CSC8634_TeraScope Report

Morgan Frodsham (210431461)

10/01/2022

Extended Technical Project: Performance Evaluation of Terapixel Rendering in Cloud (Super)computing

- what is the need for the project?
- justify your choice of response (i.e. the nature of, and plan for your project). To give strength to your argument, you should reference practice elsewhere (e.g. in academic literature or industry practices)
- What, concisely, did you do?
- How successful has it been? Provide evidence sing appropriate evaluation methodologies, and comment on the strengths/weaknesses of your evidence in answering this question.
- What are the future implications for work in this area? If applicable, which areas of extension work are now possible due to the foundational work you have performed in this project?
- A brief relfection on your personal and professional learning in undertaking this project. Here you can comment on how you found the process, what you learned about the technologies and methodologies you used, which aspects you found most difficult/straightforward, and any conclusions which will inform the way you undertake similar projects in future.

Introduction

Urban Observatories attempt to replicate the "breadth, longevity and success of astronomical observatories" in understanding how cities operate (P James (2016)). Funded in partnership with the UK Collabatorium for Research in Infrastructure and Cities ("Details of Grant" (n.d.)), the Newcastle Urban Observatory collects environmental data about the city of Newcastle-upon-Tyne (Observatory (n.d.)). This is considered to be the most expansive set of "environmental data in the UK, with: over 74 urban indicators; 7,000 observations every minute; 3,600 sensor streams; 9 billion data points; 540 CCTV sensors; and hundreds of millions of images" (Newcastle University (2021)). Newcastle University has created a scalable cloud supercomputer software architecture for visualising this data as realistic terapixels(Forshaw (2021)), which are images containing over one trillion pixels that allow the user to zoom into incredible detail (Nicolas S. Holliman (2019)).

Purpose

There is significant computational cost to producing a high quality terapixel image (Nicolas S. Holliman (2019)); Newcastle University's supercomputer architecture for scalable visualisation, however, allows the institution to only pay for what is needed while producing terapixel viusalisations that support daily updates and undertaking rigorous evaluation (Forshaw (2021)). This project uses a dataset created from the production of a terapixel image by this architecture, and aims to provide useful insights into the computation performance. The data set was created during a run using 1024 GPU nodes, and this run is split into three jobs to render the terapixel visualisation (levels 4, 8 and 12) (Forshaw (2021)).

###Data Mining Goals

Which event types dominate task runtimes? - execution time if often the best metric for accessing comuter performance (Hoefler) What is the interplay between GPU temperature and performance? What is the interplay between increased power draw and render time? Can we quantify the variation in computation requirements for particular tiles? Can we identify particular GPU cards (based on their serial numbers) whose performance differs to other cards? (i.e. perpetually slow cards). What can we learn about the efficiency of the task scheduling process?

- we should see efficientl linear scalaing as we add more compute notdes - gustafosn-Baris' law (**terapixel?**) Rendering images is considered to be a good test of hardware performance as it is capable of absorbing all available compute resource, following Gustafson-Barsis' law that the problem size scales to fill the compute capacity (Nicolas S. Holliman (2019)).
  ###Benefits

Include resources, requirement, risks, terminology, cost and benefits

###Plan

Include plan plus initial assessment of tools and techniques

##Data Understanding

###Collect Inital Data

Newcastle University has provided three data sets created from application checkpoint and system metric output for the production of a terapixel image (Forshaw (2021)).

###Describe Data

###Explore Data

###Verify Data Quality

##Data Preparation

###Select Data

###Clean Data

###COnstruct Data

###Integrate Data

###Format Data

##Modeling

##Evaluation

###Evaluate Results

###Review Process

###Determine Next Steps

#Deployment

Bibliography

"Details of Grant." n.d. https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/P016782/1.

Forshaw, Matthew. 2021. *Summary*. https://github.com/NewcastleDataScience/StudentProjects202122/blob/master/TeraScope/Summary.md.

Newcastle University, Proffessor P James on behalf of. 2021. "Written Evidence Submitted by Newcastle University (Evp0115)." https://committees.parliament.uk/writtenevidence/22873/html/.

Nicolas S. Holliman, James Charlton, Manu Antony. 2019. "Petascale Cloud Supercomputing for Terapixel Visualization of a Digital Twin." *CoRR* abs/1902.04820. http://arxiv.org/abs/1902.04820.

Observatory, Newcastle Urban. n.d. "Our Urban Observatory." https://www.ncl.ac.uk/who-we-are/vision/urban-observatory/.

P James, J Jonczyk, RJ Dawson. 2016. "The UK Urban Observatory Programme - Newcastle: Towards Slightly Less Dumb Cities." https://huckg.is/gisruk2017/GISRUK_2017_paper_15.pdf.