Big Data Earth Observation Processing

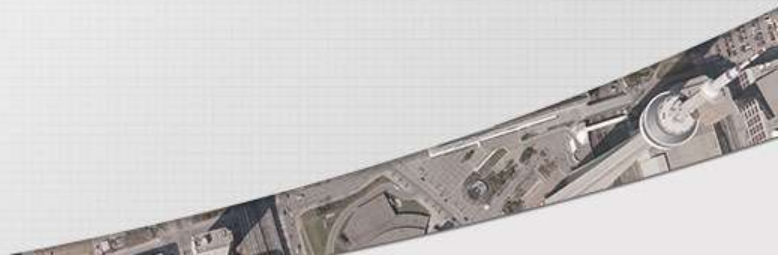
Shawn Melamed
Tech. Solution Specialist

About PCI and Big Data

- In 2007, PCI Geomatics developed it's first high volume Earth Observation (EO) processing system called the GeoImaging Accelerator (GXL)
- The GXL was initially created as a customer solution to orthorectify SPOT-4/5 imagery for all of Canada, as part of the National Imagery Project (NIP) http://www.pcigeomatics.com/pdf/case_study_NIP.pdf

About PCI and Big Data

- The National Imagery Project (NIP) was announced in 2007 (8 years ago)
- Over $2.4 million dollars was invested in this project for data acquisition, processing and dissemination
- 11.5 million km^2 of SPOT-4/5 (10m PAN and 20m MS) data was processed
- The data is freely available to all Canadians and non-Canadians at www.geobase.ca



About PCI and Big Data

- Since the NIP project in 2007, PCI Geomatics has successfully delivered dozens of GXL systems worldwide

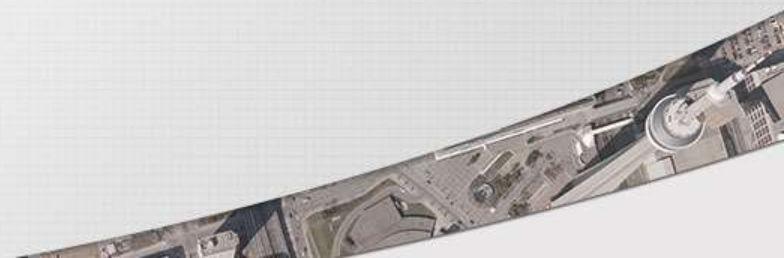
PCI GEOMATICS

Earth Observation by the Numbers

PCI GEOMATICS

EO Market by the Numbers

- In the last 10 years 179 civil and commercial EO satellites >50kg have been launched[2]
- Top nations by number of EO satellites include: China 25.5%, USA 23.5%, India 7.29%, Germany 4.69% and Russia 3.65%[1]
- It is expected that over 400 more satellites will be launched in the next 10 years (not including micro satellites)[2]



EO Market by the Numbers

- In 2024 the market for commercial EO data is expected to reach $3.5 billion[2]
- Largest growth markets, expected to be Asia, Latin America and Africa[2]
- Major applications: Natural resource management, infrastructure and defense[2]

1 Pixalytics - http://www.pixalytics.com/how-many-eo-space/

2 Euroconsult - http://www.euroconsult-ec.com/shop/earth-observation/74-satellite-based-earth-observation-market-prospects-to-2024.html

EO Market by the Numbers

- This means there will be many Exabytes (1024 Petabytes) of data
- This will require specially designed systems and algorithms to convert EO data to EO information

PCI GEOMATICS

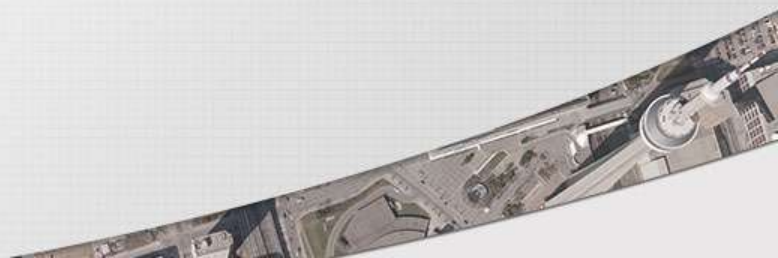
Fun Fact

What is the oldest EO satellite currently in operation?

Hint: It was expected to be operational for 1 year!

Hint: It has been operational for over 22 years and counting

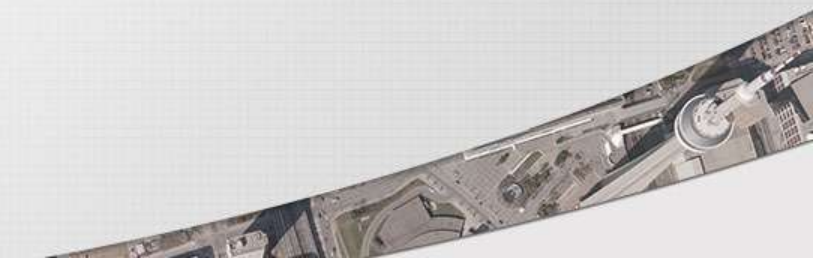
Answer: The Brazilian made Satélite de Coleta de Dados (SCD-1)



Fun Fact

What can we deduce from this?

The world needs Latin America to build more EO satellites



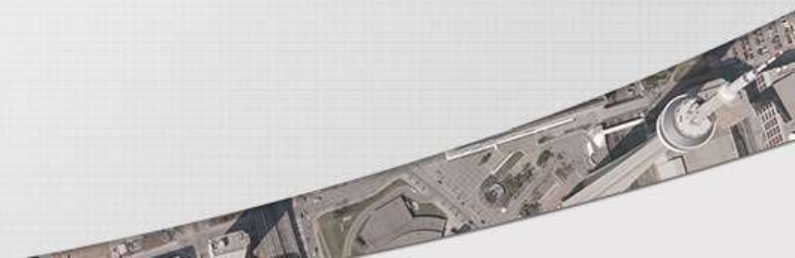
GXL

Big Data Processing Systems

PCI GEOMATICS

What is the GXL?

A high volume *photogrammetric & mosaicking* production system – GeoImaging Accelerator



Designed for Limited Operators

Distributed Production Systems

Desktop Processing

Ease of Use – Web Portal

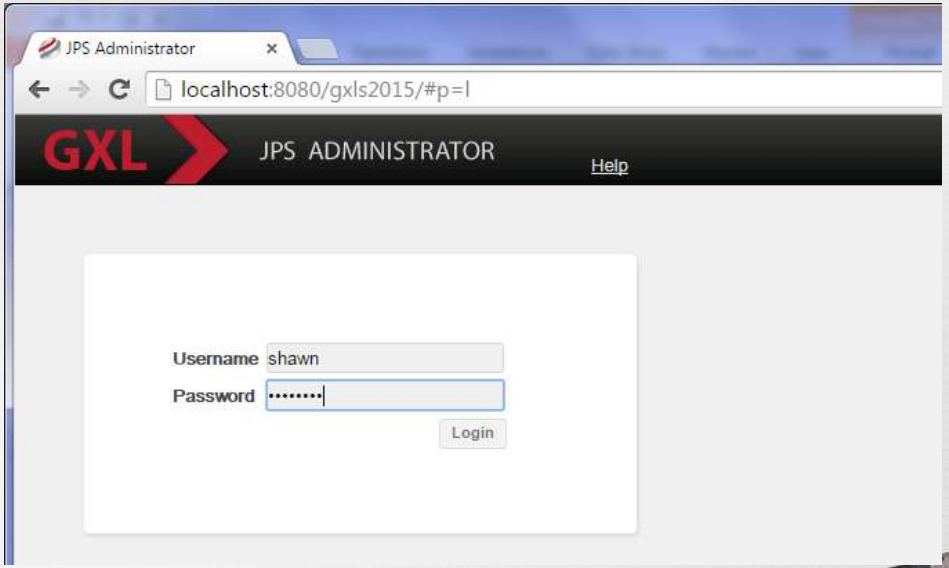
Open Configuration – Access the GXL from any computer on the web

Ease of Use – Web Portal

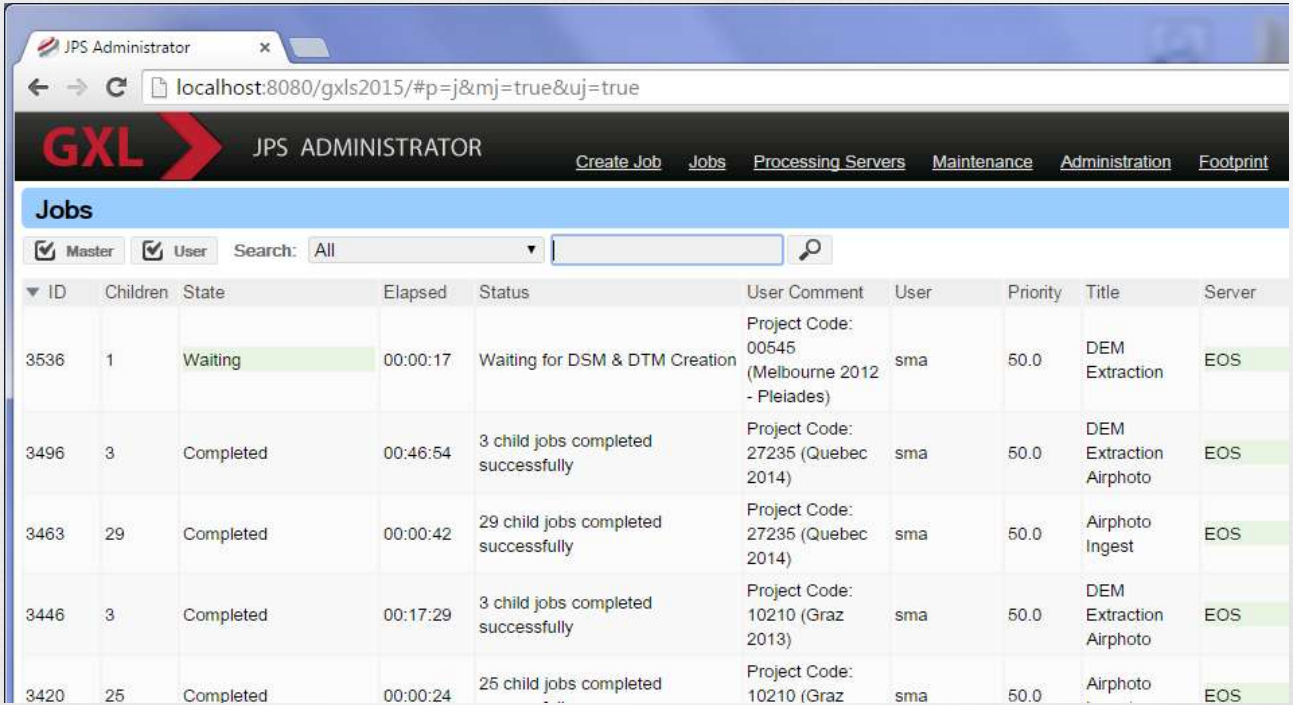
Secure Configuration – Access the GXL from any computer on your local network

Ease of Use – Web Portal

- Securely access the GXL from any computer, tablet or phone on the GXL's network
- Multiple User Access



Ease of Use – Monitor & Manage Batch Projects



ID	Children	State	Elapsed	Status	User Comment	User	Priority	Title	Server
3536	1	Waiting	00:00:17	Waiting for DSM & DTM Creation	Project Code: 00545 (Melbourne 2012 - Pleiades)	sma	50.0	DEM Extraction	EOS
3496	3	Completed	00:46:54	3 child jobs completed successfully	Project Code: 27235 (Quebec 2014)	sma	50.0	DEM Extraction Airphoto	EOS
3463	29	Completed	00:00:42	29 child jobs completed successfully	Project Code: 27235 (Quebec 2014)	sma	50.0	Airphoto Ingest	EOS
3446	3	Completed	00:17:29	3 child jobs completed successfully	Project Code: 10210 (Graz 2013)	sma	50.0	DEM Extraction Airphoto	EOS
3420	25	Completed	00:00:24	25 child jobs completed	Project Code: 10210 (Graz	sma	50.0	Airphoto	EOS

Simple table to track and access all batch processing projects (jobs)

PCI GEOMATICS

GXL

Big Data Processing Algorithms & Tools

PCI GEOMATICS

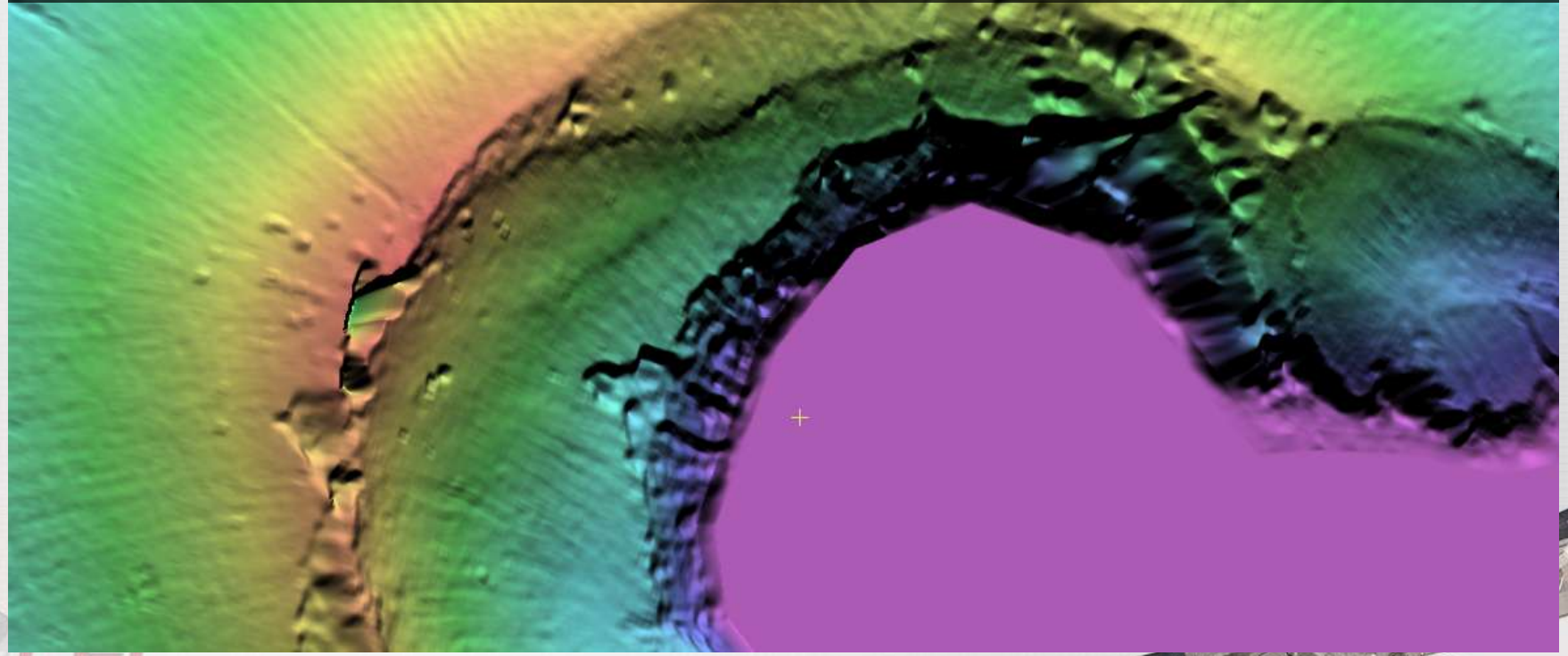
High Performance DEM Extraction

Source	Original	Current	Improvement
UltraCam airphoto	160s	40s	4x
WorldView 2 (hi-res satellite)	35m	11m	3x

- DEM extraction specifically improved (epipolar generation and geocoding unchanged)
- With all steps included total times are 1.5x to 2x faster

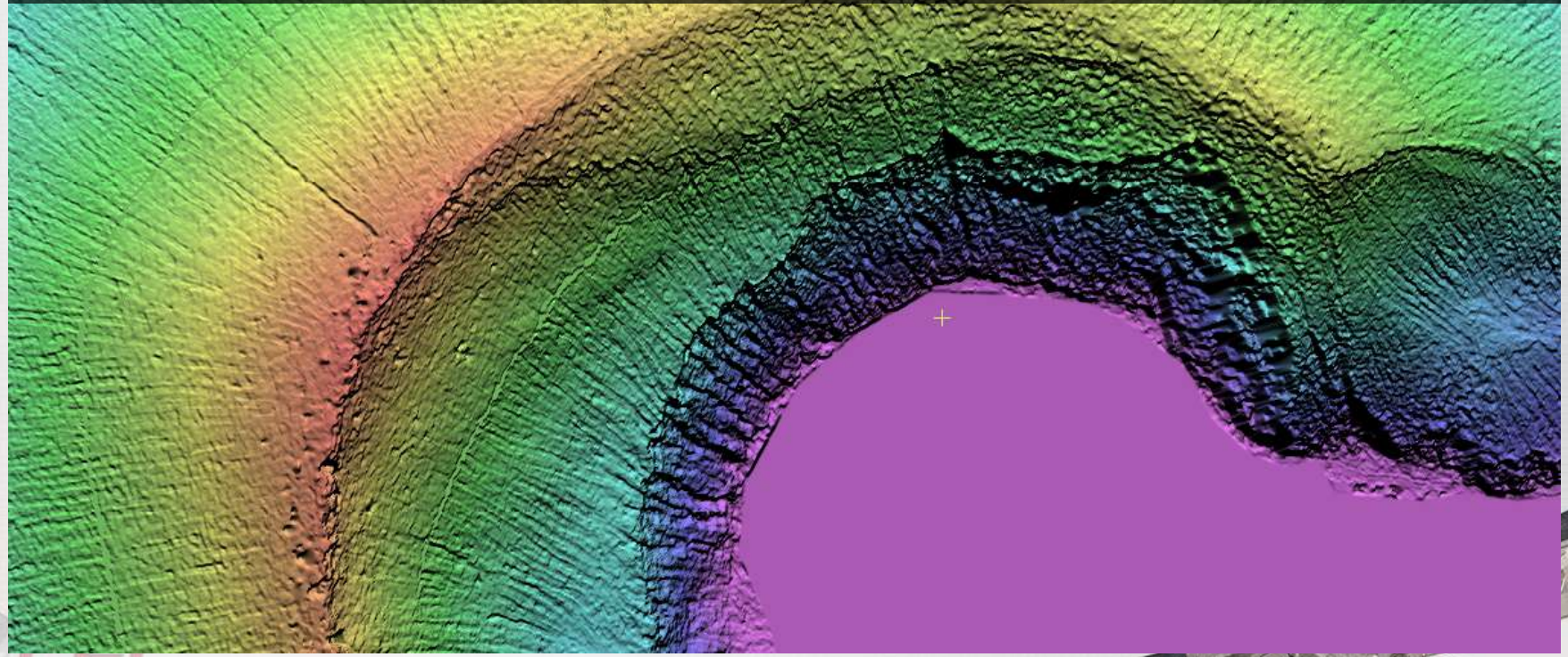
DEM Extraction - Quality

WV-3 DEM –Extracted with old algorithm



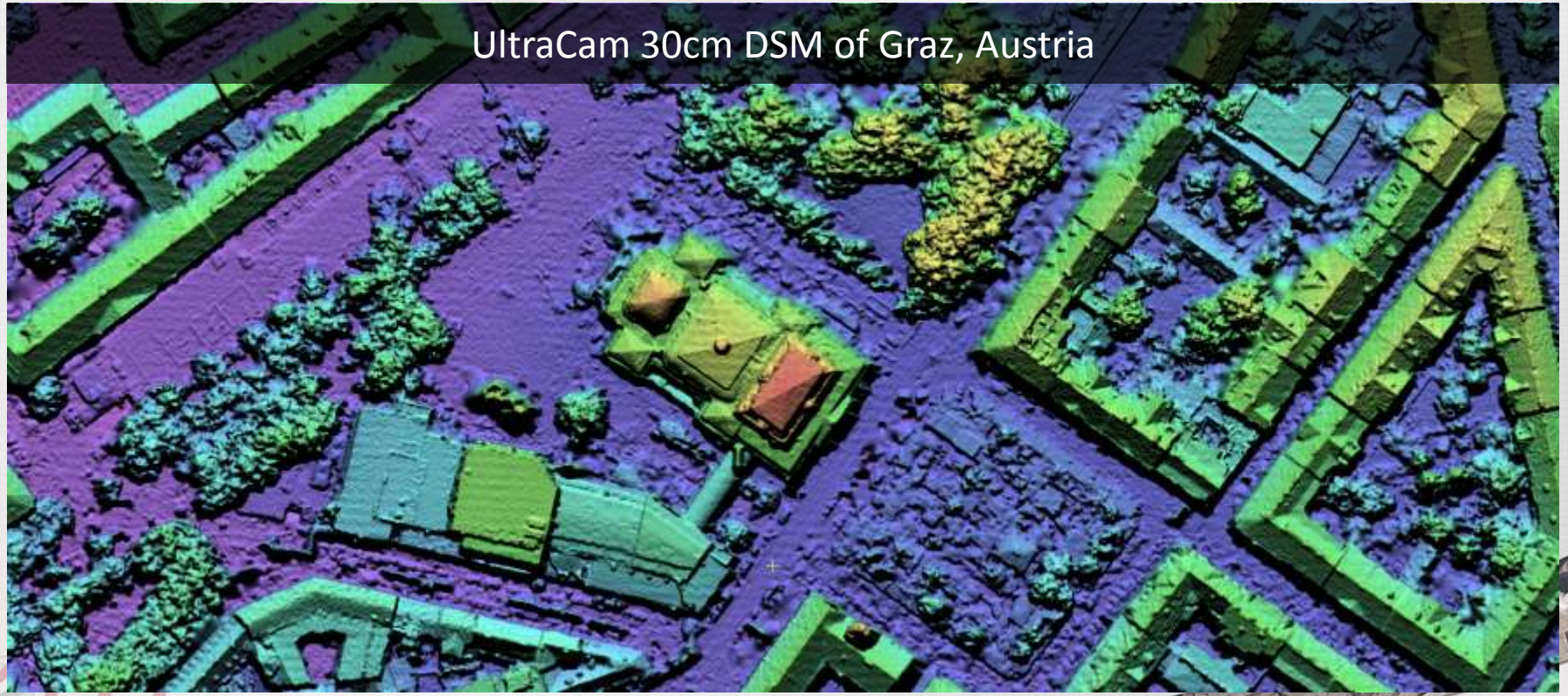
DEM Extraction – Quality

WV-3 DEM –Extracted with new algorithm

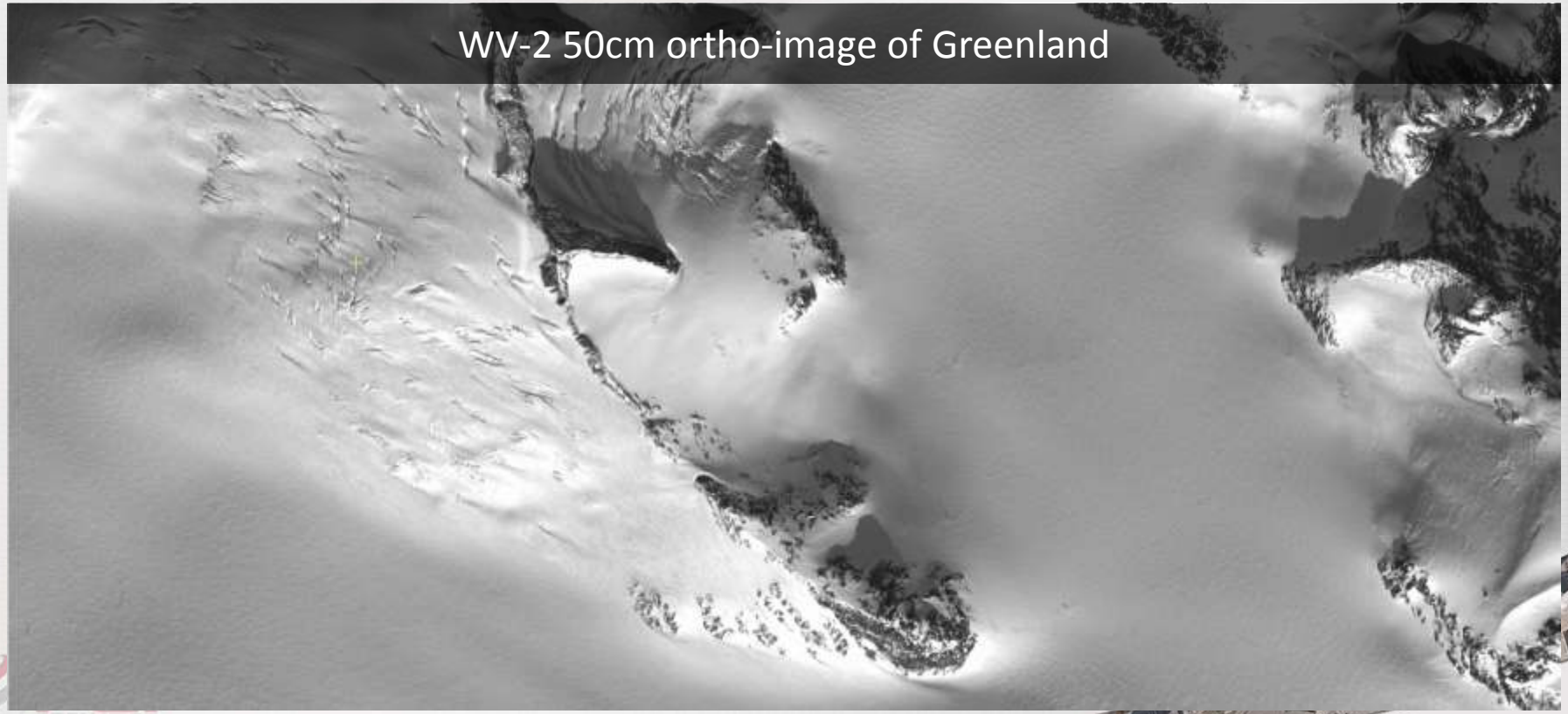


High Quality DEM Extraction

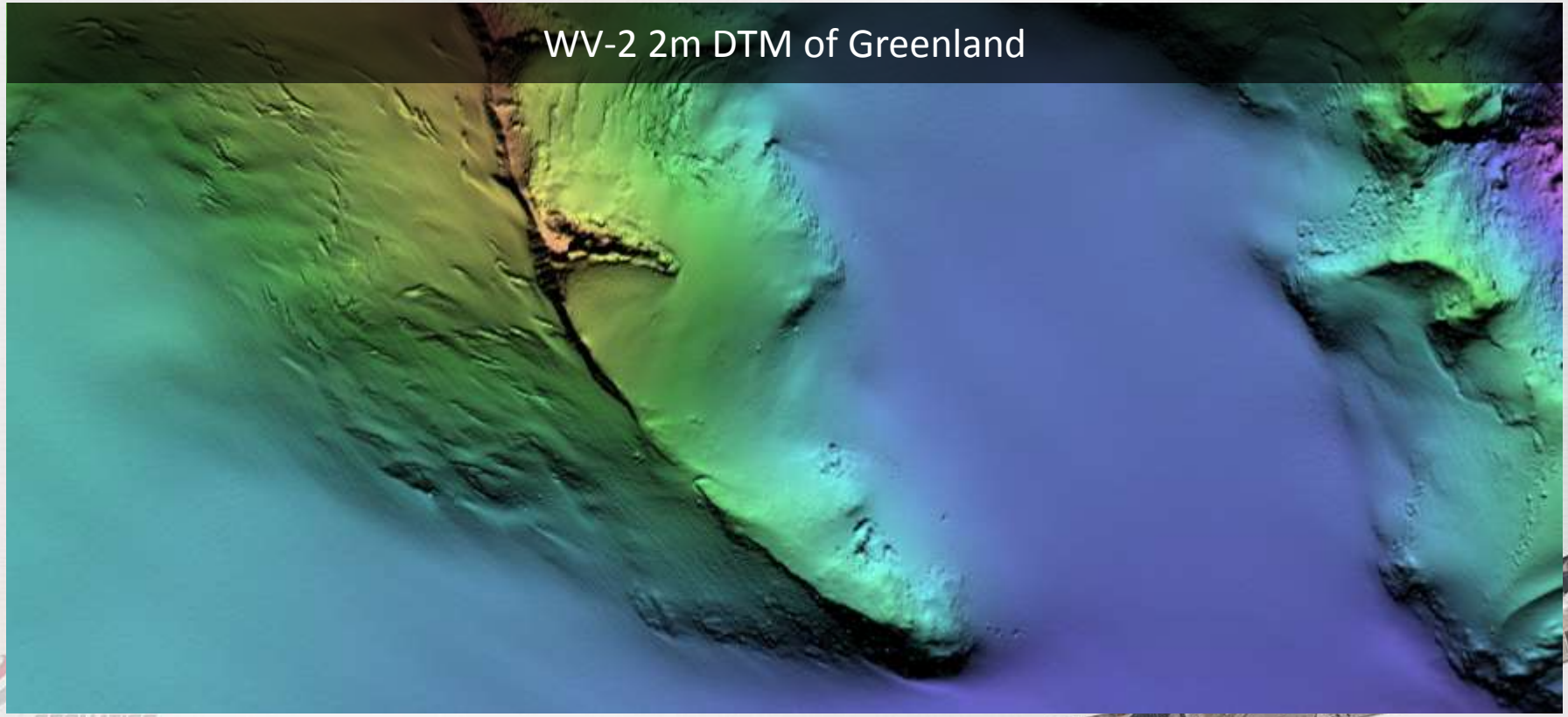
High Quality DEM Extraction



High Quality DEM Extraction



High Quality DEM Extraction



Automatic Mosaicking Quality

None

Bundle

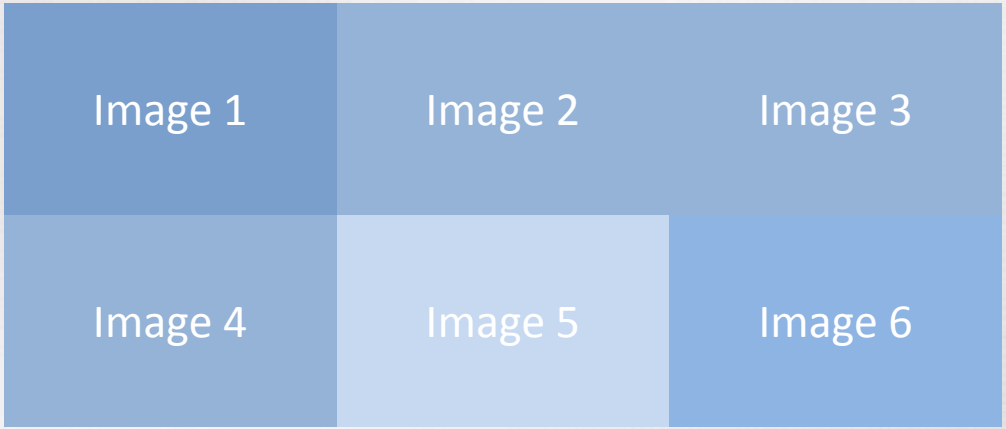
Traditional Colour Balancing Algorithms

- Traditional Method uses a single set of coefficients to define the brightness and contrast of each colour band

- Calculation and use of band specific gain and bias coefficients for color balancing

Traditional Colour Balancing Algorithms

- Traditional method is more likely to create checkerboard effect

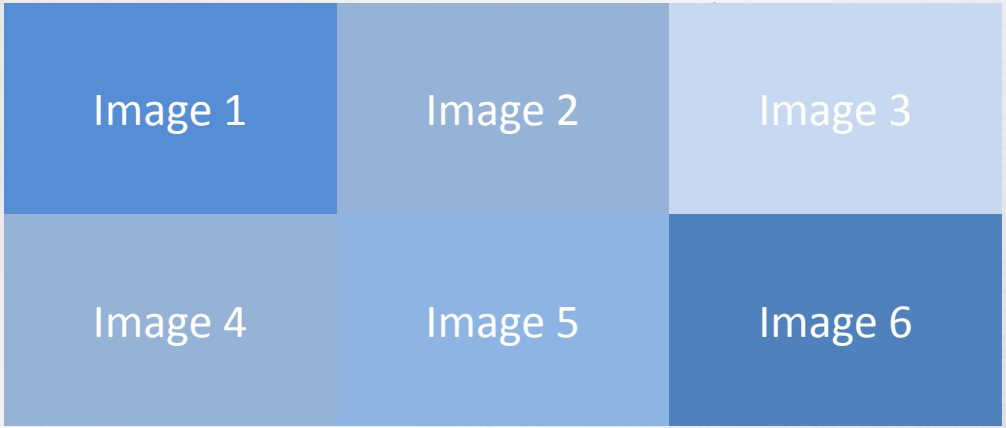


New Colour Balancing Algorithm

- New colour balancing algorithm is called ‘Bundle’
- It consists of 2 primary steps to balance a set of images for mosaicking
  1. Coarse Balancing
  2. Local Balancing (edges)

New Colour Balancing Algorithm

- The Coarse balancing step is a global operation and performs an initial balance on the entire image based on all images in the mosaic

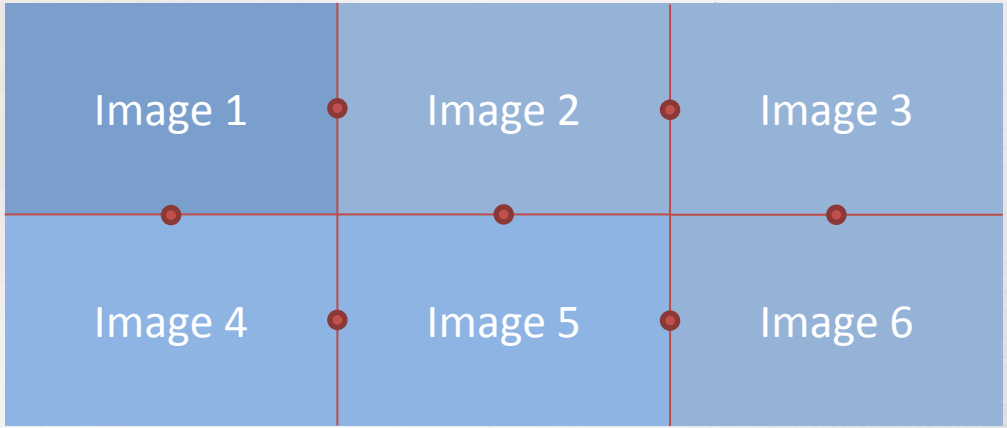


New Colour Balancing Algorithm

- This improves the balancing of the images for the final step (local balance)
- The **global mean** and **sigma** are preserved to ensure the natural appearance of the images is retained

New Colour Balancing Algorithm

- The final step uses the coarse balancing results and further refines the colours along the edges only
- A set of coefficients is created for each overlap region and then blended



- Calculation and use of band specific gain and bias coefficients for color balancing

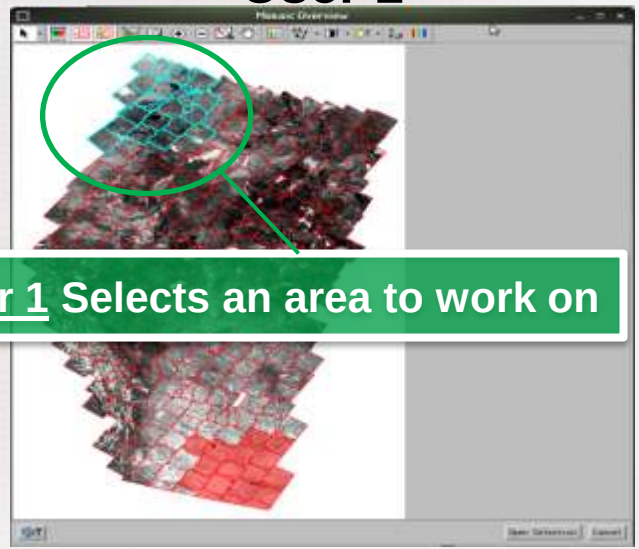
New Colour Balancing Algorithm

- Only pixels along the edges are further adjusted and then blended into the rest of the mosaic
- This improves the balancing and image fidelity

Interactive Tools – Mosaic Tool

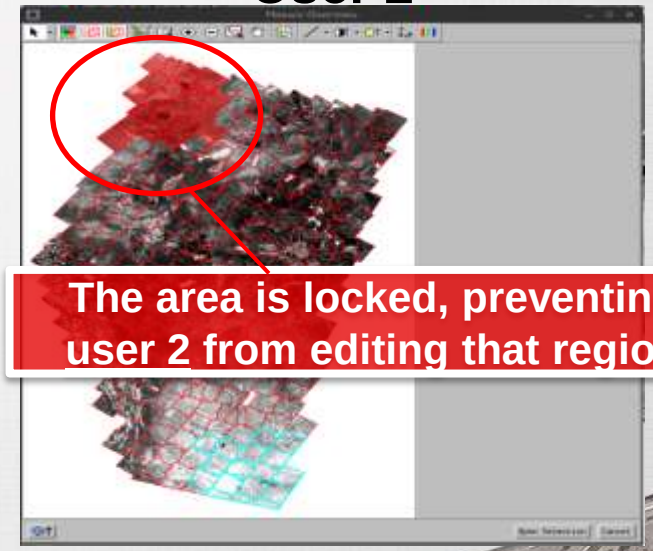
Multiple users on separate machines can quality check and edit different regions of the same mosaic at the same time

User 1



User 1 Selects an area to work on

User 2

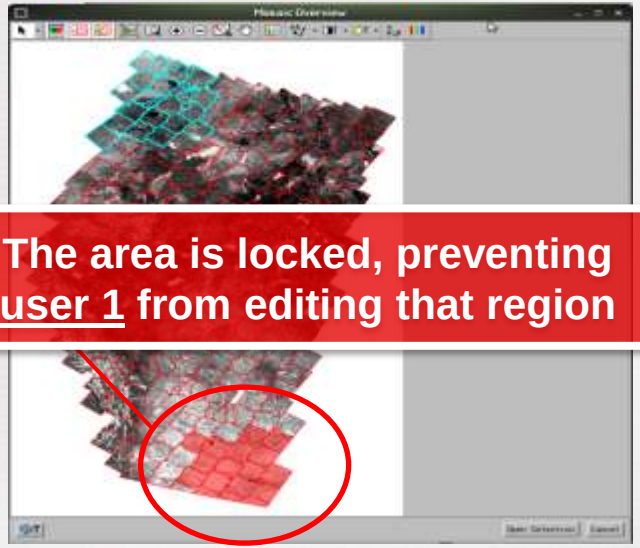


The area is locked, preventing user 2 from editing that region

Efficient Editing Tools

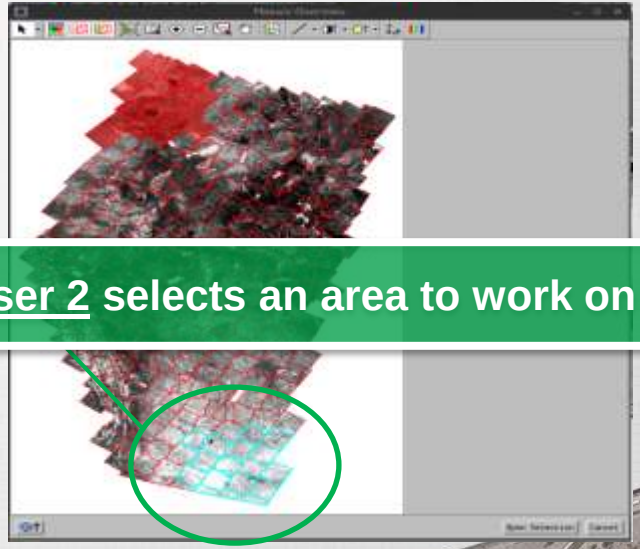
Multiple users on separate machines can quality check and edit different regions of the same mosaic at the same time

User 1



The area is locked, preventing user 1 from editing that region

User 2



User 2 selects an area to work on

GXL

Big Data Processing Architectures

PCI GEOMATICS

Performance – Metrics

GXL-Satellite

Ingest → GCP Collection → Bundle Adjustment → Pansharp → Ortho → Mosaicking

Sensor:	Output:	Area:
RapidEye*	840 GB/day	1 400 000 km^2/day (5m)
WV-2 (4-node)	1200 GB/day	256 000 km^2/day (0.5m)
Ikonos (Cloud)	3 TB/day	600 000 km^2/day (1.0m)

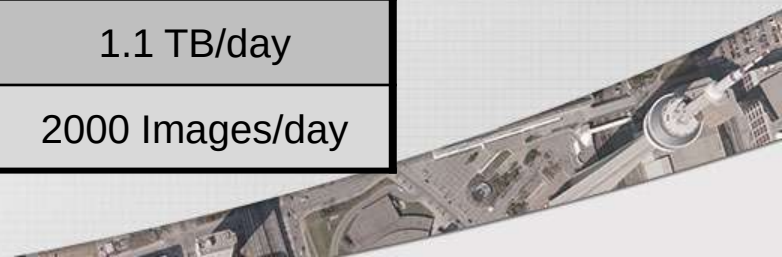
*Rapideye data cannot be pansharpened

Performance – Metrics

GXL-Aerial

Ingest → DEM Extraction → Ortho → Mosaicking

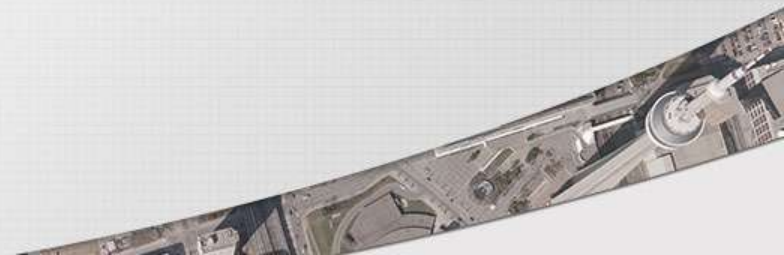
Ortho-Mosaic:	UltraCam X	UltraCam Xp
Project:	3300 Images	4500 Images
Total Time:	17.5 Hours	52.5 Hours
Output:	1.8 TB/day	1.1 TB/day
Speed:	4500 Images/day	2000 Images/day



Performance – Method Overview

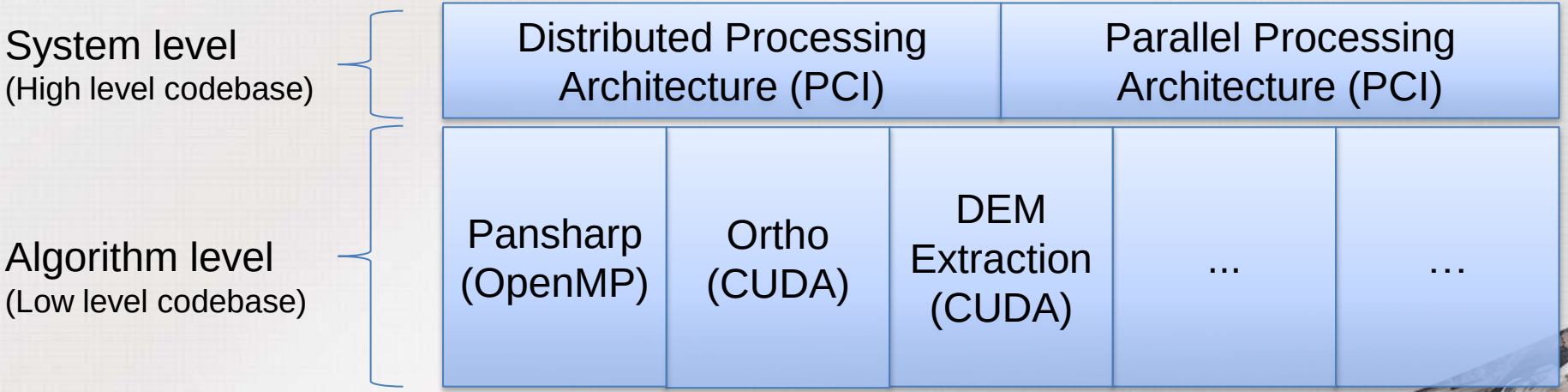
The GXL's industry leading throughput is a result of well thought-out code implemented at both the algorithm and system levels

- Multi-threaded functions (Algorithm level)
- GPU Processing (Algorithm level)
- Parallel processing (System level)
- Distributed processing (System level)



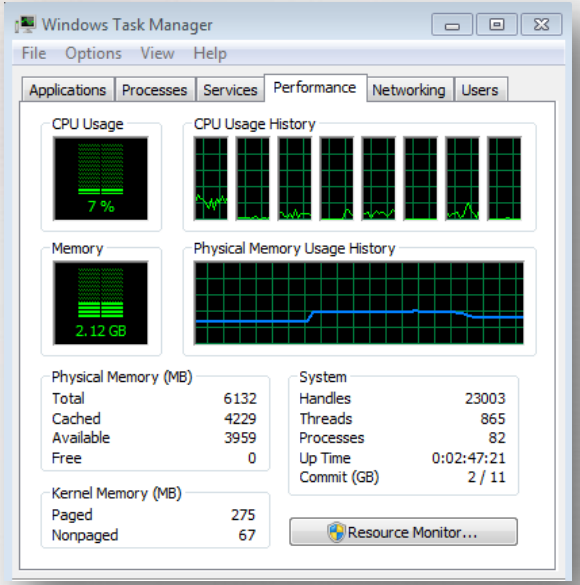
Performance – Method Overview

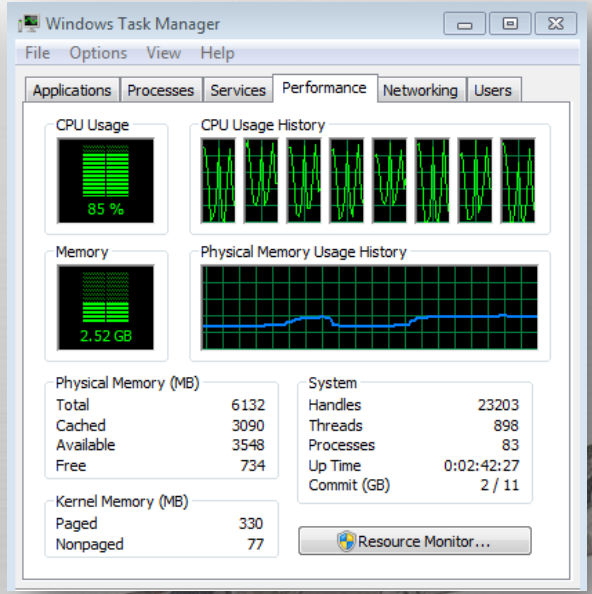
Strong emphasis on processing speed at all levels of code



Performance - Multi-Threaded Processing

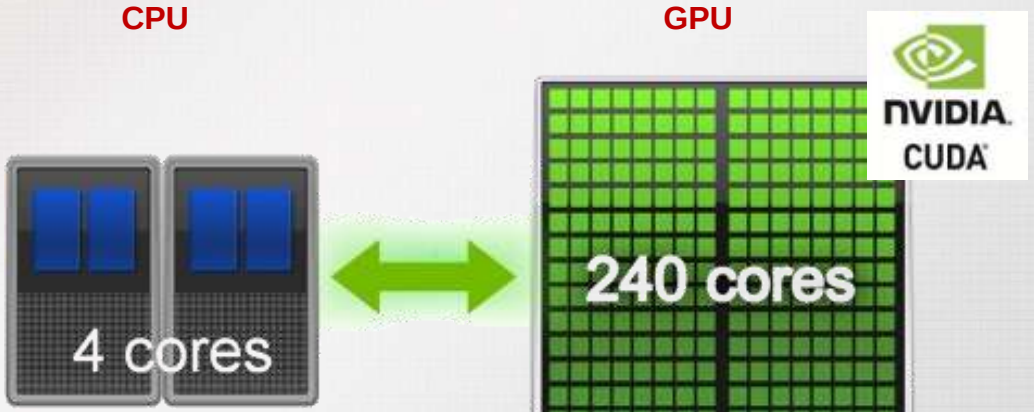
Many of the GXL algorithms are programmed using OpenMP standards to take advantage of modern multi-core CPUs



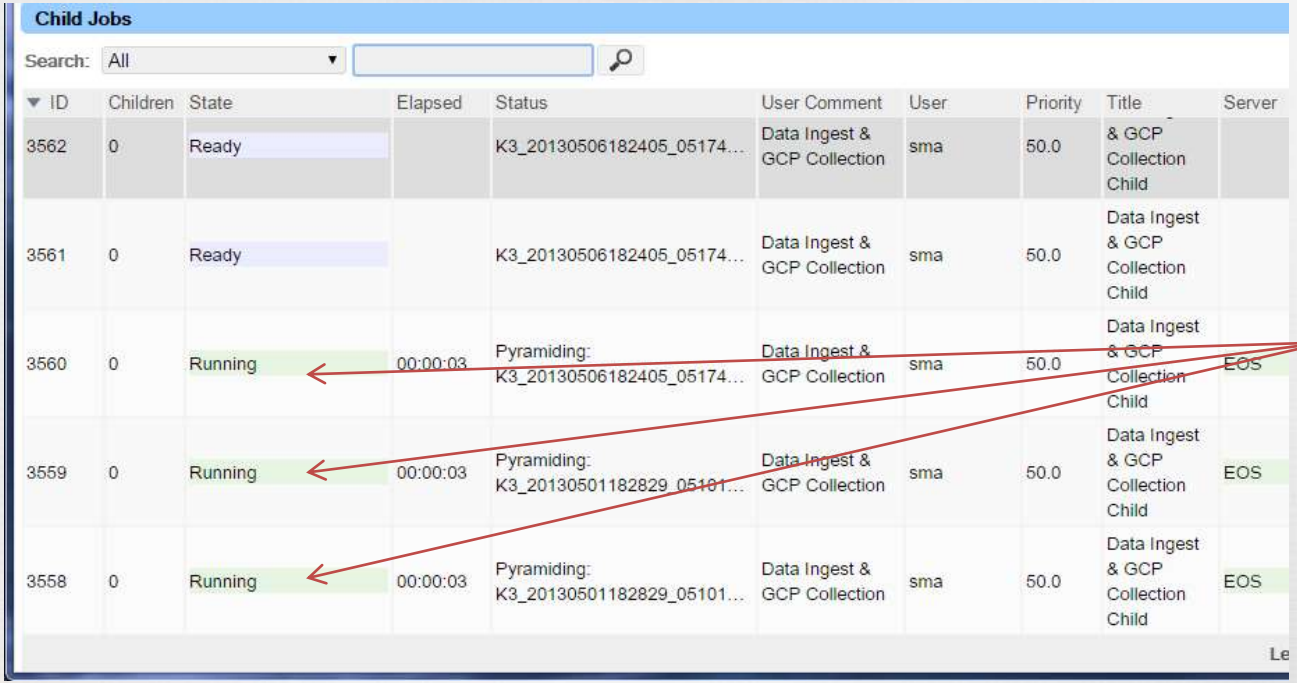


Performance - GPU Processing

- A GPU contains hundreds of cores capable of performing hundreds of identical parallel processes on different chunks of data.
- GPU processing can significantly improve the net processing speeds of certain algorithms



Performance - Parallel Processing



Child Jobs

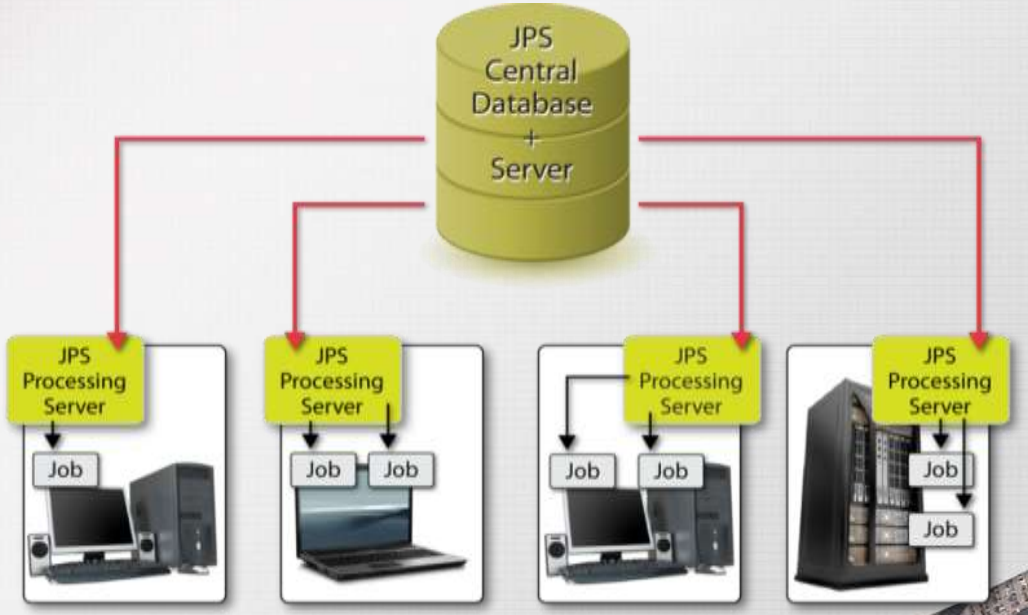
Search: All

ID	Children	State	Elapsed	Status	User Comment	User	Priority	Title	Server
3562	0	Ready		K3_20130506182405_05174…	Data Ingest & GCP Collection	sma	50.0	& GCP Collection Child	
3561	0	Ready		K3_20130506182405_05174…	Data Ingest & GCP Collection	sma	50.0	Data Ingest & GCP Collection Child	
3560	0	Running	00:00:03	Pyramiding: K3_20130506182405_05174…	Data Ingest & GCP Collection	sma	50.0	Data Ingest & GCP Collection Child	EOS
3559	0	Running	00:00:03	Pyramiding: K3_20130501182829_05101…	Data Ingest & GCP Collection	sma	50.0	Data Ingest & GCP Collection Child	EOS
3558	0	Running	00:00:03	Pyramiding: K3_20130501182829_05101…	Data Ingest & GCP Collection	sma	50.0	Data Ingest & GCP Collection Child	EOS

3 jobs running in parallel

Performance - Distributed Processing

GXL Processes can be autonomously distributed among all active processing servers in the cluster

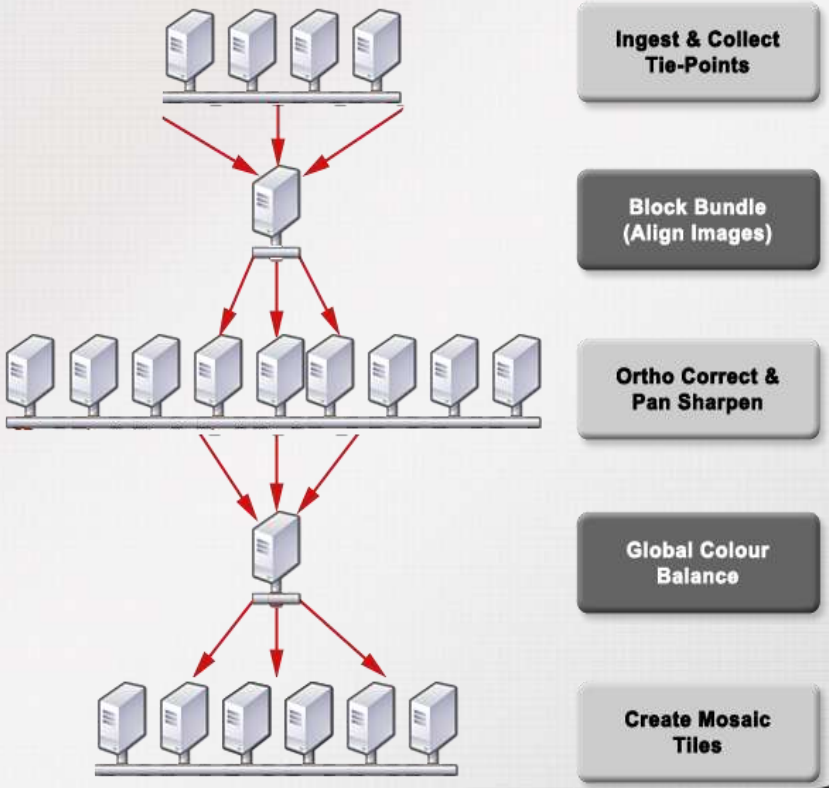


Performance – Cloud Processing

The Cloud's scalable architecture is perfect for GXL

Add nodes when you need them…

Take them away when you don't…



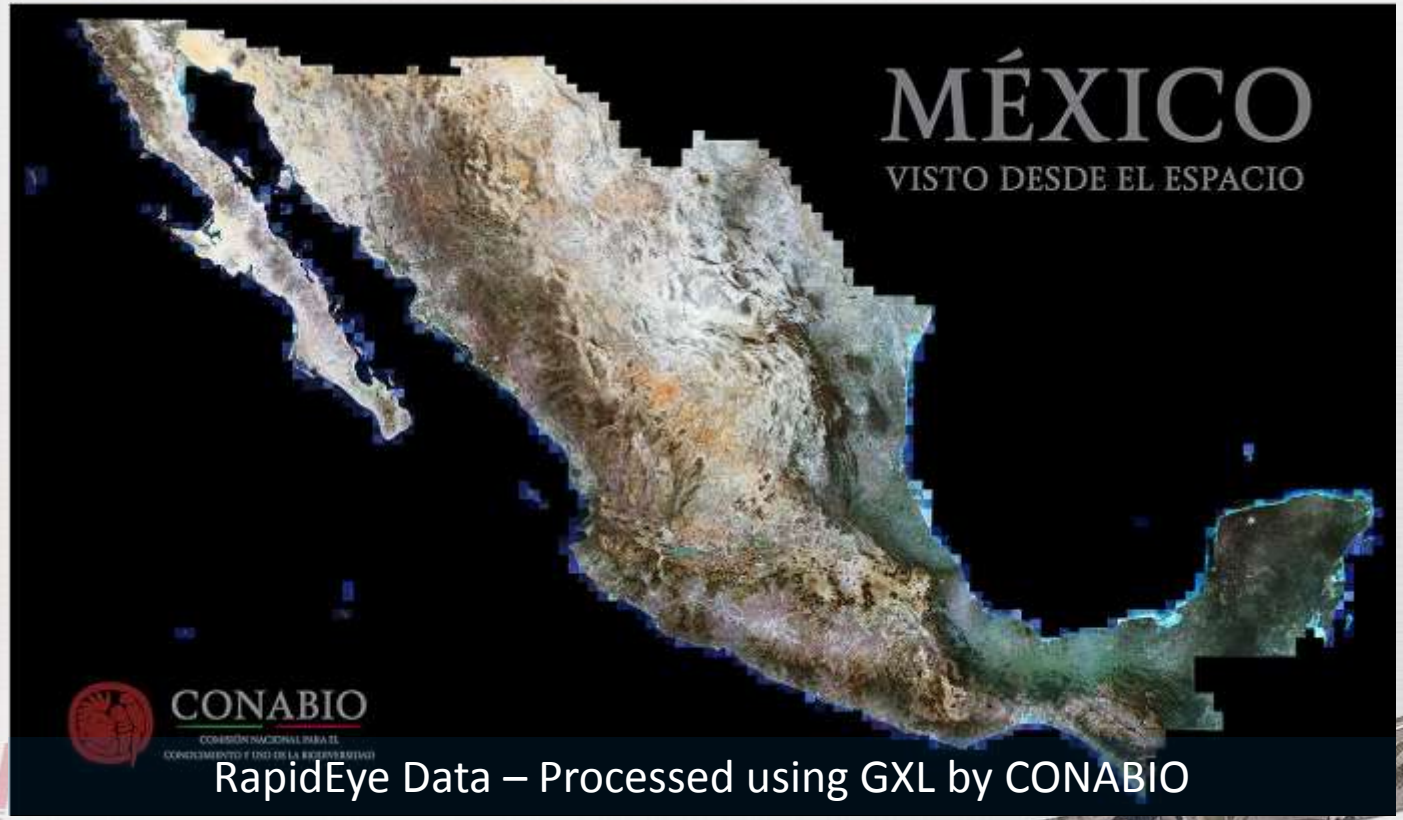
PCI GEOMATICS

GXL

Success Stories

PCI GEOMATICS

Conabio's Mosaic of Mexico



RapidEye Data – Processed using GXL by CONABIO

Conabio's Mosaic of Mexico

Component	Specification
Sensor	Rapideye
Input Images	25,000
Images Used (Cloud free)	4,338
Image Resolution	5m
Operator	1
Coverage Area	1,972,550km^2
Disk Size	>1TB

Esri’s Ortho-Mosaic of the World

Esri's Ortho-Mosaic of the World

Bundle Adjustment → Orthorectification → Automatic Mosaicking → Mosaic Editing → Mosaic Polishing

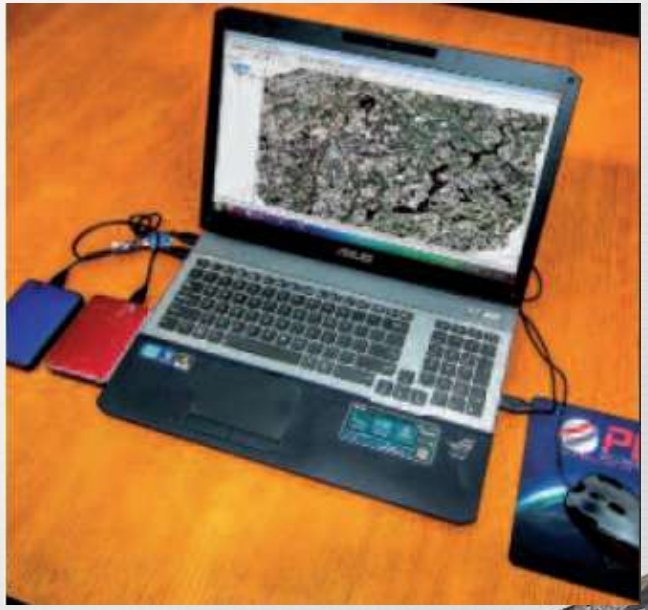
Component	Specification
Sensor	Ikonos
Input Images	>100,000
Images Used (Cloud free)	>100,000
Image Resolution	1m
Operator	3-4
Coverage Area	>50,000,000km^2
Disk Size	N/A

PCI's Challenge

Process over 4000 UltraCamX images in one weekend on a laptop with the GXL

Processing Computer Details

Components	Specification
Laptop	ASUS G75V
CPU	Intel 3610QM (4 cores)
GPU	Nvidia 670M
RAM	24GB
Internal Disk	256GB SSD + 480GB SSD
External USB 3.0 Disks	2x 2TB
Software	GXL 2014



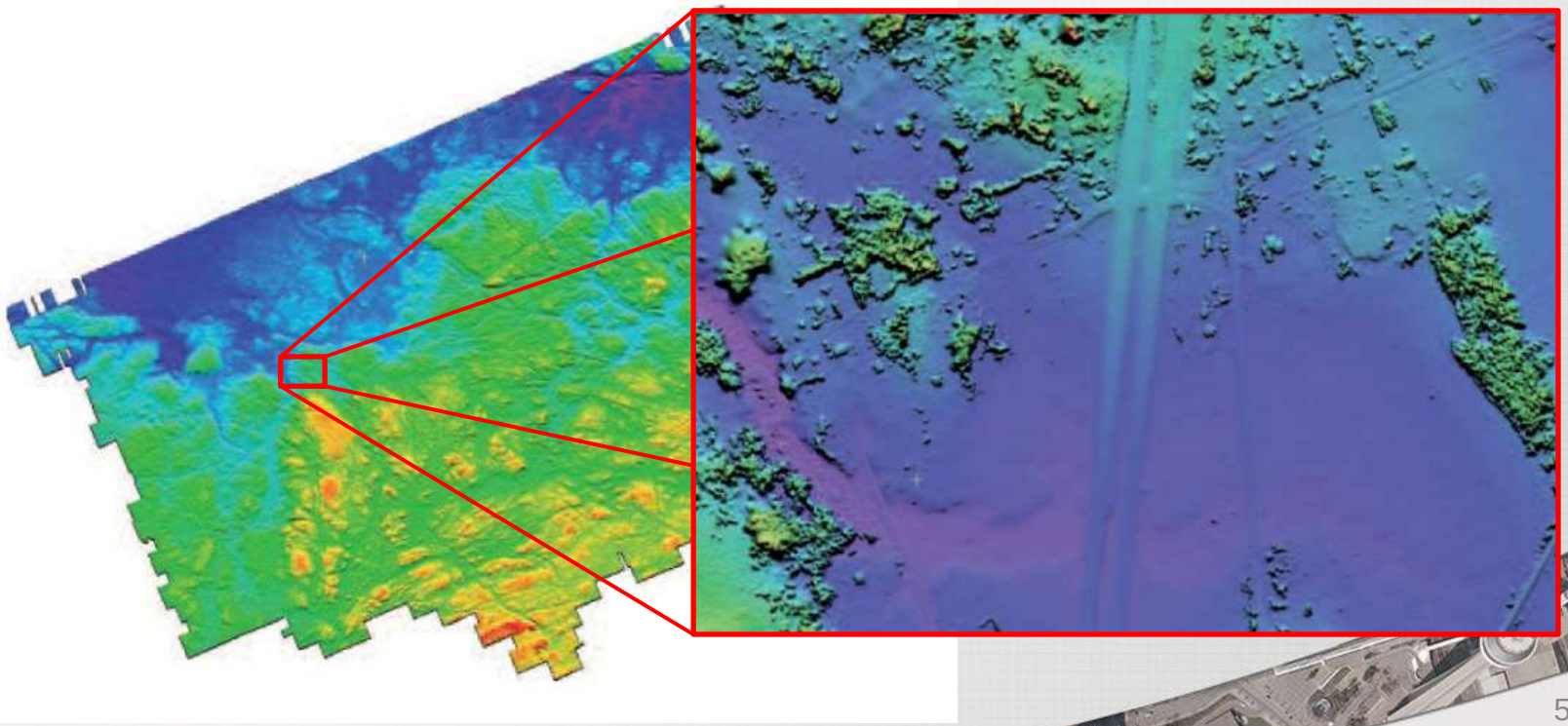
PCI's Challenge

Component	Specification
Camera	UltraCamX Large Format
Image Size	14430 x 9420 pixels
Image Characteristics	8 bit, 3 band TIF Format
Ground Sample Distance (GSD)	20cm
Overlap	70/30

PCI Challenge

Processing Step	Output Disk Size	Processing Time
Data Ingest	<20GB	14m 12s
DSM Extraction	63GB	29h 51m 45s
DSM to DTM	50GB	3h 20m 58s
Orthorectification	460GB (cropped 30%)	9h 59m 16s
Color Balancing & Cutline Generation	<20GB	1h 53m 24s
Mosaic Tile Generation (172 5km x 5km tiles @ 20cm GSD)	269GB	7h 19m 24s
Totals	**1242GB**	**52h 38m 20s**

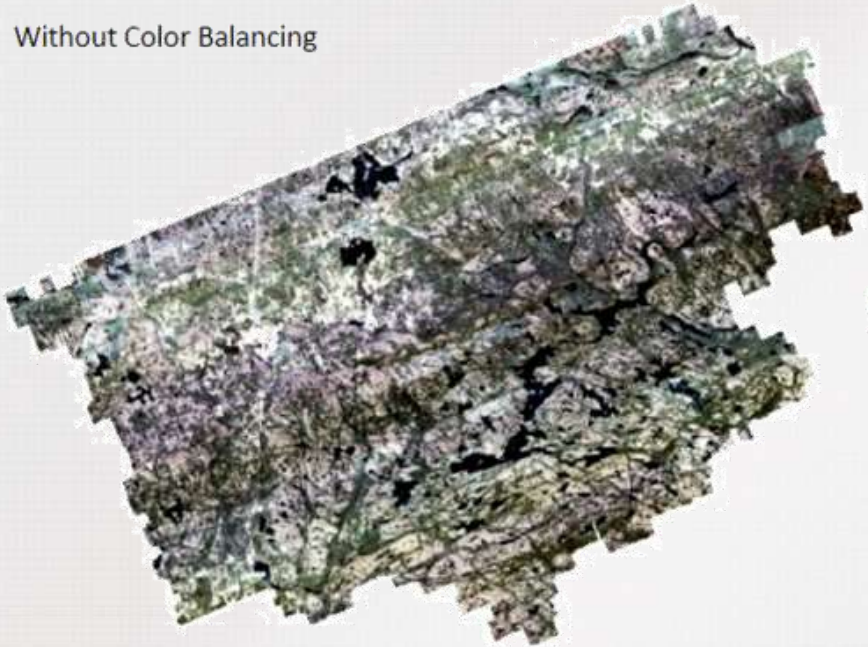
PCI Challenge



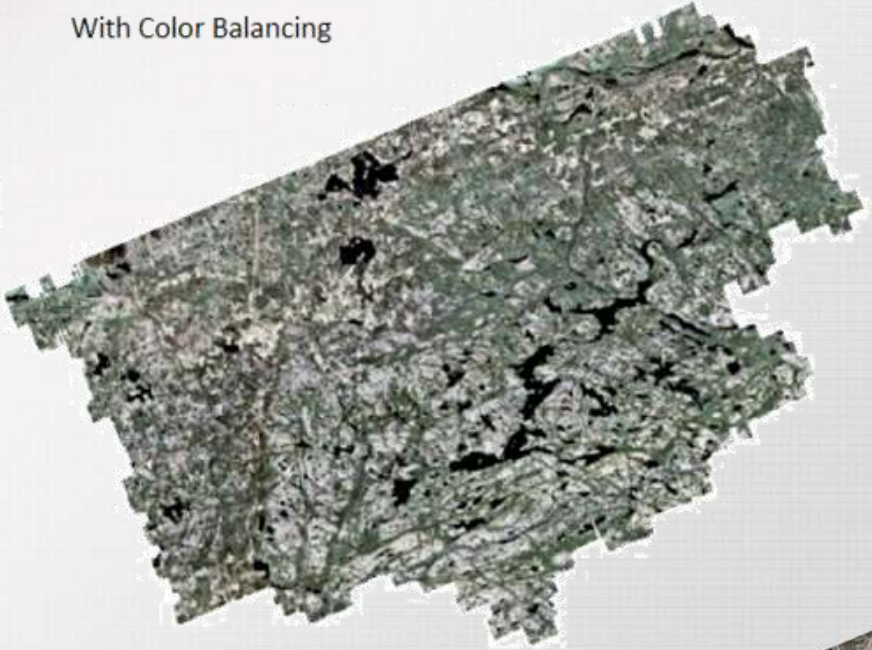
PCI Challenge

PCI Challenge

Without Color Balancing



With Color Balancing



Contact PCI Geomatics

www.pcigeomatics.com
info@pcigeomatics.com