OGSI

Adapted from

Ian Foster

Mathematics and Computer Science Division

Argonne National Laboratory

and

Department of Computer Science

The University of Chicago

Overview

- OGSI
- Data Intensive Grid Service Models
- Data Access Model

Open Grid Services Infrastructure (OGSI)

- Gives a formal and technical specification of what a grid service is.
- Its a excruciatingly detailed specification of how Grid Services work.
- It is a formal and technical specification of the concepts described in OGSA.
- The Globus Toolkit 3 is an implementation of OGSI.
- Some other implementations are OGSI::Lite (Perl)1 and the UNICORE OGSA demonstrator2 from the EU GRIP project.
- OGSI specification defines grid services and builds upon web services.

Open Grid Services Infrastructure (OGSI)

- OGSI creates an extension model for WSDL called GWSDL (Grid WSDL). The reason is:
  - Interface inheritance
  - Service Data (for expressing state information)
- Components:
  - Lifecycle
  - State management
  - Service Groups
  - Factory
  - Notification
  - Handle Map

Data intensive grid service models

Applications in the grid are normally grouped into two categories

- Computation-intensive and Data intensive
- Data intensive applications deals with massive amounts of data. The grid system must specially designed to discover, transfer and manipulate the massive data sets.
- Transferring the massive data sets is a time consuming task.
- Data access method is also known as caching, which is often applied to enhance data efficiency in a grid environment.
- By replicating the same data block and scattering them in multiple regions in a grid, users can access the same data with locality of references.

Data intensive grid service models

- Replication strategies determine when and where to create a replica of the data.
- The strategies of replications can be classified into ***dynamic and static***

Static method

- The locations and number of replicas are determined in advance and will not be modified.
- Replication operation require little overhead
- Static strategies cannot adapt to changes in demand, bandwidth and storage variability
- Optimization is required to determine the location and number of data replicas.

Dynamic strategies

- Dynamic strategies can adjust locations and number of data replicas according to change in conditions
- Frequent data moving operations can result in much more overhead than in the static strategies
- Optimization may be determined based on whether the data replica is being created, deleted or moved.
- The most common replication include preserving locality, minimizing update costs and maximizing profits .

Grid data Access models

In general there are four access models for organizing a data grid as listed here

1.Monadic method

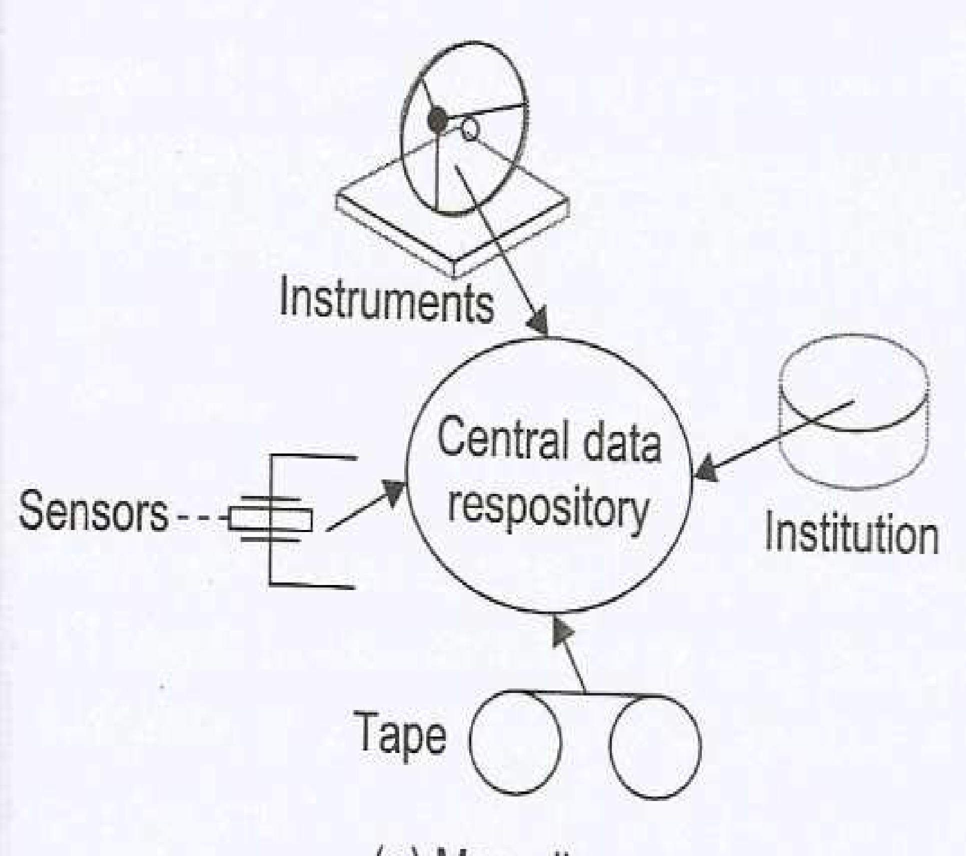
2.Hierarchical model

3.Federation model

4.Hybrid model

Monadic method

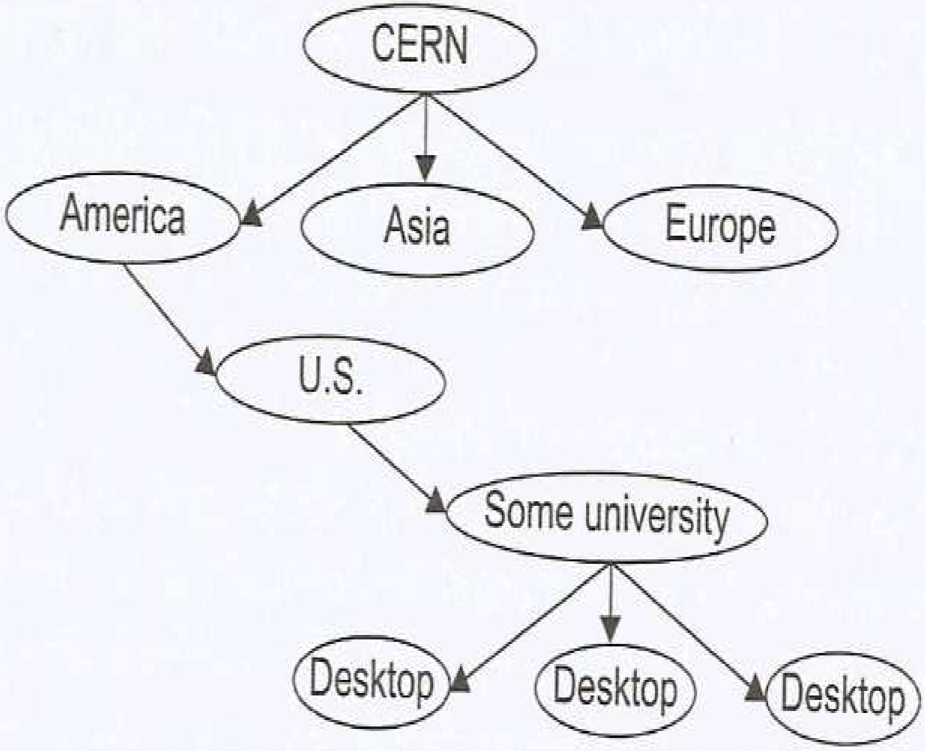
- This is a centralized data repository model. All data is saved in central data repository.
- When users want to access some data they have to submit request directly to the central repository.
- No data is replicated for preserving data locality.
- For a larger grid this model is not efficient in terms of performance and reliability.
- Data replication is permitted in this model only when fault tolerance is demanded.



(a) Monadic

Hierarchical model

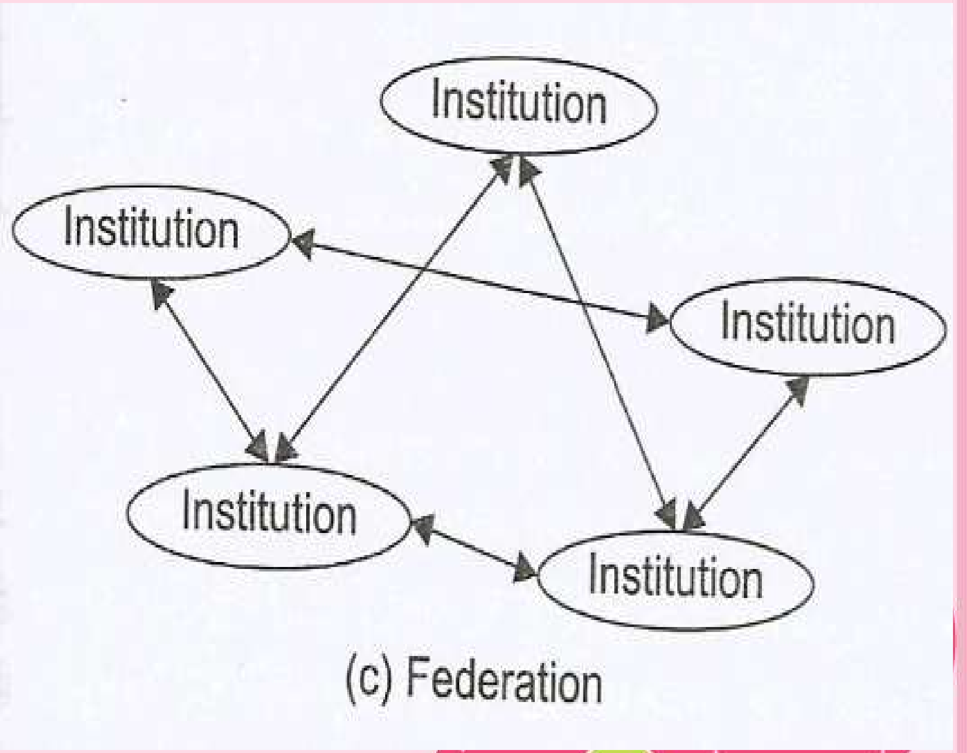
- It is suitable for building a large data grid which has only one large data access directory
- Data may be transferred from the source to a second level center. Then some data in the regional center is transferred to the third level centre.
- After being forwarded several times specific data objects are accessed directly by users. Higher level data center has a wider coverage area.
- PKI security services are easier to implement in this hierarchical data access model



(b) Hierarchy

Federation model

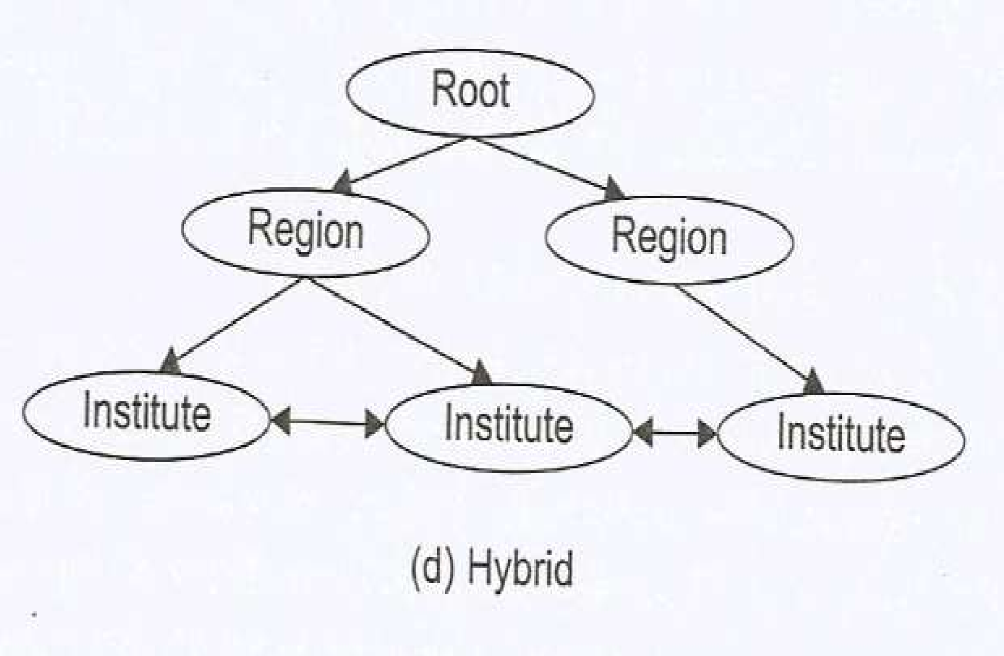
- It is suited for designing a data grid with multiple source of data supplies.
- It is also known as a mesh model
- The data is shared the data and items are owned and controlled by their original owners.
- Only authenticated users are authorized to request data from any data source.
- This mesh model cost the most when the number of grid intuitions becomes very large



(c) Federation

Hybrid model

- This model combines the best features of the hierarchical and mesh models.
- Traditional data transfer technology such as FTP applies for networks with lower bandwidth.
- High bandwidth are exploited by high speed data transfer tools such as GridFTP developed with Globus library.
- The cost of hybrid model can be traded off between the two extreme models of hierarchical and mesh-connected grids.



(d) Hybrid

Parallel versus Striped Data Transfers

- **Parallel data transfer**
  - opens multiple data streams for passing subdivided segments of a file simultaneously.
  - Although the speed of each stream is same as in sequential streaming, the total time to move data in all streams can be significantly reduced compared to FTP transfer.
- **Striped data transfer**
  - data objects is partitioned into a number of sections and each section is placed in an individual site in a data grid.
  - When a user requests this piece of data, a data stream is created for each site in a data gird.
  - All the sections of data objects ate transferred simultaneously.

Summary

- OGSI
- Data Intensive Grid Service Models
- Data Access Model

References

1. Kai Hwang, Geoffery C. Fox and Jack J. Dongarra, “Distributed and Cloud Computing: Clusters, Grids, Clouds and the Future of Internet”,
2. https://www.dcc.fc.up.pt/~ines/aulas/1213/CG/OGSA.ppt
3. http://www.computerworld.com/article/2552339/networking/open-grid-services-architecture.html
4. http://searchsoa.techtarget.com/definition/Open-Grid-Services-Architecture
5. www.cs.umsl.edu/~sanjiv/classes/cs6740/presentation/OGSA.ppt
6. www.nesc.ac.uk/news/.../OpenGridServicesArchitectureApril20021.ppt
7. www.cse.buffalo.edu/~bina/cse486/spring2011/progtutorial_0.4.3.pdf

Thank You

Questions and Comments?