Inside Track
Executive Brief

FREEFORM DYNAMICS

The fall and rise of the Mono-socket Server

Why cores matter as much as sockets

in association with

FUJITSU

and

intel

Servers still matter

All the evidence we have, along with our own analysis, says that the future of IT is hybrid. There are many reasons for adopting cloud-based services in some areas, but there are at least as many for keeping other services in house, from concerns over regulatory compliance and the ongoing cost of cloud, to the need for legacy applications that are not cloud-compatible. As a result, most SMBs will continue to invest in servers, while at the same time trying to show ever-better value for money.

For many years now, one of the things that distinguished server-class computers from PCs has been the number of sockets for physical processors: PCs typically had just one, while servers had two – and sometimes more. A single-processor (or mono-socket) PC might have been used as a basic server in a small office, but rarely for anything intensive.

Now, that is changing. More powerful processors are enabling mono-socket machines to do much, much more. In this paper we look at the rationale for the return to a single processor and at how the map of the PC server business is changing.

The socket and the core

Once upon a time, most x86 computers had a single processor in a single-socket. But then, as the demands laid on servers and workstations grew faster than processor performance, vendors developed dual-processor systems, and many buyers switched over. Software developers fairly quickly caught up with the ability to run tasks in parallel and this new configuration became a de-facto standard for many applications.

Some software vendors saw this as a profit opportunity. These dual-socket servers, and even larger ones with four or more sockets, could do more with their software, so they wanted to charge more for them. Many decided to price their software according to the number of sockets, using that as a proxy for compute power.

Then came the realization by processor engineers that they could put more than one physical processor engine (a ‘core’ in chip-speak) within the same chip package. Dual-core processors first appeared in 2005 and are now the norm for computers from smaller servers to laptops, but the chip designers did not stop there. At the time of writing there are 28 and 32-core general-purpose processors available at the very high end, with 48 and 64-core models promised for release during the course of this year.

And lurking in the background is the fact that core performance is largely maxed out. Even where systems aren’t limited in other areas and can fully stretch their processors, there is relatively little scope for silicon engineers to increase the performance of each individual processor core. Instead, processor design is aimed at making cores smaller, not faster. This is a big part of why chip designers have adopted the multi-core approach.

The mono-socket advantage

The advent of these powerful multi-core processors also means that, after years of dual-socket domination, mono-socket servers can do at least as good a job in many cases. They make it practical for small and mid-sized businesses to replace an older dual-core, dual-socket server with a modern one running a single four or six-core processor, say. There are several reasons for this:

Less complexity and cost: Taking out a socket and a processor saves some cost by itself, but it removes other expenses too. Less associated circuitry is needed, and the system design overall is less complex and resource-intensive.

Economical and quiet: The reduction in complexity and component-count possible with mono-socket servers results in a smaller power budget, and therefore less need for cooling. Not only can this reduce energy costs by perhaps 10% to 20%, it also makes fan-less, and therefore near-silent, servers possible for in-office use.

Good thermal properties: Modern multi-core processors leverage low-power technology originally developed for the mobile market. This enables server processors to run cooler and more efficiently than ever before. Cool-running systems can sustain higher ambient temperatures, which makes it practical to run servers outside the data center in many more locations that lack air conditioning.

Cool-running systems can sustain higher ambient temperatures.

Smaller size: Fewer sockets and less need for cooling also translates to a smaller mainboard and a smaller system overall. A powerful server can be housed in a mini-tower for SMB office use or a slim 1U 'pizza-box' rack enclosure for data centers.

Just as importantly for the average business, even a 'regular' server processor can easily have four or six-cores in a single package. With some versions able to run two software 'threads' per core, a small server with just one of these multi-core chips is significantly more powerful than many of the bigger servers available just a few years ago.

Cheaper licensing: Software licensing terms vary considerably from supplier to supplier and from application to application, but business software is often licensed on a per-socket basis. Mono-socket servers can therefore be significantly cheaper to license for some applications, including widely deployed tools such as Oracle and VMware vSphere. Of course, licensing terms vary and can change – some software is already licensed per core, in which case software running on a mono-socket system with a quad-core processor will cost the same as a dual-socket system with dual-core processors.

Time for change?

Even if new mono-socket servers built around multi-core processors can offer benefits, we must ask: why change now? After all, most organizations, from small and midsize businesses to large enterprises, will already have servers installed and running.

The answer is that we are at an inflection point, with at least three factors coinciding, all of them pushing in the direction of a switch to new server and processor architectures. The first is quite simply that many organizations currently have older servers that are coming up for replacement anyway.

Second, the performance and functionality advances brought about by the latest multi-core processors represent a break with the past, whether they are used in mono or multi-socket systems. In previous processor generations, the changes we saw from one generation to the next were relatively small – incremental, even. In those circumstances it made sense to keep older systems in use for longer, whether that was to get more value out of an asset, to avoid generating more electronic scrap, or other reasons.

These new multi-core processors are not just faster versions of what came before.

These new multi-core processors are not just faster versions of what came before, however. For example, they add advanced hardware features to support application privacy and security. Depending on the type of processor, these may include hardware-accelerated encryption, support for partitioning system memory so applications can run in private 'hardened' enclaves, and extensions to assist with network protection and encryption key management.

Lastly, the older server generation has not aged as well as it might have done. In recent years, we have seen numerous adverse factors accumulate, from security vulnerabilities to unsupported new technologies. While each of these problems might be tolerable on its own, in combination they may well justify server hardware refreshes. For example:

I/O and memory restrictions: Many older servers are only 10% to 20% utilized, not because they are processor-bound but because of other constraints. They may not have enough PCIe lanes to support modern networking or solid-state storage, say, or be memory-constrained – the latter is a key limiting factor for virtualization, for instance. The latest members of processor families such as AMD EPYC and Intel Xeon are therefore designed to support many more PCIe lanes and much more RAM.

Spectre/Meltdown/Foreshadow: There are fixes or mitigations for these processor vulnerabilities, either by applying software patches or disabling hardware features. The performance impact from all of this is manageable in the short to medium term, but it is another factor to consider when reviewing replacement plans.

Indirect operational costs: Many factors affect the total cost of operating an older system and therefore the return on investment from an update. Older technology may lose reliability and cost more to maintain, for example, or it might consume more space and power than a more modern alternative.

New storage technologies: Even if they have the I/O capacity, some older systems will lack support for new technologies such as NVMe, etc. This means they will not be receiving the full boost that is possible from incorporating the latest solid-state storage devices.

In conclusion

When processor chips contained just one or two-cores, adding another processor socket was the only way to make a server powerful enough to go beyond the needs of a small office, but this is no longer the case. Instead, when you also take into account the architectural advantages of modern processor designs, a server with a single modern four or six-core processor is likely to outperform a server with two dual-core processors – and do it more safely and efficiently too.

Similarly, when dual-socket was the default for a "general-purpose workhorse" server, you often had to buy a dual-socket machine even if you only put one processor in it. That's because it was the only way to get other server-class features, such as memory capacity, storage, network connectivity, and so on. All that has changed: as well as providing more processor power, modern mono-socket server designs from a reputable supplier can be extremely well equipped in all those other areas too.

Application	Dual-core mono-socket PC	Multi-core mono-socket server	Multi-socket server
Desktop	**	**	
Workstation	*	**	*
File/print/backup server	*	**	*
Virtualization server		**	***
High performance computing	*	**	***
Cloud services	*	**	**

Table 1: Each architecture has its strengths and weaknesses.

From an architectural perspective, there is still a need for multi-socket servers, but it has moved much further up the system performance curve. They may well be required in high performance computing, say, or for complex business intelligence applications, where the need for I/O and system memory outstrips what's possible with a mono-socket server (see Table 1 above). But for many lower-end jobs, such as simpler ERP and analytics roles, a mono-socket system with a four or six-core processor could fit the bill quite nicely.

About Freeform Dynamics

Freeform Dynamics is an IT industry analyst firm. Through our research and insights, we help busy IT and business professionals get up to speed on the latest technology developments, and make better-informed investment decisions.

For more information, and access to our library of free research, please visit www.freeformdynamics.com.

About Fujitsu

Fujitsu is the leading Japanese information and communication technology (ICT) company offering a full range of technology products, solutions and services. Approximately 155,000 Fujitsu people support customers in more than 100 countries. We use our experience and the power of ICT to shape the future of society with our customers.

For more information, please visit www.fujitsu.com.

About Intel

Intel (NASDAQ: INTC), a leader in the semiconductor industry, is shaping the data-centric future with computing and communications technology that is the foundation of the world's innovations. The company's engineering expertise is helping address the world's greatest challenges as well as helping secure, power and connect billions of devices and the infrastructure of the smart, connected world – from the cloud to the network to the edge and everything in between. Find more information about Intel at newsroom.intel.com and intel.com.

Intel and the Intel logo are trademarks of Intel Corporation in the United States and other countries.