Grid Data Management

Thomas Sandholm <sandholm@pdc.kth.se>

KTH VETENSKAP OCH KONST
CENTER FOR PARALLEL COMPUTERS

Outline

- **Grid File System**
- File Replication
- File Transfer
- Data Access and Storage Resource Management

Grid File System Goal

Virtual Grid Operating System

- Transparent access to data distributed in a Grid environment
- User interface to find files
- Automatic file transfer between storage systems and computing systems

Grid File System Requirements

- Secure access
- File naming and replication
- High-speed transfers
- Common access interface to data

Grid File System Secure Access

- Global authentication (single sign-on)
- Local authorization: role-based v.s. identity-based
- Sites retain full control over their resources
- 3rd party authentication
- Protect data on storage systems
- Support secure data transfers
- Protect knowledge about existence of data

Grid File System File Naming and Replication

- Global namespace
  - Unique identifier for each file
  - Mapping: Logical file name -> Physical file name(s)
  - No centralized data storage
  - Mechanisms for discovery and data storage
- File replication
  - Master file and replicas
  - Consistency: data are seldom modified (80-20)
  - Lifetime of master and replicas
  - Replication – atomic operation
    - Pre-processing
    - File transfer
    - Insert replica in the file system

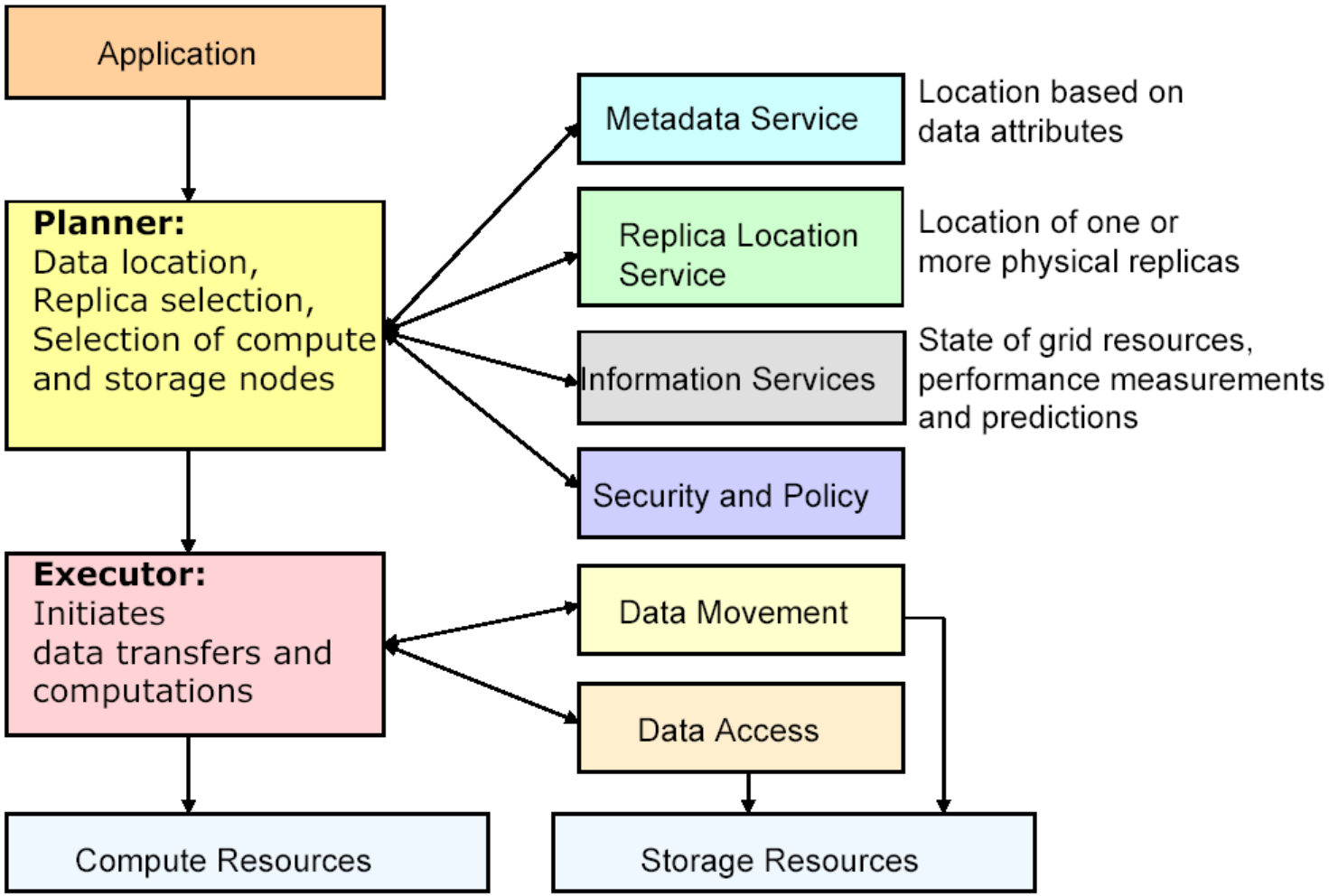
Grid File System High-Speed Transfers

- Scalability of HTTP (1.1 chunking)
- Scalability of FTP (TCP congestion control limitation) higher QoS than HTTP
- Secure data channels
- Secure control channels
- 3rd party control
- Parallel transfers

Grid File System Common Access Interface to Data

- Heterogeneous storage systems – different in architecture and administration policy
- Retain local control
- Many different access protocols lead to proprietary clients
- Leverage existing infrastructure (RDBS, File system) – add a Grid Access layer
- Handle small as well as very large amounts of data
- Decoupled from transfer protocols (may allow many different transfer QoS depending on network and payload size)

Grid File System Functional Overview



Outline

- Grid File System
- **File Replication**
- File Transfer
- Data Access and Storage Resource Management

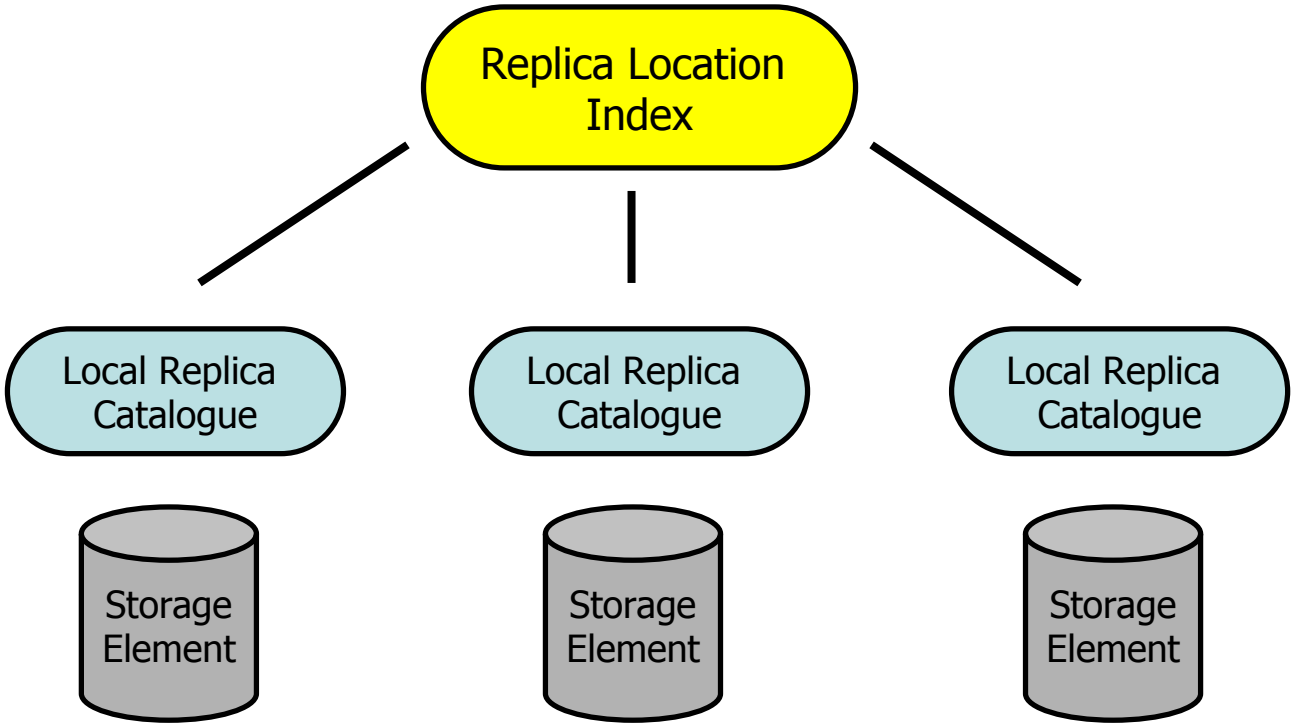
File Replication

- Data intensive applications
  - Produce TeraBytes or PetaBytes of data
- Replicate data at multiple locations
  - Fault tolerance
  - Performance: avoid wide area transfer latencies, achieve load balancing
- Issues:
  - Locating replicas of desired files
  - Creating new replicas and registering their locations
  - Scalability
  - Reliability

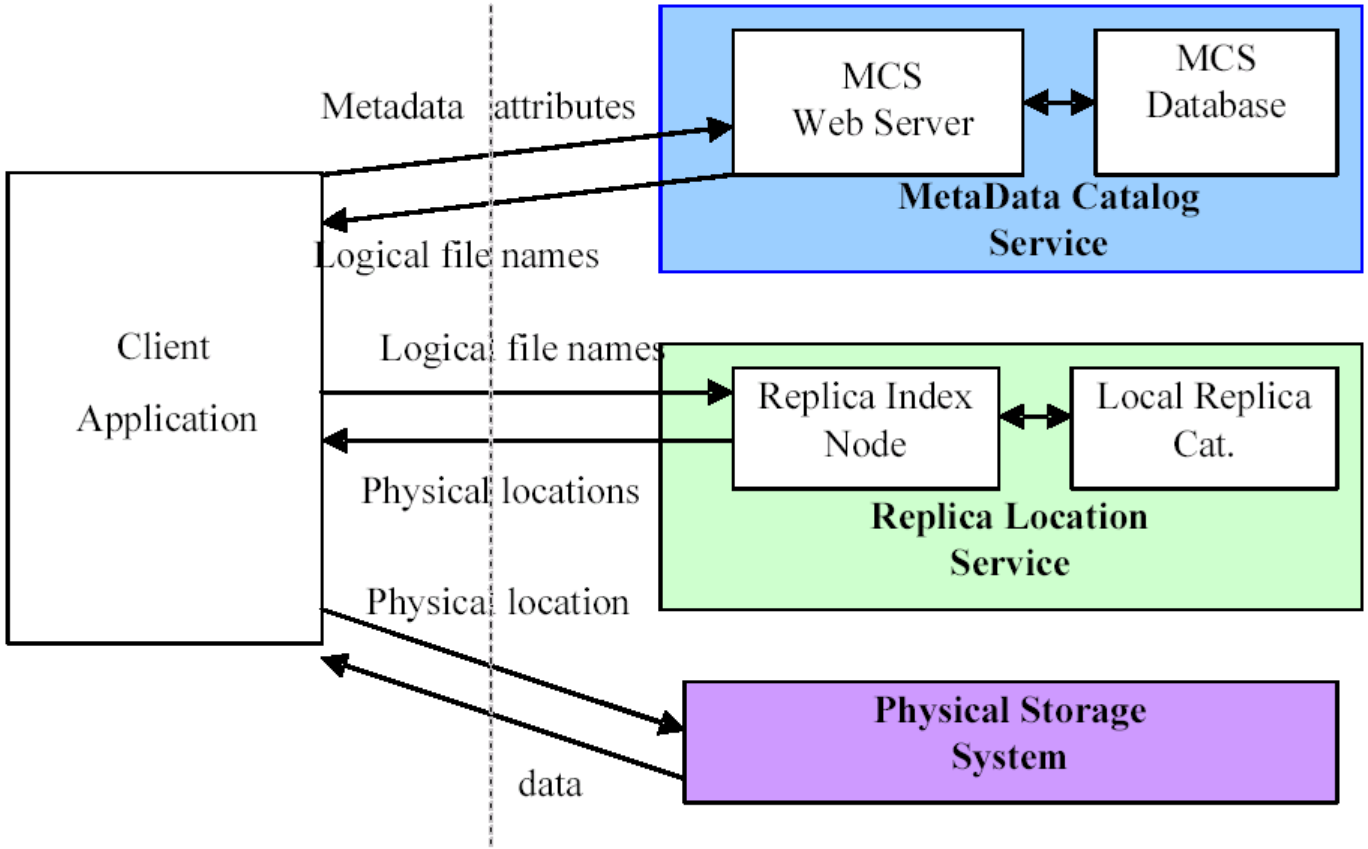
Replica Location Service

- Records locations of data copies and allows discovery of replicas
- Mapping Logical File Name-> Target Name
- Target Names may be represented by
  - Physical file name(s)
  - Logical file name(s) to create hierarchies
- Replica Location Index at a VO level
- Replica Catalogues deployed at each site with knowledge about all files at the local site
- Does **not** handle transfers

RLS Architecture



MetaData Catalogue Service



RLS and MCS implementations

- RLS
  - Globus
  - EU DataGrid (Replica Optimization)
  - Pre-WS implementations based on SQL DB
- MCS
  - GriPhyN, NVO, ESG project lead by ISI
  - WS implementation based on OGSA DAI
  - Using Globus core infrastructure but not part of the Globus Toolkit

Outline

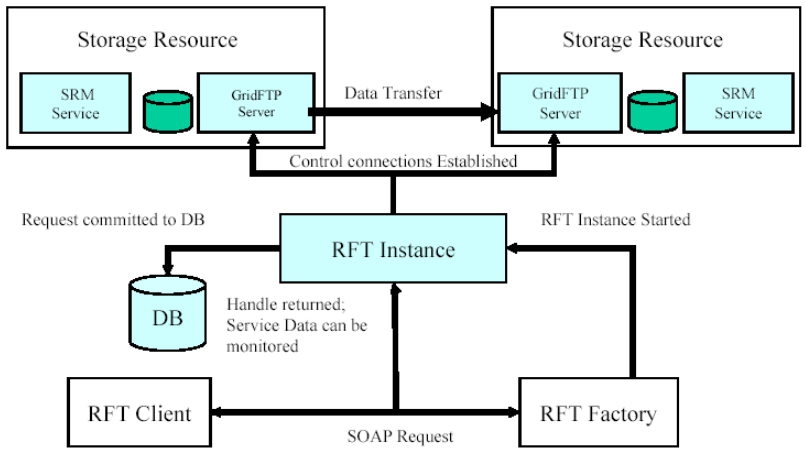
- Grid File System
- File Replication
- **File Transfer**
- Data Access and Storage Resource Management

GridFTP

- Transfer and replication of large data sets (TeraBytes, PetaBytes)
- Separation of control and data channels
- Multiple **parallel data channels**: TCP congestion window workaround
- **Striped data transfer**: parts of file transferred in parallel by multiple hosts and then consolidated
- **Partial file transfer**: off-set region based
- 3rd party transfer: url-copy
- Reliable monitored transfers: **restart and performance markers**
- **Optimization of TCP buffer size** (manual and automatic)
- **GSI** and Kerberos 3rd party authentication
- RFC 959 (FTP), RFC 2228 (security extensions), RFC 2389 (feature negotiation and command options), IETF Draft MLST-16 (stream mode restart and standard file listings) , GGF GFD.020 (GridFTP protocol)
- Pre WS Implementations – Globus, Fermilab

Reliable File Transfer Service

- Manages 3rd party transfers through WSRF interface
- Restart markers are used to provide reliability
- Failed transfer retry strategy configuration
- Set of file transfers submitted in batch
- All-or-nothing transfers optional
- Recursive directory transfers
- Globus implementation



Outline

- Grid File System
- File Replication
- File Transfer
- **Data Access and Storage Resource Management**

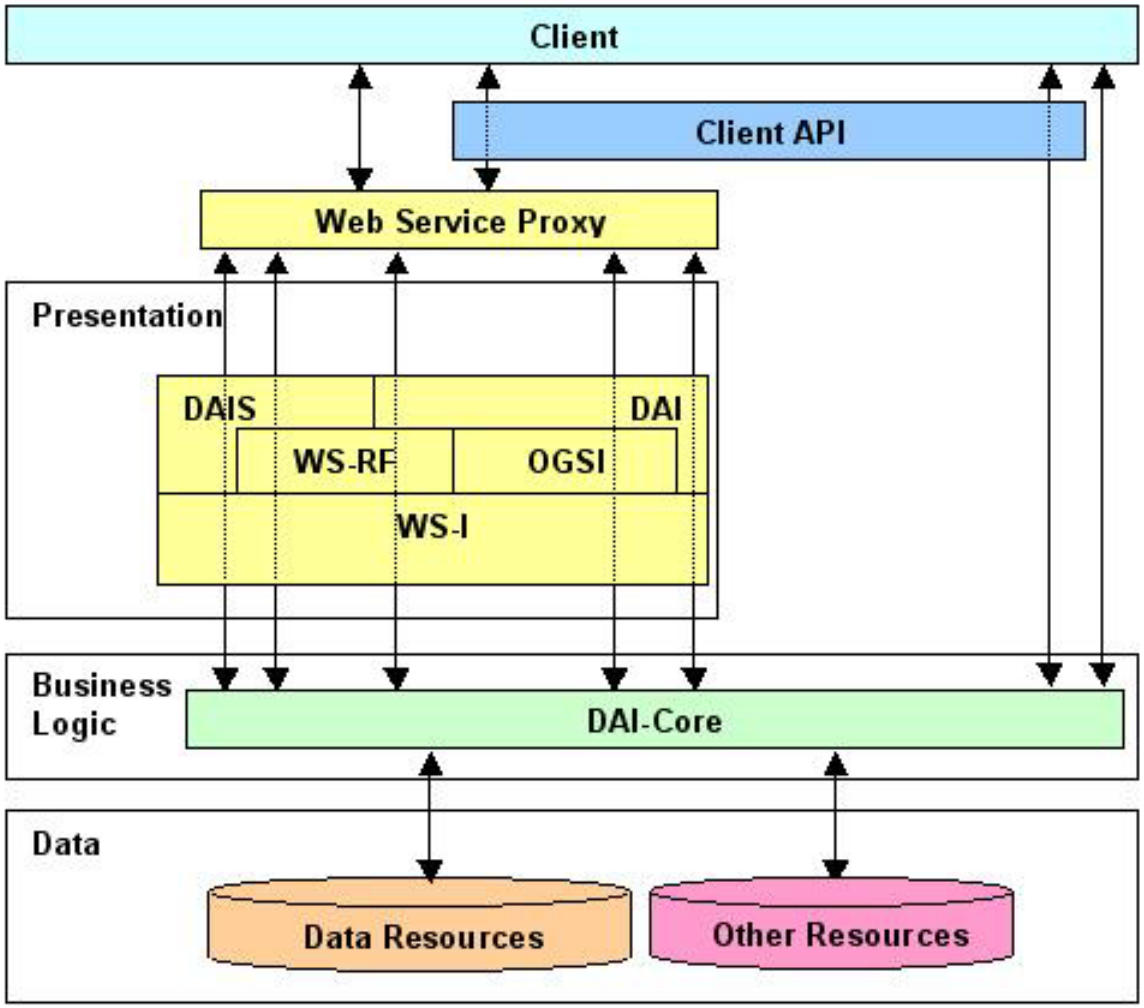
OGSA Data Access and Integration

- Extensible framework that allows access to and updating of data resources
- Uniform access to heterogeneous resources (RDBMS, XMLDBs, structured files)
- Resource capabilities (and data model) exposed – SQL, XQuery
- Abstraction of database driver technology, data formatting techniques, and delivery mechanism
- Can be used to build higher-level services such as data federation and distributed query processing

Project:

- UK e-Science Grid Core Project lead by University of Edinburgh (EPCC)
- IBM and Oracle participation
- GGF DAIS-WG standardization effort
- Globus Toolkit contribution

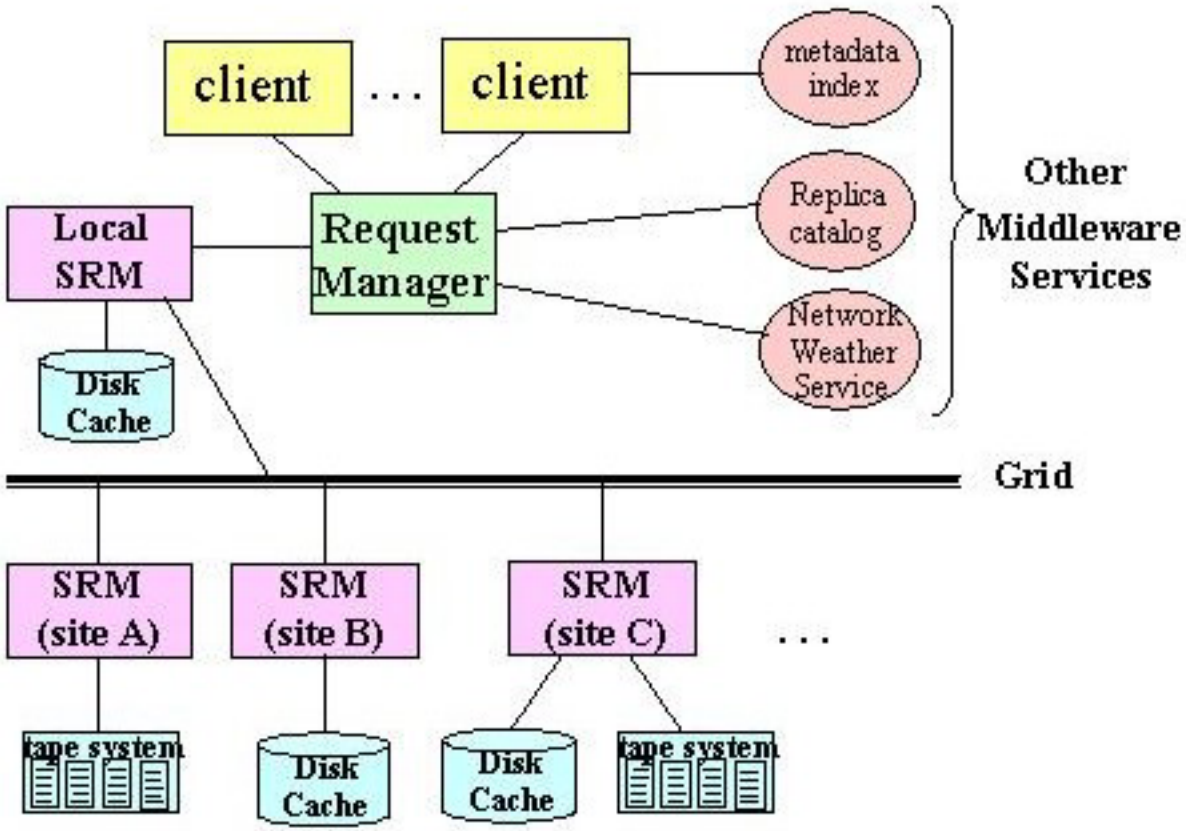
OGSA DAI Architecture



LBNL Grid Storage Resource Manager

- Grid Middleware to manage storage resources
  - Complement management of compute and network resources
  - Compute job file stage-in and stage-out services
- Coordinate distributed disk caches
  - Smart replacement
  - File pinning
- Seamless access to tape storage
  - Automate staging and archive requests in background
  - Insulate client from hardware and network failures

Grid SRM Architecture



SDSC Storage Resource Broker

- Uniform interface for connecting to heterogeneous data resources over a network and accessing replicated data sets
- Access files based on attributes as opposed to file names using a metadata catalogue
- Archiving, Caching, Synchronization and Backup
- Framework comprising:
  - Distributed File System
  - Data Grid Management
  - Digital Library
  - Semantic Web

SDSC SRB Architecture