Why Scale-Up With 4S and 8S Servers?
Article

Scale-up 4-socket (4S) and 8-socket (8S) systems can represent an attractive alternative for a variety of reasons, including IT operational cost reductions because of reduced server management staffing requirements, reduced power and cooling costs, reduced software licensing costs, and reduced IT infrastructure costs. Scale-up can also yield improved performance, including increased resource utilization, improved application performance, less unplanned downtime, and extended datacenter life.

Scale-up advantages:

- Datacenter space is limited and expensive. A smaller number of 4S or 8S servers is more suited and less costly than a larger number of 1S or 2S servers.
- The application workload itself is best performed on fewer machines, each with a high level of compute power and memory.
- Consolidation cost savings including software licensing provides cost saving benefits with the use of fewer servers
- The challenge of setting up a “scale out” servers is not justified by the workload.

Let take a further look at the business value and applications that drive scale-up servers.

Business Value of Scale-Up Deployment

Companies consolidating workloads from scale-out x86 servers to scale-up servers can experience savings in a number of categories, including server and networking cost reduction, licensing fee reduction, and avoidance of costs associated with IT staffing as well as datacenter facilities and power and cooling.

Without consolidation, these costs would have increased because the number of small servers keep increasing to support the workloads and end user support cost keeps increasing. Also, end-user productivity can be enhanced with less downtime or performance issues.

The benefits of a scale-up deployment include the following:

- **Improved performance**

  Scale-up server products are best suited for highly demanding workloads because they provide double the processors, more total cores, double the memory, more IO capabilities giving 4S servers the ability to run workloads faster and quickly access data.

- **Reduced software licensing costs**

  A major factor to consolidating onto scale-up systems included the reduction in the number of instances of supporting software such as operating system licenses and third-party licenses. Since many major software packages are licensed on a per-processor or per-core basis, reducing the number of systems can significantly lower annual software supports cost.

- **Improved IT staff productivity**

  Using scale-up servers reduces the total number of processors and servers under management. With fewer physical servers to maintain, server administrators were freed up to focus on higher-value-added tasks and expand process automation applications without adding overhead. As the number of servers increases, the number of OS instances that need to be managed grows, adding further layers of complexity and reducing IT flexibility. Automation and manageability features associated with scale-up server deployments can further ease the administrative burden and reducing server-related administrative labor.

- **Scalability**

  4S and 8S servers have a tremendous amount of compute, memory, storage and I/O scalability. You can start out with a 2S server and simply upgrade it to a 4S or in some cases even an 8S server. This allows the server to grow with the organization and workloads demands and increases.

- **Reduced energy consumption**

  With a focus on global warming and fluctuating energy cost, organizations now have an objective to reduce energy consumption in its data center. Reducing the server footprint onto fewer, more modern systems can dramatically lower the energy requirements and the power bill for an organization. In some cases, organizations are reaching limitations on both electrical power and the ability to maintain required temperatures. With reduced energy consumption, organizations can extend the useful life of the existing data center and avoid data center expansion.

- **Increased resource utilization (CPU, memory, and network)**

  Combining workloads on scale-up servers rather than having them on dedicated hardware improves utilization of valuable compute resources. In the one-application-per-server deployment model, server utilization is normally very low—averaging between 10 percent and 30 percent—which is a very inefficient use of server resources. Each server needs to be large enough to handle spikes in workload but normally will need only a small part of the server capacity. Applications that are consolidated on a larger server benefit from shared headroom, so consolidating applications can lead to much higher server utilization as excess capacity is reduced significantly.

- **Reduced Server Hardware cost**

  In some cases, by replacing multiple two socket servers with much more powerful 4 and 8 socket servers, organizations can reduce their initial hardware purchase. The benefit is not only the immediate benefit of reducing server hardware today but also deferred future hardware purchases. Reductions in server spending also reduced spending in other infrastructure areas such as networking or storage infrastructure.

- **Reduced end-user productivity losses**

  Scale-up servers with mission critical RAS capabilities can incur less unplanned downtime than scale-out. Less downtime translates to better user productivity with less lost time and less IT help desk supporting better IT productivity.

- **Accelerated Application Deployment**

  With virtualized scale-up servers, clients can accelerate the deployment of new applications initiatives. Projects that previously took weeks to size, purchase, install and configure new server resources, can now be ready in hours. By decoupling specific application requirements from physical servers, clients can set up and reconfigure application environments much more rapidly and make much better use of computing resources.

- **Extended datacenter life**

  Reducing the number of physical servers alleviates pressure on datacenter capacity, enabling organizations to defer expensive new buildouts.

The business value of deploying scale-up x86 servers needs to account for the long-term cost advantages of scale-up server reduced management costs, reduced power and cooling costs, and reduced software licensing costs that accumulate throughout the lifetime of the server deployment.

Customers should take an inventory of where their applications are running today and take advantage of the opportunity to change their traditional patterns for deployment within the datacenter. They should consider new deployment patterns leveraging scale-up x86 servers that will help them reduce the use of datacenter floor space and reduce energy and maintenance costs.

By scaling workloads on fewer, more powerful systems based on four or eight sockets, customers can achieve the very real business benefits of reduced operating costs, reduced power and cooling costs, and reduced software licensing costs. They can improve utilization, improve application performance, and extend datacenter life.

By using scale-up servers in the datacenter and computer rooms, customers can achieve measurable IT and business benefits, reducing the costs to the organization and improving service to end users.

Application Drive Scale-Up Deployments

Some application and workloads lend themselves particularly well to scale-up rather than scale-out deployments. Factors contributing to a decision to scale up rather than scale out include processing and memory requirements, the number of end users, scalability cost and administrative cost.

Scaling up allows you to add CPU, Memory, I/O and Storage to your servers, and it does not require you re-write your application to harness the new horsepower.

For large databases, scale-up architectures can provide higher levels of scalability than large numbers of scale-out distributed databases, and scale-up servers are often easier and less expensive to manage.

Lets take a look at some Enterprise applications that are better suited for 4 and 8 socket scale-up deployments:

- **SAP HANA**

  SAP HANA is a high-performance in-memory database that accelerates data-driven, real-time decision-making and actions, and supports all workloads, with the broadest advanced analytics. Storing data in main memory rather than on disk provides faster data access, queries and processing. Scale-up systems with a large amount of system memory are required to support SAP HANA.

- **SAP Applications**

  SAP is typically deployed in a 3-tier architecture, and the database

- layer cannot be scaled merely by adding servers. Properly scaling to handle larger database loads requires deploying them on hardware that is suited to house both the SAP applications and their associated databases. Scale-up can also be used to consolidate multiple instances of SAP scattered around the organization onto fewer, virtualized, scale-up servers

- **Microsoft SQL Server**

  For larger databases Microsoft designed and enhanced SQL Server to perform better on scale up rather than scale out in a single-instance scenario. While it supports many high-availability features such as clustering, a single SQL Server database is not designed to be deployed across multiple server platforms as a single virtual instance. This means that scaling is best achieved by upgrading the underlying hardware on which the database runs rather than by adding instances on separate servers.

- **Oracle Database**

  While Oracle Database software products support both scale-out and scale-up computing, deploying Oracle Databases on scale-up servers brings a number of benefits, including simplified server management, reduced server licensing costs, and reduced operating costs related to IT staff time and management. When scaling-up, Oracle Real Application Cluster (RAC) is not required. Another benefit is that machine resources (especially CPU) are instantly available for sharing.

- **High Performance Computing**

  HPC workloads such as Genomics require servers provide high throughput volumes and accelerated analytics. Accelerating the workflows and increasing throughput is key for areas such as testing, vaccine design and tracking viruses. Scale-up servers provide the compute, memory, storage and I/O capabilities required for these workloads.

- **Enterprise Custom Applications**

  Custom applications, many of which are legacy applications running on aging centralized hardware, drive many mission-critical workloads in enterprises today. And as demands on these applications grow, the underlying hardware is not always able to keep pace. There are many advantages associated with rehosting custom-developed applications on scale-up x86 servers. First is the opportunity to include workloads on standardized IT infrastructure based on x86 server hardware, running Microsoft Windows or Linux – in place of operating systems that have reached their end of life (EOL). Rehosting on scale-up x86 server hardware also brings greater IT simplicity, reduces operational costs associated with Management, downtime, and reduces server footprint in the datacenter.

Other workloads that benefit from scale-up, deployments that provide single system high performance, large memory and individual server scalability.

- In-Memory analytics
- CRM
- OLTP
- ERP
- IBM DB2
- Business intelligence
- SAS
- Virtualized consolidation
- Business intelligence
- RISC migration
- Decision support
- Large-scale, virtualized applications
- Enterprise-critical middleware
- Large memory, consolidation workloads

As workload performance increases and memory and scalability requirement increase, the need for scale-up workloads and big applications increase. This drives the need for individual system scalability and resiliency.

ThinkSystem 4S and 8S Servers

Lenovo has a family of scale-up 4S and 8S ThinkSystem Servers to meet customer needs.

ThinkSystem SR950 - The unique, modular 4U/8P ThinkSystem SR950 is a resilient, scalable system with large memory and storage capacity that runs the most crucial workloads and demanding applications.

ThinkSystem SR850 - ThinkSystem SR850 is a fast, highly reliable 2- to 4-processor system with large memory and storage capacity, designed for high-growth workloads like databases and virtualization.

ThinkSystem SR850P - ThinkSystem SR850P is a dense, fast, highly reliable four processor system engineered for maximum performance for high-growth, memory intensive workloads.

ThinkSystem SR860 V2 - The 2U/4P SR850 V2 is designed for high-growth workloads; in-memory computing, databases and virtualization, with large memory and storage capacities.

ThinkSystem SR850 V2 - The 4U/4P ThinkSystem SR860 V2 is designed for exceptional price/performance with the storage, GPU, and memory capacity to grow with your business.

Conclusion

By scaling workloads on fewer and more powerful 4S and 8S systems, organizations can achieve real business benefits of reduced operating cost, reduced power and cooling cost and reduced software licensing cost. You will also see improved utilization, application performance and extended datacenter life.

About the author

Randall Lundin is a Senior Product Manager in the Lenovo Infrastructure Solution Group. He is responsible for planning and managing ThinkSystem servers. Randall has also authored and contributed to numerous Lenovo Press publications on ThinkSystem products.

This article is one in a series on the ThinkSystem SR850 V2 and SR860 V2 servers:

- Five Highlights of the Lenovo ThinkSystem SR850 V2
- Five Highlights of the Lenovo ThinkSystem SR860 V2
- **Why Scale-Up With 4S and 8S Servers?**
- Unique Intel Features Available with ThinkSystem SR850 V2 and SR860 V2
- ThinkSystem SR860 V2 is the New 4S Performance Leader
- The Value of Refreshing Your 4-Socket Servers with the ThinkSystem SR860 V2 and SR850 V2
- The Perfect 4-Socket and 8-Socket Servers for SAP HANA
- Total Cost of Ownership Comparison of Running SAP HANA on Lenovo ThinkSystem Servers

Related product families

Product families related to this document are the following:

- 4-Socket Rack Servers
- 8-Socket Rack Servers
- Mission Critical Servers

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION ”AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document, LP1350, was created or updated on July 31, 2020.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP1350
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP1350.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at https://www.lenovo.com/us/en/legal/copytrade/.

The following terms are trademarks of Lenovo in the United States, other countries, or both:
Lenovo®
ThinkSystem®

The following terms are trademarks of other companies:

Intel® is a trademark of Intel Corporation or its subsidiaries.

Linux® is the trademark of Linus Torvalds in the U.S. and other countries.

Microsoft®, SQL Server®, and Windows® are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.