Tencent Explores Datacenter Resource-Pooling Using Intel® Rack Scale Architecture (Intel® RSA)

Traditional rack servers do not meet today’s datacenter needs. Tencent worked with Intel on a proof-of-concept involving Intel® RSA technology.

Tencent, a leading provider of Internet Services in China, recently collaborated with Intel on a proof-of-concept to demonstrate that resource-pooling—even in the early stages of development—could bring better experience to users, reduce power consumption, and yield measurable total cost of ownership (TCO) savings.

This paper addresses the reasons for developing a resource-pooled server technology standard; explains some of the challenges we face; offers insights on trends in the development of this technology; and presents some of the findings from the Tencent proof-of-concept.

Over the last decade, Tencent has been involved in several efforts to improve datacenter design, including a partnership with other Internet business leaders to create a rack-design specification, called Project Scorpio. Tencent focuses on maximizing datacenter performance and increasing resource utilization; reducing hardware acquisition, operations, and maintenance costs; reducing datacenter TCO; and ultimately delivering a better user experience to customers.

"As a leading provider of Internet Services in China, Tencent has a keen interest in Intel® RSA technology. How resource-pooling can help us reduce datacenter total cost of ownership while simultaneously improving our customers’ user experience is a key question for our business.”

Herry Wang
Principal Engineer and Architect
Tencent Holdings Limited

---

1. Ninety percent of the world’s data has been created in the last five years, and we add more than 2.5 quintillion bytes to the total every day.¹

2. Worldwide mobile data traffic—projected to grow at a compound annual rate of 57% for the next four years—is expected to reach a throughput of 24.3 exabytes per month by the year 2019.²

3. Datacenters globally consume over 100 GWh per year, a figure industry analysts expect to exceed 130 GWh by 2016.³

4. As measured by the power usage effectiveness (PUE) index defined by The Green Grid, CRAC (computer room air conditioning) units alone can consume as much as half of a datacenter’s power needs.⁴
The shortcomings of rack servers
The rapid expansion of datacenters worldwide has been fueled by the explosive growth in social media and the commoditization of relatively inexpensive rack servers. Ironically, that very same proliferation of rack servers has also introduced new concerns. For example, many servers are not optimally configured for their purposes, which can result in waste and inefficiency.

The “one size fits all” server configuration does not work in today’s datacenter. In resource-intensive environments, a traditional rack server suffers from low operation efficiency and low deployment density. Some negative side-effects include the following:

- In compute-dense applications, unused memory slots, HDD (hard disk drive) slots, and expansion slots negatively affect computing density.
- In memory-dense applications, unused expansion slots and HDD slots waste server “real estate” that could be used for more memory.
- In storage-dense applications, CPUs and memory might be overprovisioned.

As Figure 1 shows, the modern compute environment features a variety of workloads, each with different compute, storage, and I/O needs. With I/O intensity on the horizontal axis and CPU and memory intensity on the vertical axis, it is difficult, perhaps impossible, for a traditional server—with its balanced configuration of computing, memory, and storage resources—to support the wide variety of applications encountered in a modern datacenter.

For years, Tencent and many other cloud service providers (CSPs) worldwide have been customizing servers to perform specific types of tasks from different workloads, provisioning the servers with different hardware configurations. As a result, server specialization has become the norm. Tencent’s datacenter, for example, houses hundreds of thousands of rack servers, more than 90% of which have been custom-provisioned for specific workloads and purposes.

This expansion of server types has partially mitigated the inflexible provisioning of traditional cookie-cutter servers, but it has also led to broader server diversity in the datacenter, which introduces new challenges to server resource management, day-to-day maintenance, and overall datacenter operations. Each new custom server type introduces an additional layer of complexity to datacenter management and maintenance.

Moreover, past efforts to compensate for different application needs and inflexible provisioning have run into unsynchronized server component lifecycles, which have also raised the TCO. For example, CPU performance has been doubling every two to three years, while storage capacity has been doubling at a rate of about five years. Misaligned technology advances such as these produce gaps in server optimization. An all-in-one server, with equally provisioned resources, will see one area of its resources become outdated more frequently than another area. This misalignment of technologies makes it difficult to upgrade to more efficient processors, memory, and storage without unnecessarily discarding still-useful resources.

Compounding this problem is the fact that businesses seldom replace servers as often as they’d like—or as often as they should. The majority of organizations have old, less efficient servers.

Homogeneously provisioned servers could ameliorate these unavoidable mismatches in technology advancements by isolating the technology to specific resource pools.
To address these old and new challenges, we must devise a solution that allows us to intelligently provision servers and manage the newly configured datacenter more efficiently.

Virtualization is not enough

As with many large Internet service providers, Tencent’s datacenter currently features sophisticated virtualization and cloud services, providing end-users with a better experience and minimizing IT expenses. With such an efficient environment, why would an IT department seek to pool its servers’ resources?

Virtualization is a good midpoint phase in datacenter server optimization. With a firm footing in datacenters today, virtualization and cloud implementations have improved resource utilization for datacenter servers; however, CPUs and memory in such environments are still often underutilized, due to the granularity of server resources.

In other words, virtualization and cloud services go part of the way, but they are no substitute for resource pooling.

The first example in Figure 2 shows a 1-to-1 single-server environment, with various apps running on a single operating system on one physical server. To increase capacity, you add more servers, but it is already obvious that this leads to underutilization.

In the 1-to-n virtualized server model (center), a single physical server can divide its physical resources and allocate portions to multiple virtual machines. This does reduce idle resources, but it does not entirely eliminate inefficiencies, as different workloads require specially configured servers that virtualization alone is not equipped to handle.

Pooled resource technology (right) allows a datacenter to allocate resources with even greater efficiency by providing multiserver-to-multinode (m-to-n) resource allocation. Cloud services remain virtualized at the software level, but resource-pooled servers provide further virtualization capability at the hardware level.

These last two approaches are not contradictory, but complementary. With resource-pooled servers, we can continue to run virtualization and cloud server software on logical servers and create virtual machines for end-users. This shrinks the resource allocation hole left by cloud services, improves the utilization of hardware resources, and ultimately reduces TCO. Compared to the generic virtualization of a single server into multiple nodes, multiserver-to-multinode virtualization provides even better utilization of a pool of servers’ resources.

The biggest challenges in resource-pooling are interconnectivity/latency, provisioning, and management software.

---

**Figure 2** Virtualization overdrive. In a one-to-one environment, every physical server is a unique node. Adding users or nodes means adding servers. A virtualized server environment allows more users or nodes using the same resources as a one-to-one environment. Only in a disaggregated server environment, however, does the datacenter approach full resource utilization.
**Interconnectivity and latency**

The first challenge facing developers of a resource-pooling standard is the means of connecting the servers. Regardless of the protocol chosen, the inherent transmission latency of transmitting the data to various servers only complicates the problem.

Each of the many interconnection protocols to choose from—PCIe*, Ethernet, SAS, fabric, etc.—has advantages and disadvantages, and as a result, industry leaders who are developing resource-pooled server standards are still searching for the optimal solution. Using Ethernet as the fabric backbone appears to be a cost-effective and versatile method. However, an open solution should be protocol-neutral, as advances may occur to push another option to the forefront.

To resolve the network latency issue on the OSI physical layer, the Intel® RSA SDV incorporates photonics (light) technology (*Figure 3*), a promising high-speed, low-latency solution. Using light to move huge amounts of data at very high speeds with extremely low power over a thin optical fiber rather than using electrical signals over a copper cable has the added benefit of leaving a small footprint, which allows increased rack/interface density.

After nearly a decade of research and innovation to prove the viability of photonics networking, in September 2013 Intel demonstrated an Intel® RSA-based system with high-speed photonics components—including ClearCurve* optical fiber and the new MXC* optical connector that Intel codeveloped with Corning. A core building block for photonic communication that will help define the way datacenters are built in the future, the protocol-neutral MXC connector offers high speed with a long reach and a relatively low cost. It may be used for network links throughout a datacenter, over existing technologies as well as Terabit Ethernet.

This Intel-based photonic system allows data transfers as fast as 1.6 Tbps (terabits per second) and distances of up to 300 meters.5

Following this successful demonstration, Intel submitted a design guide to the datacenter community, which provides an overview for implementing an intrarack optical interconnect scheme that uses MXC photonic connectors and embedded optical modules to deliver high-speed data rates.

**Provisioning and management**

A resource-pooling implementation requires a new software stack to manage resource discovery, resource allocation, bookkeeping, monitoring, etc. Intel® RSA packages a complete software stack to provide resource-pooling, discovery, configuration, and management all in one (*Figure 4*).

Intel® RSA management software not only provides an API to manage low-level pooled resources, it also...
provides APIs exposed to middle-layer software offered by VMware or OpenStack. The pooled resources generate considerable amounts of provisioning and management data. Traditional IPMI interfaces and protocols, based on the I²C bus, can be used for data communication, but their low data rate and inability to carry information render them incapable of meeting resource-pooling requirements. Intel, Hewlett-Packard, and Dell have formed a joint venture, named Redfish, which will address this issue, and Intel will integrate the solution into the Intel® RSA software architecture.

Note: Intel® RSA does not alter the operation of business software from third-parties, such as Azure, VSphere, or privately developed hypervisors and management stacks. A resource-pooled server is only rearchitected in its hardware design; from the software perspective, resource-pooling is transparent.

Resource-pooling specifications

Intel® RSA is a server resource-pooling specification with a defined architecture, a clear roadmap, and working prototypes—part of Intel’s plan to rearchitect the datacenter by treating compute, storage, memory, and networking as a whole. Although the standard is still in development, Intel released software development vehicles (SDVs) to customers in early 2015, and the Intel® RSA design is on course to debut officially by 2017.

Figure 4 shows the basic concept. Intel® RSA pools compute, network, fabric, and storage resources at the bottom of the stack, and uses a multirack “pod manager” module to serve as the software/firmware that exposes the hardware underneath to the orchestration layers above that manage and enforce policies.

Resource-pooling allows a data-center architect to gather key server resource components into different resource pools, such as a computing pool, a memory pool, a network pool, and a storage pool, as shown in Figure 5).

After disaggregating processors, storage, memory, and networking resources, a datacenter administrator can then flexibly assign those resources to meet the demands of individual workloads. The pooled resources are recomposed into an “instance,” and the underlying operating system and applications run the same way they normally would on a traditional server equipped with the physical resources as if they were on the same physical server. This architecture lets administrators easily manage and scale resources to reduce the total OPEX (operational expenditures)

Note: Memory and networking are difficult to pool; Intel is still working on methods to pool these resources effectively.

---

**Figure 4** Pooled resource management. In addition to the concepts of pooled, disaggregated hardware and interconnectivity solution that overcomes traditional datacenter latency issues, the third piece of the Intel® RSA puzzle is the resource management layer. The Intel® RSA API interoperates with the hardware and various orchestration partners via a pod manager application.
Another benefit of resource-pooling is improved failure recovery. According to Tencent's internal statistic operation data, hard disk drives (HDD) account for the highest failure rate of all their server components. While HDD failure on a traditional server could very likely result in lost data, pooled storage protects against such data loss via redundancy and fast migration. A failed HDD may be replaced/repaired at IT management's convenience.

Under Intel's vision, the disaggregated rack architecture will also feature photonic-based high-bandwidth technology, which provides farther reach using fewer cables and offers extreme power efficiency compared to today's copper-based interconnects.

Tencent and Intel have contributed to similarly focused industry initiatives—such as Project Scorpio, Redfish, the Distributed Management Task Force (DMTF), and the Scalable Platforms Management Forum (SPMF)—to deliver a common baseline for improved datacenter architecture and management.

Many industry analysts expect these collaborative efforts will coalesce to meet the expectations of cloud/web-based IT professionals who seek scalable platform hardware management with existing toolchains, as well as the expectations of end-users who want simple, modern, and secure management of scalable platform hardware.

Emerging datacenter architectures will improve operational efficiency through increased resource utilization and interoperability, and bring significantly increased performance and lower total cost of ownership.

To support these claims, Intel and Tencent partnered to conduct a proof-of-concept that would highlight resource-pooling in a datacenter environment.

### Tencent's proof-of-concept

Tencent implemented a proof-of-concept (POC) in 2014 to demonstrate the benefits of resource-pooling, using a relatively simple "shared boot" exercise. As shown in Figure 6, the POC setup consisted of 20 servers—with their internal hard disk drives removed—functioning as the "computing pool" and one storage server acting as a simplified "storage pool". By allocating 20 compute nodes as the boot drives, the storage pool was able to assemble up to 20 logical servers.

**Note:** The maximum number of server nodes allowed is limited by network bandwidth (10 Gbps in this POC), as shown in Figure 7. On a network with higher throughput, the same quantity of LUNs would be able to support more server nodes.

To serve as the interconnect between the two "pools", Tencent connected the diskless CPU servers with the LUNs on the aggregated single-server storage pool over a 10 Gbps Ethernet network using

---

**Figure 5 Intel® RSA architecture.** Different types of servers for different resource pools. Under the Intel® RSA model, datacenter servers would be optimally provisioned according to the pool they will inhabit. In this example, a typical CPU server—configured with multiple high-performance CPUs and limited or no storage capabilities—would inhabit the computing pool, while a typical storage server—with several large HDDs and light processing resources—would inhabit the storage pool. The Intel® RSA standard will treat differently configured servers appropriately.
iSCSI. The compute servers then attached to the virtual, network-accessible drives, and booted without needing any local storage. Tencent ran a simulated workload on the 20 servers in the computing pool to demonstrate full capability, and found performance to be comparable to a traditional one-to-one 20-node environment.

A shared boot scenario such as this is just one use case to show benefit from the resource-pooling concept. Based on prior internal testing and the actual results from this POC, Intel and Tencent estimate that a typical datacenter could reduce hardware deployment by 6 to 10% operating under the exact same workload.

That does not take into account other projected savings from resource-pooling, such as reduced OPEX, reduced administration staff, better resource utilization, and better workload matching, to name a few.

Moreover, this POC indicates that a datacenter could reduce its number of SKUs by standardizing on fewer server “classes” for deployment to the various resource pools.

Also note that the storage pool in the Tencent POC employed global RAID design, a common feature in most datacenters and SAN farms. This allowed Tencent’s datacenter administrators to use RAID management software (Figure 8) to monitor storage node bandwidth.

Figure 7: Storage node network bandwidth. Tencent’s POC network, with 20 server nodes, required a network bandwidth of at least 1200 Mbps.

Figure 6: POC setup. The Tencent proof-of-concept included a computing pool of 20 servers with HDDs removed and a storage pool consisting of a single server with its HDDs configured into 20 LUNs (logical unit numbers)—one for each server in the computing pool. These devices were connected via a “network pool” consisting of a 10 Gbps Ethernet network using iSCSI.

Tencent was founded in Shenzhen in 1998 and went public on the Main Board of the Hong Kong Stock Exchange in 2004. The Company is one of the constituent stocks of the Hang Seng Index. Tencent seeks to evolve with the Internet by investing in innovation, providing a hospitable environment for partners, and staying close to users.

For more information, visit www.tencent.com/en-us.
the physical drives in the storage pool. If a HDD failure were to occur, there would be no impact on the service layer, as the data would be mirrored across several other physical drives. Additionally, datacenter administrators would be able to fix this type of resource failure with software, rendering 24/7 physical presence of onsite engineering support unnecessary. In this type of environment, whenever the administration staff does decide to replace failed units, physical HDD hotswapping can take place at the IT department’s convenience.

Other cost-saving benefits include independent HW/ SW upgrades, simplified administration, interoperable solutions, and modularity.

Summary

Resource pooling brings flexible and elastic deployment, component redundancy capability, better resource utilization, and lower TCO. Although much remains to be explored, Tencent’s POC demonstrated clear advantages of resource-pooling and highlighted its value in cloud computing.

Tencent will continue efforts to exploit the newest computing technology and to work with partners to arrive at a rack design that would provide a highly competitive server solution for Tencent’s customers. Tencent will also continue to lead and contribute to the Scorpio 3.0 project, to ensure the industry will share this technology’s benefit.

Figure 8 HDD failure in a fault-tolerant RAID volume. RAID logical volume management software allows Tencent datacenter administrators to identify HDD failures and repair/replace faulty drives at regularly scheduled maintenance visits, rather than at the time of failure.

Endnotes

1. IBM estimates, 2013.
3. Based on Forest & Sullivan market survey.

For more information on Redfish and the Intel® RSA standard, visit: http://www.dmtf.org/standards/redfish

Learn how the Intel® DCM SDK can help you address real-time power and thermal monitoring issues in your data center at software.intel.com/datacentermanager.

Find a business solution that is right for your company. Contact your Intel representative or visit the Reference Room at intel.com/references.