 06, 2017, from AWS: [http://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/CWL\\_GettingStarted.html](http://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/CWL_GettingStarted.html)
- AWS. (n.d.). *What is Amazon CloudWatch Logs*. Retrieved August 06, 2017, from docs.aws.com: <http://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/WhatIsCloudWatchLogs.html>
- Bigelow, S. J. (2013, February). *TechTarget*. Retrieved August 13, 2018, from How to choose the right server performance monitoring tools: <http://searchitoperations.techtarget.com/tip/How-to-choose-the-right-server-performance-monitoring-tools>
- Bigelow, S. J. (2015). *Log analysis tools join data center management arsenal*. Retrieved March 5, 2016, from <http://searchdatacenter.techtarget.com/>:

<http://searchdatacenter.techtarget.com/feature/Log-analysis-tools-join-data-center-management-arsenal>

Buttolph, B. (2016, March 16). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets*, 26(2), 173-194. Retrieved August 13, 2018, from <https://link.springer.com/article/10.1007/s12525-016-0219-0>

CallMiner. (2013, December 5). *Why Real-Time Monitoring Is So Important in the Contact Center*. Retrieved april 02, 2016, from callminer.com: <http://callminer.com/real-time-monitoring/>

Campbell, S. (2014). What is Qualitative Research. *Clinical Laboratory Science*(27.1). Retrieved August 16, 2017, from <https://search.proquest.com/docview/1530677717?pq-origsite=gscholar>

checkit. (2016). *Automated Monitoring*. Retrieved April 10, 2016, from <http://www.checkit.net/about/>: <http://www.checkit.net/products/automated-monitoring/>

Cloudera. (2013). *Big Data Use Cases for Telcos*. Retrieved June 5, 2016, from cloudera.com: <https://www.cloudera.com/content/dam/cloudera/Resources/PDF/solution-briefs/Industry-Brief-Big-Data-Use-Cases-for-Telcos.pdf>

Cloudera. (2016, February 18). *Cloudera Gains Momentum Across Global Telecommunications Market*. Retrieved June 5, 2016, from gobenewswire.com: <https://globenewswire.com/news-release/2016/02/18/812096/0/en/Cloudera-Gains-Momentum-Across-Global-Telecommunications-Market.html>

Conley, T. (2016, February 11). *Proactive IT Monitoring Prevents System Slowdowns*. Retrieved May 14, 2016, from galileosuite.com: [galileosuite.com/industry-views/proactive-monitoring-prevents-system-slowdowns/](http://galileosuite.com/industry-views/proactive-monitoring-prevents-system-slowdowns/)

Croll, A., & Power, S. (2009). Could They Do It?: Real User Monitoring: Chapter 10 - Complete Web Monitoring. In *Complete Web Monitoring*. O'Reilly. Retrieved April 10, 2016, from <http://archive.oreilly.com/pub/a/web-development/excerpts/9780596155131/chapter-10.html>

Cser, A. (2015, June 2). *Market Overview: Cloud Workload*. Retrieved April 30, 2016, from cloudpassage.com: <https://pages.cloudpassage.com/rs/857-FXQ-213/images/forrester-market-overview-cloud-workload-security-management-solutions-automate-or-die.pdf>

Di Martino, B., Aversa, R., Cretella, G., Esposito, A., & Kolodzeij, J. (2014). Big data (lost) in the cloud. *InderScience online*, 1(1-2), 3-17. doi:10.1504/IJBDI.2014.063840

Dumbill, E. (2013). Making Sense of Big Data. *Big Data*, 1(1), 1-2. doi: 10.1089/big.2012.1503

- Eberhardt, C. (2014, March 24). *The Art of Logging*. Retrieved August 8, 2015, from codeproject.com: <http://www.codeproject.com/Articles/42354/The-Art-of-Logging>
- Farooq, R. (2016, April 7). *Pfizer Inc and IBM Combine Pharma and Tech For Innovation*. Retrieved April 9, 2016, from businessfinancenews.com: <http://www.businessfinancenews.com/28500-pfizer-inc-and-ibm-combine-pharma-and-tech-for-innovation/>
- Fronza, I., Sillitti, A., Succi, G., Terho, M., & Vlasenko, J. (2013, January). Failure prediction based on log files using Random Indexing and Support Vector Machines. *Journal of Systems and Software*, 86(1), pp. 2-11. Retrieved February 7, 2016, from <http://www.sciencedirect.com/science/article/pii/S0164121212001732>
- Fronza, I., Sillitti, A., Succi, G., Terho, M., & Vlasenko, J. (2013, January 1). Failure prediction based on log files using Random Indexing and Support Vector Machines. *Journal of Systems and Software*, 86(1), 2-11. Retrieved February 27, 2016, from <http://www.sciencedirect.com/science/article/pii/S0164121212001732>
- Fuchs, M., Hopken, W., & Lexhagen, M. (2014, December). Big data analytics for knowledge generation in tourism destinations – A case from Sweden. *Journal of Destination Marketing & Management*, 3(4), pp. 198-209. Retrieved February 7, 2016, from <http://www.sciencedirect.com/science/article/pii/S2212571X14000353>
- Gandomi, A., & Haider, M. (2015, April). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. doi:10.1016/j.ijinfomgt.2014.10.007
- George, G., Hass, M., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, 57(2), 321–326. Retrieved March 26, 2016, from [http://www.amhj.aom.org/uploadedFiles/Publications/AMJ/Apr\\_2014\\_FTE.pdf](http://www.amhj.aom.org/uploadedFiles/Publications/AMJ/Apr_2014_FTE.pdf)
- Gill, A. Q., & Hevary, S. (2016). Cloud Monitoring Data Challenges: A Systematic Review. *Neural Information Processing*, 9947, pp. 72-79. Retrieved August 12, 2018, from [https://link.springer.com/chapter/10.1007/978-3-319-46687-3\\_8](https://link.springer.com/chapter/10.1007/978-3-319-46687-3_8)
- Google. (2015, March 19). *Leverage the power of log data to drive operational and business decisions through Google Cloud Logging*. Retrieved August 27, 2016, from cloudplatform.googleblog.com: <https://cloudplatform.googleblog.com/2015/03/leverage-the-power-of-log-data-to-drive-operational-and-business-decisions-through-Google-Cloud-Logging.html>
- Guddemi, L. (2014, October 22). *Why 85% Of Companies Fail At Creating A Configuration Management Database (CMDB)*. Retrieved November 14, 2015, from forbes.com: <http://www.forbes.com/sites/sungardas/2014/10/22/why-85-of-companies-fail-at-creating-a-cmdb-and-how-to-escape-their-fate/>

- Halcyon. (2016, February 17). *Is Your IT Systems Management Reactive or Proactive?* Retrieved May 8, 2016, from helpsystems.com:  
<http://www.helpsystems.com/halcyon/resources/articles/your-it-systems-management>
- Heath, G. (2011). *The benefits of IT monitoring and the early warning of system issues*. Retrieved August 27, 2016, from ncc.co.uk:  
<http://www.ncc.co.uk/article/?articleid=16890>
- Hernantes, J., Gallardo, G., & Serrano, N. (2015, July-Aug). IT Infrastructure-Monitoring Tools. *IEEE*, 32(4), 88-93. doi:10.1109/MS.2015.96
- Hernantes, J., Gallardo, G., & Serrano, N. (2015, July). IT Infrastructure-Monitoring Tools. *IEEE*, 32(4), 88-93. Retrieved August 12, 2018, from  
<http://ieeexplore.ieee.org/abstract/document/7140697/>
- Hesse-Biber, S. (2015, December 04). Qualitative or Mixed Methods Research Inquiry Approaches: Some Loose Guidelines for Publishing in Sex Roles. *Sex Roles*, 74(1-2), 6-9. Retrieved August 16, 2017, from <https://link.springer.com/article/10.1007/s11199-015-0568-8>
- Hilson, G. (2016, January 15). *Rocana uses big data to proactively troubleshoot IT problems*. Retrieved May 21, 2016, from itworldcanada.com:  
<http://www.itworldcanada.com/article/rocana-uses-big-data-to-proactively-troubleshoot-it-problems/380006>
- IBM. (2010). *Three Must-Read Big Data Case Studies*. Retrieved February 13, 2016, from ingrammicroadvisor.com:  
[https://www.ibm.com/smarterplanet/global/files/ca\\_\\_en\\_us\\_\\_healthcare\\_\\_smarter\\_healthcare\\_data\\_baby.pdf](https://www.ibm.com/smarterplanet/global/files/ca__en_us__healthcare__smarter_healthcare_data_baby.pdf)
- IBM. (2013, June). *Leverage Big Data to fight claim fraud*. Retrieved June 5, 2016, from IBM.com: [http://www-935.ibm.com/services/multimedia/Exploiter\\_le\\_Big\\_Data\\_pour\\_lutter\\_contre\\_la\\_fraude\\_aux\\_sinistres\\_Juin\\_2013.pdf](http://www-935.ibm.com/services/multimedia/Exploiter_le_Big_Data_pour_lutter_contre_la_fraude_aux_sinistres_Juin_2013.pdf)
- IOT. (2016, April 6). *Zebra Technologies and Bosch Collaborate on Food Safety Reference Solution*. Retrieved April 10, 2016, from iot.do: <http://iot.do/zebra-bosch-food-safety-2016-04>
- Iteris. (n.d.). *Leveraging Big Data for Congestion Management Programs*. Retrieved May 21, 2016, from iteris.com: <http://www.iteris.com/products/services/big-data-performance-monitoring>
- Jeske, D., Blackhaus, J., & Stamov Robnagel, C. (2013, June). Self-regulation during e-learning: using behavioural evidence from navigation log files. *Journal of Computer Assisted*



- Learning*, 30(3), 272-284. Retrieved February 7, 2016, from <http://onlinelibrary.wiley.com/doi/10.1111/jcal.12045/full>
- Kalakota, R. (2015, May 25). *Big Data Analytics Use Cases*. Retrieved February 13, 2016, from <http://practicalanalytics.co/>: <http://practicalanalytics.co/2015/05/25/big-data-analytics-use-cases/>
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014, July). Trends in Big Data Analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561-2573. doi:doi:10.1016/j.jpdc.2014.01.003
- Kepes, B. (2015, August 5). *Alert Logic launches cloud-native vulnerability and configuration management tool*. Retrieved April 30, 2016, from networkworld.com: <http://www.networkworld.com/article/2952856/cloud-security/alert-logic-launches-cloud-native-vulnerability-and-configuration-management-tool.html>
- Killi, A. (2015, December 25). *4 Good Open Source Log Monitoring and Management Tools for Linux*. Retrieved April 24, 2016, from tecmint.com: <http://www.tecmint.com/best-linux-log-monitoring-and-management-tools/>
- Kim, W., Jeong, O.-R., & Kim, C. (2014). A Holistic View of Big Data. *IGI Global*, 10(3), 56-69. doi:10.4018/ijdw.2014070104
- Kiryakov, A. (2016, April 7). *Practical Big Data Analytics For Financials*. Retrieved June 5, 2016, from ontotext.com: <http://ontotext.com/practical-big-data-analytics-financials/>
- Kumar, N., & Bhatnagar, V. (2015, December). Big Data Analytics. *4th International Conference, BDA* (pp. 1 - 274). Hyderabad: Springer. Retrieved August 13, 2018, from <https://link.springer.com/content/pdf/10.1007/978-3-319-27057-9.pdf>
- Kusnetzky, D. (2012, January 11). *Are log files the beginning of Big Data?* Retrieved December 12, 2015, from zdnet.com: <http://www.zdnet.com/article/are-log-files-the-beginning-of-big-data/>
- Li, S., & Gao, J. (2016, March 4). Security and Privacy for Big Data. *Big Data Concepts, Theories and applications*, 281-313. Retrieved August 13, 2018, from [https://link.springer.com/chapter/10.1007/978-3-319-27763-9\\_8](https://link.springer.com/chapter/10.1007/978-3-319-27763-9_8)
- Logentries. (2014, July 9). *Logentries and AWS Partner to Centralize CloudTrail, CloudWatch and Log Data Monitoring*. Retrieved June 26, 2016, from logentries.com: <https://logentries.com/logentries-and-aws-log-data-analysis/>
- Lurie, A. (2014, May 19). *Top 47 Log Management Tools*. Retrieved April 16, 2016, from blog.profitbricks.com: <https://blog.profitbricks.com/top-47-log-management-tools/>

- Mao, Y., Chen, W., Chen, Y., Lu, C., Kollef, M., & Bailey, T. C. (2012, June). *An Integrated Data Mining Approach to Real-time Clinical Monitoring and Deterioration Warning*. Retrieved April 9, 2016, from nd.edu:  
<http://www3.nd.edu/~dwang5/courses/spring15/papers/medical/p3.pdf>
- Millman, R. (2016, April 8). *Cambridge City Council unveils cloud-based food safety monitoring system*. Retrieved April 10, 2016, from publictechnology.net:  
<https://www.publictechnology.net/articles/news/cambridge-city-council-unveils-cloud-based-food-safety-monitoring-system>
- Moga, J. A. (2014). *Management of Information Systems*. Retrieved August 27, 2016, from referenceforbusiness: <http://www.referenceforbusiness.com/management/Log-Mar/Management-Information-Systems.html>
- Mowlem, S. (2015, July 14). *Small IT, big problems: Log data reveals the unknown*. Retrieved February 27, 2015, from infoworld.com:  
<http://www.infoworld.com/article/2945593/systems-management/small-it-big-problems-log-data-reveals-the-unknown.html>
- Mullich, J. (2013, September 09). *Drivers Avoid Traffic Jams with Big Data and Analytics*. Retrieved June 5, 2016, from sap.com:  
[http://www.sap.com/bin/sapcom/en\\_ca/downloadasset.2013-09-sep-23-16.drivers-avoid-traffic-jams-with-big-data-and-analytics-pdf.html](http://www.sap.com/bin/sapcom/en_ca/downloadasset.2013-09-sep-23-16.drivers-avoid-traffic-jams-with-big-data-and-analytics-pdf.html)
- n-able. (2015). *N-Central is software for better IT Service*. Retrieved August 2, 2015, from n-able.com: <http://www.n-able.com/products/n-central>
- N-Able. (n.d.). *Deployment, Operations and Best Practices*. Retrieved August 9, 2015, from applied-imagination.com: [http://www.applied-imagination.com/aidocs/n-able/v6.7/N-central\\_Essentials\\_Guide\\_65-SP2.pdf](http://www.applied-imagination.com/aidocs/n-able/v6.7/N-central_Essentials_Guide_65-SP2.pdf)
- Nemati, H. (2016, May 5). *Monitoring and Analyzing Virtual Machines— Resource Overcommitment Detection and Virtual Machine Classification*. Retrieved August 12, 2018, from Polytechnique Montréal:  
<https://hsdm.dorsal.polymtl.ca/system/files/5May2016.pdf>
- Nemschoff, M. (2014, February 26). *7 Important Types of Big Data*. Retrieved November 9, 2014, from SmartDataCollective:  
<http://smartdatacollective.com/michelenemschoff/187751/7-important-types-big-data>
- Nemschoff, M. (2014, June 28). *A Quick Guide to Structured and Unstructured Data*. Retrieved November 21, 2015, from smartdatacollective.com:  
<http://www.smartdatacollective.com/michelenemschoff/206391/quick-guide-structured-and-unstructured-data>

- Neos. (2015, November 20). *How big data is transforming telecom industry?* Retrieved June 5, 2016, from neos.hr: <http://www.neos.hr/how-big-data-is-transforming-telecom-industry/>
- Pandian, K., & Chinnathurai. (2016, December 08). Sustaining quality of services through service reliability and availability. *Accelerated Stress Testing & Reliability Conference (ASTR), 2016 IEEE*. doi:10.1109/ASTR.2016.7762294
- Peng , W., Li, T., & Ma, S. (2005). Mining Logs Files for Data-Driven System Management. *SIGKDD Explorations*, 7(1), 44-51. Retrieved December 5, 2015, from [http://kdd.org/exploration\\_files/7-Peng.pdf](http://kdd.org/exploration_files/7-Peng.pdf)
- Power, D. J. (2014, January 21). Using 'Big Data' for analytics and decision support. *Journal of Decision Systems*, 23(2), 222-228. doi:10.1080/12460125.2014.888848
- Prescott, A. (2015, January 8). *The Importance of being Licensed*. Retrieved November 14, 2014, from oxford-knowledge.com: <http://www.oxford-knowledge.com/importance-licensed/#.VkaCL7crJ9A>
- Raghupathi, W., & Raghupathi, V. (2014, February 7). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*. Retrieved May 21, 2016, from <http://hissjournal.biomedcentral.com/articles/10.1186/2047-2501-2-3>
- Rashid, F. Y. (2016, April 13). *WhatsUp Gold v16*. Retrieved April 30, 2016, from au.pcmag: <http://au.pcmag.com/whatsup-gold-v16/1122/review/whatsup-gold-v16>
- Ronk, J. (2014, September 1). *STRUCTURED, SEMI STRUCTURED AND UNSTRUCTURED DATA*. Retrieved November 21, 2015, from jeremyronk.wordpress.com: <https://jeremyronk.wordpress.com/2014/09/01/structured-semi-structured-and-unstructured-data/>
- Rouse, M. (2014, July). *Big Data*. Retrieved October 25, 2014, from TechTarget.com: <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>
- Sharp, M. (2015, March 5). *Proactive Monitoring*. Retrieved May 21, 2016, from imperators.com: <http://www.imperators.com/proactive-monitoring/>
- Sheppard, D. (2015, October 27). *Do you need a CMDB for cloud-based systems?* Retrieved November 14, 2015, from itworldcanada.com: <http://www.itworldcanada.com/blog/do-you-need-a-cmdb-for-cloud-based-systems/377870>
- Shiomoto, K. (2013, November). *Feature Articles: Applications of Big Data Analytics Technologies for Traffic and Network Management Data*. Retrieved February 13, 2016, from ntt-review.jp: [https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201311fa1.pdf&mode=show\\_pdf](https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201311fa1.pdf&mode=show_pdf)

- Shoor, I. (2015, April 11). *Log Management Tools Face-Off: Splunk vs. Logstash vs. Sumo Logic*. Retrieved April 24, 2016, from javacodegeeks.com:  
<https://www.javacodegeeks.com/2015/04/log-management-tools-face-off-splunk-vs-logstash-vs-sumo-logic.html>
- Smith, A., & Gasan, A. (2015, March 10). *Big Data for Fraud Mitigation*. Retrieved June 5, 2016, from equifax.com: [http://www.equifax.com/assets/IFS/15-3903\\_topten.pdf](http://www.equifax.com/assets/IFS/15-3903_topten.pdf)
- Solar Winds. (2015). *N-Central is software for better IT service*. Retrieved June 21, 2015, from n-able.com/products/n-central: <http://www.n-able.com/products/n-central>
- Solar Winds N-Able. (2013). *Solarwinds N-Able*. Retrieved June 21, 2015, from LinkedIn: <https://www.linkedin.com/company/n-able-technologies>
- Splunk. (2016). *Logging Overview*. Retrieved February 27, 2016, from dev.splunk.com: <http://dev.splunk.com/view/logging/SP-CAAADP5>
- SQream Technologies. (2013, December 10). *Retailers Using Big Data: The Secret Behind Amazon and Nordstrom's Success*. Retrieved February 7, 2016, from sqream.com: <http://sqream.com/how-retailers-are-using-big-data-to-improve-sales-and-customer-service/>
- Stringfellow, A. (2015, March 10). *24 MARKETING EXPERTS REVEAL THE MOST EFFECTIVE WAYS TO LEVERAGE BIG DATA IN YOUR MARKETING STRATEGY*. Retrieved May 21, 2016, from ngdata.com: <http://www.ngdata.com/how-to-leverage-big-data-in-your-marketing-strategy/>
- Subramanian, S. R. (2015, October 7). *Getting More from Analytics by Emphasizing 'Quality' over 'Quantity' in Data Logging*. Retrieved January 23, 2016, from dbta.com: <http://www.dbta.com/Editorial/Trends-and-Applications/Getting-More-from-Analytics-by-Emphasizing-Quality-over-Quantity-in-Data-Logging-106246.aspx>
- Sumo Logic. (2015). *About Sumo Logic*. Retrieved June 21, 2015, from sumologic.com: <https://www.sumologic.com/company/who-we-are/>
- Sumo Logic. (2015). *About Sumo Logic*. Retrieved August 2, 2015, from sumologic.com: <https://www.sumologic.com/company/who-we-are/>
- Sumo Logic. (2015). *How-it-works*. Retrieved June 21, 2015, from sumologic.com: <https://www.sumologic.com/how-it-works/>
- Sumo Logic. (2015). *Pricing*. Retrieved July 19, 2015, from sumologic.pricing.com: <https://www.sumologic.com/pricing/>

- Sumo Logic. (2015). *Setting up a Hosted Collector*. Retrieved July 18, 2015, from service.sumologic.com: [https://service.sumologic.com/help/Setting\\_up\\_a\\_Hosted\\_Collector.htm](https://service.sumologic.com/help/Setting_up_a_Hosted_Collector.htm)
- Sumo Logic. (2015). *Where should I set up Installed Collectors?* Retrieved July 18, 2015, from service.sumologic.com: [https://service.sumologic.com/help/Deciding\\_Where\\_to\\_Install.htm](https://service.sumologic.com/help/Deciding_Where_to_Install.htm)
- Sumo Logic. (2016, February). *Machine Data Analytics with Sumo Logic*. Retrieved June 26, 2016, from sumologic.com: <https://www.sumologic.com/resource/white-paper/machine-data-analytics-with-sumo-logic/>
- Talia, D. (2013). Scalable Big Data Analytics. *IEEE Computer Society*, (pp. 98-101). Retrieved July 09, 2016, from <http://xa.yimg.com/kq/groups/16253916/1476905727/name/06515548.pdf>
- Teradata. (2015, May 28). *More Grocers Need to Leverage Big Data Analytics*. Retrieved February 27, 2016, from blogs.teradata.com: <http://blogs.teradata.com/industry-experts/grocers-need-leverage-big-data-analytics/>
- Treynor, B., Dhalin, M., Rau, V., & Beyer, B. (2017, May 17). The Calculus of Service Availability. *Acm Queue*, 15(2), 49-67. Retrieved August 12, 2018, from <http://queue.acm.org/detail.cfm?id=3096459>
- van der Lande, J. (2013, May). *Using Big Data to build value for Operators*. Retrieved June 5, 2016, from asiainfo.com: [https://www.asiainfo.com/Portals/0/New\\_Branded\\_Collateral/White\\_Papers/AsiaInfo\\_AnalysysMason.pdf](https://www.asiainfo.com/Portals/0/New_Branded_Collateral/White_Papers/AsiaInfo_AnalysysMason.pdf)
- van Rijmenam, M. (n.d.). *A Short History Of Big Data*. Retrieved June 7, 2015, from Datafloq.com: <https://datafloq.com/read/big-data-history/239>
- Vyas, A. (2016, March 1). *Big Data in Healthcare – Proactive Monitoring and Clinical Intervention*. Retrieved May 29, 2016, from blog.persistent.com: <https://blog.persistent.com/index.php/2016/03/01/big-data-in-healthcare-proactive-monitoring-and-clinical-intervention/>
- Wamba, S. F. (2015). Big data analytics and business process innovation. *Business Process Management Journal*, 23(3), 470-476. Retrieved August 13, 2018, from <http://www.emeraldinsight.com/doi/full/10.1108/BPMJ-02-2017-0046>
- Weiss, T. (2014, April 23). *The 7 Log Management Tools Java Developers Should Know*. Retrieved April 24, 2016, from blog.takipi.com: <http://blog.takipi.com/the-7-log-management-tools-you-need-to-know/>

- Wills, M. J. (2014). Decisions Through Data: Analytics in Healthcare. *Journal of Healthcare Management*, 59(4), 254-262. Retrieved July 09, 2016, from <http://search.proquest.com/docview/1550823687?pq-origsite=gscholar>
- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Muzaffer, U. (2015, January). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44. doi:doi:10.1016/j.ijhm.2014.10.013
- Yilmaz, K. (2013, June). Comparison of Quantitative and Qualitative Research Traditions: epistemological, theoretical, and methodological differences. *European Journal of Education*, 48(2), 311-325. Retrieved August 16, 2017, from <http://onlinelibrary.wiley.com/doi/10.1111/ejed.12014/full>
- Zenz, S. (2015, June 13). *Category - N-Central Features*. Retrieved November 7, 2015, from [blog.n-able.com: http://blog.n-able.com/n-central-features/](http://blog.n-able.com/n-central-features/)
- Zulkernine, F., Martin, P., Zou, Y., Bauer, M., Gwadry-Sridhar, F., & Aboulmaga, A. (2013, July 2). *Towards Cloud-Based Analytics-as-a-Service (CLaaS) for Big Data Analytics in the Cloud*. Retrieved July 03, 2016, from IEEE: <http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=6597120>

## Glossary

Abbreviations	Meaning
API	Application programming interface
AWS	Amazon Web Services
CLaaS	Cloud based analytics as a Service
CLI	Command line interface
CMDB	Configuration management database
CPU	Central processing unit
CSP	Communications Service Provider
CSV	Comma separated values
Devops	Developer and operations role blended
EC2	Elastic compute
FDA	Food and Drug Administration
HyperV	Microsoft Hypervisor
IaaS	Infrastructure as a Service
IAM	AWS identity and access management
ISO	Software image file
IoT	Internet of things
IT	Information technology
JSON	JavaScript Object Notation
LAN	Local area network
MSP	Managed Service Provider
NSA	National Security Agency
PaaS	Platform as a Service
PCI DSS	Payment Card Industry Data Security Standard
Perfmon	Microsoft performance monitor
PS	Process status
RSM	Remote support manager
S3	Amazon storage service
SaaS	Software as a service
SIEM	Security information and event management
SLA	Service level agreement
SQL	Structured query language
Taskmgr	Microsoft task manager
WAN	Wide area network