Data Grids

- Address computational and data intensive applications

- Combine huge amounts of data and resources which are geographically distributed

- Provide very high availability and reliability

Issues in Data Grids

- Resource scheduling
- Data access optimization
- Network storage
- High speed data transfers
- Data Management
- Security

Data Grid Services

- In most cases Data Management involves file handling
- Typically Data Grids need to handle various types of data
  - Files
  - File Collections
  - Relational databases
  - XML Databases
  - Virtual Data Sets
- Data must be identified using a mechanism –Grid Data Handle (GDH)

Grid data Handle (GDH)

- Similar to a GSH in OGSA.

- Unique identifier to locate and retrieve data

- GDH is immutable, facilitates location, tracking, transmission etc

Grid Data Reference and Data Registry

- GDH carries no protocol specific or instance specific information.

- The Grid Data Reference holds all protocol and instance specific information

- GDR includes data location, access protocols supported, data lifetime and other meta data.

- Data Registry holds the mapping from GDH to GDR

Data Grid and OGSA

- Factories
  - Increase robustness and availability of services
  - Transactions are made easy by building on functionality of factories
  - Each Component can have a dedicated factory and be instantiated.

Data Grids and OGSA ..

- Registries
  - Where should registry interfaces be deployed?
  - Scalability issues restrict registries to be kept in as few nodes as possible
  - Must be kept in nodes which have high QOS.

Data Grids and OGSA ..

- Service Lifetime Management
  - Keep a set of factories for very long lifetime and keep them alive using OGSA lifetime extension mechanisms
  - Setup factories which create new services on demand. (Applications are responsible for keeping services alive)
  - Redirect incoming applications to existing services based on loads.