UDC 004.05

Construction of Proactive Monitoring Model using Forecasting Techniques in the SCOM Software Complex

Kateryna Soldatova[1], Svitlana Nosok[1]

[1] *1National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Educational and Research Institute of Physics and Technology*

Abstract

The majority of companies depend on their information systems, the stability of infrastructure operations and the failover of computing resources. Various monitoring tools are mostly used to automate the benchmarking process of company.

The company that has a large distributed infrastructure should pay close attention to this process, as it makes the state of operations difficult to maintain, and increases the probability of the loss of functionality for errors or even shutdown of some servers. The one of solutions is reactive monitoring.

Reactive monitoring is a technique where system administrators use monitoring tools to continuously collect data that determine the active and current status of information system environment. Measurements obtained from real-time monitoring tools illustrate the performance data of current information environment.

However, if we discuss the main metrics of system resources, such as the level of processor load, RAM or disk usage, their change can be quite fast. And for servers that are responsible for critical functions, the problem of full resource usage is important. This problem can be solved with proactive monitoring.

The purpose of this article is to construct the proactive monitoring model with time series forecasting hybrid method for the resources load. The final solution will be used in the management pack of software complex System Center Operations Manager (SCOM). The forecasting methods such as Least squares, SMA and EMA were considered in this work.

Keywords: Proactive monitoring, SCOM, Forecasting, Management pack

Introduction

Relevance. The effectiveness of modern company depends on its ability to collect, process, transmit and store large amounts of information. Therefore, the activities of many companies require a developed information system. Depending on the size of the company and the magnitude of the information flow, it can have a rather complex structure and a large size.

If there is a large number of servers, it becomes more critical to monitor the state of operations. To automate the process of finding problems, speed up their resolution and reduce the number of required specialists, companies should use various monitoring tools[3].

However, reactive monitoring on its own is not always enough. For servers that are responsible for the most important functions of a business process or the system itself, the problems of full resource loading are critical.

With reactive monitoring, the resolution of such an incident would begin at the moment when the problem is already happening, from the second the system administrators are notified, and will take a certain amount of time. If at one hundred percent processor load there is a possibility of services shutdown, which will affect the loss of functionality for the end user, then such a problem would be best predicted with the help of proactive monitoring.

Task setting. The proactive monitoring model for the resources load should be constructed. For effective forecasting the hybrid method should be used. The results of chosen forecasting methods should be combined.

The purpose of this article. Construct the proactive monitoring model for the resources load. Choose methods of time series forecasting.

Combine them in one hybrid forecasting method. Describe the structure of management pack with this method that can be implemented in SCOM.

1. Related works

For the software complex SCOM, that is widely used by companies with large infrastructure[2], no similar solutions have been found.

The software complex SCOM uses reactive monitoring to forecast system resources. It also creates alerts if thresholds are crossed. The thresholds are set by system administrators and are stored in the operational database. In most cases, it is enough, but not in the case of servers that are responsible for critical functions of a company in large infrastructure environment. Therefore, the proactive monitoring model for the resources load using forecasting techniques is needed.

Proactive monitoring is a set of monitoring tools that not only collect real-time information, but also predict possible failures before they impact end users[1].

In the article [7], the advantages of proactive monitoring over reactive monitoring are described. The article [8] deals with fields where proactive monitoring can be used, such as business monitoring, infrastructure monitoring, monitoring of monitoring, that is checking the operations of the monitoring process.

The article [9] deals with the main problems caused by CPU overruns: server restart, process termination. The author also discusses the reasons why CPU monitoring is needed. The author advises to constantly monitor the CPU, if its overload can affect end users.

In the article [10], Kevin Holman examines the features of how the CPU is monitored in the SCOM software complex. It looks at what CPU metrics exists in SCOM and how we should use them. We can view the data we need in this study either immediately in the operational console on performance graphs or make a request to the operational database.

The article [11] uses the ARIMA model for forecasting workflow. Also, in the article [12] we can see using a time series model to predict network traffic patterns.

The difficulty of forecasting task in our case lies in the need to properly address the specifics of our data. The resources load changes quickly. And these changes can have different profile. We should precisely understand under which conditions the chosen methods need to be used and apply them correctly.

The efficiency of forecasting method depends on many factors. In practice, the researcher has quite a lot of freedom to choose not only the type of model, but also the number of parameters entered into it. The following criteria are usually taken into account[6]:

- the amount of effort spent on building the model and the availability of ready-made machine programs;
- the speed with which the method detects significant changes in the behavior of the series, for example, a sudden shift in the mathematical expectation or an increase in the angle of inclination of the trend line;
- existence of serial correlation in errors;
- immutability of primary data;
- the full volume of work in some areas of activity – when thousands of rows need to be updated every month, low costs and speed are of primary importance;
- urgency of forecasting.

2. Implementation

The program complex SCOM generally provides for the following monitoring scheme: the monitoring agent is installed on each end server of the system that is being monitored, which collects data from the server and sends it to the management server. In turn, management servers receive relevant data and send them to the operational database, where they are stored for a week, and then transferred to the storage database. Each management server determines which metrics need to be collected based on the configurations of the management packages stored in the operational database. Data processing may also take place on the management servers.

The following components are proposed for our proactive monitoring subsystem:

- the input request module, which will be obtained from the operational database of the SCOM and assume the value of the load of computing resources at certain moments of time.
- the module for calculating predicted values that will be calculated using the hybrid

forecasting method up to a certain forecast horizon.

- the results evaluation module, which involves comparing the obtained values with the specified level, which should not be exceeded.
- the notification module that will provide alerts for system administrators to the operational console of the SCOM when the forecast of the specified limit level is exceeded

It is also advisable to add the configuration file to the structural diagram, which will contain the list of servers for monitoring, monitoring metrics, their forecast horizons and limit values.

It is assumed that the input request module first receives the data stored in the configuration file on the management server. It gets the domain name of the servers that are the objects of monitoring and monitoring metrics. Next, this module constructs a request to the operational database of SCOM to get the necessary data.

The received data is transferred to the module for calculating the predicted values. This module obtains the forecast horizon from the configuration file and calculates the values up to the forecast horizon or until the time when the resources should run out.

The results evaluation module receives the results and compares them with the limit values from the configuration file. Next, this module passes the positive or negative evaluation status to the notification module and a message for notification if it is needed.

Next, the notification module creates an alert for system administrators and sends them to the operational console of the SCOM. And the occurrence of this alert starts the process of checking the operation of the server, the resources of which may soon run out.

The corresponding scheme of the subsystem of proactive monitoring for SCOM is shown in figure 1.

This subsystem interacts with the software complex SCOM when requesting input data in the input request module. Also, the created modules use computing resources of the management servers of the software complex.

3. Research of forecasting methods

Information about the resource load on the system servers is collected by the software complex SCOM automatically, if the monitoring agent is running on these servers. It represents the value of the resources load at the corresponding moments of time.

To more effectively forecast the resources load, we should combine best methods in the context of our data and its properties. In the study, the following methods are presented: Least squares, SMA, EMA.

The obtained data of the resources load is represented as the time series with corresponding load values for CPU, RAM and disk usage.

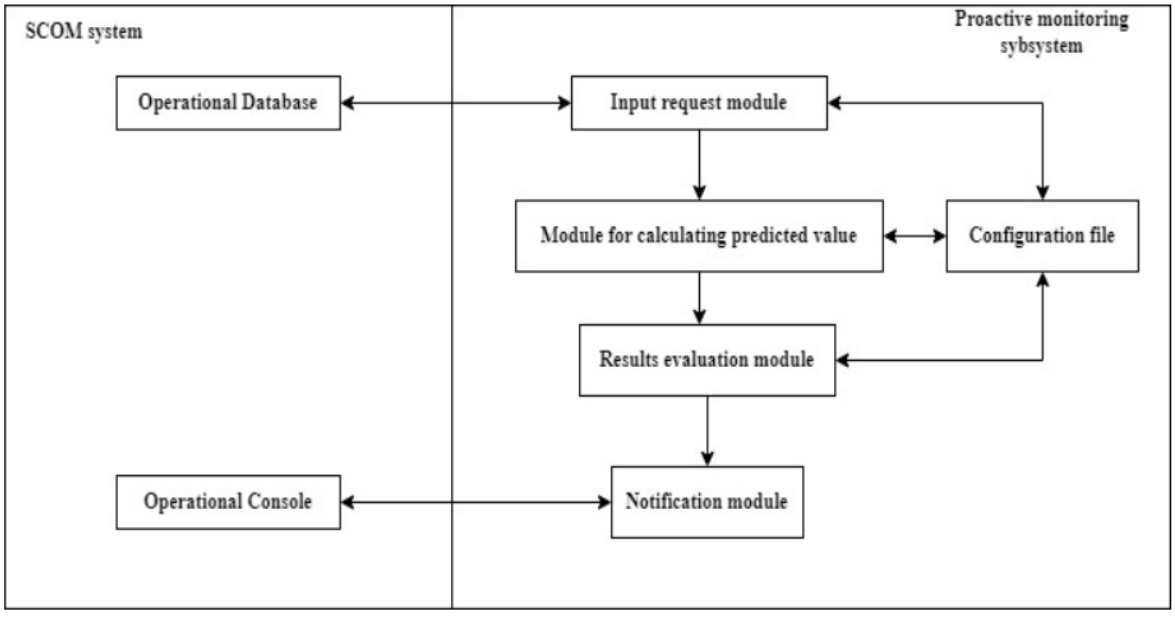

Figure 1: Proactive monitoring subsystem scheme

The classic method for substantiating forecasting methods is the least squares method. It is used to estimate unknown quantities based on the results of measurements or experiments for an approximate representation of the function. Therefore, it is useful for our data when the change rate of resource load is nearly stable[5].

Suppose we have initial statistics. If we define X as time and its elements are presented in increasing order, then we have a certain dynamic series. And the Y value describes a certain dependence:

$$\bar{Y} = f(X), \quad (1)$$

This dependence results from the minimization of the sum of the squares of the deviations of the values of y_i from the calculated values of y_i:

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \quad (2)$$

where $\hat{y}_i$=f(x_i).

The resulting dependence allows for the prediction of the following values.

If we assume that the dependence is linear, then the following function exists and deviations are caused by random factors:

$$\bar{Y} = a + bX, \quad (3)$$

where X is the set of time values. Also, a and b can be found using formulas:

$$\bar{b} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i{}^2 - (\sum_{i=1}^{n} x_i)^2} \quad (4)$$

$$\bar{a} = (\frac{1}{n}\sum_{i=1}^{n} y_i) - \bar{b}(\frac{1}{n}\sum_{i=1}^{n} x_i) \quad (5)$$

where n is the amount of time series values, y_i is the time series value with index i, x_i is the time value with index i.

The second used method Simple Moving Average (SMA) is an indicator that is obtained by summing data points in a given dataset and dividing the total by the number of periods. The SMA formula is written as following:

$$SMA = (\sum_{i=1}^{n} A_i)/n, \quad (6)$$

where A is the average value for a certain period, n is the number of periods.

For a forecast, we can use a simple moving average for a certain "window". For example, we have values of the time series and we choose the window size of 3. Then we calculate the moving average for values 1-3, then for values 2-4, and so on. Next, we calculate the changes between the moving averages and analyze them.

A common approach is to set the marginal value of changes in the moving average, and to add the largest increment in the analysed values to the last value of the series afterwards.

The SMA method should be used, when the observed change is increasing at a moderate rate.

Unlike the SMA method, which considers the calculations for all periods to be equal, the exponential moving average (EMA) assumes that the weight of the last period in the calculation is the largest and has the greatest impact on the forecast result[6].

Each subsequent value of the exponential moving average is calculated using the formula:

$$\mathrm{EMA_n} = (\mathrm{A_n} - \mathrm{EMA_{n-1}}) * \mathrm{counter} + \mathrm{EMA_{n-1}}, \quad (7)$$

where A_n is the last value of period n, EMA_{n-1} is the value of the exponential moving average of the previous period, and the formula for counter is written as following:

$$\mathrm{counter} = 2 / (\mathrm{numOfPeriod} + 1), \quad (8)$$

where numOfPeriod is the number of periods.

The EMA method should be used when we observe the rapidly increasing change of its indicator value.

In the hybrid method consisting of these forecasting methods, we assume the following logic. For the least squares method, we calculate the values using the formulas 4 and 5 and then calculate the predicted value.

Then the value of the SMA method indicator is calculated and it is checked whether the change in value increases more than while using the least squares method and whether the calculated values are closer to the latest actual ones. If the described conditions are met, we choose this method for forecasting.

Afterwards, we calculate the EMA indicator, and if a sharp increase is detected and the deviation from the latest actual value is the smallest, we use this method.

4. Experimental study

The chosen methods was checked on the real data to confirm that they are suitable for use in forecasting of the loading of various resources. Data with different change profile was taken and used for forecasting using the appropriate method.

Table 1

Processor load values with stable rate of change

7:00	7:05	7:10	7:15	7:20	7:25	7:30
35,5	37,5	39,7	41,3	43,8	45,6	47,7

Based on data in the table 1, we can use least squares method. The result that we obtain via this method is illustrated on the figure 2.

As we can see, with slow changes in time series data, the method of least squares gives accurate results and can be used for small fluctuations of changes.

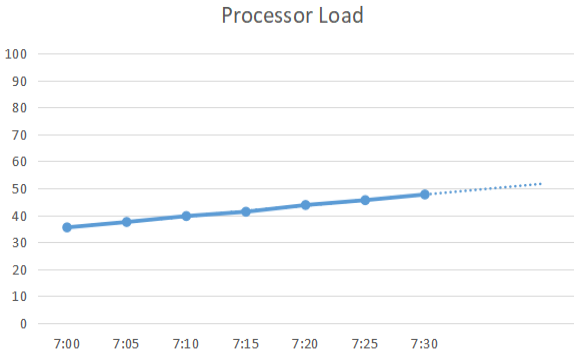

Figure 2: The result of least squares method, in %

For data from the table 2 with the window size equal to 3, the SMA method outputs the result that is illustrated on the figure 3.

Table 2

Processor load values with increasing change intensity.

10:25	10:30	10:35	10:40	10:45	10:50	10:55
35,5	37,5	40,5	45,3	52,6	61,8	72,2

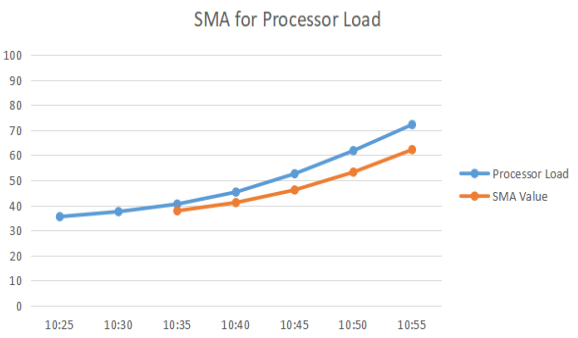

Figure 3: The result of SMA method, in %

Evidently, at a normal rate of increase in load growth the data obtained by using SMA method will be more accurate than when using the method of least squares, as we can see in the figure.

Table 3

Processor load values with quickly increasing change intensity.

Time	5:15	5:20	5:25	5:30	5:35	5:40
Load, %	26,3	26,5	26,4	27,2	27,3	28,2
Time	5:45	5:50	5:55	6:00	6:05	6:10
Load, %	29,6	29,4	28,7	29,3	31	33,2
Time	6:15	6:20	6:25	6:30	6:35	6:40
Load, %	34	35,7	37,4	39	43	44,3
Time	6:45	6:50	6:55	7:00	7:05	7:10
Load, %	48	55,2	58	50	42,7	48
Time	7:15	7:20	7:25	7:30	7:35	7:40
Load, %	44,7	50	56,6	63,2	74	88

As we can see on the figure 4, for data from the table 3 EMA is better for addressing the quick increase in usage of resources. And in this case, we will more accurately get the timing when our resources run out. And the time to solve the problem will be obtained more accurately.

As the methods described above should be used in different situations, we combine them to improve the forecasting result.

So, a combination of methods in the one script can be used in the management pack for SCOM to forecast resources usage.

Our management pack should make a query to the operational database of SCOM and get last load levels. The class, the virtual essence of our

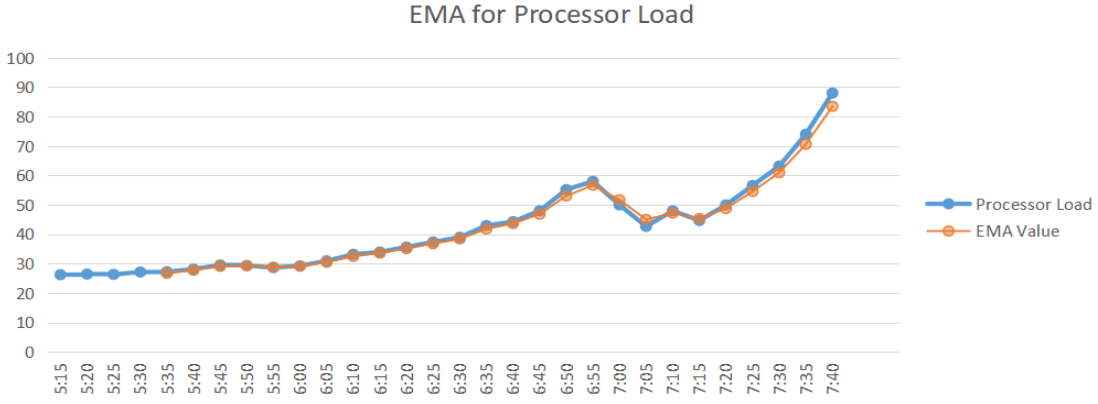

Figure 4: The result of EMA method, in %.

monitoring, should have the name of the server for which requests will occur as a property.

Let's aim it at the resource pool. Accordingly, due to the fact that we target more than one management server with the class, we get fault tolerance of our monitoring.

The objects of the class should be created dynamically using the discovery. The script of the discovery should obtain the list of servers as a parameter and create appropriate objects.

To create notifications under certain conditions, monitors should run a script to receive the date and time when resources are expected to run out and the status of the script in response.

Conclusions

Therefore, proactive monitoring of the resources load helps prevent the loss of critical functionality when the server fully utilises them.

Henceforth, the subsystem of proactive monitoring for computing resources can be built in the software complex SCOM system that will allow to response to possible problems more quickly. This subsystem uses following forecasting methods for its work: the least squares method, the SMA method, the EMA method.

The combination of chosen forecasting methods helps implement such proactive monitoring and can be used in SCOM, a monitoring tool widely used in large infrastructure business environments. The management pack that was created using these methods can be deployed to improve monitoring process of computing resources.

For further improvement of forecasting process, it’s appropriate to create even more complicated logic of methods interaction.

References

[1] B.Kirsch, Combine reactive, proactive monitoring for optimal IT visibility, 2019. URL:https://www.techtarget.com/searchitoperations/tip/Combine-reactive-proactive-monitoring-for-optimal-IT-visibility.

[2] Microsoft Corporation. Operations Manager 2022 Guide, 2022.

[3] E.Moreno. How to monitor your IT infrastructure effectively, 2021. URL:https://adam.es/en/blog/how-to-monitor-your-it-infrastructure-effectively/.

[4] H.Abdi, The Method of Least Squares, 2006. URL:https://www.semanticscholar.org/paper/TheMethod-of-Least-Squares-Abdi/6955b8683ceac56c80b80c2a50173660faeba2b7.

[5] C. Mayaski, Understanding Exponential Moving Average vs. Simple Moving Average, 2021. URL:https://www.investopedia.com/ask/answers/difference-between-simple-exponential-moving-average/.

[6] G.Malytska. Time series: lecture notes. Vasyl Stefanyk Precarpathian National University, 2017.

[7] C.Cooney, Proactive Monitoring vs. Reactive Monitoring, 2022. URL:https://coralogix.com/blog/proactive-monitor-vs-reactive/

[8] T.Theakanath, Proactive Monitoring, 2015. URL:https://devops.com/proactive-monitoring/

[9] SentinelOne, How and Why to Monitor Server CPU Usage, 2021. URL:https://www.sentinelone.com/blog/how-and-why-to-monitor-server-cpu-usage/

[10] K.Holman, How does CPU monitoring work in the Windows Server 2016 management pack, 2019. URL:https://kevinholman.com/2017/05/13/how-does-cpu-monitoring-work-in-the-windows-server-2016-management-pack/

[11] R. N. Calheiros, E. Masoumi, R. Ranjan, and R. Buyya, “Workload prediction using arima model and its impact on cloud applications qos,” IEEE Transactions on Cloud Computing (2015): 449–458.

[12] R. S. Kalyanaraman, Y. Xiao, and A. Yl-jski, “Network prediction for energy-aware transmission in mobile applicatioins,” in International Journal on Advances in Telecommunications, 3:7282, 2010.