Summary Exploratory Analysis

Peiqiang Li - 200987503

1/19/2022

Business Understanding

The rendering performance of terapixel visualization is consistent with many factors of supercomputer. For example, the temperature, performance and power draw of GPU which constituting supercomputing cluster, and the efficiency of computer task scheduling and so on. Therefore, finding the relationship between the efficiency of rendering task execution and GPU types, performance and power draw is very important to improve the computing power of supercomputers.
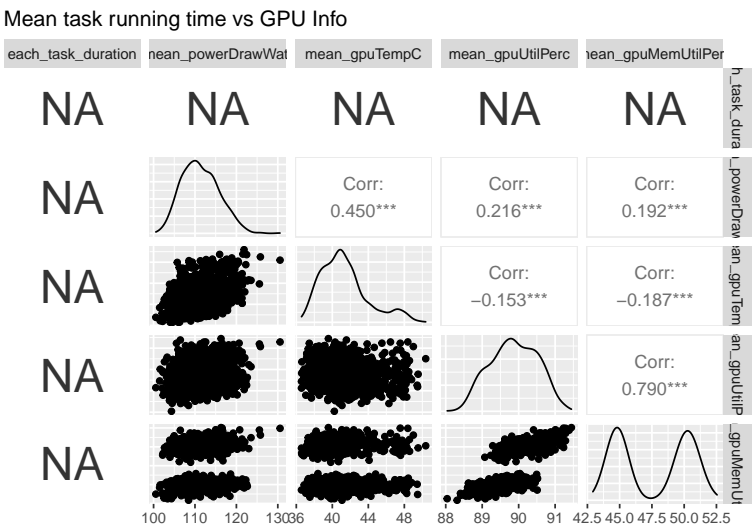
Data Understanding and Preparation

The data of this project comes from the records of 1024 GPU nodes performing rendering tasks during the run. This run is about 3 rendering jobs with levels 4, 8 and 12 of the visualisation output. In addition, details of the application's performance timings, the performance of GPU cards, and details of which part of the image was being rendered in each task are recorded in 3 datasets(`application.checkpoints`, `gpu`, `task.x.y`) when executing this 3 rendering jobs.

After removing duplicate data. We got 657930 records in `application.checkpoints` dataset, 65793 records in `task.x.y` dataset and 1543672 records in `gpu` dataset. Then, we found the following relevant information:

- We have 5 event types: "Tiling", "Saving Config", "Render", "TotalRender", "Uploading".
- The duration of "TotalRender" event is the sum of the durations of the other types.
- Each task corresponds to one part of the image in certain level.
- Each task corresponds to the `START` and `STOP` states of the 5 event types.
- **jobId** "1024-lvl12-7e026be3-5fd0-48ee-b7d1-abd61f747705" is the rendering of level 12. This job has 65536 task records, and 1024 gpu rendering.
- **jobId** " 1024-lvl4-90b0c947-dcfc-4eea-a1ee-efe843b698df"is the rendering of level 4. This job has 1 task records, and 1 gpu rendering.
- **jobId** "1024-lvl8-5ad819e1-fbf2-42e0-8f16-a3baca825a63" is the rendering of level 8. This job has 256 task records, and 256 gpu rendering.

Data Analysis

Mean task running time vs GPU Info

each_task_duration	mean_powerDrawWat	mean_gpuTempC	mean_gpuUtilPerc	mean_gpuMemUtilPer

					h_task_dura
NA	NA	NA	NA	NA	
NA		Corr: 0.450***	Corr: 0.216***	Corr: 0.192***	_powerDraw
NA			Corr: −0.153***	Corr: −0.187***	an_gpuTem
NA				Corr: 0.790***	an_gpuUtilP
NA					_gpuMemUt

100 110 120 13036 40 44 48   88 89 90 91   42.5 45.0 47.5 50.0 52.5

Warning: Removed 328965 rows containing non-finite values (stat_boxplot).

Event types occupy task runtimes(s)

Render   Saving Config   Tiling   TotalRender   Uploading