IBM Global Services
November 2003

Optimizing Linux environments for performance and scalability

Contents

Introduction

Linux is a reliable, security-rich and scalable operating system designed to deliver high performance across an increasing range of business applications. Organizations are moving to Linux to take advantage of the sizeable benefits of its open source operating system, including significant cost and performance benefits over proprietary operating systems. Linux and open source applications can be implemented across a wide range of technology platforms, greatly expanding hardware and integration choices for budget-conscious organizations. In addition to the cost advantages that this flexibility can afford, Linux also provides a solid and scalable infrastructure designed to support movements to grid computing and e-business on demand.™

With Linux, as with any other operating system, achieving optimal performance takes effort. Out-of-the-box functionality doesn't equate to "out-of-this-world" performance. Linux operating system distributors and application software vendors build their products to satisfy a broad range of customers. What's more, vendors typically configure and tune their products for their own test lab environments. In order to realize the optimal performance required to support business-critical applications and satisfy end users, companies must invest in planning, configuring and tuning Linux implementations for their unique IT environments.

Scalability planning to help ensure capacity for future growth is also important to the success of Linux implementations. Planning for scalability is much more than over-provisioning hardware or implementing load balancing servers.

Highlights

Accurate planning requires a focus on the whole picture—understanding how individual components of the infrastructure relate to each other and are affected by the design of existing and planned applications. Considering performance from an application and transaction perspective when planning for future capacity of your Linux-based solutions can help prevent costly productivity and availability problems down the road.

This paper provides a guide to proven performance and scalability planning methods that IBM has developed to successfully deploy high-performing business systems in Linux environments.

Charting your course

The first phase of planning is developing a clear vision of your ultimate destination. The key to is to start with tangible business goals, and then map those objectives into measurable IT goals. Those IT goals will, in turn, define the infrastructure requirements for business success.

Begin the planning process by identifying tangible business goals, then translate those into measurable IT goals.

For example, a financial services company may have the business goal of growing market share by offering innovative Web brokerage services. The availability and performance of this new application will be critical to maintaining customer satisfaction during peak trading periods. This business goal, therefore, could translate into an IT goal of providing network coverage for up to 40,000 simultaneous users. The result is a clearly defined target for IT performance and scalability planning.

Highlights

Planning in reverse—making infrastructure decisions based strictly on IT goals such as hitting budget targets, adhering to current platforms or achieving desired application response times—may in fact provide cost savings or better performance. But if those IT goals do not fall in line with business objectives, the project can fail to produce the desired return on investment.

Creating a baseline

An accurate baseline of your current operating environment is a crucial step to achieving realistic projections.

A ship that is slightly off course will not reach its planned destination. The same is true with scalability planning for Linux environments. Having an accurate picture from which to start the journey—in this case, a baseline of the current operating environment—is the essential first step to achieving accurate projections. Merely estimating actual workloads or assuming anticipated resource utilization can result in a plan that will be off track—and applications that will not achieve the performance or availability needed to support business goals.

Baseline data collection should encompass all components of the infrastructure and application mix.

The accuracy of scalability planning is directly tied to the ability to obtain a complete picture of existing application and infrastructure performance. IT organizations often have information available about workloads, but little else. A common mistake is failing to capture and correlate response time or performance statistics from each component of the IT infrastructure that supports a business solution. Be sure to collect performance metrics from all components of the infrastructure and application mix. This includes application, CPU, input/output (I/O), memory, network, storage and database data

Highlights

It is important to include both peak and average workloads to provide a valid base for capacity planning and testing.

that can give a valid representation of current infrastructure and resource usage. Using specialized tools and probes, you can obtain application profiles for unique business functions and capture end-to-end response times. Also, perform a network traffic workflow analysis, including response times, trends and application flows.

Defining both online and batch-window baselines is important because the characteristics of each are different. Likewise, it is important to take measurements by capturing actual data and transaction rates from the current production environment on both peak and average hours or days. The baseline should be representative of the most extreme peak-hour workloads to provide valid information for future capacity planning and testing. Be sure to collect a sufficient amount of data over a long-enough time period to accurately reflect the business use and stress on your infrastructure.

Although capturing baseline data is the necessary starting point for performance and capacity planning, it is just as important to implement methods to continually capture important volumetric information. No environment is static. Having a plan for ongoing capture and analysis of meaningful statistics – leveraging the tools you used to capture the baseline data – will help provide the information you need to identify emerging performance inhibitors as well as reduce the cost of future planning efforts.

Highlights

Using modeling and simulation techniques allows you to project potential loads and evaluate changes before making significant investments.

Capacity planning helps enable you to make informed decisions today and provides a road map for Linux infrastructure investments when business forecasts become a reality.

Simulating the environment

The next step to supporting high performance and scalability for Linux implementations is capacity planning, which helps determine if new loads, applications or changes in current technology will result in an aborted mission. Modeling and simulation are the proven methodologies that allow you to evaluate technology change without risking disruption to your business operations. Projecting potential loads, simulating various forecasted business demands and making technology upgrades can be done in a virtual playroom before you commit to an implementation plan or make significant investments.

Using baseline information, you can associate business and infrastructure metrics to give greater insight about the most resource-intensive business functions and their effect on the infrastructure. This provides for more accurate forecasting when applying anticipated growth rates to the model representation of the environment.

Capacity planning through modeling and simulation of applications built on Linux technologies can help you determine the feasibility of architectural options and evaluate the required hardware resources. Modeling also provides a means for comparing the expected performance of centralized and distributed Linux environments. Capacity planning helps companies make decisions to help prevent potential bottlenecks, as well as develop a strategy for growing the Linux infrastructure "wide" or "tall." It helps answer questions such as: Should we simply add to the current pool of resources, or are significant upgrades needed to address pending performance bottlenecks? Capacity planning arms the CIO and IT management with a road map to follow when business forecasts become a reality.

Highlights

Testing in a live environment can help uncover hidden performance constraints in your Linux environment.

Verifying data used for simulation and validating application design points are a valuable outcome from testing.

Testing the solution

As you are developing your Linux solution and before you begin deployment, testing in a live systems environment can help you validate performance expectations. While capacity planning is very effective for determining requirements for an IT infrastructure based on Linux technologies, the techniques employed – modeling and simulation – assume that the application design was followed exactly, which is rarely the case.

Testing stresses the actual code with user transaction loads to capture performance or scalability bottlenecks inherent in the application. Stress test results can provide a clear understanding of the capacity limits of an application, as well as constraints on the database. Failing to stress your systems with expected and peak demands will limit your ability to uncover performance inhibitors in your Linux environment, such as database locking or contention.

What's more, simulated projections are only as good as the data used to build the model. If the baseline information is not based on solid historical data or actual peak production volume metrics, the results from the simulations will be meaningless. For example, if you project 100 new users a month but you are actually adding 100 new users a week, your capacity planning will be all for naught. Testing greatly reduces this risk by supporting both verification, "I accurately heard what you are telling me," and validation, "I confirmed that what you told me is true." New or changing Linux-based applications that are highly visible to the business should always be tested to validate performance in a "real-world" environment.

Highlights

Follow these process flow steps to safely test your applications without affecting your production environment.

Develop a test plan to determine that all aspects of your Linux infrastructure and business systems are included. The following flow outlines the process used to test an environment with Web-based and batch applications:

- Develop and execute user scripts to isolate specific workloads
- Collect system performance metrics to validate the chosen Linux platform
- Use a network analyzer to develop network traffic patterns
- Collect I/O subsystem metrics
- Validate data quality
- Statistically analyze, aggregate and join the server, network and I/O subsystem performance metrics
- Develop capacity consumption profiles and characterize the online and batch workloads.

To safely test your applications, IBM advises duplicating the actual planned infrastructure resources in a security-enhanced test lab environment. This will allow you to test new or highly constrained resources with current and projected workload profiles without affecting the operational environment.

Monitoring your Linux systems on an ongoing basis provides updated data points that analysts can use for more cost-effective planning and modeling.

Sustaining performance

Maintaining a high-performance Linux infrastructure requires ongoing monitoring of the environment to capture accurate data to make informed tactical and strategic decisions. Once a good baseline and projection are set, modeling is a quicker, more cost-effective way to make further projections. Ongoing monitoring of your Linux systems provides for updated data points that an analyst can use to validate and enhance existing models. Only when you have a major application or business change do you need additional stress testing.

Highlights

Evaluate the cost of effective planning against the potential loss to your business of unplanned downtime.

Scalability and performance planning can produce significant cost savings and help avoid losses associated with poor performance or outages.

Assessing the price

Achieving high performance and scalability in your Linux environment comes with a price. Effective planning requires time and experience, as well as valuable and scarce skills. How do you decide if it's worth the investment?

Reviewing business goals will provide the guidance you need. For client-facing applications, what is the cost of downtime or poor performance? What is the potential for loss of current and future revenue? What will a major outage do to your brand credibility? For Linux solutions supporting your internal business functions, what would be the impact of reduced user productivity or the cost of rework? If the answer is "minimal," then little planning is justified. However, if there is a risk of significant impact, then planning is a small investment to make.

Planning can result in better performance, improved service-level management, more effective use of resources and smoother project implementations. Significant cost savings can be realized when you provision only those resources that are required at any point in time. At the same time, ascertaining that you have adequate capacity in place to handle growth can help limit the potential lost business opportunity and reduce the IT costs related to unplanned downtime. All of these cost savings can, in many cases, more than compensate for the investment in planning. When purchasing is constrained due to economic conditions, capacity planning and testing can also help develop a staged implementation approach that allows you to delay purchasing the optimal configuration until it is really needed.

Summary

Using the methods detailed in this paper, companies can effectively plan for scalability to support ever-changing business demands, taking advantage of Linux technologies and open source applications. To help you hit your target, begin with and keep a focus on business objectives. Follow a set process by first establishing an accurate baseline of your company's infrastructure and applications to eliminate planning errors that can result from invalid assumptions. Taking the time to obtain and maintain an accurate understanding of the existing environment results in more accurate modeling, simulation and testing results. For the highest return on investment resulting in development, implementation and support savings, commit to implementing performance methods into business planning and application development processes. Then, when the business brings new demands, your company can be confident that your Linux infrastructure will deliver.

Find out more

For more information about how IBM can help you deploy high-performing business systems in Linux environments, please contact your IBM representative or visit:

ibm.com/services

IBM Global Services
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
11-03

G510-3697-00