On Demand Satellite Image Processing

Next generation technology for processing Terabytes of imagery on the Cloud

WHITEPAPER

MARCH 2015

Introduction

Profound changes are happening with computing hardware and software. Rather than making single CPU's run faster manufacturers are moving to multi-core systems – cheap systems with 16 cores are now available for example. In addition Graphics Processing Units (GPU's based on processors used in graphics cards) offer the potential of teraflops of processing power for the cost of a video card. I/O bottlenecks are also being addressed via the emergence of Solid State Disk drives (SSD's) and fast networks (e.g., infiniband). 64 bit operating systems can take advantage of 100's of GBytes of RAM and standardized open standards for accessing multi-core CPU's and GPU are emerging.

Taken together, these emerging hardware capabilities - if properly integrated and used - can cope with the flood of image data expected from the new Earth Observation satellites by operating hundreds of times faster than the current traditional CPU desktop systems.

In addition new computational environments are also emerging – cloud computing which promises to provide vast computational resources on demand. Many corporations are moving to such environments to better share resources between compute intensive applications, to provide for surges in demand and reduce overall cost of ownership. Over time larger customers will demand applications running in these environments.

General Description

PCI Geomatics is at the forefront of one of the most challenging applications of Cloud Computing: high volume processing of Earth Observation imagery. To play a significant role in the emerging high speed computing industry as it relates to imagery, PCI has invested significantly in developing technology that is innovative, nimble, portable, scalable, automated, and most of all optimized for high speed performance.

PCI Geomatics has extensive background technology in Earth Observation processing – including geometric/ortho correction, mosaicking, atmospheric correction, image analysis and radar analysis. More than 550 algorithm modules have been implemented over the course of the company's history. Starting in 2007, PCI made strategic investments in its infrastructure, developing and implementing parallel processing, multi CPU/GPU optimized code, which can be monitored and balanced for optimal processing. The GeoImaging Accelerator (GXL) includes many benefits, including the fact that the technology has been designed to scale a system up or down, depending on throughput requirements. With the costs of hardware trending downwards, this means lower costs to deploy highly productive processing centres.

Cloud Computing Features
Some of the features of Cloud computing are listed below – their validity or specific challenges that apply to the Earth Observation industry are discussed.

Cloud Computing Features	Benefit/Challenge
Reduced cost	**Benefit:** Optimize the costs by paying only for what an organization uses (Large capital expenses are not required, pay incrementally)
Increased storage	**Challenge**: Transferring large amounts of imagery is time consuming and costly. I/O is a specific issue that is more challenging to EO industry.
Highly automated	**Benefit:** Software can be maintained automatically. Cloud services include guaranteed machine up times and redundancy.
Flexibility	**Benefit:** Ability to add/remove computing resources as required, often automatically based on demand.
More mobility	**Benefit:** Simplified, easily accessible management consoles that can be managed/ viewed from anywhere.
Allows IT to shift focus	**Benefit:** No longer having to worry about constant server updates and other computing issues – IT can focus on innovation.

As you can see above, one of the main challenges for the Earth Observation sector is to deploy technology on the cloud that is not hindered by data Input/Output (I/O). For smaller e-commerce transaction types (eBay, Google Mail, PayPal, etc…), transfer of large amounts of data is less of a challenge, as compared to processing Earth Observation Imagery. I/O issues are a key issue in multiple levels for Cloud based processing, including the initial transfer of data to the Cloud (EO Imagery is very big – a typical high resolution image (2 m multi-spectral + 50 cm panchromatic) can be very large – from 1-5 gigabytes depending on the number /types of bands included in the dataset.

During processing, I/O operations of imagery based content presents challenges, as software was not designed to read in data efficiently into the increasingly large banks of memory available – many operations in the software we designed to work on desktop processing systems, which traditionally were limited in terms of the available memory – therefore it was more efficient to implement multiple read/write operations within algorithms – this type of implementation is not well suited for the Cloud, and PCI has re-written much of its code to ensure optimization based newer Cloud based systems architectures.

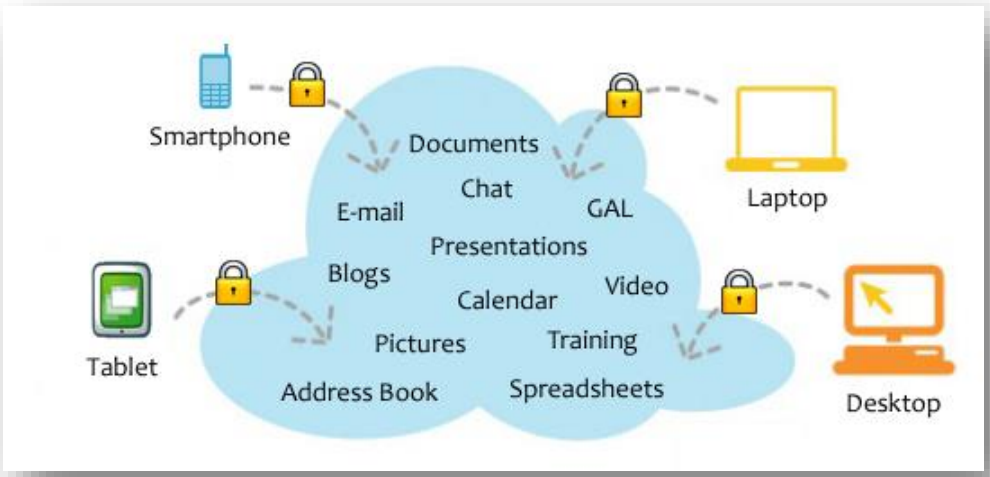

Figure 1: Cloud Computing

Basic definition: having access to all your applications and data from any network device

System Deployment

The GXL system can be deployed on the Cloud much in the same way a non-cloud based GXL system. The main difference is that there is no need to purchase any physical hardware in order to configure a GXL system that can achieve the stated throughput requirements. PCI's non-Cloud based GXL systems are typically deployed on desktop or rack mounted systems, which include a certain set of hardware specifications, including (PCI would specify which hardware to purchase):

- CPU / GPU
- RAM
- Disk Storage
- UPS
- File Server
- Network Switch
- Operating System

With Cloud based system deployment, the customer does not need to purchase any hardware, since it is provided through the Cloud Processing Services. Cloud computing ***instances*** can be created through a web console as required, and allocated to the GXL system to add/remove processing node capability. This provides unparalleled scalability and dramatically reduces the lead time for system implementation – instances can be added or removed within seconds / minutes, where non-GXL systems would require purchasing / shipping / installing / configuring specific hardware.

Case Study: Large volume processing on the Cloud

PCI Geomatics has successfully deployed its GXL System to the Amazon Cloud to process large volumes of imagery for one of its customers. The following terminology is key to understanding the GXL Cloud system and how it is deployed to the Amazon Cloud.

Gluster:	Main Data Repository (where data is stored, accessed by GXL system)
License Server :	Central node which contains all s/w licenses, dbase for GXL, QA tools
Processing Nodes:	Cloud based instances (virtual machines) that get allocated for processing, on demand
AMI:	Amazon Machine Image (preconfigured machine configuration to be used when adding new instances
S3:	Amazon Simple Storage Service – used for data storage (in the case of GXL, Gluster is the preferred method over S3 for data storage, due to more efficient handling of I/O
EC2 :	Elastic Computing – management console within Amazon Cloud Services for adding/removing computing resources
Instance:	A virtual machine – Amazon provides standard configurations that range in processing capability (i.e. Micro, Small, Large, Extra Large)

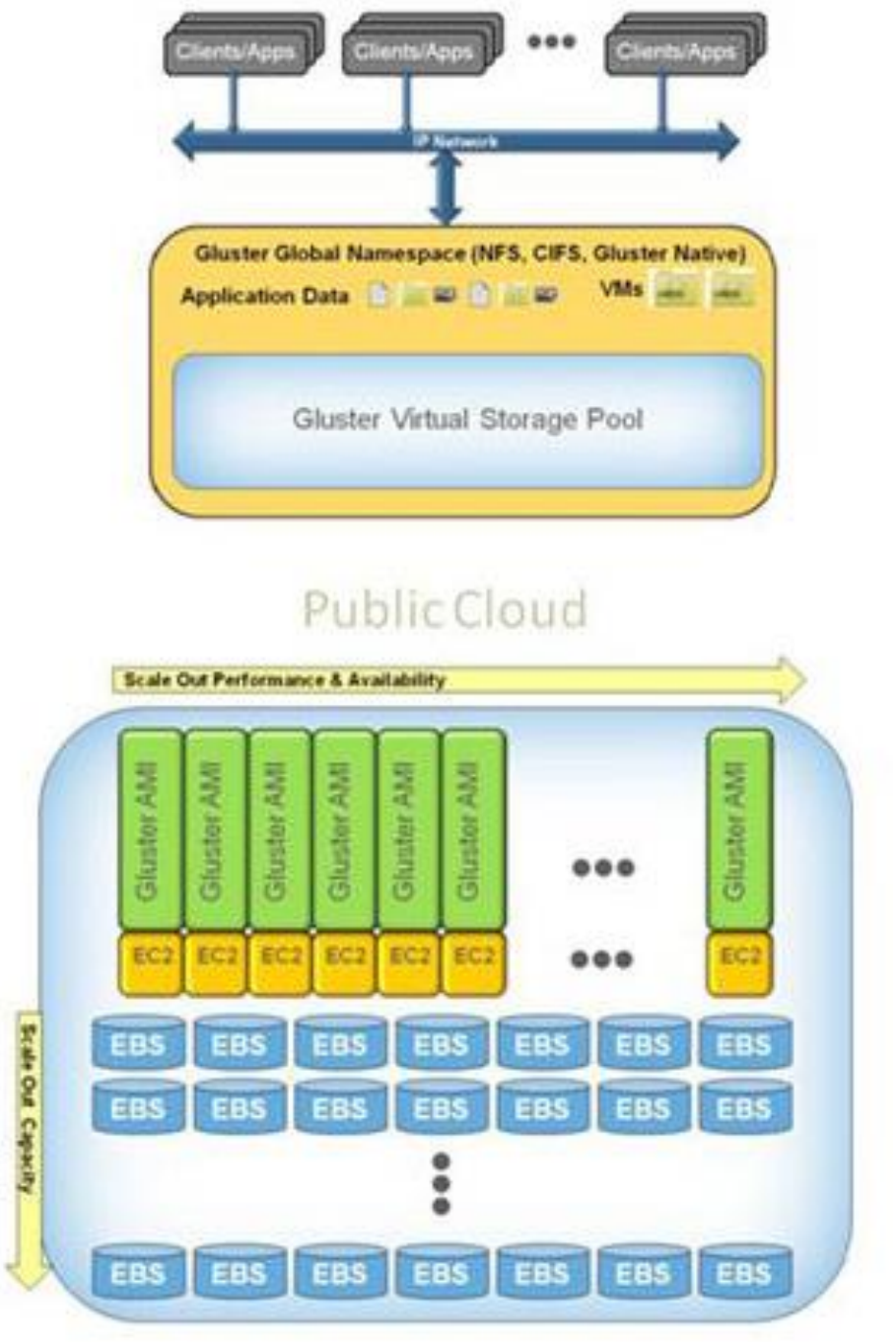

Figure 2: Amazon Cloud Based System Architecture
Similar to GXL Cloud based configuration

Once initial setup was completed, the management and use of the GXL Cloud system is very similar to the non-Cloud based GXL system, with a few exceptions. Below are a series of screen shots showing the Amazon based GXL Cloud system.

Figure 3: Login into Amazon Console
Accessible from any computer/device connected to the Internet

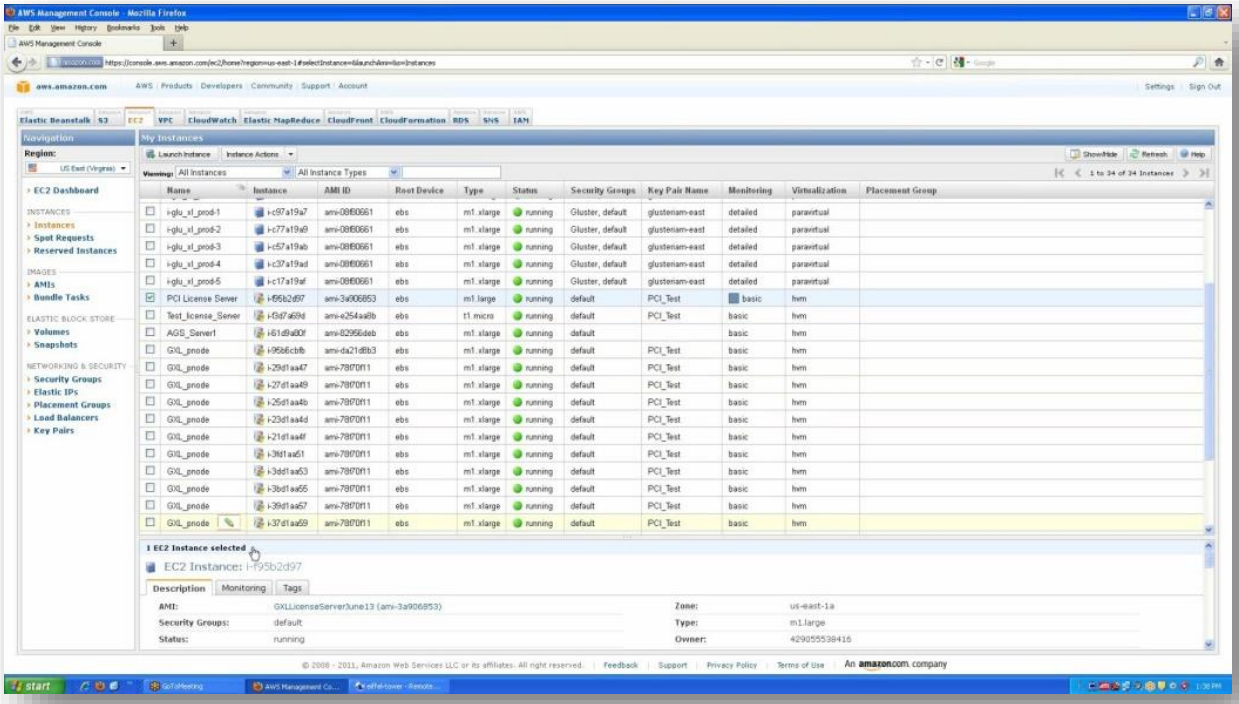

Figure 4: Amazon Management Console
Add/Remove instances, monitor usage

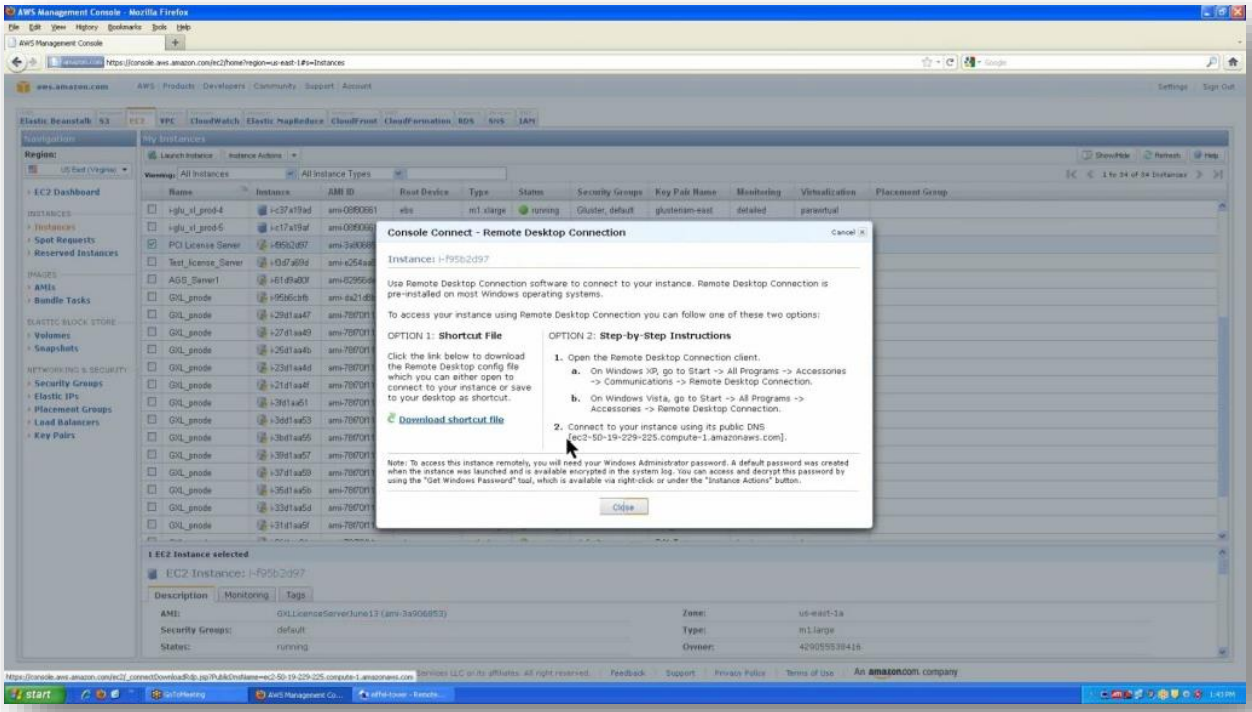

Figure 5: Accessing GXL from Amazon Console
Direct IP address, or RDP session options are available

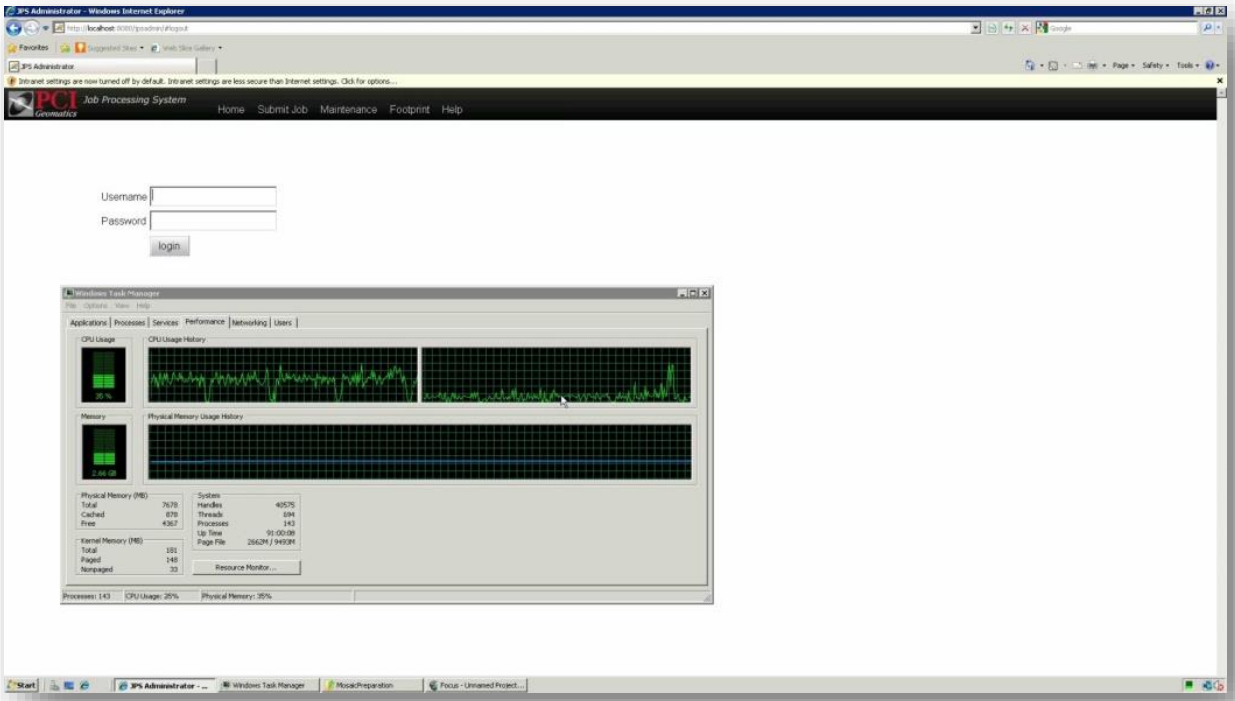

Figure 6: Logging into GXL Instance
GXL system deployed on Cloud allows monitoring/management of processing

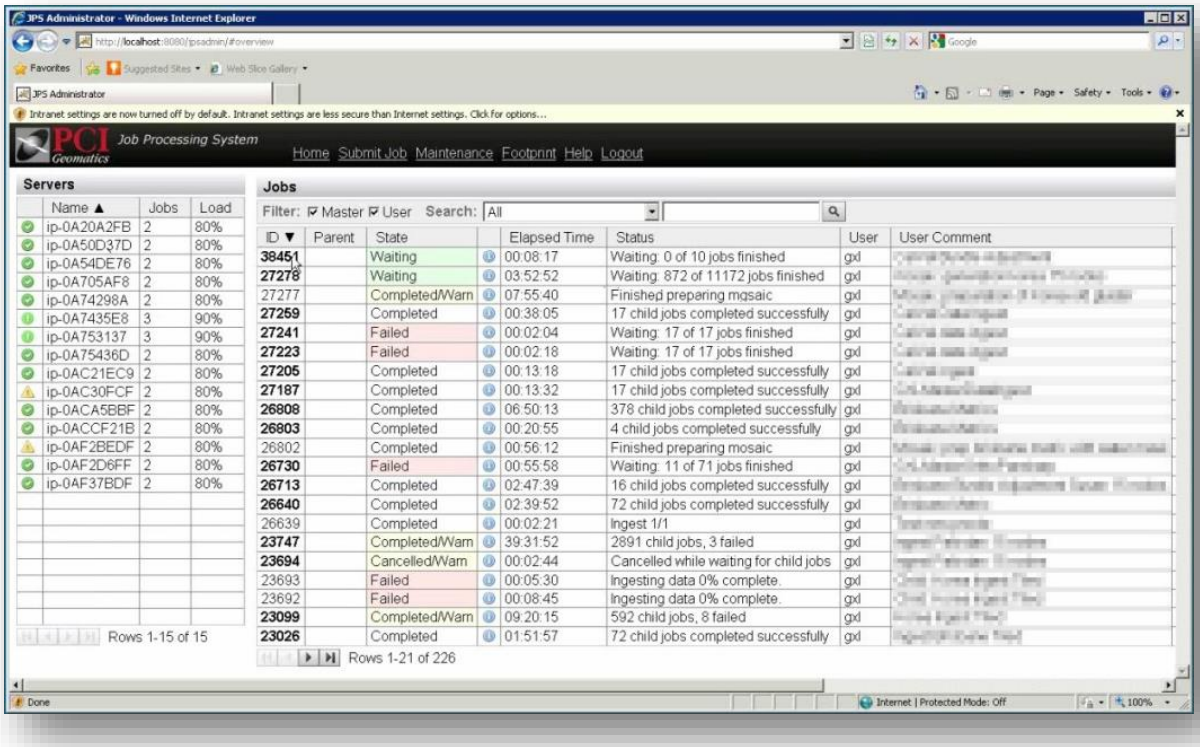

Figure 7: Processing nodes and active jobs displayed on Cloud GXL system
Note the 15 instances listed on the Server Panel.
These can be added/removed on demand, as required

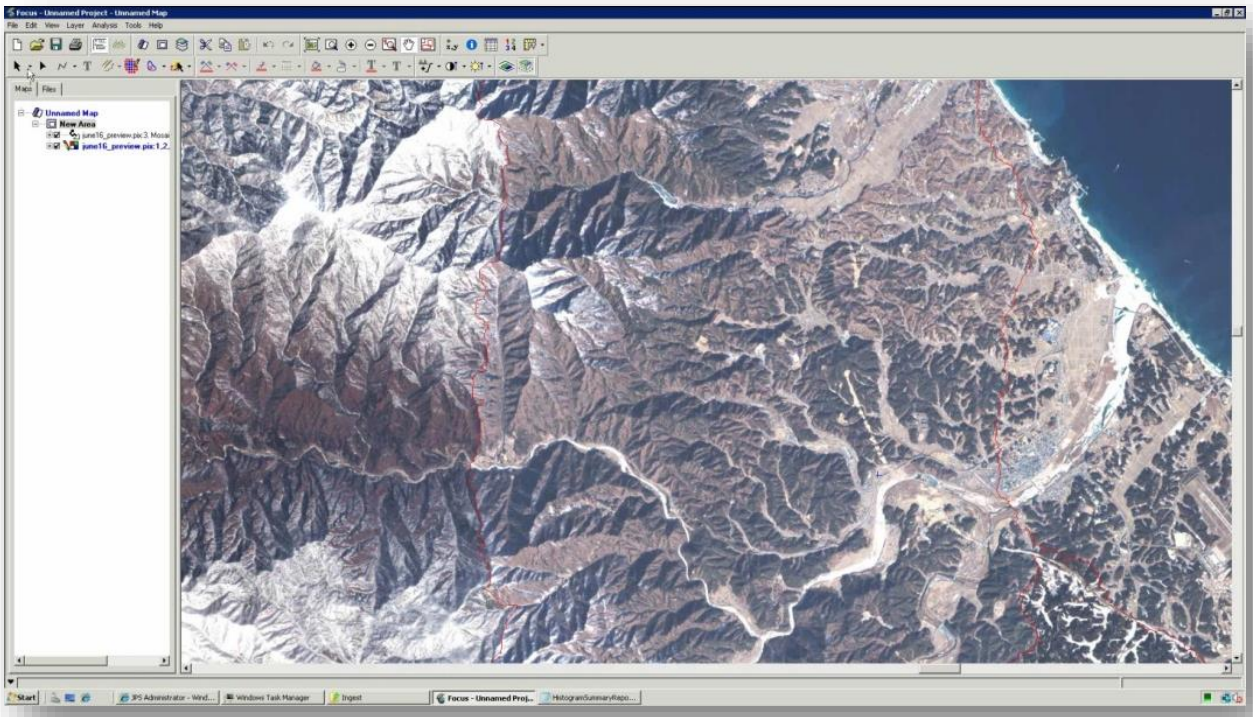

Figure 8: QA on GXL Instance on the Cloud

QA can be scaled up/down through the deployment of Cloud based instances, on demand

Business Benefits

The benefits can be summarized as follows:

Simplified, Fast System Setup

- As outlined in the paper, setting up the GXL system is very straightforward; GXL could be deployed to other Public Clouds (Microsoft Azure, for example) as we have done for the Amazon Cloud Solution. This fast, simplified setup gets an organization up and running quickly – the only caveat would be to get the initial data copied to the Cloud, where the GXL can be used to process it as described (Ortho, Pan Sharpen, Color Balance, Mosaic/Tile)

On Demand Processing

- The system does not sit idle. It is used as required, for on demand processing. This removes the headaches associated with maximizing use of purchased hardware systems to maximize the value of capital investments.

Ability to Scale / Up or down

- Accelerating throughput can be achieved by adding additional nodes, then removing them once the work is completed. The same can be said for Quality Assurance, if a large team can be deployed to work on the products to ensure high levels of quality, multiple QA instances can be created on demand to accommodate a large scale QA effort.

Summary

As this paper outlines, PCI Geomatics has made significant investments in its technology to leverage parallelization and scalability, key trends which have emerged in the IT industry. Although Cloud based services on the surface appear to be better suited to transactions that have a smaller data footprint, we have demonstrated that by staging the data directly within the Cloud, we can realize the same benefits as other Cloud based services.

PCI is continuing to experiment with Cloud based services for GXL-we are planning to deploy our GXL system to other cloud environments (Microsoft Azure) as well as experiment with other Operating Systems (Linux), in an ongoing effort to improve performance.

Contact us

PCI Geomatics Inc.
90 Allstate Parkway, Suite 501
Markham, Ontario L3R 6H3
Canada
www.pcigeomatics.com
info@pcigeomatics.com
Phone: (905) 764-0614