

02_DataUnderstanding

Data Understanding

The datasets was created during a run using 1024 GPU nodes. There are 3 levels visualisation output; 4, 8, 12.

The data provided shows performance timing of the render application, performance of the GPU card, and details of which part of the image was being rendered in each task.

application-checkpoints.csv (111.2MB)

This file contains application checkpoint events throughout the execution of the render job. Examples of the events are given for the `eventName` field.

Schema

1. **timestamp:** time of data creation. Ex: 2018-11-08T07:41:55.921Z
2. **hostname:** unique hostname created by the Azure batch system. Ex: 0d56a730076643d585f77e00d2d8521a00000N
3. **eventName** Name of the event occuring within the rendering application.
 - **TotalRender** is the entire task
 - **Render** is when the image tile is is being rendered
 - **Saving Config** is simply a measure of configuration overhead
 - **Tiling** is where post processing of the rendered tile is taking place
 - **Uploading** is where the output from post processing is uploaded to Azure Blob Storage
4. **eventType:**
 - **START**
 - **STOP**
5. **jobId:** ID of the Azure batch job. Ex: 1024-1v112-7e026be3-5fd0-48ee-b7d1-abd61f747705
6. **taskId:** ID of the Azure batch task. Ex: b47f0263-ba1c-48a7-8d29-4bf021b72043

Summary

##	timestamp	hostname	eventName	eventType
##	Length:660400	Length:660400	Length:660400	Length:660400
##	Class :character	Class :character	Class :character	Class :character
##	Mode :character	Mode :character	Mode :character	Mode :character
##	jobId	taskId		
##	Length:660400	Length:660400		
##	Class :character	Class :character		
##	Mode :character	Mode :character		

gpu.csv (208.7MB)

This file contains metrics that were output regarding the status of the GPU on the virtual machine.

Schema

1. **timestamp**: time of data creation. Ex: 2018-11-08T08:27:10.314Z
2. **hostname**: unique hostname created by the Azure. Ex: 8b6a0eebc87b4cb2b0539e81075191b900001C
3. **gpuSerial**: The serial number of the physical GPU card.Ex: "0323217055910"
4. **gpuUUID**: The unique system id assigned by the Azure system to the GPU unit. Ex: GPU-1d1602dc-f615-a7c7-ab53-fb4a7a479534
5. **powerDrawWatt**: Power draw of the GPU in watts. Ex: 131.55
6. **gpuTempC**: Temperature of the GPU in Celsius Ex: 48
7. **gpuUtilPerc**: Percent utilisation of the GPU Core(s). Ex: 92
8. **gpuMemUtilPerc**: Percent utilisation of the GPU memory. Ex: 52

Summary

##	timestamp	hostname	gpuSerial	gpuUUID
##	Length:1543681	Length:1543681	Min. :3.201e+11	Length:1543681
##	Class :character	Class :character	1st Qu.:3.236e+11	Class :character
##	Mode :character	Mode :character	Median :3.236e+11	Mode :character
##			Mean :3.240e+11	
##			3rd Qu.:3.250e+11	
##			Max. :3.252e+11	
##	powerDrawWatt	gpuTempC	gpuUtilPerc	gpuMemUtilPerc
##	Min. : 22.55	Min. :26.00	Min. : 0.00	Min. : 0.00
##	1st Qu.: 44.99	1st Qu.:38.00	1st Qu.: 0.00	1st Qu.: 0.00
##	Median : 96.59	Median :40.00	Median : 89.00	Median :43.00
##	Mean : 89.20	Mean :40.08	Mean : 63.06	Mean :33.41
##	3rd Qu.:121.34	3rd Qu.:42.00	3rd Qu.: 92.00	3rd Qu.:51.00
##	Max. :197.01	Max. :55.00	Max. :100.00	Max. :83.00

task-x-y.csv (6.2MB)

This file contains the x,y co-ordinates of which part the image was being rendered for each task.

Schema

1. **jobId**: Id of the Azure batch job. Ex: 1024-1vl12-7e026be3-5fd0-48ee-b7d1-abd61f747705
2. **taskId**: ID of the Azure batch task. Ex: b47f0263-ba1c-48a7-8d29-4bf021b72043
3. **x**: X co-ordinate of the image tile being rendered. Ex: 116
4. **y**: Y co-ordinate of the image tile being rendered. Ex: 178
5. **level**: Zoom level
 - 4
 - 8
 - 12

Summary

```
##      timestamp      hostname      gpuSerial      gpuUUID
## Length:1543681    Length:1543681    Min.   :3.201e+11    Length:1543681
## Class :character  Class :character  1st Qu.:3.236e+11    Class :character
## Mode  :character  Mode  :character  Median :3.236e+11    Mode  :character
##                                     Mean  :3.240e+11
##                                     3rd Qu.:3.250e+11
##                                     Max.   :3.252e+11
## powerDrawWatt      gpuTempC      gpuUtilPerc      gpuMemUtilPerc
## Min.   : 22.55    Min.   :26.00    Min.   : 0.00    Min.   : 0.00
## 1st Qu.: 44.99    1st Qu.:38.00    1st Qu.: 0.00    1st Qu.: 0.00
## Median : 96.59    Median :40.00    Median : 89.00    Median :43.00
## Mean   : 89.20    Mean   :40.08    Mean   : 63.06    Mean   :33.41
## 3rd Qu.:121.34    3rd Qu.:42.00    3rd Qu.: 92.00    3rd Qu.:51.00
## Max.   :197.01    Max.   :55.00    Max.   :100.00    Max.   :83.00
```

```
## # A tibble: 6 x 8
##   timestamp      hostname      gpuSerial gpuUUID      powerDrawWatt gpuTempC gpuUtilPerc
##   <chr>          <chr>          <dbl> <chr>          <dbl>      <int>      <int>
## 1 2018-11-08~ 8b6a0eebc8~    3.23e11 GPU-1d16~      132.        48         92
## 2 2018-11-08~ d8241877cd~    3.24e11 GPU-04a2~      117.        40         92
## 3 2018-11-08~ db871cd77a~    3.23e11 GPU-f459~      122.        45         91
## 4 2018-11-08~ b9a1fa7ae2~    3.25e11 GPU-ad77~       50.2       38         90
## 5 2018-11-08~ db871cd77a~    3.23e11 GPU-2d4e~      142.        41         90
## 6 2018-11-08~ 265232c5f6~    3.24e11 GPU-7176~      120.        43         88
## # ... with 1 more variable: gpuMemUtilPerc <int>
```