

Windows HPC Server 2008

User's View

Christian Terboven <terboven@rz.rwth-aachen.de>

19.03.2012 / Aachen, Germany

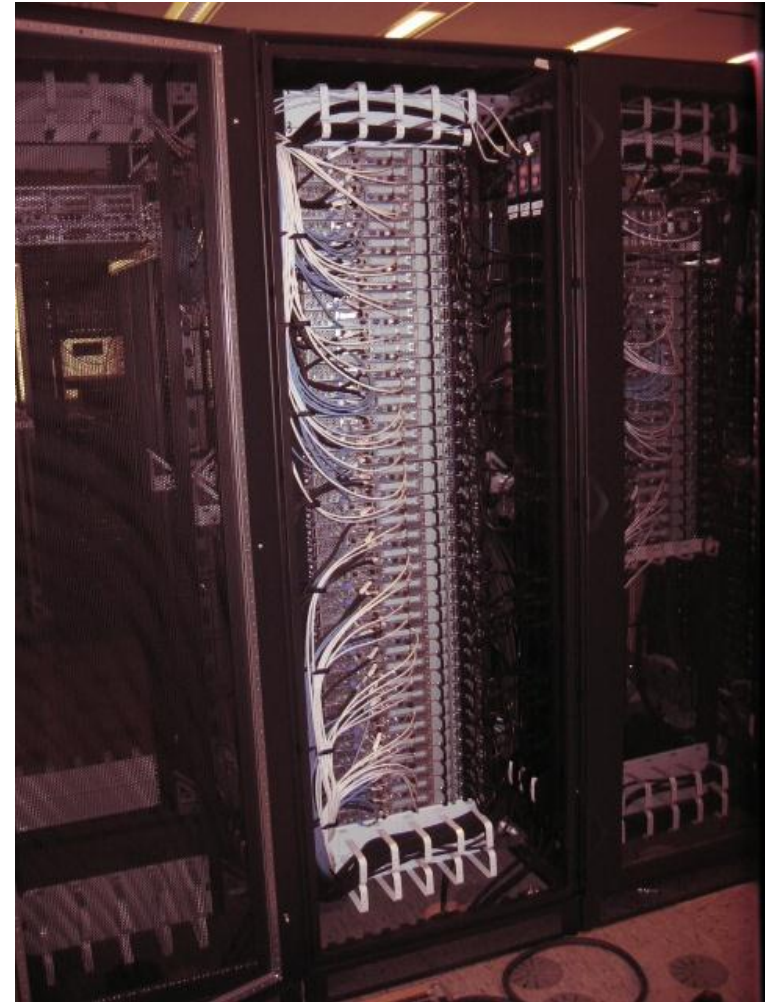
Stand: 16.03.2012

Version 2.3

- ▶ **How does a Windows-HPC-Cluster look like?**
- ▶ **Batch System**
- ▶ **Software Environment**

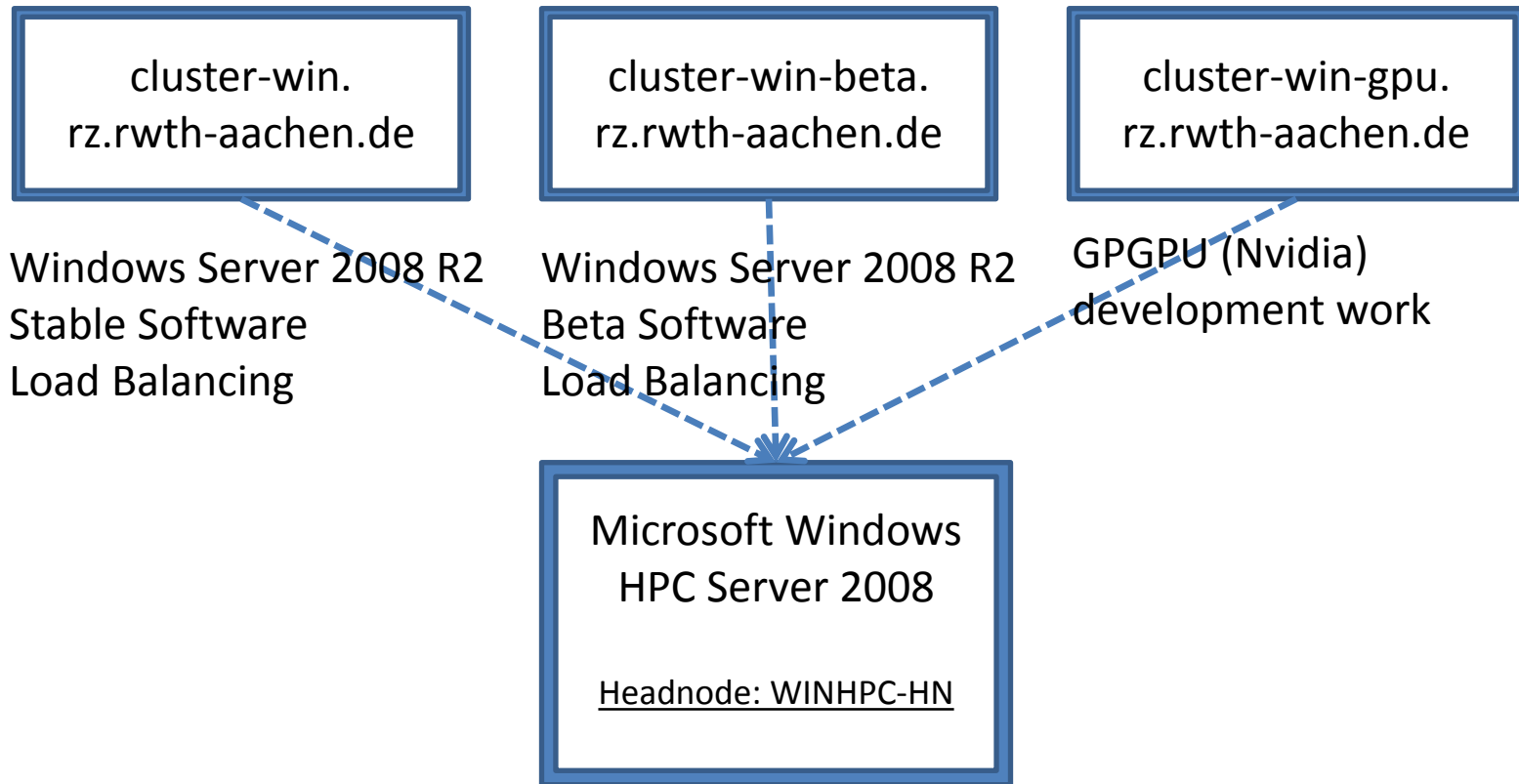
How does a Windows-HPC-Cluster look like?

- ▶ **Cluster installed in Q1/2008:**
 - ▶ Fujitsu-Siemens
Primergy RS 200 S4 servers
 - ▶ 2x Intel Xeon 5450
(quad-core, 3.0 GHz)
 - ▶ 16 / 32 GB memory per node
 - ▶ 4x DDR InfiniBand:
MPI latency: 4.5 us
MPI bandwidth: 1250 MB/s
- ▶ **Installation-on-demand: Lin + Win**
- ▶ **Rank 100 in Top500 in 06/2008!**
 - ▶ 18.81 TFlop/s with 256 nodes
 - ▶ 195 Mflop/s per Watt



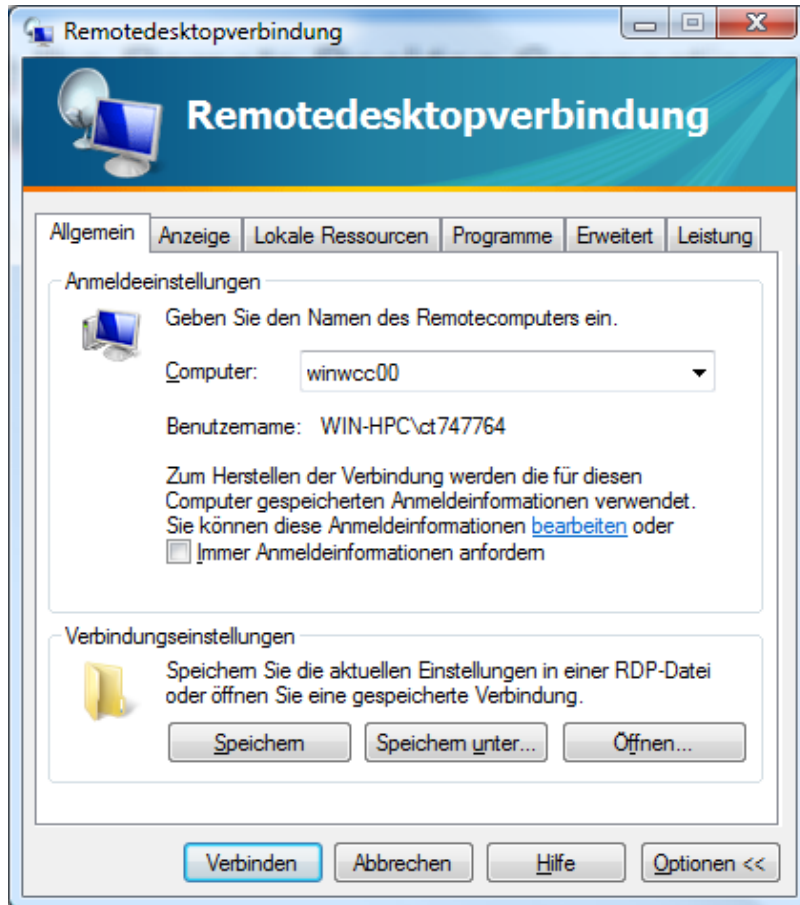
▶ Currently we are offering three Frontends for the Cluster

- ▶ Interactive Use for Software Development and the like

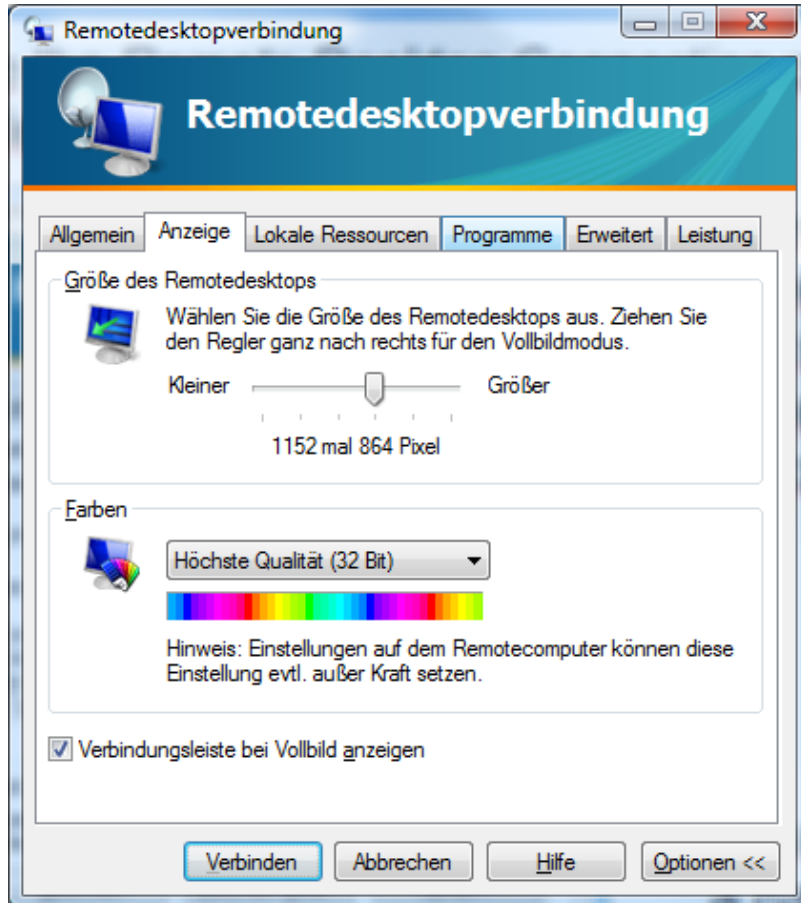


-----> Batch Job (long running Compute Job) Submission

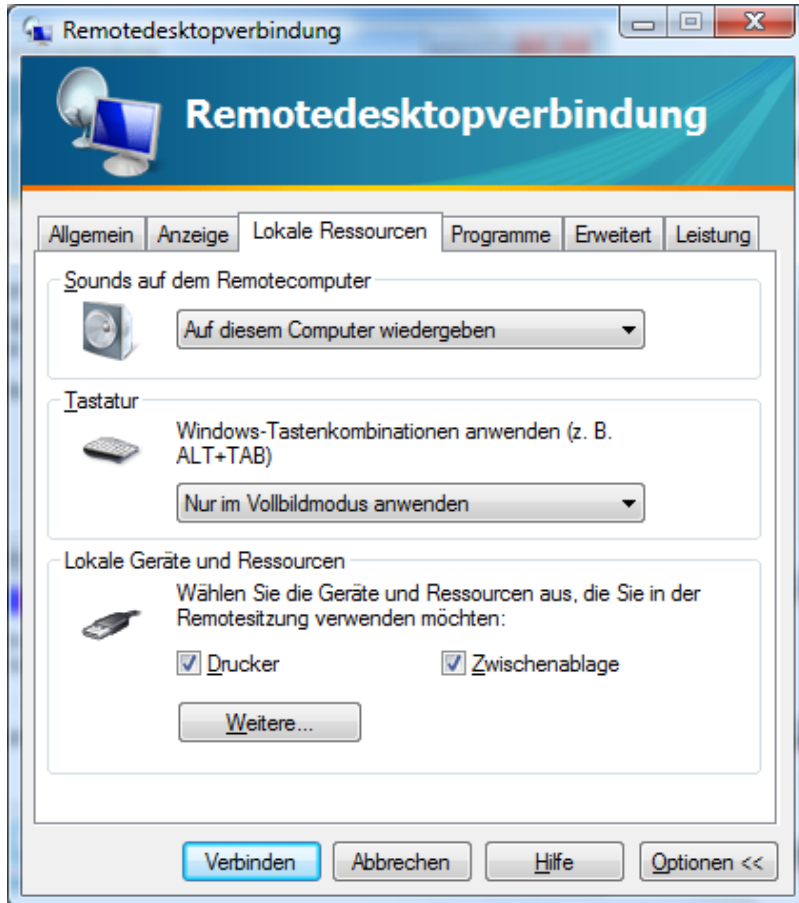
- ▶ Use the *Remote Desktop Connection* program, usually available under *All Programs* → *Accessoires* → *Communication*.



- Specify the computer name, i.e. `cluster-win.rz.rwth-aachen.de`.
- You can save a set of settings under a named profile / link.



- You can choose a resolution or fullscreen mode.
- You can choose the color depth.
- In fullscreen mode you should set this flag to ease the handling of the remote desktop window.



- Take resources of your local computer with you:
 - Audio device
 - Printer
 - Clipboard
 - Local hard disc drives
 - Locally mounted network drives

- ▶ Use the `rdesktop` program available from www.rdesktop.org, probably already included in your distribution.
- ▶ **Basic usage:** `rdesktop [options] host with`

 - ▶ `-u <user>` Login as user `<user>`
 - ▶ `-d WIN-HPC` Login to domain WIN-HPC
 - ▶ `-4` Use protocol version 4 (often needed)
 - ▶ `-g WxH` Use resolution Width x Height
 - ▶ `-f` Use fullscreen resolution
 - ▶ `-a 24` Use 24bit color depth
 - ▶ `-k de` Use german keyboard layout
 - ▶ `-r sound:local` Play sound on local system

Batch System

▶ Multi-Threading with OpenMP

- ▶ Control the number of threads to be used with env. Variable
- ▶ `set OMP_NUM_THREADS=8`
- ▶ Batch job: Reserve full node or appropriate number of cores per process and set environment variable

▶ Message-Passing with MPI

- ▶ Number of processes is determined by startup command
- ▶ `mpiexec -n 8 ...`
- ▶ Batch job: Reserve appropriate number of nodes or cores, number of processes is then determined automatically

- ▶ Find the HPC Job Manager in the menu: *All Programs* → *Microsoft HPC Pack* → *HPC Job Manager*.

Several pre-conf. views available.

„Modern“ Action Pane.

► To submit a new Job choose *Actions* → *Job Submission* →

Create New Job

Job Details

Job name:

Job template:

Project:

Priority:

Job run options

Do not run this job for more than:

Days: Hours: Minutes:

Run job until cancelled or run time expires

Fail the job if any task in the job fails

Job resources

Select the type of resource to request for this job:

Enter the minimum and/or maximum of the selected resource type that this job is allowed to use:

Minimum: Auto calculate

Maximum: Auto calculate

Use assigned resources exclusively for this job

No other jobs will be allowed to run on the selected nodes while the job is running.

Submit Save Job as ... Cancel

- You are free to choose a *Job Name* and a *Project Name* as you like.
- You might specify runtime and failure options for the job.
- Resource allocation changed significantly:
 - Per *Core*, or
 - Per *Socket*, or
 - Per *Node*.
- Saving of Job Templates possible.

▶ Resource Allocation Granularity:

- ▶ *Per Core*: Get n processor cores. No further restrictions, for example it cannot be assumed that a (sub)set of cores shares the same main memory (→ not suited for Shared-Memory).
- ▶ *Per Socket*: Get n sockets. On our cluster, currently each socket has four cores (quad-core Xeon), thus it can be used for Hybrid or Shared-Memory (up to four threads per process).
- ▶ *Per Node*: Get n nodes. On our cluster, currently each node has two sockets à four cores (dual-socket quad-core Xeon), thus it can be used for Hybrid or Shared-Memory (up to eight threads per node).

▶ If you use OpenMP: Set `OMP_NUM_THREADS` env. variable, otherwise you would get as many threads as there are cores

- ▶ `mpiexec -genv OMP_NUM_THREADS 2`

▶ A Job consists of one or more Tasks.

Enter the tasks for this job

Task Name	Command Line	Requested Resources
-----------	--------------	---------------------

Task name: MyTask

Command line:
Task.exe

Work directory: Browse...

Standard input: Browse...

Standard output: Browse...

Standard error: Browse...

Specify the minimum and maximum number of resources to use for this job. The job resource type is set to core.

Minimum: Maximum:

Buttons: Save, Cancel, Submit, Save job as..., Cancel

- *Command Line*: You can specify the full path to a program including program options or to a .bat or .cmd file.
- You have to use network paths instead of drive letters in any path.
- For MPI Tasks just include `mpiexec` in the *Command Line*, do not specify any other MPI options.

- ▶ Some restrictions for the node selection can be specified.

The screenshot shows the 'Create New Job' dialog box with the 'Resource Selection' tab selected. The dialog is divided into several sections:

- Job Details:** Includes 'Task List' and 'Licenses'.
- Node preferences:** Contains a checkbox 'Run this job only on nodes in the following node groups:'. Below it are two lists: 'Available node groups' (HeadNodes, ComputeNodes, WCFBrokerNodes) and 'Selected node groups' (empty). Buttons 'Add >>' and '<< Remove' are between the lists.
- Run this job only on nodes in the following list:** Contains a checkbox and a table of nodes.
- Hardware preferences:** Contains checkboxes for 'Minimum memory (MB):' (value 0), 'Minimum cores:' (value 0), and 'Prefer nodes with:' (dropdown menu set to 'More Memory').

At the bottom of the dialog are buttons for 'Submit', 'Save Job as ...', and 'Cancel'.

Node Name	Cores	Memory	State
<input type="checkbox"/> WINSCC002	8	16383	Ready
<input type="checkbox"/> WINSCC004	8	16383	Ready
<input type="checkbox"/> WINSCC005	8	16383	Ready
<input type="checkbox"/> WINSCC006	8	16383	Ready
<input type="checkbox"/> WINSCC007	8	16383	Ready
<input type="checkbox"/> WINSCC008	8	16383	Ready
<input type="checkbox"/> WINSCC009	8	16383	Ready
<input type="checkbox"/> WINSCC010	8	16383	Ready
<input type="checkbox"/> WINSCC011	8	16383	Ready

- Allow selected classes of nodes only.
- Allow a selected set of nodes only.
- Allow nodes with enough memory only.

Software Environment

▶ Complete set of Development Software:

▶ cluster-win:

- ▶ Visual Studio 2008 and Visual Studio 2010
- ▶ Intel Compiler Suite 12 (C/C++ and Fortran)
- ▶ Microsoft HPC Pack 2008 R2 (= MS-MPI)
- ▶ Intel Cluster Toolkit 3.1
 - ▶ Intel MPI 3.2 (= I-MPI)
 - ▶ Intel Trace Analyzer & Collector 7.2
- ▶ Intel Threading Building Blocks 2.2
- ▶ Intel Parallel Studio 2011 XE

▶ If you want to use Parallel Studio, contact us at hpc@rz.rwth-aachen.de

▶ **Complete set of Development Software:**

- ▶ cluster-win-beta: same as above, but / plus (+)
 - ▶ Intel beta (C/C++ and Fortran compilers, MPI, Tools, ...)
 - ▶ soon: Visual Studio 11 beta

▶ **Selected (and growing) list of tools and utilities:**

- ▶ All interactive machines:
 - ▶ Notepad++ editor
 - ▶ Subversion Client
 - ▶ Tortoise Subversion GUI / Explorer integration
 - ▶ X-Win32
- ▶ cluster-win-beta: same as above plus Windows Debugging / Analysis tools

▶ Selected (and growing) list of ISV-Software:

- ▶ ANSYS
- ▶ HyperWorks
- ▶ Fluent
- ▶ Maple
- ▶ Mathematica
- ▶ Matlab
- ▶ Microsoft Office 2010 + Microsoft Excel Compute Services
- ▶ MSC.Marc
- ▶ MSC.Adams
- ▶ Linear Algebra Libraries (e.g. Intel MKL 10.2)
- ▶ ...

If we have (floating) licenses and if the software is available on Windows, we try to provide it.

We make user-provided software available as well (if possible without giving administrator privileges away).