NEC Network Queuing System V (NQSV)

Migration Guide

Proprietary Notice

Preface

This guide explains how to migrate to NEC Network Queuing System V (NQSV) from NEC Network Queuing System II (NQSII), and major changes up to version R1.02 of NQSV.
It is assumed that NQSV will be introduced into a new system, and then NQSII of existing systems will be migrated to NQSV.

This document consists of the following chapters:
Chapter 1, Migration procedure.
Chapter 2, Difference from NQSII R4.00.
Chapter 3, Difference from NQSII R3.00.

Conventions

The following conventions are used throughout this document.

- Names of keys are printed as they appear on a standard keyboard, **Ctrl**, **Back Space**, and so on.
- Text strings enclosed in brackets are optional. In the following example, options may or may not be included after the command.

  gray [*options*]

Related Documents

Name	Contents
NEC Network Queuing System V (NQSV) User's Guide [Introduction]	Overview of NQSV and configuration of basic system
NEC Network Queuing System V (NQSV) User's Guide [Management]	Management functions of the system
NEC Network Queuing System V (NQSV) User's Guide [Operation]	End-user’s functions guide
NEC Network Queuing System V (NQSV) User's Guide [Reference]	Command reference guide
NEC Network Queuing System V (NQSV) User's Guide [API]	C programming interface (API) to control NQSV
NEC Network Queuing System V (NQSV) User's Guide [JobManipulator]	Administrator’s guide of JobManipulator
NEC Network Queuing System V (NQSV) User's Guide [Accounting & Budget Control]	Accounting function guide

Remarks

This document describes the functions of the following program products.

(1) This manual conforms to Release 1.00 and subsequent releases of the NQSV.

(2) All the functions described in this manual are program products. The typical functions of them conform to the following product names and product series numbers:

Product Name	product series numbers
NEC Network Queuing System V (NQSV) /ResourceManager	UWAF00 UWHAF00 (support pack)
NEC Network Queuing System V (NQSV) /JobServer	UWAG00 UWHAG00 (support pack)
NEC Network Queuing System V (NQSV) /JobManipulator	UWAH00 UWHAH00 (support pack)

(3) UNIX is a registered trademark of The Open Group.

(4) OpenStack is a trademark of OpenStack Foundation in the U.S. and/or other countries.

(5) Red Hat OpenStack Platform is a trademark of Red Hat, Inc. in the U.S. and/or other countries.

(6) Linux is a trademark of Linus Torvalds in the U.S. and/or other countries.

(7) Docker is a trademark of Docker, Inc. in the U.S. and/or other countries.

(8) InfiniBand is a trademark or service mark of InfiniBand Trade Association.

(9) All other product, brand, or trade names used in this publication are the trademarks or registered trademarks of their respective trademark owners.

Definitions and Abbreviations

Term	Definition
VE	Vector Engine The NEC original PCIe card for vector processing based on SX architecture. It is connected to VH.
VH	Vector Host for short. The x86-64 architecture machine that VE connected.
IB	InfiniBand
HCA	Host Channel Adapter The hardware to communicate with other node by using InfiniBand.
MPI	Message Passing Interface MPI is a specification for a standard library for communication.

Contents

Chapter1　Migration procedure

1.1　Migration from NQSII R4.00

1.1.1　Procedure

The outline of the migration procedure is as follows.

(1)　Prepare NQSV/BSV environment.

(2)　Unbind and stop NQSII/JobServer on the execution hosts.

(3)　Uninstall NQSII/JobServer on the execution hosts.

(4)　Install NQSV/JobServer on the execution hosts.

(5)　Start NQSV/JobServer, and bind NQSV/JobServer (execution hosts) to NQSV/BSV.

1.1.2　Note

1.1.2.1 Account Data

- The NQSII account data is not able to migrate to NQSV.
- The saved account data is referred by NQSII R4.00 account commands. Please refer to NQSII User's Guide [Accounting & Budget Control] 3.5. Saving Accounting Data.
- The account data after migration is referred by NQSV.

1.2　Migration from NQSII R3.00

1.2.1　Procedure

The outline of the migration procedure is as follows.

(1)　Prepare NQSV/BSV environment.

(2)　Uninstall and stop NQSII/JSV R3.00 packages on each execution host.

(3)　Update Linux OS of each execution host. (RHEL6->7)

(4)　Install the NQSV/JobServer packages on each execution host.

(5)　Start NQSV/JobServer, and bind NQSV/JobServer (execution hosts) to NQSV/BSV.

1.2.2 Note

- The NQSII account data is not able to migrate to NQSV.
- The saved account data is referred by NQSII R4.00 account commands. Please refer to SCACCT User's Guide [Accounting & Budget Control] 3.5. Saving Accounting Data.
- The account data after migration is referred by NQSV.

Chapter2　Difference from NQSII R4.00

2.1　Topics

2.1.1　New Concept "Logical Host"

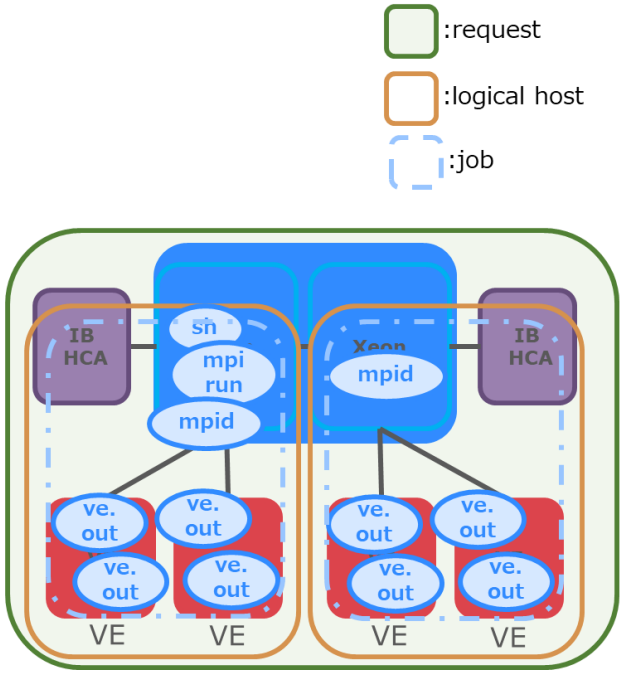
The "logical job" on NQSII changed to the concept　"logical host" on NQSV.

Request , Job and Logical host

- A request is a management unit for user jobs and is managed by NQSV.
- A job is a collection of processes to be executed on a job server and also an execution unit on an execution host.
- A logical host is formed by dividing the resources of an execution host. A single execution host can be made up of multiple logical hosts.

```
#PBS -b 2
#PBS -T necmpi
#PBS -l elapstim_req=300
#PBS --cpunum-lhost=2
#PBS --memsz-lhost=10gb
#PBS --venum-lhost=2
#PBS -use-hca=2

mpirun -nn 2 -np 8 ./ve.out
```



Function

Options to specify resource limit functions were changed from "logical job" to "logical host".

NQSII	NQSV	Description
-l cpunum_job	--cpunum-lhost	Specify the limit on the number of CPUs per logical host.
-l cputim_job	--cputim-lhost	Specify the limit of CPU occupancy time per logical host.
-l gpunum_job	--gpunum-lhost	Specify the limit on the number of GPUs that can be executed simultaneously per logical host.
-l memsz_job	--memsz-lhost	Specify the limit on maximum memory size that can be used per logical host.
-l vmemsz_job	--vmemsz-lhost	Specify the limit on maximum virtual memory size per logical host.

These NQSII options are also available on NQSV.

Manuals

NQSV User's Guide [Management] 4. Queue Management

NQSV User's Guide [Operation] 1.2.9. Resource Limit Options

2.1.2 SX-Aurora TSUBASA architecture support

Resource management specific to VE/HCA

(1) The new options for qsub(1)

Submitting of a request with the number of VEs/HCAs specified is available. The following options added:

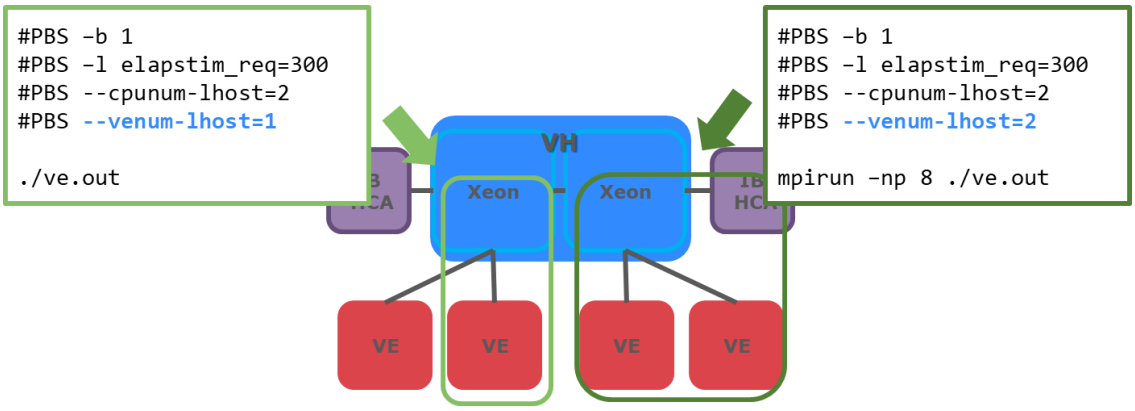
NQSV	target	Description
--venum-lhost	per logical host	Specify the limit on number of VE nodes per logical host.
--venode	per request	Specify the total number of VE nodes to be assigned to

		a request
--use-hca	per VE	Specify the number of HCA ports for each VE in the same device group. This option is effective for a request to use VE nodes.

These NQSV options are only for the environment whose execution host is SX-Aurora TSUBASA system.

--venum-lhost

- --venum-lhost is the option to specify the number of VE nodes per logical host.
- NQSV handles each VE node as a minimum assign unit, and the VE nodes are dedicated to the job on the logical host.
- It is not necessary to specify the memory size or the number of CPU cores for VE node.
- This assignment policy is similar to that for GPUs



--venode

- --venode is the option to specify the total number of VE nodes required for the request.
- If this option is specified, the number of required jobs is automatically calculated based on the default number of incorporated VE nodes specified for the queue.

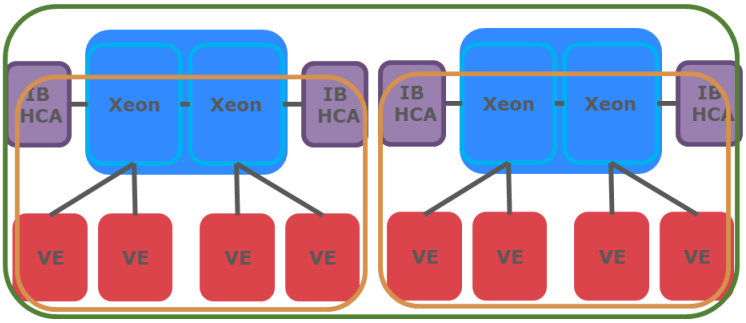
(Please refer to NQSV User's Guide [Management]

13.1 Submitting a request with the total number of VEs specified and Setting of the default number of incorporated VE nodes)

- This option and the option –b (to specify the number of jobs) cannot be specified at the same time.

```
#PSB -l elapstim_req=300
#PBS --cpunum-lhost=2
#PBS --venode=8
#PBS -use-hca=1

mpirun -np 8 ./ve.out
```



--use-hca

- --use-hca is the option to specify the use for direct communication (mode) and the number of HCA port (num).

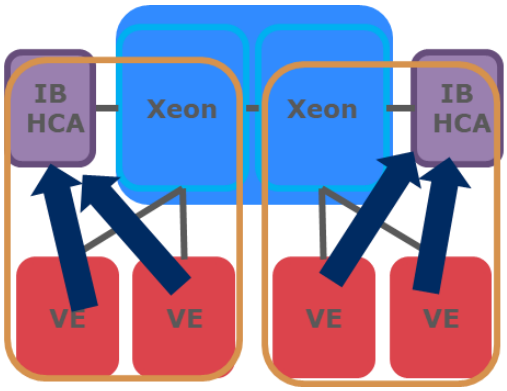
  ex) --use-hca=all:1

- In cooperation with NEC MPI, NQSV assigns the most suitable HCA port to minimize network cost of jobs.
- The device resource configuration file is necessary to use this option.
- Please refer to NQSV User's Guide [JobManipulator]
  5.4.2.2 Device Resource Configuration File

If the number of jobs is more than one in the request using the VEs, please set this option to 1 or larger.

```
#PBS -b 2
#PBS -T necmpi
#PBS --cpunum-lhost=2
#PBS --use-hca=1
#PBS --venum-lhost=2

mpirun -nn 2 -np 8 ./ve.out
```



The new option for qstat(1)

-Je and --venode options were added.

```
$ qstat -Je
```

```
JNO  RequestID       EJID  VEMemory    VECPU JSVNO VectorIsland     UserName Exit
---- --------------- ----- -------- -------- ----- --------------- -------- ----
   0 72.host.example  11366   1.00GB     2.00     10 exechost.exampl user1      -
```

```
$ qstat --venode
VectorIsland VE_No  Cores Memory Status        OS_Status
------------ ----- ------ ------ ------------ ----------
a1sb8_003        0      8    48GB ONLINE        ONLINE
a1sb8_003        1      8    48GB ONLINE        ONLINE
a1sb8_003        2      8    48GB ONLINE        ONLINE
a1sb8_003        3      8    48GB ONLINE        ONLINE
a1sb8_003        4      8    48GB ONLINE        ONLINE
a1sb8_003        5      8    48GB ONLINE        ONLINE
a1sb8_003        6      8    48GB ONLINE        ONLINE
a1sb8_003        7      8    48GB ONLINE        ONLINE
```

Manuals

NQSV User's Guide [Management] 2.6.2. Assignment of VE and HCA

NQSV User's Guide [Management] 13. VE and GPU Support

NQSV User's Guide [Operation] 1.2. Batch Request Submit

NQSV User's Guide [Operation] 5.1.2. Check of Detail Information

NQSV User's Guide [JobManipulator] 5.4 HCA Assignment Feature

Note

To allocate HCA correctly, please locate the "device resource configuration file" (/etc/opt/nec/nqsv/resource.def) on each execution host and configure it.

VE/HCA Failure Management support

(1) Function

NQSV detects VE failure automatically, and continues to operate with optimal use of remaining working VEs. NQSV can remove the execution host from operation when it detect the failure of HCA.

Manual

NQSV User's Guide [Management] 13.4. HCA failure check

NEC MPI support

(1) Function

NQSV supports NEC MPI.

Manual

NQSV User's Guide [Operation] 1.14.1. Run under the NEC MPI Environment

SX-Aurora TSUBASA topology aware scheduling

(1) Function

Assign most suitable VE group and HCA to minimize network cost of application

Manual

NQSV User's Guide [JobManipulator] HCA Assignment Feature

VE accounting function support

(1) Function

The -V option to scacctreq(1) and scacctjob(1) command added to show VE information.

Manual

NQSV User's Guide [Accounting & Budget Control]

2.1.3 sstat(1)

(1) Function

Provide additional detailed execution hosts information.

```
$ sstat -E –f

Execution Host: Host1
CPU Number Ratio = 1.000000
CPU Number Ratio of RSG = {
RSG 0 = 1.000000
}
Memory Size Ratio = 0.000000
Memory Size Ratio of RSG = {
RSG 0 = 0.000000
}
Eco Status = {
Status = EXCLUDED
```

```
State Transition Time = 2017-06-20 10:49:36
Exclude Reason = HW_FAILURE
DC-OFF Times (Day) = 0
DC-OFF Times (ACCUM) = 0
}
Hardware Failure = {
Status = CPUERR
}
Execution Host: Host2
CPU Number Ratio = 1.000000
CPU Number Ratio of RSG = {
RSG 0 = 1.000000
}
Memory Size Ratio = 0.000000
Memory Size Ratio of RSG = {
RSG 0 = 0.000000
}
Eco Status = {
DC-OFF Times (Day) = 0
DC-OFF Times (ACCUM) = 0
}
Hardware Failure = {
Status = EXCLUDED
Exclude Reason = VE_DEGRADATION
VE Degradation = YES
}
```

Manual

NQSV User's Guide [JobManipulator] 4.22 Display the Detail of the Execution Host Information

2.1.4 Maximum Number of Execution Hosts

(1) Function

The total number of execution hosts managed by a batch server increased from 2048 to 10240.

Manual

NQSV User's Guide [Introduction] 1.2. Components of NQSV

NQSV User’s Guide [Operation] 16. Limitations

2.1.5 The Upper limit of DC Power Off Operation

(1) Function

The upper limit of DC Power Off operation is increased from 12 up to 200.

Manual

NQSV User's Guide [JobManipulator] 4.16.2.5 Setting of the DC Power Off Limit

2.2 Changes

2.2.1 Daemon Management

(1) Function

The NQSV daemon management is performed via systemctl (formerly, via init.d)

Manual

NQSV User's Guide [Management] 1. Unit Management

2.2.2 Installation Path

(1) Function

The installation path of each component and configuration directory changed to be compatible with Linux manner.

ex : /usr/sbin/nqsII/nqsd -> /opt/nec/nqsv/sbin/nqs_bsvd

ex : /etc/nqsII -> /etc/opt/nec/nqsv

Directory structure

Directory	Contents
/opt/nec/nqsv/bin	Commands
/opt/nec/nqsv/sbin	Administrator’s commands and daemons
/opt/nec/nqsv/sbin/systemd_prog	Scripts to start/stop NQS services
/opt/nec/nqsv/etc	Configuration files
/var/opt/nec/nqsv	Database, log files
/opt/nec/nqsv/include	Header file

/opt/nec/nqsv/lib64	Shared library
/opt/nec/nqsv/man	Reference manual
/usr/local/lib/systemd/system	Unit definition file

Manual

None.

2.2.3 License Management

(1) Function

The license management method changed.

Manual

NQSV User's Guide [Introduction] 2.2 Installation

NQSV User's Guide [Management] 2.3.11. Getting of License

2.2.4 Scheduling Parameter Configuration Command

(1) Function

The scheduling parameter configuration command changed

ex: set plugin xxx -> set priority xxx

Manual

NQSV User's Guide [JobManipulator] 3.1.3 Scheduling Priority Chapter 6. Command Reference

2.2.5 Functions for SX Series (SUPER-UX)

Batch Server

(1) “ECO” Power saving operation by CPU stop

The following options of qmgr(1M) were deleted.

- set batch_server cpu_eco_mode
- help set batch_server cpu_eco_mode

The following items of qstat(1) were not displayed.

- qstat -Bf

  CPU Eco Mode

- qstat -Eft

  CPU Status

(2) The option of SUPER-UX's multi-node resources

The following subcommand of qmgr(1M) was deleted.

- create node_group type=multinode

(3) NQSII-BSV Agent package

(4) The SUPER-UX's kernel parameter option

The following option of qalter(1) was deleted.

- qalter -K <parameter-name>

(5) Job migration during job execution

The following options of qmgr(1M) were deleted.

- set execution_queue reserve_id
- set execution_queue per_job gpid_number_limit
- set execution_queue standard per_job gpid_number_limit
- set global_queue reserve_id
- set global_queue per_job gpid_number_limit
- set global_queue standard per_job gpid_number_limit
- set execution_queue restart_option
- set global_queue restart_option
- delete execution_queue restart_option
- delete global_queue restart_option
- help set execution_queue reserve_id
- help set execution_queue per_job gpid_number_limit
- help set execution_queue standard per_job gpid_number_limit
- help set global_queue reserve_id
- help set global_queue per_job gpid_number_limit

- help set global_queue standard per_job gpid_number_limit
- help set execution_queue restart_option
- help set global_queue restart_option
- help delete execution_queue restart_option
- help delete global_queue restart_option

(6) MPI/SX support

(7) The execution host information display for SUPER-UX

The following items of qstat(1) were not displayed.

- qstat -Ef

  Reserve ID

  RSG Resource Information

  RSG Average Information

- qstat -Sf

  RSG Number

  RSG Resource Information

  RSG Average Information

JobManipulator

(1) Setting of HW resource for multi-node MPI/SX jobs

The parameter (JID_CONTROL) in the configuration file, which is used to take control of HW resource for multi-node MPI/SX jobs, was deleted.

Setting and display of multimode resource group

Option -G to indicate a multi-node resource group was deleted from sstat(1).

The following option of smgr(1M) was deleted.

- set node_group multimode_resource

Cluster Concentration Assignment for Multi-node MPI/SX Requests

The following option of smgr(1M) was deleted.

- set queue cluster_concentration_assign

SX-specific Preferential Assign Policy of AC Power Share Node

The value ac_power_share cannot be set in the following subcommand of smgr(1M).

- set assign_policy_priority

SX-specific Preferential Assign Policy of IXS-B Column Node for Extended Cluster

The parameter (EXTENDED_CLUSTER) in the configuration file was deleted.

Chapter3　Difference from NQSII R3.00

3.1　Topics

3.1.1　Supported MPI

(1)　Function

MVAPICH2 support

Manual

NQSV User's Guide [Management] 10.3. MVAPICH2 Environment Settings

3.1.2　Group request function support

(1)　Function

Request can be executed with a particular group’s permission specified at job submission.

Manual

NQSV User's Guide [Management] 11. Group of Request

3.1.3　Resource Limit function per Group and User support

(1)　Function

Resource control on a per-group basis and per-user basis.

Manual

NQSV User's Guide [Management] 12. Limit per Group and User

3.1.4　Resource management specific to GPU

A new option for qsub(1)

(1)　Function

“-l gpunum_job” is the option to specify the number of GPU per job.

Manual

NQSV User's Guide [Management] 13. VE and GPU Support

Responsive to the number of available GPUs

(1) Function

In cases of change in the number of available GPUs, such as failure and recovery of GPU, JobManipulator performs scheduling based on the updated number of available GPUs and the requests that have been assigned to the scheduler map will be.

Manual

NQSV User's Guide [JobManipulator] 4.10 Scheduling with the change in the number of CPUs/GPUs

3.1.5 Socket Scheduling function support

(1) Function

When using a NUMA architecture scalar machine (Linux) as execution host, the most suitable resource set (CPUs and memory) is allocated by the socket unit to a job (socket scheduling). It can work together with the CPUSET function of the Linux to enable resource partitioning.

Manual

NQSV User's Guide [Management] 19. Socket Scheduling

3.1.6 Custom Resource Function support

(1) Function

The custom resource function is used to take control the custom resource to be concurrently used in accordance with the defined custom resource information.

Manual

NQSV User's Guide [Management] 18. Custom Resource Function

3.1.7 Advance Reservation (Resource Reservation Section)

(1) Function

Advance Reservation enables a system manager to set the maintenance period in which jobs cannot be executed or a user to surely execute a request by reserving a Resource Reservation Section.

The following function added:

- Reservation accounting
- Creation of reservation section for specified groups
- Creation of reservation section excluding urgent queue
- Health-check and cleanup
- Resource Reservation Section Specifying Template

Manual

NQSV User's Guide [JobManipulator] 4.7 Advance Reservation (Resource Reservation Section)

3.1.8 RunLimit

(1) Function

"Run Limit" is the upper limit of the number of requests that can be executed simultaneously. The following options added:

- Request run limit per users
- Request run limit per groups
- CPU run limit

Manual

NQSV User's Guide [JobManipulator] 2.7.1 Run Limit

3.1.9 Hook Script Function

(1) Function

The hook script function executes a script (called a hook script) defined by an administrator on a batch server host when a request transits to a certain state.

Manual

NQSV User's Guide [Management] 14. Hook Script Function

3.1.10 User’s Pre and Post Script Function

(1) Function

The User’s Pre/Post script function executes a script specified (called a UserPP

script) when submitting a request, before job execution (PRE-RUNNING) or after job execution (POST-RUNNING).

Manual

NQSV User's Guide [Management] 15. User Pre-Post Script Function

3.1.11 Setting Function of the First Stage-in Time

(1) Function

When a request with necessary file staging is assigned around the beginning of the scheduler map, there is a possibility that its scheduled start time is canceled because of delay of the stage-in. So, you can set the estimated time for the first stage-in as First Stage-in Time per scheduler.

Manual

NQSV User's Guide [JobManipulator]

4.10 Scheduling with the change in the number of CPUs/GPUs

3.1.12 Pre-Staging Function

(1) Function

This function which allows to assign requests without file staging is supported. It helps reduce the load of file system in case of simultaneous file staging for many requests at assignment or escalation.

Manual

NQSV User's Guide [JobManipulator] 4.21 Pre-Staging Function

3.1.13 Failure Detection and Power Supply Control support (Linux)

(1) Function

NQSV has two functions to detect failure of the execution host from outside the execution host, and to save power of execution host by power control function.

Manual

NQSV User's Guide [Management] 20. Failure Detection and Power Supply Control

3.1.14 Failover

(1) Function

Batch server, accounting server and JobManipulator can be duplexed, which allows sustained operation of NQSV without down time.

Manual

NQSV User's Guide [Management] 21. Failover

3.1.15 Provisioning environment in conjunction with OpenStack

(1) Function

NQSV can dynamically configure a job execution environment in an execution host in conjunction with OpenStack.

Manual

NQSV User's Guide [Management] 16. Provisioning environment in conjunction with OpenStack

3.1.16 Provisioning environment in conjunction with Docker

(1) Function

NQSV can execute a job on an isolated system (container) within an execution host in conjunction with Docker that can achieve container-based virtualization.

Manual

NQSV User's Guide [Management] 17. Provisioning environment in conjunction with Docker

3.1.17 SCACCT function integrated to NQSV

(2) Function

The accounting and budget control performed by SCACCT is integrated in the accounting server of NQSV. The differences between SCACCT and NQSV are as follows:

(1) Correspondence of each module

SCACCT	NQSV
Top server	Accounting Server
Middle Server	None
Agent	None
Monitor	Accounting Monitor
CUI	AUI

(2) Available account information

	SACCT	NQSV
Request accounting data	Yes	Yes
Job accounting data	Yes	Yes
Process accounting data	Yes	No
Reservation accounting data	No	Yes

(3) Setting unit of the accounting rate

NQSII R3.00 : a node (Agent unit of SCACCT), a queue

NQSII R4.00, NQSV R1.0x : a queue, a template of OpenStack and Docker

The following example shows a command to set the accounting rate to the template at request submission.

```
# subedit add -t template_name:CPU=0.1,MEM=0.234,DEC=0.5,ACT=0.5
rate data (template_name) add(or update) done
```

(4) Priority of budget type

It is possible to set the priority order of budget type (accounting code, user, and group) by SBU_ORDER parameter of the configuration file of the accounting server.

- Configuration file

NQSII R4.00 : /etc/nqsII/asvd.conf

NQSV R1.0x : /etc/opt/nec/nqsv/asvd.conf

```
# cat /etc/opt/nec/nqsv/asvd.conf
#RECV_PORT_FOR_ACCT=6542
```

```
#ALLOW_CLIENTS=
SBU_CHECK=ON
#RECV_PORT_FOR_SBU=4595
SBU_ORDER=AGU
#LOG_FACILITY=LOG_LOCAL0
#ACCT_DIR=/var/opt/nec/nqsv/asv/master
#LOCK_DIR=/var/opt/nec/nqsv/asv/master
```

(5) Estimated Fees

budgetedit(1M-N) can indicate and modify the estimated fee for a request or resource reservation period. In the following example, ESTIMATE indicates the estimated fees.

```
# budgetedit
=============================================================
USER                REMAIN        ESTIMATE         INITIAL
=============================================================
usr1              11223.41           10.00        12245.00
usr2            1395382.88            0.00      1399445.00
usr3             126555.98            0.00       126555.98
=============================================================
GROUP               REMAIN        ESTIMATE         INITIAL
=============================================================
grp5                 0.00            0.00         1111.00
grp4                 0.00            0.00     19874344.00
=============================================================
ACCOUNT             REMAIN        ESTIMATE         INITIAL
=============================================================
acct1                0.00            0.00         1111.00
acct2                0.00            0.00     19874344.00
```

(6) Manual

NQSV User's Guide [Accounting & Budget Control]

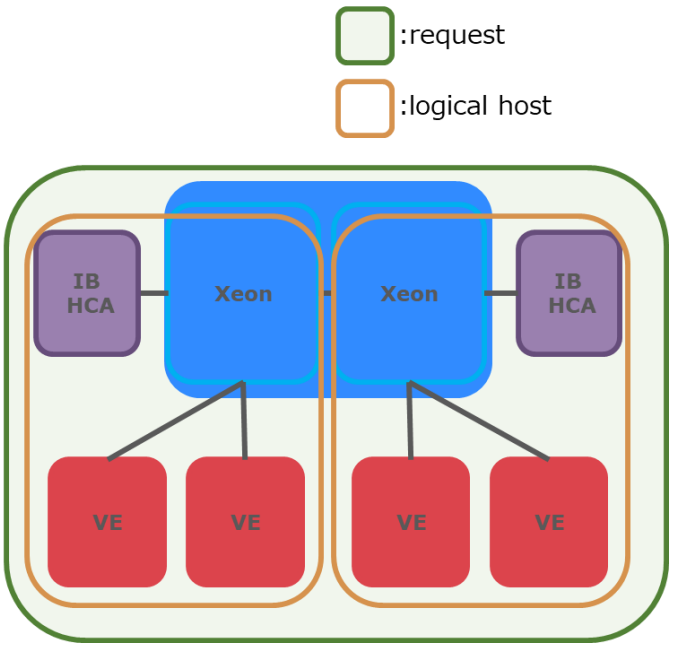
Appendix A How to submit NQSV Request

A.1 Request using VEs

The following example shows the job script of the MPI program with 8 processes, on two logical hosts, two VEs each logical hosts.

```
#PBS -b 2
#PBS -T necmpi
#PBS -l elapstim_req=300
#PBS --cpunum-lhost=2
#PBS --memsz-lhost=10gb
#PBS --use-hca=1
#PBS --venum-lhost=2

mpirun -nn 2 -np 8 ./ve.out
```

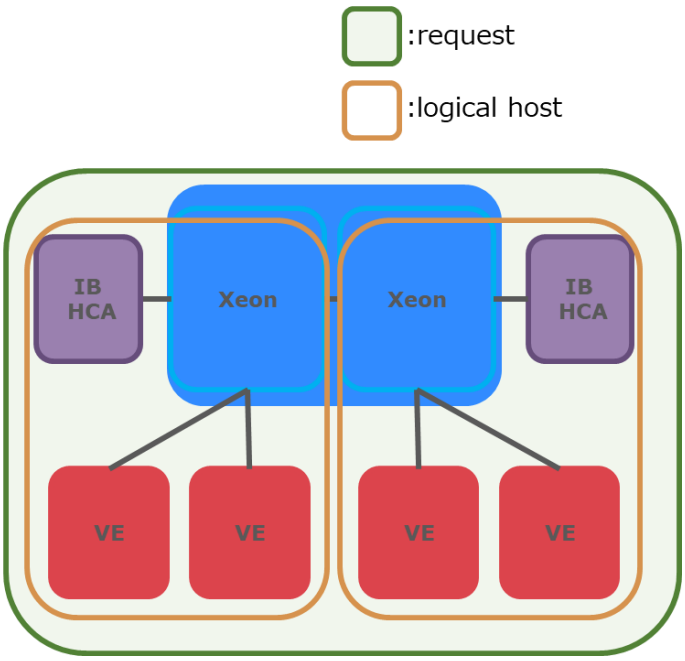


A.2 Request using x86

The following example shows the job script of the OpenMP program which uses only x86 CPUs of VH.

```
#PBS -b 2
#PBS -T necmpi
#PBS -l elapstim_req=300
#PBS --cpunum-lhost=2
#PBS --memsz-lhost=10gb
#PBS --use-hca=1
#PBS --venum-lhost=2

mpirun -nn 2 -np 8 ./ve.out
```

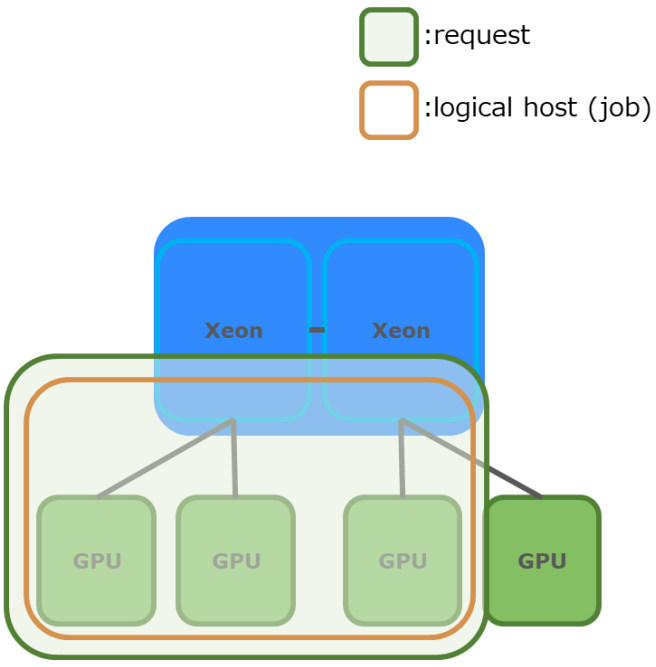


A.3 Request using GPUs

The following example shows the job script of the program using three GPUs.

```
#PBS -b 1
#PBS -l elapstim_req=300
#PBS --cpunum-lhost=8
#PBS --gpunum-lhost=3

./cuda.out
```



A.4 Resource Limit Options

Per Request		
NQSV/NQSII	Description	
-l elapstim_req	the maximum elapsed time	
Per Logical Host		
NQSV	NQSII	Description
--cpunum-lhost	-l cpunum_job	the maximum number of CPUs In case of SX-Aurora TSUBASA, it's the number of CPUs used on VH.
--cputim-lhost	-l cputim_job	the maximum CPU time In case of SX-Aurora TSUBASA, it's the CPU time used on VH.
--gpunum-lhost	-l gpunum_job	the maximum number of GPUs
--memsz-lhost	-l memsz_job	the maximum memory size
--venum-lhost	------	the maximum number of VEs
--vmemsz-lhost	-l vmemsz_job	the maximum virtual memory size SX- In case of SX-Aurora TSUBASA, it's the virtual memory size used on VH.
-l socknum_job	-l socknum_job	the maximum number of sockets
Per Process		
NQSV/NQSII	Description	
-l coresz_prc	the maximum size of core files	
-l cputim_prc	the maximum CPU time In case of SX-Aurora TSUBASA, it's the CPU time used on VH.	
-l datasz_prc	the maximum data size In case of SX-Aurora TSUBASA, it's the data size used on VH.	
-l filenum_prc	the maximum number of open file descriptors	
-l filesz_prc	the maximum file size	
-l memsz_prc	the maximum memory size In case of SX-Aurora TSUBASA, it's the memory size used on VH.	
-l stacksz_prc	the maximum stack size In case of SX-Aurora TSUBASA, it's the stack size used on VH.	
-l vmemsz_prc	the maximum virtual memory size In case of SX-Aurora TSUBASA, it's the virtual memory size used on VH.	

Appendix B Account Item List

Comparative table of NQSII R3.00 (SX-ACE, x86), R4.00 and NQSV.

B.1 Request Account

Name	Description	NQSII R3.00 SX	NQSII R3.00 x86	NQSII R4.00 x64	NQSV R1.0X
REQUEST-ID	Request ID	✓	✓	✓	✓
REQUEST-NAME	Request name	✓	✓	✓	✓
USER NAME	Submission User name	✓	✓	✓	✓
GROUP NAME	Group name	✓	✓	✓	✓
ACCOUNT CODE	Account code	✓	✓	✓	✓
QUEUE NAME	Submission Queue name	✓	✓	✓	✓
QUEUED TIME	Submit time	✓	✓	✓	✓
START TIME	Start time	✓	✓	✓	✓
END TIME	End time	✓	✓	✓	✓
CPU (SECS)	CPU consumption time (system + user) (sec.)	✓	✓	✓	✓
REAL (SECS)	Elapsed time (sec.) (*1)	✓	✓	✓	✓
REQUEST PRTY	Priority of the request	✓	✓	✓	✓
NICE	Nice value	✓	✓	✓	✓
TIME SLICE		✓	✓	✓	
REQELAPS TIME(S)	Elapse Time Limit Value (sec.)	✓	✓	✓	✓
REQCPU TIME(S)	CPU Time Limit Value (sec.)	✓	✓	✓	✓
REQCPU NUM	Number of CPU Limit Value	✓	✓	✓	✓
REQMEM SIZE(K)	Memory Size Limit Value (KB)	✓	✓	✓	✓
REQGPU NUM	Number of GPU Limit Value			✓	✓
IO (BLOCKS) MFF		✓			
IO (BLOCKS) SCD		✓			
IO (BLOCKS) SMT		✓			
FLOPS		✓			
CONCURRENT FLOPS		✓			
H/W CHECK		✓		✓	
EXIT STAT	Exit status	✓	✓	✓	✓
CHARS TRANSFD		✓			
BLOCKS R/W		✓			
KCORE MIN	Total memory consumption (KB * MIN)	✓	✓	✓	✓
MEAN SIZE(K)	Average memory consumption (KB)	✓	✓	✓	✓
MAXMEM SIZE(K)	Max. memory consumption (KB)	✓	✓	✓	✓
INSTRCT (K)		✓			

VECTOR INST(K)		✓			
VECTOR ELMT(K)		✓			
VEC-EXE (SECS)		✓			
MAX PROC		✓			
CPU RESIDENT TM(SECS)		✓			
QUE TYPE	Queue type	✓	✓	✓	✓
NUM PROCS	Number of executed processes	✓			
NODE NUM	Number of execution hosts			✓	✓
JOBS	Number of jobs	✓	✓	✓	✓
SUBREQ	Number of sub requests (only parametric request)	✓	✓	✓	✓
FPEC(K)		✓			
CMCC(SEC)		✓			
BCCC(SEC)		✓			
ICMCC(SEC)		✓			
MNCCC(SEC)		✓			
MT-OPEN COUNTS		✓			
M/S		✓			
RERUN COUNT	Rerun count	✓	✓	✓	✓
PRERUN COUNT	Rerun count of the parent request	✓	✓	✓	✓
MAX NTASK		✓			
TEMPLATE NAME	Template name (*4)			✓	✓
crname(*2)	Custom resource consumption			✓	✓
REQVE NUM	Number of requested VE nodes (*3)				✓
RSVVE NUM	Number of reserved VE nodes (*3)				✓
VE CPU(S)	CPU consumption time on VE nodes [SEC] (*3)				✓
VE KCORE MIN(K)	Total memory consumption on VE nodes [KB * MIN] (*3)				✓
VE MEAN SIZE(K)	Average memory consumption on VE nodes [KB] (*3)				✓
VE MAXMEM SIZE(K)	Max. memory consumption on VE nodes [KB] (*3)				✓

*1 Elapsed time while the request was in the RUNNING state.

*2 Specified custom resource name

*3 These items are available only for the environment where the execution host is SX-Aurora TSUBASA system.

*4 These items are NOT available for the environment where the execution host is SX-Aurora TSUBASA system.

B.2 Job Account

Name	Description	NQSII R3.00 SX	NQSII R3.00 x86	NQSII R4.00 x64	NQSV R1.0X
JOB ID	Job ID	✓	✓	✓	✓
REQUEST-ID	Request ID	✓	✓	✓	✓
REQUEST NAME	Request Name	✓	✓	✓	✓
USER NAME	Submit user name	✓	✓	✓	✓
GROUP NAME	Group name	✓	✓	✓	✓
ACCOUNT CODE	Account code	✓	✓	✓	✓
HOST-NAME	Execution host name	✓	✓	✓	✓
QUEUE NAME	Submit queue name	✓	✓	✓	✓
QUEUED TIME	Submit time	✓	✓	✓	✓
START TIME	Start time	✓	✓	✓	✓
END TIME	End time	✓	✓	✓	✓
CPU (SECS)	CPU consumption time (system + user) (sec.)	✓	✓	✓	✓
REAL (SECS)	Real time of job (sec.)	✓	✓	✓	✓
REQUEST PRTY	Priority of the request	✓	✓	✓	✓
NICE	Nice value	✓	✓	✓	✓
TIME SLICE	Time slice value	✓	✓	✓	
REQELAPS TIME(S)	Elapse Time Limit Value (sec.)	✓	✓	✓	✓
REQCPU TIME(S)	CPU Time Limit Value (sec.)	✓	✓	✓	✓
REQCPU NUM	Number of CPU Limit Value	✓	✓	✓	✓
REQMEM SIZE(K)	Memory Size Limit Value (KB)	✓	✓	✓	✓
IO (BLOCKS) MFF	Number of I/O blocks of MFF	✓			
IO (BLOCKS) SCD	Number of I/O blocks of SCSI disk	✓			
IO (BLOCKS) SMT	Number of I/O blocks of SCSI tape	✓			
FLOPS	FLOPS value	✓			
CONCURRENT FLOPS	Concurrent FLOPS value	✓			
H/W CHECK	Hardware trouble flag(Hexadecimal)	✓		✓	
EXIT STAT	Exit status (*2)	✓	✓	✓	✓
CHARS TRANSFD	Number of transferred characters	✓			
BLOCKS R/W	Number of I/O blocks	✓			
KCORE MIN	Total memory consumption (KB * MIN)	✓	✓	✓	✓
MEAN SIZE(K)	Average memory consumption (KB)	✓	✓	✓	✓
MAXMEM SIZE(K)	Max. memory consumption (KB)	✓	✓	✓	✓
INSTRCT (K)	Number of executed commands	✓			
VECTOR INST(K)	Number of executed vector commands	✓			
VECTOR ELMT(K)	Number of vector elements	✓			
VEC-EXE (SECS)	Elapsed time of executing vector commands(sec.)	✓			
MAX PROC	Max. number of concurrent processes in a job	✓			
CPU RESIDENT TM(SECS)	Processor resident time	✓			
QUE TYPE	Queue type	✓	✓	✓	✓
WAIT TIME(SEC)	Wait time (the time from scheduling start time until actual start time)	✓	✓	✓	✓
NUM PROCS	Number of executed processes	✓			
FPEC(K)	Floating-point data execution element	✓			
CMCC	Operand cache miss time	✓			
BCCC(SEC)	Bank conflict time	✓			

ICMCC(SEC)	Instruction cache miss time	✓			
MNCCC(SEC)	Memory network conflict time	✓			
MT-OPEN COUNTS	MT open count	✓			
M/S	Flag indicating a multitask or not	✓			
MAX NTASK	Max. number of created physical tasks	✓			
REQVE NUM	Number of requested VE nodes for the Job (*1)				✓
RSVVE NUM	Number of reserved VE nodes for the Job (*1)				✓
VE CPU(S)	CPU consumption time on VE nodes [SEC] (*1)				✓
VE KCORE MIN(K)	Total memory consumption on VE nodes [KB * MIN] (*1)				✓
VE MEAN SIZE(K)	Average memory consumption on VE nodes [KB] (*1)				✓
VE MAXMEM SIZE(K)	Max. memory consumption on VE nodes [KB] (*1)				✓
VE REQ NODELIST	List of assigned VE nodes for the Job (*1)				✓
VE USE NODELIST	List of used VE nodes for the Job (*1)				✓
VE RSV NODELIST	List of reserved VE nodes for the Job (*1)				✓
*1 These items are available only for the environment where the execution host is SX-Aurora TSUBASA system.					

B.3 Budget Control

Setting of the Accounting Function

	Setting of SCACCT / Accounting Server	Setting of Batch Server
NQSII R3.00 SX/x86	Set SBU_CHECK parameters of configuration files of each modules in SCACCT.	● Setting for SCACCT server set batch_server **scacct_server** ● Setting of check for over budget set batch_server **budget_check**
NQSII R4.00 x64	Set the parameters in /etc/opt/nec/nqsv/asvd.conf file.	● Added NQSII own accounting server in addition to SCACCT for the budget management server. It can be selected by following command. set batch_server **acct_func = { scacct | nqs_acct}** ● Setting for the server set batch_server **acct_server** ● Setting of check for over budget set batch_server **nqs_budget_chk**

NQSV R1.0X	Set the parameters in /etc/opt/nec/nqsv/asvd.conf file.	Nothing

Accounting Rate

Name	Description	NQSII R3.00 SX	NQSII R3.00 x86	NQSII R4.00 x64	NQSV R1.0X
CPU	Accounting rate per second for CPU consumption time	✓	✓	✓	✓
MEM	Accounting rate per unit memory usage (1KB * min.)	✓	✓	✓	✓
TRNSFR	Accounting rate per 1 KB for number of transferred characters	✓			
IO	Accounting rate per block for the number of I/O blocks (1 block = 4096 bytes)	✓			
INSTRUNCTION	Accounting rate per 1000 executed instructions	✓			
VECTOR	Accounting rate per 1000 executed vector instructions	✓			
VELEMENT	Accounting rate per 1000 vector elements	✓			
PROCESS	Accounting rate per process	✓			
JOB	Accounting rate per job	✓	✓	✓	✓
MTOCNT	Accounting rate per MT opening	✓			
VECCPU	Accounting rate per second for vector instruction execution clock count	✓			
FLOPEC	Accounting rate per floating-point data execution element	✓			
DKIOBLK	Accounting rate per block	✓			

	for the number of normal disk I/O blocks				
ADKIOBLK	Accounting rate per block for the number of array disk I/O blocks	✓			
MFFIOBLK	Accounting rate per block for the number of MFF disk I/O blocks	✓			
MASSDPSIOBLK	Accounting rate per block for the number of master data processing system I/O blocks	✓			
QTIOBLK	Accounting rate per block for the number of 1/4" CGMT I/O blocks	✓			
HCTIOBLK	Accounting rate per block for the number of 1/2" CGMT I/O blocks	✓			
DTIOBLK	Accounting rate per block for the number of DAT I/O blocks	✓			
ETIOBLK	Accounting rate per block for the number of 8mm CGMT I/O blocks	✓			
HTIOBLK	Accounting rate per block for the number of 1/2" MT I/O blocks	✓			
SCSIDKIOBLK	Accounting rate per block for the number of SCSI disk I/O blocks	✓			
SCSIMTIOBLK	Accounting rate per block for the number of SCSI MT I/O blocks	✓			
IMTIOBLK	Accounting rate per block for the number of IMT I/O blocks	✓			
HMTIOBLK	Accounting rate per block for the number of HMT I/O blocks	✓			

GPUNUM	Accounting rate for elapsed time (per GPU * sec.)			✓	✓
ELAPSE	Accounting rate for elapsed time (per job * sec.)			✓	✓
RESERVE	Accounting rate for resource reservation section (per node * sec.)			✓	✓
DEC	Weight for the declared amount of resources			✓	✓
ACT	Weight for the amount of the actually used resources			✓	✓
PRI_MAX	Weight for the maximum priority			✓	✓
PRI_MIN	Weight for the minimum priority			✓	✓
crname (*1)	Accounting rate per the custom resource consumption			✓	✓
REQVE	Accounting rate for the elapsed time of a VE node per second. Requested VE nodes is used for calculation of the accounting. (*2)				✓
RSVVE	Accounting rate for the elapsed time of a VE node per second. Reserved VE nodes is used for calculation of the accounting. (*2)				✓
*1 Specified custom resource name *2 These items are available only for the environment where the execution host is SX-Aurora TSUBASA system.					

Appendix C History

C.1 History table

October. 2019 Rev. 1

C.2 Change notes