A new imperative: real-time capacity planning

In enterprise data centers, capacity planning is often thought of as a complex, time-consuming exercise reserved for select mission-critical application stacks. Whether it’s completed as an in-house program or an outsourced service, capacity planning is in many cases treated as a series of narrowly targeted projects, each focused on a different application, and each with a clear beginning and end.

If you’re responsible for maintaining the performance of business applications, this conventional approach to planning may give you the confidence that you can meet the demands of the applications that are most important to the business. But it can also be a drain on staff resources and leave you with an inflexible environment that can’t respond quickly to changes in demand.

Today, capacity planning is growing in importance as enterprises turn to virtualization to help control data center costs. In a large data center, server utilization and application workloads can fluctuate dramatically over the course of a few days or a few weeks, or even within a single day. This makes it all the more important to be able to efficiently and quickly plan for capacity needs. What’s needed is accurate capacity planning in real time—“on the fly” and “on the ground.”

This isn’t a vision for the future. The era of real-time capacity planning is here today. This era is made possible by new software tools that capture thousands of data points each day on each server’s utilization and power consumption, and then provide insights into best-fit scenarios.

Technology silos and underutilized capacity

To fully explore current challenges in capacity planning, it helps to look at how server capacity is used in today’s data centers. For the last 15 years, it has been common practice to run one application per server. This approach tends to create technology silos that can lead to chronic underutilization of resources.
When capacity is locked into technology silos, servers assigned to some applications might use just a small fraction of their processing power. These servers take up valuable floor space and drive up power and cooling costs. And all of these challenges are intensified when IT assets are often spread across multiple and varied data centers.

The use of virtualization technologies can help you break down the silos and consolidate applications and workloads onto a smaller number of servers, each of which is better utilized. But to really make the most of your investments, you need to have a better view of your server workloads, available processing power, the utilization of your virtual and physical environment, and server energy consumption on an ongoing basis. What’s more, you need a better view of your opportunities to increase server utilization and improve application performance by shifting workloads to servers with free capacity.

Ideally, today’s common approaches to capacity planning should help you in these efforts. And they do, but only to a limited extent.

Current approaches to capacity planning

If you help keep an enterprise data center up and running, you have undoubtedly wrestled with capacity planning questions. How should we go about consolidating legacy systems? How can we re-balance workloads to improve application performance and take greater advantage of our available processing power? Where should we stack individual workloads within our existing environment?

Elsewhere within the enterprise, executive managers may be asking a complementary set of questions. How can we gain more value from our substantial server investments? How can we control rising data center costs while still maintaining an infrastructure that will support business growth? How can we increase the agility of our business technology to enable new business initiatives and drive better business outcomes?

To help answer questions like these, organizations turn to capacity planning. In some cases, data centers have teams of dedicated experts who use complex software tools to analyze capacity needs, usually for a limited number of mission-critical applications. In other cases, server administrators engage in an ad hoc process that uses spreadsheets and guesswork to develop estimates of the capacity needs of various applications.

When a formal capacity planning program is in place, a planner might produce one capacity plan per week, focusing on a single server running a single application. That plan will most likely be devoted to a mission-critical application. Applications that are less important to the business might never be part of a capacity plan.

In the absence of detailed and definitive information on historical and current server workloads, utilization levels and power usage, capacity planners and server administrators often have little choice but to over-provision resources just to be sure that systems can meet the peak loads of certain applications. So even with a dedicated capacity planning team in place, capacity waste becomes a part of day-to-day business.
To solve the waste problem, you need better knowledge of your environment, including accurate information on system trends and a better understanding of best-fit scenarios. Data alone isn’t enough. You need tools that can quickly sort through and analyze massive amounts of data and then generate meaningful information that you can put to work in your capacity planning efforts. This sharply targeted information is one of the keys to allocating virtual and physical resources in a more flexible manner, avoiding over-provisioning and meeting the peak demands the applications you support.

This is where today’s new tools for capacity planning come into place. These tools arm you with the insight you need to make capacity planning and resource optimization an everyday event in your data center.

Taking capacity planning to a new level

At HP, we have been working actively to apply advanced technologies to the challenge of capacity planning. These technologies, including algorithms developed by HP Labs, are a part of the new HP Insight Dynamics – VSE software suite. This software incorporates HP Smart Solver technology that allows you to automatically analyze and optimize server capacity and power use.

HP Insight Dynamics – VSE collects and analyzes thousands of historical data points from virtual and physical resources each day. It gathers data every five minutes to create a clear record of server utilization—processors, memory, networking and I/O bandwidth—and energy consumption. Then, using the HP Smart Solver, the software identifies the best fit for workloads and highlights opportunities to consolidate servers to improve utilization and cut power and cooling costs.

Using the HP Smart Solver, you can quickly evaluate different scenarios for consolidating older and underutilized physical servers onto virtual machines. You can simulate different workload configurations on different sets of server resources. You can automatically find best-fit solutions for dozens of servers, or even more. And you can easily see which alternatives use the least computing power, the least space and the least energy.

When you need to deploy a new workload, the five-star rating system lets you instantly see the best place to put the workload within your existing infrastructure. And on an ongoing basis, you can use HP Insight Dynamics – VSE to identify servers that often push their capacity limits, and then shift the loads to servers that are often underutilized. For example, a server that runs certain workloads at night, such as a backup application, might sit idle during the day. With this information in hand, you can try out different workload balancing scenarios, and then confidently shift certain daytime workloads to the idle server, so you can take advantage of the open computing capacity.

It’s easy to see how historical CPU utilization for a workload has performed over time.
Beyond processing power, HP Insight Dynamics – VSE helps you overcome your power and cooling challenges by enabling energy-aware capacity planning. It provides the hard data you need to intelligently plan for your server capacity and power needs, taking service levels into account.

For example, you might run scenarios that show what happens under different consolidation approaches and then compare different configurations to determine the best options for power utilization. The software’s five-star rating system shows the best choices for provisioning and redeploying servers in an energy-efficient manner.

These kinds of insights help you ease your energy pains through better utilization of power and cooling resources and better placement of workloads. By quickly identifying and consolidating inefficient servers onto fewer, more power-efficient ones, you can reduce energy consumption and cut your operational costs.

Achieve better business outcomes

The shift to real-time capacity planning can make a big difference in your data center. With this new approach to capacity planning, you’re not locked into the rigid allocation of over-provisioned resources. You can now complement your long-term consolidation initiatives with routine capacity planning and the ongoing optimization of resources.

Through this more advanced approach to capacity planning, you can increase utilization of your servers, improve the performance of the applications, and make better use of your power and cooling resources. In addition, better capacity planning can help you save valuable floor space, avoid costly expansion projects and extend the life of your data center.

These are the types of benefits that can make a difference in your company’s bottom line. Ultimately, better capacity planning can lead to better business outcomes.

To learn more about HP Insight Dynamics – VSE software and its Smart Solver technology, visit www.hp.com/go/insightdynamics.