XXL MEM node Documentation

CINES

nov. 13, 2020

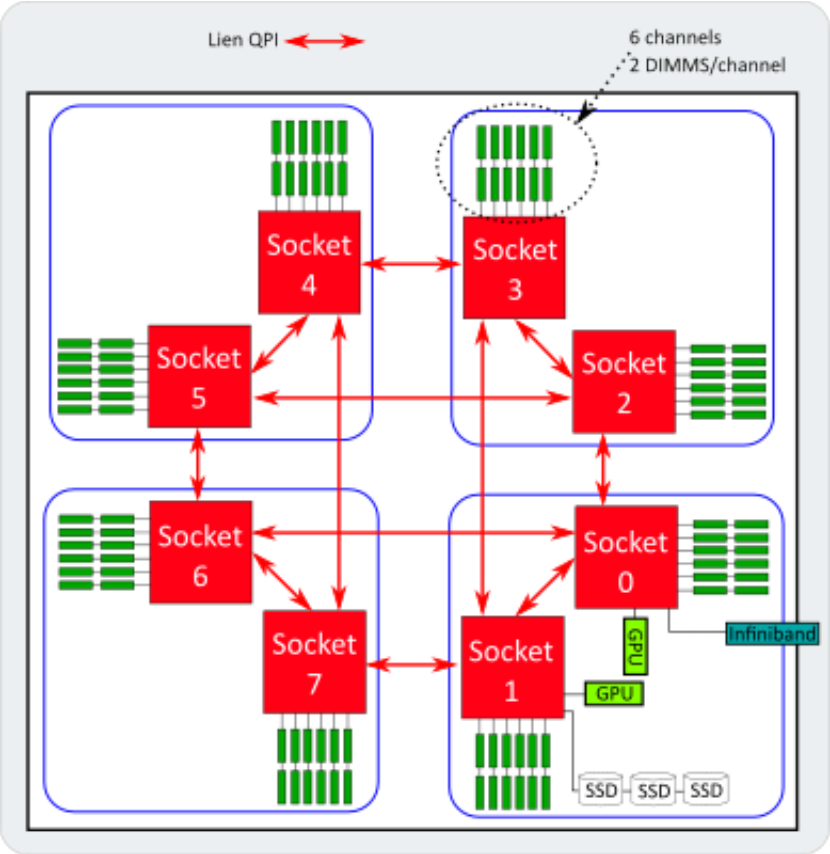
Contenu:

1 Présentation du nœud XXLMEM

Depuis novembre 2018, occigen intègre un nœud Bull x808 avec :

- 3TB de RAM accessibles en NUMA (glueless interconnect) par 8 sockets Intel® Xeon® Platinum 8176 (224 cores au total).
- 2 GPU Nvidia Tesla P100 (12GB).

Voici la représentation simplifiée de ce noeud :



2 Obtenir l'accès au nœud

L'accès au noeud s'effectue via Slurm, en rajoutant la contrainte `-C XXLMEM` dans votre script de soumission.

Voici des exemples d'utilisation d'une telle machine :

- jobs à large mémoire ou non optimaux sur les nœuds fins
- production / traitement d'images, de maillages et analyse de données hors visu interactive
- Proof of Concept : tests de logiciel GPU sans utilisation dans un mode production

Notez que dans un même travail batch, vous ne pouvez demander ce noeud SKYLAKE en même temps que d'autres noeuds HASWELL, BROADWELL, ou noeuds de visualisation.

3 Espaces disques

Le nœud a accès a quatre espaces disques :

— le $HOME
— le $SCRATCHDIR
— le $STOREDIR
— le /tmp

Les descriptions des trois espaces disques peuvent être trouvées ici.

Le /tmp local profite de la performance de disques SSD. Cet espace volatile, permet de disposer d'un espace de travail de plus de 1,2 To avec des performances élevés.

Attention, cet espace n'est pas destiné à conserver les données. Si vous souhaitez préserver vos fichiers entre les jobs, nous vous invitons à les déplacer vers des systèmes de fichiers sécurisés, comme le $HOME et le $STOREDIR.

4 Module

Environnement logiciel : module

Le CINES met à disposition de nombreux logiciels sur ses clusters, et souvent plusieurs versions de chaque logiciel.

Afin d'éviter les conflits entre différentes versions d'un même logiciel, il faut généralement définir un environnement propre à chaque version.

Les logiciels disponibles peuvent être visualisés ou chargés via les commandes suivantes :

module avail	afficher la liste des environnements disponibles
module load	charger une librairie ou un logiciel dans votre environnement
module list	afficher la liste des environnements chargés
module purge	retirer un environnement déjà chargé
module show	voir le contenu du module

Si vous ne trouvez pas le logiciel dont vous avez besoin dans la liste, contactez nous.

5 Modes d'utilisation

Le nœud peut être utilisé de trois façons différentes :
— MPI seul
— OpenMP seul
— Hybride MPI+OpenMP

Des scripts d'exemples sont disponibles dans les sections suivantes.

5.1 MPI

```
#!/bin/bash
#SBATCH -J xxlmem_mpi
#SBATCH --nodes=1
#SBATCH --ntasks=224
#SBATCH --ntasks-per-node=224
#SBATCH --time=0:40:00
#SBATCH -C XXLMEM
#SBATCH --exclusive
#SBATCH --output xxlmem_mpi.output.slurm
```

(suite sur la page suivante)

(suite de la page précédente)

```
set -e

#####Intelmpi placement auto
# module load intel/18.1 intelmpi/2018.1.163
# export I_MPI_DOMAIN=auto
# export I_MPI_PIN_RESPECT_CPUSET=0
# ulimit -s unlimited
# srun ../../../bin/hello_mpi

#####Intelmpi avec placement pour mpirun
# module load intel/18.1 intelmpi/2018.1.163
# export SLURM_CPU_BIND=NONE
# export I_MPI_PIN=1
# export I_MPI_PIN_PROCESSOR_LIST=0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,
↪20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,
↪48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,
↪76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,
↪103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,
↪124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,
↪145,146,147,148,149,150,151,152,153,154,155,156,157,158,159,160,161,162,163,164,165,
↪166,167,168,169,170,171,172,173,174,175,176,177,178,179,180,181,182,183,184,185,186,
↪187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,
↪208,209,210,211,212,213,214,215,216,217,218,219,220,221,222,223
# ulimit -s unlimited
# mpirun ../../../bin/hello_mpi


####Openmpi placement auto
module load intel/18.1 openmpi/intel/2.0.2
ulimit -s unlimited
srun ../../../bin/hello_mpi
```

5.2 OpenMP

```
#!/bin/bash
#SBATCH -J xxlmem_omp
#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --ntasks-per-node=1
#SBATCH --cpus-per-task=224
#SBATCH --time=0:40:00
#SBATCH -C XXLMEM
#SBATCH --exclusive
#SBATCH --output xxlmem_omp.output.slurm

set -e

#Make sure that OMP_NUM_THREADS = cpus-per-task * KMP_HW_SUBSET
export KMP_HW_SUBSET=1T
export OMP_NUM_THREADS=$SLURM_CPUS_PER_TASK
export KMP_AFFINITY=verbose,compact,1,0,granularity=fine

module load intel

ulimit -s unlimited
```

(suite sur la page suivante)

(suite de la page précédente)

```
rm -f *.out

srun ../../../bin/hello_omp
```

5.3 Hybride

```
#!/bin/bash
#SBATCH -J xxlmem_hybrid
#SBATCH --nodes=1
#SBATCH --ntasks=8
#SBATCH --ntasks-per-node=8
#SBATCH --cpus-per-task=28
#SBATCH --time=0:40:00
#SBATCH -C XXLMEM
#SBATCH --exclusive
#SBATCH --mem=50GB
#SBATCH --output xxlmem_hybrid.output.slurm

set -e

#####Intelmpi
# module load intel intelmpi
# export I_MPI_DOMAIN=auto
# export I_MPI_PIN_RESPECT_CPUSET=0
# #Make sure that OMP_NUM_THREADS = cpus-per-task * KMP_HW_SUBSET
# export KMP_HW_SUBSET=1T
# export OMP_NUM_THREADS=12
# export KMP_AFFINITY=verbose,compact,1,0,granularity=fine
# ulimit -s unlimited
# srun ../../../bin/hello_hybrid

#####Openmpi
module load intel/18.1 openmpi/intel/2.0.2
#Make sure that OMP_NUM_THREADS = cpus-per-task * KMP_HW_SUBSET
export KMP_HW_SUBSET=1T
export OMP_NUM_THREADS=28
export KMP_AFFINITY=verbose,compact,1,0,granularity=fine
ulimit -s unlimited
srun ../../../bin/hello_hybrid
```