Commercial Data Center (CDC)

Summary

Data centers are common in most of the big enterprises in order to consolidate the huge number of servers to reduce the total cost of ownership. Data centers play a key role in the outsourcing business where major businesses outsource their IT resource management to concentrate on their core business competence and excellence. These data centers are required to manage a huge number of IT resources (servers, storages, and networks). Since these data centers are providing resource-sharing capabilities across virtual organization, grid computing forms the technology of choice for their resource management.

In order to support such a commercial grid, the grid technology platform, middleware, and applications should possess a number of core functionalities. We identify and enlist these functionalities by defining the customers of this data center and their usage scenarios.

Customers/Providers (Actors)

- **Grid Administrator.** An administrator wants to get the maximum utilization of the resources in the data center and the management of the resource sharing to be controlled through resource policies.
- **IT System Integrator.** A system integrator wants to reduce the complexity of the distributed and heterogeneous system. Also, they are responsible for the construction of the heterogeneous system and management of service changes.
- **IT Business Activity Manager.** A business manager needs a scalable and reliable platform at a lower cost and an agreed-upon quality of service.

Scenarios

- **Multiple in-house systems support within the enterprise.** Consolidate all the in-house systems in one place and make resources available on an on-demand basis. This reduces the cost of ownership and increases resource utilization. This scenario is suitable for human resource services, customer resource management, finance, and accounting systems.
- **Time-constrained commercial campaign.** Provides the resources on demand in order to run time-constrained campaigns and levy charges on the basis of usage. Examples of these campaigns include sales promotion campaigns, game ticket sales, and so on.
- **Disaster recovery.** An essential part of the major IT systems today. Commercial GRID system could provide standard disaster recovery frameworks across remote CDC at low cost.
- **Global load balancing.** Geographically separated data centers can share high workload and provide scalable systems.

Functional Requirements on OGSA

After a thorough and careful examination of the static and dynamic behavior present in this use case, the following functional requirements of the grid architecture can be identified:

- Discovery of the available resources
- Secure authentication, authorization, and auditing on resource usage
- Resource brokering services to better utilize and use the resources and to achieve the level of quality requirements
- Scalable and manageable data-sharing mechanisms
- Provisioning of resources based on need
- Scheduling of resources for specific tasks
- Advanced reservation facilities to achieve the scale of QoS requirements
- Enable metering and accounting to quantify the resource usage into pricing units
- Enable system capabilities for fault handling and partial failure detection/correction
- Use static and dynamic policies
- Manage transport and message levels and end-to-end security
- Construct dynamic virtual organizations with common functionalities and agreements
- Facilitate resource monitoring
- Enable the facilities for disaster recovery in case of outages

Now let us move on to another use case where we will discuss a scientific research project with geographically distributed participants.

National Fusion Collaboratory (NFC)

Summary

The NFC project defines a virtual organization devoted to fusion research and provides the "codes" developed by this community to the end users (researchers). Earlier, this "code" software was installed in the end user's machine. This became a complex and unmanageable process of software management, distribution, versioning, and upgrade. Due to this change management and configuration problem the fusion community decided to adopt the ASP model, known as "network services model," where the "code" is maintained by the service provider and made accessible to the remote clients. This eliminates the burden on the end user but adds some QoS requirements on the service provider, including executing the "code" as efficiently as possible, executing within a certain time frame, and producing the results with accuracy. As you can imagine, this is the best-case usage model for a computational grid. Now, we can drill down into the usage scenarios of this grid and derive the functional requirements on Grid Computing architecture.

Customers (Actors)

Scientists. They are customers of the fusion code provided by the fusion service provider. Some of the customer requirements are:

- The ability to run the "code" in remote resources on the condition of end-to-end quality of service with a guarantee of time-bound execution.
- Availability of the resource (code execution) in the computational grid.
- A policy-based management of resources; including who can run the code, how many hardware resources are available, etc.
- Ability to use community services by getting accredited with the community rather than an individual service provider. This is a form of "dynamic account" creation and usage.

Scenarios

- A remote client (scientist at an NFC facility) can run code on a remote site within a time frame. The service provider downloads the necessary data and executes a workflow script.
- A monitoring agent starts and watches the submitted job for service-level agreement (SLA) validation. This helps the service provider to provision more resources or recover from failure conditions, etc.
- Integrate with external applications and resources for data and/or code execution and flexible delegation of rights.

Functional Requirements on OGSA

After a thorough and careful examination of the static and dynamic behavior present in this use case, the following functional requirements of the grid architecture can be identified:

- Discovery of available resources
- Workflow management for job distribution across resources
- Scheduling of service tasks
- Enabling the facilities for disaster recovery in case of outages
- Provisioning of resources based on the need
- Resource brokering services to better utilize and use the resources and to achieve the level of quality requirements
- Load balancing to manage workloads
- Network transport management
- Integration with legacy applications and their management
- Handling application and network-level firewalls
- Service-level agreement and agreement-based interaction
- Providing end-to-end security and security authorization and use policies

Next we discuss an online media and entertainment project with some highly interactive content and data sharing among participants. This is an on-demand media and entertainment system, which can be a classic representation of the next generation of on-demand applications.

Online Media and Entertainment

Summary

The entertainment contents may consist of different forms (e.g., movie on demand or online games) with different hosting capacity demands and lifecycle properties. One of the primary goals of this use case is the ability to dynamically manage the resources based on workload demand and current system configuration. Another observation with media entertainment is the change of the content during its lifecycle and changes in the roles of the actors involved.

User involvement and responsiveness with the entertainment content drives this use case into two categories:

- The consumption of the media content, movie on demand, with very limited user interaction
- Frequent user interaction with the content, as we can see in online games.

A number of new commercial consumer experiences will emerge from the economic factors of content subscription, usage-based pricing, content availability, and differentiation among competitors.

Most of online media entertainment (games and video on demand) are designed based on a stovepipe solution for each media entertainment and each solution is managed separately. This will become a cumbersome solution because of the lack of reusability and overprovisioning of the resources. The grid architecture should provide mechanisms for on-demand provisioning, new business models (pricing models), and resource-sharing models.

Actors

1. **A customer who consumes the entertainment content**

- A service provider who hosts the entertainment content
- A publisher who offers the entertainment content
- A developer who consumes the entertainment content

Scenarios

- A consumer, for example a game player, accesses the game portal and authenticates with the game server and starts the game.
- There are several providers that are working in concert to provide the required service for the consumer. For example, the network service provider offers the required bandwidth, the hosting provider provides the server and storage, and the application service provider offers common services like game engine, accounting and billing applications, and help.
- The content provider or media studio provides the content for the customer experience.

Each of the above activities is an interaction between actors.

Functional Requirements on OGSA

After a thorough and careful examination of the static and dynamic behavior present in this use case, the following functional requirements of the grid architecture can be identified:

- Discovery of resources
- Instantiating new service
- Service-level management to meet user expectations
- Enabling metering and accounting to quantify resource usage into pricing units
- Monitoring resource usage and availability
- Managing service policies
- Providing service grouping and aggregation to provide better indexing and information
- Managing end-to-end security
- Servicing lifecycle and change management
- Failure management
- Provisioning management
- Workload management
- Load balancing to provide a scalable system

We can see that the requirements enlisted in each of the use cases are complex. Providing a solution to these complex requirements is a challenging task. We will see in the coming chapter how the OGSA architecture is trying to provide some basic solutions to the above requirements.

Summary

The above use cases introduced some of the core scientific and commercial usage patterns for grid computing. After going through the above representative use cases and the functional requirements exhibited by each of them, we can classify them into four categories:

1. **Basic functions**

- Security functions
- Resource management functions
- System properties

Discussion of the details of these classifications will be covered when we discuss the platform components. Based on the above functional requirements, Open Grid Service Architecture WG started identifying the platform services and the component model definitions for each of the identified services.