National Institute of Technology Rourkela

Synopsis Seminar

Seminar Title	: Resource Management Strategies in CaaS Cloud
Speaker	: Manoj Kumar Patra (Rollno : 519cs1001)
Supervisor	: Prof. Bibhudatta Sahoo
Venue	: CS 323, 2nd Floor, Department of CSE
Date and Time	: 15 May 2024 (5.00PM)
Abstract	: Container-as-a-Service (CaaS) has emerged as a prominent cloud computing paradigm, providing developers with a convenient platform for deploying and managing containerized applications. In CaaS environments, efficient resource management is crucial for optimising performance, minimising costs, and ensuring the timely execution of tasks. Makespan, the total time required to complete tasks or jobs, is a critical metric for evaluating resource utilisation and workload efficiency. This research explores makespan-aware resource management strategies tailored explicitly for CaaS cloud environments, thoroughly analyzing CaaS clouds and emphasising their salient features, benefits, and drawbacks. It presents the architecture of CaaS systems and the basic ideas behind containerization, including the advantages of resource isolation, scalability, and portability in CaaS when deploying and managing applications using containers. The main objective of this work is to propose different approaches for minimizing makespan in the CaaS cloud while maintaining the required Quality of Service (QoS). Improvements in resource use at the server and virtual machine levels help to achieve the goal. First, a meta-heuristic approach for load balancing in the CaaS cloud has been presented to distribute incoming workload across available resources in a balanced manner, minimizing resource contention, makespan, and optimizing resource utilization. Next, a virtual machine (VM) sizing technique based on fractional pelican optimization was presented, which uses Deep-ConvLSTM to minimize the makespan, task rejection rate, and response time. Then, we came up with a Fractional Pelican Hawks Optimization (FPHO) based container consolidation method to enhance the system performance, where energy consumption, resource utilization, SLA violations, and makespan are considered the performance metrics. The simulation outcome in all three approaches signifies the better performance of the cloud system, and, in most cases, more than 97% of the tasks meet their deadline