Grid Computing 7700
Fall 2005
Lecture 14: Data Grids

Gabrielle Allen
allen@bit.csc.lsu.edu

http://www.cct.lsu.edu/~gallen
The Problem

Distributed data
Different storage systems
Different networks
Examples

- **Gravitational Wave Detectors: LIGO (10s of terabytes/yr)**
  - Laser interferometric detector picking up space-time distortions (and earthly events)
  - Continuous time series data at different frequencies

- **Observatories: Sloan Digital Sky Survey (15 terabytes)** is mapping one-quarter of the entire sky
  - Positions and absolute brightness of more than 100 million celestial objects
  - Distance to a million of the nearest galaxies
  - Distances to 100,000 quasars
Examples

- High Energy Physics: CERN CMS Detector (1 petabyte/yr)
  - Compact Muon Solenoid detector
  - Will be part of circular Large Hadron Collider (LHC)
  - High energy protons collide 800M times per second
  - Higgs boson once every 90s
  - Data already reduced by factor of 10M

- Numerical Simulations: Numerical Relativity (terabytes/run)
  - Modeling black holes, neutron stars, gravitational waves
  - Essential for understanding data from LIGO
Examples

- Climate Modelling (~1TB/run)
  - Japanese Earth Simulator!
  - Weather prediction, global warming, hurricanes, ...
  - Need to archive results
Application Types

- Simulation-driven
- Observation/experiment-driven
- Information-intensive
Technology Areas

- Workflow, dataflow, data transformation
- Metadata, data description, logical organization
- Efficient access and queries, data integration
- Distributed data management, data movement, networks
- Storage and caching
- Data analysis, visualization, integrated environments.
"Data Grid"

- Distributed access to distributed data, focusing on reading (large) datasets and creating new datasets
  - Large: terabytes to petabytes
  - Datasets:
    - Simulation data (e.g. Cactus Numerical Relativity)
    - Experimental/observational data (e.g. CERN, LIGO, NVO)
    - Information (e.g. Library of Congress)
    - Data bases

- Data Grid not a good name, because also closely involves computational resources
Data Grids

- Uniform access to data systems
- Metadata catalogues to define where files/data is located and how to access it
- Publication of peer-reviewed data (as important as peer-reviewed papers)
- Domain specific data repositories with uniform interfaces
Data-Intensive Applications

- **Scientific applications**
  - Require mechanisms to transfer, publish, replicate, discover, share and analyse data

- **Business applications**
  - Maintain database consistency, data replication, data discovery, efficient database access
Scientific Applications

- **Data Generation**
  - Running large simulations, operations of experimental, observational devices
  - Main concerns are storing data quick enough and archiving data to make space

- **Post Processing**
  - Validation, annotation, (metadata), formation of data collections

- **Analysis**
  - Combinations of data (including from different collections) are accessed and processed. Usually involves application/domain specific operations, and complex workflows
Scientific Applications

Example data analysis workflow
Business Applications

- Some are scientific in nature: engineering, aerospace, automobile, petrochemical, pharmaceutical
- Others have different characteristics (financial services, travel, fraud detection, retail supply)
  - Data mining: size of data smaller, more emphasis on exploitation of smaller resources
  - Integrated data access: support of business operations, database federation
  - Data usually in structured databases, updated frequently. Emphasis on replication and coherency, transactions.
Data Sources

- A facility which can accept or provide data
  - Software managing a storage device
  - Read-only resource
    - Instrument (experimental device)
    - Simulation
    - Distributed query evaluator

- Can be long-lived or transient
Data Storage

- Data may be held in
  - Databases
  - Mass storage system
  - File systems
  - Directory services

- Data devices
  - Disk cache
  - Networked storage
  - Disk array
  - Robotic tapes
Data Storage

- **Data types**
  - Files
  - Collections of files
  - Tabular data from relational databases
  - Structured data from XML databases
  - Semistructured data (ASCII)
  - Virtual data (abstract data)

- **Access mechanisms**
  - File system open/close/read/write
  - SQL, Xquery, Xpath queries
Data Storage

- To be shared, among people or applications, data must be structured in some known way
  - And of course everyone designs their own way
Integration

- Growing demand to combine data from different disciplines
  - E.g. NVO (National Virtual Observatory) will federate data from different telescopes as well as simulation data
  - E.g. Accurate modelling of hurricanes requires combining coastal and climate models
  - E.g. Advanced visualization tools which do not need to be rewritten for each application

- Ad-hoc integration becoming impractical
Data Discovery

- Need the ability to discover data based on metadata attributes
  - Rather than name or location of a file
  - Issues:
    - How to describe relevant attributes?
    - How to publish, index, and update metadata
    - Discovery mechanisms as for other Grid resources
    - Communities need common vocabulary and semantics
Data Movement

- Move data between storage systems or between programs or between storage systems and programs
- Data movement is the foundation of just about everything on the Grid
- Efficiency very important (large files, WANs)
- Data may be filtered before transfer (prefetch analysis ... “virtual data”)
- Reliable file transfer: maintain state on operations to retry failed operations
Data Replication

- Replicas (copies) of data to reduce access times, maintain local control, improve reliability, etc.
- Need capabilities to
  - Replicate data
  - Locate existing replicas
  - Select between replicas
  - Intelligently create replicas to satisfy demand
  - Satisfy consistency demands
- Replicated data may/may not need to be consistent.
Data Analysis

- Can require significant compute power for processing
- May require a series of processing steps with intermediate temporary or permanent data ("workflow")
- Resulting data may need to be discoverable and used for more analysis
Data Analysis

- Need to construct and execute complex series of operations
  - Planning, scheduling, monitoring
Virtual Data

- An approach to manage complexity of data analysis
  - Specify data by name or attributes
  - Underlying services construct processing steps

- Chimera virtual data system provides a language for specifying this, an interpreter to map to a plan for computing new data, an catalogues etc to discover if data exists before executing the plan.
  - Execution requires many other capabilities for moving data, executing components, storing data, ...
Provenance Information

- For derived data need to record how data was produced.
  - Programs and inputs used to create it
  - Another kind of metadata (need a standard schema)
Architecture

Tiered Model:
- Locally stored datasets
- Locally stored replicas of datasets from remote locations
- Remote datasets
Architecture

- Different classes of data-oriented services
  - Resource level services for data sources: data access (access to and update of structured data), data movement (transport, independent of access), monitoring, auditing and discovery of sources capabilities.
  - Collective services for managing data: primitive mechanisms for managing data across more than one source: 3rd party transfer, data discovery, data transformation, scheduling
  - Collective services for federating data sources: e.g. virtual data base, replication, data naming, federation (schema mediation), consistency, distributed query
  - Domain specific services: specialized capabilities
Data Access

- Data sources either file oriented or database oriented
  - Most work done with file oriented (driven by simulation science)
  - Current focus is on integrated services

- Examples:
  - GridFTP
  - Work of GGF DAI-WG
GridFTP

- Designed (Globus) as fundamental data access and data transport service
  - Uniform interface to different storage systems (e.g. hierarchical, disk, storage brokers)
  - Incompatible data access protocols by these partition data on the Grid
  - Provides extensions to FTP
Common Data Transfer Mechanism

- FTP (File Transfer Protocol) is attractive because
  - Widely implemented and well-understood IETF standard protocol
  - Well defined architecture for extensions, with dynamic discovery of extensions
  - Supports transfers between client and server, and third party transfers between two servers.
Distributed Parallel Storage System (DPSS)

Strip data server architecture
FTP

- File Transfer Protocol
  - provides the basic elements of file sharing between hosts. FTP uses TCP to create a virtual connection for control information and then creates a separate TCP connection for data transfers. The control connection uses an image of the TELNET protocol to exchange commands and messages between hosts.
FTP

---

| User | -------- |
| Interface<--->| User |
|\----^----/ |

---

| User | FTP Commands |
|------|

| User | FTP Replies |
|------|

| User | FTP Replies |
|------|

| Data | FTP Commands |
|------|

| Data | FTP Commands |
|------|

---

| User | File |
|------|

| User | System |
|------|

| User | System |
|------|

Server-FTP

USER-FTP
GridFTP

- Extension of FTP standard (by Globus team)
  - Automatic negotiation of TCP buffer/window sizes
  - Parallel data transfer
  - Third party control of data transfer
  - Partial file transfer
  - Security (GSSAPI authentication with optional integrity/privacy)
  - Support for reliable file transfer (fault recovery methods, restart of failed transfers)
GridFTP

- Can be used both for access data, and to move data.
- For accessing data, server-side processing allows the inclusion of user-written code that can process data prior to transmission.
Third Party Transfer

Data transfer

control

status

GridFTP Server

Client

GridFTP Server
Managing Data Sources

- Storage space
  - Quotas by user, group, project, ...
  - Reservation
  - Bandwidth/data throughout
  - Limited by resources local software
  - Monitoring/auditing services for remote access to information

- E.g. NeST (University of Wisconsin), Storage Resource Managers?
Data Transport

- **Multiple data object transfer services**
  - Submit a large number of simultaneous transfers and control and monitor them
- **Reliable data transfer services**
  - Maintain state of outstanding operations, monitor and restart.
Reliable File Transfer
Data Discovery

- Data items matching given characteristics or metadata
  - E.g. <URL>/blackhole/3march/run1/metric.hdf5
  - E.g. simulation data from two black hole runs from March 3rd 2004 performed by G. Allen

- How to describe data? (Semantic web/Grid)

- Many application domains develop their own metadata description (and others have none)

- SRB (Storage Resource Broker) has a metadata catalogue (MCAT) which attempts to provide an application independent metadata service (automatically).

- But also need to integrate existing databases so need automated mechanisms for adding relevent data.
Federation Services

- Should provide ways for integrating diverse data sources (of all kinds)
  - Enables sharing data between different applications
- E.g. create a virtual database, where a set of databases are presented as a single view through a federated schema (probably impossible)
- Data transparency: location, name, distribution, replication, ownership, heterogeneity, schema change
- Need to mediate between different data models, implement different access methods, ....
Data Mediation

- Transparency of data with respect to the data model
  - Renaming of attributes
  - Mapping from one data model to another
  - Important for sharing data, but also access legacy data
  - Needs well defined semantics for each data model
    - Data ontologies
Replication Services

- Replication management services: create copies and update location services
  - Based on desired availability or popularity
- Replication location services: registries to locate data by providing mapping between data object names (logical names) and storage access services (physical names)
- Consistency services: relationships between replicas
Replica Location Service: Globus RLS

- Keep track of copies (replicas) of files
  - Examples?
- Files are “registered” with RLS registry, users or services can query for the location of files.
- RLS can be distributed (protect from component failure)
- Logical file name: unique identifier for the contents of a file
- Physical file name: location of a copy of the file on a physical storage device
- One logical file name can point to multiple physical file locations
- Can also associate attributes (eg filesize)
Database Access and Integration Services

Data activities relating to the Grid have generally focused on applications where data is stored in files. However, in many scientific and commercial domains, database management systems have a central role in data storage, access, organization, authorization, etc, for numerous applications. The group seeks to promote standards for the development of grid database services, focusing principally on providing consistent access to existing, autonomously managed databases. By focusing on services, the intention is to ease application development through the provision of comparable components, and to track developments in the Open Grid Services Architecture (OGSA). The group does not seek to develop new data storage systems, but rather to make such systems more readily usable individually or collectively within a Grid framework.
Grid Forum Groups: GFS

Grid File System

Data in a grid can be of any format and stored in any type of storage systems. There can be many hundreds of petabytes of data in grids, among which a very large percentage is stored in files. A standard mechanism to describe and organize file-based data is essential for facilitating access to this large amount of data. The Grid File System Working Group (GFS-WG) will provide specifications of Grid File System Directory Services and Architecture of Grid File System Services. The specification of Grid File System Directory Services will describe and manage the namespace of federated and virtualized data from file system resources, access control mechanisms, and meta-data management. The specification of Architecture of Grid File System Services will specify the hierarchical structure to facilitate federation and sharing of virtualized data from file systems in the grid environment by providing the virtual namespace that will allow association of access control mechanisms and meta-data of the underlying physical data sources.
Grid Forum Groups: GridFTP-WG

GridFTP

At this stage, the group should focus on improvements of FTP and GridFTP v1.0 protocol with the goal to produce bulk file transfer protocol suitable for grid applications. New protocol should be backward compatible with RFC959 FTP as much as possible with new features added as (negotiable) extensions. Some desired extensions are:
- Parallel transfers
- GSI authentication
- Striped transfers
Grid Forum Groups: OREP-WG

OGSA Data Replication Services

intended to create, review and refine grid service specifications for data replication services. These specifications will conform to the Grid Services Specification being developed by the OGSI Grid Service Infrastructure Working Group. Initially, the working group will focus on a specification for Replica Location Services, which maintain and provide information about data location. In parallel to the Milestones, the Replication Research Group will discuss the following issues: 1. Agree on breadth and depth of replication services 2. Definition of replica object model; relationship to namespace and data source objects (database objects, files, streams).
Data Format Description Language

XML provides an essential mechanism for transferring data between services in an application and platform neutral format. However it is not well suited to large datasets with repetitive structures, such as large arrays or tables. Furthermore, many legacy systems and valuable data sets exist that do not use the XML format. The aim of this working group is to define an XML-based language, the Data Format Description Language (DFDL), for describing the structure of binary and character encoded (ASCII/Unicode) files and data streams so that their format, structure, and metadata can be exposed. This effort specifically does not aim to create a generic data representation language. Rather, DFDL endeavors to describe existing formats in an actionable manner that makes the data in its current format accessible through generic mechanisms.
Data Format Description Language

... The DFDL description would sit in a (logically) separate file from the data itself. The description would provide a hierarchical description that would structure and semantically label the underlying bits. It would capture: how bits are to be interpreted as parts of low-level data types (ints, floats, strings); how low-level types are assembled into scientifically relevant forms such as arrays; how meaning is assigned to these forms through association with variable names and metadata such as units; and how arrays and the overall structure of the binary file are parameterized based on array dimensions, flags specifying optional file components, etc. Further, if the data file contains highly repetitive structures, such as large arrays or tables, such a description can be very concise.
Grid Forum Groups: Ipv6-WG

IPv6

IPv6 is now emerging as a significant factor in operational networks, and continued scaling up of the Internet (and thus of Grids) will require the additional address space and management features of IPv6. It is therefore important that all GGF specifications work as well (or better) with IPv6 as with IPv4. The purpose of the working group is to identify any GGF specifications that do not meet this requirement, to provide appropriate guidelines for future specifications, and to communicate any issues discovered with IPv6 to the IETF, the Java community, etc.