

## **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**

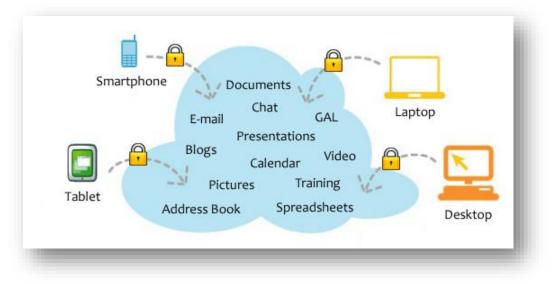
Some of the features of Cloud computing are listed below – their validity or specific challenges that apply to the Earth Observation industry are discussed.

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.

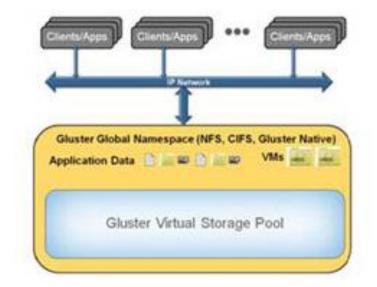


### Case Study: Large volume processing on the Cloud

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: AMI:	Cloud based instances (virtual machines) that get allocated for processing, on demand 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)





## **Public** Cloud

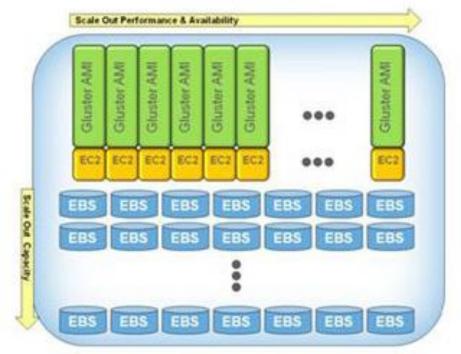


Figure 2: Amazon Cloud Based System Architecture Similar to GXL Cloud based configuration



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.

Amazon Web S	ervices Sign In	
Please enter the AWS Identit	y & Access Management (IAM) User name and password assigned by your system administrator to sign in.	
	AWS Account:	
	User Name: pci	
	Password:	
	Sign in using pur secure server	
	Please contact your system administrator if you have forgotten your user credentials.	
	Sign in using AWS Account credentials	
	Terms of Use Privacy Policy © 1996-2010, Amazon.com, Inc. or its affiliates. An amazon.com.amy	

### Figure 3: Login into Amazon Console

Accessible from any computer/device connected to the Internet



Mitps://c	onsole.sws.amazon.com/ec2/home?	region-us-east-1#sele	ctInstance+6launchA	melle=Instances							17 - C M - Grade	P .
aws.amazon.com	AWS Products Developers	Community Sup	port Account									Settings Sign Out
astic Beanstalk 53	C2 VPC CloudWatch	Flastic ManBadur	- FindEmat	FloudFormation	PDE ENE	TAN						
avigation	My Instances											the second se
egion:	Launch Instance Insta	ce Actions +										🔅 Showitkie 🗟 Ketresh 🕒 Help
US Exet (Virginia) •	Viewing: All Instances	wing: All Instances All Instance Types										€ € 1 to 34 of 34 Instances > >
EC2 Dashboard	Name	Instance	AMI ID	Root Device	Type	Status	Security Groups	Key Pair Name	Monitoring	Virtualization	Placement Group	
NSTANCES	i-glu_xl_prod-1	+c97a19a7	ami-0860661	ebs	m1_starge	anning	Gluster, default	glustenam-east	detailed	paravirtual		A
Instances	E Eglu_x1_prod-2	■ i-c77a19a9	ami-08/50661	ebs	m1_slarge	<ul> <li>nunning</li> </ul>	Gluster, default	glusteriam-east	detailed	paravirtual		
Spot Requests Reserved Instances	E i-glu_sl_prod-3	i-c57a19ab	ami-08/60661	ebs	m1_xlarge	e running	Gluster, default	glustenam-east	detailed	paravirtual		
	E iglu_xl_prod-4	■ i-c37a19ad	ami-08f80661	ebs	m1_starge	<ul> <li>running</li> </ul>	Gluster, default	glusteriam-east	detailed	paravirtual		
AMIS	iglu_xl_prod-5	i-c17a19af	ami-08€0661	ebs	m1_xlarge	<ul> <li>nunning</li> </ul>	Gluster, default	glusteriam-east	detailed	paravirtual		
Bundle Tasks	PCI License Server	A 19552097	ami-3a906853	ebs	m1.large	• running	default	PCI_Test	basic	hern		
LASTIC BLOCK STORE	Test license Server	-13d7a69d	ami-e254aa8b	ebs	t1 micro	<ul> <li>nunning</li> </ul>	default	PCI_Test	basic	him		
lolumes	AGS_Server1	161d9x80f	anvi-82956deb	ebs	m1 xlarge	<ul> <li>running</li> </ul>	default		basic	hvm		
Snapshots	GXL_pnode	19586cbfb	ami-da21d8b3	ebs	m1.starge	🔵 running	default	PCI_Test	basic	hem		
ETWORKING & SECURITY	GXL_pnode	2 i-29d1aa47	ami-78070011	ebs	m1 slarge	🔵 running	default	PCI_Test	basic	hem		
Security Groups	GXL_priode	127d1aa49	ami-78/70/11	ebs	m1.xlarge	🔵 running	default	PCI_Test	basic	tivero.		
Elastic IPs Placement Groups	GXL_pnode	🛃 i-25d1aa4b	ami-78f70f11	ebs	m1.xlarge	🥥 running	default	PCI_Test	basic	hern		
Load Balancers	GXL_pnode	iz i-23d1aa4d	ami-78/70/11	ebs	m1_xlarge	🥥 running	default	PCI_Test	basic	hem		
Key Pairs	GXL_pnode	2 i-21d1aa4f	ami-78/70/11	ebs	m1_xlarge	🔵 running	default	PCI_Test	basic	hem		
	GXL_priode	🗟 i-361aa61	ami-78f70f11	ebs	m1.starge	🔵 running	default	PCI_Test	basic	hvm		
	GXL_pnode	i-3dd1aa63	ami-78670f11	ebs	m1_xlarge	🔵 running	default	PCI_Test	basic	hem		
	GXL_pnode	i i 3bd1aa55	ami-78/70/11	ebs	m1.xlarge	nunning	default	PCI_Test	basic	him		
	GXL_pnode	🖓 i-39d1aa57	ami-78f70f11	ebs	m1.xlarge	🥥 running	default	PCI_Test	baeic	hvm		
	🖾 GXL_pnode 🦠	🛃 i-37d1aa59	ami-78/70/11	ebs	mt_slarge	🔵 running	default	PCI_Test	basic	hem		
	1 EC2 Instance selected EC2 Instance: Description Monit	I-195b2d97										1
	AMI:	GXLLicens	eServerJune13 (	ami-3a906853)				Zone:		us-east-1a		
	Security Groups:	default						Type:		m1.large		
	Status:	running						Owner:		429055538416		
				and the second second	where here addresses	Minister and	erxed.   Feedback			An	amazon.company	

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

<ul> <li>Interdio</li> </ul>	onsole avis amazon conjecz/home?v	egion-us-east-1#s-1	istances								🖞 + C 🚺 - Googe	: (حر
ees.amazen.com	AWS / Products / Developers	Community Sup	pert.   Applient									Settings Sign Or
istic Beanstalle 53	C2 VPC CloudWatch I	Jostic MapReduc	e Choulfront	Claudf armation	805 5NS	IAN						
avigation:	My bristences		_	_								
egion: US East (Vegnal •	Laurich hutseice Baters									Di Showiffide - 2 Rathash - 9 Hea		
	www.mgs All Instances		stance Types		IC C 1 to 24 of 34 Distances > >							
EC2 Dashboard		Instance	AMI ID	Rout Device		Status		Key Pair Hame			Placement Genup	
ARTANCES	Fglu_sl_prod-&	+c37A19ad		ebs	m1.slarge	Intering	Gluster, defaut	gkosterism-east	detailed	paravirtual		
Spot Requests	iglu_x1_prod-5	a +ct7at9al	ami-08/80661	Console Conn	ect - Remot	e Deskto	p Connection			Cancel #		
Reserved Instances	PCI License Server	12 - 49562057	ami-3a0685	Instance: I-f9	TRIVIAL							
	E Test_loanse_Server	10 d7,069d	ami 6254aa	tristance: P()	202032							
AMIS	AG5_Server1	105#Rb13+	ami-82955.de				re to connect to yo	our instance. Remot	e Desktop Con	nection is		
Bundle Tasks	CI GXL_pnode	12 Mibboht	am-da21d8b	pre-installed on most Windows operating systems.								
LASTEC BLOCK STORE	GRL phode	(g) +29d1#a47	anni-7817DH1	To access your	instance using	Remote D	esktop Connection you can follow one of these two options:					
Volumes	GXL_snode	+27d1aa49	smi-78970111	OPTION 1: Sho	ortcut File	File OPTION 2: Step-by-Step Instructions						
Snapshots	GXL_enode	12 +25d1aa4b	ami-78670r1		low to download 1. Open the Remote Desktop Connection client.							
ITTWORKING & SECURITY	GXL snode	E +23118+4d	smi-7807011	the Remote Desktop config file which you can either open to	1 to an Communications an Remote Desiston Connection							
Security Groups Elastic IPs	CI GXL_pnode	12 +21dtas#	ama-70870f11	connect to your to your desktop	ur instance or save							
Placement Groups	GXL_peode	131d1au51	smi-78/7011					-> Remote Deskto				
Load Balancers	GXL_pnode	( +3dd1aa53	ami-7867011	C Download st	hortcut file	2	. Connect to your				-	
Key Pairs	GXL_unode	ill +3bdTaa55	ami-78f70rt1		[ec2=50=19=229=225.compute=1.amazonaws.com].							
	GXL snode	语》A39d1aa57		Note: To access the	is instance remo	motely, you will need your Windows Administrator password. A default pessword was created and is available encrypted in the system lag. You can access and decrypt this password by rdf toul, which is available van inghtrickic or under the "Instance Adjoins" button.						
	GC_pnode	(2 +37d1aa59	am-7047011	using the "Get Win	was launched as dows Password'							
	GXL prode	-35d1aa5b	ami-7607011					1				
	GXL_pnode	12 +33dtau5d	Ami-790'Or1 1				Cidse	Cidue				
	GRL enode	12 +3tillau5f	ami-787/011	-		-			1			
	1 EE2 Instance selected		Contra La Contra									
	EC2 Instance:											
	Description Monit											
	AMI		eServerJune 13	anv-3a906853)				Zone:		us-mait-1a		
	Security Groups: Status:	default						Type: Owner:		m1.large 420055538418		
	secon. medDownloa®do.jsp70.bicDesta											

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



### WHITE PAPER

### March 2015

🔊 💌 🕅 http://kashast.000/ppssdwir/#lago.t	🗾 📄 🎋 🗙 🕅 Groups	. I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I
Setter S Constant for * 2 web the later *		
i Advesto dar	[5] + [3] + □1 mit + Page + Safety + 1	fools + 📦
areit settings are now turned off by default. Intri area settings are less secure than internet settings, Click for options		
POLY Job Processing System Home Submit Job Maintenance Footprint Help		
Geomatics Home submit up Martenance Footprint, Help		
Usemanne		
Pessword		
login		
Diffusions Task Manager		
The cytoric fair help		
Applications Processes Services Performance Networking Users		
OPU Lisage CPU Lisage History		
I ANALY AND A ALL AND A		
Memory Hrysical Hemory Usage Holoxy		
Physical Hessory (HB) System and		
Total         7028         Handles         40575           Cadred         878         194         194           Prese         492.7         Processor         193		
mes         stat         125           Tup Tracessis         243           Tup Trace         500.000           Tup Trace         500.000           Tup Trace         500.000		
Total 181 Paged 148 ,		
Noregand 30 Resource Nonitor,		
Processes 143 CPULIaces 29% Process Reserve 29%		
Processes 1G DVUsage 2% Physical Henery 2%		
Processes: 1-10 (2010kage: 29% Physical Mesoury: 29%		
Prozense: 14 (29) Ulage: 29% Physical Pleaser: 39%		
Prozenen: 14 (29) Ulage: 29% Physical Pennery: 39%		
Processes 10 (70) Ukage 29% Physical Phenery: 29%		

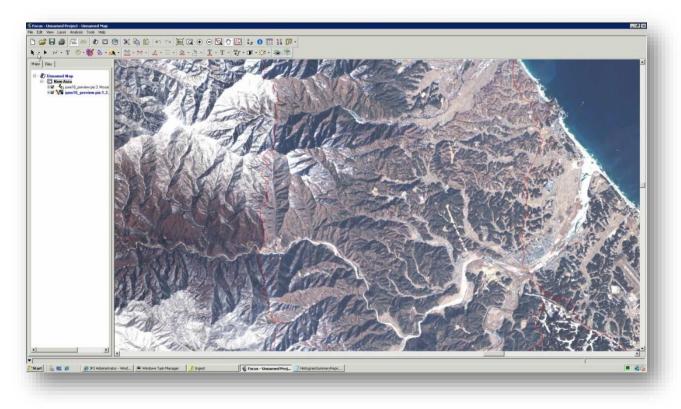
## Figure 6: Logging into GXL Instance

GAL system deployed on	Cloud allows monitoring/mana	gement of processing

a	avorites 🛛 🝰 🌄 Sug	gested Site	s • 🙋 Web	Sice Gallery	6							
37	PS Administrator											👌 • 🔂 - 🛄 📾 • Page • Safety • Tools • 👔
2.27		turned off I	by default. Intr	ranet settings	are less secur	e than Internet settings. C	ick fo	r options				
	- International Arriston											
	Geomatics	Process	sing Syste	m He	ome Sub	mit Job Maintenar	ice	Footprint Help	Logout			
			_									
Se	ervers	i Nor V		Jobs								
	Name A	Jobs	Load	Filter:	Master	♥ User Search:	A	1		a		
9	ip-0A20A2FB		80%	DV	Parent	State		Elapsed Time	Status		Us	er User Comment
0	ip-0A50D37D		80%	38451	1 Grone	Waiting	10	00.08.17		of 10 jobs finished	ad	
0	ip-0A54DE76		80%	27278		Waiting		03:52:52		372 of 11172 jobs finished	ax	North Automatications Philipped
0	ip-0A705AF8		80%	27277		Completed/Warn		07:55:40		preparing mgsaic	ax	More physical and the second physics
0	ip-0A74298A ip-0A7435E8		80% 90%	27259		Completed		00:38:05		obs completed successfully	ad	Carry Colorador - Colorador - Colorador - Colorador
9	ip-0A7435E8		90%	27241		Failed		00:02:04		17 of 17 jobs finished	and	Calify the later of the
0	ip-0A75436D		80%	27223		Failed		00:02:18		17 of 17 jobs finished	and	Calification and a second
0	ip-0AC21EC9		80%	27205		Completed		00:13:18		obs completed successfully	and	Californiagen
	ip-0AC30FCF		80%	27187		Completed		00:13:32		obs completed successfully	and	Civil Adapted Campbelland
0	ip-0ACA5BBF		80%	26808		Completed		06:50:13		jobs completed successfully		Relationshipseling
õ	ip-0ACCF21B		80%	26803		Completed		00:20:55		os completed successfully	and	Personal Address
À	ip-0AF2BEDF		80%	26802		Completed		00:56:12		preparing mosaic	gpd	Moule prop bridgets made with subscription
0	ip-0AF2D6FF		80%	26730		Failed		00:55:58		11 of 71 jobs finished	and	Col Administratives, Paretones
0	ip-0AF37BDF		80%	26713		Completed		02:47:39		obs completed successfully	and	fersional factor inputtion factor filmates
				26640		Completed	0	02:39:52		obs completed successfully	and	Paraulturi Antia
				26639		Completed		00:02:21	Ingest 1/1		gpd	Tanat menuarita ita
_				23747				39:31:52		d jobs, 3 failed	gpd	ingentification illinoiste
_			-	23694				00:02:44		d while waiting for child jobs	god	ingentifiate dan Unreden
-				23693		Failed	0	00:05:30	Ingesting	data 0% complete.	gxd	Chief increase ingent (Time)
-				23692		Failed	۲	00:08:45	Ingesting	data 0% complete.	gxl	Civil Proves Figers 7 line)
				23099		Completed/Warn	0	09:20:15	592 child	jobs, 8 failed	gid	include Report Title?
	Rows	s 1-15 of	f 15	23026		Completed	0	01:51:57	72 child jo	obs completed successfully	gpd	Real Provide Trial
				let.bet.le	N Ro	ws 1-21 of 226						
Do							_					Internet   Protected Mode: Off

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

