**CH 5: Cloud Resource management**

**1 Resource - Provisioning**

Identify adequate resources for a given workload based on QOS requirements described by consumers.

**Resource Discovery:** Identifies suitable physical resources in which the virtual machines are to be created matching user's request.

**Resource Selection:** Selecting the most appropriate resource from the practical set of resources **(**multi-criteria decision-making problem).

availability, trust, cost, responsiveness, reliability, and capability effects resource selection.

**Virtual Resource Pools Integration**

**XenServer Pool:** Allows resource provisioning over private Cloud.

**VMWare Pool:** Allows resource provisioning over private Cloud.

**Amazon EC2 Pool:** Allows resource provisioning over public Cloud.

**2. Resource Scheduling**

Mapping and execution consumer workloads on available VMs based on selected resources through provisioning**.**

Goal: Maximize System utilization and Overall performance

Effective resource scheduling reduces execution cost, execution time, energy and other QoS: reliability, security, availability, and scalability could be considered.

**Scheduling Prospective:**

Provider wants to earn as much as possible with lowest investment and maximize utilization of resources.

Consumer wants to execute workload(s) with minimum cost, minimum execution time, and guarantee QoS.

**3.Resource Monitoring**

Monitoring the performances of both physical and virtual infrastructure during the operation. security monitoring to achieve confidentiality, integrity, and availability of data.

﻿﻿**Resource Management and Scheduling** affects→ Functionality, Performance, Cost.

**Scheduling in a computing system**: deciding how to allocate resources of a system: CPU cycles, memory, secondary storage space, I/O, and network bandwidth; among users' workloads.

Policies and mechanisms for resource allocation.

Policy→ principles guiding decisions. Mechanisms→ the means to implement policies.

A diagram of a resource planning

Description automatically generated

**﻿Tradeoffs**

To reduce cost and save energy we may need to concentrate the load on fewer servers rather than balance load among them, execution Time. And operate at a lower clock rate; the performance decreases at a lower rate than does the energy.

**﻿Scheduling**

PaaS model: Workflow (job) Scheduling on available VMs

IaaS model: Virtual Machines (VM) Scheduling on available Physical Machines (PM)

Effective scheduler: Reduce operational costs, queue waiting time, Increase resource utilization.

**Resource Scheduling Vs. Resources Allocation**

**Allocation:** is the assignation or reservation of resources at the time of request.

**Scheduling:** is a request for allocation of resources at a specific time or time period.

The result of schedule request should result in the allocation of cloud resources. So, cloud resources should be ready and available at the scheduled time.

**Cloud Scheduling Algorithms**

A server can be shared among several virtual machines.

A virtual machine could support several applications.

An application may consist of multiple threads (tasks).

A scheduling algorithm should be efficient, fair, and starvation- free.

**• The objectives of a scheduler:**

**Batch system** → maximize throughput and minimize turnaround time.

**Real-time system** → meet the deadlines and be predictable.

**Common algorithms for best effort applications:**

1.**Round-robin**: assigns a fixed time slice to each task in a cyclic order. After a task has used up its allocated time, it is moved to the back of the queue, and the next task in the queue is executed.

2.**First-Come-First-Serve (FCFS):** tasks are processed in the order they arrive.

3.**Shortest-Job-First (SJF):** selects task with the shortest execution time to run. non-preemptive (once task starts, it runs to completion), preemptive (Shortest Remaining Time First SRTF).

4.**Priority algorithms:** assigns a priority level to each task, and the scheduler selects tasks based on their priority. can be preemptive or non-preemptive.

﻿❖ An optimized Task Scheduling Algorithm is needed to design which satisfies:

✓ Proper resource utilization.

✓ Reduces makespan (i.e., Execution Time).

✓ Reduces overhead from resource.

✓ Balance proper load over resource.

✓ Increase throughput and system performance.

✓ Better CPU utilization.

✓ Minimize turnaround, waiting and response time.

✓ Save Energy

**﻿SCHEDULING STEPS**

1.Resource Discovering and Filtering: Data center Broker discovers the resources present in the network system and collects status information related to them.

2. Resource Selection: Target resource is selected based on certain parameters of task and resource.

3. Task Submission: Task is submitted to resource selected.