中国人民大学物理系高性能计算集群“Kohn”介绍

刘　凯

2010.6

一．系统简介

（1）硬件部分

目前 Kohn 集群共有 73 个节点（1 个管理节点，4 个存储节点，68 个计算节点），592 个核，约 1TB 内存，理论计算峰值为 5.5 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 20Gb DDR Infiniband 网络，存储节点经由光纤交换机连接至存储设备 IBM DS3400+EXP3000（双控制器，提供 4 条 4Gb 光纤链路，24 块 450GB 15K 转 SAS 硬盘构建 Raid 阵列）。

图 1. Kohn 集群正面图

图 2. Kohn 集群背面图

管理节点 1 个，型号为 IBM X3650M2 机架式服务器， 配置 2 个四核 Intel Xeon E5530 系列 CPU（主频 2.40GHz，L3 缓存 8MB，Intel QPI 5.86GT/s）；12G DDR3 内存；2 个 146G 15K 转 SAS 硬盘，硬盘阵列 Raid 1。

存储节点 4 个，型号为 IBM X3650M2 机架式服务器， 配置 2 个四核 Intel Xeon E5530 系列 CPU（主频 2.40GHz，L3 缓存 8MB，Intel QPI 5.86GT/s）；12G DDR3 内存；2 个 146G 15K 转 SAS 硬盘，硬盘阵列 Raid 1；20Gb DDR Infiniband HCA 卡；4Gb 光纤 HBA 卡。

计算节点 68 个，分布于 5 个 IBM BCH 刀片机箱中，其中 66 个计算节点为 IBM HS22 刀片，2 个为 IBM LS22 刀片。共包含以下四种刀片类型：

数量	型号	CPU	内存	硬盘
38	HS22	Intel Xeon E5530 2.4GHz 四核	6*2GB DDR3	146GB SAS 6Gb/s 15Krpm
14	HS22	Intel Xeon X5550 2.66GHz 四核	6*2GB DDR3	146GB SAS 6Gb/s 15Krpm
14	HS22	Intel Xeon X5550 2.66GHz 四核	6*4GB DDR3	146GB SAS 6Gb/s 15Krpm
2	LS22	AMD Opteron 2435 2.6GHz 六核	4*2GB DDR2	146GB SAS

68 个计算节点之间通过 2 台 36 口 Voltaire4036 Infiniband（40Gb QDR）交换机无阻塞连接，交换机中的 4 个口通过 20Gb DDR Infiniband 连接至存储节点。

（2）操作系统、编译器、并行环境、数学库和开源软件

集群部署采用 Rocks 5.3，作业调度批处理系统为 Grid Engine，编译器版本管理采用 Modules Enviroment 模块化环境。

操作系统：Linux，内核版本 2.6.18-164

编译器：

通过输入 module avail 命令查看可用编译器；module list 查看已加载的环境；通过 module load 和 module unload 加载和卸载所需要的环境。系统已默认加载 intel 编译器，用户可以在管理节点编译程序：

Fortran 为 ifort，绝对路径为/opt/intel/Compiler/11.1/064/bin/intel64/ifort

C 为 icc，绝对路径为/opt/intel/Compiler/11.1/064/bin/intel64/icc

C++为 icpc，绝对路径/opt/intel/Compiler/11.1/064/bin/intel64/icpc

并行环境：

Openmpi，绝对路径/share/apps/compilers/openmpi-1.3.3-intel-11/bin/，可通过 module load openmpi/intel-11 加载该环境。

Mvapich2，绝对路径/share/apps/compilers/mvapich2-1.5rc1-intel-11/bin

可通过 module load mvapich2/intel-11 加载该环境。

注意：以上并行环境不能同时加载，用户可以自行比较以上并行环境对自己程序的影响，选择合适的并行环境。

数学库：

Intel MKL：绝对路径为/opt/intel/Compiler/11.1/064/mkl/lib/em64t/

其他数学库全部安装在/share/apps/libs 文件夹中，包括

GotoBLAS2：当前最快的 BLAS 库

Lapack：开源 BLAS 和 LAPACK 库

ACML：针对 AMD CPU 优化的数学库

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 GotoBLAS2 库和 Intel MKL 函数库：

其中 GotoBLAS2 库的连接方法为

-L/share/apps/libs/GotoBLAS2/ -lgoto2_nehalemp-r1.13

Intel MKL 函数库的连接方法为

-L/opt/intel/Compiler/11.1/064/mkl/lib/em64t/ -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/share/apps/software/espresso-4.2openmkl/文件夹下装有用 openmpi 编译的 espresso 软件；

在/share/apps/software/espresso-4.2mvamkl/文件夹下装有用 mvapich2 编译的 espresso 软件；

在/share/apps/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	内存	理论峰值（TFlops）	实际峰值（TFlops）	Linpack 效率
38	HS22	Intel Xeon E5530 2.4GHz	6*2GB	2.918	2.564	88%
14	HS22	Intel Xeon X5550 2.66GHz	6*2GB	1.192	1.109	93%
14	HS22	Intel Xeon X5550 2.66GHz	6*4GB	1.192	1.146	96%

注：系统性能实测数据由 IBM 工程师完成。