AI-driven cooling optimization in data centers: Reinforcement learning for dynamic workload placement and HVAC control â?? Volume 12 issue 5

Description

International Journal of Computer Techniques

ISSN 2394-2231

default watermark

DOI Registered

Author

Deepak Tomar , Kismat Chhillar , Saurabh Shrivastava , Alok Verma

Abstract

The rapid expansion of data centers has led to unprecedented energy demands, with cooling systems accounting for a significant portion of overall power consumption. Traditional rule-based methods for workload placement and HVAC (Heating, Ventilation, and Air Conditioning) management often fail to adapt dynamically to fluctuating workloads and thermal profiles, leading to inefficiencies and increased operational costs. This paper proposes an AI-driven framework leveraging reinforcement learning (RL) to jointly optimize workload distribution across servers and fine-tune cooling parameters in real time. By modeling the data center environment as a dynamic system, RL agents learn adaptive policies that minimize power usage effectiveness (PUE) while ensuring service-level agreement (SLA) compliance. Experimental evaluations using simulation-based workload traces demonstrate that the