

te testing experience

The Magazine for Professional Testers

printed in Germany

print version 8,00 €

Performance Testing

free digital version

www.testingexperience.com

ISSN 1866-5705



We turn 10!

Load and Performance Testing for Software Applications

by Hansjörg Senn

Performance might not be the first thing you think of when using a software application, but at the latest when software response time taxes your patience you will think about it. This article shows you the motivation for performance testing and points out the topics you have to take into account. Besides the performance objectives, the use of adequate tools and the appropriate strategy and procedures are other important aspects. If you are interested in these topics, please read on.

1. The purpose of performance tests

Do you remember your last performance problem? I remember my last performance problem well, and I also remember that fortunately the performance bottleneck was discovered in the testing and not in operational phase, which simplified the situation at the time. For most software development projects, performance is a crucial success factor. Project management should plan for performance even in the initial design and development phases, because with the first production run performance will be on the agenda at the latest. Most web applications have a good performance as long as only a few users are active. But what happens if a large number (e.g. thousands) of users work with the application simultaneously? At this point „Load & Performance Testing (L&P)“ becomes a major issue. By simulating a large number of web users, we can generate load and network traffic on the system, and the application as well as various servers are driven to their performance limits.

There are several reasons to perform L&P tests:

- There is no 100% certainty that multi-user software makes all its functions and methods available to all users, and is therefore multi-user capable. But with performance testing it is possible to find a high percentage of the faults which were not found by individual manual testing, independent of whether unit tests or system tests were performed.
- Performance testing allows a realistic check of the configuration of interfaces to other systems or subsystems and computers. In addition, performance tests can also detect bottlenecks in host systems..

Performance tests essentially concern the following topics:

- Consumption of resources
- Response times
- Large load / number of users

Especially people from operations are mainly interested in the system resource consumption. They want to know what to expect

with the introduction of an application to the production system. The adequacy of servers and instances is of interest. On the other hand, application users are interested in response times. With a large load created by many users, there will be an increase in response times and resource consumption.

2. New challenges

Application development used to be done using machine-oriented programming languages. This way the programmer had a more precise view of the target platform, on which the application code was carried out. For this reason, he could see certain performance-related risks more clearly and also earlier than today. New methodologies and technologies promote the use of distributed applications. This offers high flexibility for the interaction between components and for the composition of services to construct application systems. On the other hand, the new technologies are increasingly abstract and complex, which leads to an almost infinite number of possible combinations. This increases the risk of hidden performance problems.

3. Test types

Contrary to the various functional test procedures for testing the functional and technical correctness of software code (e.g. during unit testing, integration testing, regression testing), load and performance testing belongs to the category of non-functional test procedures. The technical literature [1] generally differentiates between the following types of non-functional tests:

- Performance test
- Load test
- Volume test / mass test
- Stress test

Of course, it strongly depends on the test objectives, which of the non-functional test types are to be accomplished. With the above-mentioned test types we will be able to detect faults in the following areas:

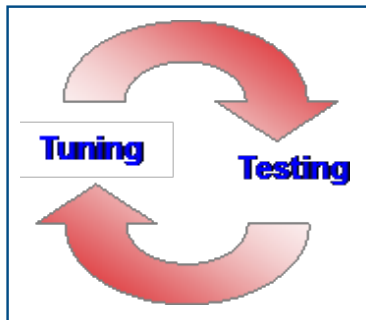
- Memory problems
- Performance issues
- Concurrency problems
- Excessive use of system resources
- Lack of storage space

From now on, we will summarize these kinds of tests under the generic terms „load and performance tests“ or simply „performance tests“.

4. Lifecycle of performance tests

An important basis of performance testing and the tuning of a system is the lifecycle of testing.

1. Each monitor which is installed and each logging mechanism which is used to measure the performance of a system has an effect on the system itself, whether intended or not. Therefore each monitor is also an interference factor. The aim should be to keep the impact of the monitors on the system as small as possible without any negative effects on the quality of measurement results.
2. Every test run is different. Even with a complete conformity of all configuration files and settings, we will receive different test results from one test run to another. Therefore the aim is to keep the deviations as small as possible.
3. After each tuning and after each parameter change, a new test run is required. Furthermore, the results must be analyzed and the system optimized and/or repaired after each test. If more than one parameter setting is modified at a time, we run the risk of not being able to understand why the system or some modules have become slower or faster. In the worst case, the effects of parameter changes cancel each other out and system behavior remains unchanged.
4. This cycle (testing, tuning, testing,...) is a never-ending story and could be continued indefinitely without reaching the maximum system performance. Thus it is important to find a trade-off, i.e. a good ratio between effort and yield.



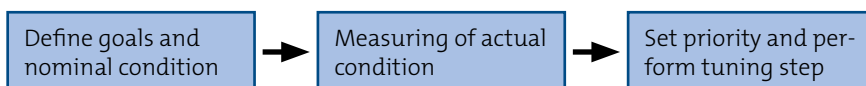
5. Procedure of load and performance testing

Test procedures are an essential part of quality assurance in the software development process. This applies to both the functional and the non-functional requirements of a system. Load and performance tests are an essential part of the non-functional test requirements of a system.

A very important point of planning is therefore to...

Define your objectives before taking any action!

In practice this rule often is violated, especially if the project is under time pressure. Therefore the following procedure is recommended:



5.1 Definition of goals and specification of metrics (benchmarks)

It is important to define the benchmarks before performing the tests. The benchmarks are the basis for professional optimizati-

on and for the selection of the appropriate test techniques. The benchmark should fit the application's requirements, i.e. a complex intranet application should not have to deal with e.g. 10,000 concurrent users. The following metrics are useful:

- Maximum number of users
- Maximum number of simultaneously logged-in users
- Average user think-time per page
- Average of max. response time per page
- Number of transactions per time unit (throughput)

In addition, at least the following conditions should be met:

- Test data based on realistic data
- Reproducible baseline, i.e. a recoverable baseline for every test run (regression), so that we can guarantee the comparability of different test runs
- Realistic test environment to obtain test results that are relevant.

Performance requirements must be defined during the specification phase and not only during the acceptance test phase.

Performance requirements should be checked for consistency and validity. As a guideline, we recommend using the so-called SMART criteria (Mannion and Keepence 1995, [2]). SMART is an acronym and means:

Specific

Are the requirements formulated in a clear and consistent way, so that they can be understood by the contractor?

Measurable

Are there measurement procedures to identify and analyze the appropriate performance data?

Attainable

The requirements can be achieved in theory and are not impossible to physically implement.

Relevant

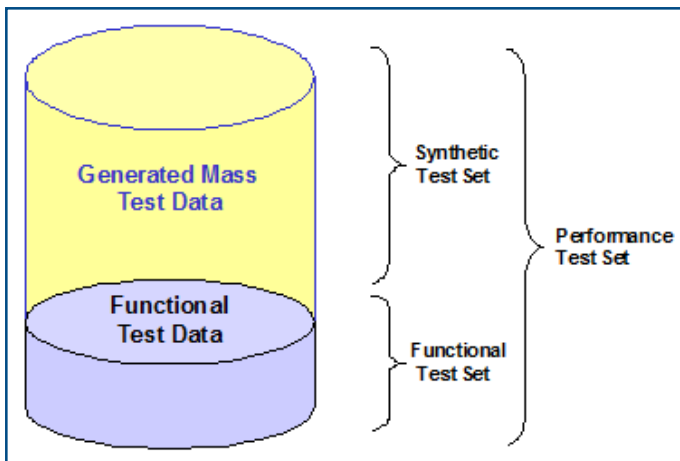
Is it possible to meet the demands in practice, or are they contrary to basic conditions (time, budget, hardware, etc.)?

Traceable

Does the implemented system meet the requirements? This means that the performance management is able to monitor and control the achievement of performance requirements with suitable test methods.

5.2 Test data

Test data is a very important topic, which is often neglected. Professional performance testing should be done with realistic data volumes, i.e. volumes which are comparable to production data volumes. In the majority of cases it is not possible to get test data from the production system, either because the application is newly developed or due to the protection of data privacy. In such cases, the mass of test data has to be generated synthetically. In practice, such mass data is generated on the basis of a functionally correct data subset (e.g. with SQL-scripts). The following graph illustrates this approach.



Realistic data volumes are the basis for meaningful performance tests, because database-related performance bottlenecks can often only be detected with a mass data load. The setup of appropriate test databases can be tool-supported (e.g. with Data Factory from Quest).

5.3 Reporting and presentation of test results

For reporting and presentation of L&P test results, it is important to know in advance how and which data has to be logged. There is no general procedure for data analysis and data collection, because this is highly dependent on the information needs of the measurement plan. In any case, the following data should be part of every test reporting:

- Statistics on response times measured on the client side compared to system load.
- Statistics on throughput as a function of the system load.
- Number of faults occurring during test
- Utilization of processors and disk volumes in the test environment.
- Network utilization and network latency.

With each performance test you will obtain large data volumes for analysis. Therefore it is strongly recommended to use tools for data handling, because we will often not find the relevant information in a single data set but in the relationships between large amounts of data. Well-established technical tools for analyzing and processing of data and handling data correlation apply statistical methods, clustering techniques for data aggregation and techniques for data filtering. The results can be saved as MS-Excel spreadsheets. The reports must contain both the actual test results and all relevant system parameters. Of course, graphical reports are useful and essential to get a quick outline on the test results.

5.4 Organizational aspects

To ensure a smooth flow of the performance test, some organizational measures have to be taken. Sometimes it is necessary to perform L&P-tests outside of normal working hours. Then it has to be ensured that access to workplaces and server rooms is guaranteed. The people involved must be on site or must at least be reachable. These kinds of issue are best organized with the help of checklists.

A very important point regarding the team organization is the coordination with all parties involved. These are usually:

- Developers
- External / internal testers
- System engineers
- Project manager

Often, performance tests are only meaningful if they are tested in

production environments. In this case we must bear in mind that, inevitably, other systems can be influenced, e.g. by network load, database access, etc.

For the implementation and monitoring of Load & Performance testing, qualified specialized staff must be available. This is also because of the special tools which are often used.

5.5 Test tools

For load and performance tests, appropriate tools are essential. These are primarily tools for performance monitoring, for load generation and possibly for data analysis, visualization and report generation. In recent years the market of performance testing tools has changed significantly. Some tool providers no longer exist or were bought by a competitor, while some have established themselves as newcomers. The choice of adequate test tools is a very important issue. It would therefore be useful to test the tools in a pilot project before rolling out in your organization.

6. Evaluation of L&P test tools

New methodologies and technologies promote the use of distributed applications. On the one hand, this provides high flexibility for the interaction between components and for the composition of services for application systems, but it also results in high complexity of the systems. Therefore it is a challenging task to evaluate the suitable tool for L&P tests. In [1] (Florian Seyfert, "Testwerkzeuge für Last- & Performancetests") we will find an evaluation method based on a cost-benefit analysis ("Nutzwertanalyse", C. Zangemeister [3]), which is a useful method to identify adequate tools.

6.1 Setup requirements catalogue

The requirements catalogue consists of two parts:

- Pre-selection criteria (ko-criteria)
- Mandatory criteria

6.1.1 Pre-selection criteria (ko-criteria)

In a first step, the selection of the possible test tools is restricted by pre-selection criteria (ko-criteria). These are as follows:

- ko-1: The tool must be suitable for load and performance testing
- ko-2: The tool must be available in the local market
- ko-3: A free demo version must be available
- ko-4: The product must be up-to-date (not older than 5 years)

There are a number of tools listed in a study by IDC¹ (International Data Corporation, 2006). This list is the basis for pre-selection.

6.1.2 Mandatory criteria and weightings

The mandatory criteria were determined by consulting reference books and by a survey among experts. The weightings are as follows:

- 5 = extremely important
- 4 = very important
- 3 = important
- 2 = less important
- 1 = unimportant

The mandatory criteria were divided into two parts, a theoretical and a practical part. The theoretical part rates the provider's information, mainly based on white papers or other product information. The practical part is an assessment of the practical tests runs with the different tools.

¹ International Data Corporation (IDC), with U.S. headquarters in Framingham, is an international consulting and event supplier in the fields of IT and telecommunications industry with offices in about 40 countries.

Altogether the criteria are grouped into 6 themes, the first two (A, B) represent the theoretical part, the others (C, D, E, F) belong to the practical part:

- A: Provider information (weighting: 4)
- B: Product information (weighting: 4)

- C: Scripting (weighting: 5)
- D: Scenario design (weighting: 5)
- E: Scenario process flow (weighting: 4)
- F: Scenario analysis (weighting: 5)

6.2 Test tools and market study

An IDC study (see [1], Florian Seyfert, "Testwerkzeuge für Last- & Performancetests") did a global market study of current L&P-test tools for the period from 2005 to 2010. The study was based on market figures from 2004 to 2005 and analyses sales trends of manufacturers and the growth of the market. At least two of the following requirements must be fulfilled:

- Annual sales in 2005 greater than 1 million U.S. Dollar
- Market share above 1% ...
- ... or growth in sales greater than 15%

Based on the above-mentioned requirements, the manufacturers list is as follows:

- HP / Mercury
- IBM
- Compuware Corp.
- Empirix
- Segue Software (since 2006 Borland)
- Borland Software Corp.
- RadView Software

Finally, only tools that pass the ko-criteria are included in the shortlist, which is as follows:

- Loadrunner (HP / Mercury)
- e-Load (Empirix)
- SilkPerformer (Borland)
- Rational Performer Tester (IBM)

The open-source testing tool OpenSTA was also considered in the evaluation, but sometimes it was difficult to apply certain criteria to the tool.

6.3 Test conditions

For each tool a test script was implemented to perform the practical test. This is the basis for rating criteria C to F:

- C: Scripting
- D: Scenario design
- E: Scenario process flow
- F: Scenario analysis

Please note that „practical test“ does not mean a real performance test in this context, but test for evaluation of L&P-test tools. Therefore it is extremely important that the test conditions for each tool are the same, i.e. identical test environment and test scenarios. A PC was used as the test environment and the test scenario was a web application for flight reservations, which was launched on a local web server. Six different users logged-in one after the other (ramp-up) to book different flights. The measure-

ment focus was on throughput (RPS, requests or hits per second) and the response times of transactions. The user think-times were from 3 to 18 sec, depending on the web site on which the test was done. Each scenario was repeated for each tool under the same conditions in order to identify possible discrepancies. In addition, the test scenario included the simulation of a severe failure to verify how each tool handles this case.

6.4 Evaluation result

Referring to [1] the evaluation result, the test winner was Loadrunner. However, if you are not prepared to pay the relatively high license costs, e-Load and SilkPerformer may be an alternative.

The ranking list is:

1. Loadrunner
2. e-Load
3. SilkPerformer
4. Rational Performer Tester
5. OpenSTA

When used by highly qualified experts, OpenSTA may also be an alternative, particularly in terms of cost. Please note, however that its functionality is limited in contrast to the winners.

Basically for each project we should define the performance objectives in detail to decide which tool is suited best. The present evaluation was done in 2007. Therefore, it is always advisable to check the current version of any tool concerning its suitability for a project. However, future versions of the tools will not offer new features, but will be enhanced in terms of stability and design.

7. Conclusion

L&P testing is a non-functional test type which should not be underestimated. L&P testing should be taken into consideration early in the project, as early as in the design phase. New methodologies and technologies provide increasing opportunities, but also more complexity for application systems. This fact increases the risk of hidden performance problems. L&P-testing is no one-man-show. In fact, the cooperation of a professional team of experts is necessary. The definition of the objectives is as important as a structured approach. In this context, we would like to point out the so-called SMART criteria (Mannion and Keepence 1995, [2]). L&P-testing is a cycle of "Testing - Tuning - Testing...", and it is very important to change only one parameter at a time to see what the result of this action is. Moreover, there are two other issues that are also significant: these are „test data“ and "data analysis / reporting". There is no L&P testing without tool support. Therefore this paper gives a rough overview of well-established tools available in the market and an evaluation result. In any case, it is essential to verify whether a tool is suitable for a specific application and for the associated performance objectives.

8. Bibliography

- [1] Florian Seyfert, "Testwerkzeuge für Last- & Performance-tests", VDM Verlag Dr. Müller, January 2008, ISBN: 978-3-8364-5559-6
- [2] Mannion and Keepence, SMART criteria, 1995
<http://www.win.tue.nl/~wstomv/edu/zip30/references/smart-requirements.pdf>
- [3] C. Zangemeister, "Nutzwertanalyse in der Systemtechnik", Wittmannsche Buchhandlung, 1976



Biography

Hansjörg Senn holds the Diploma in Electrical Engineering from the University (Fridericiana) of Karlsruhe (Germany). Since 1996, he has been working for COMIT AG (Zürich) as an expert in the area of software testing (Test and Project Manager). He focuses on banking software.



Subscribe at
te testing
experience
www.testingexperience.com