NEC Network Queuing System V (NQSV)

Release Notes

R1.16

NEC Corporation

This is the release notes for NQSV R1.16.

1 Outline

The NEC Network Queuing System V (NQSV) is a batch processing system for high-performance cluster system, which enables the maximum utilization of computing resources.
NQSV has job queuing, resource management and job execution functions and supports the most optimal assignment of resources to a job and job execution using them.
NQSV also supports accounting and budget control functions.

2 Product composition

NQSV consists of the following program product.

Product series numbers: UWAF00, UWHAF00 (Support Pack)

Product name	Package name	Package file name and the function contents
NEC Network Queuing System V/ ResourceManager	NQSV/Resource Manager	NQSV-ResourceManager-1.16-[release].x86_64.rpm Batch server function. NQSV request acceptance, execution management, resource management, and multi-cluster management. Node agent function and Accounting server function.
	NQSV/Client	NQSV-Client-1.16-[release].x86_64.rpm A command interface function and user agent.(CUI)
	NQSV/API	NQSV-API-1.16-[release].x86_64.rpm NQSV Application Program Interface.

Product series numbers: UWAG00, UWHAG00 (Support Pack)

Product name	Package name	Package file name and the function contents
NEC Network Queuing System V / JobServer	NQSV/JobServer	NQSV-JobServer-1.16-[release].x86_64.rpm Job server function. Job execution control and resource information collection.

Product series numbers: UWAH00, UWHAH00 (Support Pack)

Product name	Package name	Package file name and the function contents
NEC Network Queuing System V/ JobManipulator	NQSV/JobManipulator	NQSV-JobManipulator-1.16-[release].x86_64.rpm The batch scheduler.

2.1 Install and uninstall the package

About details of install and uninstall the package, please refer to NEC Network Queuing System V (NQSV) User's Guide [Introduction].

2.2 Manual list

The manual of NQSV is composed by following files.

File name	Contents
g2ad01-NQSVUG-Introduction.pdf	NEC Network Queuing System V (NQSV) User's Guide [Introduction] (Japanese)
g2ad01e-NQSVUG-Introduction.pdf	NEC Network Queuing System V (NQSV) User's Guide [Introduction] (English)
g2ad02-NQSVUG-Management.pdf	NEC Network Queuing System V (NQSV) User's Guide [Management] (Japanese)
g2ad02e-NQSVUG-Management.pdf	NEC Network Queuing System V (NQSV) User's Guide [Management] (English)
g2ad03-NQSVUG-Operation.pdf	NEC Network Queuing System V (NQSV) User's Guide [Operation] (Japanese)
g2ad03e-NQSVUG-Operation.pdf	NEC Network Queuing System V (NQSV) User's Guide [Operation] (English)
g2ad04-NQSVUG-Reference.pdf	NEC Network Queuing System V (NQSV) User's Guide [Reference] (Japanese)
g2ad04e-NQSVUG-Reference.pdf	NEC Network Queuing System V (NQSV) User's Guide [Reference] (English)
g2ad05-NQSVUG-API.pdf	NEC Network Queuing System V (NQSV) User's Guide [API] (Japanese)
g2ad05e-NQSVUG-API.pdf	NEC Network Queuing System V (NQSV) User's Guide [API] (English)
g2ad06-NQSVUG-JobManipulator.pdf	NEC Network Queuing System V (NQSV)

	User's Guide [JobManipulator] (Japanese)
g2ad06e-NQSVUG-JobManipulator.pdf	NEC Network Queuing System V (NQSV) User's Guide [JobManipulator] (English)
g2ad07-NQSVUG-Accounting.pdf	NEC Network Queuing System V (NQSV) User's Guide [Accounting & Budget Control] (Japanese)
g2ad07e-NQSVUG-Accounting.pdf	NEC Network Queuing System V (NQSV) User's Guide [Accounting & Budget Control] (English)

3 License

NQSV works in cooperation with license server, it needs license and installation of license management library. About details of license and license management library, please refer to HPC software license management guide.
It is necessary to configure the license server host that batch server connects on the batch server host. About details of it, please refer to NEC Network Queuing System V (NQSV) User's Guide [Introduction].

4 Important notice

None.

5 Remarks

1. You must prepare a file system which has enough free space for the job server database (/var/opt/nec/nqsv/jsv) where various files are created, input/output files for the job, the restart file, and so on.
2. You must prepare a file system which has enough free space for /var/opt/nec/nqsv of the batch server host and accounting server host, because the log files and database is placed in this directory.
3. When updating the NQSV, it is recommended that there be no jobs. In addition, each component of NQSV should be updated at the same time. In this case, update the components running on the batch server host first. For NQSV/JobServer on the execution host, if you cannot update at the same time as the batch server side, there is no problem event if you update sequentially

later.

4. The user-level checkpoint function is an experimental function for running specific applications. If you want to use it, please contact NEC support.
5. The top priority execution of the failure encounter request is a function adjusted specifically for the operation of a specific site. It is not recommended for general operation because it greatly affects the order of request execution and the utilization of the entire system.
6. The setting of process account aggregation "Generate the INDEX information and set the periodic update of the INDEX information on BSV" is removed from NEC Network Queuing System V (NQSV) User's Guide [Accounting & Budget Control]. If this operation executes periodically, the accounting database may be broken. Please remove setting of the periodic update of the INDEX information.
7. Please note the following when updating to R1.08-553 (September 2021 release).
    - JobManipulator has changed the BSV host name referent from the database of the machine ID to the IP address. Please make sure that the host names are unified if your BSV server uses the virtual IP address function and the host name of the virtual IP address is different from the host name in the machine ID database. Please restart JobManipulator after setting up.
8. **Please note the following when updating to R1.09-200 (December 2021 release).**
    - **It is recommended to update with no requests.**
        - **In case you are using UserExit script to switch VE NUMA on each request, please make sure to update without any request.**
    - **If you need to update NQSV while keeping the request, you need to do the following.**
        - **After updating, configure /opt/nec/nqsv/sbin/venuma_chg.sh before restarting operation. For details of the setting method, refer to "NEC Network Queuing System V (NQSV) User's Guide [Management] 12.7 Automatic switching of VE NUMA mode".**
        - **In addition, if the VE partitioning mode was enabled on the system before the update, compile and run the VE NUMA state change program for the requests listed in the appendix**

and change the request information so that VE NUMA mode is enabled.

9. Please note the following when updating to R1.11-360 (September 2022 release).
   - It is recommended to update with no requests.
   - In particular, be sure to terminate requests with a Request ID sequence number of 16777216 or higher before updating.
10. Multiple jobs cannot be executed on a single execution host in the case of OpenMPI jobs.
11. The resource.def file placed on the job server can contain only GPU or VE descriptions. It cannot be listed at the same time.
12. Memory usage of the job includes file cache if "enable_memory_cgroup" is "on" in /etc/opt/nec/nqsd.conf. This covers the following features.
    - "qstat" displays memory usage. Please refer to the manual below to find out which specific values are applicable.

      [Operation] 1.3.1 Check of Basic Information
      1.3.2 Check of Detailed Information
      2.2 State Check of Parametric Request
      3.2 State Check of Interactive Request
      4.1.1 Check of Basic Information
      4.1.2 Check of Detail Information
    - Memory usage of the job account and the request account. KCORE MIN, MEAN SIZE(K) and MAXMEM SIZE(K) are applicable.
    - Budget function. In case charge for actually used memory usage.
13. When suspending a normal request using VEOS's Partial Process Swapping function when executing a NECMPI program with an urgent request in SX-Aurora TSUBASA's VE node, the process manager for NECMPI should be targeted to use Hydra for both urgent and normal requests.

    For more information on setting up NECMPI's process manager, please refer to the following manual.

    [Manager] 9.5 NEC MPI Environment Settings
14. To update from R1.11-360 (released in September 2022) to R1.11-365 (released in October 2022) while in operation, replace the qsub command in the rpm package to /opt/nec/nqsv/bin/qsub. Otherwise, please update by yum command. The replacement procedure is as follows
    - Download NQSV-Client-1.11-365.x86_64.rpm from the repository.

- Convert the rpm package to an archive by rpm2archive or rpm2cpio command.
- Unzip the archive and copy the qsub command in the archive to /opt/nec/nqsv/bin/qsub.

15. To use the PPS performance improvement feature enhanced in R1.14-184 (released in June 2023), VEOS 3.1.1 (released in June 2023) or higher are required. VEOS older than 3.1.1 (released in June 2023) will not enable this feature.

6 Restrictions

None.

7 Compatibility Notes

R1.16-73

The method for referring to host files during execution of the mpirun command in OpenMPI has been changed. In OpenMPI versions 4 or earlier, the host file specified by the option was referenced. However, in OpenMPI version 5, PBS_NODEFILE is referenced regardless of the specified options.

When PBS_NODEFILE is referenced during the execution of mpirun, the number of executable processes is one process per line. In PBS_NODEFILE versions R1.15 or earlier, only one line per execution host name was listed, so only one process could be executed per host. If you want to use multiple processes per host, you need to list the same host name multiple times.

In response to the changes for OpenMPI version 5, the format of PBS_NODEFILE when using OpenMPI will be changed in NQSV R1.16 and later as follows:

- Before change: Each host listed only once per line
  host1
  host2
- After change: Each execution host listed according to the number of execution processes
  host1
  host1
  host2

host2

R1.15-88

None.

R1.15-81

The number of characters in request name output by the scacctreq and scacctjob commands has been increased from a maximum of 15 characters to a maximum of 63 characters. Expansion of the display is performed when the --long-request-name option is specified. There are two compatibility notes with this.

1. In the -3 option of the scacctreq and scacctjob commands, if request name is longer than 15 characters, request name was previously maximum of 15 characters, now changes print * on the 16th character of the request name.
2. In the -R option of the scacctreq and scacctjob commands, the number of characters in request name was previously maximum of 15 characters, but this will be changed to display up to a maximum of 63 characters.

R1.14-196

None.

R1.14-184

None.

R1.13-143

None.

R1.13-130

None.

R1.12-186

None.

R1.11-365

None.

R1.11-360

1. The Dynamic JSV Priority settings have been changed from the JobManipulator config file (nqs_jmd.conf) to the smgr command. The previous settings are no longer in effect. Please reconfigure according to the enhanced interface.

R1.09-200

1. qstat -[R]f doesn't display following items.
   Checkpoint Interval, Restart File Directory
2. qstat -Bf and -Qf, sstat -Sf and -Qf don't display following items.
   Restarting Request, Checkpointing Request
3. The description of "NUMA Node" displayed by qstat -Ef is changed as following.
   - Before change
     NUMA Nodes = {
     Socket 0 (Cpus: 0-3) = Cpu: 4/4 Memory: 15.9GB/15.9GB
     }
   - After change
     NUMA Nodes = {
     Node 0 (Cpus: 0-3) = Cpu: 4/4 Memory: 15.9GB/15.9GB
     }
4. The option -l memsz_prc (physical memory per a process) that does not work on the current Linux kernel is obsoleted from the commands qalter, qlogin, qrsh and qsub. Because the function to limit physical memory per a process was obsoleted on Linux kernel since 2.4.30 and will not revive.
   If the option -l memsz_prc is given to the command qsub, qsub shows following warning.
   "memsz_prc" has no effect. Ignore it.

R1.08-553

None.

R1.08-471

1. The calculation method of the average memory usage of VE and the maximum memory usage of VE displayed by scacctjob -V has been changed as follows.
   - Before change
     Average memory: Σ (VE memory amount * CPU consumption time (user)) / Σ CPU consumption time (user)
     Maximum memory: Maximum value of the total memory of each VE node used

by the job

- After change

  Average memory: Σ VE memory amount / number of sampling times

  Maximum memory: The amount of memory used with the largest value among each VE node that the job is using

2. Previously, when a multi-node request failed PRE-RUNNING and returned to QUEUED, the job account of the job execution was output and billed for the slave jobs that did not have an anomaly. From this version, the job account of the slave job which does not have the abnormality is excluded from the billing target without outputting because the execution has failed as a request.
3. In the VE Process Account function of ASV, the type of the value stored in the SQLite DB after aggregation has been changed. Note that CSV files output from previous versions can be read in this version as well. There is no change in the display contents of the items related to process accounts displayed by scacctjob and scacctreq commands.

8 Appendix

- VE NUMA state change program in the request

```
venuma_attr_chg.c:
/*
VE NUMA-related commands
    Change VE NUMA Mode for a submitted request

Compile:
    gcc -g -Wall -DLINUX -I/opt/nec/nqsv/include -L/opt/nec/nqsv/lib64 -
Wl,--rpath=/opt/nec/nqsv/lib64      venuma_attr_chg.c      -lnqsv      -o
venuma_attr_chg

Usage:
    venuma_attr_chg
        -h <bsv host>
        -r <rid>
        [--venuma { 0 | 1 | -1 }]

    -h <bsv host>
```

```
        Specify the hostname of BSV

    -r <rid>
        Specify the ID of the request for which you want to change the VE
NUMA Mode

    [--venuma { 0 | 1 | -1 }]
        Specify ON/OFF for VE NUMA Mode
        case 0    VE NUMA Mode = OFF
        case 1    VE NUMA Mode = ON
        case -1   VE NUMA Mode isn't set
        case without this option    VE NUMA Mode isn't set

    1. venuma_attr_chg -h host -r rid --venuma 0
    2. venuma_attr_chg -h host -r rid --venuma 1
    3. venuma_attr_chg -h host -r rid  or
       venuma_attr_chg -h host -r rid --venuma -1

How to confirm:
    qstat -f rid | grep "VE NUMA Mode"
    case 1.    VE NUMA Mode = OFF
    case 2.    VE NUMA Mode = ON
    case 3.    Nothing

*/

#include <sys/types.h>
#include <stdio.h>
#include <libgen.h>
#include <stdlib.h>
#include <unistd.h>
#include <string.h>
#include <getopt.h>
#include "nqsv.h"
```

```
void usage(char *argv[])
{
    fprintf(stderr, "¥nUsage: %s -h <bsv host> -r <rid> [--venuma { 0 |
1 | -1 }]¥n¥n", basename(argv[0]));
    exit(1);
}


int main(int argc, char **argv)
{
    nqs_res res;
    nqs_rid rid;
    nqs_alist ad;
    nqs_aid aid;
    char *bsv;
    int c, sd, priv=PRIV_MGR;

    int venuma_mode = -1;
    struct option long_options[] = {
        {"venuma",    required_argument, 0, 'N'},
        {0,           0,                 0,   0}
    };
    int option_index = 0;

    bsv = NULL;
    ad = -1;
    rid.seqno = -1;
    rid.subreq_no = -1;

    while   ((c   =   getopt_long(argc,   argv,   "h:r:",   long_options,
&option_index)) != -1) {
        switch (c) {
            case 'h':
                bsv = optarg;
                break;
            case 'r':
```

```
                rid.seqno = atoi(optarg);
                break;
            case 'N':
                venuma_mode = atoi(optarg);
                break;
            default:
                usage(argv);
        }
    }

    if (!(venuma_mode >= -1 && venuma_mode <= 1) || bsv == NULL ||
rid.seqno < 0) {
        usage(argv);
        exit(1);
    }

    /* Connect to BSV */
    if ((sd = NQSconnect(bsv, 0, priv, &res)) < 0) {
        fprintf(stderr, "NQSconnect: %s¥n", res.msg);
        usage(argv);
        exit(1);
    }

    /* Set ATTR */
    aid.type = ATTR_VENUMA;
    aid.scope = SCPE_REQ;
    rid.mid = NQShname2mid(bsv, &res);
    printf("Now, venuma of %d.%s is %d ¥n", rid.seqno, bsv, venuma_mode);

    if ((ad = NQSalist(ad, &aid, &res)) < 0) {
        fprintf(stderr, "NQSalist: %s¥n", res.msg);
        goto done;

    }

    if (NQSaadd(ad, &aid, &venuma_mode, sizeof(int), &res) < 0) {
```

```
        fprintf(stderr, "NQSaadd: %s¥n", res.msg);
        goto done;
    }

    if ((sd = NQSattrreq(&rid, ad, ATTROP_SET, &res)) < 0) {
        printf("[ERROR] Update_schmesg: NQSattrreq: %s¥n", res.msg);
        goto done;
    }

 done:
    NQSafree(ad, NULL, &res);
    if (NQSdisconnect(&res) < 0) {
        fprintf(stderr, "NQSdisconnect: %s¥n", res.msg);
        usage(argv);
        exit(1);
    }

    exit(0);
}
```

- Compile and run

  The following is an example of enabling VE NUMA mode for a request. The program should be compiled on a host where NQSV-API-1.09-200 is installed. This tool should be run by a user with NQSV MGR privileges.

```
$ cc -L /opt/nec/nqsv/lib64/ -o venuma_attr_chg -lnqsv -I
/opt/nec/nqsv/include/ venuma_attr_chg.c
$ ./venuma_attr_chg -h bsvhost -r rid --venuma 1
$ qstat -f rid | grep "VE NUMA Mode"
    VE NUMA Mode = ON
```

9 Change Log

January 2025 R1.16-73

- OpenMPI 5.0.x is now supported.
- Support for OpenMPI 4.0.x or earlier has ended. However, support for 4.1.x continues.

- Bug fixes
  1. Problem that urgent or special requests may begin executing before swap out operation has not completed when swap out operation of the partial process swapping and termination of the another job on the same execution host occur simultaneously.

October 2024

- Red Hat Enterprise Linux / Rocky Linux 8.10 is now supported.

July 2024 R1.15-88

- Bug fixes

  Problem that JobManipulator may abort in the Dynamic Power-saving Function.

March 2024 R1.15-81

- Enhancements
  1. Previously, the maximum number of characters for the request name output by the scacctreq and scacctjob commands was 15 characters, but this has been expanded to a maximum of 63 characters when the --long-request-name option is specified.
  2. Supported automatic rerun and accounting exclusion function in case of HW failure. By enabling this feature, you can rerun jobs that completed normally on the failed node and exclude them from accounting.
  3. Supported selection of behavior when submit request limit is exceeded during routing requests.
  4. Added a script to the package that needs to be run periodically when the accounting feature is enabled.
- Bug fixes
  1. Problem that nqs_bsvd parent or child process aborts when running qstat -Ef command.
  2. Problem that qstat command not responding when executing qstat -f command on request just before end.

September 2023 R1.14-196

- Red Hat Enterprise Linux 8.8 is now supported.
- Bug fixes
  1. Problem that incorrectly displays the VE process account when a VE request is

rerun.

2. Problem that the qstat command aborts inside the container when run with the -J -u, -T -u, -R -u, -R -q option.

Jun 2023 R1.14-184

- Enhancements

  PPS was enhanced. It is expectable to reduce the time for swapping-out by specifying the appropriate value to the limit on maximum VE memory size per VE node of an urgent request.

- Bug fixes

  1. Problem that when using a GPU that supports multi-instance GPU(MIG), unnecessary messages are repeatedly recorded in messages.
  2. Problem that an error may occur when submitting a hybrid request to the interactive queue.
  3. Problem that if a normal request suspended by an urgent request, the normal request may not be resumed at the scheduled resume time. Furthermore, depending on the timing, JobManipulator may abort.

May 2023 R1.13-143

- Bug fixes

  qsub -v OMP_NUM_THREADS is ignored and the limit value for the number of CPUs in the request is set.

March 2023 R1.13-130

- Enhancements

  1. Supported a function to forcibly stop a job when the size of standard output or error output is exceeded.
  2. Supported a function to limit the total number of submitted VEs when the execution queue is submitted.
  3. Enhanced the function to exclude from dynamic power saving nodes that stall with power on and cannot be restored.

- Bug fixes

  1. Problem that too many "Database is locked." messages may be output to the system log.
  2. Problem that the job submitted by specifying qsub -V option in the job script sometimes failed to execute.

3. Problem that the number of re-runs increased when jobs were moved to other queues by qmove -f.
4. Problem that the request with more than 2 jobs in distributed job type sometimes does not terminate.
5. Problem that REAL value of request account data of distributed job type with more than 2 jobs may be incorrect.
6. Problem that the health check script is executed multiple times.
7. Problem that the queue name and accounting rate displayed by the scacctreq command incorrectly shows the same as before qmove, for requests moved by the qmove command.
8. Problem that the number of jobs in scacctreq command is sometimes incorrect.
9. Problem that a hybrid request is not transferred from the routing queue to the execution queue.
10. Problem that 4th user EXIT rollback is not executed when 4th user EXIT script terminated abnormally in PRERUNNING state.
11. Problem that JobManipulator aborts abnormally in rare cases when suspend by an urgent request interrupt fails.

January 2023 R1.12-186

- Enhancements
  1. Supported multi-instance GPU (MIG).
  2. Improved file output performance of User Agent and added functions.
  3. Improved cgroup memory and process management and added functions.
  4. Supported EXPRESSCLUSTER X 5.0.
  5. Enhanced the function of reducing the non-swappable memory area of MPI processes when using PPS (Partial Process Swapping). To use this function, NEC MPI Version 3.2.0 or later, Version 2.23.0 or later, and VEOS Version 2.14.1 or later must be installed.
- Bug fixes
  1. Problem of disappearing the request whose sequence number of the request ID is greater than 16777215 when the batch server restarts.
  2. Problem of incorrectly reruning the suspendable request.
  3. Problem of request overcommitment or failure to start execution of request after power-saving startup failure due to node change.
  4. Problem when the mode of Realtime Scheduling is set to "always", requests other than urgent requests do not start executing after scheduling of early

execution.

5. Problem of not terminating the user level check point script when timeout occurs.
6. Problem of occurring error on mpirun -npernode of OpenMPI.
7. Problem of aborting a request in post-running state when using UserEXIT.

October 2022 R1.11-365

- Bug fixes
  1. Problem of incorrect output path of standard error file when a file is specified by relative path as a parameter of qsub -e option.

September 2022 R1.11-360

- Enhancements
  1. Enhanced Dynamic JSV Priority function,
  2. Caching of non-schedulable requests function is supported. This improves scheduling performance when requests with the same conditions that cannot be started at a specific start time are submitted consecutively.
  3. Enhanced User Level Checkpoint function.
- Bug fixes
  8. Problem of qsub without -e and -o cannot specify the path containing "%".
  9. Problem when the function to check ratio of memory size and CPUs per a job is enabled, displayed memory size in the error messages is larger than usable maximum size.
  10. Problem when used memory size exceeds warning value of memory size with memory cgroup, SIGTERM is sent to the job.
  11. Problem when the jmm_openmpi processes of the OpenMPI slave jobs start lately than the process of the master jobs, these jobs work abnormally.
      This bug fix applies not only to OpenMPI but also to other MPIs.
  12. Problem JSV aborts when GID of the owner of the job isn't defined on the execution host.
  13. Problem BSV aborts when the request is deleted by qdel or is rerun by qrerun during staging out.

June 2022 R1.10-183

- Enhancements

1. Exit Delay Scheduling function is now supported. When SIGKILL does not work and a job does not terminate due to I/O failure, etc., subsequent requests are rescheduled and the host in question is excluded from the scheduling process.
2. Enhanced user-level checkpointing to allow the use of environment variable PBS_O_PATH in scripts. It also enhances the function to perform qsig -s SIGSTOP_SWOV on its own requests.
3. qstat --planned-start-time option is now supported. The planned start time is displayed.
4. enhanced sstat command acceptance even when the scheduling process takes a long time.

- Bug fixes
  1. Problem where the rerun count of the request account is reset to 0, when the request is moved to the transfer queue by qmove.
  2. Problem when switch-over, normal requests may terminate without rerunning.
  3. Problem with normal request job accounts not being viewed, when urgent or special requests interrupt normal requests using the user-level checkpoint function.
  4. Problem of starting a request even though JSV is unbound.
  5. Problem with UserEXIT process not being included in job's managed process list when enable_memory_cgroup is enabled.
  6. Problem that slave job "orted" may not be started and user program may not be executed, when starting OpenMPI request execution.
  7. Problem where stage-out sometimes takes a long time due to port connection behavior during stage-out.
  8. Problem that BSV sometimes stopped when JSV heartbeat timeout occurred during job execution.

March 2022

1. Red Hat Enterprise Linux 8.5 is now supported.
2. Intel MPI 2021.4 is now supported.
3. Open MPI 4.1.1 is now supported.
4. EXPRESSCLUSTER X 4.3 is now supported.

December 2021 R1.09-200

1. VE NUMA mode auto-switching function is now supported. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's

Guide [Management] 12.7 Automatic switching of VE NUMA mode.

2. The resource management of jobs by cgroup is enhanced (for scalar). For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [Management] 2.3.19 Memory management with memory cgroup.
3. Request assignment mode function of requests is now supported. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [JobManipulator] 4.24 Request Assignment Mode.
4. Dynamic JSV priority function is now supported. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [JobManipulator] 5.7 Dynamic JSV Priority.
5. Non-swappable memory reduction function on using partial process swapping function is now supported (This is a feature that works with NEC MPI and VEOS).
6. Overtaking control function is enhanced. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [JobManipulator] 2.7.7 No Overtaking Control at Pick-up.
7. "Wait time for execution from becoming assignable" is added to the scheduling priority elements. In case of this element the time that isn't subject to scheduling (for example, HELD state of serial type request connection) isn't added to the scheduling priority. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [JobManipulator] 3.1.4 Scheduling Priority.
8. JSV records the information of the received signals to the log.
9. Following bugs are fixed.
   - Job accounts may not be aggregated into request accounts.
   - The accounting database may be broken if the periodic INDEX processing for accounting database executes during the system operation. Therefore "Generate the INDEX information and set the periodic update of the INDEX information on BSV" is removed from NEC Network Queuing System V (NQSV) User's Guide [Accounting & Budget Control].
   - Requests cannot be submitted by qsub with custom resources.
   - If concurrent requests are submitted with --parallel, those requests may not execute correctly or daemon of JobManipulator may terminate abnormally.
   - Daemon of JobManipulator may terminates if the resource reservation area for different type is created.
10. The option -l memsz_prc (physical memory per a process) is obsoleted from each

commands qalter, qlogin, qrsh and qsub.

September 2021 R1.08-553

1. Supports user-level checkpoint function. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [Operation] 17.3 User-level checkpoint.
2. Supports top priority execution of the failure encounter request. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [JobManipulator] 4.11.5 Top Priority Execution of the Failure Encounter Request.
3. Improved that if the job server on the cloud node starts first before the cloud node attaches to BSV in cloud bursting, the job server will LINKUP waiting for the node to attach.
4. Added human readable option (-3) to scacctreq(1)/scacctjob(1).
5. Fixed the following bugs.
   - Problem that requests cannot be transferred within BSV when submit number limit per batch server is reached.
   - Scheduling may stop when running a lot of sstat.
   - Problem that node health check mistakenly recognizes that a failure has occurred even on a normal node.
   - When suspending a VE job, the normal job that is interrupted by the urgent job is not suspended and is rerun.
   - When using cgroups in GPU-CPU Affinity, job fails PRE-RUNNING and cannot be executed.
6. Remove the restriction on R1.08-471, "If a request with NECMPI topology exceeds warning value of some resource limit, the request may terminate even if it handles SIGTERM."

July 2021 R1.08-471

1. Supports node health check function. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [Management] 19.5 Node Health Check function.
2. Supports hydra method as a process manager when running batch requests with NEC MPI. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [Management] 9.5 NECMPI Environment Settings.

3. Supports auto mode as a scheduling method at VE degradation. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [JobManipulator] 5.3 Scheduling in VE Node Problem.
4. Supports function to limit the number of reruns for a request. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [Management] 5.1.6 Rerun Limit function.
5. Enhanced the function to delete stalled jobs due to execution host failure.
6. Fixed the following bugs.
   - Overflow problem in VE process account aggregation and aggregation performance problem.
   - JobManipulator segfaults when submitting a workflow parallel request.
   - The problem of duplicate job accounts being registered due to node failure at the time of job termination.
   - Request can be submitted to the routing queue by specifying the number of VEs as 0.
   - When a failover of EXPRESSCLUSTER occurs, account related files are accumulated and account registration and billing stop.
   - Cloud nodes may not be stopped if JSV does not LINKUP when using cloud-bursting.

May 2021 R1.08

1. Supports the GPU-CPU Affinity function. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [Management] 18.3 GPU-CPU Affinity function.
2. Supports the cloud bursting function, which bursts and executes jobs on cloud computing resources. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [JobManipulator] 4.23 Cloud Bursting Function.
3. Supports job submission and execution functions linked to OSS (Dask-Jobqueue, etc.) and supports batch job connections. For more information about this function, please refer to NEC Network Queuing System V (NQSV) User's Guide [Management] 21 Using OSS for Batch Job Collaboration.

March 2021 R1.07

1 Platform MPI support

2 Added execution host information when specifying the -f option of the qstat

command.

3 Red Hat Enterprise Linux 8.3 support

4 The following descriptions have been removed from the manuals:
- Multi-cluster
- Light weight Batch Server

5 Fixed the following bugs.
- Requests disappear when updating batch servers
- Zombie processes are swapped out when suspend interrupting VE jobs with partial process swapping.
- qstat -S option shows incorrect host average information when using socket scheduling.

Jan 2021 R1.07

1 Improved the packages so that scripts that were updated before the package update are saved.

2 In the SUSPEND/RESUME of VE jobs by the Partial Process Swapping function, it has been improved so that an error does not occur in switch over processing when a scalar job that does not use VE is submitted with VE specified.

3 Fixed the following bugs.
- The information of the execution host with the job server number of 2048 or more is not displayed by qstat -F ehost.
- After the VE is degraded, if the number of VEs is restored by restarting the execution host and the job server is UNBIND at the same time, even if the job server is BIND again, it will be scheduled according to the amount of resources at the time of degradation.

Dec 2020 R1.07

1 The maximum value of the request ID sequence number has been expanded to 99999999. Please refer [Management] 2.3.16 Setting the Maximum Sequence Number of Request ID for details.

2 Enhancement of scheduling function
- When executing an urgent request, the assign prohibition function for the execution host that has an unexecuted urgent request can now be enabled or disabled. Please refer [JobManipulator] 4.1.1 Block of Assignment by Urgent Request for details.
- The past usage record of VE nodes can be accumulated, and the actual value

can be reflected in the scheduling priority. Please refer [JobManipulator] 3.1.3 Usage Data Collection and Adjustment and 3.1.4 Scheduling Priority for details.

3 Added support for accounting virtual memory usage for jobs. Please refer [Accounting & Budget Control] Chapter 3. Accounting for details.

4 When using the Socket Scheduling function, all the installed memory of the execution host including the memory of other sockets can be used.

5 The exit code in qwait when the resource limitations are exceeded have been enhanced. Please refer [Reference］ 1.18 qwait(1) for details.

6 Supports job execution by singularity container. Please refer [Operation ］ 17 Submitting a request using a provisioning environment in conjunction with singularity for details.

Aug 2020 R1.06

1 Added support for using the Request Immediate Execution feature and the pickup overtaking control function.

2 Enhancement of VE process account function
Added support for aggregation and display of VE process accounts even for parametric requests. Moreover, the display error of MFLOPS and Vectorization ratio was corrected.

Jul 2020 R1.06

1 Support for aggregation and display of VE process accounts
It supports the function to aggregate and display the process account information output by VEOS.
Please refer [Accounting & Budget Control] 3.6 VE Process Accounting for instructions on how to set this up.

2 Support SUSPEND/RESUME of VE requests by interrupting urgent requests
Support for suspending VE jobs when an urgent request is executed and resuming it after the urgent job ends.
Please refer [JobManipulator] 5.6 Suspend Jobs Using VEs for details.

3 Support limits the number of VEs that can be executed simultaneously
It supports the function for limiting the number of VEs per user and per group.
Please refer [JobManipulator］ 2.7.1 Run Limit for details.

4 Enhanced billing amount change function

5 Various MPI Support Updates
For details of the supported version, please refer [Management] Chapter 10 MPI Request Execution Environment Settings.

6 Support user-specific execution
Support the function to submit requests to specific other user. Please refer [Operation] 1.2.25 User Specifying and [Reference] qsub section for details.

7 Removing restrictions on Redhat EL 8.1 support
Support various MPI and Docker on RHEL 8.1.

Jan 2020 R1.05

1 Resource limits enhancement
Support the function for resource limitation of VE CPU and Memory. Please refer [Management] 4.1.2 Batch Queue Configuration (1) Resource Limit and [Operation] 1.2.9 Resource Limit Options for details.

2 Supporting VE on Docker environment
Support the function to use VE and GPU on the provisioning function with Docker environment. Please refer [Management] 17. Provisioning environment in conjunction with Docker for details.

3 Suspend function of VE job
Support the function to execute the urgent job which uses VEs by using the Partial Process Swapping function. Please refer [JobManipulator] 5.6 Suspend Jobs Using VEs for details.

4 Performance improvement
Support the functions to improve the request processing performance of batch server. Please refer [Management] 2.3.14 DB updating without sync and 4.1.2. Batch Queue Configuration (14) Disabling the stage-out for details.

5 Accounting function of HW failure
Support the function to record the job server link down while job executing to request account and job account file. Please refer [Accounting & Budget Control] 3.1 Referencing Request Accounting Data and 3.2. Referencing Job Accounting Data for details.

6 qcat improvement
Support the function to display the appended data for stdout/error on qcat command. Please refer [Operation] 1.15 Display Output file in Request Running for details.

7 Red Hat Enterprise Linux 7.6/7.7 support
Red Hat Enterprise Linux 7.6/7.7 is added to supporting OS.

Oct 2019 R1.04

1 VE/Scalar hybrid execution function

The function to execute MPI programs between different resources (VE+CPU, GPU+CPU, etc.). Please refer [Operation] 16. Hybrid Request function for details.

2 Network topology enhancement

The network topology function has been enhanced to support the scheduling function considering the multi-layer network switch and the minimum network topology node group selection function. Please refer [JobManipulator] 3.1.7 Assign Policy and 4.23 Node group selection function for minimum network topology for details.

3 Resource limits enhancement

Custom Resource function has been enhanced to support the resource limit function and billing function by custom resource consumption (actual value). Please refer [Management] 18. Custom Resource Function for details.

4 VE concentrated assignment function

Supports the VE concentrated assignment function that assigns jobs until the available number of VEs in VH host. Please refer [JobManipulator] 5.5 VE concentrated assignment for details.

Aug 2018 R1.02

1 HCA failure check

Failure detection feature of HCA which installed on the SX-Aurora TSUBASA system. Please refer [Management] 13.4. HCA failure check for details.

2 Light Weight Batch Server

The batch server to execute a large amount of small-scale job efficiently. Please refer [Management] 2.11. Light Weight Batch Server for details.

3 Exclusive execution request

The function to execute the request exclusively on the execution host. Please refer [Management] 4.1.2. Batch Queue Configuration (13) Allowance for exclusive execution request and [Operation] 1.2.23. Exclusive execution for details.

4 Red Hat Enterprise Linux 7.5 support

Red Hat Enterprise Linux 7.5 is added to supporting OS.

May 2018 R1.01

1 8VE installed model support

Support the 8VE installed model of SX-Aurora TSUBASA system for the execution host.

2 HCA assigning function

Support the scheduling feature for the HCA which installed on the SX-Aurora TSUBASA system.

Feb 2018 R1.00

Initial release.