Hanoi University of Science and Technology
School of Information and Communication Technology

Task Scheduling In Real-time System

Quang Khanh
khanh.tq170083@gmail.com

22nd March 2021

Content

Entities in the Task-Scheduling Problems

Cluster of machines

Cloud computing is built over a group of machines in order to provide services to users.

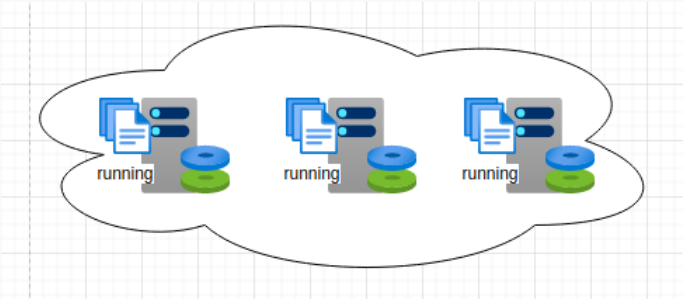

Figure: Virtual machines

Entities in the Task-Scheduling Problems

Tasks

Users submit tasks at random time to the cloud by API web services.

Notice

Users submitting tasks is a stochastic process.

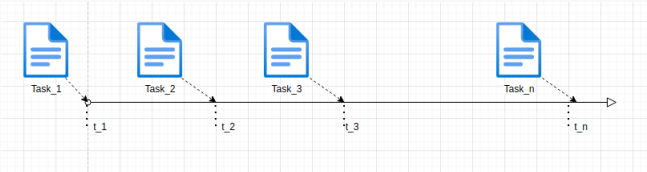

Figure: Task arrivals

Workflow of task scheduling

The order of task scheduling process

- Users submit tasks to the datacenter.
- Tasks are dispatched to waiting-queue.
- Scheduling finds matched VM for each task in waiting-queue.

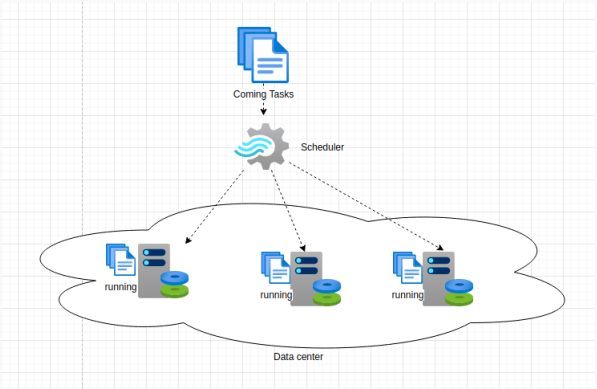

Figure: tasks are scheduled

Workflow of task scheduling

The order of task scheduling process

- Users submit tasks to the datacenter.
- Tasks are dispatched to waiting-queue.
- Scheduling finds matched VM for each task in waiting-queue.
- Tasks are dispatched to matched VMs.

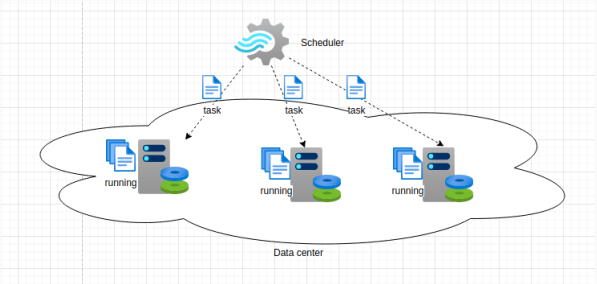

Figure: tasks are dispatched

Mathematical Model

Task i

- cpu_request(i): quantities of cpu unit
- cpu_cores_request(i): number requested cores of cpu
- ram_request(i)
- disk_request(i)
- priority(i)

VM j

- Cpu(j)
- Cpu_cores(j)
- Mips(j)
- Ram(j)
- Disk(j)

Mathematical Model

Waiting queue

$Q = \{T_i, ..., T_j\}$ is the set of waiting tasks

Trace

Trace(i, j) = 1 if task i is matched with machine j

Running queue

$R(M) = \{T_{x_1}, ..., T_{x_k}\}$ is a set of running tasks on the virtual machine M

Defered tasks

Defered(i, j) = 1 if task i is running in machine j then instantly cancelled.

Mathematical Model

Avialable resources of virtual machine j

$$A\_cpu(j) = Cpu(j) - \sum_{i \in R(j)} cpu\_request(i) * (1 - Defered(i, j))$$
$$A\_ram(j) = Ram(j) - \sum_{i \in R(j)} ram\_request(i) * (1 - Defered(i, j))$$
$$A\_disk(j) = Disk(j) - \sum_{i \in R(j)} disk\_request(i) * (1 - Defered(i, j))$$

Mathematical Model

Constraints

$$\forall\ j = 1, \ldots, N_M,$$

$$\sum_{i=1}^{N_T} Trace[i,j] * cpu\_request(i) < A\_cpu(j)$$

$$\sum_{i=1}^{N_T} Trace[i,j] * ram\_request(i) < A\_ram(j)$$

$$\sum_{i=1}^{N_T} Trace[i,j] * disk\_request(i) < A\_disk(j)$$

Mathematical Model

Priority constraint

$\forall i = 1, \ldots, N_T$

$$priority(i) > 6 \implies \sum_{j=1}^{N_M} Defered(i,j) = 0$$

Objectives

Defered Rate

Minimize

$$\frac{\sum_{i,j} Defered(i,j)}{N_T}$$

Make-span

The time from tasks are submitted to all tasks are completed

Problems

Instructions

Cannot estimate the total instructions of a task.

Machines' state

Cannot estimate the states of all machines in the next time step, which leads to ineffective scheduling-plan.

Proposal

To solve the problems

We need a strategy to estimate status of a task after its duration

Correlated features

Features correlated to the status of a task include:

- cpu_request
- ram_request
- MIPS of the machine in which it's executed
- duration

Thank you for listening!