HPC Advances in LIDAR E2E Processing

Anthony Galassi
NGA/InnoVision

September 21, 2011
2011 HPC Challenge

Anthony.K.Galassi@nga.ic.gov

Agenda

- HPC Processing
- SOA in the Cloud
- Hybrid Cloud Approach
- Deployed Concept

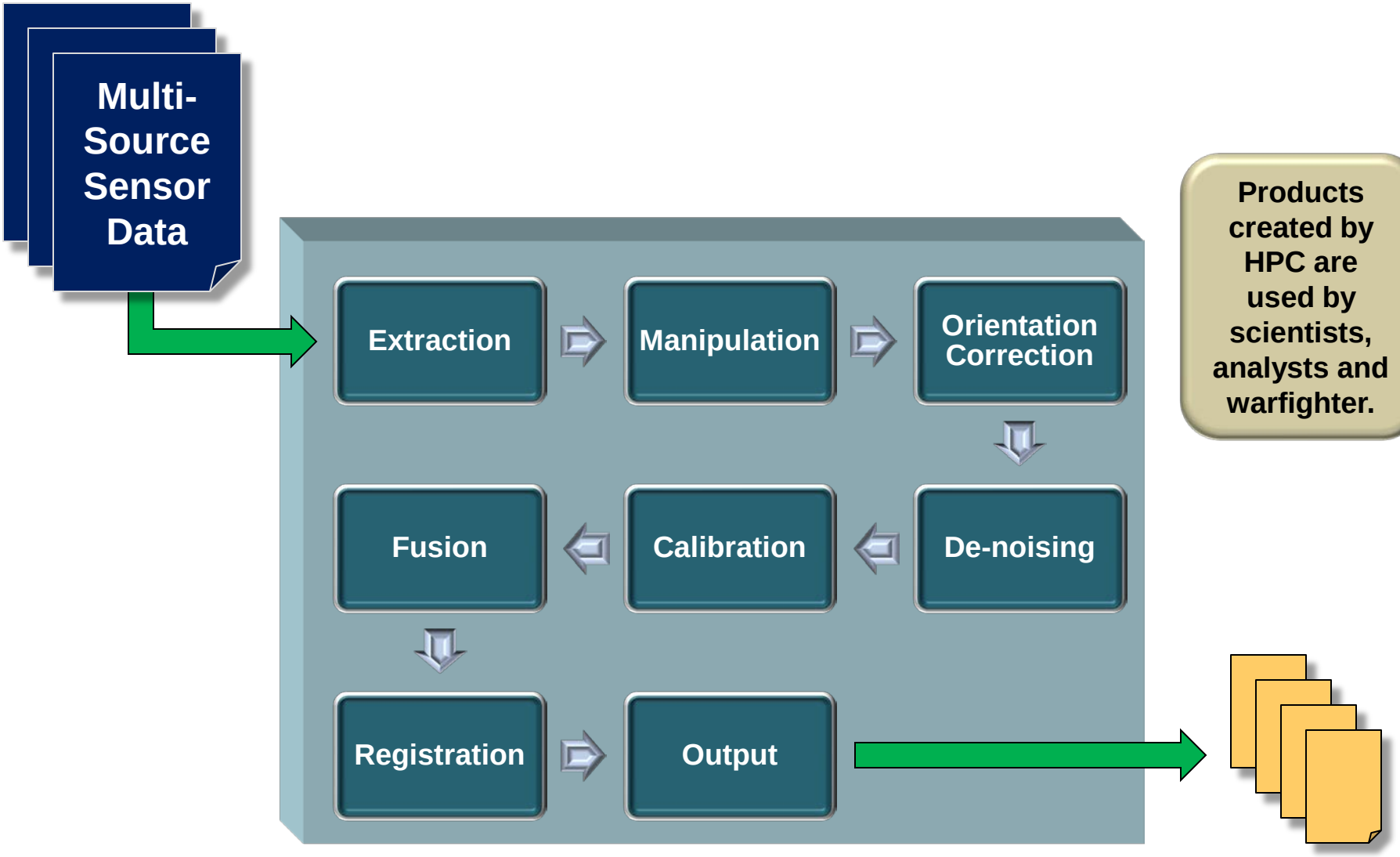
National Geospatial-Intelligence Agency (NGA)

NGA Mission:

To provide timely, relevant, and accurate GEOINT in support of national security

**** NGA is the lead federal agency responsible for Geospatial Intelligence (GEOINT)***

HPC Processing Implementation



LIDAR CMMD Implementation

Metadata and Wideband data are carried together throughout the LIDAR Processing Chain

Metadata
Sensor Data
HDF5/XML/KLV file

ALIRT
HALOE
Bathymetric
Other LIDAR Data

Extraction → Tiling → Swath Rotation → De-noising → PSF Calibration → Flattening → Registration → Geolocation

LIDAR Processing Chain

LIDAR Data Model

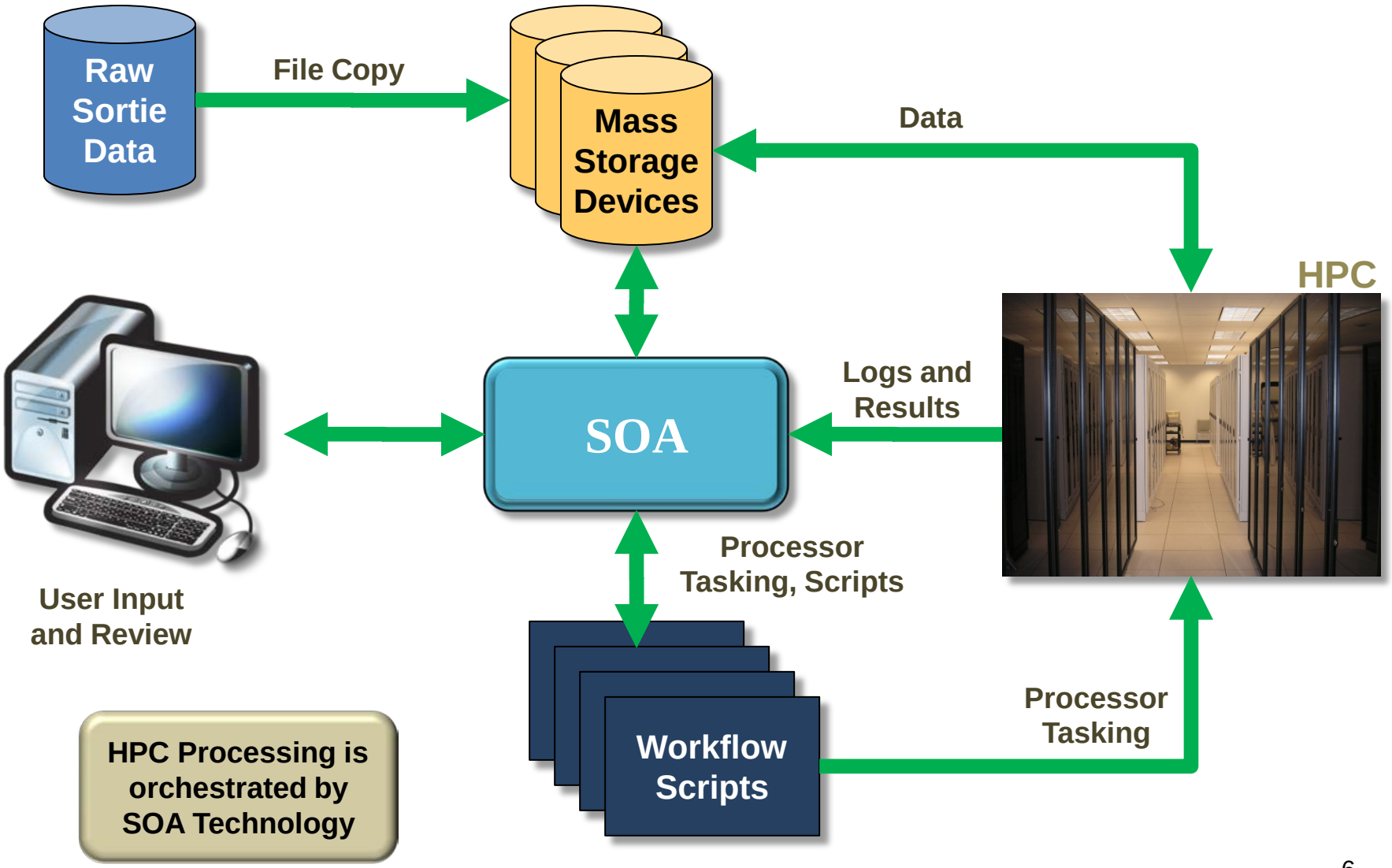
LIDAR Data Dictionary

Necessary metadata (profiles) are determined by analyzing products created by LIDAR scientists, exploitation/tool requirements, and V&V activities

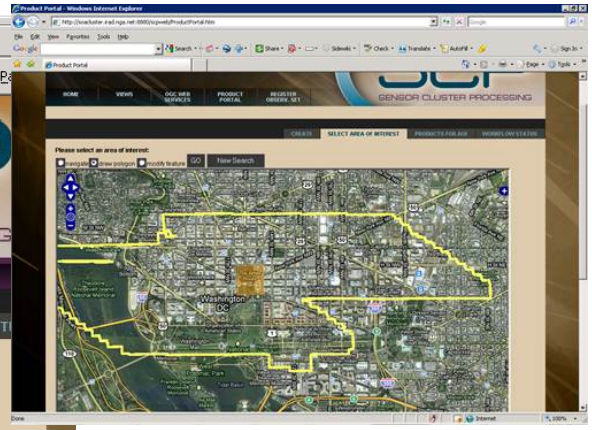
LIDAR Data Model and dictionary encompass the processing chain

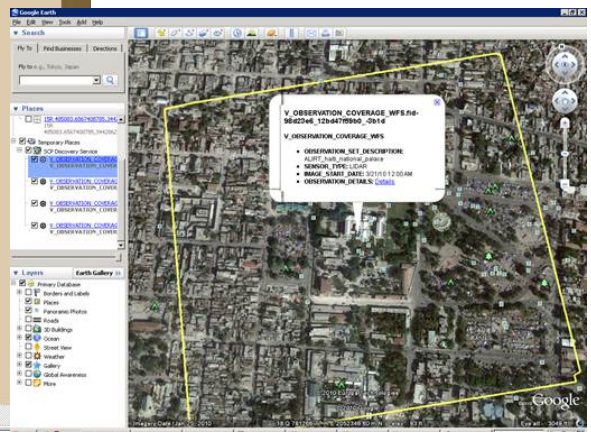
End-to-End Processing

SOA on HPC


Raw Sortie Data
File Copy
Mass Storage Devices
Data
HPC
SOA
Logs and Results
User Input and Review
Processor Tasking, Scripts
Workflow Scripts
Processor Tasking
HPC Processing is orchestrated by SOA Technology

SOA Processing of Large Data Sets





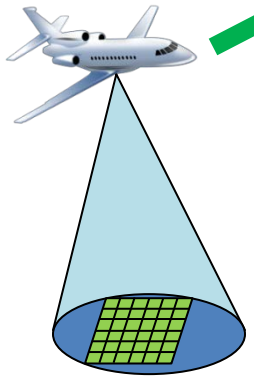
Custom Products are made using the SCP interface

Hybrid Cloud Approach

Heavy Iron & Cloud Components

Classified Time Dominant (In Theater)

Data delivered to processing environments via transportable disk-drives



Processing of Terabytes of data occurs in theater and is available remotely via web services/cloud

Government Networks (Sensitive Data)

Products are made available to the warfighter through web services

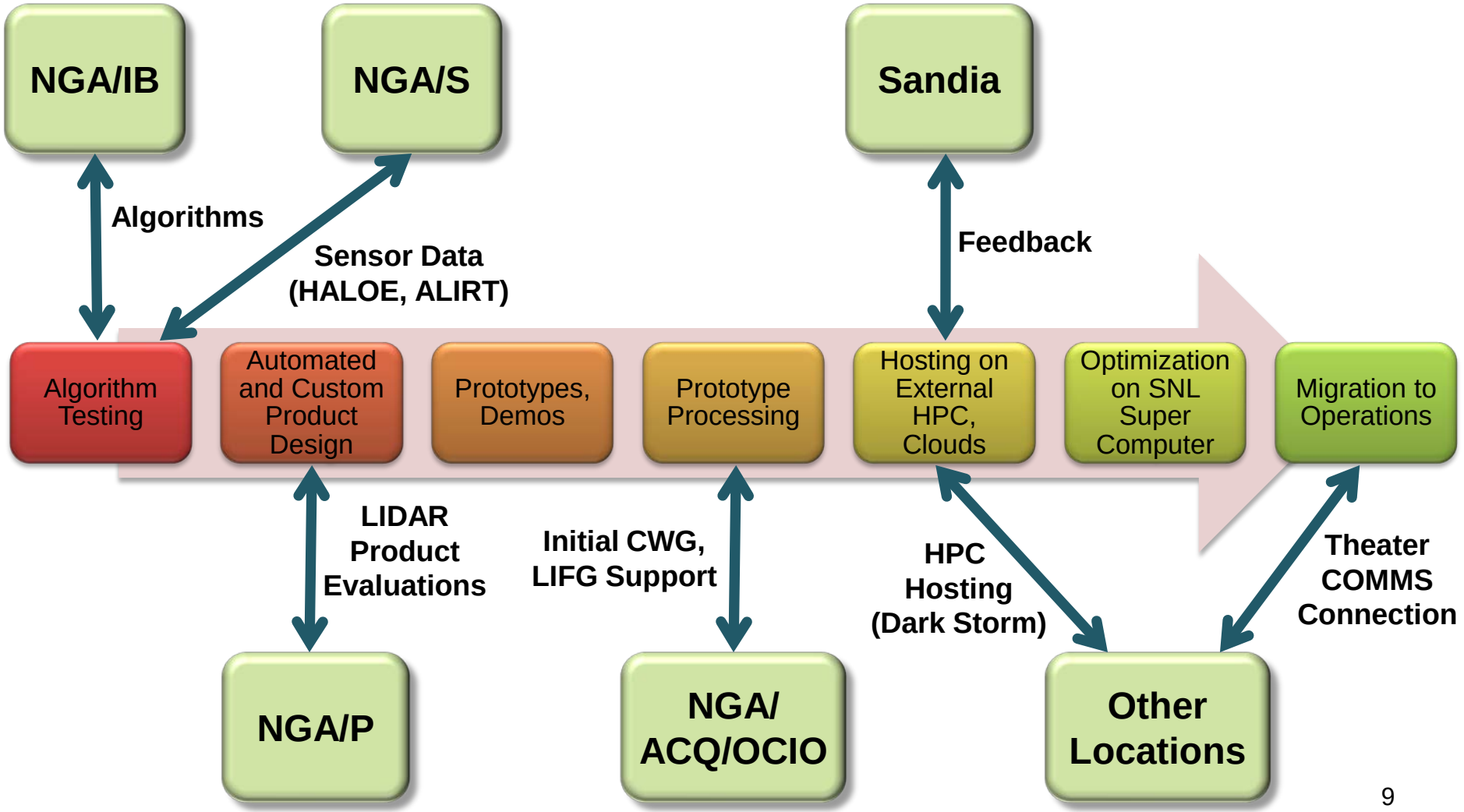
Data is processed and accessed by analysts in CONUS through web services and analyst tools in the "cloud"

Government Networks (Sensitive Data)

Further Processing of Gigabytes of data occurs at CONUS Cloud Computing facility and results are made available remotely via web services/cloud

Classified Long-term (Mapping)

NGA/IID Leads the Migration of HPC Technology



Amazon Cloud Experiment

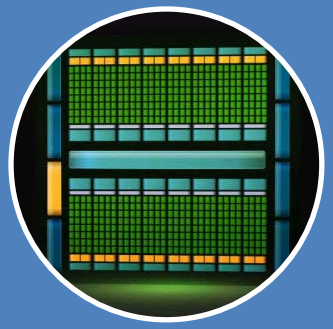
Proved Remote Access

- **Using dedicated hardware solutions for High Performance Computing (HPC) is very expensive and time consuming.**
- **The cloud approach allows computing resources to be shared among numerous programs, allows dormant or unused resource to be allocated to immediate needs, and allows more computing resources to be added as needed.**
- **The raw (L1) data is too large to move and is not readily available to analysts/analyst tools. Provide access for the analyst (and tools) to the L2 data and the product to be analyzed. Don't waste time moving raw data around.**
- **For unclassified research, commercial vendors such as Amazon can be used to build on-demand HPC clusters and analyst workstations. For sensitive data, NGA Cloud environments can be developed/leveraged.**
- **Data from Sortie 33, Partition 25, unclass collection over Wash. D.C.**

Level	Data Type	Typical Size
L0	Raw Data	41GB
L1	3D Point Clouds	460 GB (10-12x)
L2	De-noised 3D Point Clouds	96GB (down 4.5x)
L3	Geo-registered L2 3D Point Clouds	~96GB (1x)
L4	DEMs, viewsheds or other standard products	10-20 MB (down 500x)
L5	Specialized Products	10-20 MB (1x)

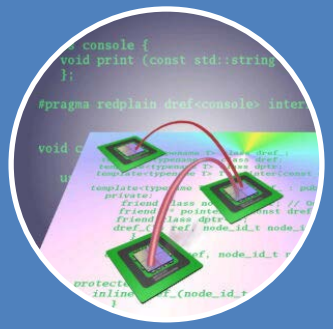
The Next 15 years

HPC-based LIDAR Processing in the National Interest



GPUs

Faster Access Storage



Distributed Processing for extremely large data sets

Cloud Computing

For More Information

Anthony Galassi
C/Advanced Data Interoperability
NGA/IID

- anthony.k.galassi@nga.mil
- 571-557-5373

www.nga.mil