CS533 - Homework 1

Ray Framework and Intel® AI DevCloud

1 Preliminaries

The objective of this assignment is to get you setup and familiar with the Intel DevCloud and to give you some basic familiarity with the Ray framework for distributed programming. To work on this assignment, you need to have a DevCloud account.

- **Follow the instructions in the document "DevCloud Setup Instructions"**

In addition to setting up an account the document discusses how to connect to DevCloud through a terminal, upload files and run jobs. You should review this information carefully to be able to do the rest of the homework. Especially note the requirement to pip-install an older version of the Ray framework.

2 Ray Framework Tutorial (75 points)

The objective of this section is to get you started with the Ray Framework. You are provided with initial skeleton code in the file `ray tutorial.py`. This file contains a description of the Ray framework. You should work through the file, read the comments, and complete the sections annotated as EXERCISE. To complete this part of the assignment:

1. Copy `ray_tutorial.py` to your home directory on DevCloud: `scp /path/to/local/file colfax:/path/to/remote/directory/`.

2. Connect to DevCloud login node: `ssh colfax`

3. Connect to a compute node: `qsub -I -lselect=1`

4. Implement the exercises in `ray_tutorial.py`

5. Test your code: `python ray_tutorial.py`

The final implementation should execute successfully and pass all the assertions.

3   Evaluating The Performance of Ray (25 points)

The objective of this section is to measure the performance of Ray for a simple parallel program. A python file named `map_reduce.py` is provided with this assignment. This program generates an array of size 8, increments each element by one, computes the sum of the elements and finally outputs the running time of the program. The only input of the program is the number of the CPUs that it can use. In this part, you should run this program with different numbers of CPUs (1, 2, 4, 8) and report the running time of the program in each case. The running time is printed to the output as `total time:  #`.

4   Deliverables

You should submit a zip file containing 1) the completed `ray_tutorial.py` file, 2) a graph of time against number of CPUs for the `map_reduce.py` program (this can be in png or jpeg format) .