IT@Intel Brief

Intel IT
IT Best Practices
Employee Productivity and IT Innovation
November 2010

Optimizing Engineering Productivity with Workstation-centered Computing

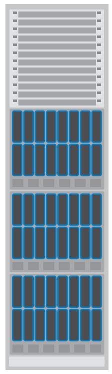
- Uses workstations to extend the benefits of high-performance computing to small design sites.
- Provides performance equal to that of a local data center while reducing power usage by 10 percent.
- Provides additional local compute capacity without capital expenditures for building or retrofitting data centers

Silicon design engineers—located at multiple Intel sites worldwide—use a variety of design applications running on more than 50,000 servers. But not all sites have adequate local data center space to host the compute capacity required to support business-critical design engineering. In particular, some small design sites must host their interactive applications at remote data centers. Slow interactive application response times negatively impact engineers at these sites.

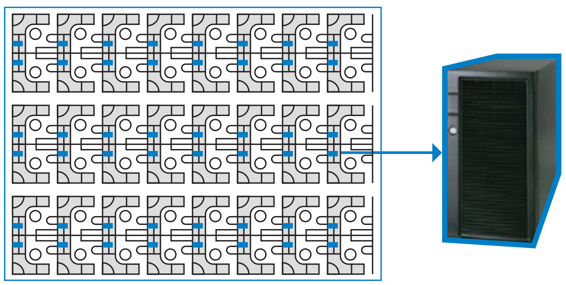
To address this problem, we devised an innovative way to deliver local access to high-performance computing (HPC): workstation-centered computing. Clustered workstations located in office cubicles form a "virtual rack" (see Figure 1), similar to a physical rack of servers in the data center. The result is faster design cycles as well as the ability to quickly meet business needs for compute capacity.

In a pilot project, we found that workstation-centered computing can provide performance, remote manageability, and data security equivalent to that of servers hosted at local data centers. In addition, with design computing requirements growing about 45 percent year over year, we can improve business agility by delivering required computing capability faster and at an overall lower capital investment compared to building new data centers or retrofitting existing ones. We also found that the workstation-centered computing approach can reduce energy-related expenses.

Based on our pilot, we believe that workstation-centered computing can make HPC performance more widely available to companies in other industries and lines of business that don't have available data center space or the expertise to build a specialized HPC environment.



Physical Rack of Servers
- 48x 2-socket blade servers
- Requires 24 square feet of data center space
- 17.2 KW power usage – 358 W per server
- Specialized data center cooling, such as chillers
- Additional data center depreciation



Virtual Rack of Workstations
- 48x 2-socket workstations
- Requires no additional office space
- 15.4 KW power usage – 320 W per workstation (1.8 KW savings)
- Existing office cooling is sufficient
- No additional depreciation

Figure 1. Intel IT aggregated workstations in standard office cubicles to create a virtual high-performance computing (HPC) cluster that delivered local HPC to silicon design engineers. Not to scale.

Business Challenge

Intel's approximately 20,000 silicon design engineers use a variety of design applications to create silicon products. These design applications use two computing methodologies:

- Batch computing, where the design application runs in an automated or scripted mode and can complete thousands of test cases to verify or correct chip designs.
- Interactive computing, where the design engineer interacts with a silicon design in real time using text and graphical modes.

The performance of batch computing is measured by throughput time, with the capacity and performance of the computing environment determining throughput effectiveness. Because engineers are not interacting with the application during batch jobs, the majority of Intel's design batch computing needs can effectively be met by either local or remote data centers. However, a small portion of batch computing must be local so that design engineers can efficiently verify designs with a small set of functional tests before running hundreds of thousands of regression tests.

The performance of interactive computing is measured by an application's response time to the design engineer's input. So long as a design engineer is not waiting for an application to respond, network latency is not impeding engineer productivity. But productivity ratings of interactive silicon design indicate that round-trip network latency greater than 10 milliseconds (ms) slows application response time and impedes design engineer productivity—even with simple two-dimensional graphics. If the network latency exceeds 20 ms, which is possible for some remote design sites, application response time becomes even worse.

Therefore, we concluded that some batch computing and all interactive computing should be co-located with design sites to enhance engineer productivity. However, our design sites are located all over the world—and restricting design sites to only those sites with data centers would reduce our flexibility and operational agility in meeting business demands. Moreover, activities such as building or retrofitting data centers have long lead times and require significant capital expenditures.

Providing our design engineers with tools that help them increase productivity offers a competitive advantage by enabling faster chip design. We therefore needed to find a way to provide the compute power required by engineers at small design sites that don't have data centers—without increasing data center costs.

Solution

We challenged ourselves to take advantage of available technologies to provide local computing capacity and devised a new HPC model called workstation-centered computing. With this approach, we place workstations in office cubicles and interconnect them to form a virtual rack—similar to a physical rack of servers in the data center—offering a highly effective local compute environment for both batch and interactive computing.

TEST SETUP

To test our new model, we conducted a pilot study that placed large-memory workstations in office cubicles, using existing power, cooling, and network resources. The workstations functioned as a virtual rack and were configured to the same specifications as their data center-hosted server counterparts. We repeated the distributed cluster of workstations at three design sites, involving about 100 design engineers.

To make workstation-centered computing an operational reality, the clustered workstations had the following characteristics:

- Cooling. The workstations were able to work with ambient office cooling.
- Acoustics. Workstation noise was compatible with an office work environment.
- Manageability. We could remotely manage the workstations out-of-band, and each featured a single Ethernet connection for both data and system management.
- Physical security. The internals of the workstations were completely locked; we also physically locked the workstations to the cubicle structure.
- Other security items. We disabled all external USB ports and OS hot-keys, installed a locked front panel with the power switch behind the panel, and secured the network ports.
- Intellectual property protection. We encrypted the data stored on the drive. To allow unattended re-boot, we stored the security keys on a centralized server that was inaccessible if the workstation was not connected to the network.
- Reliability. We used multiple power supply connections to help ensure that an accidental cable disconnection did not cause system downtime.

We used simple, innovative solutions to resolve key challenges:

- We worked with the system supplier to provide a single master key for the IT department to allow locking and unlocking of system internals. This prevented the IT administrator from having to manage multiple keys for individual systems and made hardware replacement easier.
- We used a combination-based monitor-port lock, which allowed us to lock a system with a combination instead of managing multiple keys.
- To address intellectual property concerns, we used OS-based secure encryption for the data partition, with additional automation to help ensure that keys were not stored locally, but rather on a centralized secure server.

TEST METHODOLOGY

We tested the runtimes of various workloads using the clustered workstations and compared the results to the runtimes of the same jobs on servers located in a local data center. The workloads included one of or a combination of the following silicon design job types:

- Section router
- Block layout
- Waveform viewing
- Chip layout editing
- Design entry
- Text editing

During the tests, multiple users evaluated and subjectively assigned a productivity rating to their experience with application responsiveness for each type of job. Ratings fell into one of three categories including "crisp," "noticeable slowness," and "annoying slowness." In our experience with these applications, we have found that that improving application response time from "annoying slowness" to "crisp" can sometimes mean as much as a tenfold increase in productivity for some tasks, because the engineer is getting results faster as well as avoiding a high level of frustration.

RESULTS

In the pilot, the clustered workstations' performance of design tasks was virtually identical with similar workloads run on servers located at a local data center, with a network latency of less than 1 ms.

As shown in Table 1, application response time, which is dependent on network latency, is a major factor in design engineers' perception of productivity. Local data center computing provides the baseline data, with network latency of less than 1 ms. When network latency approaches 10 ms, some productivity ratings for design tasks begin to degrade toward "noticeable slowness." And when network latency reaches 20 ms or more, the productivity ratings decline even further—some into the "annoying slowness" range.

Workstation-centered computing is capable of providing design engineers at small, remote design sites with the same application performance as that experienced by engineers with access to a local data center.

Other Results

The pilot also showed that two other factors, noise and temperature, were non-issues. The ambient temperature reading in the cubicles was 74° Fahrenheit (23° Celsius), which was no different from cubicles that did not contain a workstation.

Noise levels are measured in A-weighted decibels (dBA), which is a measure of sound as perceived by the human ear. Noise levels associated with workstation-centered computing were between 45 and 50 dBA, which correlates to a noise level somewhere between a residential area in the evening and a private business office—quite acceptable to the engineers.

FUTURE PLANS

We have already deployed workstation-centered computing at three design sites. We plan to expand the deployment to additional sites in 2011. In selecting sites, we will carefully evaluate whether workstation-centered computing is

Table 1. Effect of Network Latency on Design Application Response Times

Productivity Rating (lower is better)
Assigned by several Intel design engineers during testing

1 – Crisp Application Response
2 – Noticeable Application Slowness
3 – Annoying Application Slowness

	Application Type					
	Section Router	Block Layout	Waveform Viewing	Chip Layout Editing	Basic Digital and Analog Design Entry	Text Editor
Local Computing Resource Local data center or local workstation cluster with round-trip network latency of less than 1 ms	1	1	1	1	1	1
Remote Data Center Computing Resource Round-trip network latency of less than 10 ms	1.8	1.8	1	1.5	1	1
Remote Data Center Computing Resource Round-trip network latency of greater than 20 ms	3	3	1.2	2	2	1.5

Note: Intel internal measurements, January 2007 and March 2010.

Potential Business Impact of Workstation-centered Computing

Although we do not expect to implement workstation-centered computing at every design site for every design engineer, it does have the potential for significant cost-savings in energy use and capital expenditure.

Energy use. If we deployed workstation-centered computing for about 50 percent of Intel's design engineers—approximately 10,000—the clustered workstations would equate to about 200 racks of servers in data centers. These 200 data center racks would consume about 3 megawatts (MW) of power at any given moment, plus another 1.5 MW of power for cooling, assuming a very efficient data center power usage effectiveness (PUE) of 1.5.

In contrast, the virtual racks of workstations would consume only about 2.7 MW—about 10 percent less than the data center servers—and would completely avoid an expensive special cooling infrastructure.

Capital expenditure. If we had to build new data centers to provide locally hosted compute capacity to engineers at small design sites, to keep their network latency to less than 10 ms, this would involve a large capital expenditure. Similarly, retrofitting existing data centers to provide additional compute capacity is also expensive. By locating local compute resources in office cubicles, workstation-centered computing avoids these types of expenses. Workstation-centered computing also avoids the depreciation cost associated with data center rack space.

the correct HPC solution for addressing specific business needs and improving business and employee productivity.

We are also exploring ways to reduce asset management costs by using location-based technologies to automate workstation inventory and location identification.

Conclusion

Workstation-centered computing is feasible and provides much-needed local access to HPC at design centers with constrained data center resources. With faster application response times, design engineers can explore more design options in less time, potentially designing more new products and reducing time to market. Running jobs on workstations delivers timely local results for small and medium-size jobs, while freeing data center HPC resources for quicker throughput on large-scale jobs. Workstation-centered computing can also reduce capital expenditures for building or retrofitting data centers.

IT administrators can manage clustered workstations with the same tools they use to manage data center servers, and the workstations use standard power sockets in existing cubicles along with existing network infrastructure. The workstations consume about 10 percent less energy than their data center-hosted server and do not require the specialized cooling necessary in data centers with high server density. And, with proper standards in place, the workstations provide the necessary physical hardware and data security.

While we have found workstation-centered computing useful for increasing silicon design productivity, it is applicable to any industry experiencing a growing need for large amounts of compute capacity but without the traditional means of getting it from a local data center.

For more information on Intel IT best practices, visit www.intel.com/it.

AUTHORS

Shesha Krishnapura
Senior Principal Engineer, Intel IT

Vipul Lal
Senior Principal Engineer, Intel IT

CONTRIBUTOR

Wesley Shimanek
Workstation Product Manager,
Intel Architecture Group

Printed in USA
1110/KAR/KC/PDF

Please Recycle
324164-001US