Virtual Resource Management (VRM) in Cloud Environment

draft-Junsheng-Cloud-VRM-00

B. Khasnabish, Chu JunSheng, Meng Yu

Friday 28 Jan 2011
Introduction

● Applications/services and basic functions provided by cloud are based on the Virtual Resources which are abstracted from Physical Resources.

● Virtual Resources include V-CPUs, V-Storages, V-Networks etc. V-Networks can be further divided into V-Routers, V-Switches, V-Firewalls, VPNs, V-Interfaces, V-Links which are abstracted from the physical Router/Switch/NIC equipments.

● Virtual Resources are integrated as a whole (Resource Pool) and supplied to upper layers. Virtual Resources can be grouped into hierarchies, allocated and released on demand.

● In Cloud environment, Virtual Resources can be further divided into many Child Resource Pools which own some of the parent’s resources and can, in turn, have a hierarchy of child resource pools to represent successively smaller units of computational capability.

● VRM can allocate resources for applications/services on multiple Virtual Machines which may be deployed on several Physical Servers in the same Cloud, or even different Clouds.
Configuration of VRM

VRM: Virtual Resource Manager
VRP: Virtual Resource Provider
VRC: Virtual Resource Client

Our Scope
Virtual Resource Management

- **Resource Manager**
  - Resource Manager is the registry of virtual resources to resource pools from resource providers and the allocation of virtual resource to clients (consumers), which arises from the need of overcommitment resources—that is, more demand than single host/cluster capacity and from the fact that demand and capacity vary over time.
  - Resource manager allows you to dynamically reallocate resources, so that you can more efficiently use available capacities.

- **Resource Providers**
  - Hosts and clusters are providers of physical resources.
  - Resource pools can be considered as resource providers for they provide resources to child resource pools and virtual machines.

- **Resource Consumers**
  - Virtual machines are resource consumers.
  - Resource pools can be considered as resource consumers for they consume their parents’ resources.
Virtual Resource Description

<Resource Pool>
- Total Capacity (CPU, Mem, etc)
- Reserved Resource Capacity (CPU, Mem, etc)
- Available Resource Capacity (CPU, Mem, etc)
- Virtual Machines (0...n1)
- Virtual Networks (0...n2)
- Child Resource Pools (0...n3)
- Hosts (1...n4)

<Virtual Machine>
- Shares (High: Normal: Low, such as 4:2:1)
- Resource Limitation (CPU, Mem, Storage, etc)
- Resource Reservation (CPU, Mem, etc)
- Number of Processors
- Number of Cores per Processors
- Network Connection (bridge, NAT, host-only, none)
- Host Information

<Virtual Network>
- Shares (High: Normal: Low, such as 4:2:1)
- V-Switch/V-Router
- V-Network Interface/V-Network Link
- V-Firewall
- VPN

<Child Resource Pool>
- Total Capacity (CPU, Mem, etc)
- Resource Limitation (CPU, Mem, Storage, etc)
- Resource Reservation (CPU, Mem, etc)
- Reserved Resource Capacity (CPU, Mem, etc)
- Available Resource Capacity (CPU, Mem, etc)
- Reservation Type (Extensible, fixed)
- Virtual Machines (0...n5)
- Virtual Networks (0...n6)
- Child Resource Pools (0...n7)

<Host>
- Processors
- Number of Cores per Processors
- Total Capacity (CPU, Mem, etc)
- Reserved Resource Capacity (CPU, Mem, etc)
- Available Resource Capacity (CPU, Mem, etc)
View of Hierarchical Resource Organization
VRM provides Virtual Resource Register & Deregister, Virtual Resource Provision function to VRP and VRC.

- When powered on, VRP registers to VRM immediately to deliver resources.
- When powered off, VRP deregisters to VRM to cancel delivered resources.
- When VRC needs resources to perform services, it inquires VRM to provide demand resources.
- VRM checks the required resources from delivered resource pool and responses eligible virtual resources to VRC.
- VRC interacts with VRP to bind and utilize the allocated resources for allocated time.
Features of Virtual Resources in cloud

- Physical equipments are virtualized and grouped into virtual resources.
- Resources are organized hierarchically and have their own unique identifiers.
- Resource can be operated with the generic connector interface.
- Operations on resources have no change to their identifiers.
- Resources registration, requests and bindings can be designed stateless.

Based on the features described above, REST (Representational State Transfer) can be used for virtual resource management.

- Identify hierarchical resource with URIs.
- Operate distributed resources with HTTP requests.
- Represent resource attributes with XML, JSON, XHTML etc.

REST API can be used for registering, managing and applying cloud resources, including compute, storage, and networking components. Detailed resource types include (child) resource pool, virtual machines (backups), virtual network, public address, storage, network interfaces/links, etc.
Virtual Network Resource (VNR)

- Physical Switches/Routers/NICs (Network Interface Cards) can be further abstracts into Virtual Network Resources, such as
  - V-Switch
  - V-Router
  - V-Firewall,
  - VPN
  - V-Network Interface
  - V-Network Link

- Based on business demands, these Virtual Network Resource can be used to provide specific service-levels or meet particular business needs.

- By defining a set of standardized/generalized Virtual Network Resources, Applications/Services can be easily realized, deployed and extended by VRM.
Resource Admission Control (RAC)

- Aggregate computing/storage/network capacity across a collection of servers into logical resource pools.
- Share resources with another cloud, RAC also has the capability to aggregate another cloud resources into local resource pools.
- Create one or a group of Virtual Machine(s), when Applications/ Services are ready to be deployed on the Cloud.
  - RAC will check if there are enough unreserved resources in the shared resource pools can be provided to the Virtual Machine(s).
  - If enough unreserved resources are available, the virtual machine is powered on. Otherwise, an Insufficient Resources warning appears.
Resource Authentication and Authorization (RAA)

- When Application/Services request RAA to allocate (extra) resources, it will check whether they are authorized the capabilities.
- When the Cloud is requested to provide resources to another Cloud, it will authenticate the entity who initiates the inter-cloud resource request. If passed, the required amount of resources will be provided.
- Also, in case of new host or another Cloud providing resources to the Cloud, it will check the security of the resources. If reliable, it will authorize the resources and put them into the resource pools.
Resource Scheduler (RS)

- RS defines rules and policies to decide how resources should be prioritized among virtual machines and intelligently allocates available resources from virtual resource pools among the virtual machines based on the pre-defined rules that reflect Applications/Services requirements and changing priorities.
- Consolidate workloads during off-peak hours and power off hosts to make datacenters more energy efficient.
- Bring powered off hosts online to meet virtual machine requirements either at a pre-defined time or when the utilization of virtual machine increases.
Inter-Cloud Resource Control (ICRC)

- Resources in one cloud can be shared with another cloud in certain circumstances, that is, with ICRC the cloud has
  - the capability to provide resource to another cloud
  - the capability to consume resource from another cloud
- ICRC may provide the following functions
  - Signaling with other cloud
  - Control and monitoring the resources in other cloud
  - Control and provisioning the resources requested by other cloud
  - Transport and security with other cloud
  - Naming, Addressing and Translation if they have different format.
Conclusion

• Grouping virtual resources into hierarchies is the efficient way to manage and schedule virtual resources.

• Virtual Network Resources can be flexibly used to realize virtual network, telecom network virtualization, and VLAN related services.

• Delivery of Virtual Resources to VRM needs to be considered when dynamically construct Virtual Resource Pools.

• Based on pre-defined policy and rules, VRM schedules resources in a efficient way among virtual machines to satisfy business requirements.

• VRM mainly focuses on allocation of Virtual Resources efficiently, securely, flexibly and reliably.
Next Step

- Further develop the requirements
- Define the interfaces and protocols
- Develop the profile for the protocols
- Work on the draft
  - draft-Junsheng-Cloud-VRM-00
Comments/Suggestions
Thanks!