

USING REGRESSION MODEL TO ANALYZE THE WORKLOAD OF EXAMINATION DATABASE SERVER

Junaid Fazlani¹, *Aftab Ahmed Chandio², Qamar-ul-Nisa Chandio³

^{1,2}Institute of Mathematics and Computer Science, University of Sindh, Jamshoro, Pakistan.

³Government Degree Boys College, Qasimabad, Education and Literacy Department, Government of Sindh, Hyderabad, Pakistan.

*Corresponding Author: (chandio.aftab@usindh.edu.pk)

DOI:(<https://doi.org/10.71146/kjmr873>)

Article Info

Abstract



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license
<https://creativecommons.org/licenses/by/4.0>

This study is focused on the workload pattern analysis of database servers running Examination Management Systems (EMS) in universities. Tests and examinations in schools and colleges are a wide range of assessment activities planned by teachers to measure, monitor, and mark the performance of students in studies, learning, and skill acquisition in a specific subject. As education administration becomes increasingly digitized, these core academic processes are now dominated by advanced, computer-based EMS systems. Therefore, it is imperative to know the workload pattern of the underlying database servers so that the performance can be optimized, the system can be made stable at peak usage hours, and future upgrades in educational technology infrastructure can be planned. This study aims to assess the workload on the database server of a Public Sector University, with a case study of the University of Sindh, Jamshoro, during official working hours. A large number of human resources with robust computer-based infrastructure are needed because of the complexity of examination process including time-table generation, attendance tracking, mark tabulation, and the issuance of academic documents (i.e., mark sheets, certificates, and degrees). In this research, statistical analysis techniques including regression are employed to evaluate system workload. In this way, the results help to enhance performance, reduce failure risk, and ensure smooth operation of the EMS.

Keywords:

Regressions, Analysis, Workload, Characterizations, Examination, Automation, Database.

Introduction:

Currently, Universities are called modern academic institutions since database and web servers are employed as the backbone of IT infrastructure, where a large number of processing and data handling concerning students, lecturers, administrative staff, academic records, and organizational functions. The functions of these servers are further complexed and enhanced when population of information oriented dynamically increased [13]. Moreover, these servers enable the academic and administrative human resources to communicate information to function smoothly. Resultantly, a strong and highly scalable database structure is needed for the educational system in order to achieve advanced technology needs, which supports high standards in the fields of administration, teaching and research as well.

In the modern technology era, it is very important to work smartly rather than working on traditional systems that leads any organization to lengthy procedures and inaccuracy. This study is focusing on examination system therefore computerization of examination record keeping, tabulation and results are very important to record digitally. The end-users of any examination system may be the students, teachers or administrators are using this system frequently, it simplifies the examination process and maintain the tabulation process of marks by entering award lists in the system digitally and expedites the result announcement process. While designing the system structure, it is necessary to include workload assessment for the given period of time that we are going to implement anywhere. As this process is very important as data grows, we face difficulty of manage system complexity. An examination system handles student records and online assessments. It slows down with the increase in data with time. To handle these types of performance degradation we need proactive plans to tune it. To mitigate the workload complexities, we need to use difference tools to identify bottlenecks in the system and analyze the increasing impact of current system through some tools and techniques. It is also necessary of predict the workload for the future maintenance of the system, the factors can be involved in this load may be complexity of system structure, workload assessment, end-user demands to enhance or make a change in the system. The objective to face these challenges is to assess workload and mitigate the performance issues on timely manner [12]. This approach will make the system to work more efficient, reliable and scalable. This approach is to be followed by examination system to work smoothly and allows end-users fast and robust environment to work on.

Any system can be assessed during peak working times and at that slot it may encounters performance deterioration [11] Major workload may lead to slow down the EMS when there are maximum academic activities in a system. During the examination season the number of end-users like students, teachers or administrator login simultaneously and performs huge tasks and examination system personals also executes the multiple jobs to entering award lists and perform tabulation process to generate ledgers, it may decline the overall performance, resultantly users

may face slow throughput of system. In this case the examination system of Public Sector University needs to evaluate the performance impact and classify it to assess the deep insights and identify the actual cause that impacts the load on system. These problems may be caused by insufficient resources that need to be addressed. This study is total focusing on diagnosis of aforementioned issues of examination system to handle and optimize such system where the load is on a peak. Smoothly execution of system is a basic target of any digital infrastructure in this regards there are different approaches may be applying to make the system able to run any task with efficiency and accuracy. There are so many approaches to assess and analyze the load. In this study we use workload assessment by which we get the log files of system where the connection logs generated by examination system. These log files contain detailed information like timestamp, IP addresses of clients where from the request has been initiated, address of webpage accessed, HTTP method (GET or POST etc.), HTTP Status codes, bytes transmit and receives etc. these factors are necessary to identify the workload on system.

The study in concerning about the performance enhancement examination management system which is using database as a back-end installed at a public sector university. The goals regarding this research are listed as: (a) to study the workload brought by various processes in the EMS, such as result compilation, ledger generation, and grade input, during normal and heavy usage periods. (b) for learning about query, transaction, and user activity distribution so that one can see how the system works under various conditions and determine what resources are affected the most due to changes in workload. (c) in order to quantify the database server's performance during periods of peak academic activity such as result finalization, report generation, and grading. (d) to identify performance bottlenecks in the system, including CPU utilization, memory capacity, slow query execution, and other factors affecting system efficiency and responsiveness. (e) in order to use statistical models and data mining algorithms to analyze historical data and predict future patterns of workload. (f) to create a precise model to forecast peak load times, transaction volume, and database query activity to improve capacity and resource planning. Following areas have been identified to go through in this research, initiating with findings and then analyze the (EMS) database server at the University of Sindh, Jamshoro, this includes: We examine and define the activities carried out in examination system of this university like compilation of results of all academic years, award list entries, generating ledger entries either manual of job execution etc. The performance, efficiency, and scalability of the database server both during peak and off-peak periods of the academic year will be tested.

We organize this paper in the major three sections. In section I, the literature review is addressed, while Section II describe the research methodology and the real-world dataset. In the last, Section III discuss the results collected from workload characterization and prediction regression model.

Literature Review

This section critically examines the related work, summarized in Table-1 with the research data, techniques tools, the domains and key findings.

Table 1. Summary of Related Research on Workload Characterization in DBMS

Research	Research Data	Techniques Used	Domain	Key Findings
Zhibin et al. [1]	Real-time log data from OLTP systems	Temporal and statistical analysis	Financial & Enterprise DBs	Temporal trends and bursts affect performance; real-time monitoring helps preemptively optimize systems.
Raza et al. [2]	Query logs from university database servers	Profiling and clustering techniques	Academic/University systems	Different user groups impose varied workloads; role-based optimization enhances performance.
Brown & Patterson [3]	Synthetic workloads & TPC-C benchmark	Benchmarking, regression modeling	E-commerce, transactional DBs	Workload thresholds can guide indexing and caching for performance tuning.
Zhang et al. [4]	Query logs from cloud-based DBaaS	Machine Learning (SVM, clustering)	Cloud databases	ML-based prediction improves auto-scaling and resource allocation.
Kumar & Singh [5]	University examination system logs	Heatmap analysis, time-based profiling	Academic ERP systems	Identified peak load periods; most load aligns with exams and results.
Lee et al. [6]	Logs from healthcare information systems	Workload segmentation, anomaly detection	Medical/Healthcare DBs	Early classification of workload anomalies helps maintain system stability.
Alam & Shah [7]	Database logs from a Pakistani public university	System monitoring, log analysis	Student Information System (SIS)	Performance lag observed during registration and grading periods; limited to general system load.
Kumar et al. [8]	SQL Profiler logs from a technical university	Query analysis, slow query identification	Academic ERP	Identified slow queries but did not relate performance to user roles or time-based variation.
Sultana et al. [9]	User logs from private university LMS	Time-based profiling, usage tracking	Learning Management System	Peak usage before exams caused system lag; suggested dynamic scaling.
Mehta & Joshi [10]	ERP system logs from an Indian university	Module-level query profiling	Academic ERP	Grading and finance modules had high latency; recommended query and index optimization.

Zhibin et al. in [1], used a real-time log data collected from online transaction processing (OLTP) to provide a comprehensive workload analysis. In the findings, the authors revealed that the system performance critically effected by temporal patterns and workload surges. Raza et al. in [2] conducted the study on query logs collected from the university database servers to analyze the different kind of human resources involved in interaction with academic procedure. The techniques

related to profiling and clustering were used. In the findings, authors revealed that diversity in workload is a real issue in academic databases. Different categories of users have different access patterns and query types. Role-based workload optimization was highly effective in improving database responsiveness and user satisfaction.

The study of Brown & Patterson in [3], in order to simulate typical e-commerce and transactional database system loads of normal transactions, synthetic workloads and TPC-C benchmarks were employed, and benchmarking and regression modeling were used in this research to quantify performance across various simulated loads. Findings: They found that the users' workload levels significantly affect system behavior. Their most significant contribution was in finding optimal caching and indexing parameters that could be preconfigured according to predicted workloads.

Zhang et al. [4] published a paper on cloud Database-as-a-Service (DBaaS) query log analysis. The authors predicted cloud database workload patterns using machine learning models like Support Vector Machines (SVM) and clustering. The authors provided the findings that the predictive models could efficiently predict workload spikes and further it enables the efficient auto-scaling and dynamic resource allocation.

The studies including [5-13] are related to our study that characterize the workload in different aspects in order to provide the recommendation and prediction as shown in the above table. The above cited studies provided valuable contributions in the domain, they provided (a) analysis of the real-time impact of concurrent users (b) predictive modeling to forecast future workload trends (c) role-based workload optimization strategies. However, it remains a gap in domain-specific workload studies particularly in examination systems. We provide workload characterization and predictive analysis of examination system in the University.

Methodology:

In this section, we provide a complete description of methodology including analysis and characterization of the workload, which is collected from a real-world application. In the last, we use the regression models to prediction of workload.

Data Collection

In this study, the used dataset is collected from a real-world application known as examination management system (EMS) of the University of Sindh, Jamshoro. The complete description of the workload is shown in Table 2. The key variables of the datasets are: HTTP Method: (GET/POST), Response Status Code:(200, 404, 304, etc.), User Roles: (Faculty, Coordinator, Admin), Execution

Time and Request Frequency

Table 2 Data Collection

Source:	Apache server log files generated by the EMS web interface and backend database server.
Format:	Standardized HTTP log entries including IP address, timestamp, request method, URL, protocol version, status code, and response size.
Period:	1st January 2020 to 21st January 2020 — a critical timeframe covering result finalization and examination periods.
Total Records:	442,436 log entries.
GET	4,099
POST	438,337
Status 200	439,426
Status 404	2,782
Status 304	228
Unique Users	9

Data Preprocessing and Tools

A comprehensive preprocessing technique was implemented. In the parsing and filtering process, we used GSPLIT and MS Excell to split the log files to processing and storing purpose. In the cleaning process, we have cleaned the log files by removing irrelevant entries with proper time stamps. In the structuring and field extracting process, Timestamp, HTTP Method, Status Code, Response Size, URL, are considered. For compatibility process, final dataset was structured in CSV format with MS Excel, and WEKA. Table 3 shows the tools and platforms were used in the study.

Table 3 Tools and Software Utilized

Category	Tools / Platforms
Log Parsing & Cleaning	Log Expert, GSPLIT
Statistical Analysis	R Language, Microsoft Excel
Data Regression	WEKA
Visualization	Tableau, Microsoft Power BI
Database Interaction	Oracle 11g, TOAD for Oracle, Oracle AWR Reports
System Monitoring	Oracle V\$ Dynamic Performance Views, Trace Files

Regression Model:

Regression is a statistical technique to determine the relationship between independent and dependent variables and it is also used for prediction. In this study, we used linear regression. The general form of regression is:

$$y = \alpha + \beta x$$

Where,

y = dependent variable

α = intercept (value of y when x=0)

β = slope

x = independent variable

Regression Analysis Results and Discussion

In this section, we present the findings revealed from the analysis of workload collected from the database servers of EMS of the University of Sindh.

The complete database activities of workday are depicted in Figure 1. The figure reveals that the clear peaks in server activity are happened during 9:00 AM to 12:00 PM. The day is particularly dedicated for scheduling the result compilation and ledger generation.

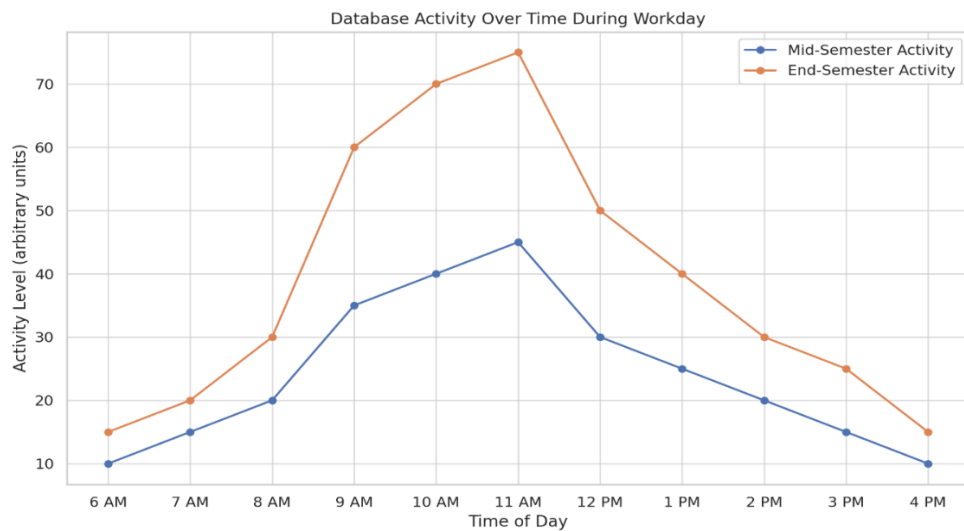


Figure 1. Database activity over time workday

Overall Workload: In this section, the overall workload is depicted in Fig. 2. In this figure, x-coordinate shows number of request/ hits and y-coordinates shows total sum of data requested by each request.

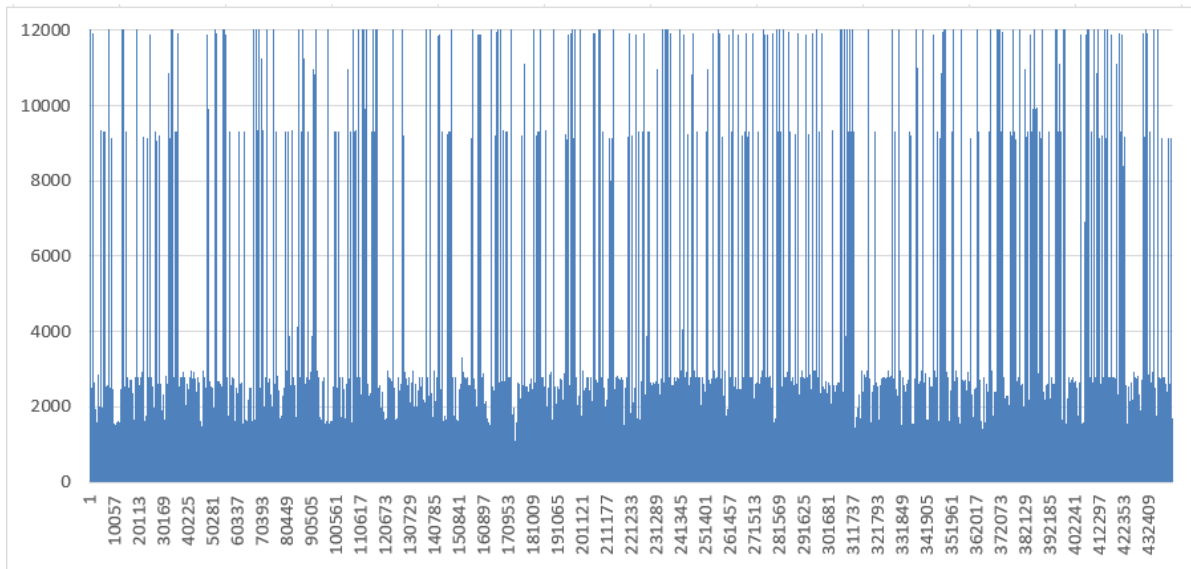


Figure 2. Workload in data (bytes) format by each request

Workload distribution based on Users: In this section, the overall workload based on hits by each user is depicted in Fig. 3, while the overall workload based on data required by each user is depicted in Fig. 4. In the figure 3, x-coordinate shows number of users and y-coordinates shows total hits requested in thousand. Furthermore, In the figure 4, x-coordinate shows number of users and y-coordinates shows sum of data requested in thousand.

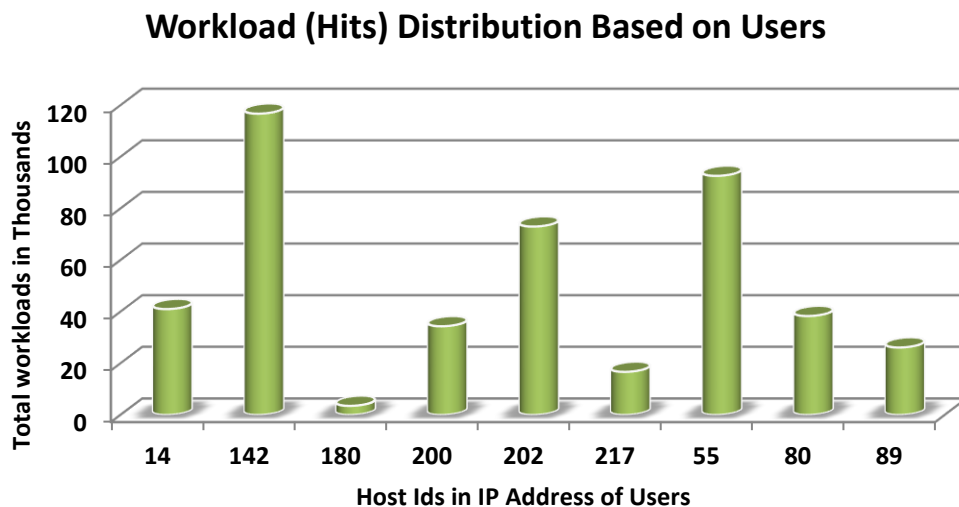


Figure 3 Workload (Hits) Distribution Based on Users

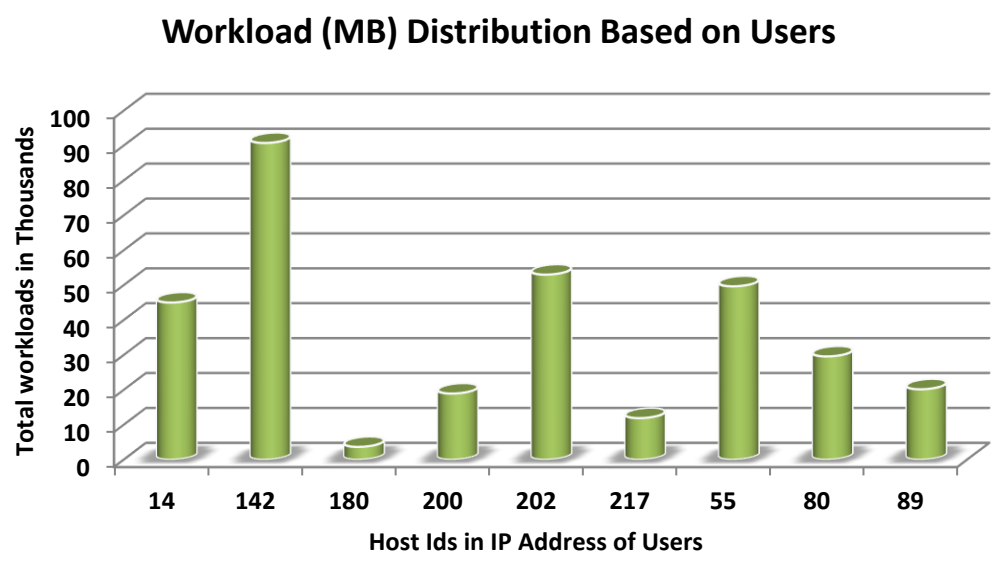


Figure 4 Workload (MB) Distribution Based on Users

Daily Workload distribution based on Users: In this section, the overall Daily workload based on hits by each user is depicted in Fig. 5, while the overall daily workload based on data required by each user is depicted in Fig. 6. In the figure 5, x-coordinate shows hours and y-coordinates shows total hits requested in thousand. Furthermore, In the figure 6, x-coordinate shows hours and y-coordinates shows sum of data requested in thousand.

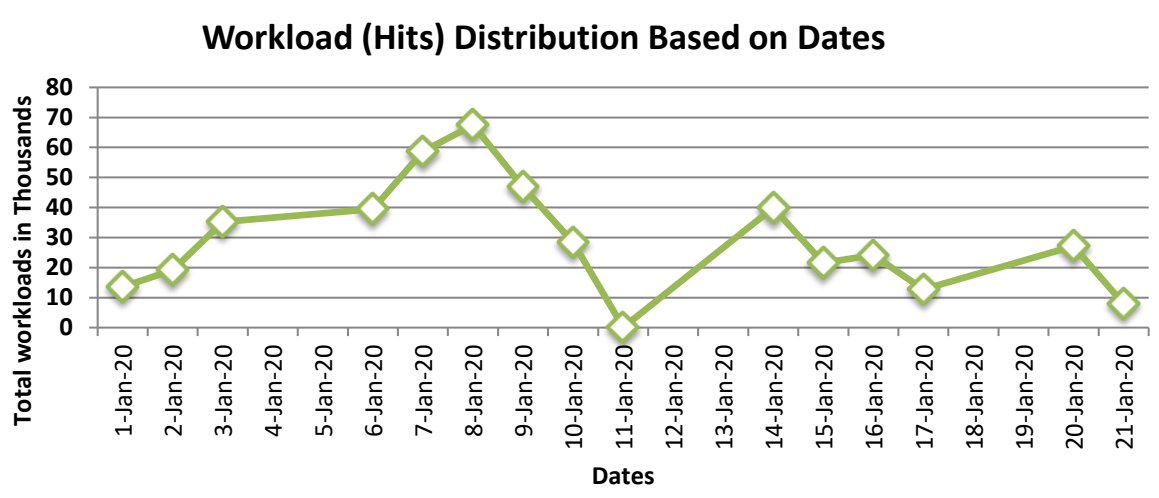


Figure 5. Hourly Workload (Hits) Distribution Based on Dates

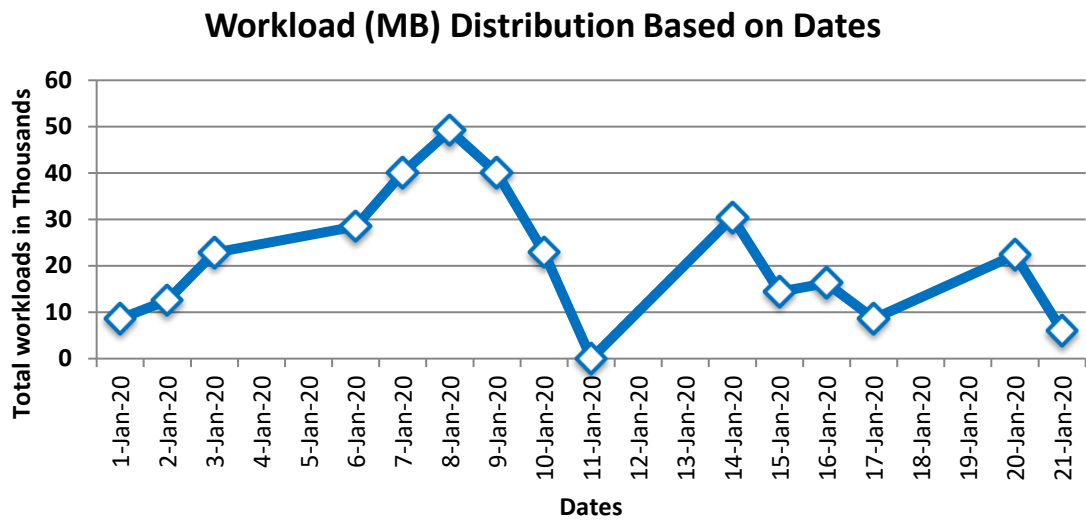


Figure 6. Hourly Workload Data (MB) Based on Dates

Hourly Workload distribution based on Users: In this section, the overall hourly workload based on hits by each user is depicted in Fig. 7, while the overall hourly workload based on data required by each user is depicted in Fig. 8. In the figure 7, x-coordinate shows hours and y-coordinates shows total hits requested in thousand. Furthermore, In the figure 8, x-coordinate shows hours and y-coordinates shows sum of data requested in thousand.

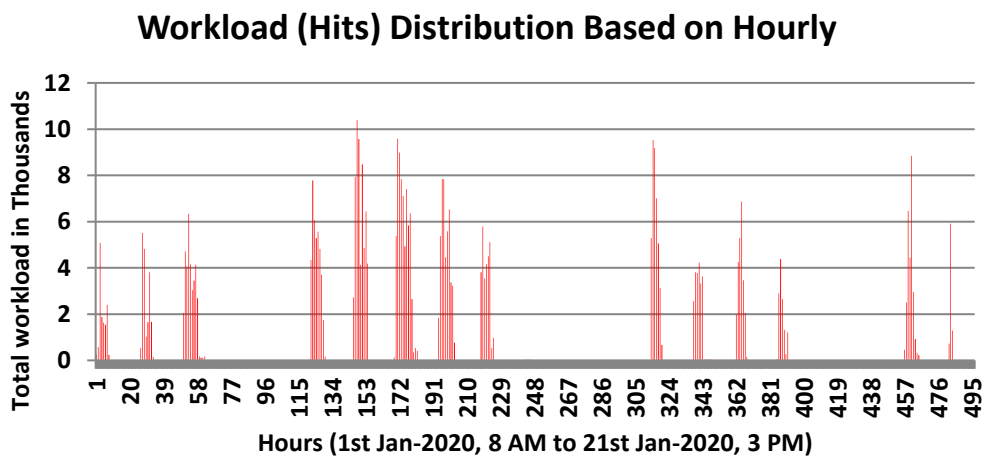


Figure 7. Hourly Workload Data (Hits) Based on Dates

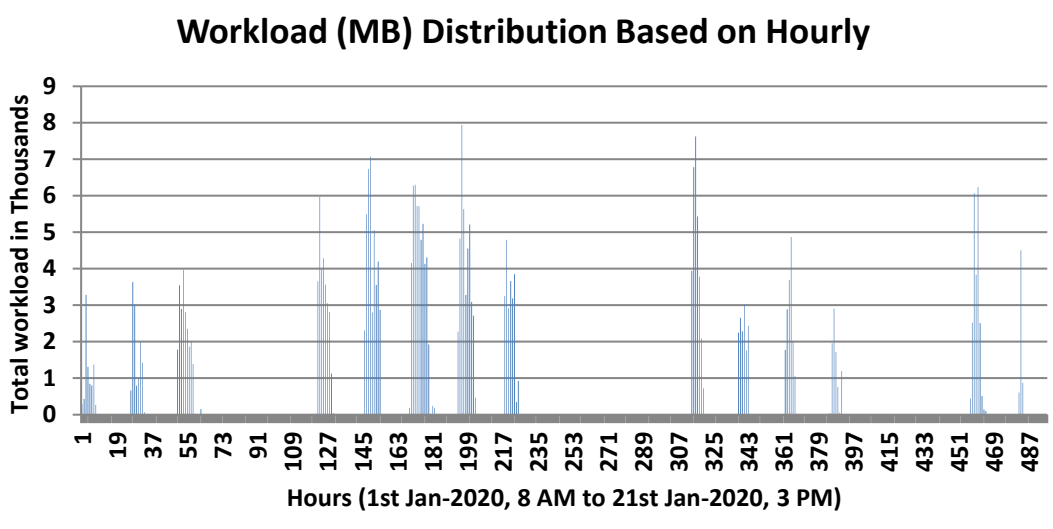


Figure 8. Hourly Workload Data (MB) Based on Dates

Overall Workload Count Distribution: In this section, the overall workload data distribution based on user hits the system is depicted in Fig. 9. In this figure, x-coordinate shows categories of data in MB and y-coordinates shows total number of hits requested by users.

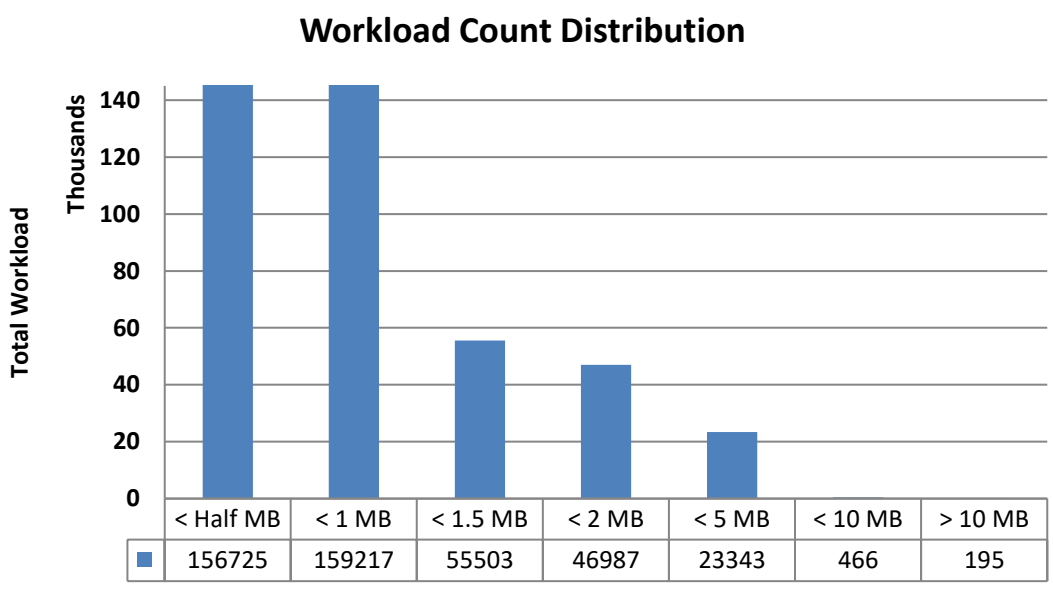


Figure 9. Workload data distribution-based user hits the system

REGRESSION ANALYSIS: The fig. 10 shows the results of regression method employed on two variables i.e., number of users in specific time shift, 8:00 to 15:00 and total sum of data requested

in the specific time shift, 8:00 to 15:00. In the figure, x shows number of hits and y shows total sum of data. The results revealed that r square shows strong relation between both variables.

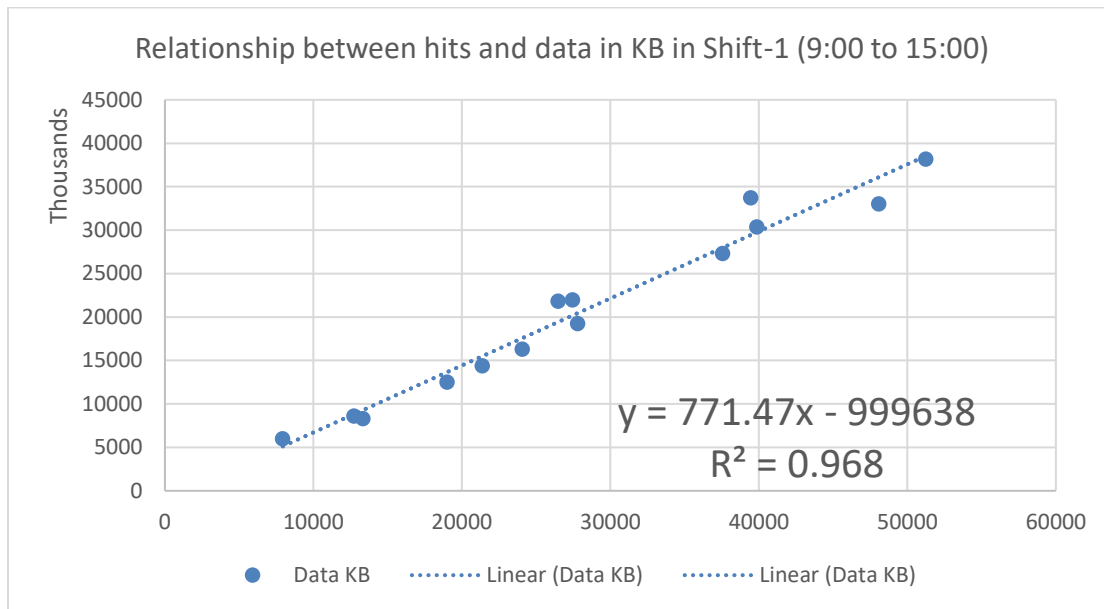


Figure 10 Relationship between hits and data in KB in Shift-1 (9:00 to 15:00)

Conclusion

The study provided a characterization of the examination database server workload. It identified a key performance bottleneck, and provided workload forecasts using regression models. The highlighted insights can directly support future performance improvements and inform server infrastructure planning in academic institutions. The workload on the examination database server is noticeable on the peak time during key activities such as result compilation, mark entry, and ledger generation, especially between 9 AM and 12 PM. This research has contributed to a better understanding of the workload dynamics of an academic examination database server. The insights gained and recommendations proposed aim to support university IT departments in managing their examination systems more efficiently and reliably, both in the short and long term. The collected and preprocessed log data was analyzed to understand the nature and intensity of operations performed on the examination database server.

Acknowledgment

Junaid Fazlani 's work was supported for his MPhil studies at University of Sindh, Jamshoro, Pakistan. The Log Dataset received from the ITSC University of Sindh Jamshoro granted NOC approval to carry out the study.

References

- [1] References Zhibin, W., Zhang, Y., & Chen, H. (2015). Workload analysis in OLTP systems using real-time log mining. *Journal of Database Systems*, **32**(4), 245–259.
- [2] Raza, S., Ahmed, F., & Jamil, M. (2017). Role-based workload profiling for academic database systems. *International Journal of Information Systems and Education*, **25**(2), 101–112.
- [3] Brown, A., & Patterson, D. (2012). Benchmarking and regression modeling of synthetic workloads in e-commerce databases. *ACM Transactions on Database Systems (TODS)*, **37**(3), 54–70.
- [4] Zhang, T., Li, X., & Wang, J. (2016). Predictive modeling of workload patterns in cloud databases using machine learning. *IEEE Transactions on Cloud Computing*, **4**(2), 131–142.
- [5] Kumar, R., & Singh, A. (2018). Visual analysis of examination server logs using heatmaps and temporal profiling. *Journal of Educational Technology and Systems*, **29**(1), 87–95.
- [6] Lee, D., Park, J., & Kim, S. (2020). Anomaly detection in healthcare databases using workload segmentation. *Health Information Science and Systems*, **8**(3), 110–122.
- [7] Alam, M., & Shah, N. (2018). Impact of academic activities on performance of student information systems: A case study. *Pakistan Journal of Computer and Information Systems*, **15**(2), 59–68.
- [8] Kumar, V., Sharma, R., & Das, A. (2020). SQL profiler-based performance assessment of academic ERP systems. *Indian Journal of Information Systems*, **12**(4), 173–180.
- [9] Sultana, R., Khan, M., & Imran, A. (2019). Temporal analysis of LMS usage in private universities. *International Journal of E-Learning and Educational Technologies*, **9**(1), 35–44.
- [10] Farooq, N., Ahmed, S., & Malik, U. (2022). Performance evaluation of examination systems using load balancing techniques. *Journal of Higher Education Computing Research*, **19**(2), 89–101.
- [11] Qureshi, Saleem Razzak, Chandio, Aftab Ahmed and Chandio, Qamar-ul-Nisa. " Characterization and Forecasting the Workload of DRLAB Medical Database Server Based on the Shift-Wise and Machine Learning Approaches." *The Science tech* 7, no. 1 (2026).
- [12] Bhutto, Afia, Aftab Ahmed Chandio, Kirshan Kumar Luhano, and Imtiaz Ali Korejo. "Analysis of Energy and Network Cost Effectiveness of Scheduling Strategies in Datacentre." *CYBERNETICS AND INFORMATION TECHNOLOGIES* 23, no. 3 (2023).
- [13] Jhatial, Soomal, & Chandio, Aftab Ahmed. "Analysis of Academic Web Server Traffic and Workload Characterization for Performance Evaluation." *Sindh University Research Journal - SURJ (Science Series)*, 55(02), 45-55. (2023) <https://doi.org/10.26692/surj-ss.v55i02.6491>