

CMPE 230 FALL 2016 HOMEWORK 2

AMAZON WORKFLOW SCHEDULER

Batuhan Demir

4 ARALIK 2016

1 Specification

Design and implement simple scheduler that executes workflows on Amazon cloud instance. After that sorting workflows topologically and run programs on the remote instance.

2 Project Implementation

My implementation consists of a single Python file that is 'wdfsched.py'. In the implementation, firstly I have parsed the wdf file and created 2 dictionaries.

- **Workflow dictionary** – where keys are name of nodes in wdf file and values are programs to execute. This is the main list of program; it holds workflows topologically and is runned.
- **Adjacency List** – Keys are nodes in graph and values are adjacency nodes. It is used to represent graph and help topologically sorting it.

Also I have used topological sorting algorithm using queue. After done the steps ,now the program is ready to run on Amazon instance. To accomplish this, I have used boto3 library to connect my amazon instance. I transferred files using “scp” and run the program using “ssh”. After the completion done I transferred files back to main machine.

After sorting programs according to topological sorting I have created a temporary bash script to run programs on Amazon instance. This bash file shortened my lines of code instead of individual ssh command lines. But as a drawback I couldn't know when they started and ended, so in verbose mode this functionality is missing.

After creating amazon instance using boto3 I am waiting 60 seconds for instance to run without problems. After everything is settled I transfer workflow directory compressed by tar command. Then I extract the directory on the instance and run my bash script there. After that I transfer the directory back and replace existing one. And finally I terminate the instance.

Implementing verbose was not so difficult. But since I run programs on the instance using bash script that I've created, I couldn't print information on when programs started and ended. But beside that, I have successfully print times of

- Instance created
- Workflow transferred to instance
- Workflow execution is started
- Workflow execution ended
- Workflow is downloaded from instance
- Instance is terminated