SOLUTION BRIEF

Communications Service Providers
Intel® Rack Scale Design

End-to-End 5G Adaptive Edge Over Intel® Rack Scale Design

Intel® Rack Scale Design architecture provides scalable network services for MNOs as well as access to specialized resources including FPGAs and Intel® Movidius™ Vision Processing Units.

Introduction

Multitiered data centers that can flexibly deliver services utilizing compute nodes at the edge of the network are an important development in the provisioning of dynamic services and low latency content delivery.

Many services have peak busy times and customers have had to provision services to meet these peak levels—with the capacity not being used in off-peak hours. Multitiered data centers allow dynamic resource provisioning based on the customer demand, service use case, or the service level agreement (SLA). With these architectures, the customer can provision services that meet their needs dynamically.

In addition, multitiered data centers allow communications service providers (CommSPs) to offer access to specialized processing resources on a dynamic basis. Specialized video analytics or field programmable gate array (FPGA)-based acceleration can be served to a customer for a specific time duration—at night, for example, when security camera video is analyzed or for a specific number of days or weeks to support a compute intensive business project.

For mobile network operators (MNOs), the advent of high bandwidth 5G networks will provide the network capacity to support delivery of these dynamic services. MNOs are now deploying multi-access edge computing (MEC) nodes to deliver cloud, content, and other data center services from cell towers and other locations at the very edge of the network.

At Mobile World Congress 2019, Intel and Lenovo* joined forces to demonstrate a multitiered end-to-end 5G data center environment that would serve the needs of MNOs. The solution utilized Lenovo ThinkSystem* servers that are based on Intel® Rack Scale Design (Intel® RSD) architecture. Some of the design parameters for the solution included the following:

- Flexible platform definition that could include additional servers, CPUs, or FPGAs
- Orchestration and resource management for compute resources located both in the data center and in the core network
- Utilization of standards including Redfish* management APIs and standardized components
- Scaling up and down capability that is driven by application needs

Key Technologies

The demonstration made innovative use of Intel RSD, Lenovo ThinkSystem servers, and technology from a number of other providers. Here are some of the key technologies used in the demonstration:

Intel Rack Scale Design Architecture

Intel RSD architecture is an open industry blueprint for the transition to a software-defined, composable, disaggregated infrastructure (CDI) for data center computing.

In an Intel RSD-based data center, rack-installed disaggregated resource pools comprising compute modules, non-volatile memory modules, hard disk storage modules, FPGA modules, and networking modules can be logically assembled as a composed node (see the complete architecture in Figure 1). Using high-speed, low-latency interconnects, such as 100 Gbps Ethernet, makes disaggregation and composition possible without sacrificing performance. The composed nodes are created using orchestration software and can utilize resources from any of the pools within the rack for a dramatic increase in flexibility in contrast to the fixed ratios of compute, storage, and networking resources that exist in today's servers.

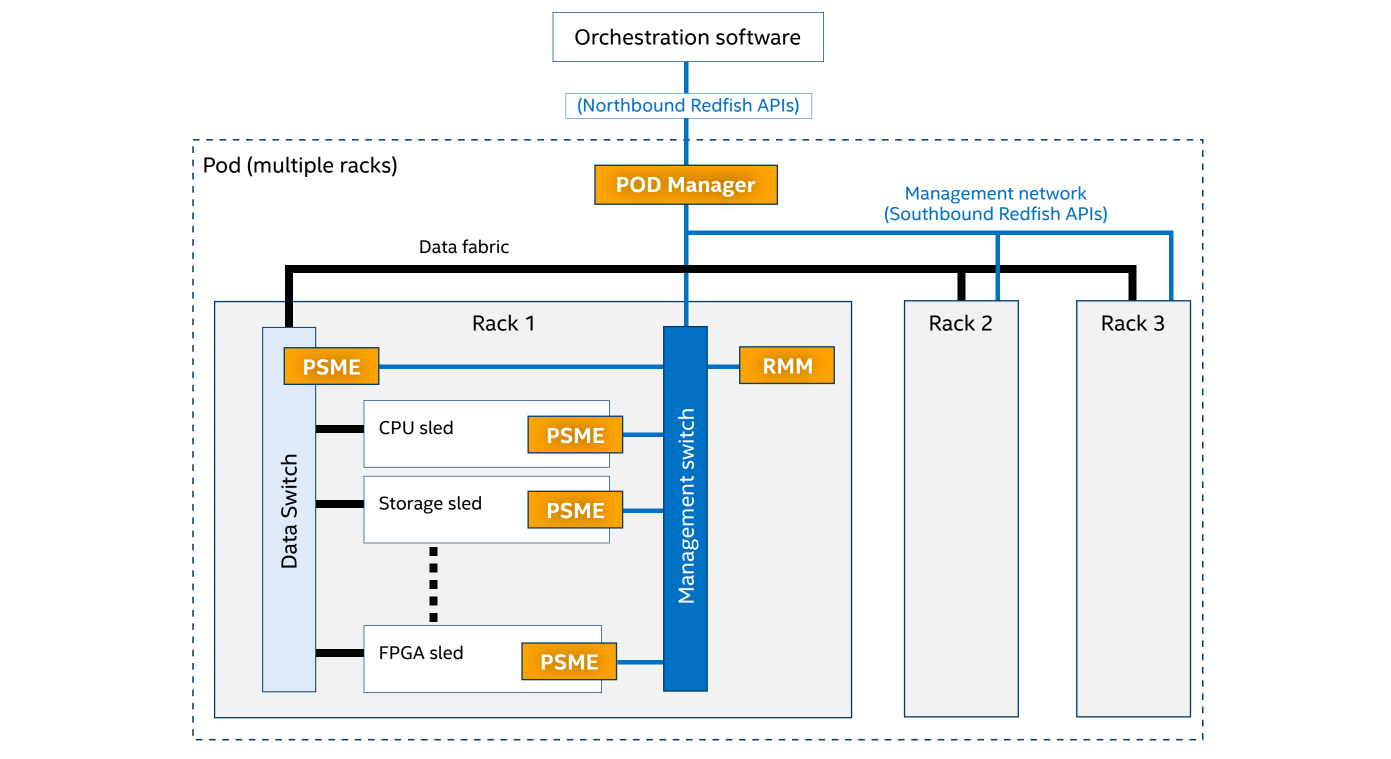


Figure 1. Intel RSD software at the module and rack and pod level communicate to facilitate the creation of composable nodes that can be flexibly configured and reconfigured.

Critical to the composition of these nodes are the Intel RSD management elements, which include the Pod Manager (PODM), the Rack Manager Module (RMM), and the Pooled System Resource Manager (PSME). Every hardware resource in a rack is assigned a PSME that works with the RMM to report availability of resources. Multiple rack systems are called pods and the PODM communicates with the PSMEs via southbound DMTF Redfish* APIs to build the inventory of resources that is exposed to an orchestrator via northbound Redfish APIs. The orchestrator can then request the PODM to compose nodes based on these standard management APIs to meet the application requirements.

When a composed node is no longer needed, its resources can be released back to the resource pools for use by another workload. This enables more efficient use of hardware resources and the ability to quickly react to changes in workload requirements.

Lenovo XClarity* and ThinkSystem Servers

The demonstration solution was based on Lenovo ThinkSystem servers managed by the Lenovo XClarity pod manager, an Intel RSD-based PODM. For the demonstration, the Lenovo XClarity pod manager worked in conjunction with the Lenovo PSMEs and RMMs to provide a single management plane interface to the orchestrator. This allowed for a flexible and dynamic allocation of resources to match workload requirements to optimize efficiency and utilization.

In this demonstration, the solution dynamically allocated disaggregated Non-Volatile Memory Express (NVMe)* storage drives and FPGA resources based upon workload requirements. The orchestration layer instructed the XClarity PODM to logically assign these remote pooled devices to the compute node. This logical connection of the devices was accomplished across the high-speed, low-latency RDMA 100 Gbps Ethernet data fabric. This provided for a very flexible and scalable configuration of system resources as compared to resources that are statically configured in the local server resulting in stranded or insufficient resources in many cases.

Figure 2. Lenovo ThinkSystem* SR630 (left) and SR650 (right).[1]

Servers used in the demonstration included the Lenovo ThinkSystem SR650, the company's two-socket, 2U rack server for service providers needing outstanding reliability, management, and security features, as well as excellent performance and scalability for future growth. The other server used in the demonstration was the Lenovo ThinkSystem SR630, a two-socket, 1U rack server for service providers also needing outstanding reliability, management, and security features inside a compact 1U mechanical form factor.

Both the Lenovo ThinkSystem SR650 and the SR630 servers are designed for a wide range of services including video analytics, speech analytics, retail, and content delivery networks (CDN). Virtual network functions, including firewalls (vFW), broadband network gateways (vBNG), enhanced packet core (vEPC), and IP multimedia subsystems (vIMS) can also be supported with these servers.

Redfish*

Redfish is an open source, open standard API for simple and secure management designed for a software defined data center (SDDC), but can also be used to manage edge cloud data centers. Redfish development is managed by the Distributed Management Task Force (DMTF) and uses common internet and web services standards including RESTful interface, JSON, and OData to expose information to the tool chain used by the data center administrator. Designed to reduce vendor lock-in and increase the productivity of system administrators, DMTF's Redfish API is an open industry standard specification and schema that helps enable simple, secure, and interoperable management of modern scalable platform hardware.

Intel® Distribution of OpenVINO™ Toolkit

The Intel® Distribution of OpenVINO™ toolkit allows the development of applications that emulate human vision. The OpenVINO toolkit utilizes a library of convolutional neural networks (CNN) and provides a common API to enable vision processing workloads across a wide range of Intel® CPUs, including FPGA accelerators and the Intel® Movidius™ Neural Compute Stick.

Intel® Movidius™ Vision Processing Units

Intel® Movidius™ Vision Processing Units (VPUs) are specialized processors for computer vision applications that combine highly parallel programmable compute with workload-specific hardware acceleration. By colocating these components on a common intelligent memory fabric, Intel Movidius VPUs are able to deliver high performance with a very low power footprint. Intel Movidius VPUs can be incorporated into servers, like those in the demonstration, to bring deep neural network and computer vision processing capabilities to these systems.

Intel® Arria® 10 FPGA

Intel® Arria® 10 FPGA family consists of high-performance and power-efficient 20 nm mid-range FPGAs that offer acceleration of networking, storage, and computing workloads. The benefits of using Intel Arria 10 FPGAs for workload acceleration include very fast performance, low power usage, and low total cost of ownership (TCO). This is especially true in complex and data-intensive applications that are pushing the boundaries of data center capabilities.

Other providers of key demo capabilities and systems include:

NBYCOMP
NearbyComputing

Nearby Computing*

Nearby Computing's orchestration engine allows the unified management of wireless and wired compute environments including cloud, MEC, fog, edge, and internet. The software allows simplified deployment processes and life-cycle management of complex and ambitious IoT applications.

A((elleran

Accelleran*

Accelleran provides Mobile RAN/vRAN software and small cell product solutions addressing the real-world challenges of increasing data volumes and 5G applications, calling for hyperdense networks. Accelleran's unique software architecture is genuinely independent from any hardware platform. Customers can leverage one software solution for integrated small cells and for disaggregated and virtualized RAN networks with slicing and edge capabilities. With a global cellular industry's leading design team, Accelleran is the technology choice for carrier grade, reliable, performant small cells and radio access network software solutions for fixed wireless access, public mobile, private network, neutral host or IoT/vertical solution providers.

Qwilt

Qwilt*

Qwilt's Open Edge Cloud* solution for MNOs provides content delivery close to customers to remove latency caused in the transport network. The Qwilt solution includes a virtual CDN (vCDN), deployed in the MEC, which reduces service provider network infrastructure costs and improves quality of experience for streaming video and other real-time applications.

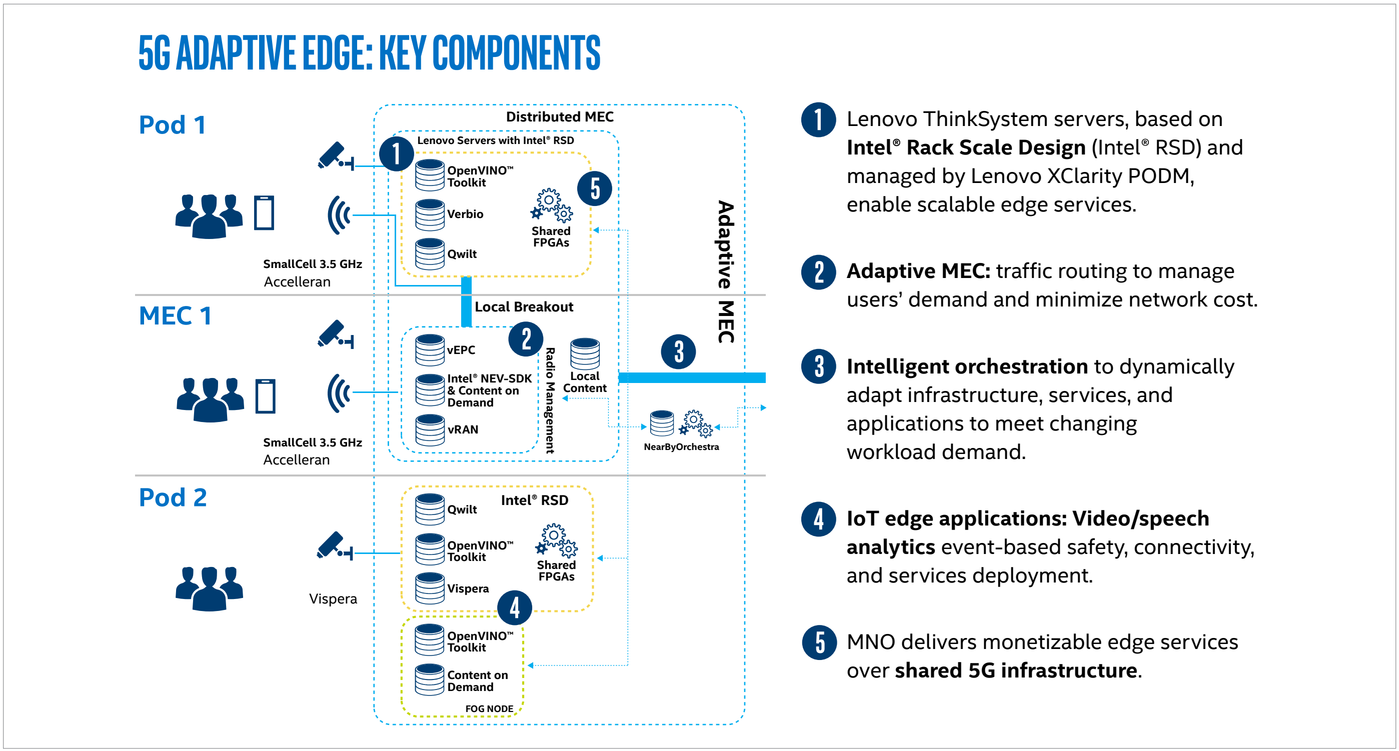
Verbio*
Verbio, a Spanish company that provides deep neural network speech recognition, natural language understanding (NLU) and machine learning technology solutions, contributes with an advanced voice cognitive system integrated with advanced speech analytics deployment at the edge. There are many potential use cases where Verbio speech analytics software running at the edge can deliver tremendous benefits for a real-time experience. Recent breakthroughs in artificial intelligence-based algorithms, including continuous speech recognition (CSR), natural language processing (NLP), speech synthesis or test-to-speech (TTS), and voice biometrics (VB), are now enabling real-time speech analytics.

Cellnex*
Cellnex Telecom is Europe's leading operator of wireless telecommunications infrastructures with a total portfolio of 28,000 sites. Cellnex's business is structured in four major areas: telecommunication infrastructures services; audiovisual broadcasting networks; security and emergency service networks; and solutions for smart urban infrastructure and services management (smart cities and the Internet of Things (IoT)). The company is listed on the continuous market of the Spanish stock exchange and is part of the selective IBEX 35 and EuroStoxx 600 indices. It is also part of the FTSE4GOOD and CDP (Carbon Disclosure Project) and "Standard Ethics" sustainability indexes.

Demonstration Network Details

The demo was designed a show a multitiered distributed edge network architecture based on two pods and a 5G-enabled multi-access edge compute (MEC) node all connected with fiber optic cable. Figure 3 shows the architecture of the adaptive edge solution along with the data flows. There were a number of services deployed in each of the locations meant to support smart city and streaming video delivery use cases.

Pod 1
Pod 1 both delivered CDN and speech analytics services and acted as a resource for extra compute power for video analytics servers in the other pod and MEC. For added video analytics compute resources, the pod featured a composed node with an Intel® Programmable Acceleration Card with Intel® Arria® 10 GX. An OpenVINO toolkit video analytics VNF was running on that node, which could leverage the extra compute power in times of high video traffic, which increases demand for analytics. Also featured was a node with several Intel Movidius VPU-based accelerators for image recognition compute power. These resources were available to all of the nodes in the MEC when video data traffic required additional processing resources.



1 Lenovo ThinkSystem servers, based on **Intel® Rack Scale Design** (Intel® RSD) and managed by Lenovo XClarity PODM, enable scalable edge services.

2 **Adaptive MEC:** traffic routing to manage users' demand and minimize network cost.

3 **Intelligent orchestration** to dynamically adapt infrastructure, services, and applications to meet changing workload demand.

4 **IoT edge applications: Video/speech analytics** event-based safety, connectivity, and services deployment.

5 MNO delivers monetizable edge services over **shared 5G infrastructure**.

Figure 3. 5G adaptive edge architecture.*

In addition, this pod supported the Lenovo XClarity PODM that communicates with the pod manager and node controllers in the other pod and MEC to manage the service deployments and scaling for all of the services.

Pod 2
The services offered in Pod 2 included a CDN, image recognition, visual search, and video analytics. A bank of security cameras that were part of the demo fed live video stream into the demonstration for these services. In addition, an Intel® Core™ i7 processor-based industrial laptop simulated a Fog Node that would be installed in an outdoor IoT environment. On that PC was an IoT gateway for aggregating data from IoT sensors as well as an OpenVINO toolkit analytics VNF for connected security cameras.

MEC 1

The MEC server in this implementation provided the wireless infrastructure for the network. This included virtualized radio access network (vRAN) and evolved packet network (vEPC) running in their own separate containers and connected to cellular base stations in this MEC and in Pod 1.

Results

The demonstration described in this paper showed how a multitiered data center solution based on Intel RSD and Lenovo ThinkSystem servers can be used to deliver dynamic services in a 5G environment. Using this technology, MNOs can provision smart city or video streaming services that can adjust to meet customer demand. MNOs can also make specialized FPGA or VPU-based compute resources available dynamically to applications or customers as demand increases. The flexibility and application interoperability demonstrated in this multitiered data center highlights the ability to install this infrastructure and then utilize technology from Intel® Network Builders ecosystem partners to provision new services and respond to customer demand.

About Lenovo

Lenovo is a global leader in providing innovative consumer, commercial and data center technology with 50,000+ employees operating in 160 countries. The company's products and services include PCs, workstations, servers, storage, networking, software, smart TVs and a family of mobile products like smartphones, tablets and apps. Lenovo servers, storage, networking and solutions work in all data center environments from the simplest to the most complex for all types of workloads. Visit us at http://www.lenovo.com.

About Intel® Network Builders

Intel Network Builders is an ecosystem of infrastructure, software, and technology vendors coming together with communications service providers and end users to accelerate the adoption of solutions based on network functions virtualization (NFV) and software defined networking (SDN) in telecommunications and data center networks. The program offers technical support, matchmaking, and co-marketing opportunities to help facilitate joint collaboration through to the trial and deployment of NFV and SDN solutions. Learn more at http://networkbuilders.intel.com.

[1] Figure 2 provided courtesy of Lenovo.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No product or component can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com.

Cost reduction scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

0219/DO/H09/PDF ♻ Please Recycle 338721-001US