Satellite Image Processing Applications in MedioGRID

Ovidiu Mureșan, Dorian Gorgan
Florin Pop, Valentin Cristea

{Ovidiu.Muresan, Dorian.Gorgan}@cs.utcluj.ro
{florinpop, valentin}@cs.pub.ro
Outline

- Introduction
- MODIS Satellite Imagery
- Grid Image Processing in MedioGRID – General Architecture
  - Data Management System
  - Command and Result Dissemination Component
- Communication and data flow in MedioGRID
- Grid Monitoring Solutions
- Experimental results
- Conclusions
Objectives

- Provide a general purpose satellite image processing system for extracting relevant environmental and meteorological parameters
- Use GRID computing resources and real-time MODIS satellite imagery
- Implement fire detection and water coverage detection (used for flooded area extent estimation)
- Modular architecture

Research institutions:

- Universitatea Tehnică Cluj-Napoca (UTCN)
- Universitatea Politehnica Bucureşti (UPB)
- Universitatea Politehnica Timișoara (UPT)
- Universitatea de Vest Timișoara (UVT)
- iQuest Technologies (iQuest)
- Universitatea Babeș-Bolyai Cluj-Napoca (UBB)
- Administrația Națională de Meteorologie R.A. (ANM)
MODIS Satellite Imagery

- Produced by sensors onboard the Terra and Aqua satellites
- Covers the entire surface of the Earth
- 36 observational channels
- 250m to 1km spatial resolution
- Data distributed by the NASA DAAC
Grid Image Processing in MedioGRID – General Architecture
Data Management System

- Requirements
  - Robustness
  - Efficiency
  - Transparency

- Components
  - Data Mirroring and Indexing Component
    - Realtime fetching and indexing of MODIS data
    - Data processing operations
      - Split data granules into the composing spectral bands
      - Index associated XML data
      - Generate full color JPEG images
  - Metadata Catalog Service
    - On top of the Replica Location Service
    - Answers data queries
    - Uses OGSA-DAI
  - Data Access Component
    - Provides separate access to the MODIS data layers
Command and result dissemination component

- Integrated system for commanding GRID processing operations and result dissemination
- Creates the interface between the user and the GRID computing resources
- Should include a GIS component:
  - Integrates both topology and satellite imagery data
  - Realtime display of processing results
Communication and data flow

1. Command a GRID processing operation
2. Query the metadata service
3. A list of logical file names
4. Schedules GRID jobs
5a. Query RLS service
5b. A list of physical file names
5c. Access data from a specific GridFTP location
5d. Store/update spatial data as result of GRID processing operations
5e. Send job completion notification
**Grid Monitoring Solutions**

- **GridICE**: different aggregations and partitions of monitoring data are provided with a different abstraction level of a Grid:
  - the Virtual Organization level,
  - the Grid Operation Center level,
  - the Site Administration level and
  - the End-User level.

- **R-GMA** (Relational Grid monitoring Architecture):
  - three components: Consumers, Producers and directory service
  - part of the "Enabling Grids for E-science in Europe"

- **Ganglia**: toolkit for monitoring clusters and hierarchical aggregations of clusters
  - collects system status information and makes it available via a web interface.
  - Ganglia status can be subscribed to and aggregated across multiple systems.

- **MonALISA** (MONitoring Agents using a Large Integrated Services Architecture)
  - designed as an ensemble of autonomous multi-threaded, self-describing agent based subsystems which are registered as dynamic services, and are able to collaborate and cooperate in performing a wide range of information gathering and processing tasks.
MedioGrid Monitoring Solutions

- **Ganglia**
  - access to each node in MedioGRID cluster and
  - request information about state of node: load, CPU usage, etc.

- **MonALISA**
  - Agent based support for collecting system and job data from each node in MedioGRID network.
  - Centralize date and view it in Client and/or Repository.

- **ApMon**
  - APIs in C, C++, Java, Python, Perl
  - Library for collect dynamic (and complete) information about jobs and systems and send it to a MonALISA database.
  - Used in MedioGRID to create a “check-point” mechanism for Satellite Image Processing Applications
The optimal performance is achieved when using 3 to 15 GridFTP parallel data streams.

Depending on the file size, larger values for the number of parallel connections offer better performance when transferring data over the WAN links.

Data transfer performance is directly influenced by the number of parallel data streams.
Conclusions

- Proposed a high-level architecture for grid-based satellite image processing
- Studied the GridFTP LAN and WAN transfer characteristics
- Development directions
  - Integrate the system with a powerful GIS engine for efficient result dissemination
  - Implement specialized GIS tool providing for the evaluation of flood and fire evolution over time