Analysis of Devops Tools using the Traditional Data Mining Techniques

R. Vaasanthi	V. Prasanna Kumari, PhD	S. Philip Kingston
Research Scholar,	HOD, MCA	Project Manager
SCSVMV University	Rajalakshmi Engineering	Infosys, Mahindra City,
Kanchipuram	College, Chennai	Chennai

ABSTRACT
DevOps - a portmanteau of Development and Operations is a set of practices focused on using latest generation tools to automate the configuration process for system resources and application components [2]. Process efficiency improves from hours and days to seconds and minutes. IT performance strongly correlates with well-known DevOps practices such as use of Version Control and Continuous Delivery. The longer an organization has implemented — and continues to improve upon — DevOps practices, the better it performs [7]. And better IT performance correlates to higher performance for the entire organization. Today, enormous amount of tools are present in DevOps space.

Keywords
Data mining, DevOps and Classification

1. INTRODUCTION
The main objective of this paper is to apply Data Mining (*The practice of examining large preexisting databases in order to generate new information*) techniques in DevOps tools space, as multiple DevOps tools and solutions are available and new tools are being released in the market.

As application development lifecycle involves multiple stages [11], DevOps too is spread across the following phases - Source Code Management, Build and packaging, Code quality Analysis, Code Review, Unit Testing, Functional testing, System & Performance testing, Environment Provisioning & Deployment and System monitoring. There are multiple tools serving a specified purpose and those that provide a combination of Devops capabilities across the phases specified.

2. DATA MINING:
Data Mining is the process of analyzing data from multiple perspectives and extracting useful information [4]. Data Mining involves extraction, transformation and loading transactions into a Data Warehouse. Data mining process includes anomaly detection, association rule, learning, classification, regression, summarization and clustering. Data Mining concepts and methods can be applied across various fields like Marketing, Medicine, Real Estate, Customer Relationship Management, Engineering, Web Mining, etc. This paper categorizes DevOps tools using the *Classification technique.*

3. DEVOPS
DevOps is supported by a culture of Collaboration, Automation, Measurement and Information sharing and Web service usage. DevOps benefits both development and operations performance. It also has positive effects on web service development and quality assurance performance. DevOps delivers tangible results with many improvements, depending on the current Devops maturity state of an Organization. A recent survey reveals - 46% of IT decision makers are already experiencing an increased Deployment frequency in their software and services, with another 44% expecting the same results in the near future. 39% are seeing an increased number of End user or customers using their software and services, and 39% said they are currently seeing increased collaboration between departments. More than one-third of the IT organizations have vouched that they have seen the quality and performance of their software improve by (36%) and (34%) respectively [5]. They were also reportedly able to reduce the time spent in fixing and maintaining applications.

The results indicate that as DevOps matures in the industry, implementations will mature and deliver upon metrics that continue to be important to both the IT organizations and their business counterparts. Today 88% of those surveyed indicated they had already adopted or planned to adopt DevOps within the next five years. However, the challenge is in identifying the right tool set for the pipeline; a wide range of tools performing similar operations are available in the market, from Source Code Management process up to Deployment phase. The purpose of this study is to identify an optimized tool set for the Continuous Integration and Continuous Deployment pipeline.

4. CLASSIFICATION
Classification is the processing of identifying a set of models (or functions) which describe and distinguish data classes or concepts. It consists of assigning a class label to a set of unclassified class. There are two types of classification; supervised classification and unsupervised classification. The difference is the set of possible classes is known in advance for former and set of possible classes is not known in later and after the classification a name is assigned to the class which is often referred as *Clustering Technique* [9].

The objective of supervised classification is to analyze the input data and to develop an accurate description or model for each class using the features present in the data.
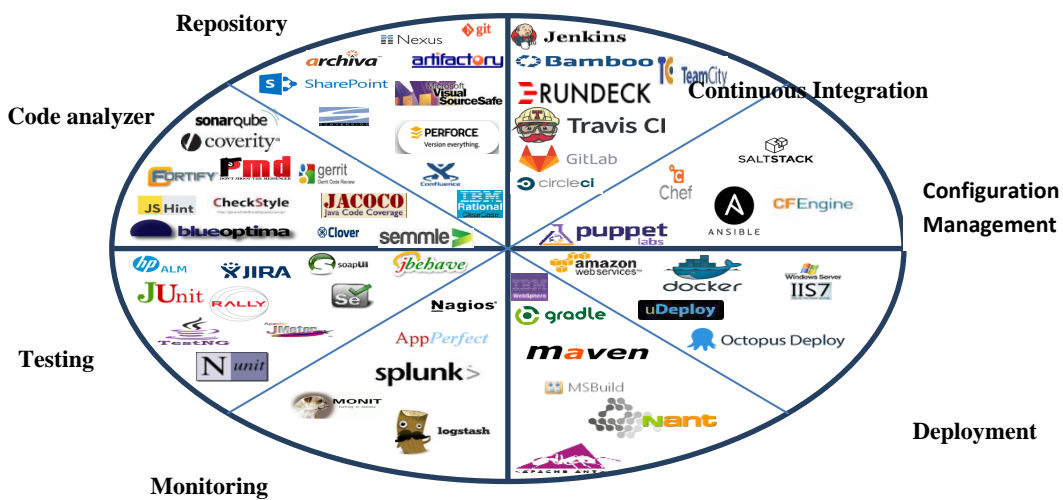
Fig 1. Devops Space*

Fig 2. Supervised Classification

Classification maps the data into predefined targets [10]. It is a supervised learning as targets are predefined. The aim of the classification is to build a classifier, based on common behavior or on attributes used to describe the object or identify attributes to describe a group of these objects. The data collected is based on the frequency of the tools usage in implementing Continuous Integration and Continuous Deployment pipeline. Tools from DevOps space are classified using labels based on the predominant functionality they are designed to operate, this ensures that each and every tool falls exactly in one subset, along with the group of similar tools. The tools are classified under different labels as Software Repository, Continuous Integration, Configuration Management, Deployment tools, Build tools, Monitoring tools, Testing tools and Code analyzer/review tools as predefined targets. These labels can be further classified to deeper levels as below. Example, considering the label Repository this can be further classified into four levels as below.
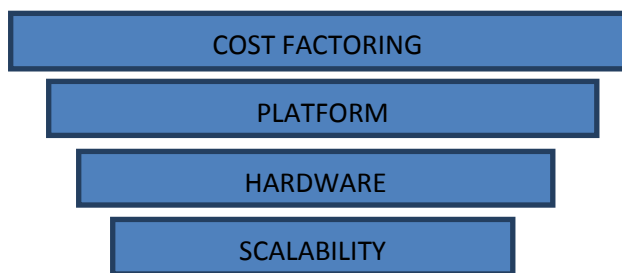

Fig 3. Levels of Repository Classification

Repository labels can further be classified to four different levels and each level can be classified further. For example, Cost factoring can classify the tools based on License or Freeware, Platform will classify the tools under Windows or UNIX based platforms. The Hardware based classification can group tools under Cloud Based or On-Premise based setup [3]. Scalability would classify the tools based its capability to a system, network, or process to handle multiple services.
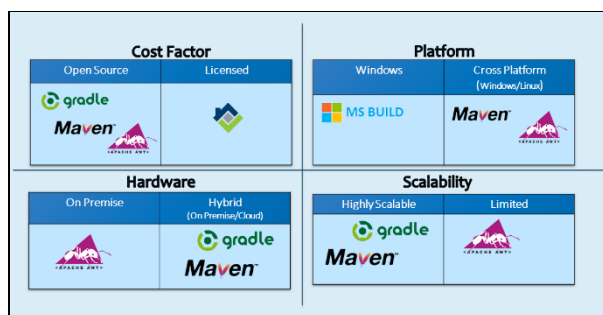

Fig 4. Build Tools Classification *

* Logos are used for illustrative purpose only

5. CONCLUSION

In this paper, a new method is proposed to categorize tools in the DevOps Space by applying the Classification Data Mining technique. The proposed methodology creates deeper levels of classification more efficiently; the technique can be applied for other labels of the supervised classification. It also improves the quality of solution to classify the tools more accurately.

6. REFERENCES

[1] CSI Communications, Knowledge Digest for IT community ISSN 0970-647X, Volume No: 38, August 2014.

[2] Devops for Big data @ Supaket, Solution Architect at Enersys.co.th, August 2016

[3] Dell Software, "Devops for the cloud", achieving agility throughout the application life cycle, Dec 2014.

[4] Han J. and Kamber M., Data Mining: Concepts and Techniques, 2nd ed., San Francisco, Morgan Kauffmann Publishers,2001

[5] Shahrukh Teli and Prashasti Kanikar., A Survey on Decision Tree Based Approaches in Data Mining., Volume 5, Issue 4, 2015., Volume 5, Issue 4, 2015.

[6] Li, Linna, and Xuemin Zhang. "Study of data mining algorithm based on decision tree." In Computer Design and Applications (ICCDA), 2010 International Conference on, vol. 1, pp. V1-155. IEEE, 2010

[7] CA technologies, Research paper: DevOps: the worst-kept secret to winning in the application economy., October 2014

[8] Floris Erich, Chintan Amrit and Maya Daneva., Research Gate, "Report: DevOps Literature Review", October 2014

[9] Saroj, Tripti Chaudhary., "Study on Various Clustering Techniques" International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 3031-3033

[10] Shameem Fathima,D. Manimegalai and Nisar Hundewale, " A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus- Dengue" IJCSI International Journal of Computer science issues, Vol 8 Issue 6m No 3 , 2011.

[11] https://en.wikipedia.org/wiki/Systems_development_life_cycle

[12] S. Anupama Kumar and M.N. Vijayalakshmi "Relevance of Data mining techniques in efficient sector" International Journal of Machine Learning and Computing. Vol 3, No.1, 2013.