OS-Update on FLOW

Stefan Albensoeder

Contact: Stefan.Albensoeder@uni-oldenburg.de

MOTIVATION

Motivation

Current Operating System (OS)

- Scientific Linux
- Based on RedHat 5.x
- Old kernel/system libraries
  → causes problems when install new libraries/software
  → some new tools are not running any more
  → security leaks
- Old InfiniBand stack
  → parallel programs have not optimal performance
- New nodes are not supported

FLOW EXTENSION

FLOW extension

Status

- 10 Ivy-Bridge nodes à 16 cores, 64Gb memory
- Speedup (preliminary benchmark)
  - OpenFOAM: 2-3x
  - PALM: 2-2.5x
  - WRF: 2x
- Used as test system for new OS
  (Old OS not usable for the new nodes)
- Login node for test cluster:

  ```
  flow03.hpc.uni-oldenburg.de
  ```
- Full integration with OS update in January

NEW OPERATING SYSTEM ON FLOW

New Operating System on FLOW

- RedHat 6.5
  → actual software supports this OS
- Newer libraries/kernel
  → supports new compute nodes
- New InfiniBand stack
  → improved performance for parallel programs
- E.I.S. hopefully in January
- Test cluster: `flow03.hpc.uni-oldenburg.de`
- Many changes
  - New modules/old software removed
  - Integration of new nodes in FLOW
  - Changes in SGE

CHANGES IN MODULES

Changes in modules

- Sorted in groups for better overview
- Update of software releases (compiler, libraries,…)
- Deletion of unused modules
  → better overview
- Reduced dependency of modules
  → new modules try to avoid to load other modules (e.g. Intel compiler)
  → Needed dependencies still loaded (e.g. related MPI release)
- Unique naming scheme
- Avoid double installation (e.g. WRF geo data, ParaView for OpenFOAM)

Changes in modules

New modules/view

```
user@flow02:~ > module avail
------------------ /cm/shared/uniol/modulefiles/SYSTEM -------------------
hpc-uniol-env qt4/4.8.6     sge/2011.11p1

------------------ /cm/shared/uniol/modulefiles/CFD ----------------------
ansys/15.0.1                  openfoam/3.1-ext_2014_10_08
dars/2.08.009                 starccmp/8.06.007_01
foampro/2.1.2                 starccmp/9.04.009_01
nektarpp/3.4.0                starccmp/9.06.009_02
openfoam/1.6-ext_2011_09_28   wrf/3.5/em_real/DP
openfoam/1.6-ext_2013_11_15   wrf/3.5/em_real/SP
openfoam/1.7.1                wrf/3.6.1/em_b_wave
openfoam/2.0.0                wrf/3.6.1/em_heldsuarez
openfoam/2.0.1                wrf/3.6.1/em_les
openfoam/2.1.0                wrf/3.6.1/em_quarter_ss
openfoam/2.1.1                wrf/3.6.1/em_real/DP
openfoam/2.2.0                wrf/3.6.1/em_real/SP
openfoam/2.2.1                wrf/3.6.1/em_scm_xy
openfoam/2.2.2                wrf/3.6.1/em_tropical_cyclone
openfoam/2.3.0                wrfgeodata/3.6.1
openfoam/2.3.1
...
```

Changes in modules

New modules/view

```
...
----------------- /cm/shared/uniol/modulefiles/CHEMISTRY -----------------
gaussian/g09.d01 molcas/78        molpro/2010.1

----------------- /cm/shared/uniol/modulefiles/COMPILER ------------------
clang/3.5.0           ics/2013_sp1.3.174/64 pgi/12.10
gcc/4.8.1             nag_fortran/5.2       pgi/13.10
ics/2013_sp1.3.174/32 open64/4.5.2.1

-------------- /cm/shared/uniol/modulefiles/DATAPROCESSING ---------------
cdo/1.6.4                        merra2wrf/2.0
gmt/5.1.1                        merra2wrf/2.0_mod
idl/8.4                          ncarg/6.2.0
maple/18                         ncarg_highres_coastlines/2014.12
matlab/r2010b                    nco/4.4.4
matlab/r2011a                    octave/3.8.1
matlab/r2011b                    python/2.7.8
matlab/r2013a                    r/3.1.1
...
```

Changes in modules

New modules/view

```
...

---------------- /cm/shared/uniol/modulefiles/DEVELOPMENT ----------------
cmake/2.8.12.2       java/latest           scalasca/2.1/openmpi
flex/2.5.39          kcachegrind/0.7.4     valgrind/3.10.1
itac/9.0.0.028       likwid/3.1.2
java/1.8.0           netbeans/8.0

----------------- /cm/shared/uniol/modulefiles/LIBRARIES -----------------
arpack/96/intel/2013_sp1.3.174
boost/1.55.0
fftw/3.3.4/DP/serial/intel/2013_sp1.3.174
fftw/3.3.4/SP/serial/intel/2013_sp1.3.174
grib_api/1.12.3/gcc/4.4.7
grib_api/1.12.3/intel/2013_sp1.3.174
gsl/1.16/intel/2013_sp1.3.174
hdf4/4.2.10
hdf5/1.8.13/gcc/4.4.7
hdf5/1.8.13/intel/2013_sp1.3.174
metis/5.1.0
...
```

Changes in modules

New modules/view

```
...
nag/c_library/9/intel
nag/dmc_library/2
nag/fortran90_library/4/intel
nag/fortran_library/23/intel
nag/matlab_toolbox/22
nag/parallel_library/3/intel
nag/smp_library/22/intel
netcdf/4.3.2/gcc/4.4.7
netcdf/4.3.2/intel/2013_sp1.3.174
qrupdate/1.1.2/intel/2013_sp1.3.174
udunits/2.1.24
zlib/1.2.8

-------------------- /cm/shared/uniol/modulefiles/MPI --------------------
impi/5.0.0.028/32/gcc   impi/5.0.0.028/64/gcc   openmpi/1.8.2/gcc
impi/5.0.0.028/32/intel impi/5.0.0.028/64/intel openmpi/1.8.2/intel

--------------- /cm/shared/uniol/modulefiles/VISUALIZATION ---------------
gnuplot/4.6.5   paraview/3.12.0 xmgrace/5.1.23
ncview/2.1.2    paraview/4.1.0
```

Changes in specific modules

OpenFOAM

- module set standard aliases of OpenFOAM
(e.g. wmSET, foamApps, tut, ...)
- in all releases the flushing is disabled
- ParaView is installed in a separate module
(the related ParaView module will be loaded by the OpenFOAM module)

WRF

- The geographical data is installed in a separate module and has not to be installed by the user (to reduce the used disk space). The location is automatically set in the `namelist.wps` file when using the `setup_wps_dir.sh` script. Additionally the location is set in the environment variable `WRF_GEO_DATA_DIR.`

CHANGES FOR PALM

Changes for PALM

- modifications in `mrun` script
  - SGE complex `highmem` is not defined anymore and has to be removed
  - Parallel environment has to be changed to `impi` instead of `impi41`
- modifications in `.mrun_config`
  - Use new modules and login command

```
%login_init_cmd .:/etc/bashrc;:export:HOSTNAME=flow   lcflow parallel
%modules  hpc-uniol-env:sge/2011.11p1:ics/2013_sp1.3.174/64:\
          impi/5.0.0.028/64:qt4  lcflow parallel
```

  - Add new nodes

```
%host_identifier       cfdi*             lcflow
```

CHANGES IN SGE

Motivation

- Use OS update as chance to review the SGE settings
- Current disadvantages and potential error sources
  - no queue/reserved nodes for serial jobs
    - serial jobs uses usually a full node
    - with `excl_flow=false` jobs can disturb parallel jobs
  - complexes are defined to stear jobs in the right queue but it is maybe not needed, e.g `highmem, express, longrun`
  - to many parallel environments
- Low utilization of `cfdx` nodes
- New nodes are missing in the old concept

Parallel environments

- Reduction of parallel environments
  - `*_long` environments are removed
    (not needed and causes many errors by the users in the past)
  - `impi41` is replaced by `impi41` to `impi`
    (Intel MPI 4.0 is deprecated and not installed anymore)
  - Unused PEs are removed
- Valid PEs

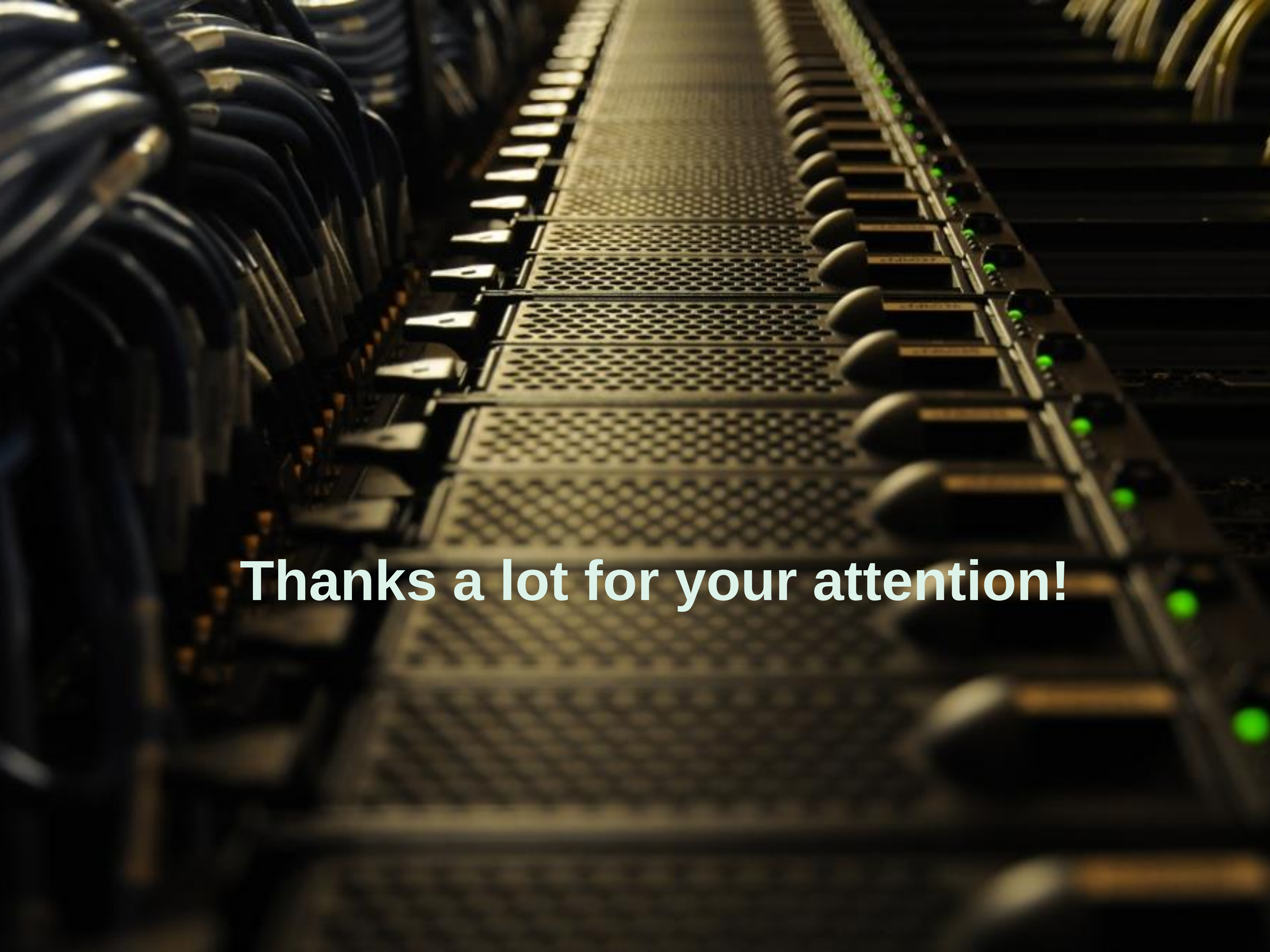
```
ansys
impi
linda
mdcs
molcas
mpich
mpich2
openmpi
openmpi_ib
smp
starccmp
```

Reordering/redefinition of queues

- `highmem`, `long` and `express` complexes are removed
  → SGE decide by justification of the queues
  → Longruns are still allowed, but maybe limited (to a certain number of slots) in future
- New time limits
  - interactive queue max. 8h
  - if `h_rt` not set limit to 24h (15 minutes for interactive jobs)
- New memory limits
  (to avoid inefficient use of most of the nodes and to allow deletion of `highmem`)
  - `cfdl*` 1850M per slot
  - `cfdi*` 3850M per slot
  - no limit only on `cfdh*`
  - no limit for express jobs

Reordering/redefinition of queues

- use `cfdx` nodes automatically for
  - serial jobs (to keep them away from other nodes)
  - express jobs (max. 2 hours, parallel jobs allowed)
  - interactive jobs with `qlogin`
- use new ivy bridge nodes only for jobs with `h_rt` up to 24h (to enable many user to speed up their runs)



Thanks a lot for your attention!