Improved network equipment real-time monitoring system based on Ganglia

Yuanwen Wang

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China

ywwang2013@163.com

Keywords: Ganglia, Network Equipment, Monitoring, Fault Tolerance.

Abstract. The data fault tolerance has become an important performance indicator in a real-time monitoring system of network equipment. Thus the data fault tolerance has become an important performance indicators. We analyse the deficiencies of ganglia monitoring system and propsed an improved scheme for ganglia systems. The test result shows that the improved scheme can make the performance of the system more stable and efficients.

Introduction

Ganglia sponsored by the UC Berkeley is an open source monitoring project, designed to monitor thousands of nodes. It can monitor the current operating state information of each node in the cluster, such as: cpu, memory, disk utilization, I/O load, network traffic, etc. Furher more, the histrrical data can also be rendered with a curve way by php program[1].

Meanwhile, the system has good scalability allowing users to customize the resource information of the monitor system, which plays an important role for the reasonable adjustment, allocation of system resources and optimization of the whole system performance. And now it has been widely used in a variety of monitoring systems for data collection.

In the ganglia cluster, every host runs a daemon named gmond to collect and send monitor data. The data is collected from the operating system and the specified host. Hosts receiving all monitoring data can display these data and passed them to the hierarchy in simplified form. Precisely because of this level of architecture model, the ganglia can achieve good extension. The system load of gmond process is very small, which makes it a piece of code running in the cluster without impacting user performance on individual servers.

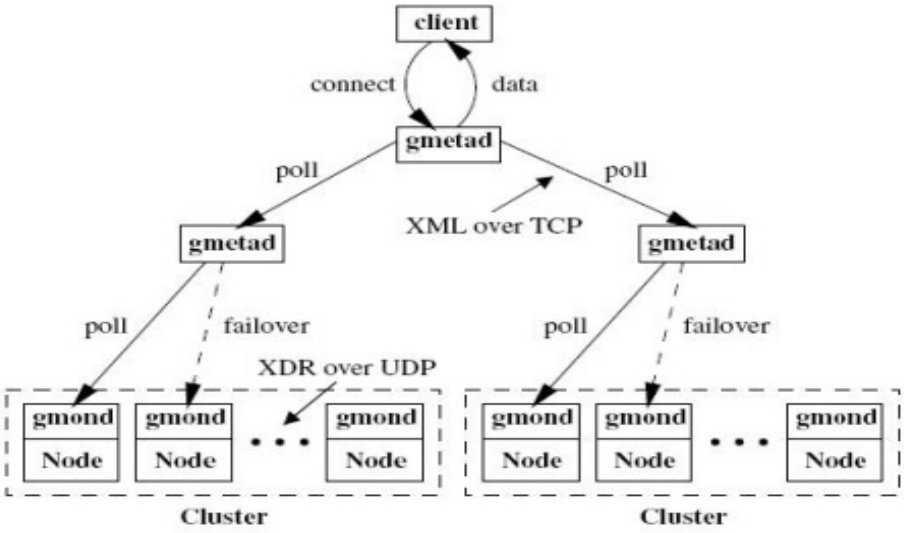

Fig. 1. Ganglia cluster function schematic

From the Fig. 1, we can see that the architecture of the Ganglia is a complex structure with multiple nodes, so the error of data transmission is difficult to be avoided totally[2][3]. Especially when an error exists in the top node in a tree structure, the bottom quark node data transfer is unable to complete, therefore the system in fault tolerance aspects of the urgent need for improved. If any of the

collection node or the line failed, the information of all the leaf nodes (i.e., the monitoring information of the corresponding cluster grid) will not be collected.

Related Problems

Deficiency about data of Ganglia system. The ganglia architecture of transfer data hierarchy lead to the situation when it transfers data layer up in each layer, if the current node is abnormal, for example, network anomaly or machine crash, web page for the current node and its child node data will cannot be effective monitoring. And it also can not make an assessment of the performance of the machine. So if an exception occurs at a critical time, it will cause an unexcepted loss in the normal operating system.

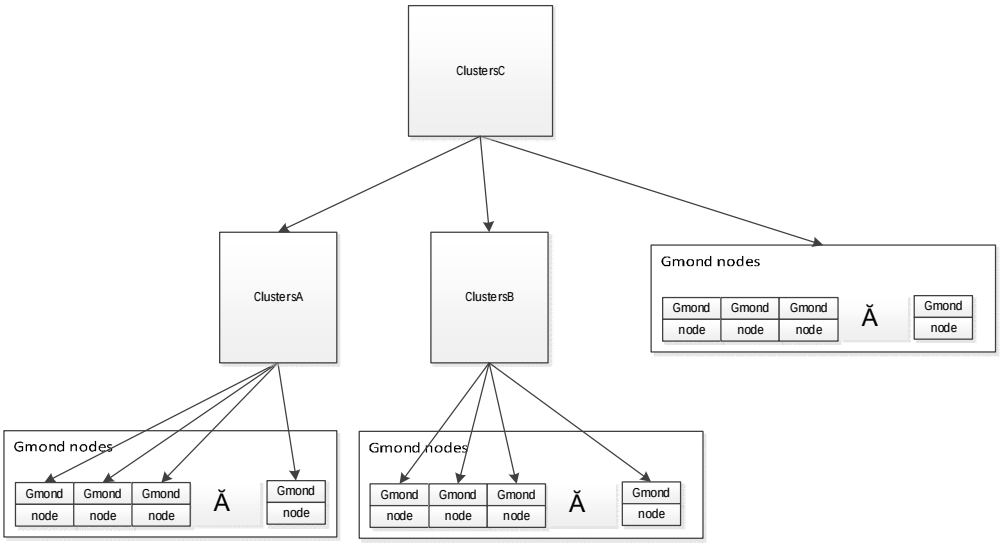

Fig. 2 Schematic diagram of Ganglia test cluster

In Fig. 2, according to the above analysis, if the network in cluster A interruptes abnormally or the machine goes down, monitoring data of a Gmond sub node in the cluster A can not be timely showed in web pages or is not effective collection, statistics and analysis, which maybe causes unpredictable consequences.

Review of the orginal scheme. The article[3] **Error! Reference source not found.**proposed in the monitored grid cluster system, there are at least two different lines between any two nodes to ensure that in cluster A the network anomaly or hosts downtime will not appear on the data missing and the state information of the child nodes in the cluster A cannot be displayed through web page.

The advantage of the fault tolerance scheme, for this enhancement system, is to start the backup link device to protect the physical interface or link when the nodes or lines in a sysstem are failure. But it also brings about inncreasement of the complexity and cost of system equipment management. In a large network system, the nodes in the grid can be hundreds or even thousands. To this point, the maintenance of the whole grid system is not guaranteed[4]. Further, for a complex network system, the complexity of the network is high already, the system will bocome high complexity, high redundancy if we add the link between the nodes to ensure the fault tolerance. This is not consistent with the system's simplicity and ease of maintenance, so it is necessary to further improve.

Improvement of monitoring system

We introduced a concept, “double live data”, which usually refers to two running machines, but they don’t provide services at the same time. When one machine is break down, another one will take over and provide services automatically during a very short time[4].

We uses the keepalived cluster to achieve "double live data". Working principle of keepalived is VRRP (Virtual Router Redundancy Protocol) which is similar to "replicate set", a multiple copies to ensure the fault tolerance[5]. Even if one copy doesn’t work and there are many other copies to switching to, within a cluster based on its election system.

In Keepalived cluster, there is no strict master node so each gmetad node in the system can achieve "double live data". And this can prevent a single point of failure (one point failure will lead to part or even the entire system architecture unavailable)[6][5]. At the same time, it can monitor data continously to get the cluster status in real time.

Experiment Results

Test environment deployment. Test environment deployment diagram in the paper as shown in Fig. 3.

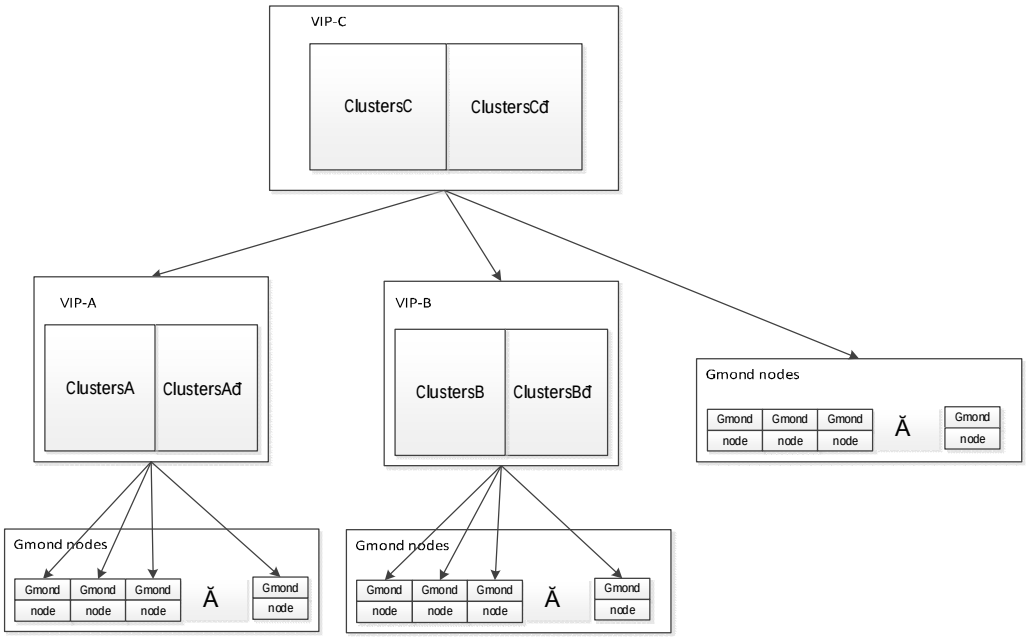

Fig. 3. Test environment schematic diagram

The same gmond adopt unicast node cluster testing. Clusters A/B/C Gmetad is data acquisition nodes, we set virtual IP: VIP-A, VIP- B, VIP-C. corresponding to clusterA/A', clusterB/B', clusterC/C' respectively through Keepalived.

Cluster Performance Test Comparison. First, during normal operation, clusterB/B' were both shut down for half an hour at the same time, in the entire grid cluster, the child nodes of the clusterB/B' will not upload data, which equivalent to no use "data double live"program, then test results are shown in Fig. 4. the entire cluster load and memory showed a" concave "shape, that is, page can not render the data of the child nodes in clusterB/B'.

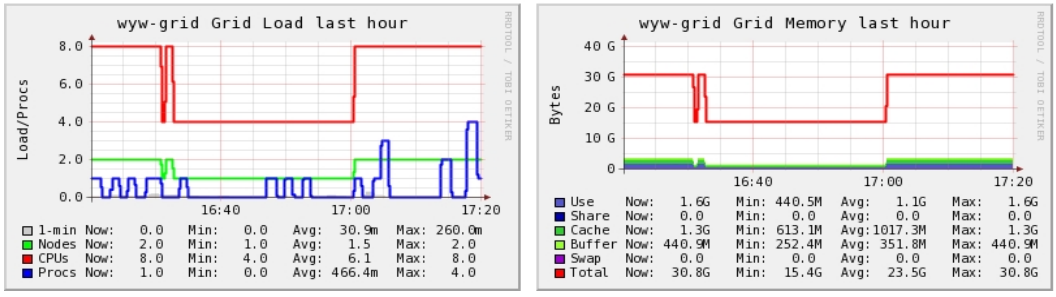

Fig. 4 [Before improvement] Monitor indicators of the entire cluste

Secondly, during normal system operation, only one machine between the cluster nodes clusterB/B'of (eg clusterB') goes down, that is a "data double live" program test, the test results showedn in Fig. 5 below, we can see that system load and memory were not affected, seamless handover between machine is achieved. and it ensure that the performance of the cluster can be monitored in real time data.

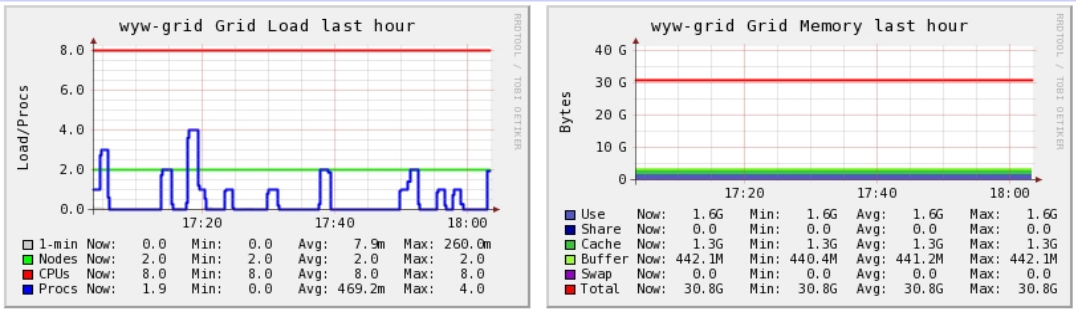

Fig. 5 [After improvement] Monitor indicators of the entire cluster

Stand-alone Performance Test Comparison. During the normal operation of system, while at the same time will clusterA/A'two machine down time at the same time, which is equivalent to not use "double live data" plan, the test results as shown in Fig. 6, but when A machine (such as clusterA') goes down forcely, so that using "dual machine scheme of hot standby system environment, test results are shown in Fig. 6 and Fig. 7.

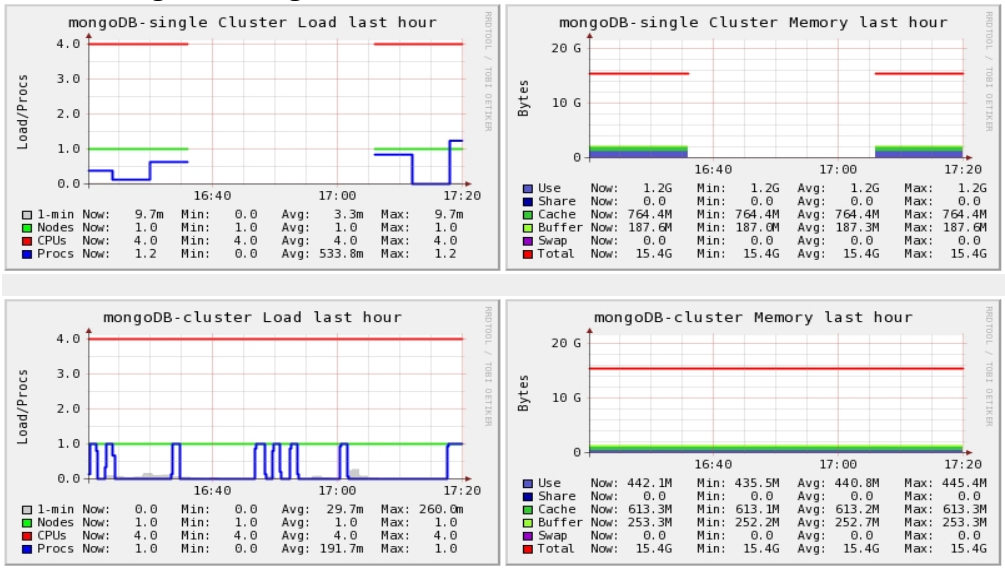

Fig. 6 [Before improvement] monitoring indicators of the single cluster

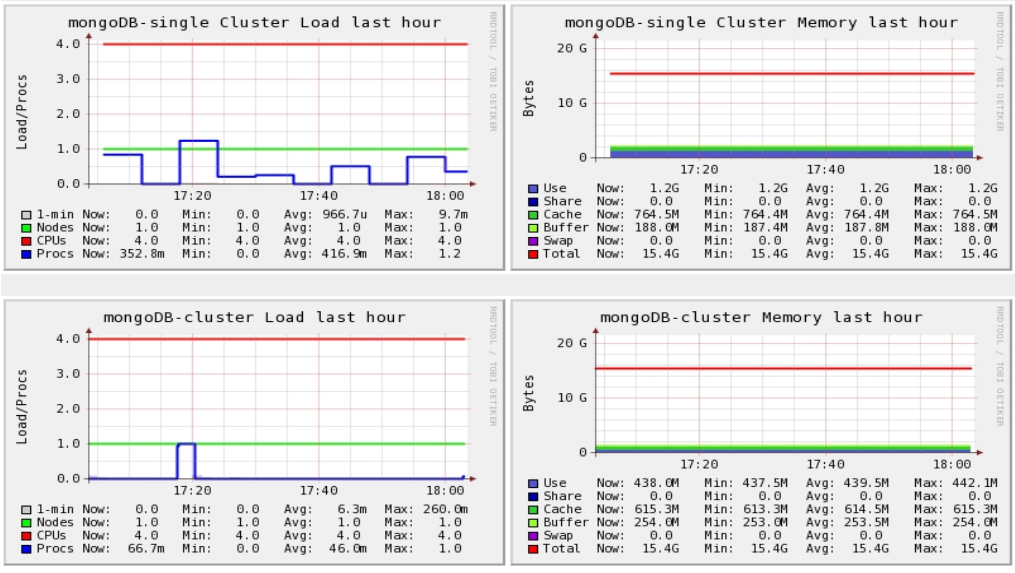

Fig. 7 [after improvement] monitoring indicatorsof the single cluster

Comparing the two figures, when it does not use "double live data", cluster A/A' child nodes cannot obtain data service from cluster C/C' Gmetad , simutaneously the web client page cannot monitor child nodes in the real-time, the loss caused by this cannot be estimated for a large network.

But useing "double live data", can avoided such problem so that, the normal function of the system running is guaranteed.

Conclusions

This paper discussed defects in transmission and fault tolerance of system at the data level, and put forward the scheme of "double machine hot standby in order to guarantee the real-time monitoring function of the system, then verified the feasibility and efficiency of the scheme. Test results show that the improved system can guarantee the stability of the Ganglia grid monitoring system and improve.

Acknowledgements

This work is supported by NSFC (Grant Nos. 61300181, 61502044), the Fundamental Research Funds for the Central Universities (Grant No. 2015RC23).

References

[1]. M. L. Massie, B. N. Chun, and D. E. Culler. “The Ganglia Distributed Monitoring System: Design, Implementation, and Experience”. Parallel Computing, 30(7), July 2004

[2]. Vyas R A, Prajapati H B, Dabhi V K. Embedding custom metric in ganglia monitoring system[C]// Advance Computing Conference (IACC), 2014 IEEE InternationalIEEE, 2014:793-797.

[3]. Li-Ping H E, Liu L C. Improved Grid Monitor System Based on Ganglia[J]. Journal of Guangdong University of Technology, 2006.

[4]. Bohm, S, Engelmann, C, Scott, S. L. Aggregation of Real-Time System Monitoring Data for Analyzing Large-Scale Parallel and Distributed Computing Environments[C]// High Performance Computing and Communications (HPCC), 2010 12th IEEE International Conference on2010:72-78.

[5]. Yang C T, Chen T T, Chen S Y. Implementation of Monitoring and Information Service Using Ganglia and NWS for Grid Resource Brokers[C]// Proceedings of the The 2nd IEEE Asia-Pacific Service Computing ConferenceIEEE Computer Society, 2007:356 - 363.

[6]. Sacerdoti, F.D., Katz, M.J., Massie, M.L., Culler, D.E.: Wide Area Cluster Monitoring with Ganglia. In: Proceedings of the IEEE Cluster 2003 Conference (2003)