Assignment 7

We intend to design a workflow scheduler to schedule workflows on a given set of cloud instances (or VMs) to minimize the makespan (i.e., the completion time of the whole workflow) subject to a given budget.

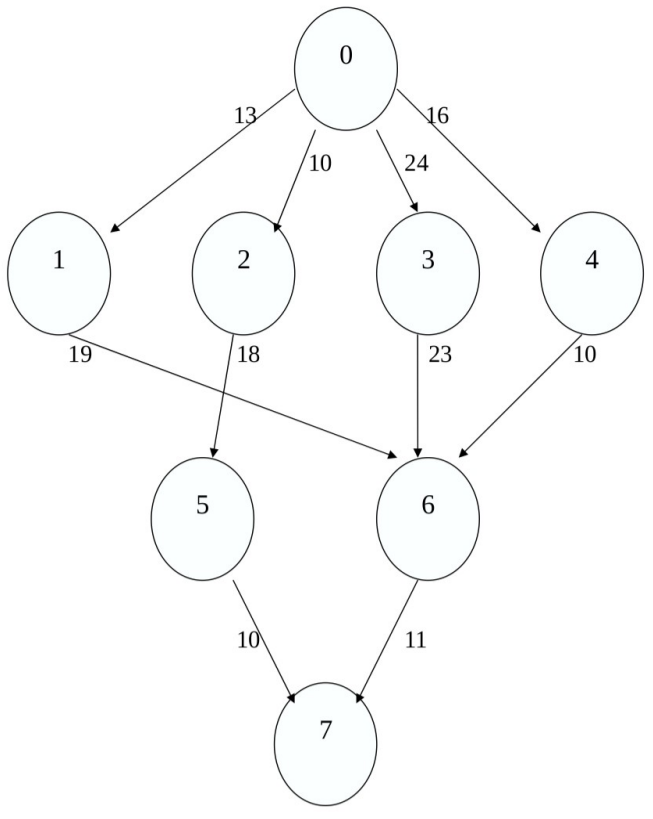
- Inputs:

1. A given workflow in the form of Directed Acyclic Graph (DAG) including a set of jobs and the links between the jobs that indicate the size of data transferring between the tasks (in Byte).
2. The execution time of each task on each instance in second.
3. The network bandwidth between the cloud instances. Let's assume the communication bandwidth between all instances are the same and it is equal to 40 Mbyte/sec.
4. The cost of running each instance in hour/sec. We target an hourly-based cost model, meaning that when an instance run for partial hour (e.g., 1.5 hour), we are charged for the full hour (2 hours). This is the cost model used by Amazon EC2 instances.
5. A given budget.

- Output:

1. The total cost of execution of the workflow, which of course should be lower that the budget.
2. The completion time (makespan) of the workflow.
3. Assignments of jobs to instances.
4. The order of execution of jobs hosting by the same instance.

- Example:
- An example of a DAG is shown below. Each node represents the job number and each link between these job represents the traffic transferred between them.



Assignment 7

- Description of files:
- Here are the required files for this assignment.
  - price.csv:
  - This file includes the price of each cloud instance in $ per hour.
  - task_time_instance.csv:
  - This file includes the execution time(sec) of each job on each cloud instance.
  - traffic_relation.csv:
  - This file shows the amount of data(bytes) that each row’s task sends to each column’s task. If there is no data, it means that the task sends nothing to the other task. This file can also show the parent-child relationship. In other words, if job A send data to job B, A is the parent of the recipient job B.

All input files are available in Github: (https://github.com/inns-apds/assignments/raw/master/assignment7_files.zip).

You can make a group (not more than two students) to do this assignment. The deadline is December 9.