DATE: 9/20/2020
COURSE: IS698
NAME: Anthony PAGAN

Project Proposal: Analyze current and historical performance metrics and predict critical system errors before degradation of system.

Introduction

Many organizations rely on monitoring systems to monitor performance of their backend systems. The expectation is that the monitoring will notify support staff when critical systems are in a critical state. Unfortunately, many time support teams are only alerted after a system is degraded. In some cases, manual monitor tuning results in false alerts.

The main goal of this project is to build a proactive monitoring system model based on performance metric calculations that can be correlated to events and allow teams to build automated resolutions. Such a system would reduce support cost and increase efficiency while minimizing false non-actionable alerts.

Relevant Research

- Monitoring and prediction Linux server performance with linear regression, Pavel Barca,Bojan Vujanić,Nemanja Maček (https://www.researchgate.net/publication/325174187_Monitoring_and_Predicting_Linux_Server_Performance_With_Linear_Regression)
    - This paper details the uses linear regression to for prediction analysis based on performance metrics and monitoring. The paper focuses on Linux systems. The paper includes details on the different approaches including linear regression, multiple linear regression, multivariate regression and non-linear regression. The approach uses Prometheus to gather the metrics and perform linear analysis, Telegraf plug-ins expose the data and Grafana to display visualizations.

- Deep Recurrent Model for Server Load and Performance Prediction in Data Center, Zheng Huang, Jiajun Peng, Huijuan Lian, Jie Guo, and Weidong Qiu (https://www.hindawi.com/journals/complexity/2017/8584252/)
    - This paper looks at user log to request as it pertains to server performance. Their method is to apply RNN-LSTM network to predict server workload and performance.

Research Question

In this research topic We are attempting to answer the following questions:

- Can we predict critical system errors before they occur by analyzing current and historical performance metrics with a high-level accuracy?
- Can we use existing ticketing systems as a recommender system to add value and automate resolutions?

Methodology

Our approach would be to collect critical alerts and key performance metrics and identify correlations. The metrics and alerting will be collected from System Center Suite of products and we will focus on Windows servers.

High-Level Performance Metrics Collection Goals:

- Calculate statistics on base metrics (CPU/Memory/Disk utilization)
- Build anomaly detection to identify metrics when they are outside a range of 2-3 standard deviation for a set amount of time
- Build correlation dashboard for performance metrics vs critical alerts with what-if analysis
- Group metrics on a cluster type of services (AD/Exchange/Lync..etc)
- Build automation/self-healing scripts based on detected anomalies

Assumptions

We will assume we can use LSMT for the metrics data and binary regression for alerting with time series data. The data will be internal data. As a result, we will need to transform the data to protect internal server names. This will not affect predictions since the official identification is not required for our analysis.