Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 7, July 2015, pg.224 – 229

RESEARCH ARTICLE

Data Analysis and Reporting using Different Log Management Tools

Kavita Agrawal[1], Reader Hemant Makwana[2]
[1]Dept. of CS, Institute of Engineering & Tech., D.A.V.V., Indore, India
[2]Dept. of IT, Institute of Engineering & Tech., D.A.V.V., Indore, India
kavitaagrawal2904@gmail.com [1], hemanturl@yahoo.com [2]

Abstract— Data analytics is the science of examining raw data with the purpose of drawing conclusions about that information. Data analytics refers to the process of assembling, consolidating and analyzing large sets of data to detect patterns and other effective information. Data analytics not only help to perceive the information contained within the data, but it also helps in identifying the data that is decisive for the business [10]. Data analytics is used in many industries to allow companies and organization to make better business decisions and in the sciences to verify or disprove existing models or theories [16]. In this paper the various log management tools are studied and compared, thus the conclusion for the improvement is proposed.

Keywords— Data analysis, log management, operational intelligence, predictive analytics.

1. Introduction

Data analysts are concerned with analyzing the data and then gaining knowledge from that data. Analysis of large volume of data, i.e. big data is typically performed using specialized software known as log management tools and applications for data mining, predictive analytics, forecasting, text mining and data optimization. Collectively these processes maybe separate but highly unified functions of high-performance analytics. Using various log management tools and software enables an organization to process large volumes of data that a business has gathered to determine which data is relevant and can be analyzed for better business decisions in the future. [9][14]

The main functions of a log management tool are

- Collect logs from all log sources, like Windows events, syslog, flat file, databases or applications
- Organize logs in integrated, scalable, and secure manner
- Enable fast, flexible search into all logs
- Systematize log archiving and retrieval for long term retention
- Search and recover archived logs in second
- Identify anomalies in applications, databases, systems and devices in real time
- Automate log classification, normalization, aggregation & correlation

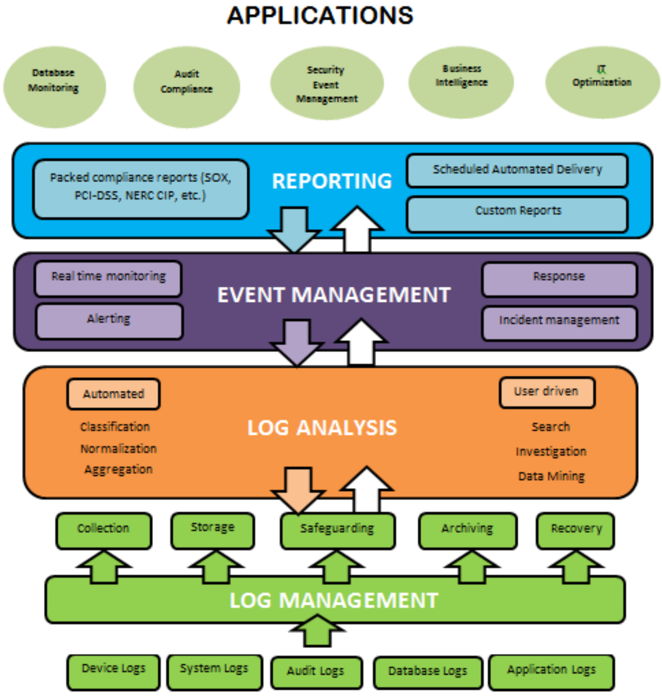

Figure 1: Flowchart of working of log management tools

Comparison of few log management tools is done on the basis of features like data input type, major application area, SDK available for languages, dashboard creation functionality, cost, real time supportability, web interface, etc. These parameters are selected as it is some of the important feature of log management tools.

2. Different Log Management Tools

Few log management tools are listed in random order, not in order of performance or capabilities.

2.1 Splunk

Splunk, an industry-leading platform for machine data, automatically indexes complete log data, including unstructured, structured and tangled multi-line log data of application. Splunk provides a deeper understanding of real-time data, with the ability to search complete data without custom connectors or scalability limitations. One can quickly search, report and diagnose operations and security issues in a fast and repeatable process. Splunk has built-in advanced reporting capabilities, with features like charts and dashboards creation, and a pivot interface to generate visual reports with drag-and-drop ease.[2][12]

Key Features:

- Resolve issues up to 70% faster
- Decrease cost escalations by up to 90%
- Provides complete picture view across the organization
- Monitor significant performance indicators and make keener decisions
- Identify trends and patterns for transactions, systems and customers
- Identify, name and tag fields and data points to add context
- Install content, add-ons and apps from Splunk app website to expand functionality
- Setup regular searches as real-time alerts
- Trigger automatic responses

2.2 Papertrail

Some of the features which make Papertrail comprehensive and streamlined solution for log management are its time-saving tools and facility of creating flexible system groups, long-term log data retention, advanced analytics with visualizations and various exporting options. In just few seconds, Papertrail enables businesses to effectively monitor their servers.[1][2]

Key Features:

- Derive value from the data which is already collected
- Works well with all common log methods and formats

- App logs, text log files, syslog can all be aggregated in one place
- Aggregates data from multiple groups of log sources
- Real-time functionality from command line, browser, or API
- Instant alerts
- Convention alerts on any event configuration
- Explore and filter with long-term data retention
- Save beneficial and frequently-used searches
- View correlated events together in dashboard
- Unlimited systems and users

2.3 Loggly

Loggly is a cloud-based log management service, which makes the log management process much less complicated. With a simple set-up process and intuitive tools, Loggly doesn't need a ton of on-ramping. Loggly provides instant value by interpreting and making sense of data pouring in from the applications, platforms and systems immediately.[2]

Key Features:
- Infinite custom dashboards based on any search
- In-built customizable alerts with triggers
- Adaptable interface with several views, pages and workspaces
- RESTful API to assimilate with other applications
- Unlimited saved searches and users
- Automated filters and event parsing
- Custom source groups
- Point-and-click trending graphs
- Text-based logs from any source

2.4 Graylog2

Graylog2 is an open-source data analytics system that's been field-tested around the globe. It collects and aggregates events from a group of sources and presents data in a streamlines, simplified interface where one can drill down to significant metrics, identify key relationships, generate powerful data visualizations and derive actionable insights.[2][4]

Key Features:
- Leverages Java, Scala and ElasticSearch technologies
- Central syslog monitoring
- Interactive API browser
- Application debugging
- Exception monitoring
- API analytics
- Intuitive search interface
- Comprehensive dashboard

2.5 LOGalyze

LOGalyze offers solutions for system admins, network managers and security teams. Network monitoring, integrated log management, real-time data collection and analysis are also some features provided by it. LOGalyze provides a range of functionality used across several disciplines to maintain tighter security and derive more accurate insights from data that's distributed in siloes across the infrastructure.[2][15]

Key Features:
- Assemble event logs from distributed Windows hosts or syslog
- Analyze log data
- Multi-dimensional statistics
- Pre-defined compliance reports
- Custom report generation
- Plug-in-style alert components
- Classifies logs by source host, severity and type
- Filters logs into groups for easy analysis
- Associate data to define events and alerts

2.6 McAfee Enterprise Log Manager

McAfee Enterprise Log Manager automates log management and analysis for all log types, signing and validating to ensure authenticity and integrity, a requirement for regulatory compliance. Compliance rule sets and reports are functional out-of-the-box, simplifying setup and configuration. All of this functionality serves to strengthen the company's security profile and improves ability to comply with more than 240 standards. McAfee Enterprise Log Manager is an integrated component of McAfee Enterprise Security Manager. Log Manager stores logs, while Enterprise Security Manager handles parsing, normalization and analysis.[2][17]

Key Features:

- Intelligent log management
- Stores the correct logs to ensure compliance
- Parses and analyzes logs for security
- McAfee supports chain of custody and non-repudiation efforts
- Adapt storage and retention by log source
- Analyze and search logs
- Differentiate logs stored for security and logs to be parses/analyzed
- Store locally or via managed SAN
- Up to 7.5 TB of usable HDD storage on the appliances
- One-click access to log files or specific log files during any point

3. Discussion

There are log management tools spanning practically over every use scenario and configuration, so this list is intended to be a central resource for comparing various appliances against some common specifications.

Table 1: Comparison among different log management tools used for data analysis [13]

Feature	Splunk	Papertrail	Loggly	Graylog2	LOGalyze	McAfee Enterprise Log Manager
Cost	FREE : Initial 60-day Splunk Enterprise 6.1 license – 500 MB/day. Enterprise license can be purchased depending on the use	FREE: 100 MB/month. Enterprise license can be purchased depending on the use	FREE : 200 MB/day. Enterprise license can be purchased depending on the use	FREE	FREE (unlimited, for everyone including commercial use)	Need to contact company for quotes
Web Interface	Simple	Simple	Simple	Simple	Simple	Simple
Support Real Time	Yes	Yes	Yes	Yes	Yes	Yes
Data Input Type	Windows event logs, web server logs, live application logs, network feeds, system metrics, change monitoring, message queues, achieve files	Operating System syslog, application log files, text data	Application & system data, any text based data	Syslog messages and files	Syslog, text file	Windows Event logs, Database logs, Application logs, Syslog

Dashboard Creation	Yes	Yes	Yes	Yes	No	No
Major Application Area	Application Management, security, compliance, operational intelligence	Log analysis	Log analysis	Log analysis	Log collection, log analysis, compliance report	Log collection, log management, log analysis
SDK available for	Java, JavaScript, Python, PHP, Ruby, C#	Java, Ruby	PHP	Java, Ruby	Java	Java
Cons	Free version is quite limited & Enterprise version pricing is based on amount of data indexed.[8]	Mostly text based.[7] No built-in graphing	It's cloud based so issues regarding the security of data is of concern[5]	Does not have the ability to read directly from syslog files. Graylog relies upon elastic search & mongodb[3].It does not match the feature list of splunk[6]	Presently not commercially preferred over Splunk. Newer versions are being made available to overcome bugs.[15]	Free Trial version is very limited & Enterprise version pricing needs to be quoted from company.[17]
Area for improvement/ updation	Performance can be improved by improving search queries for dashboard creation for bulk of regular expression	Can be enhanced in areas like security, compliance, operational intelligence	Can be enhanced in areas like security, compliance, operational intelligence	Can be enhanced in areas like security, compliance, operational intelligence	SDK can be made available for more number of languages, so that its usage can be increased. Dashboard creation feature can be added.	SDK can be made available for more number of languages, so that its usage can be increased. Dashboard creation feature can be added.

4. Conclusion

Data analytics refers to the process of assembling, organizing and analyzing large amount of data to detect patterns and further useful information [11][14]. In this survey, few log management tools are been considered. There are many more tools which can be preferred according to the requirement. According to the above discussion we came to the conclusion that Splunk is a very effective tool. Splunk Enterprise is the generator for machine data. Splunk software facilitates enterprises to gain operational intelligence by tracking, reporting and analyzing real-time machine data as well as terabytes of historical data located in cloud or on-premise. Automated searches can continuously track for irregular patterns of behavior in host, network and application data [9]. Every enterprise and organization is different, and while some tools are more flexible than others, each organization may have distinct requirements and preferences.

References

[1] Splunk, Big Data and the Future of Security http://www.decisionreport.com.br/download/WhitePaper_Splunk.pdf

[2] Top 47 Log Management Tools by Andy Lurie http://blog.profitbricks.com/top-47-log-management-tools,May 19,2014

[3] Big Data: Open Source Tools vs Splunk by Jonah Cook http://netloid.com/technology/big-data-open-source-tools-vs-splunk , May 9, 2014

[4]Graylog2 organization https://www.graylog2.org/resources/documentation/general/plugins

[5] Send Events Safely to the Loggly Cloud by Xavier Mertens http://blog.rootshell.be/tag/loggly,Dec 27,2010

[6] Graylog2 Optimization for High-Log Environments by Rich McDonough
http://server.dzone.com/articles/graylog2-optimization-high-log, July 24,2012
[7] The 7 Log Management Tools Java Developers Should Know by Tal Weiss
http://blog.takipi.com/the-7-log-management-tools-you-need-to-know, April 23,2014
[8] Splunk feels the heat from stronger, cheaper open source rivals by Serdar Yegulalp
http://www.javaworld.com/article/2101306/open-source-tools/splunk-feels-the-heat-from-stronger-cheaper-open-source-rivals, Feb 25,2014
[9]Big Data Analytics by Vangie Beal http://www.webopedia.com/TERM/B/big_data_analytics.html
[10] What is Big Data Analytics? http://www.webopedia.com/TERM/B/big_data_analytics.html
[11]Big data Analytics-Ultramax | SAS Big data analytics training
http://ultramaxit.com/big-data-analyics/
[12] Splunk_big_data_futureofsecurity http://www.slideshare.net/sbelyaeva/2-21677-splunkbigdatafutureofsecurity
[13] Intelys
http://www.intelys.net/services/managed
[14] YB Analytics :: Services
http://ybanalytics.com/big-data-analytics.php
[15] http://www.logalyze.com/ (general internet site)
[16] Data Analytics by Margaret Rouse http://searchdatamanagement.techtarget.com/definition/data-analytics
[17] (general internet site) http://www.mcafee.com/in/products/enterprise-log-manager