LOG 6309E - Intelligent DevOps

Individual Assignment: Analyzing Performance Data

Notes

- This assignment is done individually.

- Weight: 5% of the final grade.

- Deliverable: a report of 2-4 pages in PDF.

- Due: Friday, Sept 19, 11:59 PM.

Objectives

The objectives of this assignment include:

- To understand common data analytics techniques (e.g., hypothesis testing, effect size).

- To apply data analytics techniques to DevOps data (e.g., performance data).

- To apply data visualization to explore or interpret data.

- To draw conclusions from data analytics results.

Specifications

Software monitoring is a critical to ensure the quality of software systems and their delivered services. The raw monitoring data are typically analyzed to understand the states of the system or to recommend data-driven decisions. For example, the performance monitoring data can be analyzed to determine whether a performance regression exists, or whether a scaling-up is needed to improve the performance of the system.

This Google Drive folder includes the monitoring data of several versions of a software system (OpenMRS) under random workloads: *base_version*, *version_1*, *version_2*, and *version_3*. The monitoring data is aggregated at 30-second intervals. The "TOTAL_CPU" column indicates the average CPU usage of the system in each 30-second interval. Analyze the data to understand the differences among the performance (i.e., the "TOTAL_CPU" column) of the different versions. Specifically, answer the following questions:

- Q1: Are the performance data (i.e., CPU usage) of the different versions normally distributed?

- Q2: Is there a statistically significance difference in performance between *base_version* and *version_1*?
- Q3: If there is a statistically significant difference, how large is the difference?
- Q4: Considering *base_version*, does the performance change over time?
- Q5: Group and rank the different versions according to the statistical difference in their performance.

Use data analytics techniques (e.g., hypothesis testing, effect size) to answer these questions. Explain your rational for choosing each technique.

Use visualization (e.g., comparing distributions) to support your analyses.

Discuss your analysis results and highlight your conclusions.