AIST’s inter-cloud framework and cloud-related middleware

Tomohiro Kudoh
National Institute of Advanced Industrial Science and Technology (AIST)
National Institute of Advanced Industrial Science and Technology (AIST)

Ministry of Internal Affairs and Communication
- NICT (Research, Funding)

Ministry of Economy, Trade and Industry
- AIST (Research)
- NEDO (Funding)

Japan Government

Communication

Computer
AIST’s cloud related projects

• GridARS: Inter-cloud resource management framework
• SSS: Fast MapReduce system
• Yabusame: Fast VM Live Migration technology
GridARS

Inter-Cloud
Resource Management Framework
http://www.g-lambda.net/gridars/
Inter-cloud: combination of services provided by multiple cloud providers

Resource management of data centers and Network is a key issue
Required technologies for Inter-Cloud with manageable network

- A unified interface and provisioning system to request network and other resources (computers storages)
- A mechanism to set up network related application execution environment
  - Dynamic assignment of IP addresses, VLAN ids etc.
- A mechanism to provide users with monitoring capability of virtual infrastructures while keeping isolation
  - Monitoring of not physical but virtual infrastructure
AIST GridARS inter-cloud framework

- GridARS constructs a virtual infrastructure over multiple domains based on advance reservation

- GridARS functionalities
  - Discovery Service: collects static resource information and provides it to users and the Resource Management Service
  - Resource Management Service: schedules and co-allocates appropriate resources
  - Provisioning Service: constructs a virtual infrastructure for the resources at a reserved time
    - Set up IP addresses, ssh key sharing and a share file system
  - Monitoring Service: collects resource usage information of each user’s virtual infrastructure and provides it to the user
1. Request reservation

2. Resource Management Service
   Schedules and co-allocates appropriate resources.

3. Provisioning Service
   Constructs a virtual Infrastructure.

4. Monitoring Service
   Provides filtered monitoring information.

0. Discovery Service
   Collects and provides static resource information.

Receive monitoring information
GridARS key features

• Unified resource management/provisioning/monitoring system of network and compute resources
  – OGF NSI for network provisioning

• Dynamic scheduling and automatic set up of IP addresses, VLAN (host and network), ssh and file system

• Supports monitoring of virtual infrastructure
  – Provide requester with monitoring information of the provisioned infrastructure only
  – The monitoring system interoperates with the provisioning system
  – Policy based filtering
    • Provider can define information to be provided to a particular requester
SSS

Fast MapReduce system
http://sss.apgrid.org/
SSS MapReduce

Background

- MapReduce for huge data processing
- Hadoop
  - De fact Standard Open Source Implementation
  - Suitable for huge one-time job
  - Not suitable for iterative tasks
    - Iterations are commonly required for tasks that require convergence
  - Cannot handle complicated workflows

AIST SSS MapReduce

- New MapReduce software developed by AIST
- Separate Map and Reduce phases
- Owner compute
SSS MapReduce - Concept

- KVS based Implementation
  - Suitable for quick iteration
  - Complicated workflows with Mappers and Reducers

- Owner compute
  - Each node process data on the same node
  - Data distribution is done when data are written to KVS

Diagram:
- M/R
- Distributed KVS
- Local KVS
- Distributed KVS
- KV
- KV
- KV
- KV
- KV
- KV
- KV
- KV
- KV
- KV
- KV
- KV
- KV
- KV
SSS MapReduce - Performance

PrefixSpan (Data Mining)  MatMul (Page Rank)

Elapsed Time (sec)

SSS
Hadoop

Elapsed Time (sec)

SSS
Hadoop
Yabusame

Fast VM Live Migration Technology
http://grivon.apgrid.org/
Virtual Machine Live Migration – Yabusame –

- VM Live Migration
  - Migrate VM without stopping processing
  - A key technology for server consolidation
- Problems of existing Live Migration:
  - Takes long time to migrate a VM (depends on memory footprint size and memory update behavior)
  - Hard to predict how long it will take

- AIST Yabusame VM Live Migration technology
  - Fast VM migration using “post-copy”
    - Switch executing host ASAP, before the whole memory image is copied
  - Patch for Qemu/KVM
Yabusame: Postcopy Live Migration

Yabusame copies memory pages after switching the execution host. This technique minimizes the time needed to switch the execution host and also greatly shorten total migration time.

1. Stop Service
   - Host A
   - Host B

2. Copy CPU registers and device status
   - Host A
   - Host B

3. Restart Service
   - Host A
   - Host B

4. Transfer memory pages
   - Host A
   - Host B
## Precopy vs. Postcopy

<table>
<thead>
<tr>
<th></th>
<th>Precopy</th>
<th>Postcopy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before the execution host is switched, memory pages are transferred to the destination.</td>
<td>After the execution host is switched, memory pages are transferred to the destination.</td>
</tr>
</tbody>
</table>
| The time until the execution host is switched | $\text{RAM size} + \alpha$  
$\text{Network speed}$ | $200-300\text{ms}$ |
| The time until all states are removed       | $\text{RAM size} + \alpha$  
$\text{Network speed}$ | $\text{RAM size}$  
$\text{Network speed}$ |

- $\alpha$: depends on memory update speed (non deterministic!)
- Note the above values are the worst case.
  - Qemu skips zero-filled page.
A VM with 1GB RAM is live-migrated to the right PC.

Normal Live Migration

Yabusame Live Migration
(Developed by AIST)
Summary

• AIST has been developing several cloud middleware and framework
• In this talk, GridARS, SSS and Yabusame were introduced
  – All of them are (or will be) open source
• Persons in charge:
  – GridARS: Atsuko Takefusa (atsuko.takefusa at aist.go.jp)
  – SSS: Hidemoto Nakada (hide-nakada at aist.go.jp)
  – Yabusame: Takahiro Hirofuchi (t.hirofuchi at aist.go.jp)
• AIST is looking for international collaboration opportunities.