

A Basic Introduction to DevOps Tools

Akshaya H L, Nisarga Jagadish S, Vidya J, Veena K

*Department of computer science and engineering,
VTU, Belgaum, University,*

Abstract— DevOps is a practice in which operations and developers work together in each stage of DevOps life cycle (from development stage to production stage). DevOps is good for developers. Developers need DevOps because for a better quality of life, pride of ownership and for more relevant work. DevOps can be divided into four categories like log monitoring, monitoring, build and test and deployment and configuration. In this paper, authors give the brief introduction to DevOps tools.

Keywords— Log monitoring; System and network monitoring; Build and test; Deployment and configuration.

I. INTRODUCTION

DevOps is a new trend which has emerged from collision between two old trends namely 'agile system administrations' or 'agile operations' and other is the understanding the value of collaboration between development and operation staff in each and every stages of DevOps life cycle. 'Dev' means all the developers involved in production stage. 'Ops' means the term used for system engineers, system administrators, security professionals and various other disciplines or job title. So, totally 'DevOps' does not differentiate between any of the system administrator sub-disciplines. DevOps is a practice in which operations and developers participate together in each stage of DevOps life cycle like from development stage to production stage.

A. LOG MONITORING:

Logs are used to analyze the system performance and usage trends. So log data are used by developers in debugging process. When any application comes across different stages of DevOps life cycle (like testing, business analytics, production monitoring), log acts as like a valuable tool [15]. When log data applied to first stage of DevOps life cycle, log acts like debugging tool and also serves as system load and performance testing [15]. When log data applied to second stage of DevOps life cycle, logs are used for production monitoring and production trouble is shooting [15]. When log data applied to third stage of DevOps life cycle, logs are used for web analytics and business metrics [15]. When we analyze the logs it improves the QA process by catching the issues faster, identifying the issue before they grow and also having better communication between the groups [16]. Logs are considered to be critical while measuring the success of the service and also trouble shooting the issues when aroused [17]. Real scalability and security are the features that don't receive attention while managing the logs. There number of DevOps log tools for managing the log data. In these some of them are open source and some of them are paid tools.

B. MONITORING:

Monitoring tools will be used by the developers to make sure that the deploying software is performing correctly [1]. DevOps monitoring tools can be classified as system and network monitoring tools. System monitoring tools will be used for monitoring the performance of the system, collecting and storing the data, producing graphs of the collected data and it keeps track of system resources [2]. Network monitoring tools can be used to monitor the network and its resources, checks for the status of the network devices and if in case any issue arises it will notify the administrator, it can also collect and analyze the network traffic [3].

C. BUILD AND TEST:

The application of these tools is in automating common developer tasks like compilation of source code into binary code, creation of executable, running test cases and in the creation of documents. Tools falling under this category are Ant, Lvy, Gradle, Jenkins, cruise control and Maven. Hosted services like Travis CI offer additional options. The correct tool chain for DevOps will automate IT services, provide real-time visibility into system and application performance and also gives us a single source of truth. More important than an individual tool's capabilities, still, is how closely it all matches the organization's strategic goals. That's the method to maximize the chances of attaining DevOps goodness [37].

D. DEPLOYMENT AND CONFIGURATION:

System administration main task is of helping people to use computers. System administrators are domain experts who provide impedance matching between users' desires and computers. The expertise of system administrators is manifested in their choices of computer hardware and software and of system configuration. Environment of work places are constantly changing; the need for timely software updates and frequent configuration changes. Configuration and Deployment tools provide different levels of automation out of all have the same basic goal: to help in the system configuration & deployment process.

II. DEVOPS TOOLS

A. LOG MONITORING TOOLS

There are number of DevOps log monitoring tools available in market. Here in this section some of the DevOps log tools are explained.

1) Splunk:

Splunk is America Multinational Corporation situated at San Francisco, California. Splunk is a log monitoring tool which produces software via web interface for searching, monitoring and also for analyzing huge amount of data generated by machine. Also it generates indexes and relates

this with real time data which is in a searchable repository to generate graphs, reports, alerts and visualization [18].

Splunk is proprietary software which does not have any daily limits for storing log data; it is capable of storing log data of 1 day; data is kept confidential in this tool; it alerts the user or the customer whenever the log data arrives or when the retention time is over through email[19]; there are large numbers of applications and plugins to support this tool; it is also used to manage the log data of cloud environment [21]; then moving over to search option in Splunk it takes the input from user and it automatically parse the according to the input given; it also uses regular expression for searching; It conducts more number of trainings, events and courses so that people get updated[20].

2) Sumo Logic:

Sumo Logic is a log management and analytic service that converts the machine generated huge data to real time IT insights. It takes machine generated big data and converts it into real time actionable insights. Sumo logic helps the IT operation teams to conduct the analysis to detect the cause of critical IT infrastructure; it also helps the DevOps teams to quickly analyze and troubleshoot the problem; also helps the IT security team to protect the large amount of log data [22].

Sumo Logic is proprietary software which has the limit to store the log data daily; it is capable of storing log data of 28 days; log data is kept confidential in this tool; it alerts the user whenever the data arrives and also it alerts whenever the data retention time is over through email; there are large there are large number of applications and plugins to support the tool; it reduces the thousands of log data into meaningful information of one page [21]; it is also used to manage the log data in cloud environment; then moving over to search option, it takes input from user and automatically parse the log data according to input given and also supports regular expression while searching; It conducts more number of trainings, events and courses to support the users[20].

3) Logstash:

Logstash is originated from Jordan Sessils's background in DevOps and system administration. When the Logstash founder Jordan was dealing with large amount of log files, he needed the centralized system to aggregate and manage the log files. So, Logstash is a centralized log management system and also is a part of open source stack in which Elastic search is used to index and search the log data and Kibana is used for visualization of data [24].

Logstash is an open source tool which has daily limits for storing the data [23]; it has the retention time nearly about 30 days; it alerts the user whenever the log data arrives or whenever the data retention time is reached through email; it reduces the thousands of log data into meaningful information of one page; it also manages the log of cloud environment; it automatically parse the data according to the input given by the user and also support regular expression while searching.

4) Kibana:

Kibana is a data visualization engine of open source stack (ELK stack). It interacts with all the log data in Elastic search via custom dashboards. It senses the log data to

create real time dashboards and then share them with other users in organizations. Kibana converts the log data into visualization like graph, dashboards. It has strong and efficient user interface so that users can sense, search, store and analyze the log data of Elastic search. And the main function of this is to visualize the Logstash data [25].

Kibana is an open source tool which has limit for storing the daily log data; it has the retention time of at least 30days; it has the paid storage system i.e.to store the log data the user pay; it alerts to the admin whenever the log data arrives and also alerts whenever the data retention time has reached; alerting is through email; then coming to searching, it takes the input from Elastic search and it also supports regular expression while searching; it also supports on demand analysis [23].

5) Paper Trail:

Paper Trail is log management DevOps tool that is available for general availability. Now a days companies have learned that it can get business IT insights from log data. It gives the place to search, store and analyze the log data from different computers or system for very low rate. It gives long term storage capacity for user with the help of amazon's s3 storage service [26].

Paper Trail is open source software which do not limit for storing the data daily; it has the retention time of 28 days [28]; it alerts the user or the admin whenever the log data arrives or whenever the retention time is reached through email; it can also be used to manage the cloud based organization; it supports on demand data analysis; in this there is a guarantee of the query completion [23].

6) Loggly:

Loggly is a cloud based management and analytics server which is situated at San Francisco, California. It does not use any kind of proprietary software agents; it instead uses open source technologies like Elastic search, Apache lucene4, Apache kafka. Other log monitoring tools like Splunk, Sumo Logic tells administrators what all problems have occurred in system but Loggly claims why the problem occurred or cause of the problem in system. Loggly is also said to be robust log analyzer which focuses on simplicity so that it helps the DevOps audience. Loggly helps in solving cloud centric problems [27].

Loggly is an open source technologies which has daily limits for storing log data; it is capable of storing log data of 30 days; it does not alert the user or customer whenever data arrives; when compared to other open source tools it has more number of applications and plugins to support this tool; it is used only to manage the log data of cloud based systems; in search option it automatically parse the data according to the input given by user but it does not support regular expression in searching the log data; It does not conduct more number of training and events for users [28].

B. SYSTEM MONITORING TOOL

1) Graphite

Graphite is an open source tool written in python, which is used to track the values of any metric that changes dynamically [4]. It stores the data, renders the graphs and it will monitor the performance of a computer system [5]. The user has to use some existing tools (like collectD, statsD,

Gmond and so on) or write some applications for collecting the data. It handles numeric time-series data. This scalable graphing tool can also be used with the cloud environment; plugins are used to collect the data, checks are used for monitoring the data.

Mainly it has three components [4], Carbon-this is the processing backend which listens to the data points and it can handle a huge number of clients. Whisper-it is similar to RRD that offers fast and reliable storage of the received data points over time. Graphite Webapp-Django webapp will render the graphs on demand, when there is a request for graph it retrieves the data from the disk and if it is not yet been written on to the disk then it will take the data points directly from the carbon in order to produce real-time graphs.

2) *Ganglia*

Ganglia is an open source monitoring tool, which can be used by high performance computing systems like grids and clusters, and its overall performance can be monitored [6]. It will use a multicast address to send a heartbeat message to all the nodes of the cluster to know whether it is available or not and the system and application-specific data will be shared by all the nodes where all the nodes will be having an entire overview of the cluster which can be used to recover the cluster after any crash [7]. This is a scalable distributed monitoring system which can be used in the cloud environment, plugins are supported and it can be easily integrated with other tools.

It is composed of three daemons [7], Ganglia monitoring daemon-which will be installed on all the nodes that has to be monitored and it will interact with the operating system of the host to acquire the system-specific metrics. Ganglia meta daemon- it will collect the information from many gmond or gmetad sources and stores it in the RRD (Round-Robin database). Ganglia PHP Web front-end- it will present the collected data. The graphs of any metric from two or more hosts can be aggregated so that it can be viewed simultaneously.

3) *Sensu*

Sensu is an open source tool written in ruby, it is used in the cloud environment. This monitoring router runs checks on the systems that need to be monitored which will return one of the following common exit status codes: 0-OK, 1-WARNING, 2- CRITICAL, 3-UNKNOWN and based on this result handlers will take actions [8,9]. This scalable monitoring tool will support plugins, checks and handlers. It is capable of scheduling downtime, it can be integrated with other tools also. It will notify the user in case of any outages.

The main components are [8], Sensu-server- will initiate the checks on the clients, it uses Redis database in order to store the persistent data and RabbitMQ is used to pass the data between the server and client nodes. Sensu-api-will be used to access the data that is stored in the Redis database. Sensu-dashboard will be used for displaying and managing events and clients. Sensu-clients will run on all the hosts that has to be monitored and executes the check scripts and the result is sent to server through RabbitMQ.

C. NETWORK MONITORING TOOL

1) *Zabbix*

This open source distributed monitoring tool will monitor various parameters of network services, servers and other network hardware. The status of the network and the health of the servers can be accessed from the remote location through the Zabbix server [10]. It may be used by large or small organization. It offers flexible notification mechanism, reporting, data visualization, polling, trapping, availability monitoring, log monitoring, proactive monitoring, logical grouping, agentless monitoring and automatic network discovery features.

The components of Zabbix are [10], Server-it is a central repository to which the agents will report the information. Database-to store the collected data. Web interface-it is a part of the server that should be located on the same physical machine on which the server is running. Agents-it will run on the monitoring systems and it will report the data to the server. Proxy-it is optional, but it is better to use proxy so that the load of a single server can be distributed.

2) *Nagios*

It is an open source tool used for monitoring, scheduling and alerting. It cannot do anything without the plugins which will return the exit state (OK, WARNING, CRITICAL) and a message that states the meaning of the exit state [11, 12]. Network services (say SMTP, HTTP and so on) and host resources (like processor load, disk usage etc.) can be monitored using the Nagios. It will notify the user in case of outages. Checks, handlers, plugins and templates are supported. This scalable tool can be used in the cloud environment and it supports some of the features like availability monitoring, log monitoring, distributed monitoring, proactive monitoring, automatic network discovery, and logical grouping.

The components of Nagios are, Scheduler-it is the part of the server that performs checks and results are sent to handlers. GUI-for displaying the alerts, state buttons and so on. Plugins-are used to add a feature to this tool.

3) *Cacti*

It is an open source tool designed as a front-end application which collects, manages and displays the graphs [13]. It handles data gathering, data sources and round robin archives. Some of the features supported by Cacti are polling, auto padding, log monitoring, proactive monitoring, logical grouping, automatic network discovery, agentless monitoring.

The Cacti operations are [14]: Data retrieval-this can be done through poller or SNMP is used for retrieving the data from remote hosts. Data storage-the data will be stored in the rrd files of the RRD tool. Data presentation-as it uses RRD tool it will provide a graphing function for producing graphs, it can be accessed through the browser when combined with web server.

C. BUILD AND TEST:

1) *Jenkins*:

Jenkins is one open source tool to perform continuous integration. It was split from Hudson after a dispute with Oracle. The basic functionality of Jenkins is to perform a

predefined list of steps based on a certain trigger. Jenkins is an award-winning application that monitors executions of repeated jobs, such as building a software project [31].

Jenkins also monitors the execution of the steps and allows stopping the process if one of the steps fails. Jenkins allows notifying users about the build success or failure.

Jenkins is simple and easy to install, understand and use, the truth that it is Java-based must not be an impediment to .NET development shops. Jenkins provides an impressive browser-hosted project management dashboard.

Jenkins can be started via the command line or can run in a web application server. Under Linux you can also install Jenkins as a system service. It supports file fingerprinting, it also support always security performance [30][34]

2) *Gradle:*

Gradle is a project automation tool that builds upon the concepts of Apache Ant and Apache Maven and is licensed under the ASL. Introduces a Groovy-based domain-specific language (DSL) instead of the more traditional XML form of declaring the project configuration. Unlike Apache Maven, which defines lifecycles, and Apache Ant, where targets are invoked based upon a depends-on partial ordering; it uses a directed acyclic graph to resolve the order in which tasks can be run [35].

Gradle build scripts are written in Groovy, not XML. But different other approach this is not for simply exposing the raw scripting power of a dynamic language, It Architecture contains Deep API, Plugins, Build tools and Libraries. It is the first build integration tool, it supports Easy of migration, Groovy, wrapper, scales and Multiproject builds. Also including CD optimization, Performance tuning, Standardization, Plug-in development.

3) *Maven:*

Apache Maven is a software project management and comprehension tool. Based on the concept of a project object model, which can manage a project build, reporting and certification from a central part of information.

Maven's main goal is to allow a developer to know the complete state of a development effort in the shortest period of time [29].

Maven does encourage best practices, but we realize that some projects may not fit with these ideals for past reasons. Even as it is designed to be flexible, to an extent, in these situations and to the needs of different projects, it cannot provide to every situation without making compromise to the integrity of its objectives.

It contains Core Engine provides project processing, Build life cycle management, Framework for plug-ins, Plug-in provides the core operation to build user projects and plug-ins also provides one or more goals and Repositories[36].

4) *Ant and Ivy:*

Apache Ant is a Java library and command-line tool whose work is to drive processes described in build files as targets and extension points dependent upon each other. The major known usage of Ant is the build of Java applications. Ant supply a number of built-in tasks allowing to compile, assemble, test and run Java applications [29].

Ant is written in Java and Ivy is a very powerful dependency manager oriented toward Java dependency management, although it can be used to manage dependencies of any kind. Software development projects look for a result combine build tool and dependency management can use Ant in grouping with Ivy.

It contains number of main features like Clean dependency reports, Non intrusive, Extremely flexible, Easily extensible, Transitive dependencies, Strong conflict management, Out of the box maven repository support.

5) *Cruise control:*

CruiseControl is an open source tool setup specifically to perform continuous integration software builds. CruiseControl architecture helps to easily understand the modules functionality [32].

CruiseControl is composed of 3 main modules: The build loop: center of the system, it triggers build cycles at that time notifies different listeners using a variety of publishing techniques. The trigger can be internal or external. It is configured in an xml file which maps the build cycles to certain tasks, thanks to a system of plug-in. Depending on configuration, it can produce build artifacts. The JSP reporting application allow the user to browse the results of the builds and access the artifacts. The dashboard provides a visual representation of all project build statuses

D. *DEPLOYMENT AND CONFIGURATION TOOLS*

1) *Puppet:*

Puppet tool is a configuration management system that allows you to define the state of your IT infrastructure, then without human intervention enforces the correct state. Whether to control just a few servers or thousands of physical and virtual machines, this tool automates tasks that sysadmins often do manually, freeing up time and mental space so sysadmins can work on the projects that deliver greater business value. Whether you're deploying vendor-supplied applications or working with a team of internal software developers, Puppet automates every step of the software delivery process: from provisioning of physical and virtual machines to orchestration and reporting; from early-stage code development through testing, production release and updates. Puppet ensures constancy, reliability and stability. It also facilitates closer association between sysadmins and developers, enabling more efficient delivery of cleaner, better-designed code [38]. Once you install Puppet, every node in your infrastructure has a Puppet agent installed on it. Also have a server designated as the Puppet master.

2) *Chef:*

Chef is a configuration management and automation platform from Opscode. Chef helps to describe the infrastructure with code. Since infrastructure is managed with code, it can be automated, tested and reproduced with ease.

Chef is a powerful automation platform that transforms complex infrastructure into code, bringing servers and services to life. Whether the user operating in the cloud, on-premises, or a hybrid, Chef automates how applications are configured, deployed, and managed across your network, no matter its size.

Chef is a thin DSL (domain-specific language) built on pinnacle of Ruby. This approach allows Chef to provide just enough abstraction to make reasoning about your infrastructure easy. Chef includes a built-in taxonomy of all the basic resources one might configure on a system, plus a defined mechanism to extend that taxonomy using the full power of the Ruby language. Ruby was chosen because it provides the flexibility to use both the simple built-in classification, as well being able to handle any customization path that organization requires[39].

3) Rancid:

RANCID monitors a router's or device configuration, including software and hardware (cards, serial numbers, etc) and uses CVS Subversion to maintain history of changes.

RANCID also includes looking glass software. Our version has added functions, supports cisco, juniper, and foundry and uses the login scripts that come with rancid; so it can use telnet or ssh to connect to your devices[40].

4) CFEngine:

CFEngine is a suite of programs for included autonomic management of either individual or networked computers. It has existed as a software suite since 1993 and this version published under the GNU Public License (GPL v3) and a Commercial Open Source License (COSL). CFEngine is Copyright by CFEngine AS, a company founded by CFEngine author Mark Burgess.

CFEngine is an open source configuration management system, written by Mark Burgess. Its primary function is to provide automated configuration and maintenance of large-scale computer systems, including the unified management of servers, desktops, consumer and industrial devices, embedded networked devices, mobile smartphones, and tablet computers [41].

5) Ansible:

Ansible is an IT automation tool. It can configure systems, deploy software, and orchestrate more advanced IT tasks such as continuous deployments or zero downtime rolling updates.

Ansible's goals are foremost those of simplicity and maximum ease of use. It also has a strong focus on security and reliability, featuring a minimum of moving parts, usage of OpenSSH for transport (with an accelerated socket mode and pull modes as alternatives), and a language that is designed around auditability by humans – even those not well known with the program.

Ansible is appropriate for managing small setups with a handful of instances as well as enterprise environments with many thousands.

Ansible manages machines in an agentless manner. Ansible is decentralized – it relies on your existing OS credentials to control access to remote machines; if needed it can easily connect with Kerberos, LDAP, and other centralized authentication management systems[42].

III. CONCLUSION

As per the present survey of devops tools with respect to the various metrics, In log monitoring tools combination of Logstash, Elastic search and Kibana is said to be the best tool. In system monitoring tools Graphite is said to be the best tool, in network monitoring tool Zabbix is considered to be the best tool. In build and test tool Gradle and Jenkins is said to be best tools. In Deployment and Configuration tool puppet is considered to be the best tool.

REFERENCES

- [1] <http://blogs.vmware.com/cloudops/tag/devops>
- [2] http://en.wikipedia.org/wiki/System_monitoring
- [3] http://en.wikipedia.org/wiki/Network_monitoring
- [4] Chris Davis, Graphite Documentation Release 0.10.0, October 18, 2014, <http://media.readthedocs.org/pdf/graphite/latest/graphite.pdf>
- [5] http://en.wikipedia.org/wiki/Graphite_%28software%29
- [6] www.ganglia.sourceforge.net
- [7] [en.m.wikipedia.org/wiki/Ganglia_\(Software\)](http://en.m.wikipedia.org/wiki/Ganglia_(Software))
- [8] <http://sensuapp.org/>
- [9] <http://sensuapp.org/docs/0.11/handlers>
- [10] <https://www.zabbix.com/documentation/2.0/manual/introduction/about>
- [11] en.m.wikipedia.org/wiki/Nagios
- [12] Afeez Yusuff, "Network Monitoring: Using Nagios as an example tool", May 2012.
- [13] [en.m.wikipedia.org/wiki/Cacti_\(Software\)](http://en.m.wikipedia.org/wiki/Cacti_(Software))
- [14] http://www.cacti.net/downloads/docs/html/operating_principles.html
- [15] <http://devops.com/features/log-data-valuable-tool-devops-lifecycle-beyond/>
- [16] <https://blog.logentries.com/2014/12/connected-qa-selenium-log-analysis/>
- [17] <https://www.loggly.com/blog/loggly-qa-talking-james-urquhart-continuous-integration-deployment-devops-role-log-monitoring/>
- [18] <http://www.splunk.com/>
- [19] Ledion Bitincka, Archana Ganapathi, Stephen Sorkin and Steve Zhang Splunk Inc, Optimizing Data Analysis with a Semi-structured Time Series Database
- [20] <http://venturebeat.com/2014/12/04/log-management-company-splunk-gets-competitor-sumo-logic-to-change-its-messaging/>
- [21] <http://www.javacodegeeks.com/2015/04/log-management-tools-face-off-splunk-vs-logstash-vs-sumo-logic.html>
- [22] <https://www.sumologic.com/>
- [23] <http://blog.takipi.com/the-7-log-management-tools-you-need-to-know/>
- [24] <http://logstash.net/>
- [25] <https://www.elastic.co/products/kibana>
- [26] <https://papertrailapp.com/>
- [27] <https://www.loggly.com/>
- [28] <http://stackshare.io/stackups/papertrail-vs-loggly-vs-aws-cloudtrail>
- [29] <http://zeroturnaround.com/rebellabs/java-build-tools-part-2-a-decision-makers-comparison-of-maven-gradle-and-ant-ivy/>
- [30] <https://wiki.jenkins-ci.org/display/JENKINS/Jenkins+Best+Practices>
- [31] http://en.wikipedia.org/wiki/Jenkins_%28software%29
- [32] <http://cruisecontrol.sourceforge.net/overview.html>
- [33] Tim Heckel (18 Feb 2013). "Meet Travis CI: Open Source Continuous Integration". *InfoQ*. Retrieved 28 June 2013.
- [34] <http://java.dzone.com/articles/continuous-integration-aint>
- [35] <http://www.drdoobs.com/jvm/why-build-your-java-projects-with-gradle/240168608>
- [36] https://docs.oracle.com/middleware/1212/core/MAVEN/maven_versions.htm#MAVEN8855
- [37] <https://blog.newrelic.com/2014/06/02/devops-tools/>
- [38] <https://puppetlabs.com/puppet/what-is-puppet>
- [39] https://docs.chef.io/chef_overview.html
- [40] <http://www.shrubbery.net/rancid/>
- [41] <https://auth.cfengine.com/archive/manuals/cf3-Reference>
- [42] <http://docs.ansible.com/>