"Geoprocessing in the Cloud"

Brian Levy

Open Solutions Group, Inc

Defense Intelligence Agency

Who We Are

- Open Solutions Group, Inc.
- Small Business, founded in March 2008
- Locations in Northern Virginia & Southern California
- IT Solution Provider with extensive background and expertise in Enterprise Geospatial Information Systems

OPENsgi
INNOVATION. COLLABORATION. PERFORMANCE.

We embrace Open Source to…
…break low productivity habits

"Not Invented Here" Syndrome

"Not Sold Here" Syndrome

Support The Intel Community

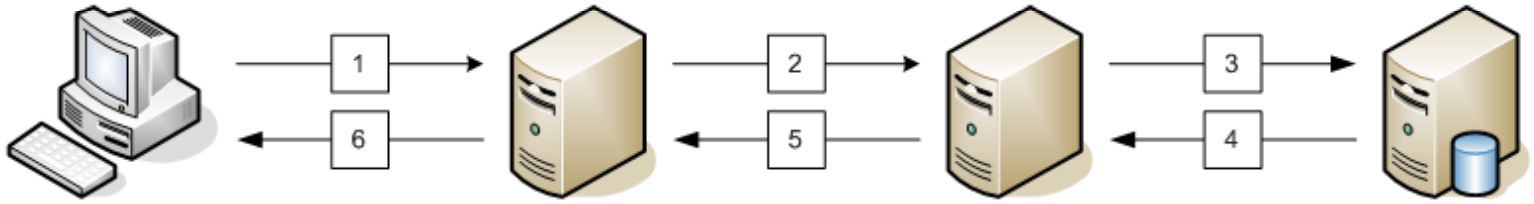
- Defense Intelligence Agency
  - DIA JWS-3 Battlespace Visualization
  - National Ground Intelligence Center
  - Anti Armor Analysis Program
  - Warfighter Imagery Server
  - JSPACE, DODIIS Portal, OPTIC

“Cache the World” Challenge

- On-going initiative with Defense Intelligence Agency to improve reliability and performance for optimal data visualization on the web…

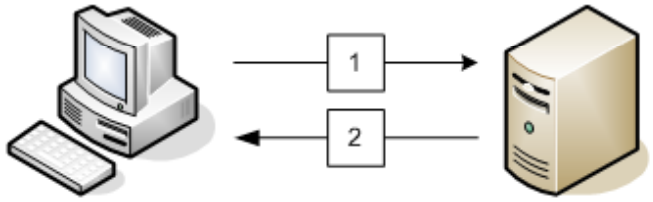
  “My Maps Must Refresh in Less than 1 Second!”

- Fighting traditional map generating bottlenecks
- Refresh rate is typically between 4 – 12 seconds
- Slow Map Services are Not Very Interesting!

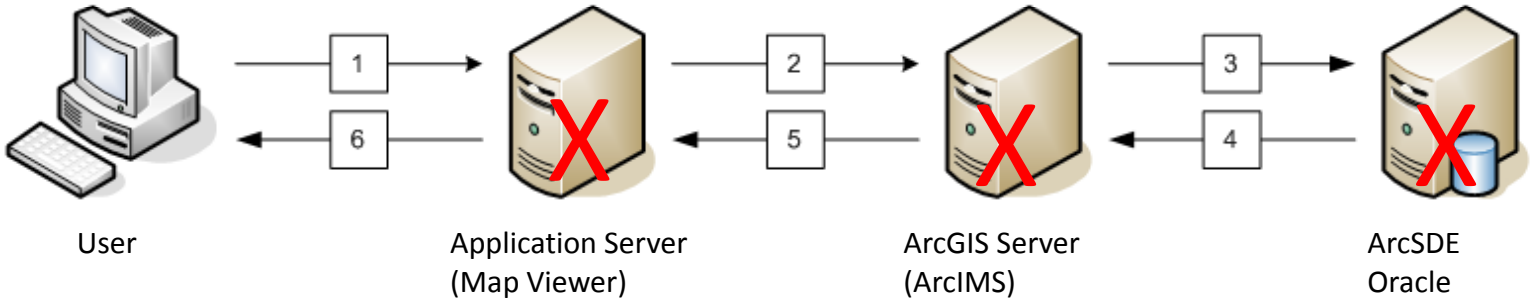


DIA Enterprise GeoCache

- Provide High Performance Access to NGA Products (Global)
  - CIB, DTED, SRTM, VMAP, NVUE, CADRG
- "Cache The World!"



- Attempted Cache Generation Against Existing GIS Infrastructure = FAILURE
- Middleware bottlenecks persisted due to constant image requests for multiple image tiles



Scanned Map Products

Map Product Name	Zoom Levels	Scales
GNC	0-6	1:221M – 1:3.5M
JNC	6-8	1:3.5M – 1:865K
ONC	7-10	1:1.7M – 1:216K
TPC	8-11	1:865K – 1:108K
JOGA	9-11	1:432K – 1:108K
TLM100	10-12	1:216K – 1:54K
City Graphics	11-16	1:108K – 1:3.5K

Imagery Products

Natural View & Controlled Image Base (CIB)

Imagery Product Name	Zoom Levels	Scales
NaturalVue	4-11	1:13.8M – 1:108K
CIB1 Meter Imagery	6-16	1:3.4M – 1:3.5K
CIB5 Meter Imagery	9-13	1:432K – 1:27K
CIB10 Meter Imagery	6-9	1:3.4M – 1:432K

Vector Products

- Collection of data from VMap0, VMap1 and Shapefiles
- Features from different layers are visible at various scales
- The new standard for Basemap

got cloud ?

Our Geoprocessing Toolbox

Geospatial Data Abstraction Library
"goo-doll"

Raster/Vector Translation Library

mrGeo

Map Reduce Geospatial

MapNik
TileCache
MapServer

Image Processing Library

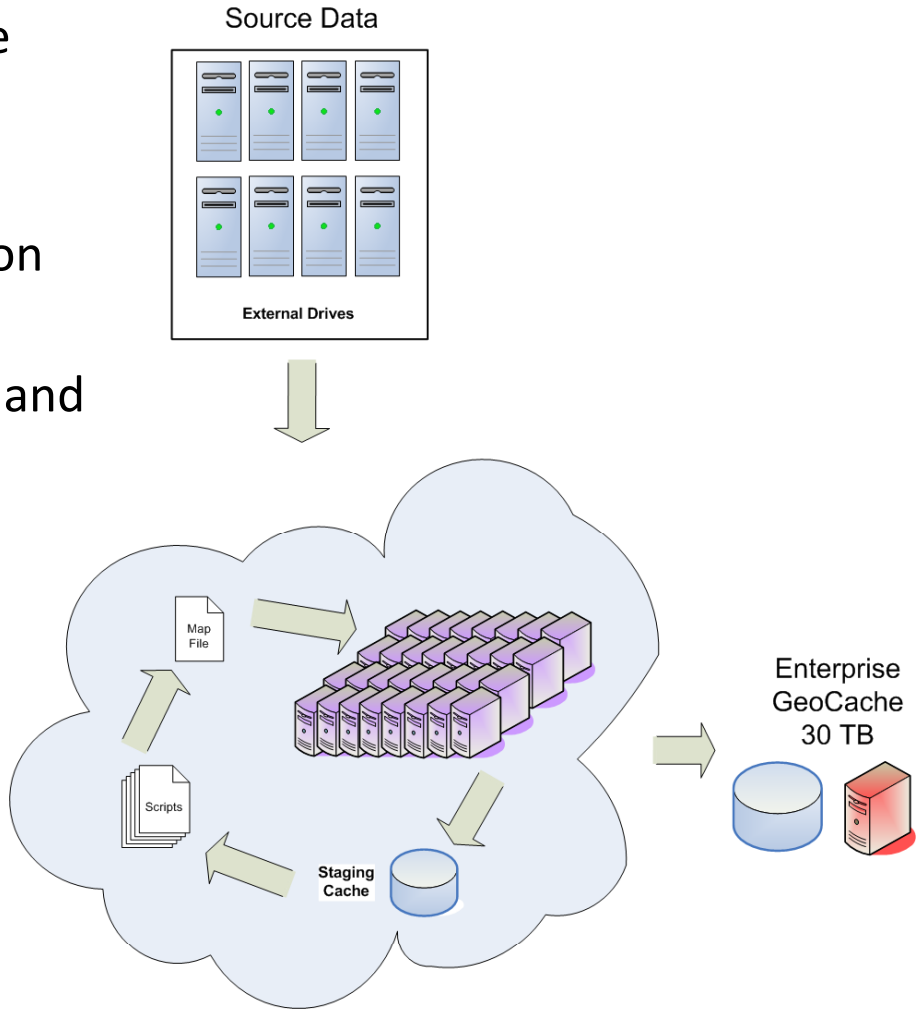
Format Conversion
Resize, Rotate, Crop
Composites
Transparency

Our Geoprocessing Warehouse

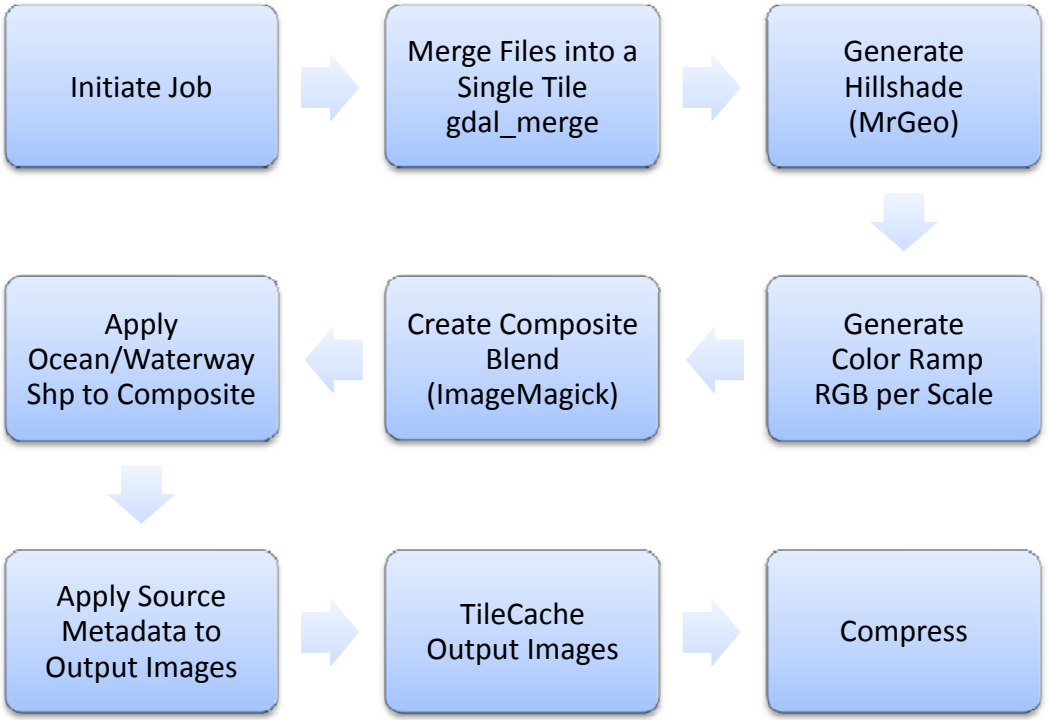
- Developed Cloud Geoprocessing Infrastructure for Server Management, Job Queue Management, Workers, and Error Reporting
- Provide Status, Tracking, and Metrics for Process Management.
- Manage 100s of thousands of geoprocessing jobs
- All available NGA Products Managed and Processed

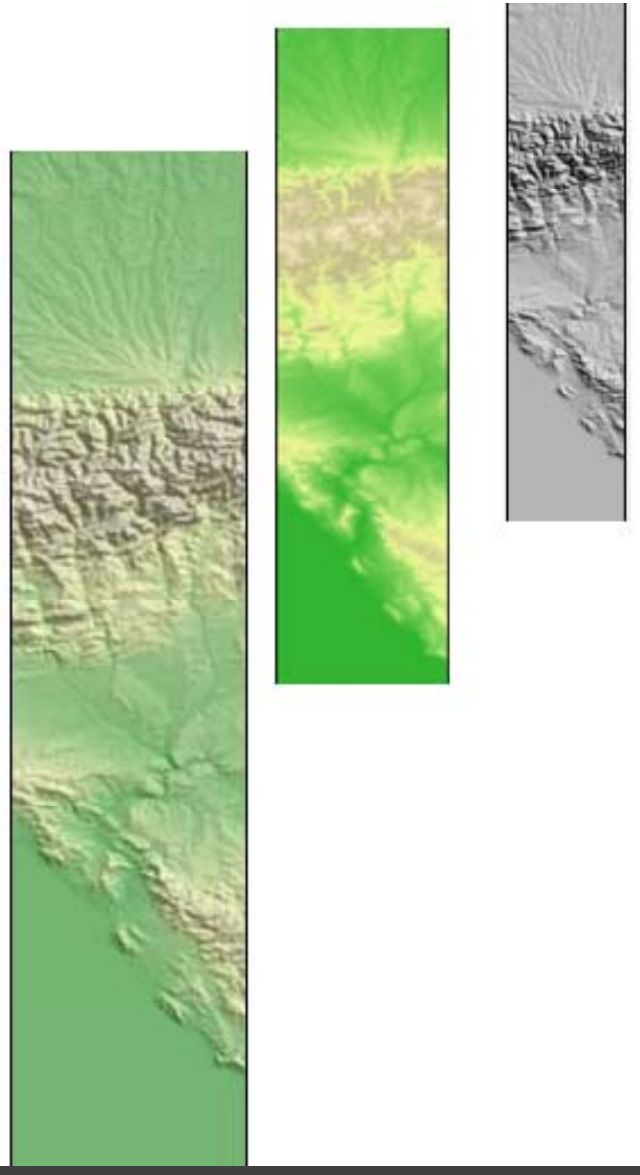
The Assembly Line

1. Data Arrives At Warehouse
2. Workers Begin Shift (Server Provisioning)
3. Jobs Assigned to Workers on Assembly Line (Queue)
4. Foreman oversees Quality and Errors (QA/QC)

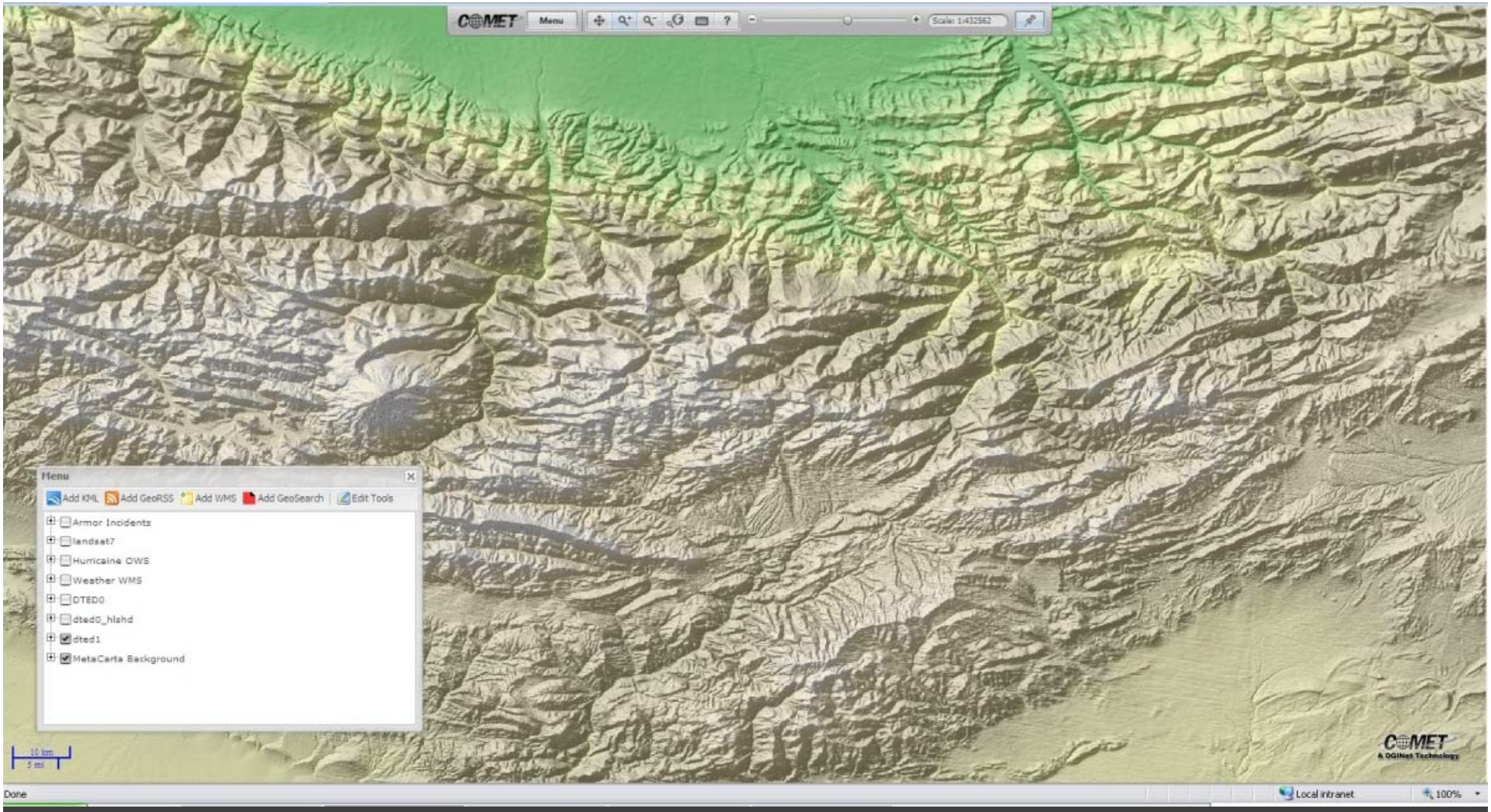


Geoprocessing Workflow (DTED)





Geoprocessing Output (DTED)



*Hillshade and Color Ramp Composite

Interoperability

- Leverage Geospatial Standards (OGC)
- Tighter Integration with MapReduce (Hadoop) to achieve parallel geoprocessing over vector datasets (vmap, osm)
- Provide multiple Output Formats to achieve interoperability and reuse
  - Completely Open Mapping Environment (COMET)
  - ArcGIS Desktop
  - Google Earth
  - WMS Clients, Tiling Clients (TMS, TileCache), Mashups
  - Adobe Flex, Microsoft Silverlight

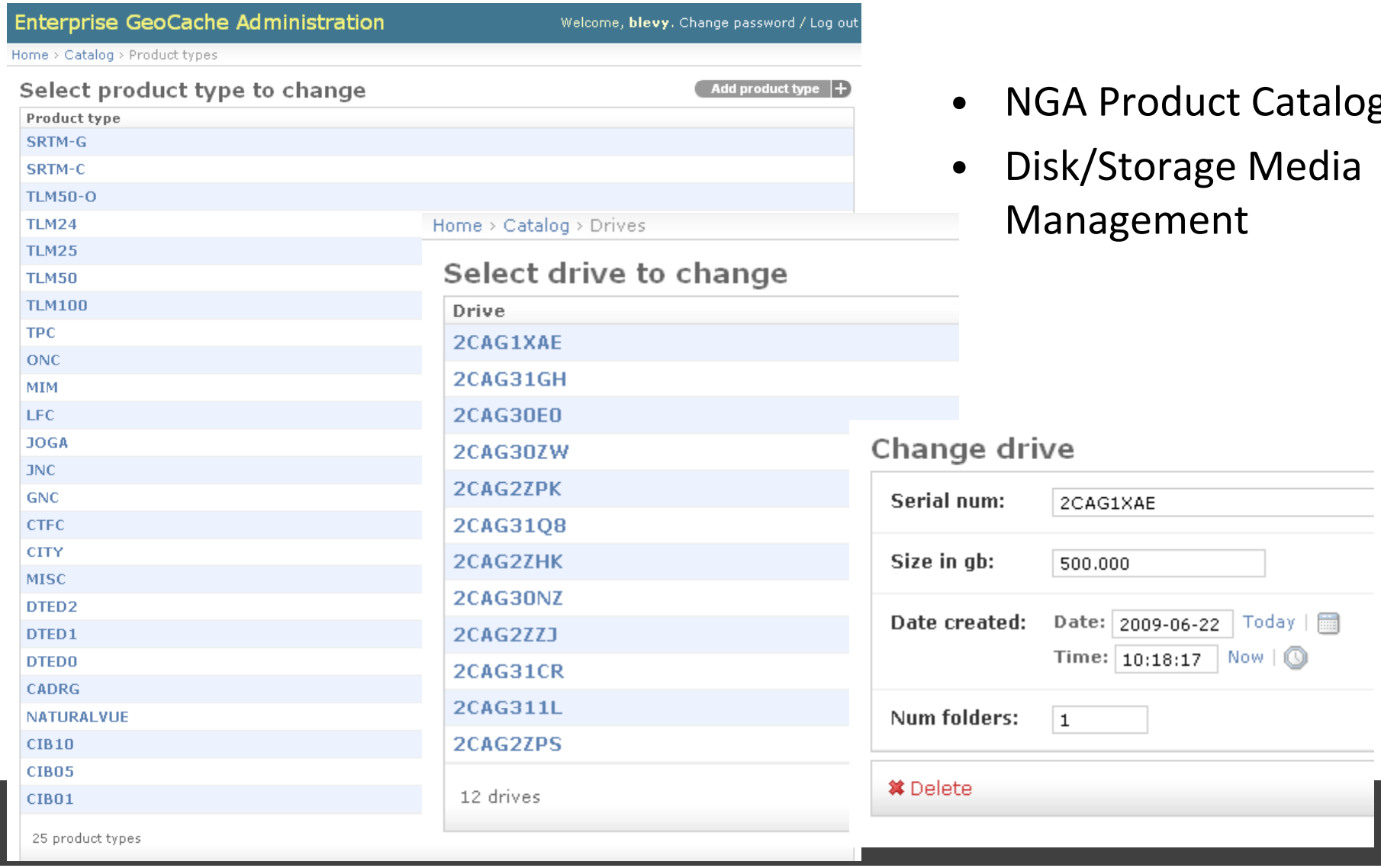
Maintain Cloud Control (QA / QC)

- Thousands of processes can (and will) generate thousands of errors during operation

  1,000,000 Geoprocessing Jobs
  0.5 % Margin
  = 5,000 Processing Errors

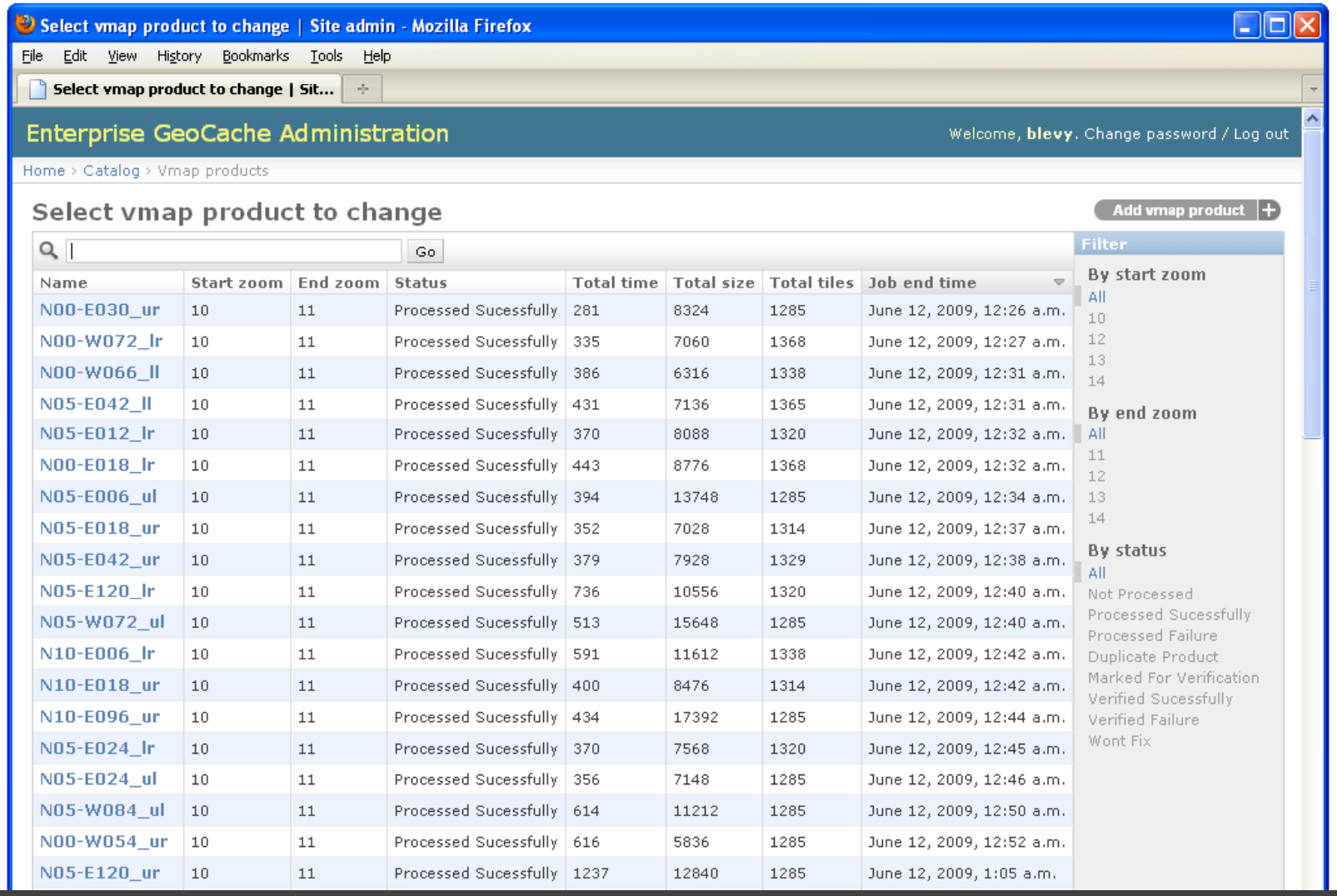
- System and resource failures will occur during high volume operations
- Each geoprocessing job must hold accountability
- High volume output must be verified by algorithms and humans

Global Catalog of Products



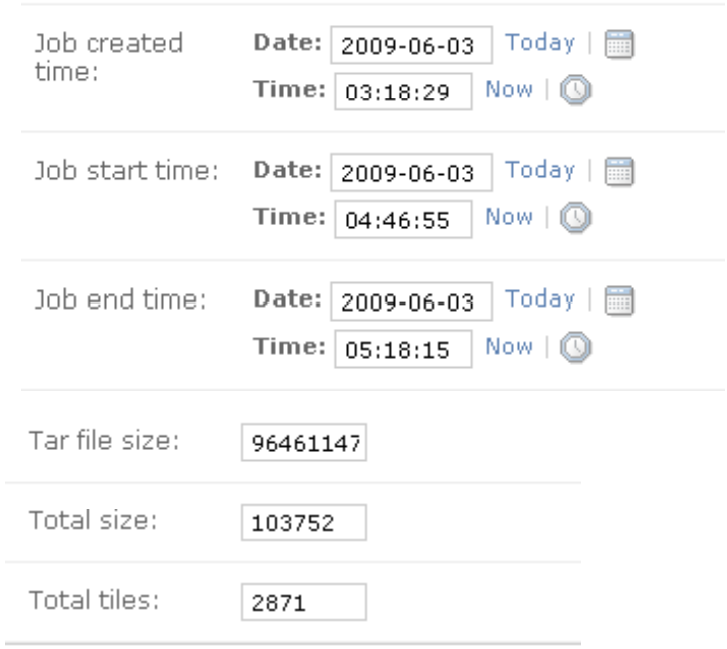
- NGA Product Catalog
- Disk/Storage Media Management

Geoprocessing Dashboard



Name	Start zoom	End zoom	Status	Total time	Total size	Total tiles	Job end time
N00-E030_ur	10	11	Processed Sucessfully	281	8324	1285	June 12, 2009, 12:26 a.m.
N00-W072_lr	10	11	Processed Sucessfully	335	7060	1368	June 12, 2009, 12:27 a.m.
N00-W066_ll	10	11	Processed Sucessfully	386	6316	1338	June 12, 2009, 12:31 a.m.
N05-E042_ll	10	11	Processed Sucessfully	431	7136	1365	June 12, 2009, 12:31 a.m.
N05-E012_lr	10	11	Processed Sucessfully	370	8088	1320	June 12, 2009, 12:32 a.m.
N00-E018_lr	10	11	Processed Sucessfully	443	8776	1368	June 12, 2009, 12:32 a.m.
N05-E006_ul	10	11	Processed Sucessfully	394	13748	1285	June 12, 2009, 12:34 a.m.
N05-E018_ur	10	11	Processed Sucessfully	352	7028	1314	June 12, 2009, 12:37 a.m.
N05-E042_ur	10	11	Processed Sucessfully	379	7928	1329	June 12, 2009, 12:38 a.m.
N05-E120_lr	10	11	Processed Sucessfully	736	10556	1320	June 12, 2009, 12:40 a.m.
N05-W072_ul	10	11	Processed Sucessfully	513	15648	1285	June 12, 2009, 12:40 a.m.
N10-E006_lr	10	11	Processed Sucessfully	591	11612	1338	June 12, 2009, 12:42 a.m.
N10-E018_ur	10	11	Processed Sucessfully	400	8476	1314	June 12, 2009, 12:42 a.m.
N10-E096_ur	10	11	Processed Sucessfully	434	17392	1285	June 12, 2009, 12:44 a.m.
N05-E024_lr	10	11	Processed Sucessfully	370	7568	1320	June 12, 2009, 12:45 a.m.
N05-E024_ul	10	11	Processed Sucessfully	356	7148	1285	June 12, 2009, 12:46 a.m.
N05-W084_ul	10	11	Processed Sucessfully	614	11212	1285	June 12, 2009, 12:50 a.m.
N00-W054_ur	10	11	Processed Sucessfully	616	5836	1285	June 12, 2009, 12:52 a.m.
N05-E120_ur	10	11	Processed Sucessfully	1237	12840	1285	June 12, 2009, 1:05 a.m.

Geoprocessing Status



- When did the Job Start?
- When did the Job Finish?
- How large is the output?
- How many tiles were generated from this single product?

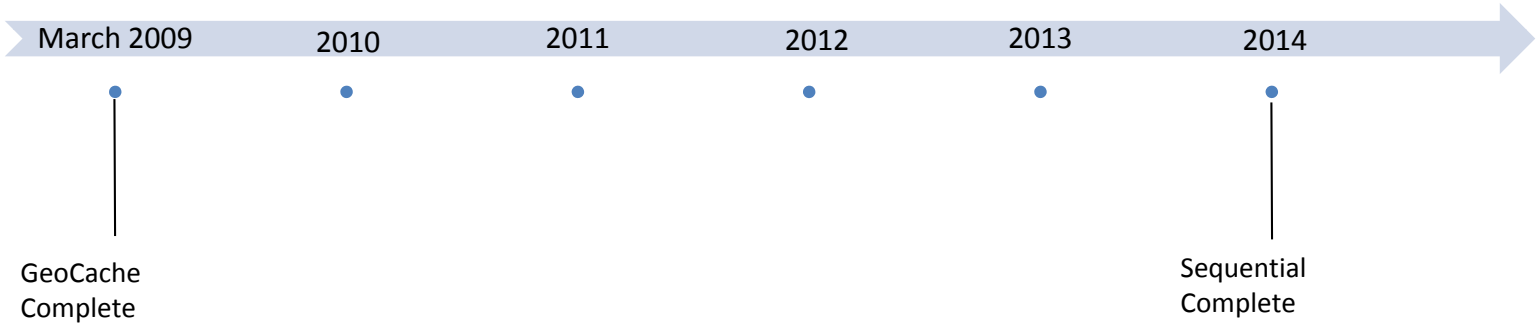
The Hard Numbers (CIB1)

- INPUT
  - 6963 Original Products @ 2.3 Terabytes
- PROCESSING
  - 9000 Processor Hours (over 1 human year)
  - Completed in 4.5 Hours
  - 2,000 Servers
- OUTPUT
  - 14 Terabytes of Processed CIB1 Imagery
  - 338 Million Image Tiles (256*256 pixels)

The Really Hard Numbers

Computing Time of 40,000 hours

- A Sequential Process Would Take 5 Years to Complete!
- We sliced processing duration to 24 Hours



- 3.5 TB of NGA Products Processed
- 30 TB of Cached Products Generated
- 1 Billion Image Tiles (256*256 pixels)

Geoprocessing On Demand

OSGI & SPADAC together are enabling "Geoprocessing On-Demand" with shared models and derivatives

- Leverage Community TradeCraft
- mrGeo Library for supporting parallel image processing (Map Reduce)
- Input Digital Elevation Models to produce GeoTIFF outputs
- Plug-n-Play with Geoprocessing Infrastructure

SRTM2 (30m) Value Added Layers

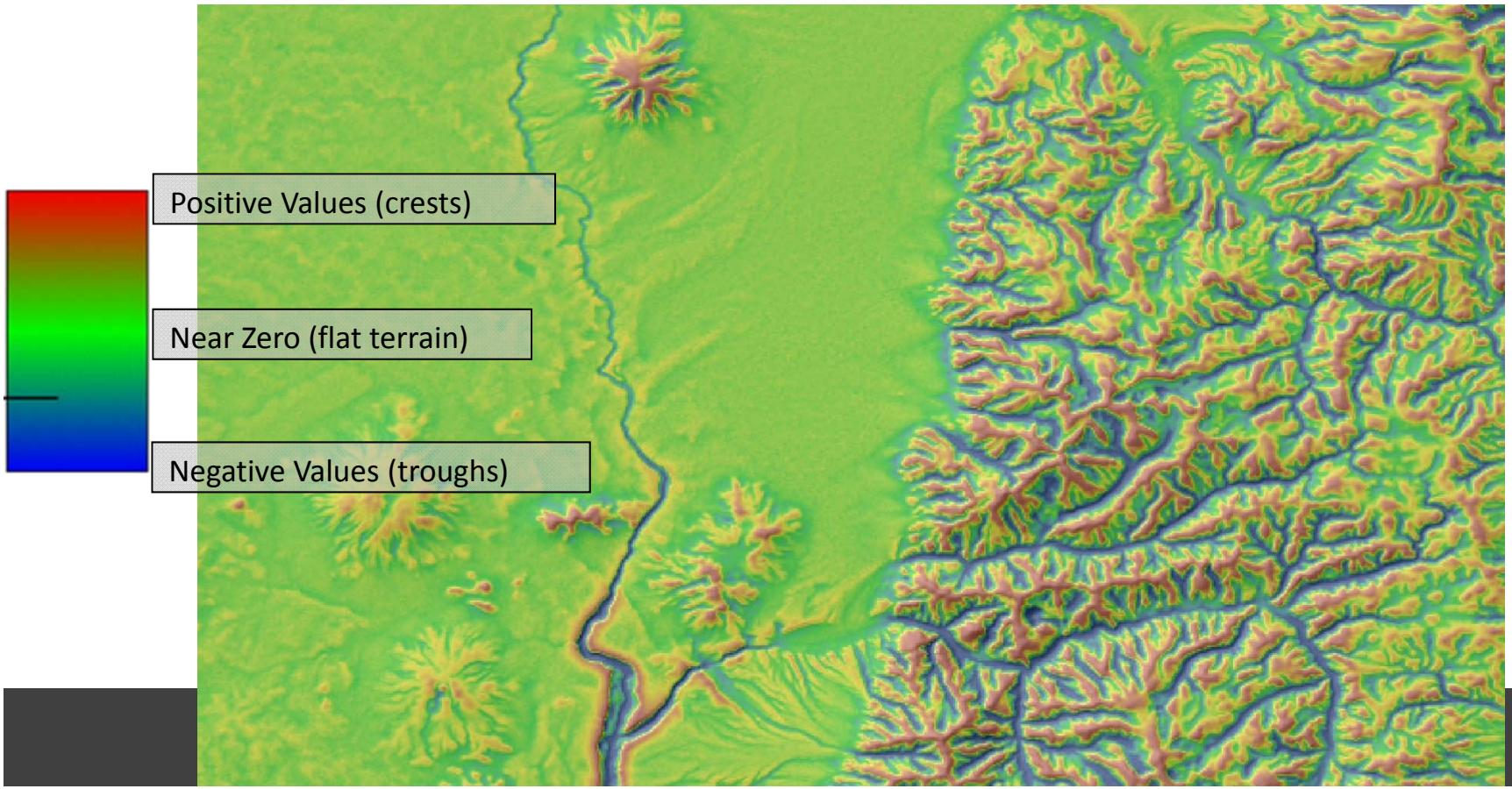
Derived Surfaces will Include:

1. Hillshade
2. Slope
3. Aspect
4. Terrain Ruggedness Index (TRI):
   - A measure of terrain variability—crucial to understanding peoples interaction with terrain.

- Relevation (Relative Elevation):
  - A cells elevation relative to its surrounding environment.

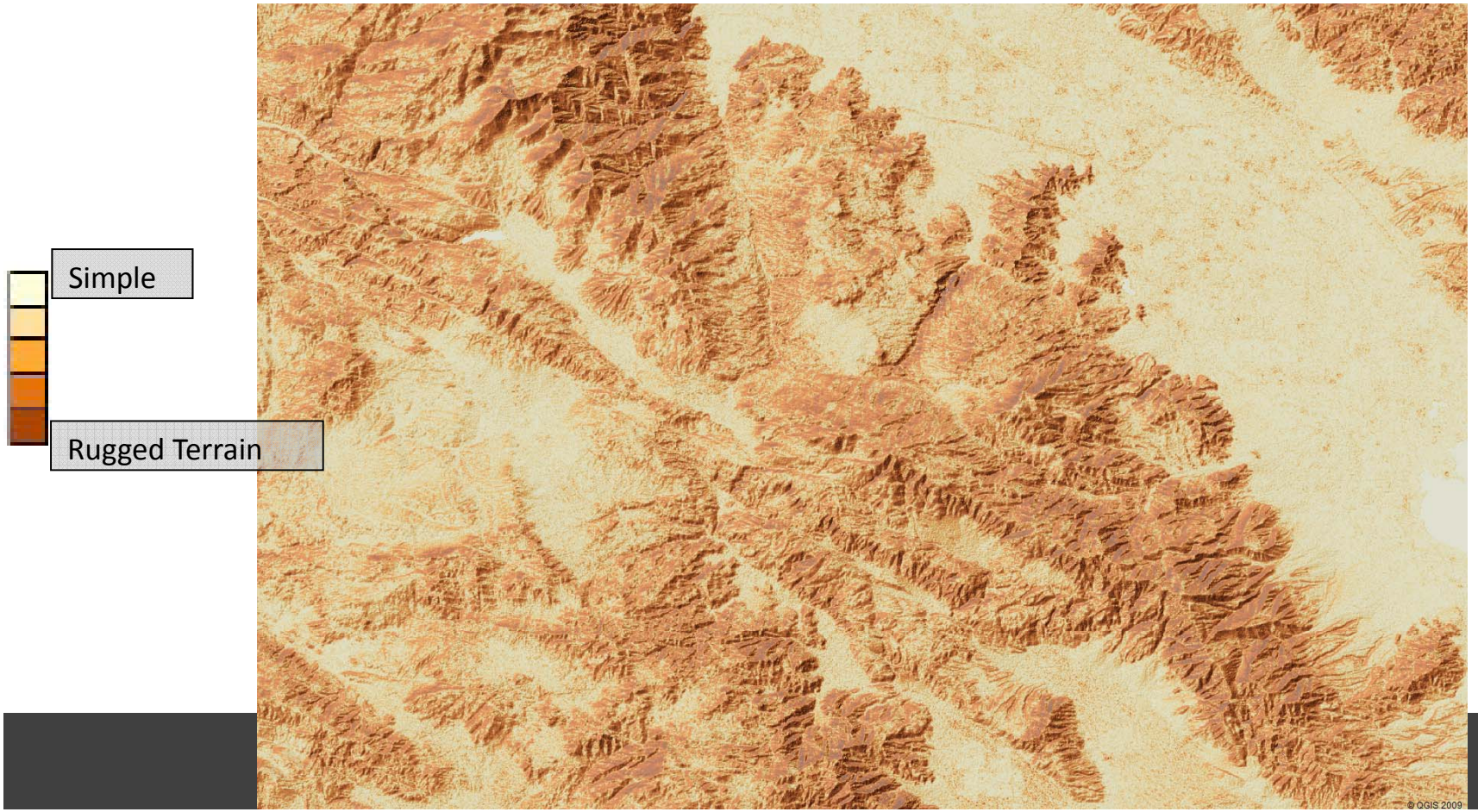
Relevation (Relative Elevation) Index Calculation

- Modification of Topographic Position Index (TPI), introduced by Weiss (2001).
- Difference in elevation between a cell and the average of its neighbors.
- Neighborhood measurements account for map distortions across all latitudes.



Terrain Rouggedness Index

- Adapted from Riley et al. (1999) to measure the variability of a landscape.
- A key parameter in distinguishing suitable/unsuitable terrain for people and infrastructure.



Geoprocessing On Demand

- Benchmarking for Iteration I
  - Input 13,000 SRTM products
  - Requires 20,000 processing hrs (Over 2 years)
  - ETA 0.5 days for completion
  - Generate 2.5 TB Output (GeoTIFF)
- Cost Estimates
  - Cheaper than a Dell PowerEdge
  - Costs less than an ArcInfo Single Use License
  - Equivalent to 5 days work of an analyst

Large Data Lessons Learned

- GeoCache Deployment
  - Moving Tremendous Amounts of Data (30TB)

    Compress, Copy, Copy, Copy Again, Uncompress
  - Groundbreaking = Limited Support
  - QA Is Difficult (the Earth?)

Future Direction

- GeoProcessing On-Demand for unique user/analyst requirements
- Identify and Leverage New Cloud Environments for Real Time Provisioning
  - JWICS
  - SIPRNET*
- Mobile Units & Portable “Local” Export
- Enterprise GeoCache Appliance
- New approaches to improve process

Questions ?

Brian Levy

President & CEO

email: brian.levy@opensgi.com

office: 858-312-5390

CORPORATE HEADQUARTERS

1394 OLD GALLOWS ROAD, SUITE 350

VIENNA, VA 22182

540-288-8466

11440 West Bernardo Court, Suite 300

San Diego, CA 92127

858-312-5390

www.OpenSGI.com