Analysis of Server Resource Consumption of Meteorological Satellite Application System Based on Contour Curve

Xiangang Zhao[a], Cunqun Fan[b, *], Manyun Lin, Lizi Xie, Xi Zhang

National Satellite Meteorological Centre, Beijing, China
[a] email:xgzhao@cma.gov.cn, [b] email:fancq@cma.gov.cn
*corresponding author

Keywords: meteorological satellite; server resource; consumption analysis; contour curve

Abstract: In order to effectively analyze the system server resource consumption, this paper proposed a server resource consumption analysis method based contour curve. Firstly, the definition of resource consumption analysis was given, and then several analysis methods of resource consumption were given, including: application resource consumption profile, application resource consumption profile standard deviation, application resource consumption frequency, application resource consumption monotonicity and so on. Finally, based on this, the total consumption, consumption mean, kurtosis, skewness, reload, resource load and so on were analyzed. The method proposed in this paper can effectively analyze the server resource consumption in depth, improve the server resource utilization and reduce the server hardware loss to a certain extent.

1. Introduction

Meteorological satellite ground application system has a large amount of data traffic. The ground system is mainly responsible for receiving, processing and transmitting data. This requires a large amount of server resources to support. Due to limited server resources, how to properly analyze the consumption of server resources will be further optimization of server resources has a critical impact. Therefore, this paper studies the analysis of server resource consumption.

Some research has been done on the analysis of server resource consumption. Paper [1] proposed a linear regression method and a maximum likelihood technique for estimating the service demands of requests based on measurement of their response times instead of their CPU utilization. The paper [2] aims to address this problem by combining knowledge of database query processing with statistical models. Paper [3] experimentally investigates the relationship between the power consumption and resource utilization of a multimedia server cluster when different load balancing policies are used to distribute a workload. In paper [4], they study the problem of identifying an assignment scheme between the interacting components of an application, such as processes and virtual machines, and the computing nodes of a cloud system, such that the total amount of resources consumed by the respective application is minimized. The paper [5] tries to establish

relation between the power usage and software aging. Paper [6] give an experimental case study about resource consumption parameters and workload parameters in an Internet Information Services. They use fitted resource consumption parameter to learn the relationship between workload parameters and resource consumption parameters through visual observation and calculation.

2. Related Definition

Resource consumption analysis is to extract valuable statistical features and quantitative indicators for important parameters of each type of resources (CPU/memory/IO/network/NFS) of an application through a certain statistical method, so as to present the overall Apply a statistical pattern of behaviour during the run.

Each type of resource contains a variety of parameters. If you analyze each parameter, it will actually bring a higher amount of calculation, and in fact, some parameters are not meaningful even if they are analyzed. After analysis and screening, the application resource consumption analysis uses the following important parameter sets to collaborate on the analysis object:

a) CPU parameters: CPU_All_Sys, CPU_All_User, CPU_All_Idle, all in percentage.
b) Memory parameters: Mem_All_MemRatio, Mem_All_SwapRatio, all in percentage.
c) IO parameters: Disk_All_Read, Disk_All_Write, Disk_All_Svctm, in bytes (Bytes).
d) Network parameters: Net_All_Recv, Net_All_Send, in bytes (Bytes).
e) NFS parameters: NFS_All_Read, NFS_All_Write, in bytes (Bytes).

3. The Method of Resource Consumption Analysis

3.1 Application Resource Consumption Profile

The resource consumption profile is used to collectively show the overall level of occupying a resource when the target application is scheduled multiple times. For example, for a CPU resource consumption profile curve, the length of the curve represents the average time that the application runs when it is scheduled multiple times, and the point on the curve represents the mean value of CPU resources occupied by the application in the case of multiple scheduling at the current time point.

For the resource S of the target application X, when the application is scheduled N times, and the scheduling time of the i-th time is T_i, the total length of the resource consumption contour corresponding to the resource S is defined as $T=\min(T_1, T_2, \ldots, T_N)$, and the resource consumption contour value of the j-th point.

$$Y_j = \frac{X_{1j} + X_{2j} + \cdots + X_{Nj}}{N} \tag{1}$$

Where X_{ij} represents the value of the S resource parameter of the target application X at the j-th time point of the i-th scheduling.

For CPU resources, this parameter is the CPU active total CPU_All_Active = CPU_All_Sys + CPU_All_User; for memory resources, physical memory footprint Mem_All_MemRatio and swap area memory footprint Mem_All_SwapRatio will calculate the resource consumption profile; for IO, Parameters such as network and NFS are calculated by using the corresponding data read/write or network transmission and reception.

3.2 Application Resource Consumption Profile Standard Deviation

The standard deviation of the application resource consumption profile is used to show the fluctuation of the corresponding resources occupied by the target application at each time point during multiple scheduling. For a certain point in time, if the application is scheduled N times, when the standard deviation of the consumption profile corresponding to the time point is large, it indicates that the amount of the specified resource occupied by the N times of the application at the time point is discrete. The distribution may be due to unstable application implementation and large data changes.

For the resource parameter S of the target application X, when the application is scheduled N times, and the i-th scheduling time is T_i, the total length of the resource consumption profile is defined as $T=\min(T_1, T_2, ..., T_N)$, and the j point resource consumption profile standard deviation.

$$Y_j = D(X_{1j}, X_{2j}, \cdots, X_{Nj}) \tag{2}$$

Where X_{ij} represents the value of the resource parameter S of the target application X at the j-th time point of the i-th scheduling, and $D(X_1, X_2, ...)$ represents the standard deviation of the data.

3.3 Application Resource Consumption Frequency

The frequency of application resource consumption is based on the resource consumption profile, and analyzes the fluctuation pattern of the application resource consumption profile curve. By performing frequency domain analysis on the resource consumption profile curve, information such as the fluctuation amplitude and phase of the contour curve consumed at each frequency can be obtained. When the consumption profile curve has a small fluctuation amplitude at all frequencies, it indicates that the hardware resources occupied by the target application during the entire workflow flow show a stable trend, which may be that the resource is occupied or occupied for a long time; The fluctuation amplitude on the frequency is much larger than other frequencies, indicating that the target application occupies an unstable state of resources.

For the target application X, if the resource consumption profile curve of the corresponding resource S is Y, the frequency of consumption of the resource S is $rFFT(Y)$, where $rFFT$ is a real sequence fast Fourier transform.

3.4 Application Resource Consumption Monotonicity

The CPU resource consumption monotonicity curve is similar to the CPU resource consumption profile standard deviation. It is also a measure of the CPU resources occupied by the target application when it is scheduled multiple times at a certain point in time. Each point on the monotonic curve indicates whether the CPU usage of the target application presents a monotonous trend during multiple scheduling at that point in time. In theory, in the case of multiple scheduling, the CPU usage at each point in time should present a discrete random distribution, and should not exhibit significant monotonicity. If the CPU usage of the multiple scheduling at a certain point in time is obviously monotonous, there may be problems, such as busy garbage collection tasks and more space allocation.

For the target application X, when the application is scheduled N times and the i-th scheduling time is T_i, the total length of the CPU resource consumption profile is defined as $T=\min(T_1, T_2, ..., T_N)$, and the j-th point The value of the CPU resource consumption monotonicity curve $Y_j=MK(X_{1j}, X_{2j}, ..., X_{Nj})$, where X_{ij} represents the value of the CPU occupancy of the target application X at the j-th time point of the i-th scheduling, $MK(X, Y, ...)$ represents the monotonic trend of discrete sequences calculated using the Mann-Kendall algorithm.

The Mann-Kendall algorithm is described in detail as follows.

(1) For time series $X = (X_1, X_2, \cdots X_n)$, the statistics for the test are as follows.

$$S(X) = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}(x_j - x_k) \tag{3}$$

Where $\operatorname{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$.

(2) When n is large enough, it can be proved that S(X) obeys a normal distribution with a mean of 0 and a variance of $var(S) = n(n-1)(2n-5)/18$. When n>10, the standard normal system variable calculation expression is as follows.

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{\operatorname{var}(S)}}, & S > 0 \\ 0, & S = 0 \\ \dfrac{S+1}{\sqrt{\operatorname{var}(S)}}, & S < 0 \end{cases} \tag{4}$$

The positive or negative of Z indicates that the original time series is in a monotonous increasing trend or a monotonous decreasing trend; the size of |Z| represents the credibility of the monotonic trend.

When |Z|>0.8, the reliability of the monotonic trend of the original time series is 80%, and the original data is considered to be monotonous.

4. Resource Type Profiling

The resource type profile, which is a further statistical analysis based on the consumption profile curve of a resource, is an overview. For each type of resource, the calculated statistical characteristics include total resource consumption, mean, peak, concentration, reload, resource consumption load, etc., and specific statistical characteristics, such as throughput, for certain resources, occupancy, etc. Resource type profiles include total consumption, consumption mean, kurtosis, skewness, reload, resource load, and so on.

4.1 Total Resource Consumption

For the consumption profile X of a resource, the total consumption S is $S = \operatorname{SUM}(X_1, X_2, ..., X_n)$.

4.2 Mean Value of Resource Consumption

For the consumption profile X of a resource, the consumption mean E is $E = \operatorname{AVE}(X_1, X_2, ..., X_n)$, and AVE is an average function.

4.3 Resource Consumption Kurtosis

Kurtosis is also known as the kurtosis coefficient. Characterization of the peak height of the probability density distribution curve at the mean. Considering the resource consumption profile as

a distribution in probability space, with the normal distribution as the reference, the kurtosis can represent the concentration of the resource consumption distribution in a short time; under the same standard deviation, the greater the kurtosis coefficient The distribution has more extreme values, then the rest of the values must be more concentrated around the mode, and its distribution is bound to be steeper, that is, there is a situation in which the resource consumption is excessively concentrated in a short period of time.

For the consumption profile X of a resource, the consumption kurtosis is as follows.

$$K = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - u)^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - u)^2)^2} \tag{5}$$

Where u is the mean of resource consumption and n is the length of the consumption profile X.

4.4 Resource Consumption Skewness

The skewness coefficient is a feature number that describes the degree to which the distribution deviates from the symmetry. The resource consumption profile is regarded as a distribution on the probability space, with the normal distribution as the reference. When the distribution is bilaterally symmetric, the skewness coefficient is 0. When the skewness coefficient is greater than 0, that is, when the heavy tail is on the right side, the distribution is right-biased. When the skewness coefficient is less than 0, that is, when the heavy tail is on the left side, the distribution is left-biased. Intuitively, when the skewness coefficient is less than 0, the heavy tail is on the left side, indicating that the overall resource consumption is in the form of "quick and slow"; when the skewness is greater than 0, the heavy tail is on the right, indicating that the overall resource consumption is presented. The form of "slow first and then urgent".

For a consumption profile X of a resource, the resource consumption skewness is as follows.

$$W = \frac{1}{n}\sum_{i=1}^{n}(\frac{x_i - u}{\sigma})^3 \tag{6}$$

Where u is the mean of resource consumption, σ is the standard deviation of the consumption profile X, and n is the length of the consumption profile X.

4.5 Resource Consumption Concentration and Reload

The concentration of resource consumption indicates the proportion of time occupied by the high consumption of resources in the overall time; the resource consumption overload degree is similar to the concentration degree, which is the total consumption of resources in the overall time.

For a class of resources, when the job is running in the whole process, it is in a high-occupancy and high-consumption mode. This indicates that the corresponding hardware resources are under great pressure, and when the time is too long, there may be a risk of downtime.

For the consumption profile curve X of a resource, the resource consumption overload degree is as follows.

$$H = \sum_{i=1}^{n} x_i \cdot I\{x_i \geq Th\} \tag{7}$$

The resource consumption concentration is as follows.

$$C = \frac{1}{n}\sum_{i=1}^{n} I\{x_i \geq Th\} \tag{8}$$

Where n is the length of the consumption profile X and $I\{\cdot\}$ is the sign function.

$$I\{\cdot\} = \begin{cases} 1, & f = true \\ 0, & f = false \end{cases} \tag{9}$$

Th is a threshold, meaning that when the occupancy rate of the resource parameter is higher than the threshold, it is considered to be high occupancy; when the resource rate of the resource parameter is lower than the threshold, it is not considered to be high occupancy. The currently set threshold $Th = 80\%$, called 80% time concentration $C_{80\%}$. In general, the value of the threshold Th can be set to correspond to different time concentrations.

4.6 Resource Load

The resource consumption load is an indicator that quantifies the resource occupancy curve and measures the load of a resource per unit time during the running of the target job. For the occupancy curve of a certain type of resources, the resource consumption load analyzes the value of the curve at each time point. Based on the consideration of reducing the overall system load, the high occupancy is given a higher weight, and the lower occupancy is given a lower value. The weight value, such that the value calculated in the form of a weighted average is called the resource consumption load.

For the consumption profile curve X of a resource, and T is the length of the curve X, the detailed algorithm for applying the resource consumption load is as follows.

(1) Set the upper bound, lower bound and increment times of the algorithm according to the range of values of the resource parameters. For a percentage parameter such as CPU_All_Idle, Mem_All_MemRatio, the upper limit is generally set to 100, the lower bound is 0, and the increment is set to 100. For a parameter with a relatively large value of Net_All_Recv, the logarithm of the parameter value is generally first, and then The upper bound is 32, the lower bound is 0, and the increment is set to 100.

(2) According to the number of increments, the lower bound of the algorithm is uniformly increased to the upper bound of the algorithm. For each increment, the incremental threshold is calculated, and how many parameter values in the time period are higher than the current threshold. For example, if the current threshold is 50, you need to count the percentage of the parameter value greater than 50 in the current time period as a percentage of the total time. This percentage is the weight corresponding to the threshold.

The weight is the threshold obtained in step (2), and the incremental threshold is weighted. If the parameter value higher than 50 accounts for 20% of the total time, and the parameter value higher than 60 accounts for 10% of the total time, the weighted sum is 20%*50+10%*60=16, which is the resource load value.

The resource consumption analysis is centered on the resource consumption profile curve, and the corresponding statistical features are added. The job is considered to be relatively normal when the contour curve exhibits the following pattern.

1) The CPU resource consumption profile average is less than 60%, the CPU load is less than 2000; the memory resource consumption profile average is less than 75%, and the memory load is less than 3000.

2) The CPU/memory contour standard deviation curve mean value is 0, indicating that the resource consumption occupied by each scheduling of the job is consistent.

3) CPU/memory/IO resource consumption concentration is less than 70%

5. Conclusions

In order to analyze the server resource consumption of satellite ground application system, this paper proposes a server resource consumption analysis method based on contour curve. Through the definition of resource consumption analysis, several analysis methods of resource consumption are given, including: application resource consumption profile, application resource consumption profile standard deviation, application resource consumption frequency, application resource consumption monotonicity and so on. On this basis, the total consumption, consumption mean, kurtosis, skewness, reload, resource load, etc. are analyzed. The method proposed in this paper can effectively analyze the server resource consumption in depth, thereby improving server resource utilization and reducing server hardware loss to a certain extent.

Acknowledgements

References

[1] Dawson S, Dawson S, Dawson S, et al. Estimating service resource consumption from response time measurements[C]. International ICST Conference on PERFORMANCE Evaluation Methodologies and TOOLS. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009:48.
[2] Li J, Narasayya V, Chaudhuri S. Robust estimation of resource consumption for SQL queries using statistical techniques[J]. Proceedings of the Vldb Endowment, 2012, 5(11):1555-1566.
[3] Dargie W, Schill A. Analysis of the Power and Hardware Resource Consumption of Servers under Different Load Balancing Policies[C]. IEEE, International Conference on Cloud Computing. 2012:772-778.
[4] Tziritas N, Khan S U, Xu C Z, et al. On minimizing the resource consumption of cloud applications using process migrations[J]. Journal of Parallel & Distributed Computing, 2013, 73(12):1690-1704.
[5] Mohan B R, Reddy G R M. Resource Consumption Analysis Using ARIMA-ARCH and ARIMA-GARCH Models for Virtualized Server Consolidation System[J]. International Journal of Software Engineering & Its Applications, 2017, 11(6):1-14.
[6] Yan Y, Guo P, Cheng B, et al. An experimental case study on the relationship between workload and resource consumption in a commercial web server[J]. Journal of Computational Science, 2018, 25(03):183-192.