Resource and Service Management on the Grid

Chapter 18

K. Czajkowski, I. Foster.
C. Kesselman

Introduction

- Central theme of grid is the ability to discover, allocate and negotiate use of network-accessible capabilities.
- Resource management describes all aspects of the processes: locating a capability, arranging for its use, utilizing it, and monitoring its state.
- Resources refers to both traditional resources and virtualized services.
- We could use "resource management" and "service management" interchangeably.

Issues

- Managed resources span administrative domains.
- Heterogeneous configurations.
- Different access policies
- Establish a mutual agreement between a resource provider and a resource consumer.
- Provider of the resource agrees to supply a capability that can be used to perform some task on behalf of the consumer.

Requirements

- Task submission
- Workload management
- On-demand access
- Co-scheduling
- Resource brokering
- Quality of service
- Provisioning capabilities; Ex: database service can be provisioned to service 10 request at a time; file space can be provisioned to allow submission of jobs requiring up to 200 processors.
- Service level agreements

Resource Management Framework

- Resource management operations: submit, acquire, bind are applicable to any resource.
- See Figure 18.1
- Available resources are discovered
- Capabilities are obtained by acquire operation
- Resources are utilized by associating submitted tasks with acquired resources using bind operations.

Service Level Agreements

- A resource consumer needs to affect the resource behavior, often requiring guarantees concerning the level and type of service being provided by the resource.
- Conversely, provider of resource may want maintain control over how the resource is used.
- A common means of reconciling these two demands is to negotiate a Service Level Agreement (SLA).
- SLA is a means by which resource provider contracts with a client to provide some measurable capability or perform a task.
- Task SLA
- Resource SLA
- Binding SLA
- See 18.2 for a management model.

Policy and Security Resource Agreement

- Resource policy controls by whom and how its resources may be used.
- This policy will govern SLAs to which resource provider is willing to agree.
- Provider will publish only policies necessary to enable discovery and later use other private policies during negotiations.
- Resource brokers are often used to carry out the negotaitions.

Resource Description

- Clients mat request resources by capability, quality, or configuration.
- Resource Description Language (RDF)
  - Parameterized resource attribute metric describing a particular property of the resource. Ex: bandwidth, latency, and space.
  - Operations to work with set of resources
  - Structures (ex: arrays) to represent a collection of resources.
- Guess the format of this language?
- Globus uses another format called RSL (Resource Specification Language).

Resource discovery and selection

- RDF is used for establishing SLAs.
- Description used for purposes of discovery is different from description used for SLAs.
- Resource selection is the process of choosing from a set of candidates provided by resource discovery.
- My opinion: this can be unified.

Task Management

- Resources are selected and action initiated.
- Then task monitoring during execution: monitoring SLAs is specially important.
  - Terminate an SLA
  - Extend an SLA's lifetime
  - Renegotiate an SLA terms
  - Create a new SLA

Grid Resource Management Systems

- Resource management in grid environment needs many improvements.
- Globus has a first generation tool: Grid Resource Allocation Manager (GRAM)
- GRAM defines resource-layer protocols and APIs that enable clients to securely instantiate a computational task.
- GRAM itself does not implement any resource management functionality but instead relies on local resource management interfaces to provide the functions.
- Wrapper on local resource management system.

General-purpose Architecture for Reservation and Allocation (GARA)

- GARA generalizes GRAM to provide for advanced reservations and end-to-end management of quality of service on different types of resources.
- Generalizes GRAM's API and protocol to provide more generic resource management.
- An interesting example, reserve and provide access to network bandwidth. (bandwidth broker, slot manager..)
- This is a evolving area.