CLOUD BASED RESOURCE PREDICTION TOOL FOR DEPLOYMENT MIGRATIONS FOR DEVOPS

A.G.N.M. Ranaweera

(179347F)

M.Sc. in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

January 2020

CLOUD BASED RESOURCE PREDICTION TOOL FOR DEPLOYMENT MIGRATIONS FOR DEVOPS

A.G.N.M. Ranaweera

(179347F)

This dissertation submitted in partial fulfilment of the requirements for the Degree of MSc in Computer Science specializing in Information System Management

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

January 2020

DECLARATION

I declare that this is my own work and this MSc Thesis Project Report does not incorporate without acknowledgment any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works.

---------------------------------- ------------------------------------

A.G.N.M.Ranaweera Date

I certify that the declaration above by the candidate is true to the best of my knowledge and that this project report is acceptable for evaluation for the MSc Thesis (CS6997).

---------------------------------- ------------------------------------

Dr. Dulani Meedeniya Date

---------------------------------- ------------------------------------

Dr. Indika Perera Date

ABSTRACT

Cloud computing is a promising paradigm for delivering computer services for industries, due to the advances in technologies such as virtualization, networking, web services, etc. As the cloud host providers offer an infinite amount of resources to their customers, the overhead of procuring traditional infrastructure resources has been reduced. In addition to the infinite resource offers, the dynamic resource scaling feature in the cloud allows the enterprise application owners to choose an appropriate pricing model. This feature in the cloud reduces the risk of over-provisioning and under-provisioning. Selecting an appropriate dynamic resource scaling model in the cloud platform is a tedious task as it depends on the accuracy of the forecasted resource provisioning values.

Based on this deliberate background, this research intends to model a POC (Proof of Concepts) for the future resource prediction in the cloud platform which assists the dynamic resource provisioning scaling. These resource provisioning predictions were done using the performance metrics gathered in the on-premises servers. The performance metrics gathered during 30 days were evaluated using the ARIMA (Autoregressive Integrated Moving Average) time series model. To improve the accuracy of the seasonality effects, the 30 days matrics have been gathered. Out of many ARIMA models, the best fitted ARIMA model was chosen based on the least RMSE (Root Mean Square Error).

The best fitted ARIMA model has predicted the future resource requirement of resources, by evaluating the gathered metrics. The model was trained based on the 75% data sets consisting of 22 days metrics and was tested using 25% data sets consisting of 8 days metrics. As a best practice, the data set splitting ratio has been selected as a 75:25 which consists of a higher number of data matrics for training. The resource provisioning model has been evaluated in the IaaS (Infrastructure as a Service) layer of the Amazon cloud. As the experimental cloud infrastructure, an EC2 (Elastic Cloud Compute) in the Amazon Web Services has been used. The observed values were evaluated using RMSE, MAPE (Mean Percentage Error), precision, recall, and F1 score factors as accuracy assessment criteria. The obtained accuracy percentages of CPU, RAM, Disk, and Network were respectively 84.07%, 93.13%, 90.47% and 71.16% which demonstrates a considerable accuracy for future forecast prediction. Further, visualization of the future forecasted resource provisioning results is provided in a dashboard. The developed dashboard has been evaluated through a survey comprised of a questionnaire that was distributed among a DevOps team which is consists of ten members. The sample of the survey consisted of both technical and non-technical users who have 4 to 14 years of work experience in the IT industry. The overall average satisfaction rate of the survey was 4.1 out of 5, which was a considerably acceptable result. Based on the survey results and accuracy percentages of assessment criteria, it can be concluded that the research work has been successful and the proposed dashboard will be used for resource provisioning management.

ACKNOWLEDGEMENTS

I would like to express my great appreciation to my research supervisor Dr. Dulani Meedeniya and the M.Sc. research coordinator Dr. Indika Perera, for their guidance, enthusiastic encouragement and useful critiques of this research work.

My grateful thanks are also extended to Mr. Dinesh Asanka and Miss Inoka Loluwagoda for guiding me in the entire academic life journey. Special thanks go to the colleagues in my office, for their help and encouragement throughout my study. I also would like to precise my gratitude to the staff members of the CSE department for supporting me in the entire academic years.

I am grateful to all of those with whom I have studied during these two years. Since my heartiest thanks go to my M.Sc. colleagues who encouraged me during the studies. Specially Mr.Tiran Wijesekara, Mrs. Trisha Amarasekara, Mr.Ryan Benjamin, Mr. Vilochane Vidyarathne, Mr.Lakshan Costa and Miss. Nithila.Without them, I would not able successfully complete this. I would also like to express my gratitude to Mrs. Damayanthi Herath, Mrs. Shyama Ranasinghe, and Miss Niroshika Wickramarathane. I always knew that you believed in me and wanted to see the best for me.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. I would like to extend gratitude to my ever-loving mother, father, sister, and brother; whose love and direction are with me in whatever I achieved.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

Abbreviation	Description
API	Application Program Interface
ARIMA	Autoregressive Integrated Moving Average
AWS	Amazon Web Services
BA	Business Analysis
BI	Business Intelligence
DB	Data Base
Dev	Development
DevOps	Development and Operations
EA	Enterprise Application System
ETL	Extraction, Transformation, Loading
FTE	Full Time Equivalent
IaaS	Infrastructure as a Service
IT	Information Technology
MAPE	Mean Percentage Error
ML	Machine Learning
PaaS	Platform as a Service
POC	Proof of Concept
QA	Quality Assurance
RMSE	Root Mean Square Error
SaaS	Software as a Service
SD	Standard Deviation
SSIS	SQL Server Integration Server
VM	Virtual Machine

1. INTRODUCTION

1.1. Background

Cloud computing has been an encouraging standard for distributing IT services, due to advances in technologies of virtualization, networking, etc.. Cloud hosting providers offer, a vast amount of resources (e.g. CPU, memory, network I/O, disk, etc.) to their customers which will be reducing the overhead of obtaining traditional infrastructure resources. Other than the unlimited resource pool for scale and elasticity, cloud providers also allow, dynamic resource scaling using diverse pricing models such as pay-per-use model, reserved instance pricing model, spot instance model, etc. Therefore, enterprise application owners can select a suitable pricing model based on workload characteristics and resource usage patterns. This reduces the risk of over-provisioning and under-provisioning, resource wasting during non-peak hours, etc. Therefore enterprise-level organizations are able to focus on providing services to their customers while consuming the required computing resources as a utility [1].

The pay-as-you-go model and dynamic resource provision model are assisting to reduce the overhead related to the static provisioning of the cloud platform. Because the static provisioning model is not deliberated as a cost-effective option as it can be led to over-provisioning and under-provisioning at any particular time. Hence eliminate this resource provisioning issue, there are many resources prediction models that are predicting the right amount of resources based on ML technologies.

Today's business environment is continuously changing with new technologies and new business trends. Therefore enterprise applications should be able to continuously adapt to the new business requirement. The digital transformation is the one way to adapt to the new requirements by improving business competitiveness. DevOps and cloud migrations are two of the ways companies can achieve this digital transformation.

DevOps has been clarified as a culture or process or practice within an organization. It is increases the communication, collaboration, and integration of the Development (which includes the QA team) and the Operations (IT Operations) teams. The ultimate intention of the DevOps is to automate and speed up the software delivery process lot more frequently and reliably [2]. This helps to increase the trust, faster software release, ability to resolve the critical issues quickly and better manage unplanned work [3].

Cloud migrations are a highly complex process and it requires, in-depth planning to effective resources management, risk management, and cost management, on-time implementation and ultimately operational success [4]. The fundamental process around the migration is to understand the benefit of a new system, measuring the gaps in the current system, planning the cloud migration approach and finally migrating to the cloud [5]. The most critical part of the cloud migration process is developing an effective and efficient planning approach. A most important task of the planning process is analyzing the architecture of existing on-premises servers and applications [6]. As cloud providers offer, on-demand resource scaling, it is a must to identify the existing resources usage patterns when provisioning the resources.

Identifying the resource usage patterns over time and workloads by prediction-based resource measurements and provisioning approaches using time series analyzation is used for dynamic resource provisioning in the cloud platform. Predictive analysis of resource usage patterns is a significant factor for the migration plan as it leads to decide workload administration, capacity arrangement, system sizing, and dynamic rule generation in the cloud platform.

The research background for this thesis is basically based on, resource provisioning for dynamic resource scaling in the cloud platform and DevOps process in the Enterprise Application Management Industry (EAM). Resource provisioning prediction will be performed based on the resource utilizations logs which have been gathering from a Business Intelligence System (BI). Different performance patterns will be identified against the different processes and different resources. Predicted results will be used as inputs for cost management and resource provision management of the cloud platform.

1.1.1. Overview for Business Intelligence System

BI (Business Intelligence) is one of Enterprise Application Systems (EPA) which is mostly used by top-level management of the company. It is supporting to make a better decision, by providing historical, current and predictive views of business operations. BI system is an analytically oriented system that presented integrated and analyzed information [7]. Figure 1.1 represents the overall view of the BI system.

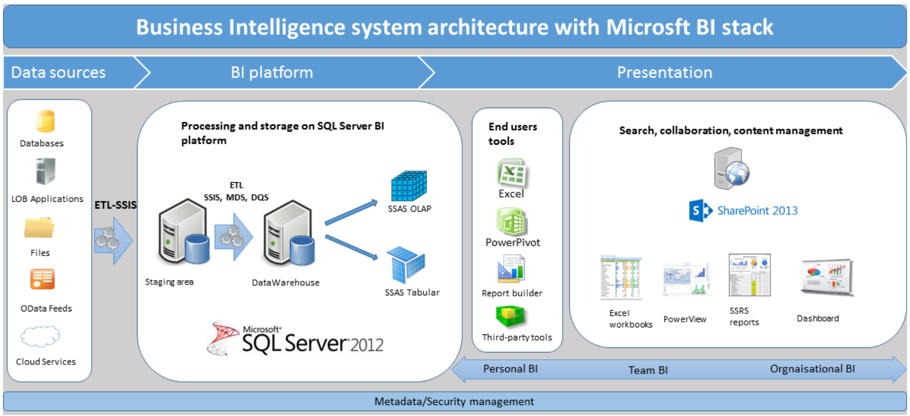

Figure 1.1: Overview of BI System, Source: [8]

BI system has four basic components which are namely source systems, ETLs (Extraction, Transformation, and Loading), data warehouses or cubes and reporting module. Source systems can be a database, a file, applications, and a cloud service.
BI system of an enterprise business has many applications like Human Resource, Finance, Resource (Employees) Management, Time Management, Project Management, etc. BI system is supposed to extract data from all these applications and load the data into the cubes for later use.

At the time of ETLs execution, data are being extracted from the source systems and transformed (the process of converting and measuring the data) into the information. This transformed data would be loaded into the data warehouses which will later use for reports and leader dashboards.

1.1.2. Overview of DevOps Practices

DevOps concept is a new trend in the IT industry which is concurrently involving the overall life cycle of the software development process. DevOps is a set of practices that automates the process between the software development team and operations team, in a manner that they can build, test and releases new products faster and more reliability [9]. This helps to increase the trust, faster software release, ability to resolve the critical issues quickly and better manage unplanned work. Figure 1.2 represents the overall practices involving in DevOps.

Before introducing the DevOps concept, there were many conflicts between the development and the operational teams. As an example, the operations team is interesting to keep system stability and security so they do not want to change the product frequently; whereas the development team is interesting to add new features, as a result, the overall system will keep changing [10].

DevOps concept has been proposed to overcome these conflicts. DevOps is a relatively new concept that is involving with every stage of the software life cycle process, from requirement analysis to product maintenance. The goal of this concept is to improve communication and collaboration among all the team members [11].

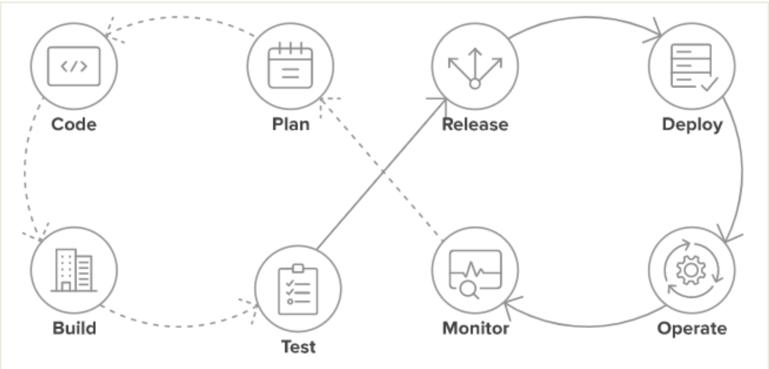

Figure 1.2: Overview of the DevOps Practices, Source: [11]

1.1.3. Overview of Cloud Migrations

Cloud technologies have recently appeared as a new trend for hosting and distributing services through the internet. Therefore most companies have been tending to migrate their EASs to cloud infrastructure as it would be more benefited. Cloud users are able to get competitive advantage since cloud services offer the most innovative technologies such as customize the applications, scale up and scale down services, easy accessibility, low infrastructure cost and maintained cost, etc. [12].

Investing for expensive infrastructure that cannot change parallel to modern technologies is a wastage. Because of this most of the companies have been choosing cloud technologies for their future. Introducing cloud technologies has a tremendous impact on computer services over the past few years. According to the explanations given by Stuart Scott who is a trainer of AWS, cloud computing can be identified as a remote and virtual group of on-demand collective resources that contribute compute, storage and network services [13].

Cloud computing provides a service-driven business model that the hardware and the platform level resources are delivered as services on-demand basis. Basically, these services can are arranged into three types. They are specifically, software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS). Other than the above business model, three different cloud deployment models can be identified namely public, private and hybrid [14].

The market share of global cloud computing has been divided among large cloud providers like Amazon, Microsoft, and Google. Other than these cloud providers Apple, Cisco, Citrix, IBM, Joyent, Rackspace, Salesforce.com and Verizon/Terremark, etc., also offers cloud services to the global market [15]. Figure 1.3 represents the market share value of cloud providers.

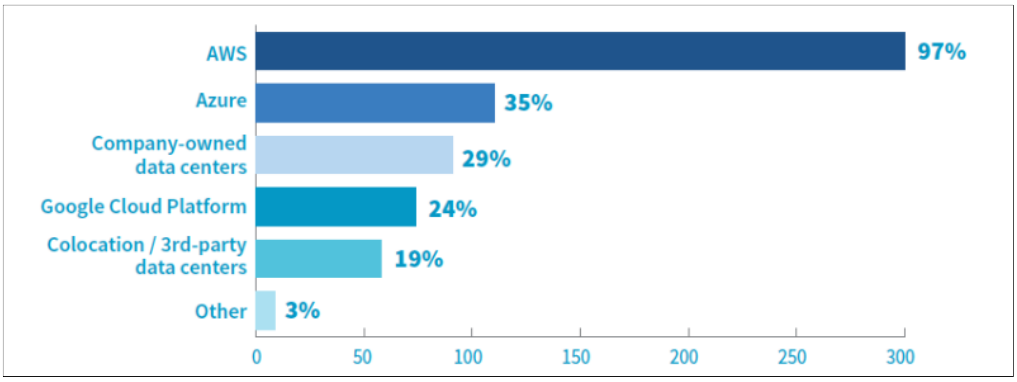

Figure 1.3: Market Share of the Cloud Vendors (the year 2019), Source: [16]

1.2. Problem Statement

The main problem addressed in this research is to develop a DevOps tool, which shall be used to predict the resource provisioning. This forecasted resource provisioning results can be applied for managing an infrastructure migration like the IaaS deployment in the cloud migration platform. This DevOps tool would be predicted the resource provisioning, based on the performance metrics gathered, during the execution of various processes in the on-premise servers.

The IaaS cloud deployment model has the capability to dynamically obtain additional or release current computing resources based on application workload patterns [17]. Therefore, there are costs and overhead in dynamic resource provisioning in cloud services. The lack of efficient resource provisioning in the cloud can lead to the risk of under-provisioning and over-provisioning.

The business should be cable of design and provision the infrastructure resources in a manner that, the capacity of the infrastructure can equal to the requirements of a new application. This research addresses the identified problems and proposes a DevOps tool to predict resource provision for infrastructure migration.

1.3. Motivation

As mentioned in Section 1.2, the business should be able to design and provide the infrastructure resources such that, the capability of the infrastructure can equal to the demands of a new application. The lack of efficient resource provisioning can lead to saturation or the waste of resources during the non-peak hours which could be considered as the most significant challenges the enterprise is faced with. The enterprises should be able to go for optimal resources provision management which is cost-effective and fully utilizes the acquired infrastructure resources.

It is rare to find a systemic way in the industry to forecast the application workload by using a DevOps tool, such that the enterprises can be used as a vendor independently. The resource provisioning tool introduced in this work forecasts the application workload based on the utilization of log patterns gathered, during the execution of various processes.

Based on the predicted resources provision results, the enterprise would be able to allocate the optimal infrastructure resources matching with new application requirements. That means, solution architects can be assigned extra resources for peak time and can be released during the low demands times.

Considering all these factors, the prediction of resource provision can be identified as an important factor to migrate the on-premises application.

1.4. Objectives

The main research objective is to address the above-stated challenges in Section1.3 while mitigating the limitations of existing methodology and tools. The ultimate outcome of this research would be a DevOps tool that supports resource provisioning management for infrastructure migration.

In addition, this research aims to achieve several sub-objectives while working to fulfill the overall research objective. These sub-objectives have been stated below.

- To research and analyze the perspectives of the DevOps team members.
- To review the literature on research in cloud migrations and DevOps tools and identify their limitations.
- To define an optimal methodology to resolve the identified problems and limitations.
- To present the predicted resource provisioning results in a dashboard.

1.5. Outline

Chapter 01

The introduction chapter comprised of the research background and the problem statement that would be addressed through the research. Finally, it described motivations and objectives that need to be achieved at the end of the accomplishment of research.

Chapter 02

The second chapter consisted of details of analyzation in similar researches, methodologies, and tools. It also reviewed the aids of DevOps tools and cloud migration processes. The association between the DevOps and the cloud also discussed in chapter two.

Chapter 03

The third chapter presented the optimal approach selected at the end of the literature review. It also described the several steps that followed during the approach. This included the methodologies for the ARIMA model, dashboard and the experimental model in the Amazon EC2 platform.

Chapter 04

The fourth chapter discussed the implementation steps with depth technical details that applied during the development. It consisted of the implementation process for performance metrics gathering, resource provision forecasting, and dashboard visualizing.

Chapter 05

The fifth chapter compromised of the evaluation results of the DevOps tool and the dashboard. The outputs evaluated based on the MAPE, RMSE, precision, recall and F1 score and the dashboard evaluated based on the questionnaire’s results.

Chapter 06

The sixth chapter included the research contribution, research limitations, and future work.

2. LITERATURE REVIEW

2.1. Introduction to DevOps

As defined in Section 1.1.2, DevOps is a collection of practices that automates the process between the software development team and the operations team. Further it will act as an interface between above two teams in a manner that they can build, test and releases new products faster and more reliability [9]. In many organizations, deployment and operation teams are two different sections and they work separately [18]. Once the software is built and tested, the developers transfer the software to the operations team with the deployment guide. But developers do not communicate much with the operations team.

As a result, deployment becomes more complex, error-prone, and time-consuming since there are no properly defined processes. DevOps address exactly these problems which emphasize the neediness for communications and collaborations between all participants in the overall process from requirement gathering to actual release. Furthermore, every single person has a reasonability to success in this overall project, not just for one individual team or one person [19]. Figure 2.1 represents collections of functions that involve an overall DevOps process [20].

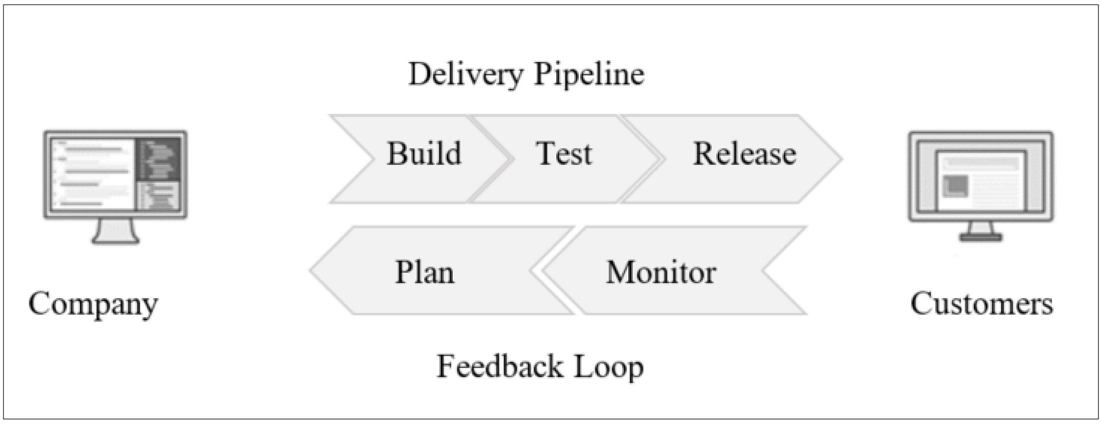

Figure 2.1. DevOps Processes, Source: [20]

2.1.1. Benefits of DevOps

Under DevOps practices, development and operations teams are no longer solid and sometimes these two teams are merged into one team and they worked across the entire application lifecycle. Basically, these two are practices to automate the processes that have been slow and inefficient historically [20]. Other than automation, there are many benefits that can be gained from DevOps.

- Speed up the delivery process through micro-services and continuous delivery.
- Rapid delivery through continuous integrations and continuous delivery.
- Ensure reliability through the monitoring and logging practice.
- Scale the infrastructure and development process through the infrastructure as code practice.
- Improved the collaborations and share many responsibilities and combine the workflows [20].

2.1.2. DevOps Process and Tools

DevOps is more about changing the existing process and simplifying the workflows in the overall process. There are many open-source tools are available which supports to simplify and automate the existing workflows [21].DevOps is consisting of many processes. Because of this reason a single tool cannot be found which supports the overall process [22]. It is more of ‘ToolChain ‘, or a ‘Suit of Tool’ that support for development, deployment or maintenance.

2.1.3. DevOps and Cloud

Companies cannot ignore the digital transformation because it directly effects to increase the competitive advantages [23]. DevOps and cloud computing have become the ways that companies can achieve this digital transformation. DevOps is related to the processes and processes improvement and cloud computing is related to technology and services. Because of this reason, the relationship between DevOps and the cloud is not easily reconciled. But there are many ways that DevOps and cloud links with [24].

The centralized nature of the cloud infrastructure provides DevOps automation with a standard platform for testing, deployment, and production. Most public and private cloud computing provide continuous integrations and continuous development tools which helps to the DevOps automation. Cloud-based DevOps platform helps to track the cost of resources and make an adjustment when needed. Resources usages can be tracked by application, developer, user data, etc. [25].

2.2. Introduction to Cloud Migration

Cloud technologies have lately appeared as a recent tendency for hosting and distributing services through the internet [26]. Many companies tend to migrate their EASs from on-premises to cloud environments because they do not want to give their competitors an advantage. Furthermore, many technology barriers are break by cloud migration and cloud offer the speed, scalability and cost savings.

Most of the companies have been understanding that there are no points of investing for expensive infrastructure that cannot change parallel to the modern technologies, because of this to the cloud has become a choice for their future [27].

Introducing cloud technologies has influenced incredibly, to the IT business during the past few years. Huge cloud providers like as Google, Amazon, and Microsoft provide additional dominant, trustworthy and cost-efficient cloud platforms. Enterprise-Level business owners are seeking to shape their business model by gaining benefits from these cloud platforms.

There are many benefits that businesses can achieve through cloud computing like no up-front investment, let down the operating cost, extremely extensible, easy to reach, minimizing business risks and maintenance cost [26].

2.2.1. Business Model of Cloud

Cloud computing provides a service-driven business model. That means the hardware and platform level resources are delivered as services on-demand basis [26]. These services can be grouped as the software as a service (SaaS), platform as a service

(PaaS) and infrastructure as a service (IaaS) [28]. Figure 2.2 represents three business models of cloud, based on the cloud platform layers.

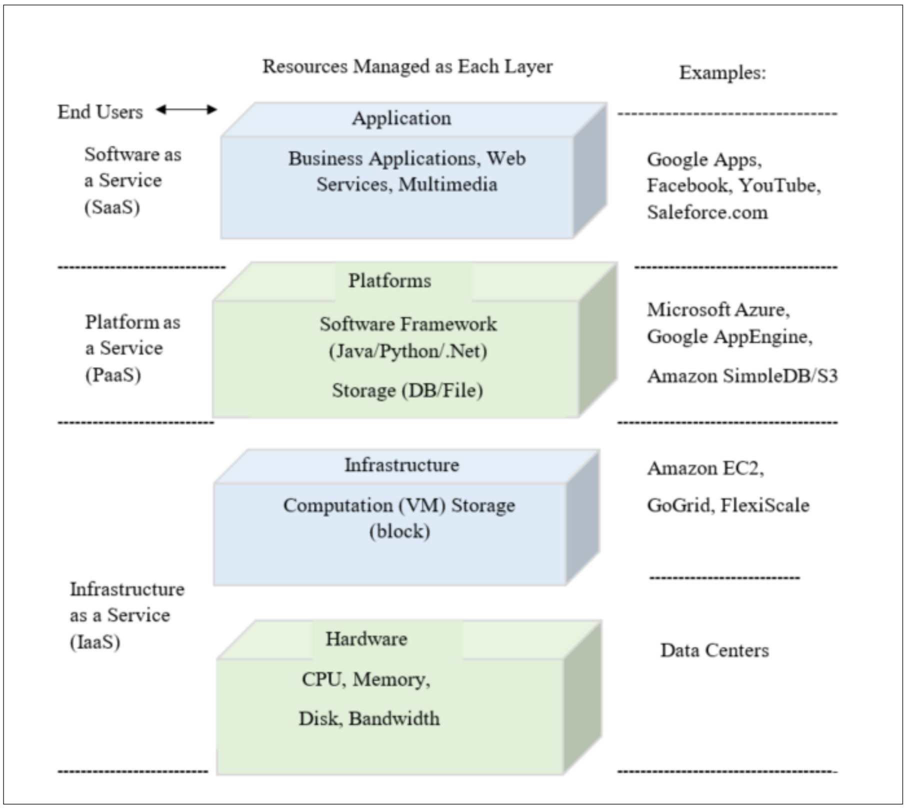

Figure 2.2. Cloud Computing Architecture, Source: [26]

- **Software as a Service (SaaS)**

Software as a service denotes to delivering on-demand applications through the internet. It trends to reduce the cost of the software by removing installations, manage and upgrade costs as well as licensing costs. Most SaaS applications are provided based on a subscription model [26].

- **Platform as a Service (PaaS)**

Platforms as a service refer to delivering platform layers resources such as operating system support and software development frameworks. Businesses can requisition resources as per their requirements and scaling as demand grows. This helps to reduce investing in hardware with redundant resources [28].

- **Infrastructure as a Service (IaaS)**

Infrastructure as a service denotes to on-demand provisioning of infrastructure resources such as VMs. These VMs are highly automated and scalable compute resources that can be self-provisioned and available on-demand. IaaS clients have direct access to their VMs and storages as same as on-premises servers but a higher order of scalability [28].

2.2.2. Cloud Migration Strategies – “6 R” Framework

Planning a cloud migrations is not an easy thing as there are many aspects a company should follow. These aspects are namely, dependencies on existing business processes, business goals, current infrastructure, current applications, etc. [29]. The “6 R” framework which is defined by AWS professional services is outlined below. This framework groups the existing application into six groups namely, Re-host, Re-factor, Re-purchase, Retire and Retain [13].

- **Re-host (Referred to as a “lift and shift”)**

Re-hosting is the process of moving on-premises servers and virtual servers to the cloud the solution based on the IaaS. The main advantage of this approach is that the system can be migrated quickly without changing their existing architecture. In this approach, the cloud is considered as just another data center.

- **Re-Platform**

In this approach, the core architecture of the applications will not be changed. As an example in this process, it will move the WebLogic to Apache Tomcat.

- **Refactor**

Refactoring is a process of running current applications on the infrastructure given by the cloud based on the PaaS. This option is allowing companies to focus on the applications without worrying about the infrastructure (Storage, CPU, and RAM).

- **Repurchase**

Certain applications need to modify more extensively before they migrating to the cloud. Some applications may require adding some functionality while others may require change existing architecture completely. The number of changes will be based on application design.

- **Retire**

In this approach; remove the applications that are no longer needed. These savings can be improved the business case and can be decreased the number of applications that has to be secured.

- **Retain**

In this process, the application that is most critical and important to the business would be kept. But these applications are required major refactoring before they can be migrated.

2.2.3. Decision Support Tools for Cloud Migration

At the 2011 IEEE 4th International Conference on Cloud Computing; has been publishing the research on ‘Decision Support tools for Cloud Migrations in the Enterprise’. This research paper has been described two tools that support to make decisions during the cloud migrations of EASs. The first model supports to produce cost estimates and the second model supports analysis of the cloud migration risk.

- **Cost Modelling**

There are several types of costs that Enterprises should be considered during the cloud of their cloud migrations process. They are IT infrastructure, data center equipment, software licenses, and software modifications, employees’ expenses, etc. These costs can be evaluated using spreadsheets. The cloud providers like Amazon have provided spreadsheets for cost assessments [30].

This cost modeling tool has simplified the process of procurement infrastructure cost estimations through different clouds. The tool can be taken by IT architects or systems engineers to generate a model of their system.

- **Benefits and Risk Assessments**

As the IT business; it is important to consider the cost savings factor in wide capacity more than other factors. Before taking the decisions about cost savings factor it is require doing benefits and risk assessments. Per the model in this paper; benefits and risk analysis assessment was done using a spreadsheet. In this research; risks are events that might occur during the cloud migrations. First, they are providing a spreadsheet includes benefits and risks that were initially identified [30].

These spreadsheet summaries the organizational, authorized, security, technical and financial profits and risks of using IaaS clouds from an enterprise perspective.

2.2.4. Prediction Models for Adaptive Resource Provisioning in the Cloud

Nowadays, due to advances in virtualization technologies, networking technologies, web services, etc., cloud computing has been a promising paradigm for delivering computing services [17]. Cloud hosting providers offers an vast amount of resources (e.g. CPU, memory, network I/O, disk, etc.) on-demand using pricing models such as pay-per-use model, reserved instance pricing model, spot instance model, etc. [31]. Therefore, application owners are able to select the most appropriate pricing models, based on the workload patterns and resource usage patterns [32].

The pay-as-you-go model and dynamic resource provision model of the cloud reduce the overhead associated with the static provisioning [33]. Because the static provisioning model is not deliberated as a cost-effective option as it can be led to over-provisioning and under-provisioning at any particular time [31]. In order to eliminate this resource provisioning issue, there are many resources prediction models that are predicting the right amount of resources based on ML technologies. [34].

The Future Generation Computer Systems has been published in the journal, which is named “Empirical prediction models for adaptive resource provisioning in the cloud”. This journal describes prediction-based resource measurement and provisioning strategies to satisfy upcoming resource demands [31]. For the prediction approach, they have used ML algorithms such as Neural Network, Linear Regression, and sliding window technique. They have trained and tested the prediction model, using generated data from running the TPC-W benchmark in the Amazon EC2 cloud. The effectiveness of the framework has been validated by using evaluation metrics like Mean Absolute Percentage Error (MAPE), PRED, etc.

Department of Systems and Computer Engineering of the University of Carleton has been published a research paper “Towards an Autonomic Auto-Scaling Prediction System for Cloud Resource Provisioning” which investigates the correctness of predictive auto-scaling systems in the IaaS layers of cloud computing [34]. They have made a hypothesis that the prediction correctness of auto-scaling systems can be improved by an appropriate time-series prediction algorithm which is based on the performance pattern over time. To prove this hypothesis, an experiment has been conducted based on performance metrics, Support Vector Machine (SVM) and Neural Networks (NN). In addition, they have used Amazon EC2 as the experiment infrastructure and TPC-W as the benchmark to generate various workload patterns.

The results of the experiments show that SVM has better prediction accuracy with periodic and growing workload patterns. Based on the results of the experiment, architecture has been proposed which supports to select the most appropriate prediction technique based on the performance patterns which gives the most precise prediction results.

2.3. Overview and Comparison of Cloud Migrations Tools

Most businesses are moving to the cloud-based platform due to its favorable features compare with the local data center technologies. Basically, the cloud platform helps to reduce cost and increase performance. Because of the benefits of the cloud, many businesses have been migrating their applications and data to the cloud. In this regards, it is important to have a migration tool that ensures effective migrations [35].

Cloud migration tools are software tools that accelerate cloud migration efforts by automating certain processes. Below stated three major cloud migration tools and services that are available today.

- AWS Application Discovery Tools
- Azure Data Migration Tools
- IBM Cloud Migration Tools

2.3.1. AWS Application Discovery Tools for Cloud Migration

AWS Application Discovery Service supports for the migration process of on-premises servers' to the AWS cloud environment. This tool is collecting usage and configuration data about on-premises servers. The simplification of the migration tracking is performing by integrating with the AWS Migration Hub. The performing discovery results can be viewed in the discovered servers, by grouping them into applications and then tracking the migration status of each application from the Migration Hub console. The discovered data can be exported for analysis in Microsoft Excel or AWS analysis tools such as Amazon Athena and Amazon Quick Sight [13].

2.3.2. Azure Data Migration Tools

Azure data migration tool helps to migrate from on-premises to Azure cloud. This tool provides a centralized hub to capture the custom assessment before actual migrations to the cloud that evaluate workload and other details. Based on these analyzed details, it also suggests business and technical plan including remodeling, right-sizing, and re-architecting by using advanced procedures and rules. Further it supports hybrid cloud migrations and it has the flexibility and faster execution [36].

2.3.3. IBM Cloud Migration Tools

IBM cloud migration tool offers standard and modularized services that help for low cost, repeatability, predictability, and high customer satisfaction. They provide multi-cloud integration with IBM Cloud, AWS, Azure, GCP, IBM API. Further it enables customers to initiate from very small and expand into more complex and higher value migration services. And also they set up disaster recovery, back-up and others to ensure BAU functionalities in case of failure [37].

Table 2.1 is represented the features and limitations of the above described cloud migrations tools and Table 2.2 is figured out the comparison of cloud migration tools.

Table 2.1: Features and Limitation of Cloud Migrations Tools

Migrations Tool Name	Features	Limitations
AWS Application Discovery [13]	Discover on-premises infrastructure	Agent tool needs to be installed in the discovered machine
	Identify server dependencies	Support only for AWS platform
	Measure server performance	
	Data Exploration in Amazon Athena	
Azure Data Migration Tool [36]	Azure readiness	Support only for Azure platform
	Azure sizing	
	Azure cost estimation	
	Dependency visualization	
IBM Cloud Migration Tool [37]	Migrate existing workloads and apps to any cloud	Agent tool needs to be installed in the discovered machine
	Modernize traditional applications using containers, microservices and API enablement	
	Multicloud and infrastructure management and data modernization	
	Built-in security from the start	

Table 2.2: Comparison of Cloud Migrations Tools, Source [35]

Tool Name	Vendor Name	Nature	Cloud type support (public, private, hybrid)	Cloud platform support	Support Migration of		
					Data	Application	Process
AWS Application Discovery [13]	Amazon	Dynamic	All	AWS	Yes	Yes	Yes
Azure Data Migration Tool [36]	Microsoft	Dynamic	All	Azure	Yes	Yes	Yes
IBM Cloud Migration Tool [37]	IBM	Dynamic	All	IBM Cloud, AWS, Azure, GCP	Yes	Yes	Yes

2.4. Conclusions Based on Literature Review

DevOps concepts are associated with process and process improvement through the automation while cloud computing provides a service-driven business model where hardware and platform level resources are provided as services on-demand basis. The relationship between DevOps and cloud is not easily reconciled. But there are many ways that DevOps and cloud links with.

The centralized nature of the cloud infrastructure provides DevOps automation with a standard platform for testing, deployment, and production. As cloud providers offer infinite resources on-demand basis resource provisioning for a cloud environment is the double-edged sword, which may lead to either over-provisioning or under-provisioning. There are many resource provisioning models and frameworks based on ML technologies like Neural Networks, Linear Regression, (Autoregressive Integrated Moving Average) ARIMA, etc. Other than these models, there can be found many vendor-specific cloud provisioning tools as an example Amazon Application Discovery tool. Even though these tools are supported for resource provisioning, there are many limitations can be found such as vendor-specific, usage cost, domain-specific, etc.

3. METHODOLOGY

3.1. Chapter Overview

The methodology followed in the research for the resource provision prediction has been described in this section. This approach has been selected based on the understanding gained through the literature review done and the DevOps experience in the software development industry.

The selected approach for the methodology consists of four main phases, namely gathering the historical performance metric values of individual workload, manipulating the historical performance metric values, analyzing the gathered workload data traces, applied the time series ARIMA model to predict the resource provisioning, presenting the resource provisioning outcome in the dashboard and implement the identified model using Amazon EC2 as the experimental infrastructure.

3.2. Overview of the Proposed Approach

The proposed resource provision methodology is basically based on the historical performance metric values of individual workload time series. Represented in Figure 3.1 is the flow chart of the methodology for predicted resource provisioning.

The historical data traces are gathered from a BI production environment which has been hosted in the on-premises data centers. Represented in Figure 3.2 is the high-level architecture of the BI system, which consists of three servers namely SSIS Server, Cube Server, and Reporting Server.

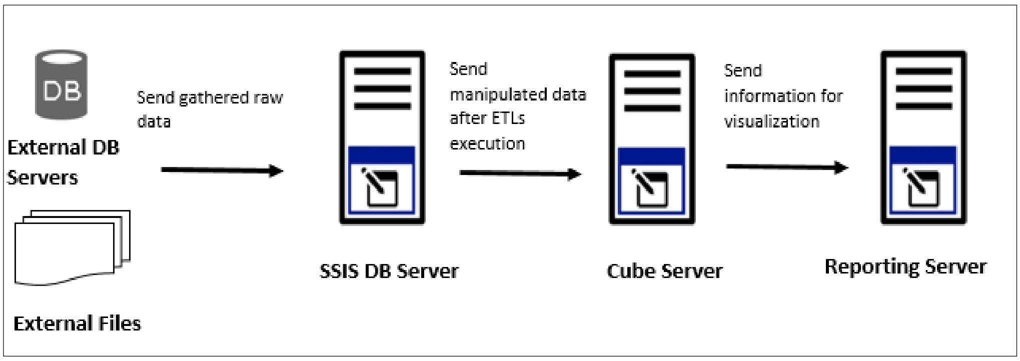

Figure 3.1: High-level Server Architecture of the BI System

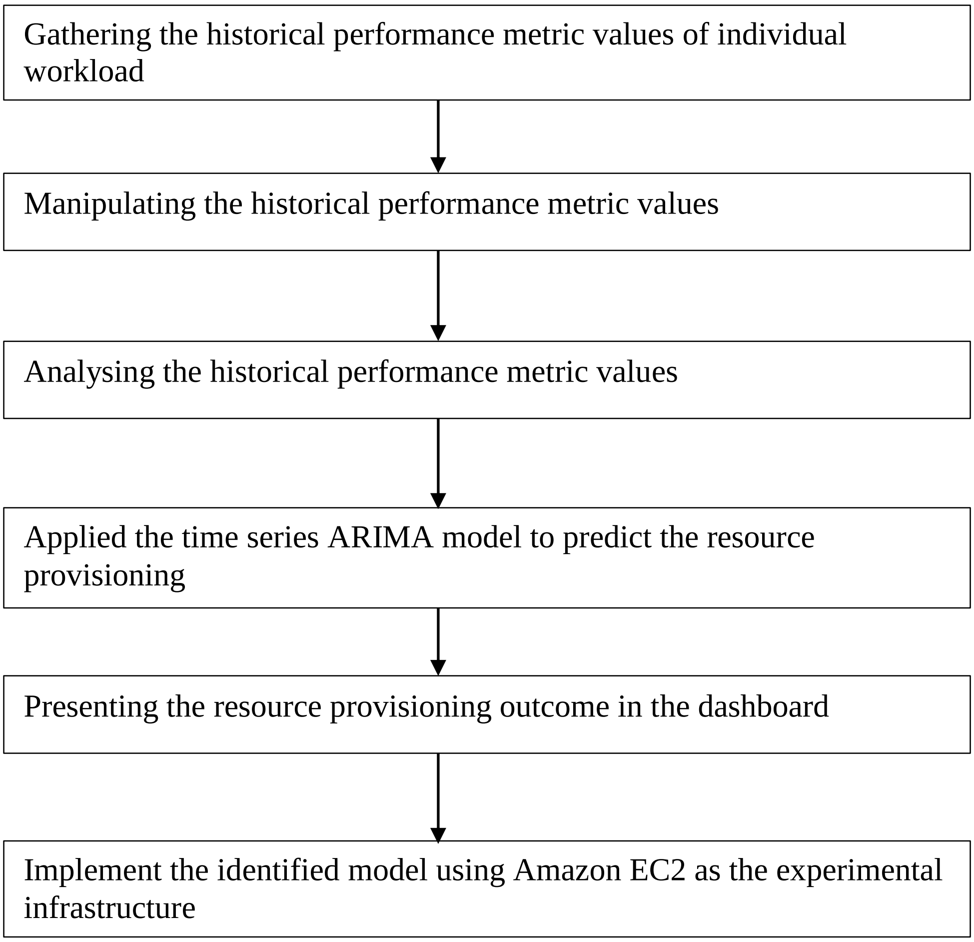

Figure 3.2: Flow Chart of the Methodology for Resource Provisioning

A performance monitoring software agent would be deployed on every server mentioned above. This software agent will periodically collect performance data, including CPU, RAM, Disk and Network utilization. These utilizations data traces are stored and would be gathered for resource provision prediction.

Gathered historical data traces would be transformed into the readable, useful data traces, at the end of the imposing of data manipulation techniques such as cleansing, mapping, splitting, joining, etc. These manipulated historical data would be analyzed based on ARIMA (Autoregressive Integrated Moving Average) workload prediction model.

The ARIMA model is a time series analysis model that offers a broad spectrum of methods to calculate workload forecast based on the historical performance metric values of workload data traces [38].

Established along with the predicted resource provisioning outcomes, it would be decided the requirement of future resource of application in conditions of resource allocation. These predicted resource provision outcomes would be presented in a dashboard such that, users are able to take a glance awareness of resource provision, even though they cause not much technical knowledge.

In the approached process mentioned above the following steps have been covered.

- Gathering the historical performance metric values of individual workload time series.
- Manipulating the individual workload time series.
- Analyzing the gathered workload data traces.
- Applied the time series ARIMA model to predict the resource provisioning.
- Presenting the resource provisioning outcome in the dashboard.
- Implement the identified model using Amazon EC2 as the experimental infrastructure.

3.3. Gathering the Historical Performance Metric Values of Individual Workload Time Series

There would be two different characters of data that could be gathered from a server, namely, dynamic data and static data. Dynamic data are designed to facilitate change of data in the run time whereas static data are fixed and cannot be changed over time [39]. Resource utilization data such as RAM (Random Access Memory), CPU (Central Processing Unit), Network and Disk are considered as dynamic data, whereas server specification (Hardware and software specification) details and running application specification details (software vendor, minimum hardware and software requirement, prerequisites) details are considered as static data.

The proposed resource provision model depends on the data traces of individual workload time-series data, which is fallen into the category of dynamic data. These statistics data traces would be considered to anticipate future resource demands. Hence the suggested approach is more interesting about the dynamic data.

The historical data traces of utilizations are gathered from the BI production environment which has been hosted in the on-premises data centers. These historical data traces would be gathered by deploying a performance monitoring software agent on every server in the BI system. As stated in section 3.2, three physical servers have been implemented on data centers in the BI production environment namely, SSIS Server, Cube Server, and Reporting Server. The performance monitoring software agent would be periodically collected the statistics of historical performance data traces in each server mentioned above.

These data traces consist of utilization statistics of RAM (Random Access Memory), CPU (Central Processing Unit), Network and Disk. The agents would be collected the statist's data traces for twenty-four hours at a minimum time granularity of 5 minutes. The data traces would be obtained a total of 30 days from the three servers with respect to the RAM, CPU, Network, and Disk.

The software agent would be collected the statistics data traces of performance, associated with resources of RAM, CPU, Network, and Disk with respect to each job execution in BI servers. It would be considered the data traces of performance for the different jobs' execution in each server. The considered modules are stated below.

2. Finance Module
3. Human Resource Module
4. Resource Management Module
5. Time Management Module
6. Projects Management Module
7. Job Portal
8. Logistic Module

At the end of each day, there would be obtained the four different data traces associated with the modules mentioned above.

3.4. Manipulating the Historical Performance Metric Values

The gathered data traces of individual workload time series need to be manipulated such that, it would be usable and readable in the analyzed phase. The data manipulation process would be applied to the data transformations process which would provide the data traces in the desired format. The data transformation process is included with data cleansing, mapping, splitting and statistical formations [40].

3.4.1. Data Cleansing

Data cleansing is the process of detecting and correcting corrupted and inaccurate records in data-trace in order to produce accurate and correct data set [41]. The data cleansing process involves removing the missing, extreme and anonymous data. At the end of the cleansing process, it would provide the quality improved data set. [42].

As stated in section 3.2, the performance agent software would be gathered the vast number of data traces with the vast number of redundant data from each server. These unwanted data should be removed and filtered in order to produce the data set which is in the desired format. Gathered historical data traces would be filtered based on the parameters such as timestamp, process ID, resource name, utilization value, etc. The data filtering methodology can be divided into two categories. In the first methodology, the required parameters can be configured, before collecting the data traces whereas in the second methodology, gathering all data traces and then filtering the traces based on the parameters mentioned above.

3.4.2. Data Mapping

The performance agent software would be collected a vast figure of data traces in each server, which is stored in an unstructured way. The data mapping process would be performed to produce the structured data traces which could be taken into consideration of resource provision aspects. Represents in Figure 3.3 are the process of data transformation which would be followed in generating the structured data set. Each data set for RAM, CPU, Network and Disk would be transformed separately into the .CSV files.

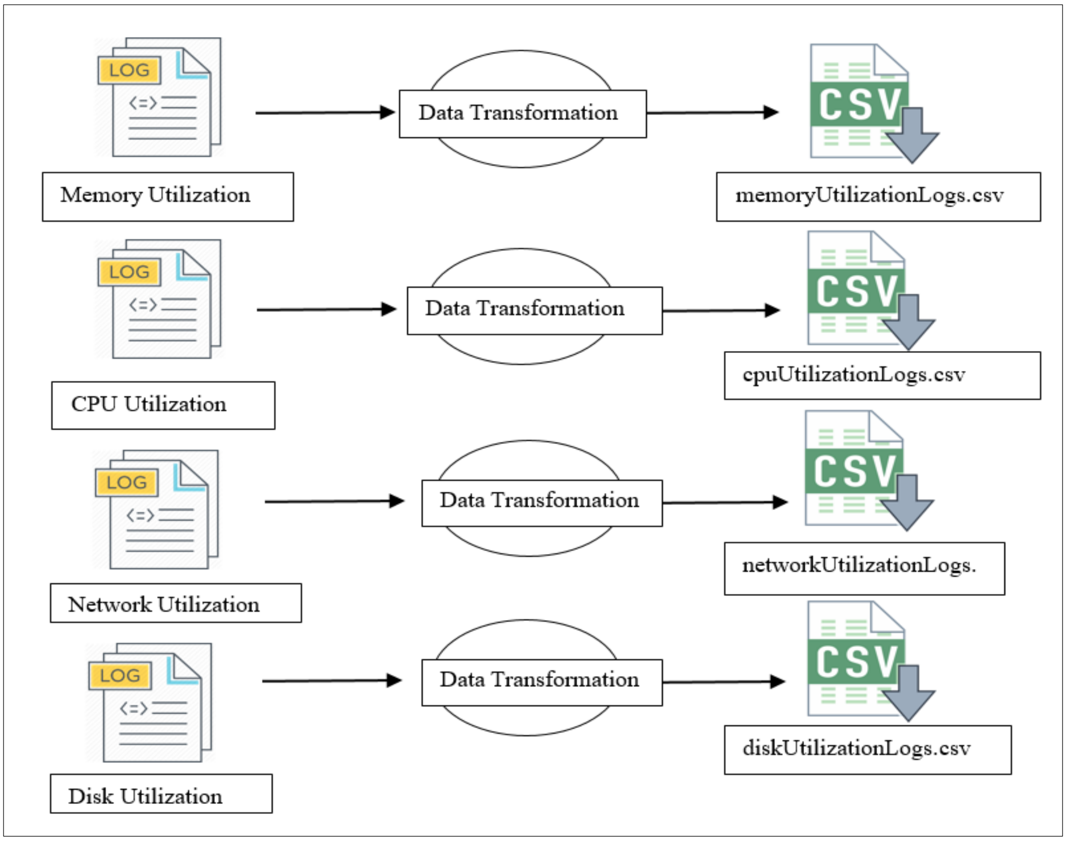

Figure 3.3: The Process of Data Transformation

The utilization data traces which would be collected by the software agent has consisted of the data traces which is associated with each resource and each process at the time of twenty-four hours. Those gathered a vast number of traces that need to be mapped with respect to each resource Id and each process Id. At the end of the mapping process, data traces would be stored in a.CSV file format, such that data analyzing phase could be considered these mapped data traces to the provision aspect.

3.5. Analyzing the Gathered Workload Data Traces

Time series data traces that were mapped in the previous section, would be analyzed furthermore to identify the optimal time series model that can be applied to the manipulated, data set.

The historical data traces were gathered during the execution of the Human Resource job in the SSIS server of BI on-premises data center. These utilization data traces include a total of thirty days of traces which have been compiled along each day. Figure 3.4 represented the line chart of the utilization data traces associated with CPU, RAM, Disk, and Network respectively. These obtained line charts explore a gradually increasing seasonality trend for each week. The seasonality suggests that the series is almost certainly non-stationary. As resource utilization has been figured out the seasonality effect of the data traces and trends, this could be considered for a time series model.

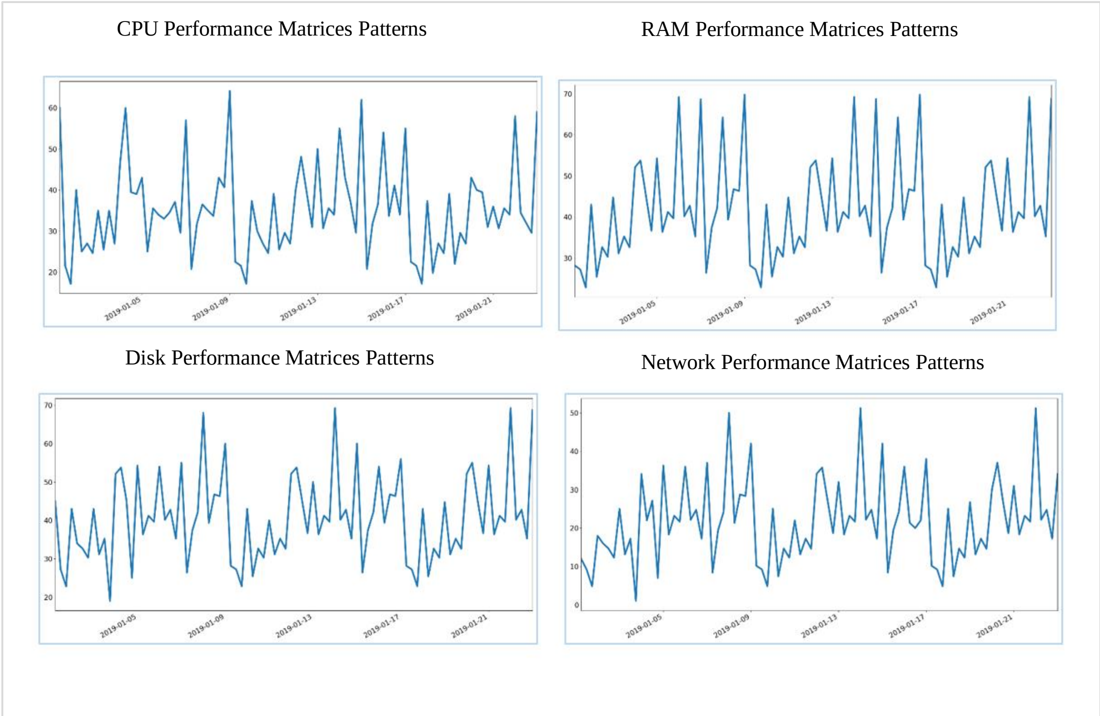

Figure 3.4: The Line Chart for the Resource Performance Metrics

The seasonality effect and the trend have been analyzed and described further based on a box plot chart produced. The box plot chart is the standardized way of representing the distribution of data traces based on the five-number summary.

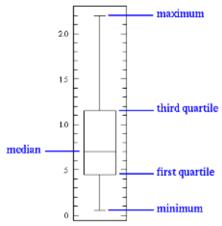

Figure 3.5: Five Numbers Summary of a Box Plot

Figure 3.5 is represented the five numbers of summary consists of a box plot, namely minimum, first quartile, median, third quartile and maximum [43].

Illustrated in Figure 3.6 is the box plot chart for the data traces distribution of the resource utilization. The box plot chart is figured out that, data traces are comprised of four seasonality, such that second and third seasons are exemplified the symmetric distribution whereas the first and fourth seasons are exemplified the asymmetric distribution. The data distribution of the box plot shows, the maximum resource utilization for the fourth season whereas the minimum resource utilization for the third season. This trend is applicable to all four resources. The results are finally demonstrated that data distribution has a seasonality effect and tend which can be considered for time series analysis.

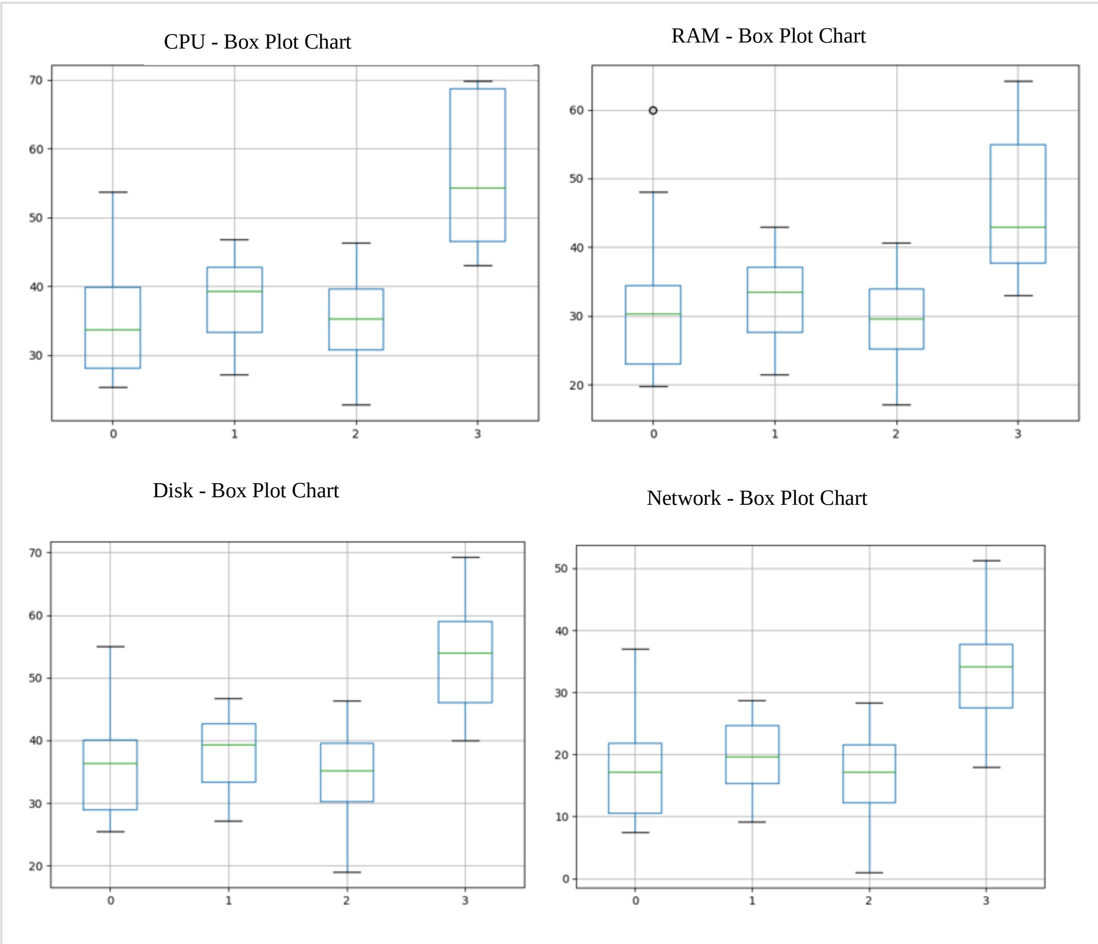

Figure 3.6: The Box Plot Chart for the Resource Utilization Data Traces

Depicted in Figure 3.7 is the variance in the seasonality effect that could be identified in the current data set. The data set is revealed that it has +20 or -20 seasonality differences that can be accepted as an in input for time series data modeling.

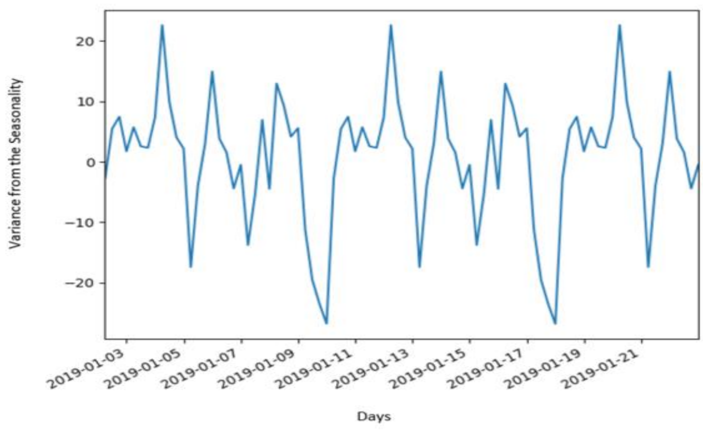

Figure 3.7: The seasonality differences of the current data set - CPU

The line chart, the box plot chart and the seasonality differences of the current data set chart are exemplified that, the data distribution of the collected traces have the seasonality effect and trend based on the job execution pattern of the server. Hence, time-series data traces of resource utilization that collected on each day, have been split into the four groups are represented in Table 3.1, namely midnight, morning, afternoon and evening.

Table 3.1: Season Identified in CPU Data Traces

Seasons	Considered Time Durations
Midnight (0)	00:00 – 05:59
Morning (1)	00:06 – 11:59
Afternoon (2)	00:12 – 17:59
Evening (3)	18:00 – 23:59

3.6. Applied the Time Series ARIMA Model to Predict the Resource Provisioning.

The historical performance metric values of individual workloads associated with the resource utilization, that gathered during the execution of HR jobs were manipulated and analyzed in the previous sections. At the end of the analyzed phase, it has been recognized that analyzed time-series data traces are revealed a seasonality effect and a trend at the distribution of data traces.

The essential concept of the proposed resource provision POC is to predict the upcoming resource utilization behavior, based on the gathered historical utilization data traces. The ARIMA model provides the approach to time series forecasting and provides a complementary approach to the requirement stated above.

The ARIMA model is cable of modeling non-stationary time series prediction that is collected of an autoregressive and a moving average model that would be applied to predict future utilization resources based on the historical utilization time series [44]. The ARIMA (p,d,q) model consists of three parameters that would be applied in the model should be configured manually.

p = order of the autoregressive part;

d = degree of first differencing involved;

q = order of the moving average part;

Represented in Figure 3.8 are the Autocorrelation Function (ACF) and Partial Autocorrelation Functions (PACF) plots that would assist exemplify the lag values for Auto-Regression (AR) and Moving Average (MA) parameters, p, and q respectively. The ACF has exemplified the lag values as a 3, whereas PACF is also exemplified the lag values as a 3. The reason for considering the lag value as 3 because the data distribution is demonstrated that there are three lags beyond the range in both autocorrelation and partial autocorrelation graph.

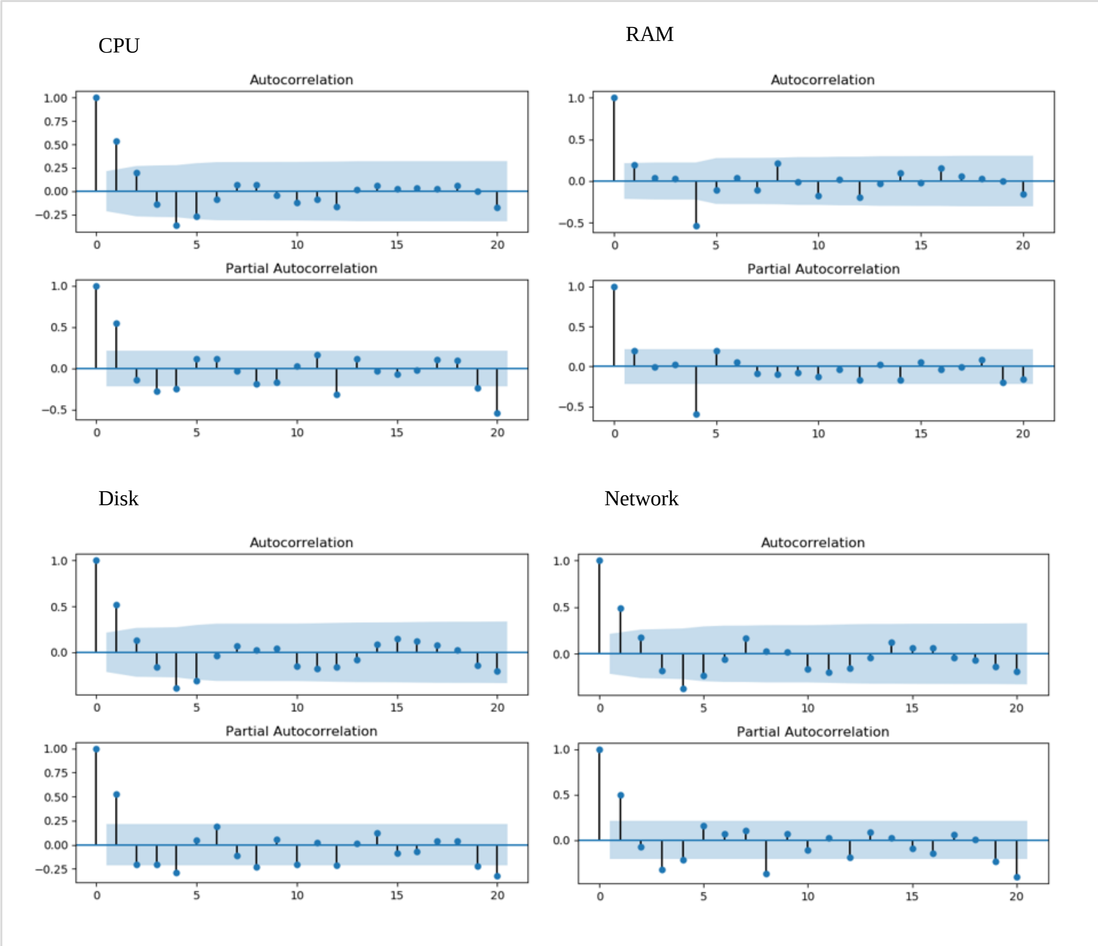

Figure 3.8: Autocorrelation and Partial Autocorrelation Plots

Table 3.2 is figured out the results of the validation of RMSE with respect to the different ARIMA parameters. Considerations of above RMSE results it has been shown that the best fitted ARIMA model parameter values are 3, 0 and 3. These best suited ARIMA models that identified using ACF plot, PACF plot and RMSE would be an input for the prediction resource provisioning approach.

Table 3.2 : RMSE Validations Results

ARIMA Values	RMSE Values
ARIMA(0, 0, 1)	RMSE=9.306
ARIMA(0, 0, 2)	RMSE=9.100
ARIMA(0, 0, 3)	RMSE=8.907
ARIMA(0, 0, 4)	RMSE=8.902
ARIMA(0, 1, 1)	RMSE=10.378
ARIMA(0, 1, 2)	RMSE=9.459
ARIMA(0, 1, 3)	RMSE=9.195
ARIMA(0, 1, 4)	RMSE=9.181
ARIMA(1, 0, 0)	RMSE=9.118
ARIMA(1, 0, 1)	RMSE=9.098
ARIMA(1, 0, 2)	RMSE=9.089
ARIMA(1, 1, 0)	RMSE=10.367
ARIMA(1, 1, 1)	RMSE=9.217
ARIMA(2, 0, 0)	RMSE=9.071
ARIMA(2, 0, 1)	RMSE=8.497
ARIMA(2, 0, 2)	RMSE=8.421
ARIMA(2, 0, 4)	RMSE=8.822
ARIMA(2, 1, 0)	RMSE=10.429
ARIMA(2, 1, 1)	RMSE=9.455
ARIMA(3, 0, 0)	RMSE=8.783
ARIMA(3, 0, 1)	RMSE=8.756
ARIMA(3, 0, 2)	RMSE=8.291
ARIMA(3, 0, 3)	RMSE=8.258
ARIMA(3, 0, 4)	RMSE=8.811
ARIMA(3, 1, 0)	RMSE=10.400
ARIMA(3, 1, 1)	RMSE=8.862
ARIMA(3, 1, 2)	RMSE=8.954
Best ARIMA(3, 0, 3)	RMSE=8.258

The gathered historical data traces of CPU utilization which are a total of 30 days would be used for train and predict the future resource requirement. Represented in Table 3.3 is the predicted values against the expected values. The RMSE value of the results is 9.351.

The results figured out in Table 3.3 can be visualized further as represented in Figure 3.9. The graph has been visualized the expected results based on the historical resource utilization data traces against the predicted results for resource utilization. Represented in red line is the predicted results whereas in blue line is the expected results. The line chart of the predicted results shows a closed relation to the expected results and it would be further evaluated in Chapter five using RMSE, MAPE, precision, recall, and F1 score.

Table 3.3 : Predicted values vs. Expected values

Predicated Values	Expected Values
Predicted=41.975,	Expected= 26
Predicted=36.486,	Expected= 37
Predicted=36.667,	Expected= 42
Predicted=72.333,	Expected= 64
Predicted=23.597,	Expected= 39
Predicted=45.631,	Expected= 47
Predicted=43.689,	Expected= 46
Predicted=66.477,	Expected= 70
Predicted=41.906,	Expected= 28
Predicted=41.406,	Expected= 27
Predicted=40.661,	Expected= 23
Predicted=62.166,	Expected= 43
Predicted=19.593,	Expected= 25
Predicted=31.126,	Expected= 33
Predicted=24.608,	Expected= 30
Predicted=46.881,	Expected= 45
Predicted=25.366,	Expected= 31
Predicted=36.552,	Expected= 35
Predicted=30.646,	Expected= 33
Predicted=46.732,	Expected= 52
Predicted=34.836,	Expected= 54
Predicted=45.788,	Expected= 45
Predicted=33.300,	Expected= 37
Predicted=54.797,	Expected= 54
Predicted=54.516,	Expected= 36
Predicted=36.960,	Expected= 41
Predicted=39.749,	Expected= 40
Predicted=55.253,	Expected= 69
Predicted=44.274,	Expected= 40
Predicted=40.235,	Expected= 43
Predicted=41.893,	Expected= 35
Predicted=67.005,	Expected= 69
Predicted=42.000,	Expected= 26
Predicted=36.319,	Expected= 37
Predicted=36.738,	Expected= 42
Predicted=72.324,	Expected= 64

The ARIMA model that has been applied in previously, is for the historical data traces of resource utilization which have been gathered during the HR jobs' execution in the server.

The POC of the resource prediction model would be illustrated the prediction of resource utilization for simplicity of exposition. However, the best fitted ARIMA model that has been proved in previously, similarly applied for another resource provisioning namely RAM, Network and Disk.

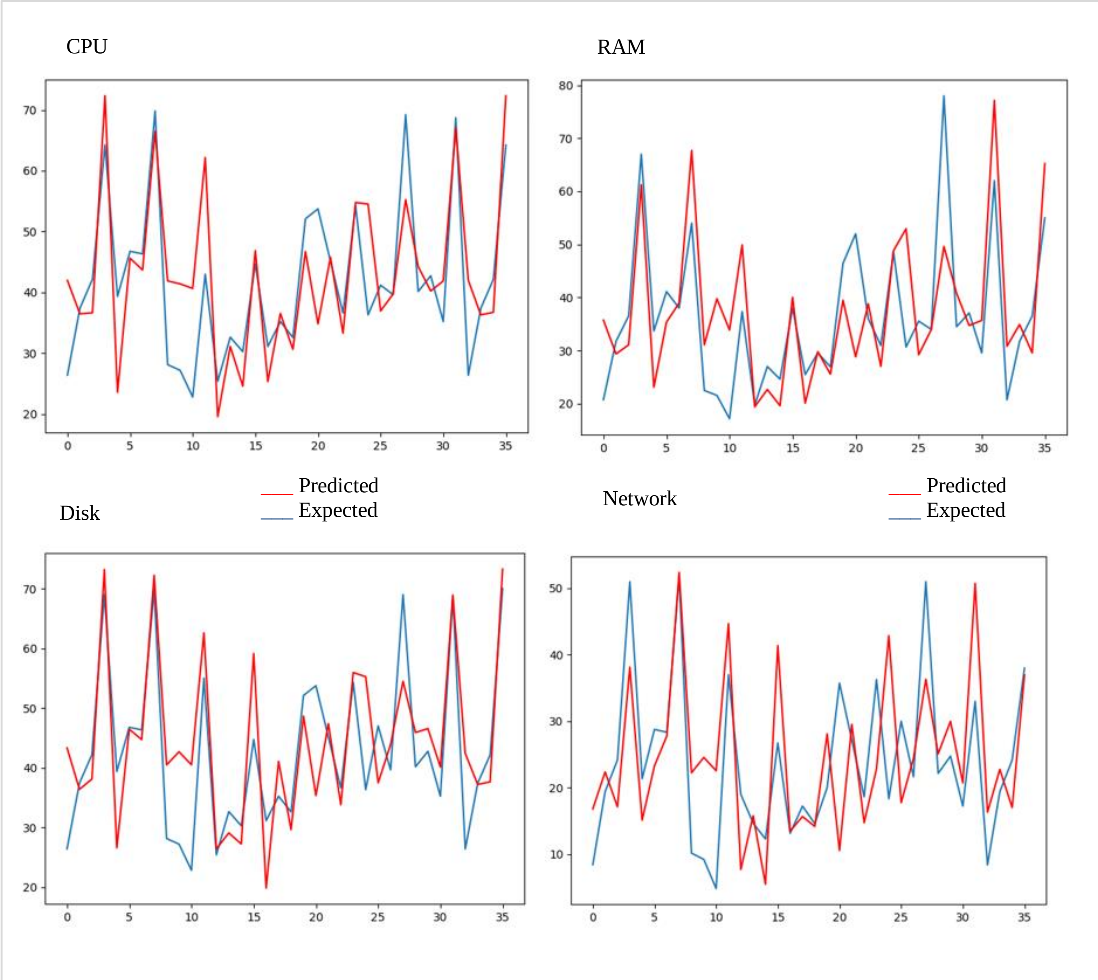

Figure 3.9: Predicted values vs. Expected values

3.7. Presenting the Resource Provisioning Outcome in the Dashboard

In the resource provisioning prediction section, multiple predictions would be made based on the historical data traces. These resource predictions results are included the future resource requirements that would be served to the future application requirement.

The predicted resource provision results are intended to filter out and visualize in a graphical dashboard in a way that would be convenient for users to be read and understand the future infrastructure requirement, regardless of users' technical knowledge. Furthermore, the management decisions such as cost and cloud migration would be made based on the visualized and filtered details in the dashboard which were resulting in the resource prediction.

3.8. Implement the Identified Model Using Amazon EC2 as the Experimental Infrastructure

As stated in section 3.6, the forested future values that have been predicted based on the historical performance metrics should be evaluated in the Infrastructure as a Service (IaaS) in the layer of the cloud platform. As the experimental infrastructure, and Elastic Cloud Compute (EC2) in the Amazon Web Services would be used.

The EC2 is web services that provide secure and scalable compute capacity in the cloud environment and is designed to facilitate the different web services for the IT industry. The web services in Amazon EC2 allow us to obtain and configure capacity with a minimum effort and provides whole control of computing resources. Amazon EC2 decreases the time required to get a new server to minutes and it allows to speedily scale capacity, both up and down as per the computing requirements change [45].

3.9. Summary

The chapter provides a six-step approach to how resource provisioning prediction would be made through the research. The proposed approach would initiate with the gathering of the historical performance metric values of individual workload time series. These values would be manipulated and analyzed further. Analyzed values would be an input for a time series forecast model. Then forecasted values would be presented in a dashboard. Finally forecasted future resource requirements would be evaluated in Amazon EC2Cloud.

4. IMPLEMENTATION

4.1. Chapter Overview

The cloud resource over-provisioning or under-provisioning are results of, correspondingly saturations or a waste of resources that would be the most significant challenge, cloud clients are faced with. Use an auto-scaling approach would be one solution to overcoming the challenges, mentioned above. The current auto-scaling approach can be classified as reactive, proactive, and predictive [34]. Out of this classified approach, the predictive approach would solve the resource saturations or wastage by predicting the future resource requirements based on the historical performance metrics of the existing infrastructure.

In this chapter, the tools and techniques used in providing a solution to predict the forecast resource provision and evaluation of forecast values in Amazon EC2 are presented. The implementation processes that break down into the steps are given below.

- Gathering Historical Performance Metric Values
- Resource Provision Prediction
- Visualized the Forecasted Resource Provisioning in a Dashboard
- Implement the Forecasted Resource Provisioning in the Amazon EC2 Platform

4.2. Gathering Historical Performance Metric Values

The prediction of the resource provision model for future infrastructure requirements highly depends on the gathered historical performance metrics (such as the number of utilization per time unit).

These historical data metrics would be gathered by implementing a performance monitoring software agent on every server in the BI system. The BI system that has been hosted in the on-premises data centers consists of three physical servers namely, SSIS Server, Cube Server, and Reporting Server. The performance monitoring

software agent would be periodically collected the statistics of historical performance data traces in each server mentioned above.
The software agent would be collected the performance metrics at a minimum time granularity of 01 minutes, a total of 30 days. The performance metrics associated with resources stored in the log files would be updated in every minute. At the end of the day, there would be four log files with respect to each resource (CPU, RAM, Network, and Disk) for individual job execution. e.g.: cpu_log_01012019.txt.

The software agent would collect 30 performance matric logs from each resource for 30 days, which would be an input for the implementation of the time series analyzing model which would be described further in section 4.3.

4.3. Resource Provision Prediction

The prediction accuracy of the predictive resource provisioning model can be increased by choosing an appropriate time-series prediction algorithm. The proposed approach has been chosen the ARIMA model as a prediction algorithm which is the most dominant prediction technique in the cloud auto-scaling area for time-series analysis.

The ARIMA is a compound of moving average and auto-regression algorithm that would be built as a model, based on the predefined training dataset. The created model would be evaluated against the testing data set. It would be used, 75% percentage of gathered historical performance metrics as the training data set whereas the 25% percentage of historical metrics as the testing data set. The accuracy of the results would be evaluated based on the RMSE.

The POC of the resource provisioning prediction has been implemented using Python, which is a high-performance scientific application. Depicted in Figure 4.1, is the versions and names of used external libraries, typically written in faster languages (like C, Fortran for matrix operations) [46]. Different usage of the Python libraries is represented in Table 4.1.

Table 4.1: Usage of Python libraries, Source [47]

Python Library Name	Usage
SciPy (1.3.0)	Linear algebra, Fourier transformation, and Random number capabilities.
NumPy (1.16.4)	Linear algebra, Integrations, and Statistics
Matplotlib (3.1.1)	Data visualization (Control line styles, font properties, formatting axes, etc.)
Pandas (0.24.2)	Data extraction and Preparation
Sklearn (0.21.2)	Classification, Regression and Clustering algorithms
Statsmodels (0.10.1)	Statistical model, Statically data explorations

A series of experiments have been conducted to verify the proposed workload prediction model by using the different performance metrics data sets.

```
Alogorithm 1: Data Validation
Input: Perfomance Metrics Data Set
Libraries: Pandas, Sklearn, Math
Results: RMSE Validation
function validation()

from pandas import Series
from sklearn.metrics import mean_squared_error
from math import sqrt
# load data
series = Series.from_csv('dataset.csv')
# prepare data
results = series.values
results = results.astype('float32')
train_size = int(len(results) * 0.50)
train, test = results[0:train_size], results[train_size:]
# walk-forward validation
history = [results for results in train]
predictions = list()
for i in range(len(test)):
      # predict
        yhat = history[-1]
        predictions.append(yhat)
        # observation
        observation_values = test[i]
        history.append(observation_values)
        print('>Predicted=%.3f, Expected=%3.f' % (yhat,
observation_values))
# report performance
mse = mean_squared_error(test, predictions)
rmse = sqrt(mse)
return RMSE
```

Figure 4.1: Data set Validation Code Snipped

The experiments are based on the 30 days of CPU, Network, RAM and Disk performance metrics data sets that were collected from the SSIS server at the execution of the HR module. Out of the 30 data sets of performance metrics, 22 data sets were considered as the training data set and 08 data sets were considered as the test data set. Represented in Figure 4.2 is the code snippet of the data validation process.

The accuracy of the model is evaluated based on the analyzing of RMSE values against different ARIMA models. There were used different ARIMA models, with different p, d, and q parameter values.

In conclusion, the least RMSE value that has been testified was 8.258. The respect ARIMA model for the least RMSE value was selected as the best ARIMA which was ARIMA (3, 0, 3) would be considered as the best fitted ARIMA model for the perdition of the collected CPU performance metrics.

Based on the finalized best suited ARIMA model, the forecast resource provisioning POC was proceeded using collected performance data sets. Represented in Figure 4.2 is the code snipped of prediction class.

```
Alogorithm 2: Data Prediction
Input: Perfomance Metrics Data Set
Results: Data Prediction
function prediction:

series = Series.from_csv('dataset.csv')
months_in_year = 4
model_fit = ARIMAResults.load('model.pkl')
bias = numpy.load('model_bias.npy')
yhat = float(model_fit.forecast()[0])
yhat = bias + self.inverse_difference(series.values, yhat,
months_in_year)
Return Prediction

# invert differenced value
def inverse_difference(self,history, yhat, interval=1):
return yhat + history[-interval]
```

Figure 4.2: Prediction Class Code Snipped

4.4. Visualized the Forecasted Resource Provisioning in a Dashboard

As stated in section 4.3, historical performance metrics of utilization would be forecasted based on the best suited ARIMA model. The forecasted future resource requirements of CPU, Network, RAM and Disks that associated with different modules such as Human Resources, Finance, Time, Resource Management, etc., would be graphically visualized in Amazon QuickSight. The Amazon QuickSight is the business analytics services that are provided by Amazon Web Services.

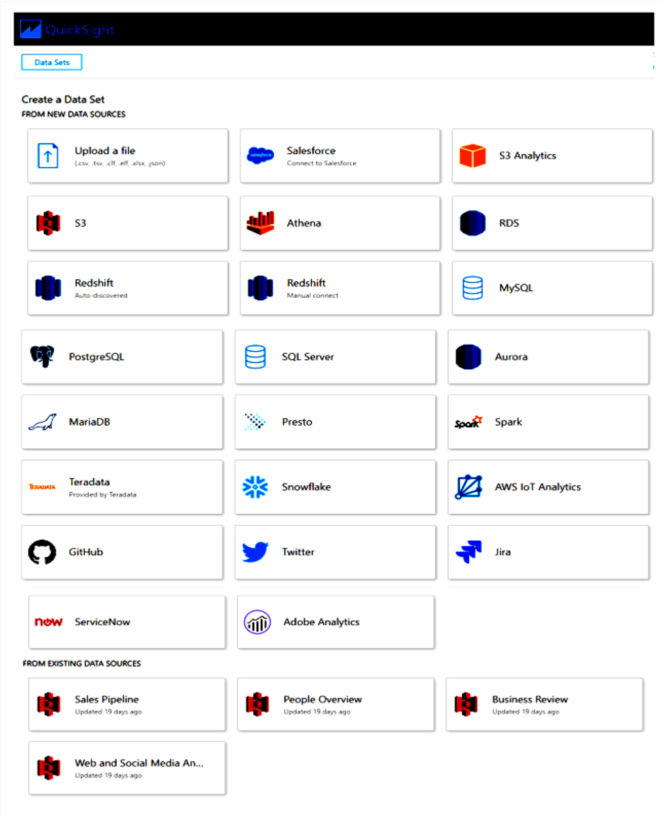

Figure 4.3: Data Uploading Interface in Quick Sight

The QuickSight can be used to build the visualizations, perform ad hoc analysis and get the business insights from data [48]. The QuickSight can be integrated with multiple applications such as SQL Server, MySQL, GitHub, Amazon S3, etc.

In additionally, it has been facilitated to connect with the multiple data sources like .csv, .tsv, .clf, .elf, .xlsx, .json. Represented in Figure 4.4 is the data upload interface of the QuickSight that can be used to upload the different data sets for visualizing.

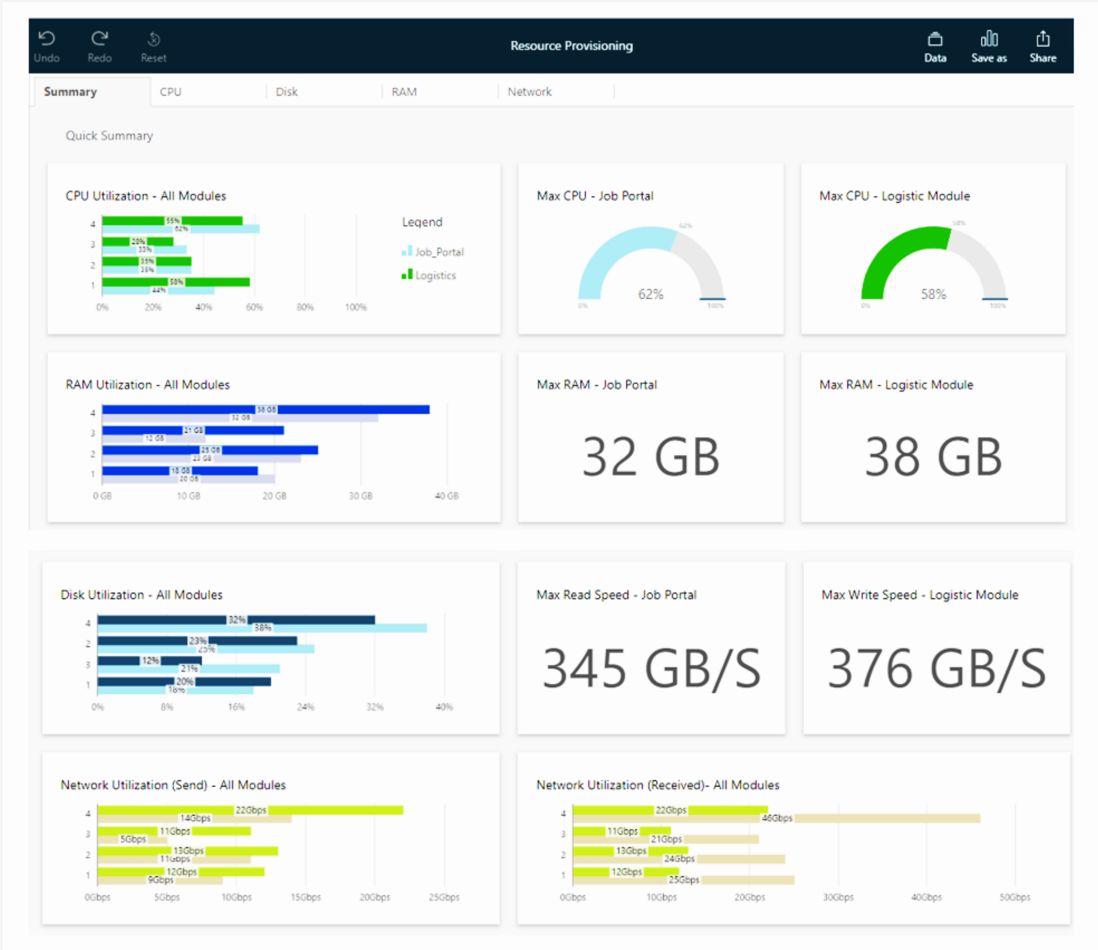

Figure 4.4: Resource Provisioning Dashboard Summary

Future resource requirements of the resources that have been predicted by the ARIMA model would be uploaded in the QuickSight. These data sets consist of the future forecasted resources values with respect to the different job executions' in the BI system. These values would be graphically visualized using different visual types such as pie chart, bar chart, line chart, pivot table, etc.

The collection of different visual types would be used to create the dashboard that would be figured out the future resources' requirements of the application at the glance.

Represented in Figure 4.4 is the resource provisioning dashboard that would be used by a company to get the high-level view of the forecasted resource provisioning without concerning about their technical knowledge.

4.5. Implement the Forecasted Resource Provisioning in the Amazon EC2 Platform

Based on the provided outcomes of the ARIMA model, the future forecasted resource requirement of the CPU would be evaluated in the Infrastructure as a Service (IaaS) layer of the cloud platform. As the experimental cloud infrastructure, and Elastic Cloud Compute (EC2) in the Amazon Web Services would be used. The EC2 is web services that provide secure and scalable compute capacity in the cloud environment and is designed to facilitate the different web services for the IT industry. The EC2 allows booting an Amazon Machine Image (AMI) to create the virtual machine.

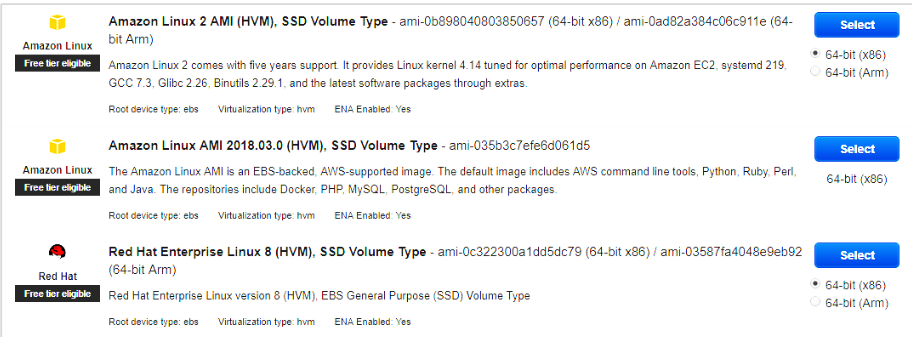

Figure 4.5: Amazon AMI Interfaces

The AMI is a pre-configured template that consists of software configurations including operating systems, and other supportive application software. Represented in Figure 4.5 is the AMI interfaces that are provided by AWS. The virtual machine that would be created using AMI, is referred to as an instance. These instances could be created, launched, stopped and terminated at any given time based on the users' requirement. Future forecasted resource requirements of CPU would be evaluated in an Amazon Linux 2 AMI (HVM) server with the parameters represented in Figure 4.6.

Amazon Linux 2 AMI (HVM), SSD Volume Type - ami-0b898040803850657 (64-bit x86) / ami-0ad82a384c06c911e (64-bit Arm)

Amazon Linux 2 comes with five years support. It provides Linux kernel 4.14 tuned for optimal performance on Amazon EC2, systemd 219, GCC 7.3, Glibc 2.26, Binutils 2.29.1, and the latest software packages through extras.

Figure 4.6: Virtual Machine Parameters

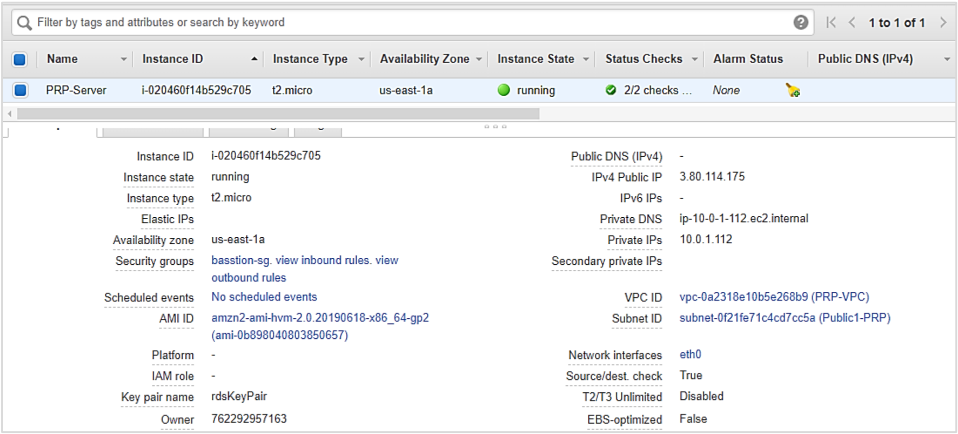

Figure 4.7: PRP-Server Details

The created instance would be referred to as a PRP-server which is depicted in Figure 4.7. This instance has been created with the minimum specifications required within the Amazon free tier. This created PRP-Server would be connected with a database instance is depicted in Figure 4.8.

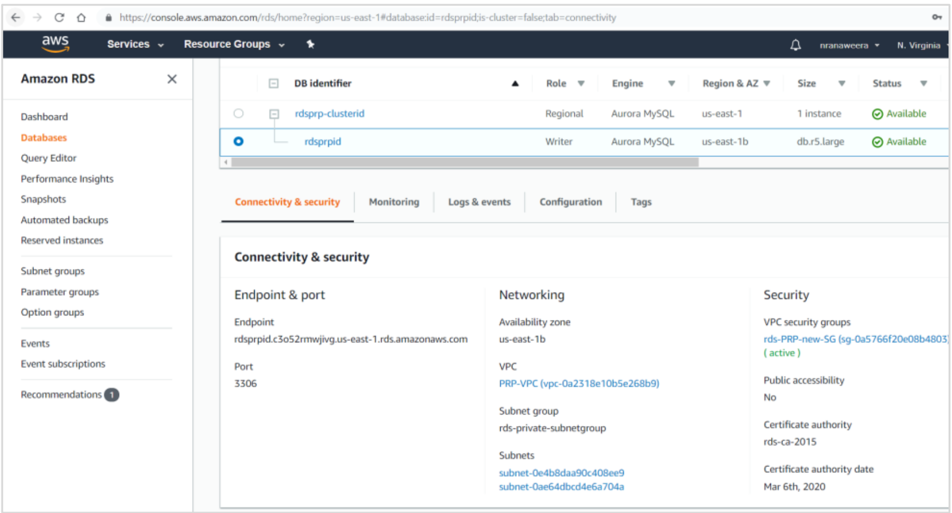

Figure 4.8: DB Instance Details

The database instance would be deployed with An Amazon Aurora with the flowing parameters that were predicted by the ARIMA model.

- Vesrion – Aurora (MySQL)-5.6.10a
- 2 vCPUs 16 GiB RAM EBS: 3500 Mbps

For the simplicity of the exposition and by considering the limitation of the services in the Amazon free tier, the evaluation process has been delimited to a selected module in the BI system.

4.6. Summary

Summarizing the steps mentioned in this chapter, first, a gathering process of historical performance metric values as described. Then the implementation of resource provisioning ARIMA model has been discussed. Next, based on the predicted values, data visualization in the dashboard has been illustrated. Finally forecasted resource provision has been evaluated on the Amazon EC2 cloud platform.

5. EVALUATION

5.1. Chapter Overview

The evaluation of the POC was carried out under several evaluation steps. The basic evaluation has been conducted based on the migrated module in the AWS. The resource (CPU, RAM, Disk, and Network) performance metrics of the migrated modules would be used for this evaluation. Another evaluation has been conducted based on the presented dashboard. These two evaluation steps have been discussed below.

5.2. Evaluate the Resource Provisioning

The key contribution of this POC is the Resource Provisioning Prediction for cloud migrations. As stated in the previous chapters, the future resource requirements of the cloud platform, have been predicted using the ARIMA model. These predictions have been made based on the gathered performance metrics from the on-premises servers.

A chosen module of the BI system was migrated to the EC2 instance in the IaaS layer of Amazon to evaluate the predicted results. The resource provisioning of the virtual machine (EC2 instance) has been made, based on the predicted results in the ARIMA model. The configurations parameters of the EC2 instance are mentioned below.

- Vesrion – Aurora (MySQL)-5.6.10a
- 2 vCPUs 16 GiB RAM EBS: 3500 Mbps
- Amazon Linux 2 AMI (HVM)

The performance metrics of the EC2 instance were collected in 10 days. These actual values and the predicted values had been evaluated based on the five different metrics namely Mean Percentage Error (MAPE) and the Root Mean Square Error (RMSE), Precision, Recall and F1 score.

5.2.1. MAPE

The MAPE is the mean or average of the absolute percentage errors of forecasting, that provides the error in term of percentages. The equation and the variable descriptions of MAPE the MAPE are stated below [49].

$$MAPE = \frac{\Sigma \frac{|A - F|}{A} \times 100}{N}$$

Equation 1: MAPE

A = Actual value

F = Forecasted value

N = Number of observation

5.2.2. Accuracy Percentage

The accuracy percentage of the predicted values can be obtained by subtracting the MAPE value from 1 and multiply by 100 as stated in the equation below.

$$\text{Accuracy Percentage} = (1 - \text{MPAE}) * 100\ \%$$

Equation 2: Accuracy Percentage

5.2.3. RMSE

The RMSE is the standard deviation of the residuals that measure of how to spread out these residuals are. Below shows the formula for the calculation of RMSE with parameter names [50].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (p_i - a_i)^2}{n}}$$

Equation 3: RMSE

p = Predicted value

a = Actual value

n= Number of observation

Observed performance metrics of the CPU, collected from the EC2 instance in the cloud were evaluated based on the performance evaluation formulas that were described in previous sections. These performance metrics consist of 09 data sets with respect to different modules, that collected during the 24 hours are evaluated. Evaluations results against the different modules were described below.

5.2.4. Precision

The precision represented the ratio of the number of correct positive predictions to the number of positive predictions. The high precision value is illustrated that the model has high accuracy with low false positives. The precision values have been evaluated based on the outcomes of the confusion matrix [51].

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Equation 4: Precision

5.2.5. Confusion Matrix

The confusion matrix is used to recognize the number of true positives, true negatives, false positives, and false negatives. Represented in Table 5.1 is the format of the confusion matrix.

Table 5.1: Confusion Matrix - Outline

	Predicted Value		
Actual Value		Negative	Positive
	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

5.2.6. Recall

The performance matrix of the recall is an essential measurement in the evaluation section. It calculates the ratio of the number of correct positive predictions to the overall number of positive predictions. The high accuracy model has a high value for the recall measurement.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Equation 5: Recall

5.2.7. F1 Score

The F1 score is the weighted average of the precision value and recall value. The best value of the F1 score is considered as the 1 and the worst value of the F1 score is considered as the 0.

$$F1\ Score = 2\ \times \frac{Precision * Recall}{Precision + Recall}$$

Equation 6: F1 Score

The performance metrics gathered at the execution of the Job Portal Modules were evaluated based on the MAPE, RMSE, precision, recall and F1 score for all four resources.

5.2.8. Evaluations of the CPU

Based on the experiment's performance metrics of CPU, calculated MAPE percentage, RMSE value, and accuracy percentage, precision, recall, and F1 score respectively were 15.92%, 6.70, 84.07%, 80%, 80%, and 80%.

$$MAPE = \frac{\sum |(P-E)/P|}{N} = \frac{5.73}{36} = 0.1592$$

MAPE Percentage = 15.92%

RMSE value = 6.70

Accuracy Percentage = 84.07%

Table 5.2: Confusion Matrix - CPU

		Predicted Value	
Actual Value		Negative	Positive
	Negative	TN (30)	FP (1)
	Positive	FN (1)	TP (4)

$$Precision = \frac{TP}{TP + FP} = \frac{4}{4 + 1} = 0.800$$
$$Precision\ Percentage = 80\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{4}{4 + 1} = 0.800$$

$$Recall\ Percentage = 80\%$$

$$F\_score = \frac{2 \times Recall \times Precision}{Recall + Precision} = \frac{2 \times 0.80 \times 0.80}{0.80 + 0.80} = 0.80$$

$$F\_score\ \ Percentage = 80\%$$

Table 5.3: Eavluation Summary - CPU

Evaluation Matric Name	Percentage Value
MAPE	15.92%
RMSE	6.70
Precision	80%
Recall	80%
F1 Score	80%
Accuracy	84.07%

5.2.9. Evaluations of the RAM

Based on the experiment's performance metrics of RAM, calculated MAPE percentage, RMSE value, and accuracy percentage, precision, recall, and F1 score respectively were 6.86%, 3.17, 93.13%, 76.47%, 92.85%, and 83.87%.

$$MAPE = \frac{\sum |(P-E)/P|}{N} = \frac{2.47}{36} = 0.0686$$

MAPE Percentage = 6.86%

RMSE value = 3.17

Accuracy Percentage = 93.13%

Table 5.4: Confusion Matrix - RAM

		Predicted Value	
Actual Value		Negative	Positive
	Negative	TN (18)	FP (4)
	Positive	FN (1)	TP (13)

$$Precision = \frac{TP}{TP + FP} = \frac{13}{13 + 4} = 0.764$$

$$Precision\ Percentage = 76.47\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{13}{13 + 1} = 0.9285$$

$$Recall\ Percentage = 92.85\%$$

$$F\_score = \frac{2\ \times Recall\ \times Precision}{Recall + Precision} = \frac{2\ \times\ 0.92\ \times 0.76}{0.92 + 0.76} = 0.83$$

$$F\_score\ \ Percentage = 83.87\%$$

Table 5.5: Eavluation Summary – RAM

Evaluation Matric Name	Percentage Value
MAPE	24.71%
RMSE	3.17
Precision	76.47%
Recall	92.85%
F1 Score	83.87%
Accuracy	93.13%

5.2.10. Evaluations of the Disk

Based on the experiment's performance metrics of Disk, calculated MAPE percentage, RMSE value, and accuracy percentage, precision, recall, and F1 score respectively were 9.52%, 5.94, 90.47%, 95%, 86.36%, and 90.47%.

$$MAPE = \frac{\sum |(P-E)/P|}{N} = \frac{3.42}{36} = 0.0952$$

MAPE Percentge = 9.52%

RMSE value = 5.94

Accuracy Percentage = 90.47%

Table 5.6: Confusion Matrix - Disk

		Predicted Value	
Actual Value		Negative	Positive
	Negative	TN (13)	FP (1)
	Positive	FN (3)	TP (19)

$$Precision = \frac{TP}{TP + FP} = \frac{19}{19 + 1} = 0.95$$

$$Precision\ Percentage = 95\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{19}{19 + 3} = 0.863$$

$$Recall\ Percentage = 86.36\%$$

$$F\_score = \frac{2 \times Recall \times Precision}{Recall + Precision} = \frac{2 \times 0.95 \times 0.95}{0.86 + 0.86} = 0.90$$

$$F\_score\ \ Percentage = 90.47\%$$

Table 5.7: Eavluation Summary - Disk

Evaluation Matric Name	Percentage Value
MAPE	9.52%
RMSE	5.94
Precision	95%
Recall	86.36%
F1 Score	90.47%
Accuracy	90.47%

5.2.11. Evaluations of the Network

Based on the experiment's performance metrics of RAM, calculated MAPE percentage, RMSE value, and accuracy percentage, precision, recall, and F1 score respectively were 28.88%, 7.75, 71.16%, 81.25%, 72.22%, and 76.47%.

$$MAPE = \frac{\sum |(P-E)/P|}{N} = \frac{10.38}{36} = 0.288$$

MAPE Percentage = 28.88%

RMSE value = 7.75

Accuracy Percentage = 71.16%

Table 5.8: Confusion Matrix - Network

	Predicted Value		
Actual Value		Negative	Positive
	Negative	True Negative (15)	False Positive (3)
	Positive	False Negative (5)	True Positive (13)

$$Precision = \frac{TP}{TP + FP} = \frac{13}{13 + 3} = 0.8125$$

$$Precision\ Percentage = 81.25\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{13}{13 + 5} = 0.722$$

$$Recall\ Percentage = 72.22\%$$

$$F\_score = \frac{2\ \times Recall\ \times Precision}{Recall + Precision} = \frac{2\ \times\ 0.72\ \times 0.81}{0.72 + 0.81} = 0.76$$

$$F\_score\ \ Percentage = 76.47\%$$

Table 5.9: Eavluation Summary - Network

Evaluation Matric Name	Percentage Value
MAPE	28.88%
RMSE	7.75
Precision	81.25%
Recall	72.22%
F1 Score	76.47%
Accuracy	71.16%

5.3. Evaluate the Dashboard

The predicted resource provisioning outcomes than predicted in the ARIMA model were visualized in the Amazon QuickSight dashboard. This dashboard was illustrated the future resource requirements against the various modules in the system. These visualization has been figured out using different charts like line chart, plot chart, bar chart, etc.

The QuickSight dashboard was evaluated through a survey comprised of questionnaires. This questionnaire was distributed among the few members in the BI team including technical and non-technical users namely project manager, senior architecture, DevOps manager, DevOps engineer, and infrastructure engineer. The survey is basically focused on the ease of use and usefulness in the dashboard. The questions in the questionnaires can be rated as 1 to 5 as mentioned below.

- Strongly agreed

2. Agreed
3. Slightly agreed
4. Slightly disagreed
5. Disagreed

5.3.1. Ease of Use

Represented in Table 5.2 is the template of the questionnaire form, given to the members of the BI team. The members consisted of both technical and non-technical users who have 4 to 14 years of work experience in the IT industry.

Table 5.10: Questionnaires Form Template for Ease of Use

Ease of use questions	1	2	3	4	5	6
1) Easy to use.						
2) User-friendly.						
3) Possible to use with fewer instructions.						
4) It requires the fewest steps possible to accomplish what I want to do with it.						
5) Possible to use in day to day life.						
6) Will it be easy to use by general users?						

5.3.2. Usefulness

Represented in Table 5.3 is the template of the questionnaire forms given to the BI team members to evaluate the usefulness of the dashboard.

Table 5.11 : Questionnaires Form Template for Usefulnes

Usefulness questions	1	2	3	4	5
1) Very suitable for people with a very busy lifestyle.					
2) The system meets my needs.					
3) It has every feature I would expect to have.					
4) The system helps me to select the most suitable Garments					
5) It saves time.					
6) It has less complexity					

5.3.3. Overall Satisfaction

Accordingly, to the gathered statistics that were associated with ease of use and usefulness are presented in Table 5.4.

Table 5.12: Results of the Questionnaires

User	Ease of Use	Usefulness	Overall Satisfaction
User 1	4	5	4
User 2	5	4	5
User 3	3	4	4
User 4	5	4	4
User 5	3	4	3
User 6	4	5	4
User 7	5	4	5
User 8	5	3	3
User 9	5	4	4
User 10	3	3	5
Average	4.2	4	4.1

Depicted in Figure 5.4 is the bar chart of the overall evaluations that are created based on the gathered statistics in ease of use and usefulness.

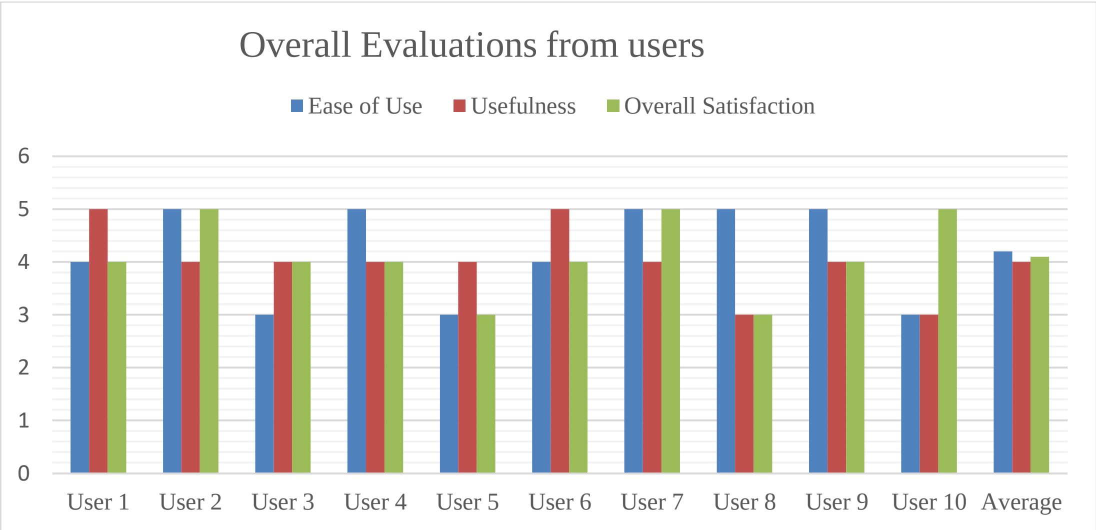

Figure 5.1: Overall Evaluations from Users

6. CONCLUSION

6.1. Chapter Overview

Due to the advances in technologies of virtualization, networking, web services, etc., cloud computing has been a promising paradigm for delivering IT services. The overhead of obtaining traditional infrastructure resources has been reduced as the cloud host providers offer an infinite amount of resources to their customers. In addition, they allow the enterprise application owners to choose a suitable pricing model based on the workload characteristics and resource usage patterns of their applications.

The dynamic resource scaling in the cloud reduces the risk of over-provisioning, wasting resources during non-peak hours. Choosing an appropriate dynamic resource scaling model for infrastructure migrations is a tedious task as it depends on the accuracy of the forecasted resource provisioning values. This research has been addressed to the identified problems mentioned above.

6.2. Research Contribution

The research intends to model a POC (Proof of Concepts) for the future resource prediction which assists for the dynamic resource provisioning scaling in infrastructure migrations. These resource provisioning predictions were done using the performance metrics gathered in the on-premises servers. The performance metrics gathered during 30 days were evaluated using the ARIMA time series model. The best fitted ARIMA model was chosen based on the least RMSE value.

The best fitted ARIMA model has predicted the future resource requirement of CPU, by evaluating the 30 days of metrics. The model was trained based on the 75% data sets and was tested using 25% data sets. The resource provisioning model has been evaluated in the IaaS layer of the Amazon cloud. As the experimental cloud infrastructure, an EC2 in the Amazon Web Services has been used. The observed values were evaluated using RMSE and MAPE factors as accuracy assessment criteria.

The obtained accuracy percentage of the resources summary were represented in Table 6.1. It is ideal to have less MAPE value as much as possible. The MAPE value of the highly accurate forecasting model is considered to have less than 10%.

Table 6.1: Evaluation Summary for Resources

	CPU	RAM	Disk	Network
MAPE	15.92%	24.71%	9.52%	28.88%
RMSE	6.7	3.17	5.94	7.75
Precision	80%	76.47%	95%	81.25%
Recall	80%	92.85%	86.36%	72.22%
F1 Score	80%	83.87%	90.47%	76.47%
Accuracy	84.07%	93.13%	90.47%	71.16%

Further, a visualizing the future forecasted resource provisioning results is provided in a dashboard. The developed dashboard has been evaluated through a survey comprised of a questionnaire that was distributed among a DevOps team which is consists of ten members. The sample of the survey consisted of both technical and non-technical users who have 4 to 14 years of work experience in the IT industry. The overall average satisfaction rate of the survey was 4.1 out of 5, which was a considerably acceptable result.

6.3. Research Limitation and Future Work

The ARIMA model of resource provision forecasting has been evaluated based on the performance metrics patterns. Because of this, the accuracy of the chosen ARIMA model highly depends on the gathered performance metrics. The ARIMA model is applicable only for a time series that has approximately constant values for mean, variance and autocorrelations.

Hence it is recommended to collect performance metrics patterns values at least for one year. But in some cases, there will be a utilization pattern in quarter wise. In this case, it is recommended to gather the data sets for at least five years. To achieve the high accuracy percentage for the forecasted results, it is recommended to train data set with a large number of data set. The proposed ARIMA model is fitted only with the environments which have the growing and the periodic performance metrics. In the future, this model can be improved to predict the resource provisioning that has unpredicted performance metrics.

In the dashboard implementation, the forecasted data was gathered as the .CSV files and uploaded to the AWS Quick Sight. As an enhancement, the AWS Quick Sight and the forecasted module can be integrated so that the data representing layer would be automated.

REFERENCES

[1] Amazon WEb Services, "An Overview of the AWS Cloud Adoption Framework," AWS, 2017 February.

[2] Riverstone Technology Unlimited, "DevOps Strategy Document," Riverstone LLC.

[3] A. v. H. W. D. Andreas Brunnert, "Performance-oriented DevOps: A Research Agenda," SPEC RG DevOps Performance Working Group, August 20th, 2015.

[4] D. K.Kearns, "Planning & Management Methods for Migration to a Cloud Environment," The MITRE Corporation, 2017.

[5] Amazon Web Services, Migrating to AWS:Best Practices and Strategies, Amazon Web Services, 2019, p. 20.

[6] AWS, "AWS Application Discovery Service User Guide," AWS, 2019.

[7] CIO Whitepapers Review, "What is Busniess Intelligence System," CIO Whitepapers Review, 2018. [Online].

[8] "https://www.guru99.com/introduction-sap-bi.html," [Online]. Available: https://www.guru99.com/introduction-sap-bi.html.

[9] Atlassian, "DevOps: Breaking the Development-Operations barrier," Atlassian, 2019. [Online]. Available: https://www.atlassian.com/devops. [Accessed 24 May 2019].

[10] Zhou, Yilei Liu & Yiran, "The Challenges and Mitigation Strategies of," Blekinge Institute of Technology,Faculty of Computing, 2017.

[11] Splunk, "Application Lifecycle Analytics," Splunk, 2017. [Online]. Available: https://www.splunk.com/en_us/it-operations/devops-app-lifecycle.html. [Accessed 10 June 2018].

[12] IBM, "Benefits of cloud computing," IBM, [Online]. Available: https://www.ibm.com/cloud/learn/benefits-of-cloud-computing. [Accessed 12 July 2019].

[13] Amazon Web Services, "AWS Migration Whitepaper," *AWS Professional Services*, p. 44, March 2018.

[14] AWS, "An Overview of the AWS Cloud Adoption Framework_version 02," AWS, 2017 February.

[15] M. L. i. B. Intelligence, "Cloud Computing Challenges Businesses Are Facing In These Days," Datapine, 10 January 2019. [Online]. Available: https://www.datapine.com/blog/top-6-cloud-computing-challenges/. [Accessed 11 July 2019].

[16] L. Dignan, "Top cloud providers 2019: AWS, Microsoft Azure, Google Cloud; IBM makes hybrid move; Salesforce dominates SaaS," 2019. [Online]. Available: https://www.zdnet.com/article/top-cloud-providers-2019-aws-microsoft-azure-google-cloud-ibm-makes-hybrid-move-salesforce-dominates-saas/. [Accessed 22 July 2019].

[17] X. L. R. S. M. G. Ta Nguyen Binh Duong, "A framework for dynamic resource provisioning and adaptation in IaaS clouds," 2011 Third IEEE International Conference on Coud Computing Technology and Science, Singapore, 2011.

[18] webopedia, "Webopedia: Online Tech Dictionary for Students, Educators and IT Professionals," webopedia, 2019. [Online]. Available: https://www.webopedia.com/TERM/D/devops_development_operations. [Accessed 13 June 2019].

[19] J. Wettinger, "Teaching Agile Development with DevOps in a Software," no. October 23, 2012.

[20] Aamazon Web Services, "What is DevOps," Amazon Web Services, 2019. [Online]. Available: https://aws.amazon.com/devops/what-is-devops/. [Accessed 30 June 2019].

[21] P. Ashoori, "10 Tools For DevOps Success - Article by Ericka Chickowski," 29th January 2016. [Online]. Available: https://www.linkedin.com/pulse/10-tools-devops-success-article-ericka-chickowski-parham-ashoori/. [Accessed 2nd April 2019].

[22] M. Natarajan, "60 Best Open Source Tools to do Devops," CenturyLink, 25 January 2017. [Online]. Available: https://www.ctl.io/developers/blog/post/60-best-open-source-tools-to-do-devops. [Accessed 12 May 2019].

[23] D. Wadhwa, "What is the Relationship between DevOps and Cloud," CCSI, 2016. [Online]. Available: https://www.ccsinet.com/blog/relationship-devops-cloud/. [Accessed 22nd June 2019].

[24] DevOps, "https://devops.com/devops-and-cloud-a-symbiotic-relationship/," MediaOPs, 2019. [Online]. [Accessed 25 June 2019].

[25] TechBeacon, "https://techbeacon.com/app-dev-testing/devops-dictates-new-approach-cloud-development," Microfocus, 2019. [Online]. [Accessed 24 June 2019].

[26] Q. Z. ·. L. C. ·. R. Boutaba, "Cloud computing: state-of-the-art and research challenges," The Brazilian Computer Society 201, 20 April 2010.

[27] Cloud Academy Team, "Understanding Cloud Migration in the Enterprise," Cloud Academy Team,, 10th October 2017. [Online]. Available: https://cloudacademy.com/blog/understanding-cloud-migration-in-the-enterprise/. [Accessed 4th April 2019].

[28] E. Colman, "Difference Between SAAS, PAAS, IAAS," 27 August 2013. [Online]. Available: https://www.computenext.com/blog/when-to-use-saas-paas-and-iaas/. [Accessed 12th June 2019].

[29] J. Woods, "Five Options for Migrating Applications to the Cloud: Rehost, Refactor, Revise, Rebuild or Replace," in *Gartner The Future of IT Conference*, Centro Banamex, Mexico City, Mexico, 2011.

[30] I. S. B. T. Ali Khajeh-Hosseini, "Decision Support Tools for Cloud Migration in the Enterprise," in *2011 IEEE 4th International Conference on Cloud Computing*, 2011.

[31] J. K. L. L. Sadeka Islam, "Empirical prediction models for adaptive resource provisioning in the cloud," *Future Generation Computer Systems,* p. 155, 2012.

[32] N. H. S. K. E. A. Nikolas Roman Herbst, "Self-adaptive workload classification and forecasting for proactive resource provisioning," *Concurrency And Computation,* p. 2053, 10 March 2014.

[33] X. Y. Arijit Khan, "Workload Characterization and Prediction in the Cloud:A Multiple Time Series Approach," IBM T. J. Watson Research Center.

[34] S. A. A. C.-H. L. Ali Yadavar Nikravesh, "Towards an Autonomic Auto-Scaling Prediction System for Cloud Resource Provisioning," 2015 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, 2015.

[35] A. B. a. D. Debnath, "Cloud Migration Tools :Overview and Comparison," Oakland University, Rochester Hills, USA.

[36] R. Wiselman, "About Azure Migrate," Microsoft Corporation, 22nd October 2019. [Online]. Available: https://docs.microsoft.com/en-us/azure/migrate/migrate-services-overview. [Accessed 06th November 2019].

[37] IBM, "IBM Services for Cloud Migration," IBM, [Online]. Available: https://www.ibm.com/services/cloud/migration. [Accessed 11th November 2019].

[38] Rodrigo N. Calheiros, Enayat Masoumi, Rajiv Ranjan, Rajkumar Buyya, "Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications’ QoS," IEEE Transactions On Cloud Computing.

[39] A computer science portal for geeks, "Static Data Structure vs Dynamic Data Structure," GeeksforGeeks, 2018. [Online]. Available: https://www.geeksforgeeks.org/static-data-structure-vs-dynamic-data-structure/. [Accessed 13th January 2019].

[40] Mathsisfun, "Probability and Statistics," Standard Deviation Formulas, 2017. [Online]. Available: https://www.mathsisfun.com/data/standard-deviation-formulas.html. [Accessed 13th January 2019].

[41] Techopedia, "Data Cleansing," Techopedia, 2019. [Online]. Available: https://www.techopedia.com/definition/1174/data-cleansing. [Accessed 14th May 2019].

[42] EDQ, "Data Cleansing," Experian, 2019. [Online]. Available: https://www.edq.com/glossary/data-cleansing/. [Accessed 13th June 2019].

[43] R. Hoffmann, "Box Plot: Display of Distribution," 1981. [Online]. Available: http://www.physics.csbsju.edu/stats/box2.html. [Accessed 13th June 2019].

[44] M. U. A. Rob J Hyndman and George Athanasopoulos, "Forecasting: Principles and Practice," April 2018. [Online]. Available: https://otexts.com/fpp2/. [Accessed 16th June 2019].

[45] Amazon Web Services, "Amazon Ec2," June 2019. [Online]. Available: https://aws.amazon.com/ec2/. [Accessed 15th July 2019].

[46] The Hitchhiker's Guide to Python., "Scientific Applications," 2019. [Online]. Available: https://docs.python-guide.org/scenarios/scientific/. [Accessed 15th May 2019].

[47] Best Python libraries for Machine Learning, "GeeksforGeeks," 2019. [Online]. Available: https://www.geeksforgeeks.org/best-python-libraries-for-machine-learning/.

[48] Amazon Web Services, "Amazon Quick Sight User Guide," 2019. [Online]. Available: https://docs.aws.amazon.com/quicksight/latest/user/welcome.html. [Accessed 18th July 2019].

[49] Springer Link, "Encyclopedia of Production and Manufacturing Management," 2000. [Online]. Available: https://link.springer.com/referenceworkentry/. [Accessed 12th July 2019].

[50] Data Science Central, "RMSE: Root Mean Square Error," 2019. [Online]. Available: https://www.statisticshowto.datasciencecentral.com/rmse/. [Accessed 13th June 2019].

[51] K. P. Shung, "Accuracy, Precision, Recall or F1," 15 March 2018. [Online]. Available: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9. [Accessed 12th June 2019].