



## A Systematic Approach to Performance Evaluation

Performance evaluation is the process of determining how well an existing or future computer system meets a set of alternative performance objectives. Arbitrarily selecting performance metrics, evaluation techniques, and workloads often leads to inaccurate conclusions.

How should one carry out a performance evaluation study? The answer is to follow a systematic approach. The methodology proposed here involves six steps:

- ◆ Understand the current environment and define goals for analysis.
- ◆ Identify and gather relevant performance metrics.
- ◆ Select the appropriate evaluation technique.
- ◆ Define characteristic workloads.
- ◆ Analyze and interpret the data.
- ◆ Present the results.

The purpose of this article is to provide the performance analyst with a systematic approach to performance evaluation and point out some common mistakes that can be avoided.

### Step One - Define Your Goals

There is no such thing as an all-purpose system performance model. Each model must be developed with an understanding of the system and the problem to be solved. Therefore, the first step in conducting a performance evaluation study is to clearly state the goals of the project and obtain a global definition of the current environment based on these goals. Setting goals involves deliberate interaction with system users and decision-makers, taking care to avoid any preconceived biases or beliefs. Embarking on a performance study to prove that one alternative is better than another is bound to be met with skepticism. It is best to base all conclusions and recommendations on the results of the model rather than on what one party wants to hear.

The definition of the computing environment may vary depending upon the goals of the study. The system definition may consist of the CPU, memory or I/O subsystem, the network, the database, or the entire computing system itself.

How the environment is defined affects the choice of performance metrics and workloads needed to compare alternative objectives. At a minimum, the overall description should include the following system components: hardware configuration, operating system, support software, and major applications. In addition, a list of services provided by these components and an understanding of existing service level agreements, user profiles, administrative controls, peak time windows and possible alternatives and problems should be part of the preliminary examination the system environment.

A large share of the analysis effort of a performance problem goes into defining the problem and stating the goals. Putting a lot of effort into this initial step helps to narrow the scope of the study and reduces the time and cost required to find the solution. Analysis without understanding the problem or goals of the study will only result in inaccurate conclusions.

## Step Two - Choosing Performance Metrics

Once the goals of the performance study have been defined and you have a clear understanding of the system or systems to be evaluated, you can move on to identifying and gathering relevant performance metrics. The choice of metrics may depend on the tools available to you for collecting and measuring system data, and vice versa. An understanding of the type of data available through accounting logs, hardware and software monitors, load generators, program analyzers, and other measurement tools is helpful in this step. Then, depending on the objectives of the study, data collection should take place during a typical (if averages are being studied) or peak (if system capacity is of concern) time period. It is also important to consider the constraints and limitations imposed by the alternatives being studied and the time available for the decision when choosing your metrics.

Most importantly, your choice of performance metrics should depend upon the goals of the study. A common mistake is to choose those metrics that can be easily computed or measured rather than those that are relevant to your objectives. Examples of commonly used performance metrics are throughput and response time. But, care must be taken when using even these standard metrics. For example, if the objective of your study is to compare the expected throughput of two alternative systems, you may run into complications during the analysis phase of your evaluation comparing the MIPS measured on one system to the VUPS reported by another. Furthermore, understanding exactly how a metric, such as response time, is measured and reported plays a role in building your performance model. Ignoring important parameters, choosing inappropriate metrics, or overlooking differences in metrics from one system to another may render your results useless.

A good way to start identifying relevant performance metrics is to make a list of all the system and workload metrics that affect the performance of the system(s) under consideration. In general, the metrics will be related to the speed, accuracy, and availability of system services. System parameters include both hardware and software parameters and may include such measurements as working set size or CPU clock speed. The global definition of the computing environment created in step one should be of help in defining the system parameters. Workload parameters are characteristics of users' requests such as traffic intensity, system load, and resource consumption. Pay special attention to those workload parameters that will be affected by your alternative performance objectives as well as those that have a significant impact on performance – CPU utilization, total I/O operations, memory usage, page faults, and elapsed time.

As a final note, it is important to understand the randomness of the system and workload parameters that affect the performance of your system. Some of the metrics you've

identified and gathered in this step will be varied during the evaluation while others will not. Observing system behavior for several days or weeks will allow you to determine realistic values and distributions for each metric included in the analysis. The ability to accurately reflect system behavior in your model through the use of relevant metrics will lead to much tighter conclusions.

### **Step Three - Selecting the Appropriate Evaluation Technique**

Performance management is the process of ensuring that a computer system meets predefined performance objectives consistently and efficiently, both now and in the future. In order to evaluate future system behavior, the analyst must use predictive models. There are four modeling techniques commonly used for performance management: linear projection; analytic modeling; simulation; and benchmarking. Each method has its own merits and drawbacks and, in some cases, choosing one technique over another may be a simple matter of economics. Knowing what data, tools, and skills are available, what level of accuracy or detail is required, how much time and money is devoted to the project, and whether the alternatives to be evaluated exist or have yet to be developed all play a key role in determining which technique should, or can, be used.

Linear projection involves collecting past data, extending or implying trends through the use of scatter plots and regression lines, and comparing the trend line with the current capacity of the system. Although linear projection is used quite frequently to make assumptions about future behavior, the performance of computer systems is far from linear. Therefore, any linear relationship determined between two system components should only be used to understand the current (or past) behavior. Linear projection is best used when you want a first approximation to a very simple model.

Analytic modeling uses queuing theory formulas and algorithms to predict response times and utilization projections from workload characterization data and key system relationships. Basically, analytic models require input data such as arrival rates, user profiles, and service demands placed by each workload on various system resources. System monitors and accounting logs can provide most or all of the required information. Most analytic models allow for the use of various assumptions in order to keep the solution simple and efficient. The affect of these assumptions on the accuracy of your conclusions must be considered when making recommendations. Analytic modeling is best used when a numeric prediction is needed, when workload forecasts are fairly accurate, when a best-case/worst-case situation holds, or when an accurate answer can save significant money, time, or headaches.

Simulation models are based on computer programs that emulate the different behavioral aspects of a system under evaluation. They allow for the investigation of system behavior at any level of detail, but not without cost. The fine level of detail (input data) generally required to build simulation models often makes them expensive to undertake. While linear projection and analytic models can be solved with pen and paper or simple PC-based applications, simulation models typically require powerful workstations or even mainframes. Simulation is best used when there are important questions about details of such things as the dispatching algorithm or application and device internals that are not answered by analytic modeling, or when the proper experience in simulation exists or can be bought.

Benchmarking is a useful approach for predicting the capacity of new computer systems or testing the impact of new workloads, databases, or operating systems on existing systems. Because of the high cost involved in conducting a benchmark study, the analysis of multiple

alternatives may be prohibitive. The accuracy of benchmarking results can vary from high (testing a real workload on an existing system) to none (testing undeveloped workloads or prototype systems). Other considerations include: generating the needed files; ensuring repeatability; representative accuracy; and time to prepare. Benchmark experiments are best used when you can't get an accurate estimate of a workload's resource usage in another way.

It is clear that the goals of a study and the data and resources available have a significant impact on what is modeled and how it is analyzed. In all cases, the accuracy of your conclusions relies on how well the model and the data used to feed the model represent the real system.

## Step Four - Defining Characteristic Workloads

Workload characterization involves studying the user and machine environment, observing key characteristics, and developing a workload model that can be used repeatedly. Once a workload model is available, the effect of changes in the workload and system can be easily evaluated by changing the parameters of the model. In addition, workload characterization can help you to determine what's normal, prepare a baseline for historical comparison, comply with management reporting, and identify candidates for optimization.

Typically, you'll find that the alternatives you're asked to evaluate are stated in business- or user-oriented units instead of machine-oriented units. For example, you'll be asked to model the performance impact of adding six new point-of-sale terminals in the northeast or of changing how the current billing application is run. The advantage of user-oriented units is that the users can easily relate to them and can frequently give a good forecast of future workload. The disadvantage is that they do not automatically relate to system resources used by a workload. As the performance analyst, you must quantify the relationships -- turn an increase in the number of terminals into an increase in total resources used by each workload class. It is not guaranteed that a good relationship exists.

Using the user-oriented information from your study of the environment (Step 1) and the system measures gathered in Step 2, you need to divide the total system workload into workload classes. A workload class is a combination of transactions that have enough in common to be considered one homogeneous workload. The goal is to have complete coverage of system workload with no overlap of classes. Depending upon the evaluation technique being used (Step 3), workloads may be characterized in different forms. Some attributes used for classifying a workload include: resource usage, application type, geographical orientation, visibility, organizational unit, processing type, and similar growth, shrinkage, or stagnation. At times, it may be appropriate to use multiple criteria to define a single class. Regardless of how you choose to characterize the workload classes, the resulting workload model should capture quantitative information about the real system workload.

In choosing the parameters to characterize each workload class, it is best to use those related to workload service demand or intensity such as arrival rate, number of terminals and think time, number of concurrent programs, or duration of request, rather than parameters that depend on the system. For example, since a transaction's response time is dependent on the system on which it is executed, response time is not a good choice for a workload parameter. Remember, you want to build a single workload model that can be repeatedly used to evaluate a variety of scenarios. After the parameters have been selected, you need to assign appropriate values to them. Two popular techniques for

calculating parameter values include averaging and cluster analysis. If you use averaging, test the variability in your measures using statistical techniques such as mean variance, standard deviation and range. Clustering analysis can be used to identify natural groupings for the parameter values based on similar traits. Various clustering algorithms are available in literature.

Like other aspects of performance management, proper selection of workload classes requires a lot of judgment by the analyst. Care should be taken to assure that your workload model accurately reflects the services exercised by the real system workload since the wrong workload may lead to inaccurate conclusions.

## **Step Five - Analyzing and Interpreting the Data**

Armed with the proper workloads and data and with a thorough understanding of the environment and goals of the evaluation, you are ready to run your model and compare it against actual system measurements. Calibration techniques may be required if your system and workload models don't readily match observed system behavior. Commonly accepted techniques to validate and calibrate performance models are available in literature. Once the model is verified and the results appear to be representative of the real system, you are ready to use this baseline model to compare alternative scenarios. The results of each run of the model will provide the information needed to draw conclusions about the set of proposed alternatives. Interpreting these results and providing supporting documentation is where the analyst truly earns his or her keep.

Erroneous analysis is as bad as no analysis. The number and type of assumptions or adjustments incorporated into your input parameters or model may have an effect on how robust your model is at reflecting actual behavior. It is important to adjust the level of confidence on the model output based on the input data or model used. Conclusions made about each alternative should take into account the variability and randomness of the results. Statistical techniques to test for variability may include regression analysis, confidence intervals and hypothesis testing.

None of your recommendations should be based on conclusions that are extremely sensitive to small changes in the model parameters. Sensitivity analysis should be used to assure that your conclusions won't drastically change if the workloads or parameters used in your model are changed slightly. Sensitivity analysis is also helpful in assessing the relative importance of various parameters and how they impact desired service levels.

In an attempt to save time and money and simplify the task of workload characterization, future workload and system behavior is typically assumed to be a modification of current or past behavior. To improve the validity of your results, any predicted effects of the strategic plans of the company or key economic indicators should be incorporated into your analysis. Furthermore, restricting how far into the future your workload predictions are made will help simplify the task of analyzing the results.

Don't waste time comparing alternatives that the decision-makers cannot or will not adopt either because they are beyond their control or don't fit into their personal agendas. Preference should be given to those alternatives that can be easily implemented, have a low associated cost or risk, or meet the qualifications defined in the statement of goals. A selection of alternatives should be analyzed and presented though, so that the decision-makers may choose the most appropriate option.

## Step Six - Presenting the Results

The final step of all performance projects is to communicate the results of your analysis to the decision-makers. Since the project report is the primary deliverable for any study, it is particularly important that it be clear and reliable. It should give specific, practical findings and recommendations, with little technical jargon and no vague generalities. The project report should serve as a valuable reference and as a convenient baseline for comparing future performance with current performance.

Acceptance of the analysis results requires developing a trust between the decision-makers and the analyst. A key factor in effectively establishing trust and communicating your conclusions is knowing your audience. Understanding their level of technical expertise and, more importantly, their motive for initiating the study will be helpful in organizing your report and assuring that the right questions are answered. The primary reasons for initiating the study may include: user dissatisfaction; external image of the company; productivity decreases; budgetary constraints; or risk of financial loss. Non-technical decision-makers are rarely interested in the modeling technique you used or how it is implemented. Their primary interest is the guidance that the model provides and the political impact the recommendation may have. Technical members may be more concerned with the assumptions and limitations of the analysis or the availability of personnel to implement a particular course of action.

The project report should include a summation of all the analysis that has been performed by the analyst, show resource utilization levels of the current system, and identify hardware and software alternatives which will satisfy future processing requirements. The report should include enough data tables and figures to support the major findings, and clearly identify the data sources that support each finding. Finally, it should provide an estimate of the time, cost, and feasibility of implementing a solution.

In addition to submitting a project report, you may have the opportunity to participate in a project briefing. A project briefing allows you to explain the results of the study to management and technical staff. It also gives you the opportunity to answer in-depth questions concerning the implementation of any corrective actions being recommended. Again, appropriate use of words, pictures, and graphs to explain the results and analysis should be used.



SightLine Systems Corporation  
11130 Fairfax Blvd., Suite 200, Fairfax, VA 22030  
703-563-3000  
877-744-4854  
[www.sightlinesystems.com](http://www.sightlinesystems.com)