



NETWORKWORLD

Debunking the myths about scale-up architectures

Given the rapid pace of server design innovation, earlier concerns about scale-up servers no longer hold water



1

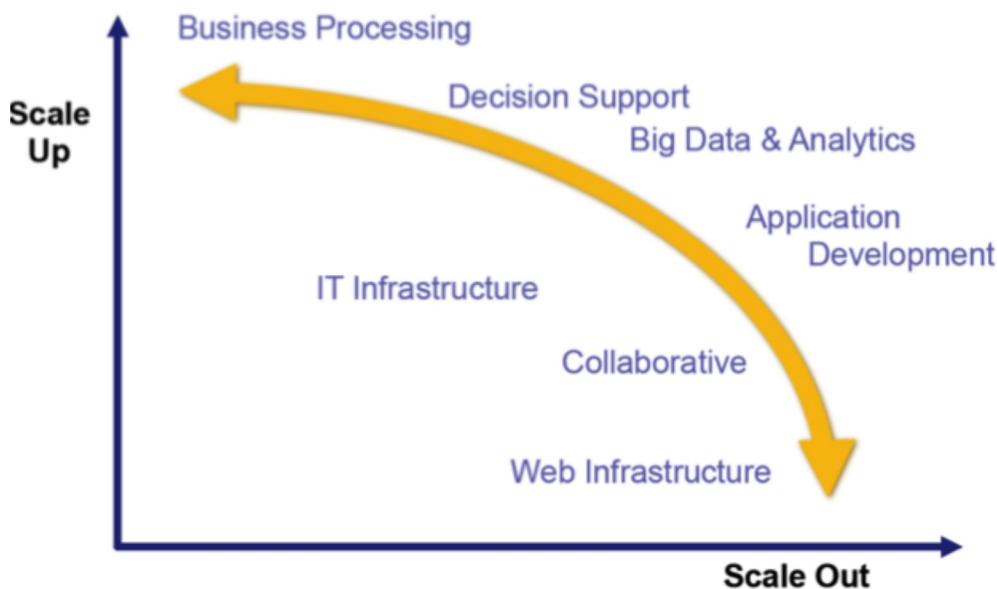
By Ferhat Hatay, Fujitsu Oracle Center of Excellence

Network World | Feb 9, 2015 11:43 AM PT

This vendor-written tech primer has been edited by Network World to eliminate product promotion, but readers should note it will likely favor the submitter's approach.

When growing capacity and power in the data center, the architectural trade-offs between server scale-up vs. scale-out continue to be debated. Both approaches are valid: scale-out adds multiple, smaller servers running in a distributed computing model, while scale-up adds fewer, more powerful servers that are capable of running larger workloads.

Today, much of the buzz is around scale-out architectures, which have been made popular by companies like Facebook and Google, because this architecture is commonly viewed as more cost-effective and “infinitely” scalable.



But, given the rapid pace of server design innovation, earlier concerns about scale-up servers no longer hold water. Newer scalable system designs blend features from both scale-up and scale-out approaches blurring the distinction between the two. Today's modern scale-up architectures bring scalability, capacity and reliability together with the economics of the scale-out model. The scale-up model should now be considered for emerging applications like Big Data and Deep Analytics, especially given its inherent advantages such as globally addressable flat memory space for In-Memory Computing, scalability with low overhead, and easier management.

Let's take a look at the facts behind common scale-up server myths:

Myth #1: Scale-up is prohibitively expensive. The higher cost of larger systems used to be a valid argument because things like special memory, I/O and other components -- while offering key benefits and higher value to the customer -- drove up the cost. Not anymore. Today, modern scale-up systems are designed to use low cost, commodity components as much as possible, debunking the “too expensive”

argument. Plus, fewer larger systems lessen the overhead and are easier to manage than hundreds or even thousands of smaller servers. This is a big win for scale-up systems, since IT departments are looking at overall operating expenses, not just initial acquisition costs.

Myth #2: Scale-out leads to higher reliability. Many IT administrators worry about systems going down and interrupting business operations. The redundancy across multiple systems in a scale-out model holds appeal because the failure of a single server is easily tolerated. Yet the challenge with sprawling, distributed systems has always been mapping workloads and applications across multiple systems and the myriad of complexities and costs it introduces.

Newer scale-up servers build high reliability into every level of the architecture, from processor to component to complete system, for continuous business operations. These systems constantly monitor themselves and can even take proactive measures to ensure uninterrupted operation such as dynamically degrading, off-lining, or replacing failed or failing components on-the-fly. Many of these newer servers also employ physical as well as software-based “partitioning” which provides levels of isolation to improve availability.

Myth #3: Scale-up offers limited scalability. The notion that a single scale-up system is limited to the resources within its physical “box” reflects conventional thinking. The capacity of these systems has grown tremendously over the years. Today’s servers can offer up to hundreds of times higher compute density than previous generations, as well as much more memory and I/O capacity – all “shrunk” into a highly scalable, compact (as small as 1U) footprint that consumes significantly less energy.

These compact, yet powerful servers feature sophisticated innovations borrowed from mainframe computing to ensure the highest levels of reliability. At the other end of the spectrum, some scale-up systems can grow to more than 1,000 processor cores, all in a single system.

Innovations in system interconnect technologies have broken architectural limitations, enabling flexible growth across physical system boundaries with modular “building blocks.” Combining the best of both worlds, dynamic scalability is a powerful feature that merges the large transaction and analytics processing power of scale-up servers with the capacity growth and economic benefits of scale-out servers. Dynamic scaling is a way of bridging to the new world of cloud computing while protecting investments in existing applications.

Unique Benefits of Scale-Up

It’s worth noting that there are additional, unique advantages that scale-up architectures offer. One big advantage is large memory and compute capacity which makes In-Memory Computing possible. This means that large databases can now reside entirely in memory, boosting the analytics performance, as well as speeding up transaction processing. By virtually eliminating disk accesses, database query times can be shortened by many orders of magnitude, leading to *real-time* analytics for greater business productivity, converting wait time to work time.

Scale-up servers that utilize an interconnect versus an external network offer accelerated processing due to reduced software overhead and lower latency in the movement of data between processors and memory across the entire system.

Is it feasible and economical to support both scale-out and scale-up workloads on the same system or class of systems? At the end of the day, it’s a question of how many nodes (scale-out) and the size of each node (scale-up).


For newer workloads like Big Data or Deep Analytics, the scale-up model is a compelling option that should be considered. Given the significant innovations in server design over the past few years, concerns about cost and scalability in the scale-up model have been rendered invalid. With the unique advantages that newer scale-up systems offer, businesses today are realizing that a single scale-up server can process Big Data and other large workloads as well or better than a collection of small scale-out servers in terms of performance, cost, power, and server density.

Hatay is product marketing lead for Fujitsu M10 servers and is part of the Fujitsu Oracle Center of Excellence (FORCE) team. The [Fujitsu M10 server](#) family delivers extreme performance, mainframe-class RAS and near-infinite scalability at extremely affordable price points.



Follow everything from Network World

➤ **Must read: 11 hidden tips and tweaks for Windows 10**

 **View 1 Comment**

YOU MIGHT LIKE

Ads by Revcontent



This Shaving Company Wants To Prove They're Worth

Harry's

Gone Girl Actress Emily Ratajkowski Is Purely Seductive

The Scene

50 Rare Historical Photos That Will Leave You Breathless

LifhackLane

4 Stages Of A Heart Attack - Are You At Risk?

Heart Attack Defender

Fraying at the Edges: Her Fight to Live With Alzheimer's

The New York Times

Replace All Your Wedges With One Club!

XE1 Golf

10 Internet of Things companies to watch

New security fabric to unite Fortinet gear with that of other

Here's what the new Intel will look like

Katy Perry's Actual Salary Will Outrage You

Forbes

Copyright © 1994 - 2016 Network World, Inc. All rights reserved.