World Scientific Research Journal (WSRJ) ISSN: 2472-3703
www.wsr-j.org

The Building and Application of High Performance Computation Platform

Zhaoyong Zhou[a], Li Li, Xijun Wang, Chengyu Cui, Wei Liu, Xilian Deng

Network & Education Technology Center, Northwest A&F University, Yangling 712100,

Shaanxi, China

[a]yzz@nwsuaf.edu.cn

Abstract: There are three methods of modern scientific research: theoretical argument, scientific experiment and scientific calculation. In recent years, advanced scientific calculation has gradually become the main method of scientific research. Establishment of school-level high performance computing platform is very necessary to adapt to modern scientific study development and improve universities' scientific research level. First, this paper introduces the concept and application prospect of high performance computing, next, it expounds HPC theory, and then gives the framework and composition of the high performance computing platform, and the main technical point of implementation has been given at the same time. At the end, it gives the test result and application situation after the platform integration. The operation result of integrated platform shows that the platform scale is appropriate, and the technology is advance, and achieves the desired objective.
Keywords: High performance computing, MPI, Platform Architecture, Building and Application

---

1. INTRODUCTION

With the rapid development of modern information technology, modern scientific research has entered the data-intensive informatization big data times. Through to excavate and analyze the great data in science, we can know inherent law about object preferably, solve the problem difficult to solve before, even the inextricable scientific problem. High performance computing or HPC for short, is a branch of computer science, which mainly refers to the technology of high performance computer development from architecture, parallel algorithm and software development aspects [1]. In recent years, HPC has been used extensively in important scientific researches and engineering application fields, such as: earth system simulation, seismic exploration, meteorology, geographic calculation, image processing, aviation, aerospace, underwater ship, theoretical chemistry, wireless and satellite remote sensing [2-12], it provides new ideas, new ways and new means to scientific exploration. HPC has gained unexpected

achievement in many aspects, and promoted the development of scientific research and advancement of engineering technology.

The Northwest Agriculture and Forestry University has a high realistic and also potential demand for high performance computing in important scientific researches and engineering application fields, such as: gene research, environmental simulation, numerical calculation, data analysis and financial engineering. To build an open infrastructure of high performance computing services with advanced hardware, complete function and abundant resources, for all the teachers and students, to provide quality services for scientific research and teaching of university, especially for the key discipline's scientific research. This will improve the innovation method and capability of university in major national requirements and academic foreland field, such as: ecological environment, climatic change, agricultural water research, biological information and high polymer chemistry, and improve university's scientific research competence.

2. HPC THEORY

Typical high performance computing system is a management system formed by I/O storage nodes, managed nodes, login nodes and compute nodes which are connected through Ethernet switch. Compute nodes connected by high speed Ethernet form a computing system, and the whole system connects with the outside world through internet by master nodes. Each node installed with Linux operating system, each node adopts TCP/IP or Infiniband protocol as communication protocol, and each node configured with MPI parallel environment [13-14]. Through management node or login node management and log into the cluster, to send order to each node on the base of Linux operating system and perform the program specified by users on compute nodes; to compute independently on each compute node, and to exchange informations, balance steps and execute the control between compute nodes. Information exchange realized through high speed local area network, the computing is very simple on relative compute node which can be ignored sometimes, thus, one task could run on multiple nodes to realize high performance computing function [15]. Parallel algorithm is the basis of parallel computing which combines with realized technology to provide solutions for effective used parallel computer.

3. HPC PLATFORM'S CONSTITUTION AND SYSTEM STRUCTURE

3.1 Platform constitution

The high performance computing platform includes hardware, operating system, parallel environment, simulation software and management system. Hardware mainly includes CPU, memory, mainboard, hard disk, network and so on. CPU is the core computing component of high performance parallel computing. Memory is the bridge of communication between CPU and external storage such as hard disk, and all programs of computer are running in internal

storage, so the memory's performance has a great influence on calculated performance and stability of computer. The mainboard connects with each component of computer, receives CPU's order to coordinates the work of each component. Hard disk is the storage medium which stores all components, data and documents of computer. Network is the essential component of high performance computing, and it couldn't be called parallel computing without network. Through network and its switch, computers with independent functions in different places and their external devices could be connected to each other, to extend the scale and capability of parallel computing effectively. In order to improve the efficiency of parallel computing and reduce the communication times between computers, the high performance computing generally use gigabit network, ten gigabit network, Infiniband network or optical network.

3.2 Platform system architecture

3.2.1 Platform requirement and design objective

(1). Hardware part

On the whole performance design, according to the users' demand, the theoretical floating point computing of platform is greater than or equal to 100Tflops (not take the computing power of CPU on CPU node and GPU into account, parallel efficiency is greater than or equal to 70 percent), speed of computing network is greater than or equal to 56Gbps, actual storage capacity is greater than or equal to 1000TB, IO measured bandwidth performance is greater than or equal to 15GB/s, as the standard, to extend based on users' other demand.

According to the design objective, it chooses the knife rack server as computing node, duplex computing node, and chooses Intel Xeon E5-2600 V3 series processors, memory is greater than or equal to 64GB, storage is greater than or equal to 300GB. It equips two fat nodes (SMP), adopts four Intel Xeon E7-4850 V3 processors, memory is greater than or equal to 1024GB, and storage is greater than or equal to 1000GB. It equips a GPU speedup node, rack-mounted server, at least two E5-2600v3 series processors, the maximum extension of memory slot onboard is 1TB DDR4 memory, it supports PCIE3.0 X16 slot, PCIE3.0 x8 slot.

(2). Software part

Software part mainly includes management software and application software. Management software: we should provide a commercial grade high-performance cluster management software which independents of hardware, with schedule, cluster management, monitoring, safety, charging and report form statistics functions; application software: the software which has been procured by school could use directly, once the system is running, we could purchase centralized based on users' actual demand, and share with the whole school.

(3). Computer room safeguard part

Computer room is the basic safeguard of high performance platform's safe operation. We should consider the following factors: circuit capacity, spatial arrangement, freezing capacity,

safety redundancy and late expansion. To plan and design system according to high standard with margins, to equip high efficiency air conditioners and power system and adopt air cooling channel layout, it can satisfy the requirement of the expansion next five years.

3.2.2 Platform system architecture

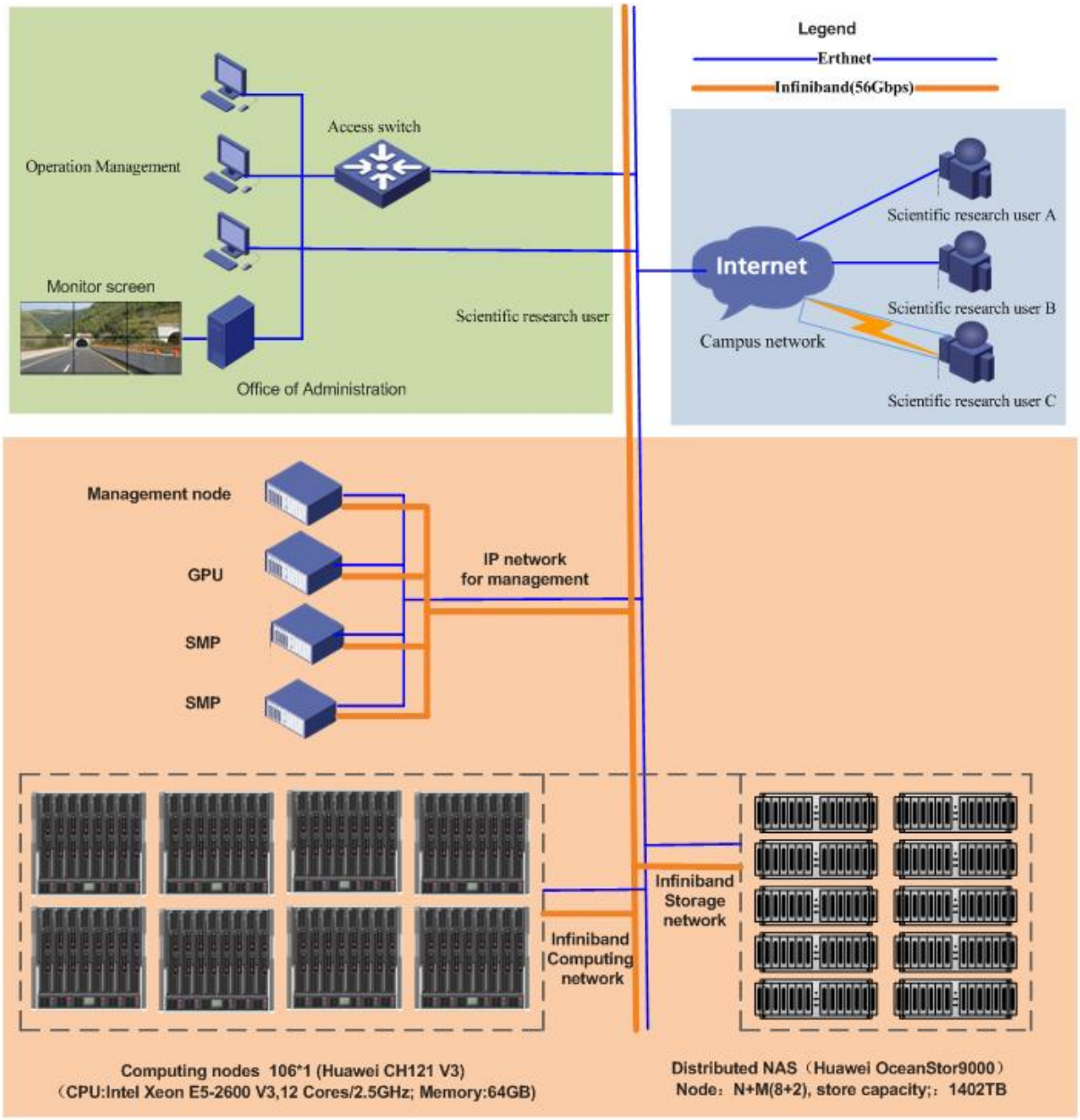

Figure 1 HPC school-level platform system architecture

Platform is formed by 106 ordinary computing nodes, one GPU node, one or two SMP nodes. The networking mode: IB whole line of speed fat tree, CPU is E5-2680 V3*2, memory is 8G DDR4*16, hard disk is 300G SAS, network card is GE*4, high speed card is FDR Infiniband*1, the operating system of all nodes adopts Redhat6.5, the MPI software environment is Intel MPI

5.0. Job scheduling, cluster management and cluster supervisory adopt cluster management system: JH UniScheduler of JH Innovation Software Company. The compositive system architecture shown as Figure 1.

4. HPC PLATFORM REALIZATION

4.1 Hardware planning and installation

(1). Equipment layout planning;
(2). Equipment installation.

4.2 Network build

(1). Oceanstor9000 storage node network;
(2). Cluster IB network connection;
(3). Cluster Ethernet network connection;
(4). Cluster IPMI management network connection.

4.3 Cluster installation and initialization after network connection

(1). Modify the IPMI address of host hardware;
(2). Operating system installation of all servers adopts Redhat6.5, x86_64;
(3). Modify the host name;
(4). Install the IB card driver;
(5). Deploy the ip address of eth0 and ib0;
(6). Close iptables、ip6tables、NetworkManager、sendmai and SELinux;
(7). Configure ssh and rsh to login without password;
(8). Synchronize /etc/hosts chart;
(9). Set the initial root username and password of cluster.

4.4 Cluster management and scheduling system installation

(1). Install JH portal application software;
(2). Install PostgreSQL database (9.1 edition) needed for platform;
(3). Install cluster safety control module JH appform;
(4). Install scheduling software module JH UniScheduler;
(5). When finished with the installation, save the server configuration and restart the server.

5. HPC PLATFORM TEST AND APPLICATION

5.1 Platform test situation

5.1.1 Linpack test situation

The test environment of hardware: Huawei FusionServer CH121 V3 (E9000 blade box)*106, networking mode: IB whole line of speed fat tree, CPU is E5-2680 V3*2, memory is 8G DDR4*16, hard disk is 300G SAS, network card is GE*4, high speed card is FDR Infiniband*1, MPI software environment is Intel MPI 5.0, test program is Intel Linpack test binary file (download URL: https://software.intel.com/en-us/articles/intel-mkl-benchmarks-suite).

Test configuration parameter shown as follows:

```
export MPI_PROC_NUM=212   //designate 106 nodes running 212 processes
export MPI_PER_NODE=2     //designate each node two processes
export NUMMIC=0           //designate node without MIC coprocessor
```

mpirun -f /stor9000/newlin2/hosts.txt -perhost ${MPI_PER_NODE} -np ${MPI_PROC_NUM} ./runme_intel64_prv "$@" | tee -a $OUT //actual order, this method is process add thread, each node two processes, each process judges CPU cores to distribute thread for itself, stress runs at full capacity.

HPL.dat configuration files' parameter setting shown as follows:

```
HPLinpack benchmark input file
HPL.out      output file name (if any)
8            device out (6=stdout, 7=stderr,file)
820000  Ns
1            # of NBs
192          NBs
1            PMAP process mapping (0=Row-, 1=Column-major)
1            # of process grids (P x Q)
53           Ps
4            Qs
16          threshold
1            # of panel fact
1            PFACTs (0=left, 1=Crout, 2=Right)
1            # of recursive stopping criterium
4            NBMINs (>= 1)
1            # of panels in recursion
2            NDIVs
1            # of recursive panel fact.
1            RFACTs (0=left, 1=Crout, 2=Right)
1            # of broadcast
```

6 BCASTs (0=1rg, 1=1rM, 2=2rg, 3=2rM, 4=Lng, 5=LnM, 6=Psh, 7=Psh2)
1 # of lookahead depth
0 DEPTHs (>=0)
0 SWAP (0=bin-exch, 1=long, 2=mix)
1 swapping threshold
1 L1 in (0=transposed, 1=no-transposed) form
1 U in (0=transposed, 1=no-transposed) form
0 Equilibration (0=no, 1=yes)
8 memory alignment in double (> 0)

The HPL performance test result of platform 106 nodes (see Figure 2) is 79.5784Tflops.

Intel E5-2600 v3 processor has two frequencies (E5-2600 v3 processor frequency introduction), Rated Base frequency 2.5GHz and AVX Base frequency 2.1GHz. Theoretical peak is equal to dominant frequency times cores times single cycle instruction count (IPC), the theoretical peak based on Rated Base frequency is 101.76Tflops.

Linapck efficiency = measured value/Rate Base theoretical peak = 79.5784/101.76 = 78.2 percent.

The test result shows the Linpack efficiency of platform 106 nodes cluster is 78.2%, achieves the requirement of platform expectant design object, Linpack efficiency ≥75 percent.

```
--------------------------------------------------------------------------------

- The matrix A is randomly generated for each test.
- The following scaled residual check will be computed:
      ||Ax-b||_oo / ( eps * ( || x ||_oo * || A ||_oo + || b ||_oo ) * N )
- The relative machine precision (eps) is taken to be            1.110223e-16
- Computational tests pass if scaled residuals are less than                16.0

================================================================================
T/V                N    NB     P     Q               Time                 Gflops
--------------------------------------------------------------------------------
WC06C2C4      820000   192    53     4            4619.09           7.95784e+04
HPL_pdgesv() start time Wed Mar 30 20:41:25 2016

HPL_pdgesv() end time   Wed Mar 30 21:58:24 2016
```

Figure 2 HPL performance test result of platform

5.1.2 Storage collective bandwidth test situation

(1). Preset conditions of test:

a. OceanStor9000 deploy has completed, configure static domain name and export NFS sharing;
b. Prepare 60 clients, and each client has installed IQzone test tool;
c. Every three clients share one 9000 static front-end business IP for mount, and each client uses different sharing in 9000.

d. Install intercommunication between nodes of 60 clients.

(2). Test process as follows:

a. Edit nodelist, and record the node name, mount point path and iozone software path of 60 clients.

b. Use command on one node: ./iozone -i 0 -i 1 -w -r 1m -s 128G -Recb /root/iozone_test_128.xls -t 60 -+m /root/nodelist –c to finish write, rewrite, read, reread test o cluster. 128G means single test file's size, /root/nodelist is absolute path of nodelist. -t 60 shows the number of client is 60, /root/iozone_test_128.xls is result exportation.

The test result shows read bandwidth reached 15GB/s, write bandwidth reached 13GB/s, has achieved the requirement of platform expectant design object.

5.1.3 Union test situation of platform operating

When the platform was built, we have tested the application softwares such as: WRF and Gromacs based on MPI operating, and Matlab, Adina, blat and bowtie which are single node operating softwares, integration and debugging are successful.

5.2 Platform application situation

After five months of commissioning, there are 66 research teams, a total of 158 users open account, the platform system availability is about 87.25 percent. The applications provided by platform include following aspects:

(1). Climatic simulation: WRF, CESM, CFS, etc.

(2). Biological information: blast, boost, bwa, blat, bowtie, bedtools, etc.

(3). Crop simulation: DSSAT;

(4). Molecular dynamics: Gromacs;

(5). Engineering computing: Matlab, Gaussian, Adina, etc.

(6). Artificial intelligence: Matlab, R, Python, etc.

6. CONCLUSION

This paper provides a new method and thought to build school-level high performance computing platform, aimed at the problems in high performance computing platform building. Combined with the actual needs of university, we adopt distributed storage system OceanStor9000 which has characteristics of high extension, high reliability, high performance and easy management; storage and computing network are built based on 56Gb Infiniband network, with characteristics of high bandwidth and low latency; the whole operating system is based on open-source linux, cluster scheduling and management system adopt JH resource management and scheduling software management system of JH Innovation Software

Company, which have good system compatibility, scalability and stability. Through test, the platform achieves the design object. The platform is used in weather prognosis, biological information and engineering computing application, the operating result shows that the platform has realized most of university's needs for high performance computing, reduced the computing difficulty of researchers greatly, shortened the running time, which has high practical value. The high performance computing platform will improve the innovation method and capability of the university in major national requirements and academic foreland field, such as: ecological environment, climatic change, agricultural water research, biological information and high polymer chemistry.

REFERENCES

[1] LI Bo, CAO Fu-yib, WANG Xiang-feng. Simple sketch of high performance computing technologies [J], Journal of Shenyang Institute of Engineering (Natural Science), 2012, 8(3):252-254 (in Chinese)

[2] WANG Bin.A typical type of high-performance computation: earth system modeling [J]. Physicals, 2009, 38(8):569-574(in Chinese)

[3] Zhang Juhua, Zhang Shengtao, Chan Lianyu. WANG Bin. Development Situation and Trend of high-performance computation: [J].Oil Geophysical Prospecting, 2010, 45(6):918-925(in Chinese)

[4] Fryza T, Svobodova J, Adamec F. overview of parallel platforms for common high performance computing [J]. Radioengineering, 2012, 21(1):436-444

[5] Van DBF, Wesseles KJ, Miteff S. HiTempo:a platform for time-series analysis of remote-sensing satellite data in a high-performance computing environment[J]. International journal of remote sensing, 2012, 33(15):4720-4740

[6] Li Jin-long. Research on Application of Image Matching Based on HPC [D], Master's degree thesis of Peking University.(in Chinese)

[7] Li Quan, Guo Zhaodian, Deng Yiju, Liao Zhenrong. CFD high-performance computing in aeronautics [J]. Huazhong Univ. of Sci. &Tech. (Natural Science Edition), 2011, 39(A1):79-82(in Chinese)

[8] PAN Sha, LI Hua, XIA Zhi-xun. High-performance Computing Application for Aerospace CFD Numerical Simulation [J]. Computer Engineering & Science, 2012, 34(8):191-198 (in Chinese)

[9] Shen Wen-hai, Discussion on application prospect of Grid computing on Meteorological high performance computing[J].Advances in Meteorological Science and Technology, 2012 , 2 (1): 48-51 (in Chinese)

[10] Zhang Yu-ting. Promoting research of Theoretical Chemistry Study on High Performance Computing [J]. Computer World, 2012, 24:1-1 (in Chinese)

[11] WU Jia-ni, LIU Lu, CHEN Luo. Parallel Implementation Approach for the GeoComputation Service Process in the High-performance Computing Environment [J]. Computer Science, 2012, 39(11):111-115 (in Chinese)

[12] Zeng Shu, Application of High Performance Computing Technology in the design of underwater-vehicle [J]. Mine Warfare & Ship Self-Defence, 2009, 17(2):40-43(in Chinese)

[13] Li Lu, Chen Bao-guo, Configuration of MPI Parallel Environment Based on Linux[J].Computer and Digital Engineering, 2007,35(11):47-48(in Chinese)

[14] Ye Mao, Miu Lun, Wang Zhi-zhang. Construction and Performance evaluation of High-performance Computing clusters Based on Linuxand MPICH2 [J], Journal of China Institute of Water Resources and Hydropower Research, 2009, 7(4):302-306(in Chinese)

[15] Chi Xue-bin, High-performance Parallel computing [R]. Computer Network and Information of Chinese Academy of Sciences, 2007 (in Chinese)