**Tentative Research Title:**

**(NIT6130 Assignment 2)**

05/2021

**Student name:** Qin Xu

**Student ID:** s4647307

**Supervisor:** Dr. Soheila Gheisari

**Unit:** Introduction to Research (NIT6130)

**Unit Coordinator:** Professor Yuan Miao

A Quantitative Report on How Cloud Based GIS Applications Can Better Support Big Imagery Data Analytics

Table of Contents

[Abstract 2](#_Toc72616339)

[1. Introduction 2](#_Toc72616340)

[2. Related Work 3](#_Toc72616341)

[2.1 Research Method 3](#_Toc72616342)

[2.2 Focused Reading of the Related Research Paper 4](#_Toc72616343)

[3. Proposed Research 9](#_Toc72616344)

[3.1 Problem statement 9](#_Toc72616345)

[3.2 Aim of the research 9](#_Toc72616346)

[3.3 Expected Outcomes and Significance 9](#_Toc72616347)

[3.4 Method and innovation 10](#_Toc72616348)

[3.4.1 Architecture of the proposed solution and workflow 10](#_Toc72616349)

[3.4.2 Cost analysis of AWS cloud 12](#_Toc72616350)

[3.4.3 Experimental imagery datasets design and workflow 13](#_Toc72616351)

[3.4.4 AWS cloud configuration of proposed architecture 14](#_Toc72616352)

[3.4.5 Experimental group of network configuration 15](#_Toc72616353)

[3.4.6 Experimental group of horizontally scaled system 16](#_Toc72616354)

[3.4.7 Experimental group of vertically scaled the system 17](#_Toc72616355)

[3.4.8 Statistical analysis 18](#_Toc72616356)

[3.5 Preliminary Study 19](#_Toc72616357)

[4. Conclusion 19](#_Toc72616358)

[5. References 20](#_Toc72616359)

[Appendix A - Literature Review, Broad Scan and Reading 21](#_Toc72616360)

[Appendix B – Use Ortho Mapping Workflow to Generate Map Products 25](#_Toc72616361)

Abstract

**Background**

Remote sensing data powered by large scale sensor networks such as satellites, UAVs (Unmanned Aerial Vehicle) has grown exponentially over the last decade and yet the speed shows no signs of slowing. In contrast, the development of application systems for processing and storing big remote sensing imagery datasets is relatively slow. This is partly due to the intense computing power requirement for big imagery data analytics. In contrast, cloud computing technologies has been developed and improved incredibly fast over the past years and it has now become a dominant form of providing on-demand computing resources. As a result, cloud-based GIS (Geographic Information System) is offering new opportunities for solving the difficulties of accessing and processing big imagery datasets which are usually costly procedures in the previous time.

**Objectives**

The proposed research will investigate how cloud computing can be utilized to deploy web applications for big imagery data analytic, specifically, using ArcGIS platform. The proposed architecture and solutions will further demonstrate the exiting research on exploring the benefits for such approaches. It will give a brief overview of why cloud based remote sensing application is an ideal platform for the coming development of remote sensing business model as well as supporting open-source initiative such as the Open Data Cube Initiative (ODC).

**Methodology**

This research will follow quantitative research methodology to analyze how different variables in the cloud environment can affect the system performance when processing big imagery datasets. By using control variate method, the result will indicate a relationship between each factor of the cloud computing resources and the processing speed. The general workflow involves building cloud environment using AWS platform, replicating the environmental configuration from the existing research, designing the input datasets, deploying the ArcGIS application, processing sample data, collecting the system performance data. finally, all the data collected will used be put into statistical analysis to generate valuable knowledge and conclusion.

**Expected Outcomes**

The proposed solution will further examine the actual performance of cloud-based ArcGIS application, giving it more degree of preciseness. By comparing the difference results based on the cost analysis of using AWS platform, it will illustrate the advantages of such approach and provide an example for future development outlook of building cloud-based GIS.

# Introduction

Our planet Earth has never been observed so extensively and thoroughly before. According to NASA, by 2022, the ingest rate of data into the EOSDIS (The Earth Observing System Data and Information System) archive is projected to grow to as much as 47.7 PB per year, according to estimates from ESDS (Earth Science Data Systems). As this ingest rate increases, the volume of data in the EOSDIS archive also is expected to grow—from nearly 32 PB today to more than 37 PB by 2020; by 2025, the volume of data in the EOSDIS archive is expected to be more than 246 PB. (Earthdata Cloud Evolution | Earthdata, 2021)

Compared to the rapidly growing data volume and analytics demand, remote sensing applications are underdeveloped in many counties. Accessing high quality image processing tools is costly and the learning curve is still steep. The major challenge lies on the fact that remote sensing analytics systems require intensive computational power and large amount of data storage. Building and deploying such application in the local server or network is extremely costly and hard to maintain.

Take Open Data Cube initiative as an example, with the support of the Committee on Earth Observation Satellites (CEOS) System Engineering Office (SEO), the goal is to reach operational Data Cubes in 20 countries by 2020, yet only three are actually operating to date which are located at Australia, Colombia, Switzerland respectively (R Rizvi, Killough, Cherry and Gowda, 2018). It partially reflects the difficulties of accessing advanced image processing systems.

The existing research from Esri, Inc. and Wuhan University provides a basic understanding of how cloud-based ArcGIS can be a full stack solution for big imagery data storage, management, processing, and application. (Huang, Gao, Zhang and Zhang, 2018)

However, there are some improvements can be made and applied to the research methodology such as increasing the dataset volume to a higher magnitude, measuring network throughput to indicate user experience, and testing the actual image production quality. The proposed research will use quantitative methodology to further analysis the result and examine the actual performance for the cloud-based ArcGIS by simulating a real word usage environment. With this experiment done, it will lay out a solid foundation for large scale commercial practices of deploying GIS in the cloud.

# Related Work

## Research Method

The method to be used in the proposed research is quantitative. That means by conducting the experiment, all the input data, processes, system performance and outputs are observable and measurable. ArcGIS is a geographic information system developed and maintained by the Environmental Systems Research Institute (ESRI, Inc.). It provides a variety of tools for map generation, revision, and change detection. These tools can be used to process aerial or satellite imagery and produce specialized maps and other related products.

In addition to monitoring and recording the system performance data, the variance of different collected data will also be calculated to achieve a better reliability. It means the analysis of variance will be conducted to show the difference between different system configuration groups. The chosen Alpha value in the proposed research is 0.05. The final P-values will also indicate the possibility that the experimental results can be used to infer real word performance of cloud-based ArcGIS.

## Focused Reading of the Related Research Paper

After literature review and going through focused reading, the article chosen is listed below. It is sourced from IEEE Xplore and accessed via Victoria University Library Portal.

* A Cloud Computing Solution for Big Imagery Data Analytics. (Huang, Gao, Zhang and Zhang, 2018)

**Current Problems in the Related Field**

The selected paper gives a brief overview of challenges and problems facing for remote sensing imagery analytics. Then, the researchers proposed a cloud computing solution which is a good model for next generation remote sensing applications. It outlines the long learning curves of processing big imagery data as well as the slow response time of the current applications when it comes to the emergency usage. The large gap between exponentially growing remote sensing data and potential costly processing the data-intensive and computation-intensive images prompted the researchers to conduct the experiment and come up with a better solution.

**Key Contributions**

To tackle these challenges, the researchers break down the problem domain mainly into three parts: data storage, data computing and data methodologies. For each of the sub-problem, they analyze how cloud computing can solve it in a cost-efficient way while achieving high scalability. They also utilize some of the great tools and features offered by ArcGIS platform which includes the support of distributed raster data storage, high-capacity image management with mosaic datasets as well as the support of distributed, scalable raster analytics. Taking advantages of all technologies, they design a new general architecture for deploying GIS in the cloud. By implementing this architecture in AWS platform, they prove that this solution is affordable and achievable.

**Key Differences in the Method (Innovation)**

The main innovation of proposed architecture is that it offers great flexibility for user to manage, process and visualize remote sensing big imagery data while achieving high degree of performance. As shown in figure 1, the Cloud Object Storage is used to stored big imagery files and related GIS data. Large size file could also be converted to optimized raster format such as CRF. GIS hosting server, image hosting server and image analytic servers together form a distributed computing system in the cloud which results in maximizing the processing speed and efficiency.

Image servers deployed in the cloud object storage also consist of more than 200 built-in tools and features that allow users to preprocess, run orthorectification and mosaicking operation against multispectral and hyperspectral imagery dataset. The ArcGIS server is used as a web service for sharing imagery and responding API request. Users can also extend the system’s analytical capabilities by writing their own custom functions or deploying different image processing algorithms. Additionally, multiple users across the same enterprise can use this system to collaborate, contribute data and sharing functions and features developed.

Diagram

Description automatically generated

Fig.1. Architecture of ArcGIS based cloud computing solution for big imagery analytics.

(Huang, Gao, Zhang and Zhang, 2018)

**Main Achievements and Significance**

As shown in the fig. 2, in the experiment, the researchers use 6 AWS m4.2xlarge EC2 instances, S3 object storage and 1 GB network along with ArcGIS Enterprise 10.6.1 to test the performance of the proposed architecture solution. The testing input data types include UAV dataset, aerial dataset, and satellite dataset. The method for compute and process these big images is ortho mapping workflow.

Diagram

Description automatically generated

Fig.2. Deployment of Experiment Environment (Huang, Gao, Zhang and Zhang, 2018).

After conducting the experiment, the result shown in the fig. 3 is inspiring. As more distributed CPU cores involved in the computation, the consumed time shows clear trend of decreasing, which means the speedup of processing time. The result also indicates as the CPU Cores number reach 20, the difference of processing time between using object storage and file storage is narrowing significantly compared to less CPU Cores. It indicates that the impact of latency caused by S3 object storage can be offset by increasing the CPU core numbers. This pattern is also consistent for all three datasets.

Chart, line chart

Description automatically generated

1. UAV Dataset

Chart, waterfall chart

Description automatically generated

1. Aerial Dataset

Chart, line chart

Description automatically generated

1. Satellite Dataset

Fig.3. Experiment Result (Huang, Gao, Zhang and Zhang, 2018)

**Existing Research Problems**

Overall, the experiment design meets the research standards and specifications. And the results are reasonable and convincing. However, there are still some underlying problems that this research does not address.

The first one is that the volume of testing imagery dataset is relatively small especially for aerial images. The total testing sample only contains 225 UAV images, 12 aerial images and 234 satellite images. In a real-world commercial environment, there could thousands of users uploading images at the same time. In this scenario, will the performance still be optimizable by involving more computing power?

Secondly, the experiment did not take network throughput as a variance to test the system performance. In many cases, network speed and capacity can dramatically affect the actual user experience. Will the larger network throughput decrease the latency which in term, results in a better user experience?

The third remaining question is the how the system can be extended to a better performance. In other words, the EC2 instance provide by AWS can be both vertically and horizonal scaled, but the impact of such change is unknown.

**Possible Improvements**

Therefore, there are some possible improvements that can be applied to obtain a broader understanding of the actual performance of the proposed cloud computing solution.

Firstly, the input volume of dataset can be increased to thousands level. This will simulate a more common real-world scenario in which multiple users are sending data process request.

Secondly, another experiment configuration which includes a different network throughput can be used to identify the impact of the performance.

Finally, we can set up two more contrast groups which have either horizontally scaled or vertically scaled system configuration. More specifically, by either increasing the number of EC2 instances or use more powerful instances, we can further examine the how the system can be scaled to adapt different performance requirements.

Further details will be discussed in the subsequent proposed research section.

# Proposed Research

**A Quantitative Analysis of the Cloud Based GIS Application Performance**

## Problem statement

The rapidly growing volume of remote sensing imagery data has prompted an unprecedented demand for big imagery analytics systems and applications. However, developing a system for processing big imagery data never been easy. Over the past decade, the fast-evolving cloud computing technologies have offered a new path to solve this challenge.

In 2018, a research team published a paper through IEEE (Institute of Electrical and Electronics Engineers), providing a full stack cloud implementation by deploying ArcGIS onto AWS (Amazon Web Services) to achieve easy access of processing big imagery dataset. Utilizing cloud computing services and data storage, the proposed architecture can achieve higher performance as more computing power involved in the system. This research will further examine the usability and actual performance of the existing architecture by simulating a real-world scenario. It will also provide more understanding of the existing solution and possible improvement of the cloud-based ArcGIS.

## Aim of the research

This research is mainly aiming to obtain a deeper understanding of the ArcGIS performance deployed in the cloud. More specifically, the objectives reside in the following aspect:

* To examine the actual system and GIS algorithm performance when a large volume of big imagery dataset is put into the system.
* To test how the different network throughputs can affect the overall system performance.
* To identify the different performance when the cloud computing environment is either vertically scaled or horizontally scaled.
* To do a statistical analysis of the different result sets and identify any possible errors and deviations by which will increase the accuracy and reliability of the research.

These objectives will be achieved by carefully designing the experiment, the final collection of data and the final scientific analysis of data.

## Expected outcomes and significance

As cloud computing become a major technology trend in IT, how it can be used to help developing powerful remote sensing imagery processing system remains unclear. By doing the proposed research, the result will provide a unique perspective to answer this question.

More specifically, the key benefits of conducing this research will be as following:

* After the experiment, we will be able to find out how the distributed system performs when it encounters thousands of imagery datasets at the same time.
* The impact of the latency caused by the network throughput of AWS storage system will also be clear and more accurate. It will give us a general idea on how to improve the system design for achieving better user experience.
* For the same amount of input data, the vertically or horizontally scaled system will show us how to better balance the performance and cost.
* Finally, by doing statistical analysis on all the result data, we can gain more reliability and confidence for potential business development for this solution.

Therefore, the proposed research will further demonstrate the usability and reliability of the ArcGIS based cloud computing solution and extend the knowledge of why it is ideal for remote sensing big data analytics.

## Method and innovation

This research will follow quantitative research methodology to analyze how different variables in the cloud environment can affect the application performance when processing big imagery datasets. By using control variate method, the result will indicate a relationship between each factor of the cloud configuration variable and the overall system performance. The general workflow involves:

* building cloud environment using AWS platform,
* replicating the cloud environment from the existing research,
* designing the imagery datasets,
* deploying the ArcGIS application,
* change the configuration of the cloud environment,
* processing input data,
* collecting the performance data.

The following sections outlines how we will implement the existing architecture, run the experiment and finally collecting the result datasets. It mainly includes three parts:

* From 3.4.1 to 3.4.4, these sections are about the architecture of the ArcGIS based solution and a comprehensive analysis of AWS cloud environment configuration.
* From 3.4.5 to 3.4.7, these sections include the design of different groups of cloud environment and the experiment workflow.
* Section 3.4.8 is about the method to use in statistics analysis.

### Architecture of the proposed solution and workflow

By deploying the geographic information system (GIS) in the cloud, users can have easy access to high performance integrated imagery processing application. Fig. 4 shows a variety of big data imagery services can be built in the cloud which includes spatial data storage, software platform, GIS servers and GIS application.

Diagram

Description automatically generated

Fig.4. GIS Cloud Architecture (Barsukov, 2021)

In this experiment, we will be using ArcGIS Enterprise 10.6 edition. It supports AWS cloud virtual machine and can be deployed using specialized tools that Esri provides. The ArcGIS Enterprise Cloud Builder also provides a graphical user interface application and Command Line Interface for Amazon Web Services. Fig.4 shows the server site on AWS with an ArcGIS Server and an Amazon RDS instance, along with two additional GIS Server installations on AWS instances available when CPU usage exceeds a specified threshold.

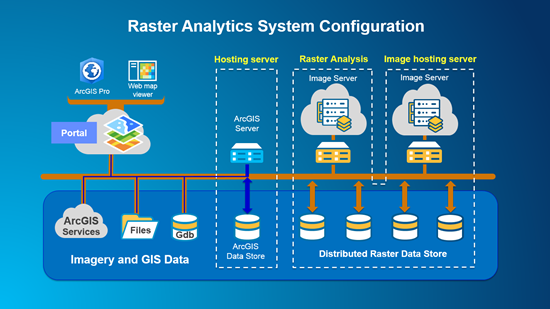


Fig. 5. Architecture of ArcGIS based cloud computing solution for big imagery data analytics

(Configure and deploy ArcGIS Enterprise for raster analytics—Portal for ArcGIS | Documentation for ArcGIS Enterprise, 2021)

For data storage, AWS provides reliable, scalable, and secure data storage system such as virtual machines (EC2) and virtual drives (EBS). For data computing, elastic computing offered by AWS giving users a complete control of computational resources such as distributing load across machines (ELB), scaling the services using an auto-scaling group (ASG).

### Cost analysis of AWS cloud

As Table 1 shows, AWS offers very competitive pricing for all their services in the industry. In terms of EC2, we will be choosing m5.8xlarge and m5.16xlarge as our clustered computing system. To take advantages of all CPU cores and memory, large process will be divided into smaller tasks which can allow us to horizontally scale the system.

Table

Description automatically generated with medium confidence

Table 1. EC2 Spot Instances Pricing on AWS Asia Pacific (Sydney) Region

To set up the contrast group for different throughputs, we will be choosing EBS for long-term stored datasets. EBS has higher network connection performance and lower latency. As discussed in the previous section, network throughput is a key indicator of the user experience because the latency and delay. Table 2 shows the different capacities between S3 object storage and EBS block storage.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Type** | **Latency** | **Price** |
| EBS | Block Storage | Low | $0.12GB/Month |
| S3 | Object Storage | High | $0.026GB/Month |

Table 2. Differences between ASW Storage Types

### Experimental imagery datasets design and workflow

For the input big imagery datasets, we will be using three different kinds of sensor data:

1. UAV dataset: 100 images captured from UAV which have the same dimensions and in JPEG format.
2. Aerial dataset: 100 images captured from aerial which also have the same dimensions and in TIFF format.
3. Satellite dataset: 100 big volume images captured from satellite which will be three-line stereo product. Like the other two datasets, all will have the same dimensions and in TIFF format.

For the workflow, we will be using orthophoto mapping workflow. Preprocessed orthophotos (or orthomosaics) are images with map-like accuracy, generated from satellite, aerial, or drone imagery, that have been orthorectified and mosaicked together for a single project. Using a mosaic dataset configured to manage orthophotos makes it easier to visualize, query, and analyze large collections of orthophotos. (Managing Preprocessed Orthophotos—Imagery Workflows | Documentation, 2021) It is a key algorithm for producing any remote sensing mapping imagery product.

Generally, the orthophoto mapping workflow is both data-intensive and computation-intensive. It involves three steps:

(1) Block adjustment. Extract tie points from raw images and use bundle adjustment to correction sensor model.

(2) Extract DEM (Digital Elevation Model). Extract dense point cloud using Semi-Global Matching algorithm then filter trees, buildings, and other above-ground features to produce DEM.

(3) Generate ortho mosaic product. Use refined sensor model and extracted DEM to create a planimetrically correct image product.

Fig. 6 shows a sample product of ortho mapping using ArcGIS pro desktop edition. More details of workflow and involved algorithms are in the appendix B. (Huang, Gao, Zhang and Zhang, 2018)

Graphical user interface, application, PowerPoint

Description automatically generated

Fig. 6. Sample Product of Ortho Mapping (Satellite image © 2020 Maxar Technologies)

### AWS cloud configuration of proposed architecture

Table

Description automatically generated

Table 3. Technical Specifications of AWS m5. EC2 Instances

Table 3 shows the technical specifications of AWS m5. EC2 instance. These instances provide flexible choices of compute, memory, and networking resources. They are ideal for deploying the GIS web servers and code repositories. In the following sections, we will be configuring more groups of experiment environment to collect the necessary data that can be used to analyze the system performance.

Diagram

Description automatically generated

Fig. 7. Configuration of the Experiment Environment

As fig. 7 shows, the hardware environment consists of 7 AWS m5.8xlarge EC2 instances, EBS block storage, and 10GB network. One is used as a GIS hosting server; other 6 instances will be used as distributed image servers to query the data database and execute the image processing workflow and algorithms. The input dataset will be stored in the block storage. File storage and geodatabase are used to stored other related GIS data.

As for the big imagery processing system, ArcGIS Enterprise 10.6 will be deployed in the configured environment and used to test the performance of the proposed solution. Additionally, three more groups of different deployment will be configured to collect the performance data needed to analyze the overall performance.

### Experimental group of network configuration

Amazon EBS allows users to create storage volumes and attach them to Amazon EC2 instances. Once attached, you can create a file system on top of these volumes, run a database. We will choose EBS as our storage system. It provides a range of options that allow users to optimize storage performance and cost for the workload.

To evaluate how different throughput can affect the speed of imagery data processing and overall system performance, we need to change the network throughput in existing configuration. The maximum throughput of m5.8xlarge allowed by AWS is 20GB and the rest of system configuration will remain exactly the same as before. Fig. 8 shows the different network throughput for the system.

Diagram

Description automatically generated

Fig. 8. Experiment Environment of 20GB Network

### Experimental group of horizontally scaled system

A "horizontally scalable" system is one that can increase capacity by adding more computers to the system. This is in contrast to a "vertically scalable" system, which is constrained to running its processes on only one computer; in such systems the only way to increase [performance](https://wa.aws.amazon.com/wellarchitected/2020-07-02T19-33-23/wat.pillar.performance.en.html) is to add more resources into one computer in the form of faster (or more) CPUs, [memory](https://wa.aws.amazon.com/wellarchitected/2020-07-02T19-33-23/wat.concept.memory.en.html) or storage. (Horizontal scaling - AWS Well-Architected Framework, 2021)

According to AWS official document, Horizontally scalable systems are oftentimes able to outperform vertically scalable systems by enabling parallel execution of [workloads](https://wa.aws.amazon.com/wellarchitected/2020-07-02T19-33-23/wat.concept.workload.en.html) and distributing those across many different computers. (Horizontal scaling - AWS Well-Architected Framework, 2021)

To find out how horizontally scaled system can potentially speed up the imagery data processing time. We will add six more instances of m5.8large as image servers and monitor the processed speed change due to the increased number of CPU cores and computing power. Fig. 9. shows the result of horizontally scaled hardware environment. The network throughput will remain at 10GB.

Diagram

Description automatically generated with low confidence

Fig. 9. Horizontally Scaled Experiment Environment

### Experimental group of vertically scaled system

AWS provides various tools and features to achieve vertically scaling without re-deploying the GIS completely. According to official document, the Ops Automator can be used to replace the instance with a new, resized instance instead of restarting the existing instance. This tool is also integrated with Elastic Load Balancing to automatically register the new instance with load Balancers. Fig. 10 shows the process of vertically scaling. (AWS Ops Automator v2 features vertical scaling (Preview) | Amazon Web Services, 2021)

A picture containing diagram

Description automatically generated

Fig. 10. Vertically Scaling Using AWS Ops Automator (AWS Ops Automator v2 features vertical scaling (Preview) | Amazon Web Services, 2021)

To avoid the complexity of scaling each image server directly, we will replace the image servers with m5.16xlarge instance. As table 3 shows, compared to the m5.8xlarge, m5.16 has two times more CPU Cores and memory. Fig.11 shows the vertically scaled hardware environment. The network throughput and the number of instances will remain the same to allow us to collect the concise data for the following statistical analysis.

Diagram

Description automatically generated

Fig. 11. Vertically Scaled Experiment Environment

### Statistical analysis

Statistical analysis is a key part for analyzing quantitative research data. We will calculate various probabilities and models to test predictions based on the collected data. The data we will collect mainly includes imagery data processing time, number of distributed CPU cores, and network throughput. After that we will calculate the mean, and median and evaluate for each experiment group and then evaluate the differences.

The main methods used to do the comparison and identify the differences between each different experimental groups will be T-test. Below is the mathematical formula T-test Calculation.

**T-test formular:**

​Text

Description automatically generated with low confidence

​**where**: mean1 and mean2 = The average values of each of the sample sets s(diff) = The standard deviation of the differences of the paired data values n = The sample size (the number of paired differences) n – 1 = The degrees of freedom​

(Hayes, 2021)

We have designed a total of 4 different hardware configuration groups. The main hypothesis is that as more CPU cores or involved in the computation, the big imagery processing time will be dramatically decreased. The data will also be used to investigate the relationship of each system configuration variable and the processing time to achieve the research objectives stated in the beginning of proposed research section.

## Preliminary Study

Cloud computing is a supercomputing paradigm based on Internet, which makes use of computer nodes in the cloud cluster through a network to complete a computing task in parallel. The use of online mapping and spatial search has become ubiquitous, with hundreds of millions of desktop and smartphone users regularly accessing mapping services (Smith [2016](https://link.springer.com/chapter/10.1007/978-3-030-26626-4_10#CR35)).

There are various platforms, that allow creating online maps without the need of coding knowledge. The GIS applications used in proposed research is available at <https://www.arcgis.com/home/index.html>,. ArcGIS is not just a tool for creating online maps, it gauges users with options to make web apps or 3D web scenes – with none or limited coding skills. (Pánek and Burian, 2019).

In the proposed research, as more computing power put into the system, we will be expecting a dramatic change of decreasing processing time. In other words, by combining the advantages of cloud computing and latest development of GIS done by Esri, Inc., it will demonstrate the efficiency of the proposed solution. A further improvement of getting a comprehensive cost analysis can also be done by other researchers who are interested in this area.

The proposed research will expand the knowledge of cloud-based solution for GIS application and inspire more developers and engineers to develop and deploy their own version of GIS tools. It will also encourage more major computing platforms to make more optimizations to GIS deployment and thus make it more user friendly and easy to interact.

# Conclusion

This report analyzes the problem we are facing as the need and requirement for big imagery data analytics are rapidly growing. The solution and architecture proposed by a research team from Esri, Inc. and Wuhan University is creative and valuable. In their paper, they utilized could computing and proposed a full stack solution by deploying and ArcGIS on AWS.

In the proposed research, this report extends the existing solution and evaluates the cost of AWS. By setting up 3 more contrast groups which includes different network throughput, vertically scaled cloud environment, and horizontally scaled environment, more experiment data about the actual system performance can be collected. Through statistical analysis of these result data, we can gain a more comprehensive understanding of the cloud-based GIS system. Additionally, more drawbacks of the design and bugs can be detected to avoid some possible scenarios that would trigger system failure.

The preliminary study indicates that cloud computing is an ideal platform to run GIS system. And it solves the difficulties of accessing efficient big imagery data processing service. The maps produced by the GIS system can be used in a lot of critical industry such as public safety, transportation, natural resources etc. It also shows that the practicality of the proposed solution and the possible improvements. The future study can focus on testing the AWS Cocker Containers for on-demand computing instances, testing the potential of deploying a complete data cube for a particular region and more.

Overall, the ArcGIS based cloud computing architecture is a great example for the development of next generation big imagery data analytics system. The quantitative analysis done by the proposed research will provide great evidence to support this hypothesis.

# References

Barsukov, A., 2021. Steps to Consider Before Moving Your GIS to the Cloud - Intellias. [online] Intellias. Available at: <https://www.intellias.com/steps-to-consider-before-moving-your-gis-to-the-cloud/> [Accessed 19 May 2021].

Doc.arcgis.com. 2021. Managing Preprocessed Orthophotos—Imagery Workflows | Documentation. [online] Available at: <https://doc.arcgis.com/en/imagery/workflows/resources/managing-preprocessed-orthophotos.htm> [Accessed 20 May 2021].

Doc.arcgis.com. 2021. Ortho mapping workflow—Imagery Workflows | Documentation. [online] Available at: <https://doc.arcgis.com/en/imagery/workflows/best-practices/historical-imagery-ortho-mapping-workflow.htm> [Accessed 21 May 2021].

Earthdata.nasa.gov. 2021. Earthdata Cloud Evolution | Earthdata. [online] Available at: <https://earthdata.nasa.gov/eosdis/cloud-evolution> [Accessed 17 May 2021].

Enterprise.arcgis.com. 2021. Configure and deploy ArcGIS Enterprise for raster analytics—Portal for ArcGIS | Documentation for ArcGIS Enterprise. [online] Available at: <https://enterprise.arcgis.com/en/portal/10.7/administer/linux/configure-and-deploy-arcgis-enterprise-for-raster-analytics.htm> [Accessed 22 May 2021].

Hayes, A., 2021. T-Test Definition. [online] Investopedia. Available at: <https://www.investopedia.com/terms/t/t-test.asp> [Accessed 22 May 2021].

Pánek J., Burian J. (2020) Online Visualisation. In: Pászto V., Jürgens C., Tominc P., Burian J. (eds) Spationomy. Springer, Cham. https://doi.org/10.1007/978-3-030-26626-4\_10

Smith, D. A. (2016). Online interactive thematic mapping: Applications and techniques for socio-economic research. Computers, Environment and Urban Systems, 57, 106–117.

S. R Rizvi, B. Killough, A. Cherry and S. Gowda, "Lessons Learned and Cost Analysis of Hosting a Full Stack Open Data Cube (ODC) Application on the Amazon Web Services (AWS)," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 8643-8646, doi: 10.1109/IGARSS.2018.8518084.

Wa.aws.amazon.com. 2021. Horizontal scaling - AWS Well-Architected Framework. [online] Available at: <https://wa.aws.amazon.com/wellarchitected/2020-07-02T19-33-23/wat.concept.horizontal-scaling.en.html> [Accessed 21 May 2021].

Y. Huang, P. Gao, Y. Zhang and J. Zhang, "A Cloud Computing Solution for Big Imagery Data Analytics," 2018 International Workshop on Big Geospatial Data and Data Science (BGDDS), 2018, pp. 1-4, doi: 10.1109/BGDDS.2018.8626847.

# Appendix A - Literature Review, Broad Scan and Reading

**Round 1**

Over the past decade cloud computing has eventually become the dominating form of on-demand computing services delivery. It provides a model for provisioning and consuming IT capabilities on a need and pay by use basis. (Dhar, 2012). I am very interested in the development and evolution of cloud computing technology over the 5 years.

Therefore, in the first round, I used “cloud computing” for the first keyword. Then I also applied “2016 - 2021” as a condition so that I can get the latest research results. A total of 46975 papers have been returned.

Besides, I am also interested in how full stack environment has interacted with cloud computing. Therefore, I chose full stack as a second keyword. The total result is 92 papers.

**Keywords: “cloud computing”, “full stack”**

**Source database: IEEE Xplore**

**Number of results returned: 92** *(Filters Applied: 2016 - 2021)*

List of candidate articles for focus reading (with abstract)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Towards a full-stack devops environment (platform-as-a-service) for cloud-hosted applications | Zhenhua Li;Yun Zhang;Yunhao Liu | Tsinghua Science and Technology | 2017 |
| **Abstract:** If every programmer of cloud-hosted apps possessed exceptional technical capability and endless patience, the DevOps environment (also known as Platform-as-a-Service, or PaaS) would perhaps become irrelevant. However, the reality is almost always the opposite case. Hence, IT engineers dream of a reliable and usable DevOps environment that can substantially facilitate their developments and simplify their operations. Current DevOps environments include Google App Engine, Docker, Kubernetes, Mesos, and so forth. In other words, PaaS bridges the gap between vivid IT engineers and stiff cloud systems. In this paper, we comprehensively examine state-of-the-art PaaS solutions across various tiers of the cloud-computing DevOps stack. On this basis, we identify areas of consensus and diversity in their philosophies and methodologies. In addition, we explore cutting-edge solutions towards realizing a more fine-grained, full-stack DevOps environment. From this paper, readers are expected to quickly grasp the essence, current status, and future prospects of PaaS. | | | | |
| 2. | Dionysios Diamantopoulos;Florian Scheidegger;Stefan Mach;Fabian Schuiki;Germain Haugou;Michael Schaffner;Frank K. Gürkaynak;Christoph Hagleitner;A. Cristiano I. Malossi;Luca Benini | Dionysios Diamantopoulos;Florian Scheidegger;Stefan Mach;Fabian Schuiki;Germain Haugou;Michael Schaffner;Frank K. Gürkaynak;Christoph Hagleitner;A. Cristiano I. Malossi;Luca Benini | 2020 IEEE Symposium in Low-Power and High-Speed Chips (COOL CHIPS) | 2020 |
| **Abstract:** The performance improvement rate of conventional von Neumann processors has slowed as Moore's Law grinds to an economic halt, giving rise to a new age of heterogeneity for energy-efficient computing. Extending processors with finely tunable precision instructions have emerged as a form of heterogeneity that tradeoffs computation precision with power consumption. However, the prolonged design time due to customization of the supported framework for a system-on-a-chip may counteract the advantages of transprecision computing. We propose XwattPilot, a system aiming at accelerating the transprecision software development of low-power processors using cloud technology. We show that the total energy-to-solution can be significantly decreased by using transprecision computations, whereas the proposed system can accelerate the energy-efficiency evaluation runtime by 10.3×. | | | | |

**Round 2**

**Keywords: “cloud computing”, “full stack”, “web service”**

**database: IEEE Xplore**

**Number of results: 10** *(Filters Applied: 2016 - 2021)*

List of candidate articles for focus reading (with abstract)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | Lessons Learned and Cost Analysis of Hosting a Full Stack Open Data Cube (ODC) Application on the Amazon Web Services (AWS) | IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium | IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium | 2017 |
| **Abstract:** The Open Data Cube (ODC) initiative, with support from the Committee on Earth Observation Satellites (CEOS) System Engineering Office (SEO) has developed a state-of-the-art suite of software tools and products to facilitate the analysis of Earth Observation data. This paper presents a short summary and cost analysis of our experience using Amazon Web Services (AWS) to host one such software product, the CEOS Data Cube (CDC) web-based User Interface (UI). In order to provide adaptability, flexibility, scalability, and robustness, we leverage widely-adopted and well-supported technologies such as the Django web framework and the AWS Cloud platform. The UI has empowered users by providing features that assist with streamlining data preparation, data processing, data visualization, and the sub-setting of Analysis Ready Data (ARD) products in order to achieve a wide variety of Earth imaging objectives. | | | | |
| 2. | SaW: Video Analysis in Social Media with Web-Based Mobile Grid Computing | Mikel Zorrilla;Julián Flórez;Alberto Lafuente;Angel Martin;Jon Montalbán;Igor G. Olaizola;Iñigo Tamayo | IEEE Transactions on Mobile Computing | 2018 |
| **Abstract:** The burgeoning capabilities of Web browsers to exploit full-featured devices can turn the huge pool of social connected users into a powerful network of processing assets. HTML5 and JavaScript stacks support the deployment of social client-side processing infrastructure, while WebGL and WebCL fill the gap to gain full GPU and multi-CPU performance. Mobile Grid and Mobile Cloud Computing solutions leverage smart devices to relieve the processing tasks to be performed by the service infrastructure. Motivated to gain cost-efficiency, a social network service provider can outsource the video analysis to elements of a mobile grid as an infrastructure to complement an elastic cloud service. As long as users access to videos, batch image analysis tasks are dispatched from the server, executed in the background of the client-side hardware, and finally, results are consolidated by the server. This paper presents SaW (Social at Work) to provide a pure Web-based solution as a mobile grid to complement a cloud media service for image analysis on videos. | | | | |

**Round 3**

**Keywords: “cloud computing” “full stack” “big data”**

**database: IEEE Xplore**

**Number of results: 62**

List of candidate articles for focus reading (with abstract)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. | A Cloud Computing Solution for Big Imagery Data Analytics | Yan Huang;Peng Gao;Yongjun Zhang;Jie Zhang | 2018 International Workshop on Big Geospatial Data and Data Science (BGDDS) | 2018 |
| **Abstract:** Remote sensing data has grown explosively in the past decade. However, remote sensing application systems evolve slowly due to the challenges on big imagery data storage and processing. The cost of accessing current remote sensing images is still very high. The learning curve of using remote sensing images and image processing tools is still steep. The response to emergencies is not fast enough. The development of big data and cloud computing technologies in recent years has brought new opportunities and challenges for remote sensing application. The remote sensing business model based on web and cloud computing environment is a clear trend. This paper gives a brief overview on the challenges faced during building big imagery analytics system for remote sensing, and proposes a cloud computing solution for big imagery data analytics. Experimental results of cloud computing with space, traditional film, and UAV images are also presented in this paper. It is a full stack solution based on ArcGIS platform for remote sensing imagery storage, management, processing and application, which can be an ideal platform for the next generation remote sensing application business model. | | | | |
| 2. | Lessons Learned and Cost Analysis of Hosting a Full Stack Open Data Cube (ODC) Application on the Amazon Web Services (AWS) | Syed R Rizvi;Brian Killough;Andrew Cherry;Sanjay Gowda | IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium | 201 |
| **Abstract:** The Open Data Cube (ODC) initiative, with support from the Committee on Earth Observation Satellites (CEOS) System Engineering Office (SEO) has developed a state-of-the-art suite of software tools and products to facilitate the analysis of Earth Observation data. This paper presents a short summary and cost analysis of our experience using Amazon Web Services (AWS) to host one such software product, the CEOS Data Cube (CDC) web-based User Interface (UI). In order to provide adaptability, flexibility, scalability, and robustness, we leverage widely-adopted and well-supported technologies such as the Django web framework and the AWS Cloud platform. The UI has empowered users by providing features that assist with streamlining data preparation, data processing, data visualization, and the sub-setting of Analysis Ready Data (ARD) products in order to achieve a wide variety of Earth imaging objectives. | | | | |

# Appendix B – Use Ortho Mapping Workflow to Generate Map Products

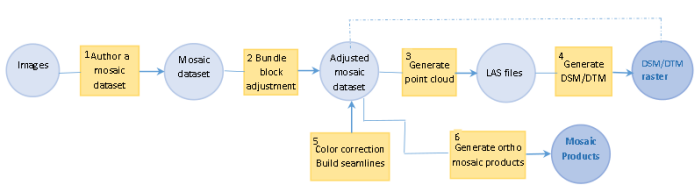
**Ortho mapping workflow**

Ortho mapping refers to the processes of generating ortho products, such as an image mosaic, a digital terrain model (DTM), and a digital surface model (DSM), from a set of remote-sensing images.

When many images are required to fully cover your study area, you need to mosaic these images together into an orthomosaic dataset. These satellite and aerial images come with geometric errors, as well as misalignments between adjacent image scenes due to systematic and nonsystematic factors. To improve geometric accuracy and reduce the misalignment, bundle block adjustment, also known as image triangulation, needs to be performed on these images.

The digital elevation model (DEM) used in the image orthorectification plays a very important role in the geometric accuracy of the final image mosaic, especially for areas with diverse terrains. In cases where high geometric accuracy of the output image mosaic is needed, you should derive high-resolution DSMs and DTMs from stereo pairs.

ArcGIS Desktop supports bundle block adjustment of satellite and aerial imagery. The software can generate ortho mapping products such as orthomosaics, DTMs, and DSMs from the bundle block adjustments. Below is a diagram of the high-level workflows for generating ortho mapping products.



There are six tasks in the ortho mapping workflow:

* Author a mosaic dataset.
* Perform bundle block adjustment.
* Generate a point cloud from a mosaic dataset.
* Generate a DTM and DSM.
* Balance color and reduce seams.
* Generate ortho mapping products.

Once the ortho mapping workflow is complete, you can generate orthomosaics or DEMs, or use the image collection (a mosaic dataset referencing the original images).

In the Output section of the Ortho Mapping tab, you can create DEMs (DSMs or DTMs) from the imagery.

In the user interface, you can specify either a DSM (top surface extracted from point cloud) or a DTM (estimated bare earth with filters applied to remove vegetation and buildings).

If you plan to use the DEM for orthorectification, you should specify DTM.

There are different algorithms available to generate the point cloud used to create the DEM: Extended Terrain Matching (ETM), Semi-Global Matching (SGM), and Multi-View Matching (MVM). It is typically best to test with the default (ETM) but consider MVM if your DEM does not show enough detail, or SGM if your imagery content shows urban buildings.

DEM generation is a fully automated process. Once created, it is advisable to review the hill shade to gauge the quality and determine if corrections are required. DEM editing tools are provided in Pixel Editor (ArcGIS Image Analyst required) if you need to remove artifacts.

(Ortho mapping workflow—Imagery Workflows | Documentation, 2021)