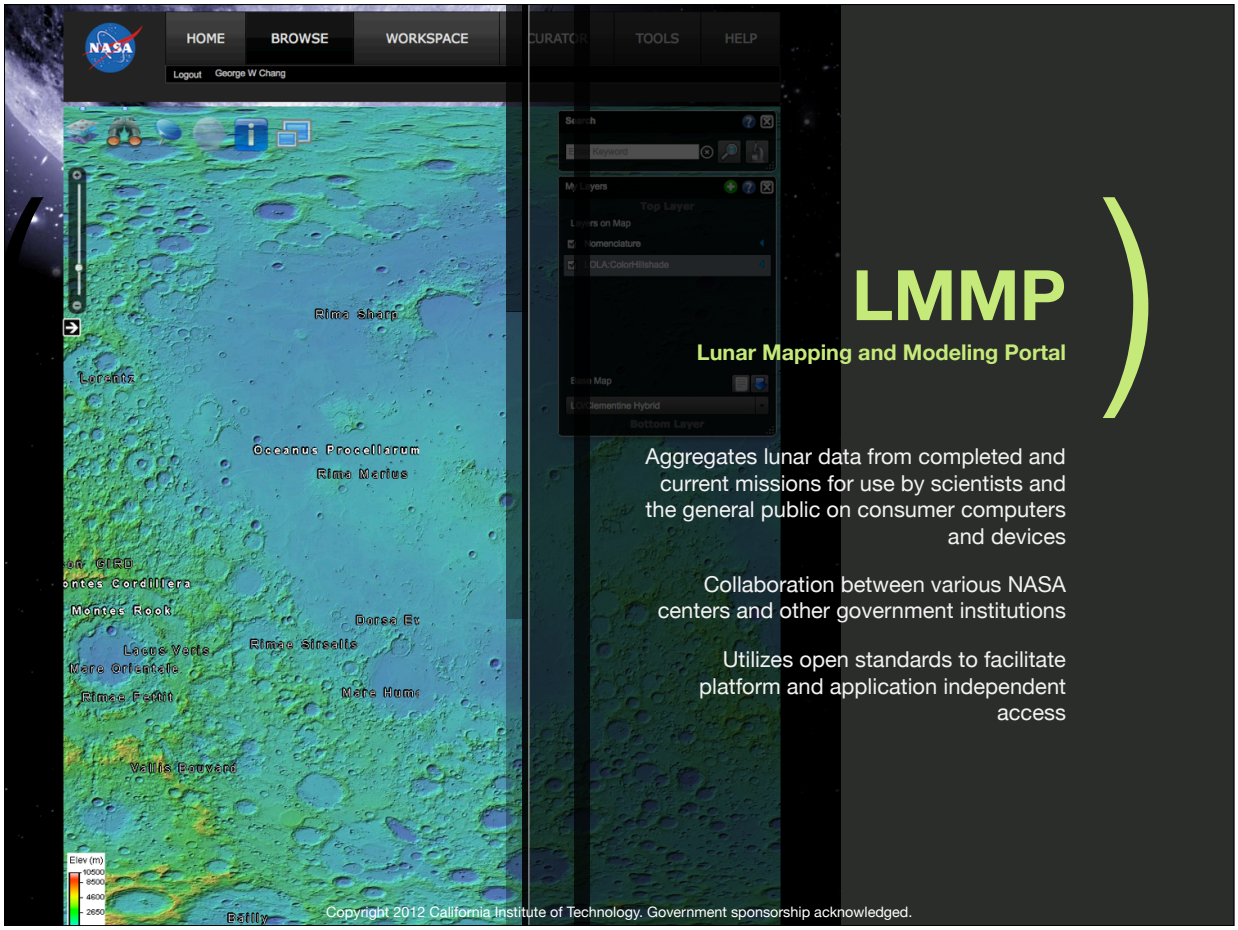
Design and Implementation of a GIS in the Cloud

George Chang
NASA/Jet Propulsion Laboratory

Use Cases

Hybrid Cloud Design for a GIS Portal

Hadoop and Cloud Computing for Image Tiling



LMMP

Lunar Mapping and Modeling Portal

Aggregates lunar data from completed and current missions for use by scientists and the general public on consumer computers and devices

Collaboration between various NASA centers and other government institutions

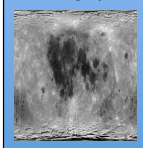
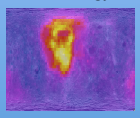
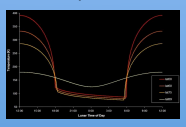
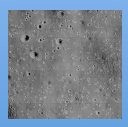
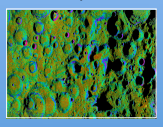
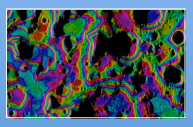
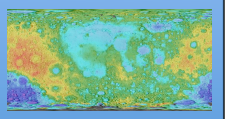
Utilizes open standards to facilitate platform and application independent access

Hybrid Cloud Architecture

Data Products

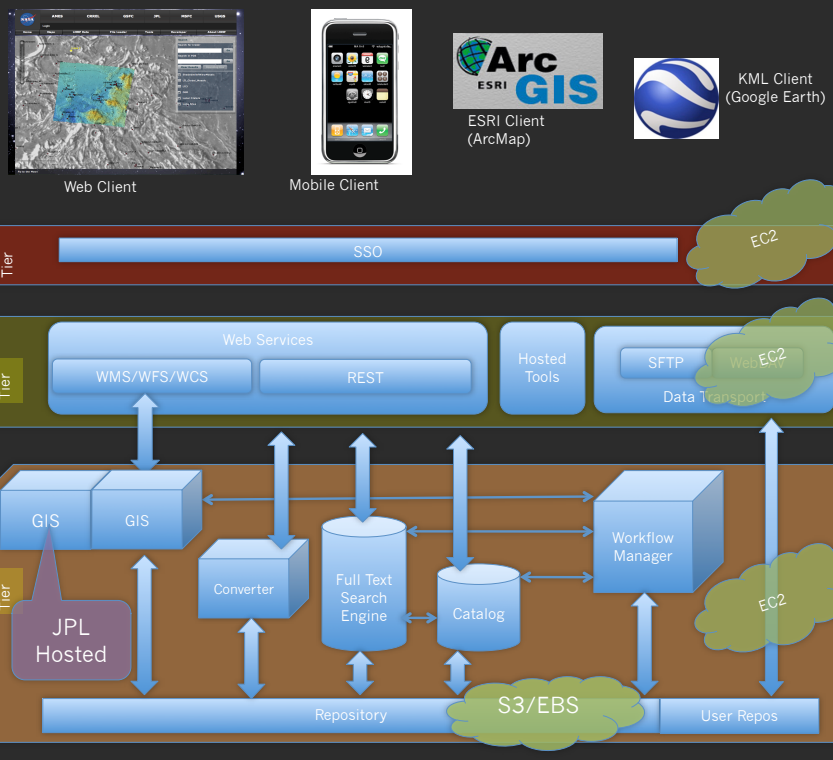
Nomenclature | Gravity | **Data Management (Products)** | Hazards | Surface Roughness

Imagery | Mineralogy | Temperature | Rock & Craters | Slopes | Contours | Elevation

Architectural Principles

Principles	Implementation
Usability	Tiered Architecture
Scalability	Service Oriented
Maintainability	Seamless Security
Open	Cloud-enabled
Data driven	
Standardization	
Reliability	
Security	



Data Clients

Available on multiple user interfaces

Loose coupling between UI and data services

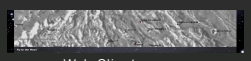
Same URL for Cloud or Privately Hosted Data

Security

Multiple LDAP Sources

Exposes REST Wrappers

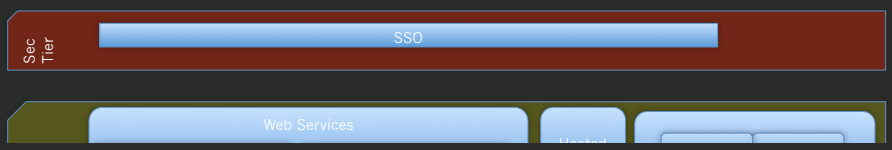
Single Sign-On Across All Domains



Web Client

Mobile Client

(ArcMap)



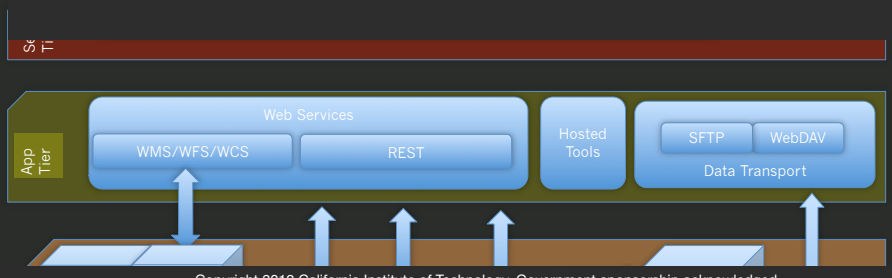
Service Application

GIS information provided by standard APIs, XML meta-catalog for GIS sources

All features invoked via REST

https://dev.lmmp.nasa.gov/LMMP/rest/transform/latlon/subset/stream/png?src=9a05d86b-ffa4-4d5d-be54-9a8e90b27bce&ulx=43.4268&uly=-29.7554&lrx=43.8398&lry=-30.0454

WebDAV, SFTP access to User Storage



Data

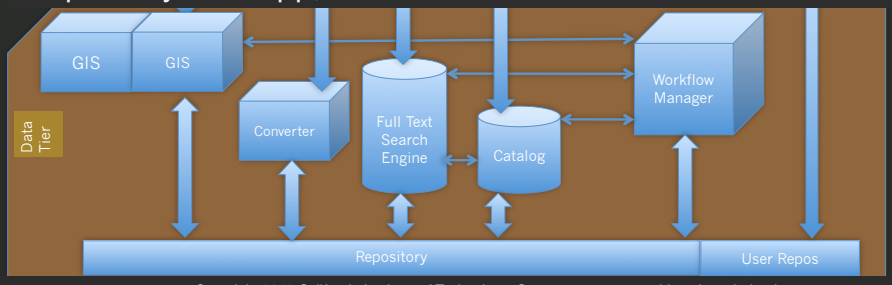
GIS Servers – JPL (T)WMS, ArcGIS

Data Converters – GDAL, Java ImageIO

Search Engine, Catalog – Solr, OODT, BerkeleyDBXML

Workflow – Jabber, Amazon SNS/SQS

Repository – Netapp, Amazon EBS/S3



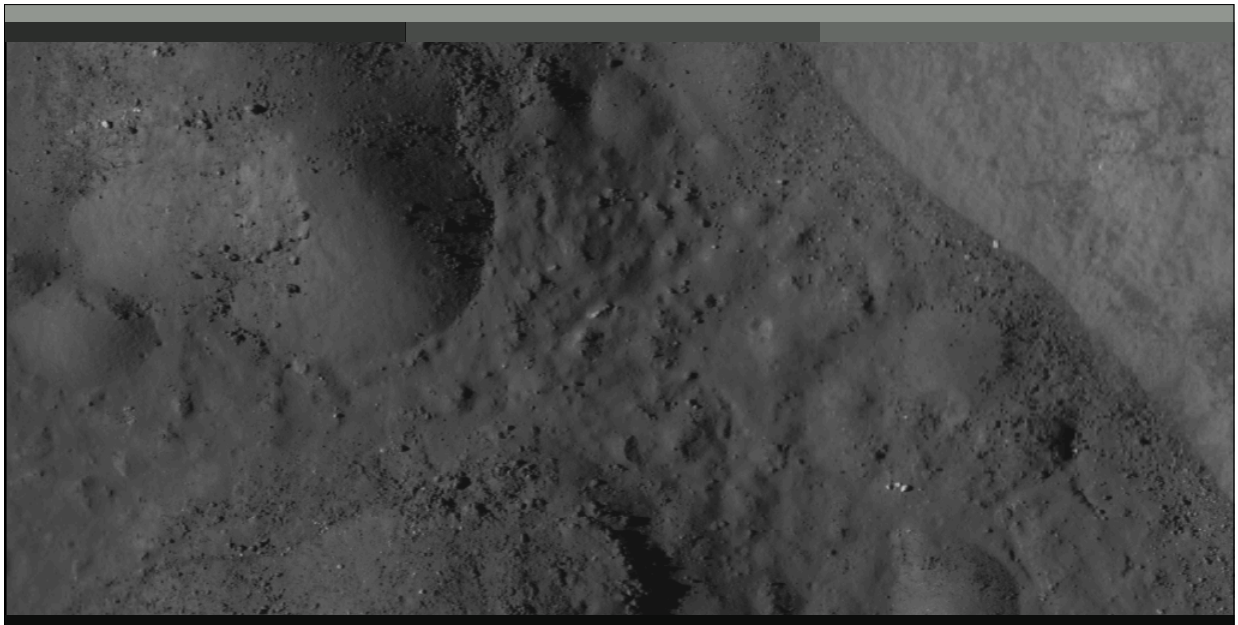
Key Takeaways

LMMP system design allows for modularized component design, vendor independence

Utilizes cloud infrastructure where appropriate, make access invisible to the user

Security model transparently grants access to data and protects sensitive products

Image Tiling on the Cloud



The Challenge

How do we make these large images usable by desktop computers, mobile devices, and other memory constrained products?

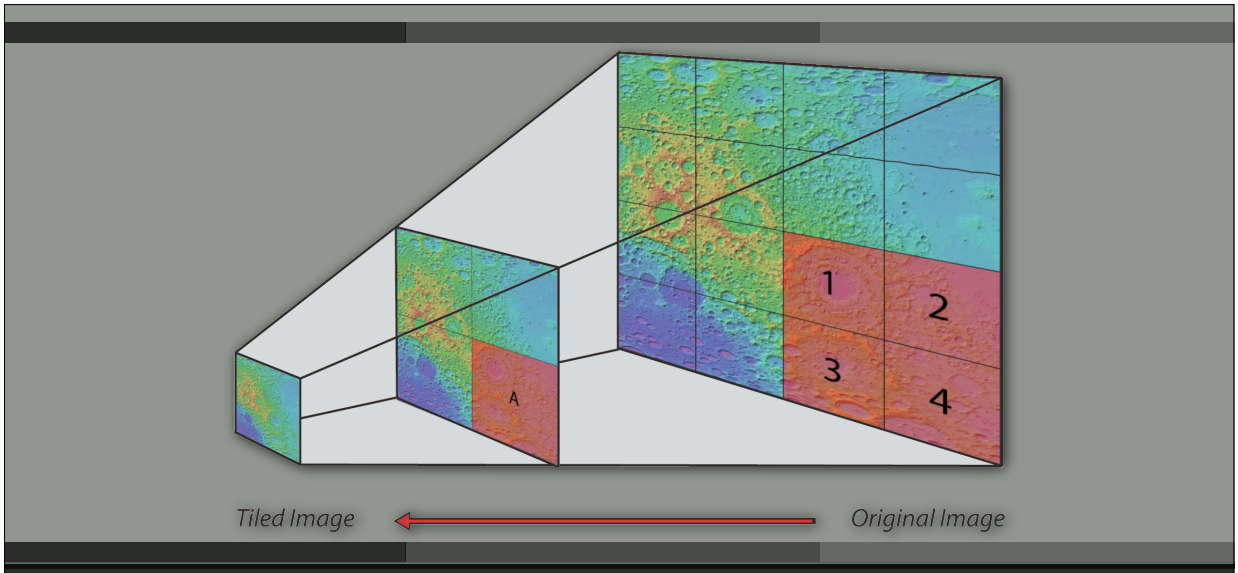
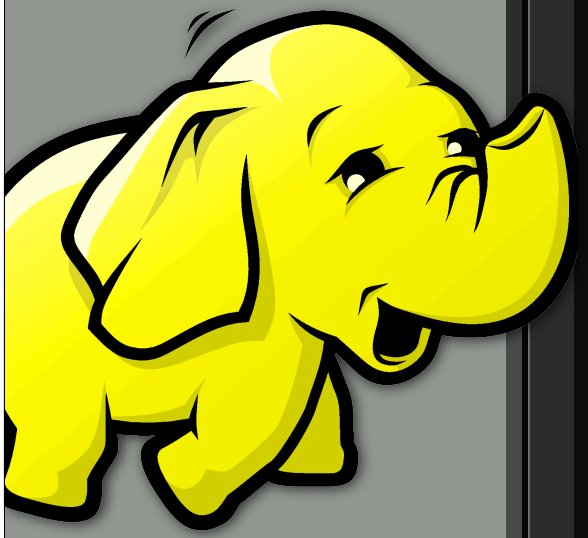


Image Tiling Process

Starting from the full resolution image, divide images into small tiles.

Combine and shrink tiles to create tiles for the next zoom level

Continue process until the final zoom level has only one tile

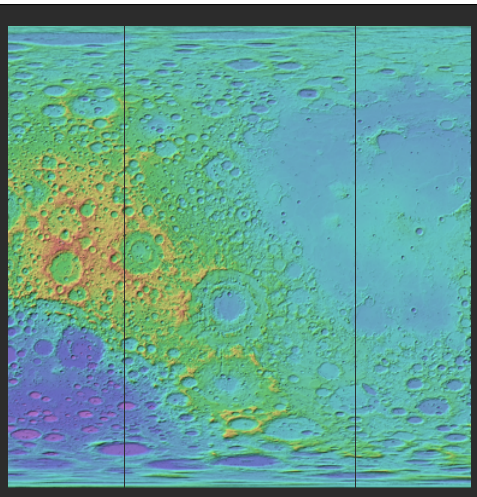


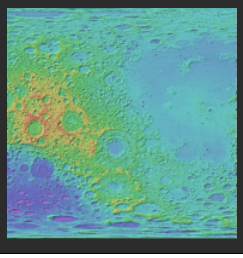
Hadoop

How We Use It

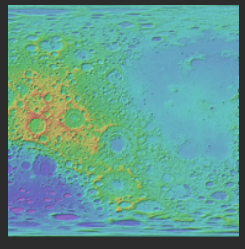
Map Function – Takes image tiles and adds metadata about its position as well as scaling them for the next zoom level

Reduce Function – Collects the 4 adjacent tiles and merges them to create a new tile





Iterate several times
and we're done!



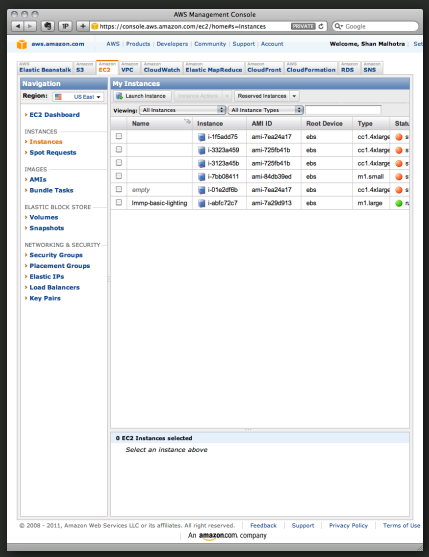
Cloud Implementation Using Amazon EC2

Test image, 2.77 gigabytes LRO LOLA (Lunar Orbiter Laser Altimeter) colorized digital elevation map which produced 9.1 gigabytes set of tiles

Amazon EC2 is a cloud computing infrastructure allowing users to “rent” virtual machines. Tested the “Large” and “Cluster Compute” instances

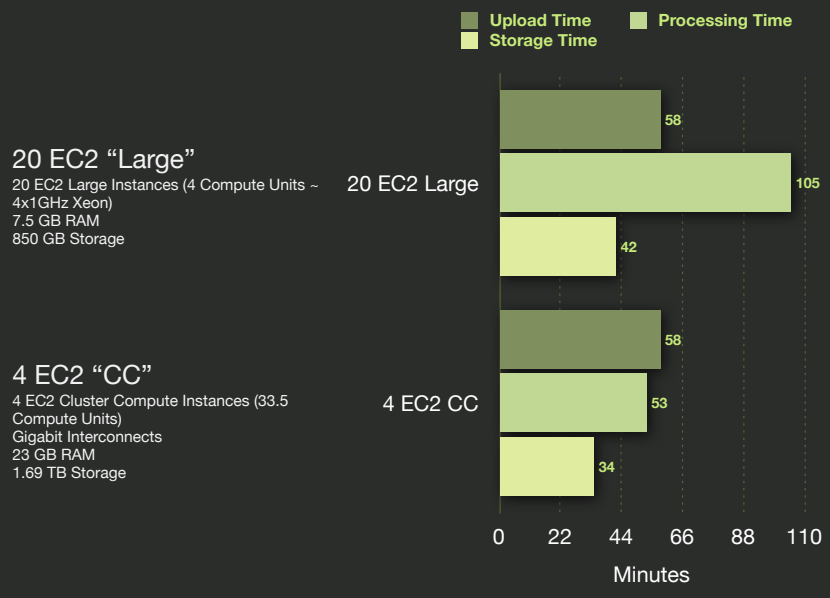
Installed Hadoop framework on a number of EC2 instances

Output image files stored on Amazon S3, a cloud storage system



Test Results

Performance Times (circa 2011)



Key Takeaways

Cloud computing provides an easily quantifiable cost on data product generation

Choice of cloud services dependent on the specific use

Hadoop framework provides a simple programmatic interface for developing distributed computing applications for problems that are parallelizable

Thank You!

http://www.lmmp.nasa.gov/

George Chang | @gchang | gchang@jpl.nasa.gov