SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(IoT and Cybersecurity Including BCT)

COURSE NAME : cloud service management

IV YEAR / VII SEMESTER

Unit II-

Topic : Demand and Capacity matching

- Cloud computing offers many benefits, such as scalability, flexibility, and cost-efficiency. However, to fully leverage these advantages, you need to match your cloud demand and capacity. This means ensuring that you have enough resources to handle your workload, but not too much that you waste money and energy. In this article, you will learn some strategies and tools to help you achieve this balance.

Why is it important?
- Matching cloud demand and capacity is important for several reasons. First, it can improve your performance and user experience, by avoiding issues such as latency, downtime, or errors. Second, it can reduce your costs, by optimizing your resource utilization and avoiding overprovisioning or underprovisioning. Third, it can enhance your sustainability, by minimizing your environmental impact and carbon footprint.

What are the challenges?
- Matching cloud demand and capacity is not always an easy task, due to the various factors and variables involved.
- To successfully do so, you must first estimate how much traffic, data, and processing power you will need at different times and scenarios, which can be affected by seasonality, trends, events, or user behavior.
- Then you need to decide how much resources to allocate to your cloud services such as compute, storage, network, or memory, taking into account availability, performance, reliability, or security.

- **Determine which infrastructural and application architectures align with your needs**.

- For instance, if you have a virtual server with routinely fluctuating capacity demands, you might find that serverless functions would be a better way to host that workload.

- Serverless functions allow you to allocate large amounts of resources for short periods in a more cost-effective and easy-to-manage way than is possible with virtual servers. **Arrange the right people and tools**.

- This is a step beyond your team knowing how many resources to allocate to workloads. It's important to find out if you have the organizational resources necessary to assign those resources.

- [You'll need staff on hand](#) to perform the necessary provisioning, and those workers should have the requisite skills to work with the tools you use to manage resource allocation.

- **Avoid disruptions to users**. Wrong-sized workloads can create problems for the people who expect a specific application to be ready for them when they need it.

- When your workload capacities are well managed, you minimize your risk of having applications or servers fail.