



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established by Government of Gujarat under Gujarat Act No. : 20 of 2007)

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી

(ગુજરાત સરકારના ગુજરાત અધિનિયમ ક્રમાંક : ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Accredited with A++ Grade by NAAC

Abstract of the Thesis



Name: Naimisha Shashikantbhai Trivedi

Enrollment No.: 179999913019

Discipline/Branch: Computer/IT Engineering

Title of the Thesis: Proactive Auto-Scaling Of Containerized Applications
In Cloud Environment Using Ensemble-Based Workload Prediction

Abstract

Cloud-native applications increasingly rely on microservices and containerized deployment models to achieve scalability, flexibility, and efficient resource utilization. In such environments, application workloads are highly dynamic and may fluctuate rapidly due to bursty traffic, periodic demand cycles, or continuous growth in user requests. Managing resources elastically is therefore essential to maintain Quality of Service (QoS) and ensure reliable application performance. Traditional reactive auto-scaling mechanisms, which provision additional container replicas only after predefined utilization thresholds are exceeded, are often limited by scaling lag. This delay can lead to transient Service Level Agreement (SLA) violations, increased latency, request rejection, and inefficient utilization of computing resources. This thesis presents a proactive auto-scaling framework for containerized microservices that uses workload prediction to provision container replicas before demand surges occur.

The proposed framework integrates a clustering-enhanced hybrid ensemble workload prediction model with a MAPE-K (Monitor, Analyze, Plan, Execute, and Knowledge) control loop deployed on Docker Swarm. The hybrid prediction model combines K-means clustering, Prophet, Long Short-Term Memory (LSTM), and Linear Regression, using inverse-error-based weighting to capture seasonal, non-linear, and steady workload patterns. The prediction model is evaluated using RMSE, MAE, MAPE, and R^2 metrics,



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established by Government of Gujarat under Gujarat Act No. : 20 of 2007)

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી
(ગુજરાત સરકારના ગુજરાત અધિનિયમ ક્રમાંક : ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Accredited with A++ Grade by NAAC

while auto-scaling performance is assessed in terms of SLA compliance, latency, request rejection, throughput, replica count, and scaling actions. Experiments are conducted using two real workload traces, which include NASA HTTP logs and ClarkNet WWW traces, and two synthetic workload patterns that have monotonically increasing and periodic traffic patterns to evaluate the effectiveness of the proposed framework.

The results show that the proposed proactive auto-scaling framework consistently outperforms reactive auto-scaling across all workload patterns. Compared with the reactive approach, it improves SLA compliance by approximately 5.3% and reduces mean P95 latency by 6.5–43.3%. Request rejection is also reduced by 68.2% for the NASA workload, 47.8% for ClarkNet, 9.9% for the increasing workload, and 31.2% for the periodic workload. These improvements are achieved with only a moderate resource overhead of approximately 3% on average across all workload patterns compared with the reactive approach. The hybrid ensemble prediction model achieves an MAE of 2.492363, an RMSE of 3.415395, and an R^2 of 0.859265 on the NASA dataset, and the lowest MAE of 5.806198 and RMSE of 7.3380937 on the ClarkNet dataset among the compared baseline models. These results confirm that proactive auto-scaling reduces scaling delay, improves application performance, and supports efficient resource management in containerized microservice environments.

This PhD Thesis would be useful for researchers, cloud architects, and DevOps engineers working on proactive auto-scaling, workload prediction, and efficient resource management of containerized microservices in dynamic cloud environments.

List of Publications

1. Naimisha S Trivedi & Ajay N. Upadhyaya. (2026). "Hybrid Ensemble Approach for Accurate Workload Prediction in Dynamic Cloud Environments." *SSRG International Journal of Electronics and Communication Engineering*, 13(2), 180–194.
<https://doi.org/10.14445/23488549/IJECE-V13I2P114>

e

Winn

Rank"

❖ "1" Rank in Gujarat State Institutional Ratings Framework (GSIR) 2019 ❖ National Leadership Excellence Award 2019 in Technical Education" ❖ "India's Most Trusted State University 2019" ❖ "Outstanding University in West India 2019"

Head office : GTU Campus, Nr. Visat Three Road , Visat-Gandhinagar Road , Chandkheda, Ahmedabad - 382 424. Gujarat, India.

Phone :- +91 079-23267521/570

e-mail : info@gtu.ac.in

Website : www.gtu.ac.in



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established by Government of Gujarat under Gujarat Act No. : 20 of 2007)

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી

(ગુજરાત સરકારના ગુજરાત અધિનિયમ ક્રમાંક : ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Accredited with A++ Grade by NAAC

2. Naimisha S Trivedi & Shailesh D. Panchal (2023). "Optimization Enabled Elastic Scaling in Cloud Based on Predicted Load for Resource Management." *Multiagent and Grid Systems*, 19(4), 289–311. <https://doi.org/10.3233/MGS-230003>



Winners of: "Most Promising University of India 2020" ❖ "1st rank Among All Universities in Gujarat 2020 by Uni Rank"

❖ "1st Rank in Gujarat State Institutional Ratings Framework (GSIRF) 2019" ❖ "National Leadership Excellence Award 2019 in Technical Education" ❖ "India's Most Trusted State University 2019" ❖ "Outstanding University in West India 2019"

Head office : GTU Campus, Nr. Visat Three Road , Visat-Gandhinagar Road , Chandkheda, Ahmedabad - 382 424. Gujarat, India.

Phone :- +91 079-23267521/570

e-mail : info@gtu.ac.in

Website : www.gtu.ac.in