変える力を、ともに生み出す。
NTT DATAグループ

Live Migration Support

2nd Feb. 2011

NTTDATA Corporation Japan
Kei Masumoto, Muneyuki Noguchi and Masanori Itoh

0. Live migration

- Providing the scheme to migrate running VM instances from a physical machine to others with:

  1. Almost no visible downtime
  2. No transaction loss

➢ Why live migration?

1. Maintenance

ex. Upgrade/installing the patches to hypervisors/BIOS.

ex. One of HDD volumes RAID / one of bonded NICs is out of order.

ex. Regular period maintenance.

2. Distributing high-load

ex. when many VM instances are running on a specific physical machine.

3. Saving power

VM instances are too much scattered, move VM instances to a physical machine!

The following 4 programs are inevitable to achieve live migration.

1. Live migration

```
# nova-manage instance live_migration i-00000001 compute-node2
                                      ec2_id     destination
```

2. Get a VM instances list running on the physical machine

   ✓ euca-describe-instances is available (no need to implement)

3. Get a physical host list (to choose destination)

   ✓ nova-manage service list is available (no need to implement)

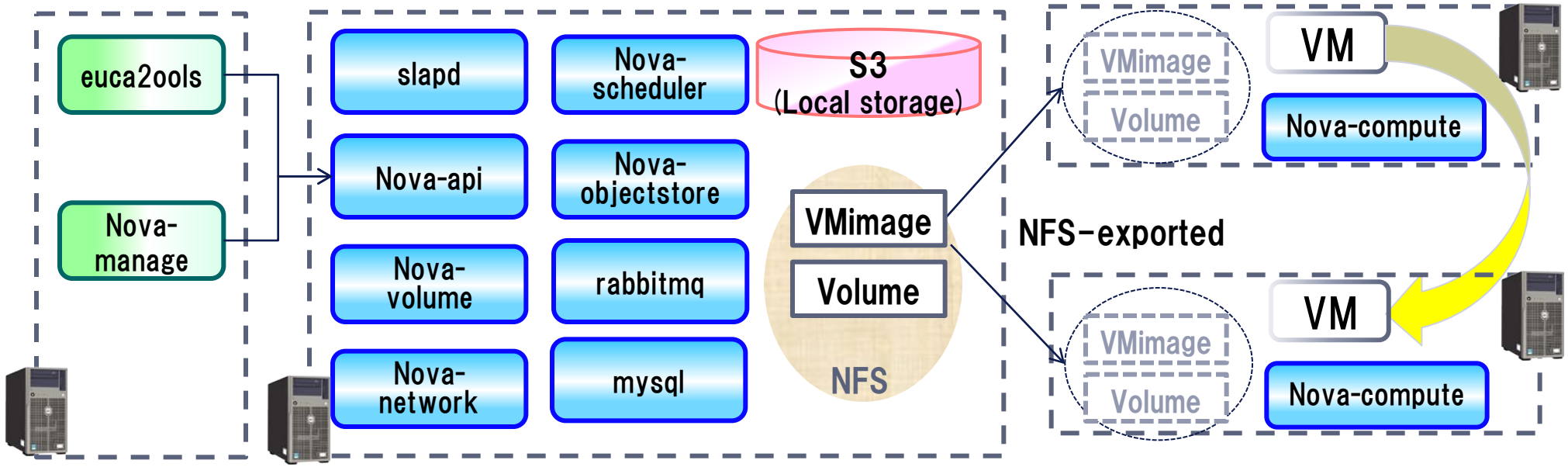
4. Provide info which physical machine still has enough machine

```
# nova-manage service describeresource compute-node2
Compute-node2  total          10     20480   1000
Compute-node2  avail          3      1024    200
Compute-node2  proj1          3      10240   300
Compute-node2  proj2          1      4096    100
node-name      projectname    vcpu   memory  hdd
```

3. Requirements and Assumptions

	Explanation
OS	Ubuntu maverick 10.10 (physical machne) 10.04 (VM)
Hypervisor	KVM (other hypervisors support may be discussed later)
Storages	The directory which instances are running and creating volumes must be a part of shared storage (using NFS)
Networks	Targeting VlanManager/FlatManager/FlatDHCPManager
Network connectivity	Source /destination physical machines must belong to same availability zone (segment) .

Developing environment



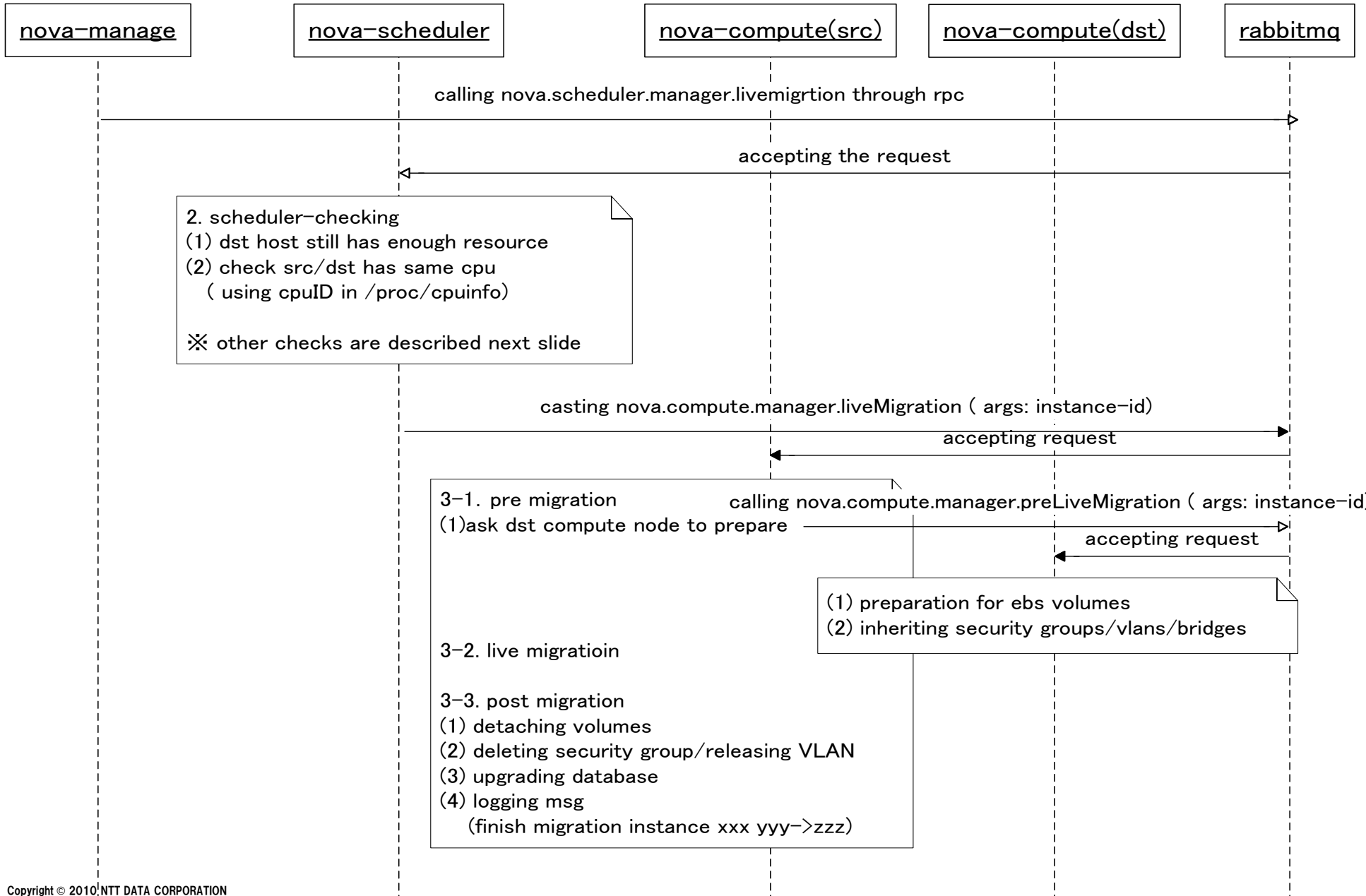
4. Other requirements

変える力を、ともに生み出す。
NTT DATAグループ

requirement	explanation
Authorization	Executed from nova-manage. No authorization.
Deciding to destination machine	Administrator must decide which physical machine a VM instance moves to. (Future consideration may be system automatically decide destination, if it is necessary.)
Floating/Fixed_ip	VM instances must use same fix/floating ips.
EBS volume	VM instances must continue mounting volumes. (targeting ISCSI/AOE only. Other network manager support will be discussed later version)
VLAN	VM instances must belong same VLAN.
Security group	VM instances must use same security group including filtering rules.

requirement	explanation
Test requirements	SSH connection is not terminated after live migration (High-load situation may be considered later)
Performance requirements	No requirements because it depends on storage/network performance.

5. Design consideration (live migration entire flow)



Below points are checked at scheduler, and live-migration is started only when all checkpoint has been confirmed.

No	Explanation
1	Instance is running
2	Source/destination host is alive
3	Source/ destination is not same
4	Destination has enough capabilities ※ memory checking only. Local hdd is not checked since total amount of hdd is not change (source/destination mounted same shared storage)
5	Source/destination hypervisor is same (not like KVM->XenServer)
6	Source/Destination hypervisor version is same.
7	Source/Destination CPU is compatible.
8	Source/Destination mount the same shared storage
9	Nova-volume is running (only when any volumes are attached to the instance)

5. Design consideration

Category	Explanation
Instance state	Instance state should be changed ‘running -> migrating’ before live migration, and ‘migrating-> running’ after live migration.
concurrent request	Scheduler should lock the destination compute service before live migration starts (i.e. Service.disabled == True), since other runinstance request may come between calculating destination resource and starting live migration, then host resource may be full.
EBS Volumes	Before starting live migration, destination host (nova-compute) confirms : ✓ For AOE volume, (1) aoe kernel modules are inserted (if fail, stop to live migratioin) (2) vblade is alive (if fail, logging messages, and continue live migration. Since volumes cannnot be used anymore) ✓ For ISCSI volume: (1) Login to iscsi-server.(if fail, stop to live migration) After live migration, source host (nova-compute) comfirms : ✓ For ISCSI volume: (1) Log out from iscsi server

5. Design consideration (DB Schema change)

変える力を、ともに生み出す。 NTT DATAグループ

✓ To notice how much resources a physical host has to cloud admins (see P3), nova-compute register below information to service table on DB when nova-compute launches.

✓ "*_used" info must be updated periodically. "nova-manage service updateresource" command let compute node update such infomation.

Adding column	Type	Explanation
vcpu	Integer	total number of cpu in a physical machine
memory mb	Integer	total amount of memory in a physical machine
local gb	Integer	total amount of local disk in a physical machine
vcpu_used	Integer	total number of used vcpu in a physical machine
memory_mb_used	Integer	total amount of used memory in a physical machine
local_gb_used	Integer	total amount of used local disk in a physical machine
hypervisor_type	Text	hypervisor type
hypervisor_version	Integer	hypervisor version
cpu_info	Text	json string converted from libvirt.virConnect.gertCapabilities ()

```
# nova-manage service updateresource compute-node2(nodename)
```

✓ CRUD analysis

Adding column	C	R	U	D
vcpu	1. nova-compute launches	1. live-migration starts 2. nova-manage service describeresource	-	-
memory mb	Same as above	Same as above	-	-
local gb	Same as above	Same as above	-	-
vcpu_used	Same as above	Same as above	nova-manage services updateresource	-
memory_mb_used	Same as above	Same as above	nova-manage services updateresource	
local_gb_used	Same as above	Same as above	nova-manage services updateresource	-
hypervisor_type	Same as above	live-migration starts	-	-
hypervisor_version	Same as above	live-migration starts	-	-
cpu_info	Same as above	live-migration starts	-	-

6. The policy for handling exceptions/errors.

- For unrecoverable errors (from Hypervisor point of view)
  1. Logging error messages.
  2. Doing the same way as terminating instances.

     (Delete any records about the instance from DB. If DB records for terminated instances remains for a while, please let me know.)

7. Future Considerations

The below considerations will not be included for Bexar release.
The discussions for future versions may be begun later.

- Other hypervisors support
- Using VPN considerations
- Block Migration（ a live migration that shared storage is not necessary.）
- RBDDriver/SheepDogDriver support.

変える力を、ともに生み出す。

NTT DATAグループ