Research on performance dependence of cluster computing system based on GPU accelerators on architecture and number of cluster nodes

D. Akhmedov, S. Yelubayev, T. Bopeyev, F. Abdoldina, D. Muratov, R. Povetkin

Institute of space technique and technology, Almaty, Republic of Kazakhstan

akhmedov.d@istt.kz, s.elubaev@gmail.com, bopeyev.t@mail.ru, farida_mail@mail.ru, m_daulet@mail.ru, povetkyn.r@gmail.com

Abstract. *This paper presents results of computing experiments for verifying correctness of the choice and clarification of technical solutions for hybrid computing system based on GPU accelerators. Explained results of testing performance of hybrid computing system consisting of two and three nodes in Linpack benchmark. It was changing the number of GPU accelerators Nvidia Tesla M2090 for each node during testing. Also defined optimal values of RAM for six variants of hybrid computing system.*

Keywords

Parallel computing, high-performance computing, cluster computing system, GPU accelerator, CUDA technology.

1 Introduction

Appeared possibilities of modern component base for parallel computing and especially dramatic increasing performance of high-end GPU accelerators make possible independent development of supercomputers in Kazakhstan. One of the perspective ways of development supercomputer technologies is to use computational capability of GPU accelerators Nvidia Tesla for development personal and cluster computing systems. Cluster solutions based on hybrid structure is optimal by cost/performance criteria.

Cluster computing system (CCS) based on GPU accelerators is cluster net with nodes which represent *personal computing systems* [1], adapted for work in cluster.

Within the budgetary project “Development of cluster computing system based on GPU accelerators” of Ministry of Education and Science of Republic of Kazakhstan developed experimental model of CCS based on GPU accelerators Nvidia Tesla. The next stage of engineering development of CCS is to create prototype of CCS. For this purpose on the first step performed tests experimental model of CCS for verifying correctness of the choice of technical solutions and clarification some technical characteristics. The following steps of design of CCS include development of all necessary documentation for creation and testing CCS with a peak performance more than 8Tflops of double precision and 16 TFlops of single precision.

2 Experimental model of cluster computing system based on GPU accelerators

Experimental model of CCS is made on the Beowulf cluster technology. The feature of such cluster is scalability, means that it is possible to increase performance of computing system by increasing amount of computing nodes. Amount of nodes of CCS is limmited by the model of switch Mellanox Infiniscale IV which allows to create cluster system containing from 2 to 8 nodes.

To provide testing of CCS it was made 6 configurations of it: 1 – two nodes with two GPU accelerators, 2 - two nodes with three GPU accelerators, 3 - two nodes with four GPU accelerators, 4 - three nodes with two GPU accelerators, 5 - three nodes with three GPU accelerators, 6 - three nodes with four GPU accelerators.

Each node consist of following items: CPU Intel Xeon E5-2620 2,0 GHz; motherboard Supermicro X9DRG-QF; GPU accelerators Nvidia Tesla M2090 (amount depends on CCS configuration), 32Gb DDR3 RAM and network adapter Mellanox ConnectX-2 VPI.

Experimental model of CCS running an open-source operating system – Linux RHEL 6.1. Message passing

interface (OpenMPI ver. 1.6) is used for distribution of computational load between nodes. Specialized software: parallel computing platform Nvidia CUDA 5.0 for acceleration scientific and engineering tasks; library of mathematical application software Intel® Math Kernel Library (Intel® MKL 10); resource and queue manager for cluster systems TORQUE. As a benchmark is chosen Cuda Accelerated Linpack 2.0 v16 [2] as it is used for making list of top supercomputers all over the world and countries of CIS.

3 Tests results and its analysis

In this section we made analysis and tried to justify tests results as well as define optimal value of RAM for every CCS configuration.

Experimental model of CCS consisting of two nodes

Peak performance of 1st CCS configuration is 2640GFlops (double precision), 3960GFlops for 2nd configuration and 5280Gflops for 3rd configuration. Tests allowed to define actual CCS performance dependence of it's RAM.

Table 1 contains tests results which include values for actual CCS performance, efficiency of CCS and cost-performance relationship. Specific GFlop cost reduces as CCS efficiency grows up. Cost of GFlop is presented in Kazakhstan's currency – KZT, which is equal 153.5 US dollars at the time of this writing.

Table also contain value of RAM, used by CCS during computations. RAM usage depends on Linpack's task dimension.

Tab.1. Tests results for 1st, 2nd and 3rd CCS configurations

RAM, Gb	Task dimension	1st CCS configuration			2nd CCS configuration			3rd CCS configuration		
		Actual performance, GFlops	Efficiency, %	Specific GFlop cost, KZT	Actual performance, GFlops	Efficiency, %	Specific GFlop cost, KZT	Actual performance, GFlops	Efficiency, %	Specific GFlop cost, KZT
4,2	14 273	540,2	20,5%	2 536,1	479,0	12,1%	4 279,7	299,9	5,7%	9 103,0
6,1	20 822	757,7	28,7%	1 808,1	739,2	18,7%	2 773,3	505,5	9,6%	5 400,6
...	...	...	...	...	...	...	...	...	...	...
42,7	69 715	1 367,0	51,8%	1 035,1	1 673,0	42,2%	1 252,2	1 676,0	31,7%	1 655,7
46,3	72 930	**1 380,0**	**52,3%**	**1 029,0**	1 707,0	43,1%	1 230,2	1 779,0	33,7%	1 562,7
50,1	76 146	1 370,0	51,9%	1 040,1	1 708,0	43,1%	1 232,4	1 830,0	34,7%	1 521,9
54,7	79 361	1 367,0	51,8%	1 046,1	1 739,0	43,9%	1 213,3	1 865,0	35,3%	1 496,0
58,8	82 577	1 350,0	51,1%	1 063,0	**1 750,0**	**44,2%**	**1 208,6**	**1 923,0**	**36,4%**	**1 453,5**

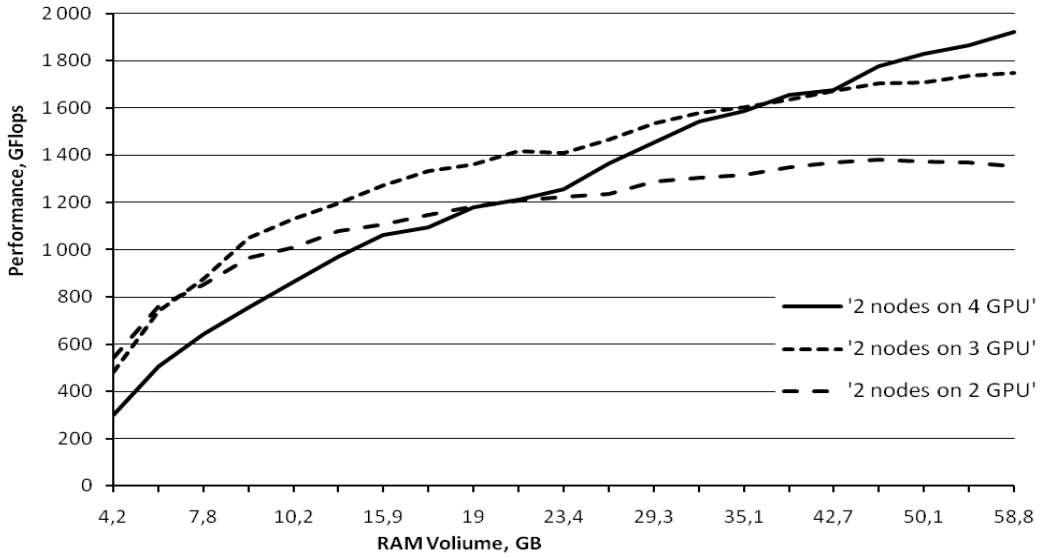

Fig. 1. Performance comparison for 1st, 2nd and 3rd CCS configurations

Comparing tests results for first three CCS configurations, it's clear that highest performance in Linpack is 1923Gflops for 3rd configuration (Figure 1). However highest CCS efficiency, which is 53.2% reached in 1st configuration, where only two GPU accelerators Nvidia Tesla M2090 in each node. Efficiency of 2nd CCS configuration is 44.2%, 3rd – 36.4%.

Obtained results for CCS efficiency as we may consider is within the normal range, since the powerfull supercomputers in the world have efficiency from 27.04% to 71.4% according to TOP500 project [3]. For example, supercomputer on the 100th place of the TOP500 with peak performance 1049.5TFlops showed actual 293.9TFlops

performance and 28% efficiency. This supercomputer contains 23040 GPU accelerators Nvidia Tesla M2090. Another supercomputer on 113 place with peak performance 373.9TFlops that have 3328 GPU accelerators Nvidia Tesla M2090 reached 71.4% efficiency.

After analysis of results we can say that actual CCS performance directly depends on RAM. The CCS performance grows up as RAM increases. However this process is not infinite and it is clear from 1st CCS configuration. Performance reached it's maximum when RAM is 46.3Gb and further increasing of RAM does not lead to growing up performance. We can make conclusion that for 1st CCS configuration optimal value of RAM is 46.3Gb.

As for 2nd CCS configuration the maximum performance is not reached. According to figure 1 performance practically took horizontal direction and would reach it's maximum point after several steps of increasing RAM. To predict actual CSS performance with increasing RAM we defined logistic trend with function $y = 427.03 \cdot 0.76 \cdot \ln(x) + 442.05$. Using this equation we defined optimal RAM volume equal 90Gb when efficiency of 2nd CCS configuration would reach 48.0%.

Analyzing the performance dynamics of 3rd CCS configuration we can conclude that it's not enough RAM to reach higher performance and efficiency. Increasing RAM at least up to 128Gb would give up to 43.0% efficiency. For prediction actual performance of 3rd CCS configuration it was used logistic trend function $y = 558.06 \cdot 0.82 \cdot \ln(x) + 57.358$.

Experimental model of CCS consisting of three nodes

The peak performance for 4th CCS configuration is equal to 3960GFlops (double precision), 5th – 5940Gflops and 6th – 7920Gflops. The overall CSS RAM was changing from 4.2Gb up to 85.1Gb. Each node has approximately 33% of overall RAM volume. Tests results are presented in table 2.

As we can see the highest performance in Linpack (2622GFlops) reached the 6th CCS configuration (Figure 2). However the highest efficiency, which is equal to 48.3% reached the 4th CCS configuration, where each node has 2 GPU accelerators Nvidia Tesla M2090. The maximum efficiency of 5th CCS configuration is 40.8% and for 6th – 33.1%.

Tab.2. Tests results for 4th, 5th and 5th CCS configurations

RAM Gb	Task dimen-sion	4th CCS configuration			5th CCS configuration			6th CCS configuration		
		Linpack perfor-mance, GFlops	Effici-ency, %	Specific GFlop cost, KZT	Linpack perfor-mance, GFlops	Effici-ency, %	Specific GFlop cost, KZT	Linpack perfor-mance, GFlops	Effici-ency, %	Specific GFlop cost, KZT
4,2	14 273	431,8	10,9%	6 970,8	229,2	3,9%	19 677,1	433,6	5,5%	13 860,7
...	...	...	...	...	...	...	...	...	...	...
58,8	82 577	1 808,0	45,7%	1 700,8	2 176,0	36,6%	2 102,5	2 385,0	30,1%	2 547,2
62,4	85 793	1 807,0	45,6%	1 704,5	2 278,0	38,4%	2 010,5	2 429,0	30,7%	2 503,1
66,8	89 008	1 816,0	45,9%	1 698,8	2 271,0	38,2%	2 018,9	2 437,0	30,8%	2 496,9
71,3	92 224	1 799,0	45,4%	1 717,6	2 295,0	38,6%	2 000,0	2 519,0	31,8%	2 417,6
75,3	95 439	1 860,0	47,0%	1 664,0	2 390,0	40,2%	1 922,6	2 569,0	32,4%	2 372,5
80,9	98 655	**1 912,0**	**48,3%**	**1 624,0**	2 396,0	40,3%	1 922,0	2 608,0	32,9%	2 340,9
85,1	101 870	1 897,0	47,9%	1 639,4	**2 421,0**	**40,8%**	**1 904,2**	**2 622,0**	**33,1%**	**2 330,3**

Performance saturation is for 4th CCS configuration is reached at 80.9Gb of RAM and further RAM increasing doesn't lead to growing up the CCS performance. For this CCS configuration optimal value of RAM is 80.9Gb with respect to cost-performance relationship. This result confirms the assumption (which was made in [4]) that each node has to have at least 12Gb of RAM.

As for 5th and 6th CCS configurations, the performance saturation is not reached. Using the logistic trend function $y = 758.02 \cdot 0.769 \cdot \ln(x) - 185.3$ defined that actual performance would equal to 2643.0GFlops at 128Gb of RAM and efficiency would eaual to 44.5%. At 162Gb of RAM actual performance would grow up by 5% and will equal to 2780.3GFlops. Efficiency also would increase by 2.3% and will equal to 46.8%. So we can conclude that for this CCS configuration it is enought 128Gb of RAM, where each node has 14.2Gb of RAM.

It was used logistic trend function $y = 758.02 \cdot 0.769 \cdot \ln(x) - 185.3$ to predict actual performance of 6th CCS configuration. At 144Gb of RAM actual performance would grow up by 11% and will equal to 2920GFlops. Efficiency also would increase by 3.8% and will equal to 36.8%. Further increasing of RAM up to 216Gb will increase performance up by 7% and will equal to 3149.9GFlops. Efficiency would grow up by 2.9% and will equal to 39.8%. At 288Gb of RAM actual performance would equal to 3313GFlops i.e. 41.8% from peak performance. As we can see that CCS configuration require at least 216Gb of RAM, where each node would have 18Gb of RAM.

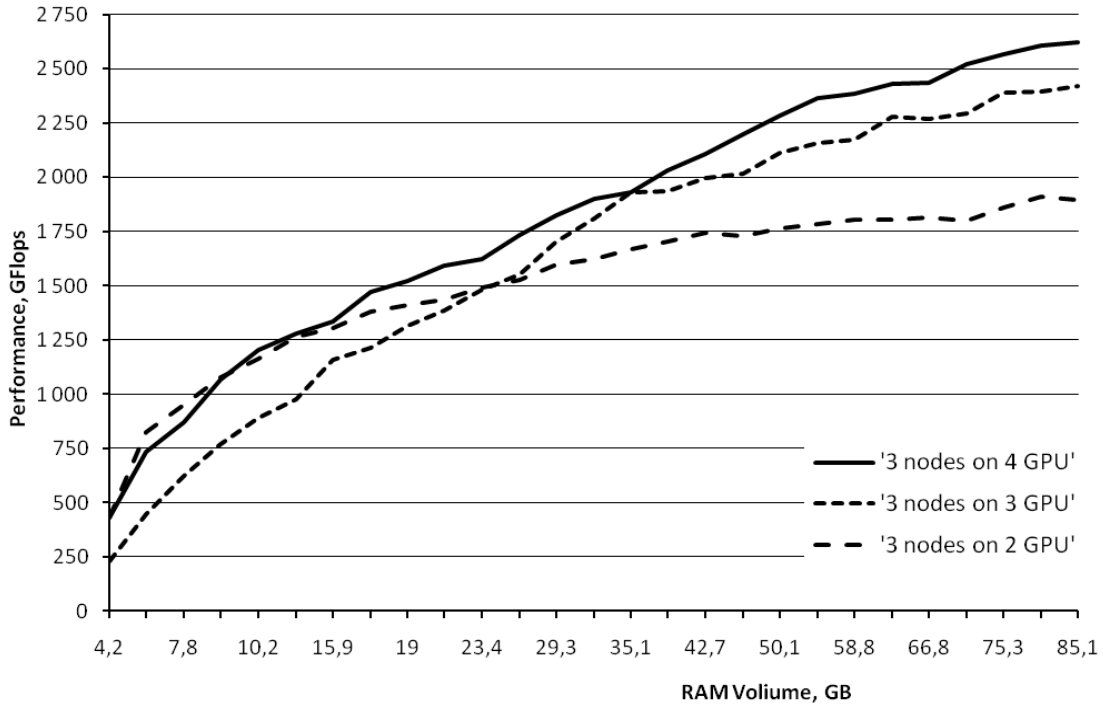

Fig. 2. Performance comparison for 4th, 5th and 6th CCS configurations

Conclusion. Logistic trend allows to approximatly define performance increasing but it doesn‘t allow to determine RAM volume where that performance reach it’s saturation. This requires additional experiments.

It can be assumed that for 2nd, 3rd, 5th and 6th CCS configurations RAM clock frequency can have influence on CCS performance. For these test RAM clock frequency was 1333MHz. Also more powerfull CPU can increase CCS performance.

CCS Scalability. For CCS configurations where each node include 2 GPU accelerators, adding 1 node lead to increasing CSS performance up by 34% and by 24% for configurations with 3 and 4 GPUs. This means that every additional node is able to increase CSS performans at least up by 24%.

4 Additional performance tests results

It was made additional tests to compare predicted values of CCS performance with it’s real performance with higher values of RAM. For this tests were selected two CCS configurations that consist of 2 and 3 nodes. Each node have 64Gb of RAM (16 RAM units of 4Gb).

Comparision of predicted efficiency for 3rd CCS configuration with experimental results shows the difference beetwin them equal to 0.38% (Table 3). Difference between experimental results and predicted efficiency for 6th CCS configuration is equal to 0.36% (Table 4). Results of this expirement allow to make conclusion that assumptions in previous sections is quite valid.

Tab.3. Comparison between predicted and real performance for 3rd CCS configuration

RAM, Gb	Task dimension	Experimental values of performance		Predicted values of performance	
		Linpack performance, GFlops	Efficiency, %	Performance, GFlops	Efficiency, %
62,4	85 793	2 007,0	38,0%	1948,9	36,9%
66,8	89 008	2 010,0	38,1%	1980,1	37,5%
71,3	92 224	2 043,0	38,7%	2009,9	38,1%
75,3	95 439	2 031,0	38,5%	2034,9	38,5%
80,9	98 655	2 066,0	39,1%	2067,7	39,2%
90,6	101 870	2 121,0	40,2%	2119,6	40,1%
109,8	111 517	2 191,0	41,5%	2207,5	41,8%
115,8	114 733	2 238,0	42,4%	2231,9	42,3%
121,2	117 948	2 236,0	42,3%	2252,7	42,7%
127,2	121 164	2 240,0	42,4%	2274,8	43,1%

Tab.4. Comparison between predicted and real performance for 6th CCS configuration

RAM, Gb	Task dimen-sion	Experimental values of performance		Predicted values of performance	
		Actual performance, GFlops	Efficiency, %	Performance, GFlops	Efficiency, %
90,6	101 870	2 622,0	33,1%	2657,3	33,6%
98,4	105 086	2 646,0	33,4%	2704,2	34,1%
105	108 301	2 779,0	35,1%	2741,0	34,6%
109,8	111 517	2 792,0	35,3%	2766,3	34,9%
115,8	114 733	2 845,0	35,9%	2796,5	35,3%
121,2	117 948	2 839,0	35,8%	2822,3	35,6%
127,2	121 164	2 877,0	36,3%	2849,7	36,0%
133,2	124 379	2 892,0	36,5%	2875,8	36,3%
140,7	127 595	2 898,0	36,6%	2906,9	36,7%
146,7	130 810	2 896,0	36,6%	2930,6	37,0%
152,7	134 026	2 943,0	37,2%	2953,3	37,3%
159,3	137 241	2 969,0	37,5%	2977,3	37,6%
166,5	140 457	2 979,0	37,6%	3002,4	37,9%
173,4	143 673	2 997,0	37,8%	3025,4	38,2%
183	146 888	3 004,0	37,9%	3055,9	38,6%

5 Conclusion

Performed tests allowed to define optimal values of RAM for 1st and 4th CCS configurations and enough values or RAM for 2nd, 3rd and 5th CCS configurations with respect to number of GPU accelerators in CCS nodes. It was made conclusions about configurations CCS that will allow to reach higher performance and efficiency.

References

[1] Akhmedov D. Personal hybrid computing system. Performance test / Akhmedov D., Yelubayev S., Abdoldina F., Bopeyev T., Muratov D., Povetkin R., Karataev A. // Collection of scientific papers Second International conference "Cluster Computing 2013", –Lvov, 2013. –P. 7-11. (in Russian)

[2] NVIDIA Developer zone: [Electronic resource]. URL: http://developer.nvidia.com.

[3] Top500 Supercomputer sites: [Electronic resource]. 2000-2013. URL: http://www.top500.org/

[4] Golovynskyi A.L. The architecture of the GPU- nodes of computing cluster / Golovynskyi A.L., Malenko A.L., Gorenko S.O., Bandura A.U. // International Conference "High-performance computing" HPC-UA'2011, –Kiev, 2011. –P. 70-75. (in Ukrainian)