

中国人民大学物理系高性能计算集群“Newton”介绍

刘 凯

2013.3

一. 系统简介

(1) 硬件部分

目前 Newton 集群共有 80 个节点（1 个管理节点，4 个存储节点，70 个双路计算节点，5 个四路胖节点），共 1280 个计算核心，约 3.5TB 内存，40TB 集中存储空间，理论计算峰值为 26.4 万亿次。管理网络采用千兆以太网；计算网络采用 40Gb QDR Infiniband 网络；计算节点到存储节点为 40Gb QDR Infiniband 网络，存储节点经光纤交换机连接至 DDN 存储设备（双控制器，提供 4 条 8Gb 光纤链路，Raid6 阵列）。Newton 集群拓扑结构图如下：

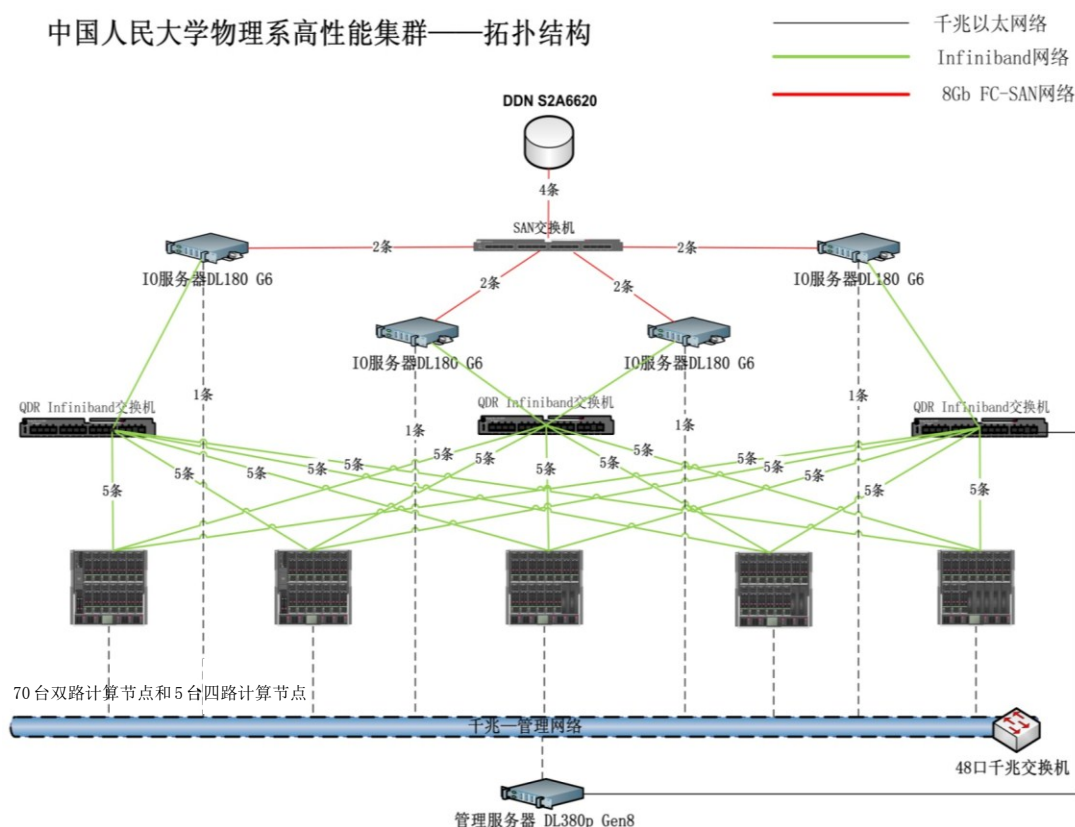


图 1. Newton 集群拓扑结构图（由惠普公司提供）

管理节点 1 个，型号为 HP DL380p G8 机架式服务器，配置 2 个八核 Intel Xeon E5-2650 系列 CPU（主频 2.0GHz，L3 缓存 20MB，Intel QPI 8.0GT/s）；8 条 4GB DDR3 内存；2 个 600G 10K 转 SAS 硬盘，硬盘阵列 Raid 1；1 块 40Gb/s QDR Infiniband HCA 卡。

存储节点 4 个，型号为 HP DL180 G6 机架式服务器，配置 2 个四核 Intel Xeon E5620 系列 CPU（主频 2.40GHz，L3 缓存 12MB，Intel QPI 5.86GT/s）；6 条 4G DDR3 内存；2 个 300G 10K 转 SAS 硬盘，硬盘阵列 Raid 1；1 块 40Gb/s QDR Infiniband HCA 卡；1 块 8Gb 双口光纤 HBA 卡。

计算节点 75 个，分布于 5 个 HP c7000 刀片机箱中，其中 70 个双路计算节点为 HP BL460c Gen8 刀片，5 个四路胖节点为 HP BL660c Gen8 刀片。具体参数如下：

数量	型号	CPU 规格	内存	硬盘
70	BL460	Intel Xeon E5-2670 2.6GHz 八核*2	32GB DDR3	300GB 10Krpm SAS
5	BL660	Intel Xeon E5-4640 2.4GHz 八核*4	256GB DDR3	2*1TB 10Krpm SAS

75 个计算节点之间通过 3 台 36 口 Mellanox Infiniband（40Gb QDR）交换机无阻塞连接，管理节点和存储节点也接入 40Gb QDR Infiniband 网络。

（2）操作系统、编译器、并行环境、数学库和开源软件

集群部署和作业调度系统采用联科 CHES 软件，编译器版本管理采用 Modules Environment 模块化环境。

操作系统： Linux，内核版本 2.6.32-220

编译器：

通过输入 `module avail` 命令查看可用编译器和数学库；`module list` 查看已加载的环境；通过 `module load` 和 `module unload` 加载和卸载所需的环境。系统已默认加载 Intel C/C++/Fortran 编译器和 MKL 数学库，用户可以在管理节点编译程序：

Fortran 为 ifort; C 为 icc; C++为 icpc。

并行环境：

openmpi, 可通过 `module load mpi/openmpi/1.6.3` 加载该环境

impi, 可通过 `module load mpi/impi/1.6.3` 加载该环境

mvapich2, 可通过 `module load mpi/mvapich2/1.9` 加载该环境

数学库：

Intel MKL, 默认已加载（通过 `module load intel/12.1.4`），其他数学库

全部安装在/opt/libs 文件夹中，可通过 module avail 查看，module load ***加载数学库，module list 确认是否加载成功，除 Intel MKL 外主要包括以下数学库

Lapack: 开源 BLAS 和 LAPACK 库

FFTW: 快速傅里叶变换

GSL: 开源的 C 和 C++数学库

NETCDF: 处理科学数据的软件库，能生成独立于机器的格式

推荐使用 Intel MKL 函数库:

其中 Intel MKL 函数库默认已经加载，比如 fortran 程序链接方法为

ifort code.f90 -lmkl_intel_lp64 -lmkl_sequential -lmkl_core -lpthread

开源软件:

Quantum ESPRESSO (开源的第一性原理计算软件):

在 /opt/software/espresso-5.0.2/ 文件夹下装有用 openmpi 编译的 espresso 软件;

在 /opt/software/espresso_pseudo/ 文件夹中有赝势库。

GNUPLOT (开源绘图软件):

集群已安装 GNUPLOT，用户可以直接输入命令 gnuplot

(3) 系统性能

由于集群中计算节点的硬件类型不同，我们分别做了 Linpack 测试:

节点数量	型号	CPU 规格	内存	理论峰值 (TFlops)	实际峰值 (TFlops)	Linpack 效率
70	BL460	Intel Xeon E5-2670 2.6GHz 八核*2	32GB	23.296	20.46	87.8%
5	BL660	Intel Xeon E5-4640 2.4GHz 八核*4	256GB	3.072	2.763	89.9%

注: 实测数据由联智科技(北京)有限公司工程师完成。