Enabling Science in the Cloud: A Remote Sensing Data Processing Service for Environmental Science Analysis

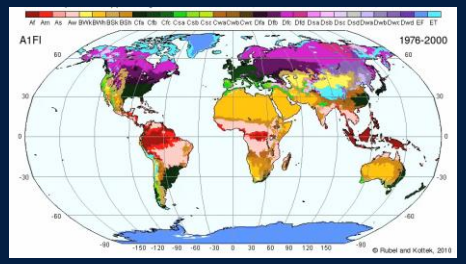
Catharine van Ingen[1], Jie Li[2], Youngryel Ryu[3], Marty Humphrey[2], Deb Agarwal[4], Keith Jackson[4]

[1]Microsoft Research; [2]University of Virginia;

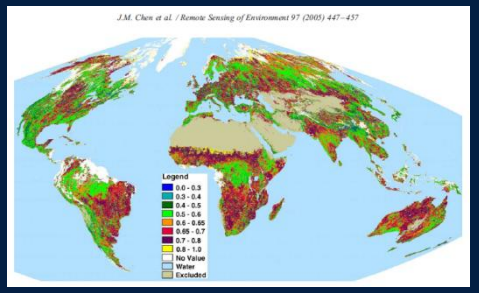
[3]University of California, Berkeley; [4]Lawrence Berkeley Laboratory

AGU Meeting of the Americas, August 2010 IN21A-07

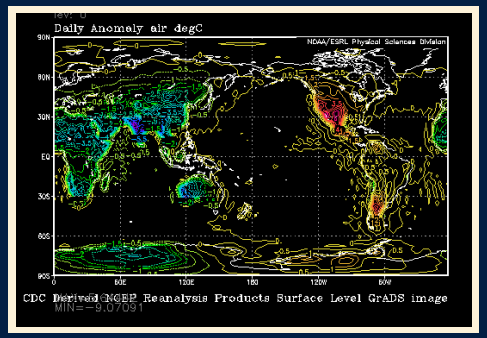
Synthesizing Imagery, Sensors, Models and Field Data



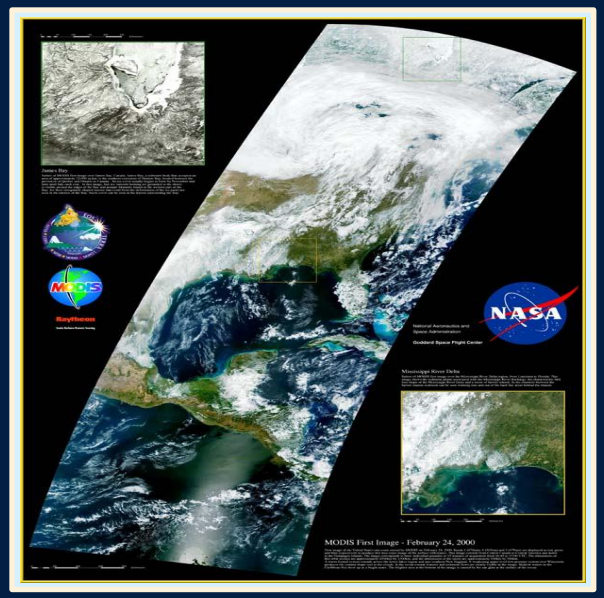
Climate classification
~1MB (1file)



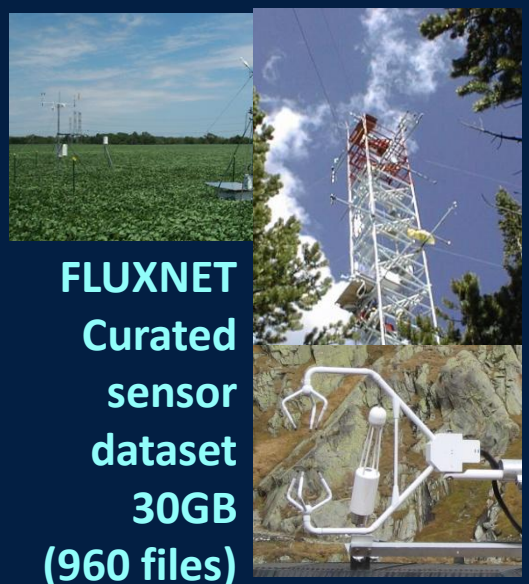
Vegetative clumping
~5MB (1file)



NCEP/NCAR ~100MB
(4K files)



NASA MODIS imagery archives
5 TB (600K files)



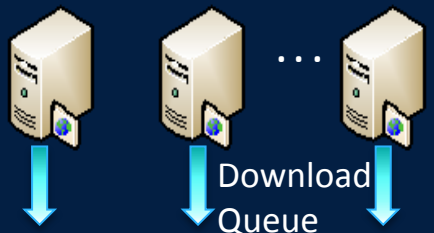
FLUXNET
Curated
sensor
dataset
30GB
(960 files)

FLUXNET
curated field
dataset
2 KB (1 file)

Sizes given are 1 US year
20 US year ~ 1 global land surface year

Four Stage Image Processing Pipeline

Source Imagery Download Sites



Download Queue

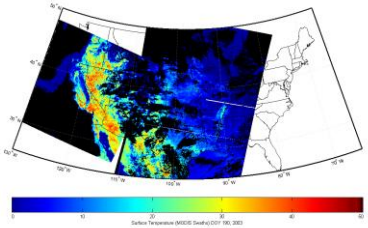
Source Metadata

Request Queue

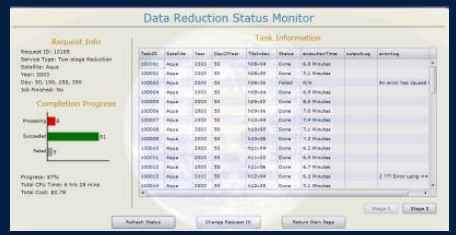
Scientists

MODISAzure Service Web Role Portal

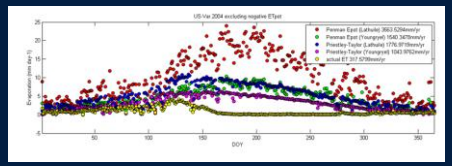
Data Collection Stage



Reprojection Queue

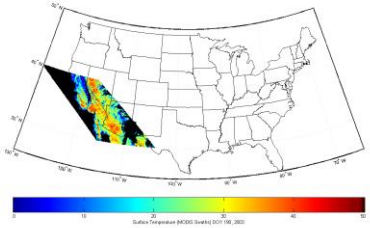


Science results



Scientific Results Download

Reprojection Stage

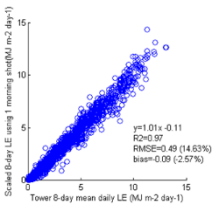


Reduction #1 Queue

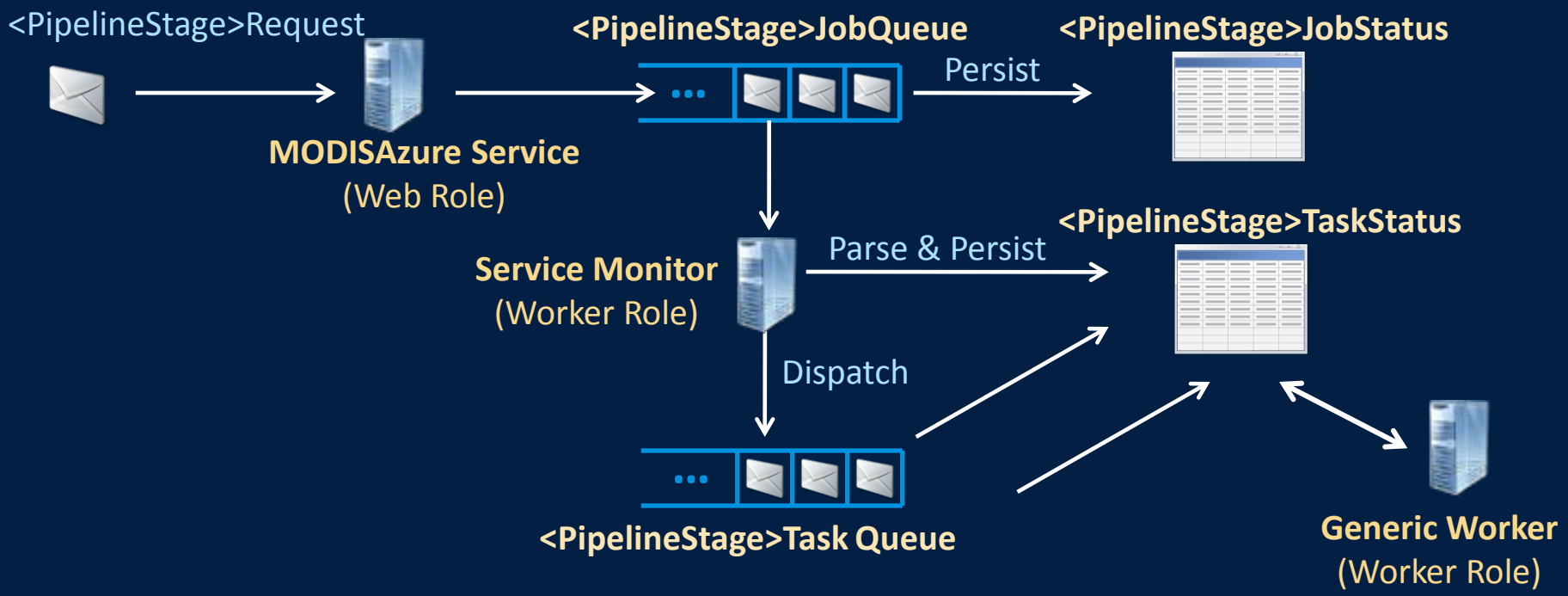
Derivation Reduction Stage

Reduction #2 Queue

Analysis Reduction Stage



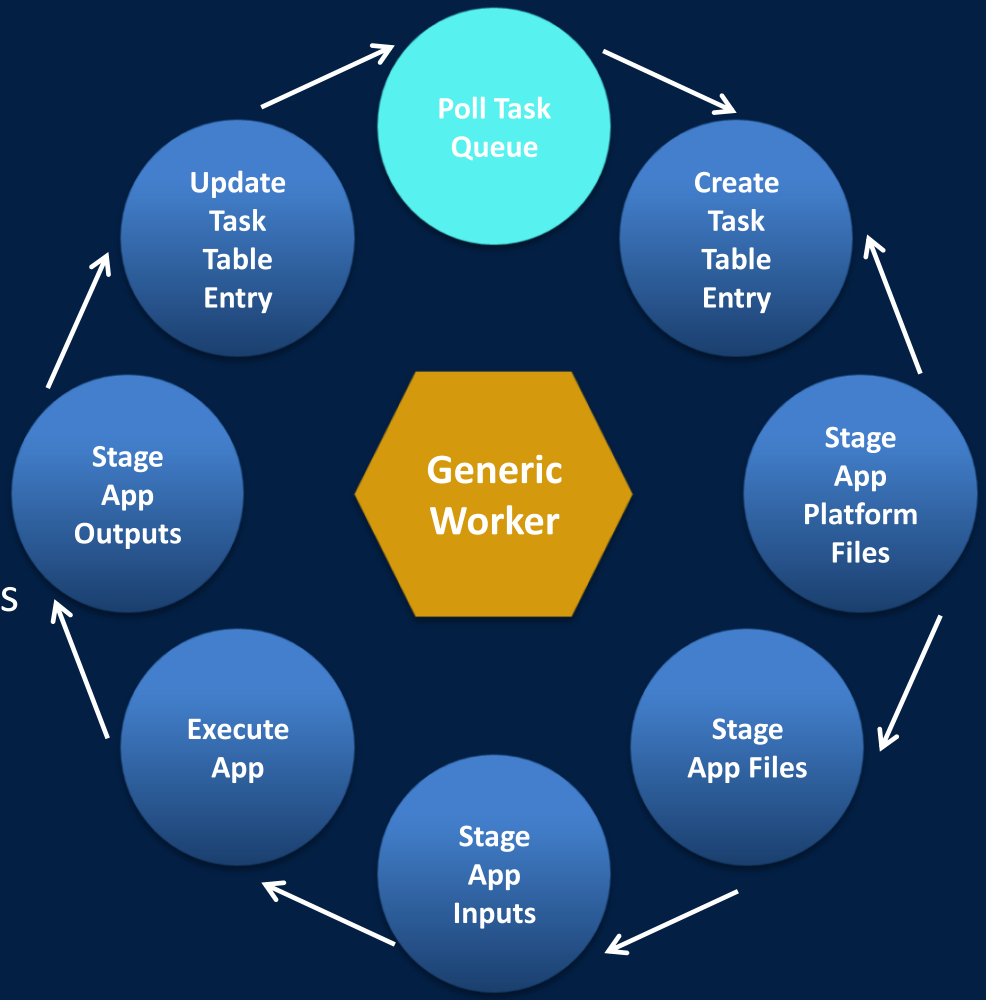
MODISAzure: Architectural Big Picture



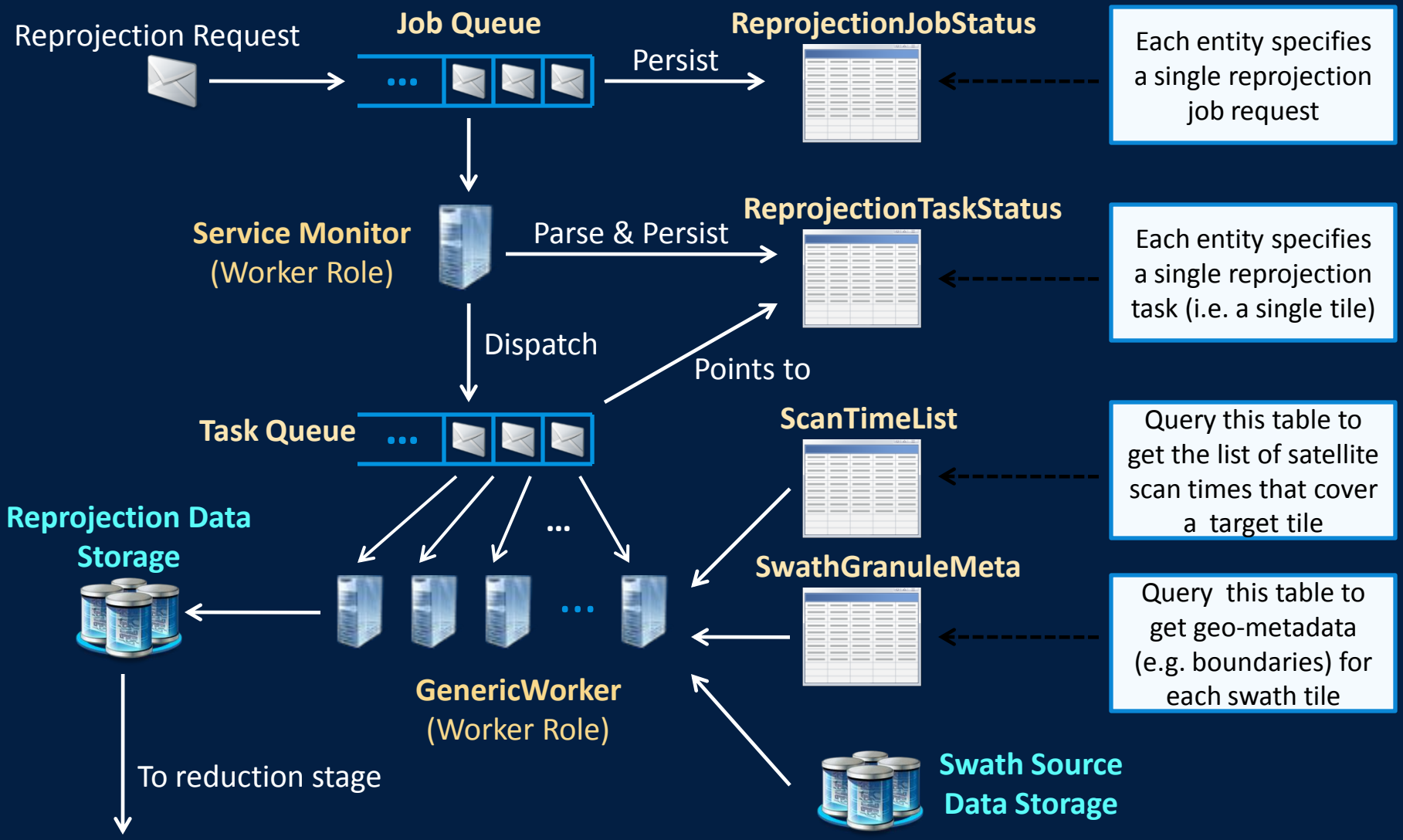
- **ModisAzure Service** is the Web Role front door
  - Receives all user requests
  - Queues request to appropriate Download, Reprojection, or Reduction Job Queue
- **Service Monitor** is a dedicated Worker Role
  - Parses all job requests into tasks – recoverable units of work
  - Execution status of all jobs and tasks persisted in Tables

Inside A Generic Worker

- Manages application sandbox
  - Ensures all application binaries such as the MatLab runtime are installed for “known” application types
  - Stages all input blobs from Azure storage to local files
  - Passes any marshalled inputs to uploaded application binary
  - Stages all output blobs to Azure storage from local files
  - Preserves any marshalled outputs to the appropriate Task table
- Manages all task status
  - Dequeues tasks created by the Service Monitor
  - Retries failed tasks 3 times
  - Maintains all task status
- Simplifies desktop development and cloud deployment



Reprojection Service



Storage Management

Source

- Original source image download
  - Can be deleted when all dependent reprojections complete

- Reduction results
  - Older results can be aged out over time
  - A zip file blob is created for each job to simplify download

Reduction Storage

- Reprojection results
  - May include the same target tile at different spatial resolution

Reprojection Storage

Metadata Storage

- Metadata includes geospatial lookup, known application library binaries, etc
  - Necessary for service function
  - Never directly accessed by scientist code

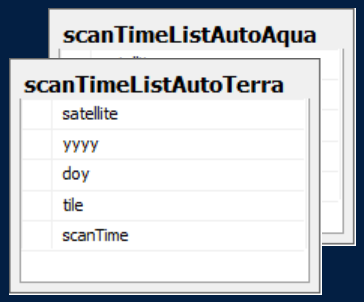
Storage separated by usage to simplify management policies

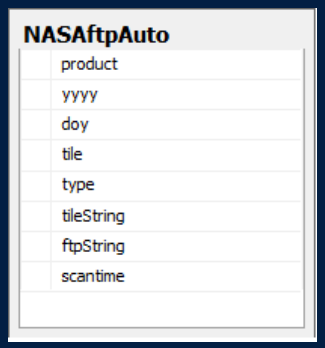
Pipeline Stage Priorities and Interactions

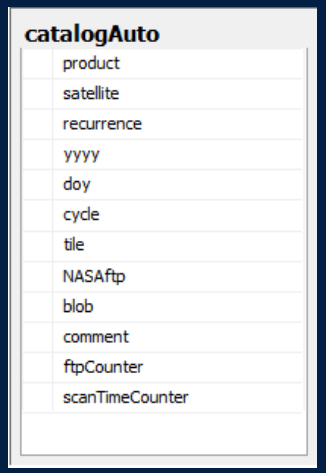
- The Web Portal Role, Service Monitor Role and 5 Generic Worker Roles are deployed at most times
  - 5 Generic Workers are sufficient for reduction algorithm testing and development ($20/day)
  - Early results returned to scientist while deploying up to 93 additional Generic Workers; such a deployment typically takes 45 minutes
  - Deployment taken down when long periods of idle time are known
  - Heuristic for scaling number of Generic Workers up and down
- Download stage runs in the deep background in all deployed generic worker roles
  - IO, not CPU bound so no competition
- Reduction tasks that have available inputs run preferentially to Reprojection tasks
  - Expedites interactive science result generation
  - If no available inputs and a backlog of reprojection tasks, number of Generic Workers scale up naturally until backlog addressed and reduction can continue
  - Second stage reduction runs only after all first stage reductions have completed

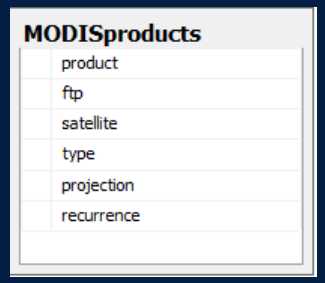
Image Catalog

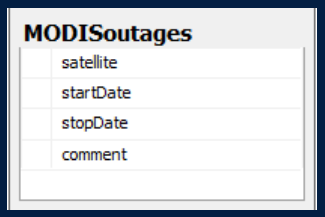
- Necessary to track and reconcile contents of source ftp sites, source blob store and reprojected blob store
  - Handles cases of intermittent download failures that are not addressed by simple retry
  - Identifies source files that are missing due to satellite outage
  - Enables choice of Terra or Aqua product based on scan time and availability

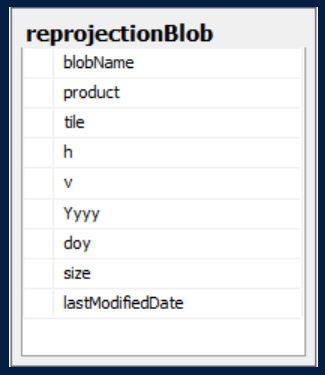


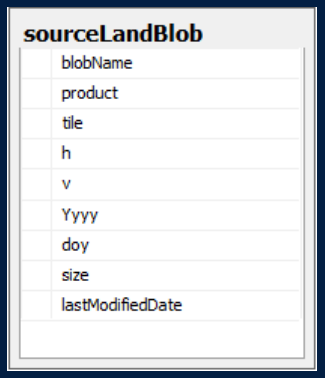








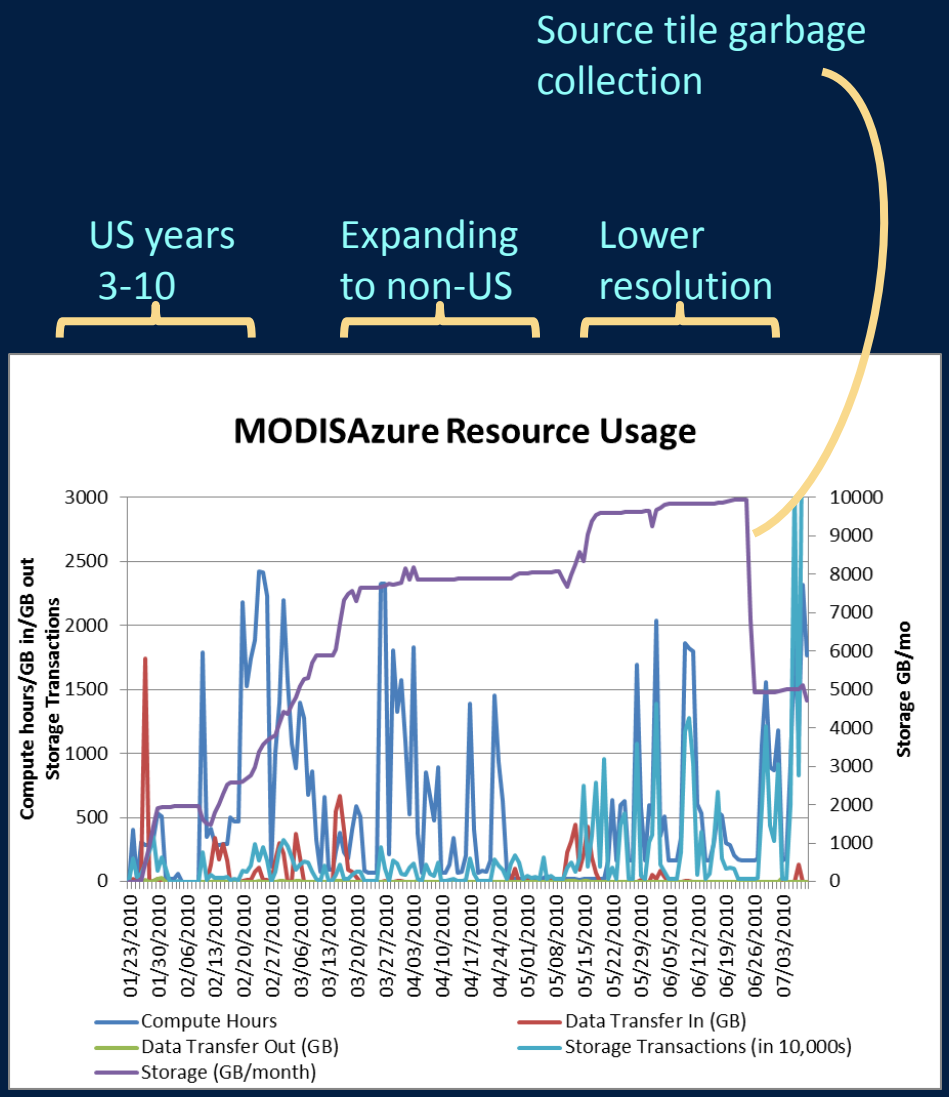




Current Status (8/2010)

Encouraging science results lead to changing resource needs

- Fine scale computation expanded to cover more of the globe: 2x compute requirements and 2x (transient) storage requirements
- Lower resolution global computation added: .5x compute requirements and 2x (transient) storage and higher IOP/cpu reprojection
- Now underway: geo-spatial validation with yearly aggregate: shifts reduction to IO intensive



Learnings

- Lowering the barriers to use remote sensing data can enable science
  - NASA makes the data accessible, not science ready
  - At AGU 2009, we learned that a cloud service that just made on-demand jpg mosaics would help tremendously
- Science and algorithm debugging benefit from the same infrastructure as both need to scale up and down
  - Debugging an algorithm on the desktop isn't enough – you have to debug in the cloud too
  - Whenever running at scale in the cloud, you must reduce down to the desktop to understand the results
- Scaling up requires additional work as even a 0.01% failure rate is time consuming to understand
  - Bake in the faults for scaling and resilience Bake in the catalog for end:end reconciliation of sources and results