Predictive Analysis and Development of Dimensioning Tool

Thesis submitted in partial fulfillment of the
Requirements for the degree of

Master of Engineering

ME in Big Data and Data Analytics

By

Uday T

(181046016)

Internship start date: 03/06/2019

Under the Guidance of

External Guide:
Mr. Ravi Rajaram
R&D Specification Specialist
Nokia Networks
Bengaluru

Internal Guide:
Mr. Arockiaraj S
Assistant Professor
MSOIS, MAHE
Manipal

MANIPAL
ACADEMY of HIGHER EDUCATION
(Deemed to be University under Section 3 of the UGC Act, 1956)

MANIPAL SCHOOL OF INFORMATION SCIENCES

(A Constituent unit of MAHE, Manipal)

MANIPAL
ACADEMY of HIGHER EDUCATION
(Deemed to be University under Section 3 of the UGC Act, 1956)

MANIPAL SCHOOL OF INFORMATION SCIENCES

(A Constituent unit of MAHE, Manipal)

CERTIFICATE

This is to certify that this thesis work titled

Predictive Analysis and Development of Dimensioning Tool

Is a bonafide record of the work done by

Uday T

(181046016)

In partial fulfillment of the requirements for the award of the degree of **Master of Engineering - ME (Big Data & Data Analytics)** under MAHE, Manipal and the same has not been submitted elsewhere for the award of any other degree. The dissertation does not contain any part / chapter plagiarized from other sources.

External Guide:
Mr. Ravi Rajaram
R&D Specification Specialist
Nokia Networks
Bengaluru

Internal Guide:
Mr. Arockiaraj S
Assistant Professor
MSOIS, MAHE
Manipal

Prof. Ramashesha C S
Director,
Manipal School of Information Sciences
MAHE, Manipal

CONTENTS

LIST OF FIGURES

ABSTRACT

In this era where data and voice services are available at a push of a button, service providers have virtually limitless options for reaching their customers with value-added services. The changes in services and underlying networks that this always-on culture creates make it essential for service providers to understand the evolving business logic.

There are various procedures that must be taken care of in order to provide a seamless fault-tolerant and highly available Network Management System (NMS) and Element Management System (EMS).

In telecom networks the NOKIA NetAct continues to evolve with the new technologies and services in mobile broadband with larger data of gigabits and terabits. Network Management is used to recognize the best practices to improve the operational efficiency, communication in the operation teams, reducing the operational risks and it ensures operational resilience.

Network management features support the life cycle of networking technologies by testing some practices in simulation environment before deploying to the real environment. In these test scenarios with various loads and configurations huge data is generated which can be analyzed to find trends, patterns and deploy machine learning models.

CHAPTER 1

INTRODUCTION

NetAct

Nokia NetAct is a new generation network management system for multi-vendor and multi-technology networks.

NetAct serves both as a network management system and as an element management system. NetAct offers a wide range of unified operation and maintenance capabilities for network elements in core, radio and transport networks both for managing physical network elements as well as virtualized network functions. It consists of many tools for handling several network elements and expanding networks. It is designed for handling an increase in both complexity of the network and the amount of traffic and data.

To facilitate its integration in multi-vendor environments, NetAct provides several open northbound interfaces that offer to operators a unified management capability for the entire network. For this purpose, a mediation application function provided by a so-called Northbound Interface Agent performs the mapping between the management information received by NetAct or generated in NetAct and the information required by the upper level management systems.

In addition to NetAct applications there are also standalone tools that are dedicated to specific management functions and provide advanced management features, for example, NetAct Advanced Configurator for configuration management. These standalone tools need a dedicated hardware platform and are sold separately.

NetAct Architecture

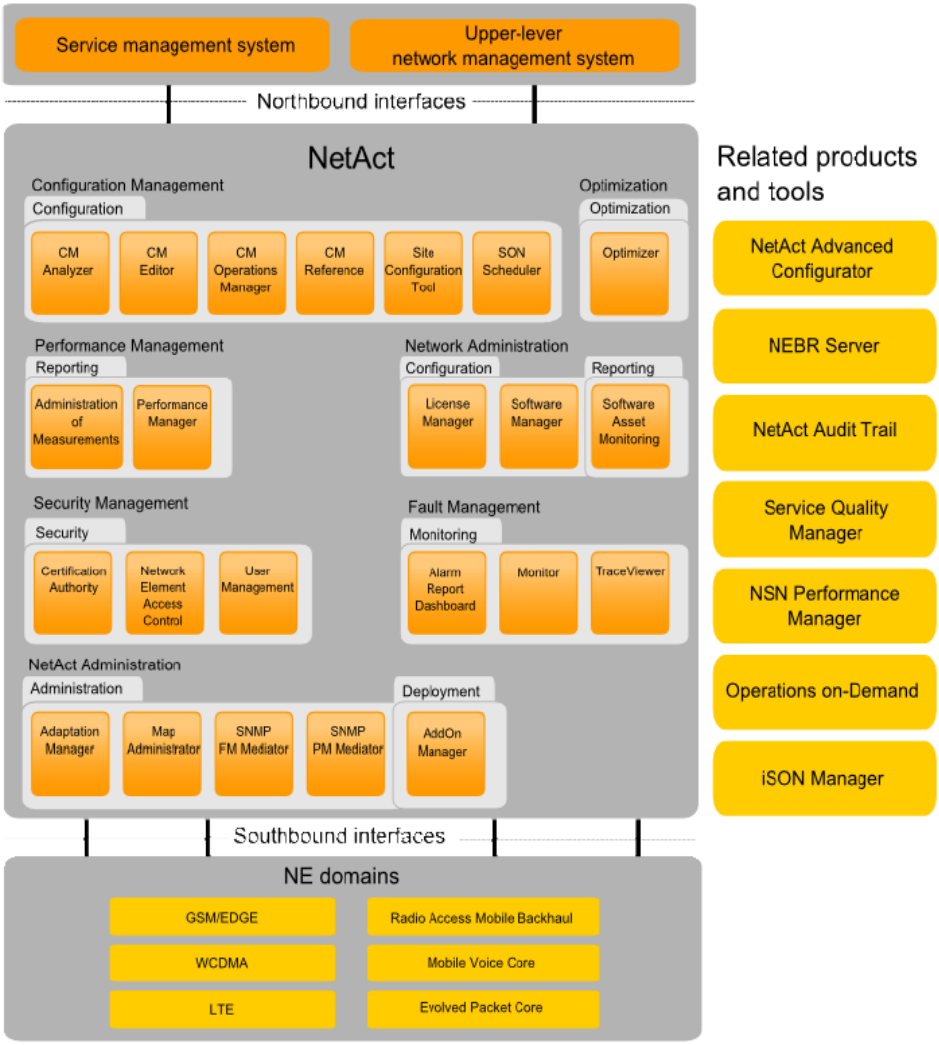

Fig 1. NetAct Architecture

FCAPS in NetAct

Network Management System (NMS) is a set of systems and applications which allows monitoring of a single Network Element (NE) or the entire network. It deals mainly with execution of the set of functions required for planning, allocating, controlling, deploying, monitoring and coordinating the resources of a network including performing functions such as performance management, fault management, accounting management, security management, configuration management and bandwidth management. The core management functions in a Network management system can be broadly classified into five management functions (FCAPS):

- **F**ault Management
- **C**onfiguration Management
- **A**ccounting Management
- **P**erformance Management
- **S**ecurity Management

With NetAct both the network and the services within the network are managed centrally enabling the operator to have an overall view of network element failures, service quality indicators, and network traffic from single screen.

In NetAct, system reliability and availability are ensured through the following system properties and features:

- NetAct virtualization ensures the availability of NetAct applications and services. It features the automated failover of the virtual machines inside the NetAct and aims at eliminating single points of failure in power, disk, memory, CPU, network connectivity, virtual machines, or service availability
- Load balancing is responsible for the distribution of server load within a WebSphere Application Server cluster and distribution of resources within the virtual infrastructure
- Online and offline backup and recovery solutions using either vSphere Data Protection or other backup solutions
- System Self-monitoring for internal faults and performance management
- Preventive Health Check tool helps verify the NetAct system status
- Resource redundancy
- Self-monitoring framework allows NetAct components to report failure issues as alarms processed in the NetAct FM pipe and report the system performance status as measurements that are then processed by the NetAct PM pipe.

System administrators are the most privileged users of the NetAct system and are responsible for setting up and maintaining the operating environment for other users. They carry out routine maintenance tasks to ensure that the resource utilization is optimized, check network performance to respond better to resource contention, and to do basic troubleshooting in case of service or network problems.

NetAct Configurations and Hardware

NetAct is delivered on several configurations using HP blade servers.

1. Small, Compact
2. Mainstream (M)
3. Large, XXL

The NetAct Small and NetAct Compact both consists of 25 VMs but the Compact houses 2 servers and Small houses 4 servers.

The NetAct Mainstream configuration consists of 31 VMs with 6 servers and eight core processors.

The NetAct Large and XXL configuration consists of 37 VMs with 6 servers and 12 core processors. The only difference is XXL consists of larger primary storage options.

These are the configuration options that are provided to the customers. The customers can upgrade or downgrade their hardware at any point of time as NetAct is scalable for all configuration options.

Overview of Monitoring and Analytics Tools

Grafana

Grafana allows you to query, visualize, alert on and understand your metrics no matter where they are stored. Create, explore, and share dashboards with your team and foster a data driven culture:

- **Visualize:** Fast and flexible client-side graphs with a multitude of options. Panel plugins for many ways to visualize metrics and logs.
- **Dynamic Dashboards:** Create dynamic & reusable dashboards with template variables that appear as dropdowns at the top of the dashboard.
- **Explore Metrics:** Explore your data through ad-hoc queries and dynamic drilldown. Split view and compare different time ranges, queries and data sources side by side.
- **Explore Logs:** Experience the magic of switching from metrics to logs with preserved label filters. Quickly search through all your logs or streaming them live.
- **Alerting:** Visually define alert rules for your most important metrics. Grafana will continuously evaluate and send notifications to systems like Slack, PagerDuty, VictorOps, OpsGenie.
- **Mixed Data Sources:** Mix different data sources in the same graph! You can specify a data source on a per-query basis. This works for even custom data sources.

ZABBIX

Zabbix is a highly integrated network monitoring solution, offering a multiplicity of features in a single package. It has various features built in with the tool, here are some prominent ones

Data gathering

- Availability of system
- Performance metrics
- Metrics with desired time intervals

Highly configurable alerting

- Can set the threshold for peak usage of performance metrics
- Can call triggers
- Can send notifications

Real-time graphing using Web monitoring capabilities

- All monitored metrics are immediately graphed using built in graphing functionality
- Can check functionality and response of metrics in the front-end window

Extensive visualisation options

- Time series data is visualized for each metrics
- can hold data at different time intervals
- high-level view of monitored resources

Historical data storage

- data stored in a database
- configurable history

The Analysis and Reporting System is an amalgamation of the online-oriented Performance monitoring and the offline-oriented Performance Reporting as shown in Figure1.1. This System is deployed on Zabbix server and accessed from a web-client.

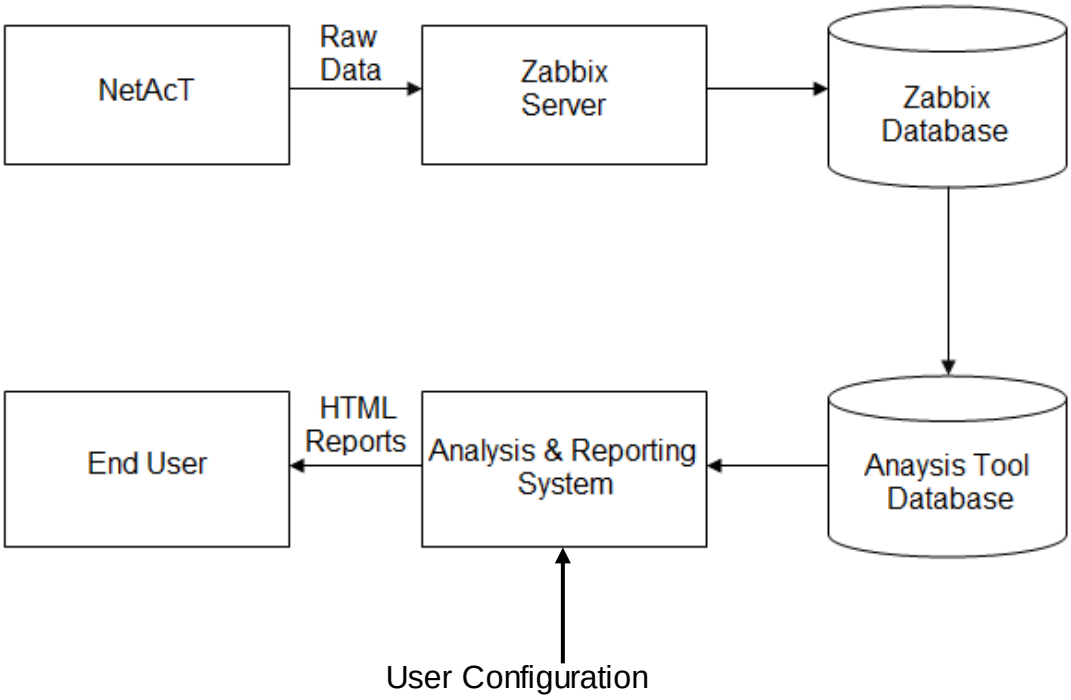

Figure 2. Analysis and Reporting System in Zabbix

It is a two-tier architecture with one-to-one mapping between the Zabbix database and the system's database. Zabbix collects the network wide data from the NetAct into the database. This raw data is fetched from Zabbix Database and the configuration information is provided as a user input to perform the required analysis and depending on the customer requirements, Testcases generate the final analyzed report for the end User. The results from this report provide measured and analyzed performance characteristics of the NetAct.

Similarly, Grafana is another network monitoring tool in which the data is transformed using each DCFW metric directly to a front-end visualization window. There is standard path and file naming over all clusters which makes fetching of data simple to PAF(Grafana). Files are stored in HDFS.

System Performance and Dimensioning Tool

The System Performance team is responsible for executing various test-cases and test-environments in simulation before deploying the build/version/upgrade to a customer. This teams ensures that there are no escape defects from R&D development to customers. These tests ensure that the customer faces no issues when they deploy NetAct.

There are two scenarios in deploying NetAct to customers site:

1. Greenfield deployment
2. Network Expansion

Greenfield: Greenfield deployment is where a completely new Network along with NetAct is deployed in the customer site. This is when a customer would opt to use NetAct where deployment of a network is done where none existed before.

A customer would provide the network details and they would like to use NetAct as their NMS and EMS tool. NetAct support teams will estimate the overall configuration details required to setup the network by obtaining the customer requirements. Since there are three primary configurations in NetAct, using primary KPIs estimation and the dimensioning tool, the NetAct support teams can provide customer with the optimal type of supported configuration.

Network Expansion: When customers wants to increase their existing capacity by adding new Network elements due to the new users, NetAct also needs to be planned for increase in capacity to support additional Network elements, then NetAct support teams approaches NetAct for increase in the supported capacity (This includes increase in Network Elements, Network technologies, tools and other hardware elements).

The purpose of the project is to perform exploratory data analysis for different use - cases on these test results and tabulate the KPI data results. After tabulating the data of various KPIs for different use – cases, we implement a machine learning model to predict the configuration details for both greenfield deployment and capacity increase.

CHAPTER 2

DETAILS OF WORK DONE

1. Time Series Analysis.

In this approach the data of a metric was captured from Grafana for a time interval. The goal was to predict the metric's behavior for a time period. For this regression algorithms were initially implemented, but due to varying and non-linear pattern there were inconclusive results. Hence **ARIMA** approach was used.

ARIMA (Auto **R**egression **I**ntegrated **M**oving **A**verage) is a Time series-based algorithm which tries to capture the seasonality and trend in the data using Regression and Averaging method and tries to predict the values for future time periods.

The first process is to obtain data from Grafana, pre-process the data and then implement Auto Regression and Moving Average on it. After this process if there is stationarity (tested using statistical methods) in data, we implement the ARIMA model. This data was not stationary and hence to remove the stationary we use several methods such as rolling mean, moving average, weighted average, 1st and 2nd order differentiation etc.

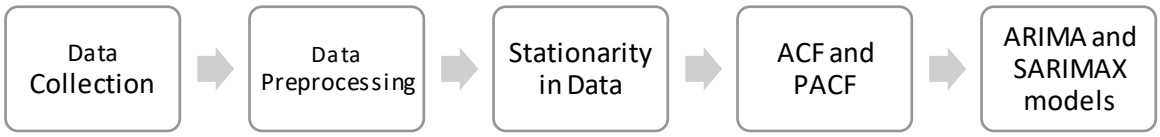

Fig 3. Workflow of time series analysis.

After this we use seasonal decomposition plots to see if there is trend and seasonality in data. Using ACF and PACF (Auto Correlation and Partial Auto Correlation) Plots we determine the order of the algorithm, then we build the model for training data (70:30) and deploy the same for test data.

After this we predict the values for test set and obtain conclusive results, we forecast the values for future data.

2. Exploratory data analysis on test results for dimensioning tool.

The System Performance team conducts various tests on the simulation environment on which there are different loads, various management functions, different configurations on different categories of hardware.

A NetAct Dimensioning Tool contains around 5 - 10 KPIs (Knowledge Performance Indicators) through which a customer or vendor can know if the configuration hardware that he either wants to buy will support the metrics or not.

The tool also helps upgrading from a small configuration to higher configuration or downgrading to a smaller configuration from higher configuration with the help of threshold values set for each configuration. However, these values are obsolete and can't be adjusted or modified. We try to obtain variable values for these KPIs using exploratory data analysis.

Every test that is being conducted will have its data integrated to ZABBIX and GRAFANA. The task here is to extract the data from these tools for the test that are conducted periodically and do exploratory data analysis in order to find patterns, anomalies relations, optimum values etc. using the data.

Significant values on this analysis are recorded in a tabulated form using either statistical methods or visual methods. This process is done for all the major tests that are carried out regularly.

Then the tabulated data is compared for different tests and recorded to help with the dimensioning tool.

Fig 4. Workflow of EDA for test results

There are different use cases for this analysis,

2.1 Peak Value Measurement.

Here we measure the peak value of a metric at a regular interval period and compare the data of the same metric for different tests.

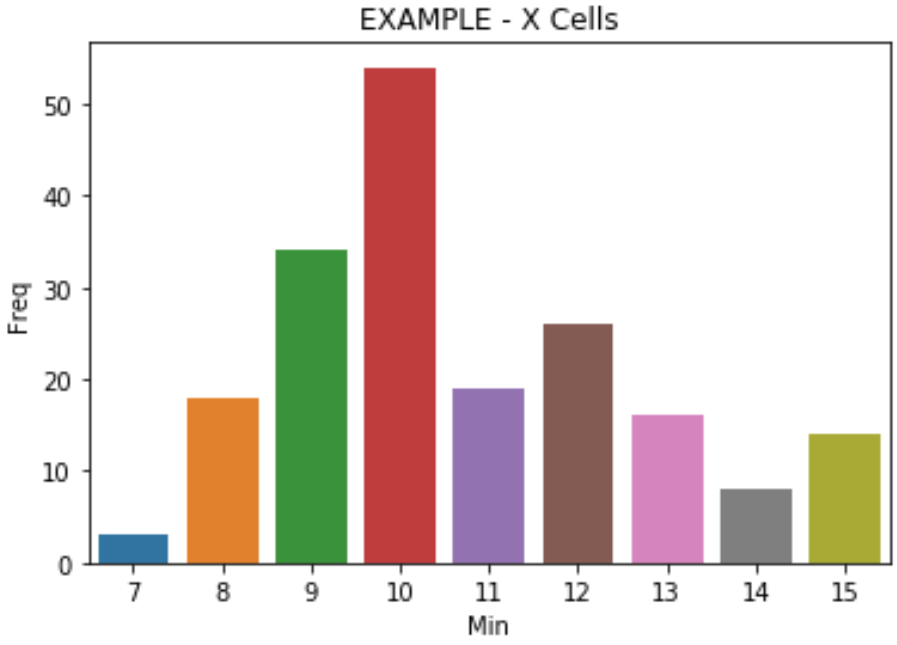

Fig 5. Peak value measurement of a metric.

For example, suppose we are measuring CPU usage of a VM where the task is to find the number of minutes the CPU stayed greater than a certain threshold.

Inference: For X cells at every interval, the greatest number of times CPU stayed greater than 20% was 10 Minutes

2.2 Impact of various metrics due to a primary feature.

Here collect the data of various metrics and then compare the change in those values w.r.t to different test scenarios using visual methods. After this we tabulate the change in those values.

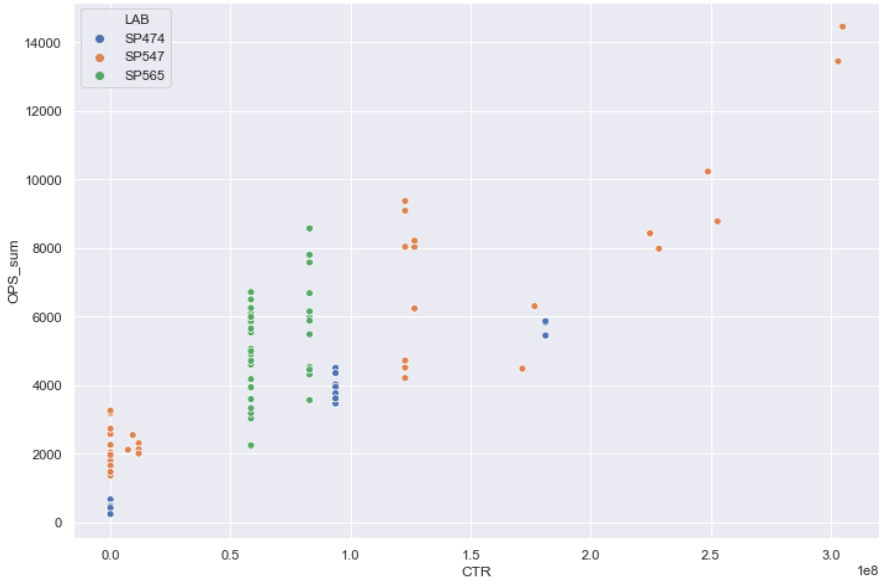

Fig 5. Impact of metric due to a primary feature for different test results.

2.3 Impact on metrics due to change in Configuration.

There are different kinds of configurations used in simulation environment. Some examples include Oracle striped configuration, 3PAR, 2PAR and non-striped labs. The same amount of data is pumped into labs with different configurations and the change in data is tabulated and detailed analysis is done on the data.

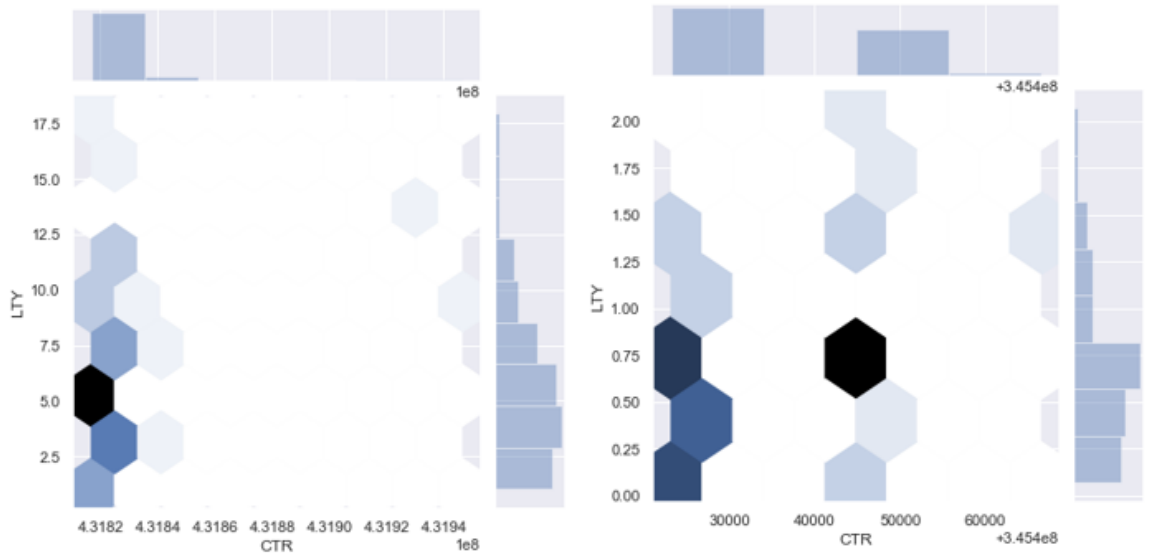

Fig 6. Change in Latency due to Counters for Striped vs Non-Striped labs

CHAPTER 3
DETAILS OF PROPOSED WORK

These are the proposed objectives for upcoming iterations,

1. Conduct similar exploratory data analysis and record the significant trends and values for upcoming tests.
   1.1. Perform EDA on the tests that are being conducted regularly.
   1.2. Tabulating the resultant optimum values for test metrics.

2. Analyzing Customer Data to gain insights that change w.r.t simulation environment and real-time.
   2.1. Customer use-case will vary compared to simulation environment.
   2.2. Various factors (backup, alarms, failures) impact the overall data.
   2.3. EDA on this data can be compared with the simulation data to gain insights.

3. Analyzing 10 unique KPIs for different tests and build ML model to predict the KPI values for different use cases.
   3.1. Obtain a tabulated data of all KPIs w.r.t to different tests
   3.2. Store all these data in a single database.
   3.3. Analyze and obtain a relationship between KPIs
   3.4. Build a ML model to predict the KPI outcomes for different test configurations.

CHAPTER 4
LIBRARIES / TOOLS USED

- Programming Languages : Python and R
- IDE : Jupyter (Anaconda) and R Studio
- Libraries : Sklearn, Pandas, Numpy, Seaborn, Statsmodels
- Analytics and Monitoring : Grafana and Zabbix
- Others : PET Analysis Tool
  (Performance Engineering and Testing)

CHAPTER 5
BIBLIOGRAPHY

- ARIMA and Time Series

  https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

  https://www.machinelearningplus.com/time-series/time-series-analysis-python

- NetAct Documentation
- Nokia, “NetAct Virtualized OSS that goes beyond network management”,
- Flextronics Software System, "FCAPS", White Paper.
- Nokia Siemens & Network, “OSS architecture –the next generation”,
- Grafana

  https://grafana.com/docs/

  https://github.com/grafana/grafana
- Zabbix SIA

  https://www.zabbix.com/documentation/2.4/manual

  https://github.com/zabbix/zabbix