Accelerate Your AI Deployment
Advantech SKY-5240 – Intel Select Solution for HPC & AI Converged Clusters
intel select solution

Deliver the compute-intensive resources needed to run artificial-intelligence (AI) workloads on existing high-performance computing (HPC) with Advantech SKY-5240

Solution Brief

Intel® Select Solutions- High Performance Computing and AI
2nd Generation Intel® Xeon® Scalable Processors
Nov. 2019

intel select solution

Overview

Enterprises believe simulation and modeling, artificial intelligence (AI), and big data analytics can help them achieve breakthrough discoveries and innovation. Enterprises will increasingly seek a new approach to delivering the compute infrastructure needed by AI workloads, with high levels of performance and cost-effectiveness, and without adding the complexity of managing separate, dedicated systems.

Contributing to this belief, one challenge is that fundamental difference in how workloads request resources and how HPC systems allocate them. AI and analytics workloads request compute resources dynamically, an approach that isn't compatible with batch scheduling software used to allocate system resources in HPC clusters. Also, enterprises might not realize that adding these workloads to an existing GPU- based HPC cluster is feasible.

What enterprises need is the ability to run HPC, big data analytics, and AI workloads within the same HPC infrastructure and with optimized resource scheduling software

that helps save time and reduce computing costs.

Advantech, the leading provider for industrial intelligent system, has developed a core competency in reliable and high performance server grade-IPC for over 30 years. Now Advantech has chosen to partner up with Intel and together, SKY-5240 is verified as Intel® Select Solutions for HPC & AI Converged Clusters. This enables users of the systems to benefit from experience in HPC, big data analytics, and simulation markets with workload-optimized performance from 2nd Generation Intel® Xeon® Scalable processors.

This solution brief outlines how Intel® Select Solutions and Advantech SKY-5240 can serve the HPC market with a highly scalable, compact and dense compute cluster utilizing Magpie for AI optimizations. In addition, the 2nd Generation Intel Xeon Scalable processors with its newly added AVX-512 AI instruction set provides a compelling reason to utilize the same infrastructure for HPC and AI Convergence. This speaks very loudly for classic High Performance Computing (HPC) use cases such as global climate change detection, investigations of drug efficacy, cancer diagnosis and classification, and drug screening are taken to the next level of research, where AI Neural networks and AI Deep learning tool chains are rapidly becoming a driving force enabling these new transformation and breakthroughs.

Challenge

The first challenge is a fundamental difference in how workloads request resources and how HPC systems allocate them. AI and analytics workloads request compute resources dynamically, an approach that isn't compatible with batch scheduling software used to allocate system resources in HPC clusters.

The second challenge is the pattern of using computing systems based on graphics processing units (GPUs) as dedicated solutions for AI workloads. Enterprises might not realize that adding these workloads to an existing HPC cluster is feasible without the use of GPUs.

Enterprises can deliver the compute infrastructure needed by AI workloads, with high levels of performance and cost-effectiveness, without adding the complexity of managing separate, dedicated systems.   What they need is the ability to run HPC, big data analytics, and AI workloads within the same HPC infrastructure and with optimized resource scheduling that helps save time and reduce computing costs.

Creating a converged platform to run HPC, AI, and analytics workloads in a single cluster infrastructure such as the ADVANTECH SKY-5240 supports breakthrough innovation, made possible by the convergence of these workloads, while increasing the value and utilization of resources. As a certified Intel Selection Solutions for HPC

& AI Converged Clusters [Magpie*], the SKY-5240 Intel Select Solutions verifies hardware and software stacks optimized across compute, storage, and networking resources for specific workloads. Built on Intel® Xeon® Scalable processors, Intel Select Solutions help ensure enterprises get the performance, agility, and security they require.

Intel Select Solutions for HPC & AI Converged Clusters [Magpie]

These solutions combine Intel Xeon Scalable processors, Intel® Solid State Drives (SSDs), a high-performance parallel file system for storage, and Intel® Omni-Path Architecture (Intel® OPA) to deliver support for multiple types of workloads in the same infrastructure. This multi-workload support means:

- Customers can start their AI journeys on existing HPC infrastructures and potentially reduce the total cost of ownership (TCO) for HPC because Intel Xeon Scalable processor–based HPC environments do not require specialized hardware to run AI workloads

- Faster time to insights with improvements in AI inferencing

- No more burden of data transfer between multiple environments, reducing the time to results for data analytics and AI training runs

- Hybrid workflows supported in the same infrastructure, a capability that allows the solutions to make use of resources and improve efficiency across HPC, AI, and data analytics workloads in a single environment

The SKY-5240 Intel Select Solutions for HPC & AI Converged Clusters, support advanced capabilities to run machine learning, deep-learning training models, and data analytics on the same HPC cluster.   For example, the solutions help users to run Intel® Optimization for TensorFlow* models on an HPC system. TensorFlow is a deep learning framework based on Python* and designed for ease of use and extensibility on modern deep neural networks (DNNs), and it has been optimized for use on Intel Xeon Scalable processors. In addition, Apache Spark* support in the solutions helps with machine learning and data analytics. The solutions also provide a cohesive HPC and AI software stack with integrated open source tools for batch scheduling, which can reduce system complexity and licensing costs and can support hybrid workloads in the same HPC infrastructure.

Intel Select Solutions are verified solutions that combine Intel Xeon Scalable processors and other Intel® technologies into a proven architecture based on the Intel® HPC Platform Specification. This specification defines common industry practices and requirements for building Intel-based clusters. As an architectural foundation, the specification provides a consistent and stable platform, enabling development and deployment of a wide variety of compute-intensive and data-intensive workloads. Included in the foundation are the Intel software performance libraries and runtime environments that allow applications to experience optimized value from the underlying Intel processors and technologies. The Intel HPC Platform

Specification enables organizations to achieve high performance with flexibility, scalability, balance, and portability.

The Intel Select Solutions for HPC & AI Converged Clusters [Magpie] simplify the challenge of building an HPC cluster and are designed to provide optimized performance for highly demanding hybrid workloads. In addition, the solutions are validated to ensure they:

- Include key components and technologies to deliver performance and scalability

- Comply with industry standards and best practices for Intel-based clusters, as defined in the Intel HPC Platform Specification

- Meet or exceed defined performance levels in targeted characteristics important to HPC applications

What Are Intel® Select Solutions?

Intel Select Solutions are pre-defined, workload optimized solutions designed to minimize the challenges of infrastructure evaluation and deployment. Solutions are validated by OEMs/ODMs, certified by ISVs, and verified by Intel. Intel develops these solutions in extensive collaboration with hardware, software, and operating system vendor partners and with the world's leading data center and service providers. Every Intel Select Solution is a tailored combination of Intel data center compute, memory, storage, and network technologies that delivers predictable, trusted, and compelling performance.

To refer to a solution as an Intel Select Solution, a vendor must:
1. Meet the software and hardware stack requirements outlined by the solution's reference-design specifications
2. Replicate or exceed established reference-benchmark test results
3. Publish a solution brief and a detailed implementation guide to facilitate customer deployment

Solution providers can also develop their own optimizations in order to give end customers a simpler, more consistent deployment experience.

Hardware and Software Selections

Intel Select Solutions for HPC & AI Converged Clusters [Magpie] include several key hardware and software components. The solutions are built on top of Intel Select Solutions for Simulation & Modeling, with hardware that provides the right performance for converged HPC, AI, and big data analytics workloads.

Compute

These solutions use Intel® Xeon® Gold 6126 processors or Intel® Xeon® Gold 6226 processors in the "Base" configuration, and Intel® Xeon® Gold 6252 processors in the "Plus" configuration. Both configurations offer all the benefits of Intel Select Solutions; however, the Plus configuration delivers higher performance levels. Intel Xeon Scalable processors feature significant enhancements that can benefit HPC applications, including improvements in input/output (I/O), memory, fabric integration, and Intel® Advanced Vector Extensions 512 (Intel® AVX-512):1
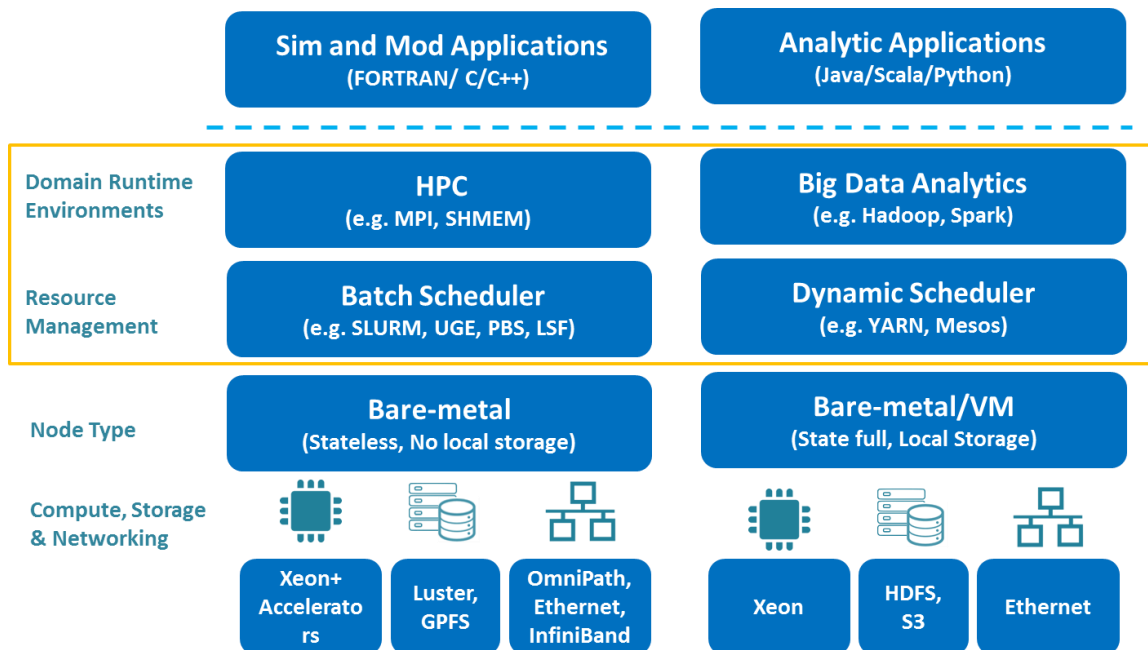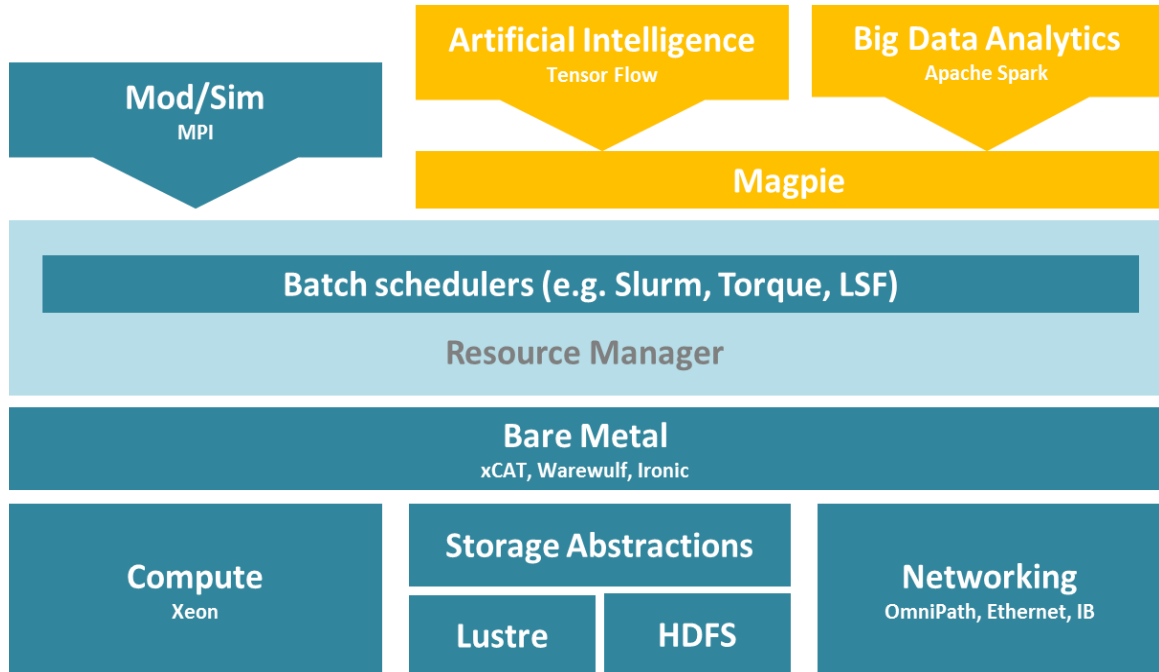
- For HPC users adopting AI, the Intel® Deep Learning Boost (Intel® DL Boost) capability makes the configurations even more compelling because it accelerates AI workloads, increasing Int16* and Int8* peak operations/second. Intel DL Boost was designed to accelerate performance of AI deep learning (inference) workloads (for example, speech recognition, image recognition, object classification, machine translation, and others).

- Existing Intel AVX-512 fused-multiple add (FMA) instructions deliver significant performance for floating-point operations. However, with Intel DL Boost, the performance acceleration extends to integer operations and handles dense computations characteristic of convolutional neural network (CNN) and DNN workloads. The Base and Plus configurations use the following additional hardware:

  ✧ SSDs: Intel® SSD DC 3520

  ✧ Storage: High-performance parallel file system

  ✧ Message fabric: Intel® Omni-Path Host Fabric Interface (Intel® OPA HFI) Adapter 100 Series

  ✧ Management network switch: 10 gigabit Ethernet (GbE) switch

Fabric

Intel OPA provides 100 gigabits per second (Gbps) bandwidth and a low-latency, next-generation fabric for HPC clusters. The 48-port switch chip delivers a 33 percent increase in density over the traditional 36-port switch application specific integrated circuit (ASIC) historically used for InfiniBand* networking, which reduces the number of required switches. Intel OPA can also reduce cabling-related costs, power consumption, space requirements, and ongoing system-maintenance requirements.

Software

Software in the solutions includes a batch scheduler that supports Magpie on Simple Linux* Utility for Resource Management* (SLURM*). As open source software, Magpie is less intrusive to the production software stack than its closed-source counterparts, and it supports multiple resource managers. Additional software in the solution includes the Linux operating system, Intel® Cluster Checker, OpenHPC*, Intel® Omni-Path Software (Intel® OP Software), Intel® Parallel Studio XE 2019 Cluster Edition, Apache Spark, TensorFlow, and Horovod*.

Mod/Sim MPI	Artificial Intelligence Tensor Flow	Big Data Analytics Apache Spark

Magpie

Batch schedulers (e.g. Slurm, Torque, LSF)

Resource Manager

Bare Metal
xCAT, Warewulf, Ironic

Compute Xeon	Storage Abstractions Lustre · HDFS	Networking OmniPath, Ethernet, IB

	Sim and Mod Applications (FORTRAN/ C/C++)	Analytic Applications (Java/Scala/Python)
Domain Runtime Environments	HPC (e.g. MPI, SHMEM)	Big Data Analytics (e.g. Hadoop, Spark)
Resource Management	Batch Scheduler (e.g. SLURM, UGE, PBS, LSF)	Dynamic Scheduler (e.g. YARN, Mesos)
Node Type	Bare-metal (Stateless, No local storage)	Bare-metal/VM (State full, Local Storage)
Compute, Storage & Networking	Xeon+ Accelerators · Luster, GPFS · OmniPath, Ethernet, InfiniBand	Xeon · HDFS, S3 · Ethernet

Verified Performance through Benchmark Testing

Advantech SKY-5240 Intel Select Solutions has been verified to meet a specified minimum level of workload-optimized performance capabilities. Intel Select Solutions for HPC & AI Converged Clusters [Magpie] use the same performance watermarks as the Intel Select Solutions for Simulation & Modeling, which demonstrate optimized capabilities for HPC applications. The SKY-5240 exceeds design and testing standards across eight well-known industry benchmarks that cover important system aspects and indicate potential scaleup and scale-out performance for big data and AI workloads. Intel Select Solutions for HPC & AI Converged Clusters [Magpie] also use the following benchmarks to verify performance: the TensorFlow ResNet 50* benchmark and the Spark-Bench* suite of tests.

The Base configuration specifies the minimum required performance capability. The Plus configuration delivers higher performance for running AI workloads. Core capabilities in Intel Select Solutions for HPC & AI Converged Clusters [Magpie] are delivered by a solution that runs AI workloads within an HPC environment. The architecture enables HPC batch schedulers to run all workloads—including simulation and modeling, big data analytics, and AI—on a common HPC infrastructure. It also enables partners to help customers build upon existing HPC investments to start running AI and big data workloads.

Joint Validation Collaboration and Solution Delivery

Advantech and Intel has worked together to give accelerate HPC and AI solution delivery through a hyper-converged cluster approach with virtualization and containers as the underlining technology to converge HPC and AI into one homogenous harmonized software/hardware framework using Intel Select Solutions for HPC and AI Converged Clusters on Magpie.   By using AI methods with traditional HPC workflows, scientific discovery and innovation processes can be driven faster.

The ADVANTECH SKY-5240 Intel Select Solutions delivers HPC and AI software frameworks in a compact 2U hyper-converged high compute density low foot print and scalable platform. Magpie greatly simplifies running Big Data and AI frameworks on HPC. The Select Solutions is available for download from github.com/LLNL/magpie.

Performance can be demonstrated and total cost of ownership can be reduced as significant incremental investment in Intel's portfolio has enabled the vision to have a converged workload approach reducing the need to consider countless hardware and setup.

AI Begin with Intel Select Solutions

So, as a rule of thumb, most enterprises are just beginning their AI journey with Intel Xeon Scalable processors. It's already the standard for deep learning inference in the data center, and is now more capable than ever for deep learning training thanks in large part to all the software optimizations in the past 1-2 years. Such examples

would be the continuous extension of Intel Xeon Scalable processors AI performance that comes with each new generation, especially now that new AI features are being built into the silicon architecture.

Intel® Select Solutions also make it easier to evaluate and deploy hardware, by providing standardized specifications optimized for particular workloads, which OEMs such as Advantech then create products for.

The Intel® Select Solution for BigDL on Apache Spark is a turnkey Analytics/AI solution integrated with Intel Xeon Scalable processors and Intel SSDs. This solution will help to deliver excellent total cost of ownership by leveraging general purpose IT and standard Big Data platforms for scalable analytics and AI solutions. This solution will also accelerate Apache Spark-based Analytics/AI time to market with a rich developer toolset, optimized libraries and analytics pipelines.

Key benefits of investing in an Intel Select Solution from your preferred data center solution provider include:

- **Simplified evaluation.** New workload integration and the transition to software-defined infrastructure are two areas where IT managers spend more and more time and money sorting through endless options, searching for optimal solutions.   Intel Select Solutions are tightly specified in terms of HW and SW components to eliminate guesswork and speed decision-making.
- **Fast and easy deployment.** With pre-defined settings and rigorous system-wide tuning, Intel Select Solutions are designed to increase efficiency in IT testing process, speed time to service delivery, and increase confidence in solution performance.
- **Workload-optimized performance.** Intel Select Solution configurations are designed by Intel and our partners to deliver to a performance threshold for the workload, and are built on the latest Intel architecture foundation including the recently-launched Intel® Xeon® Scalable platforms.

Solution delivery partners are required to verify that they have matched the specified configuration - and have met or exceeded specified performance benchmark thresholds - by submitting their results to Intel engineers.   In addition to validating their results, solution delivery partners can also add unique features and variations to fit their customers' needs, and are expected to publish detailed implementation guides that significantly reduce infrastructure evaluation and deployment time and expense.

> **The Advantech system with Advantech SKY-5240 are verified Intel Select Solutions for HPC & AI Converged Clusters that benefit end users and developers with faster access to optimized and stable platform configurations to accelerate use and development of high performance computing solutions.**
>
> **Ready to ship as pre-configured platforms and also available for benchmarking in Advantech's Remote Evaluation Service labs, interested users can utilize these platforms to conduct testing and modeling of solutions that will define next-generation compute, simulation and modeling capabilities.**

Configurations

Infrastructure solutions are a key target for today's complex workloads, based on 2nd Generation Intel Xeon Scalable processors with high speed network interconnects. Also, the new Intel Select Solution for HPC & AI Converged Clusters eases multiple customers and end users through complex selection processes to help them make smarter and faster price-performance choices based on data. The Intel Select Solution for HPC & AI Converged Clusters performance requirements have been established using HPL, HPCG, DGEMM, STREAM, IMB and PingPong benchmarks. Intel Select Solutions consists of select hardware, various 2nd Generation Intel® Xeon® processor technologies, Intel® Omni-Path interconnect along with optimized software and firmware configurations. It consists of the following components.

The system must be comprised of at least one Advantech SKY-5240 as compute node.

Hardware	Intel Reference Platform	Advantech System
CPU	2 x Intel® Xeon® Gold 6126 processor (2.60 GHz, 12 cores, 24 threads), Intel Xeon Gold 6226 processor (2.70 GHz, 12 cores, 24 threads), or a higher model number Intel Xeon Scalable processor	Intel® Xeon® Gold 6150 x2
Memory	192 GB or higher with 2 GB/core or higher, 1 DPC 2666 MHz	32GB 2666MHz x12
Message Fabric	Intel® Omni-Path Host Fabric Interface (Intel® OP HFI) Adapter 100 Series	Intel OP HFI Adapter 100 Series x1
Data Drive	Intel® SSD DC S3520 Series or better, or Intel® SSD DC P3520 Series or better	Intel® SSD DC S3610 400GB x1

The Advantech SKY-5240 server, it is a highly configurable and high performance server designed to balance server-class processing with flexible I/O and offload density in a 20" depth chassis. The system is a cost effective, robust platform optimized for high reliability in network, edge and industrial computing.

It is specifically designed for high density PCIe card payloads where maximum I/O connectivity is needed or the integration of industry leading offload and acceleration technology is essential. Equipped with flexible I/O options, it is easy to upgrade to 1G/10G/40G/100G LAN via daughter boards.

Architected around the Cutting edge technologies with Intel Xeon Scalable processor family and support up to 24 DIMMs per Node (Intel® Xeon® Gold 6150 processors in Intel Select Solution for HPC & AI Converged Clusters)

Advantech SKY-5240 Overview

- 2U High-density Multi-Node Server (2U, 4 Node)
- Support 8 2nd Generation Intel Xeon Scalable processor (up to 140W, LGA3647 socket)
- 24 x2.5" SASIII / SATAIII hot-swap bay (6 HDD per node, 2 of them support NVMe)
- Power Supply: 1+1 2200W Redundant Power Supply2x M.2 storage
- System fan: 6x80x38 fan (2 for CPU, 4 for card carge)
- Certificate: CE/FCC class A

- Specification per Node:
  *24xDDR4 DIMM slot support ECC-REG 2400/2133, Max. 768GB per Node.
  *2x M.2 storage
  *2x Intel X557 10Gb LAN
  *Remote Management 1x dedicate IPMI LAN
  *2x Low Profile PCIex16 Expansion
- Specification per System (with 4 Node)
  *96*DDR4 DIMM slot support ECC-REG 2400/2133, Max. 3072GB per System
  *8x M.2 storage
  *8xIntel X557 10Gb LAN
  *Remote Management: 4x dedicate IPMI LAN
  *8x Low Profile PCIex16 Expansion

	PWR on/off all Nodes in sequence
	Mute the buzzer, turn off location LED
	PSU LED
	Fan fault LED
	Overheat LED
	Location LED
	Node Status LED
	Node Alert LED

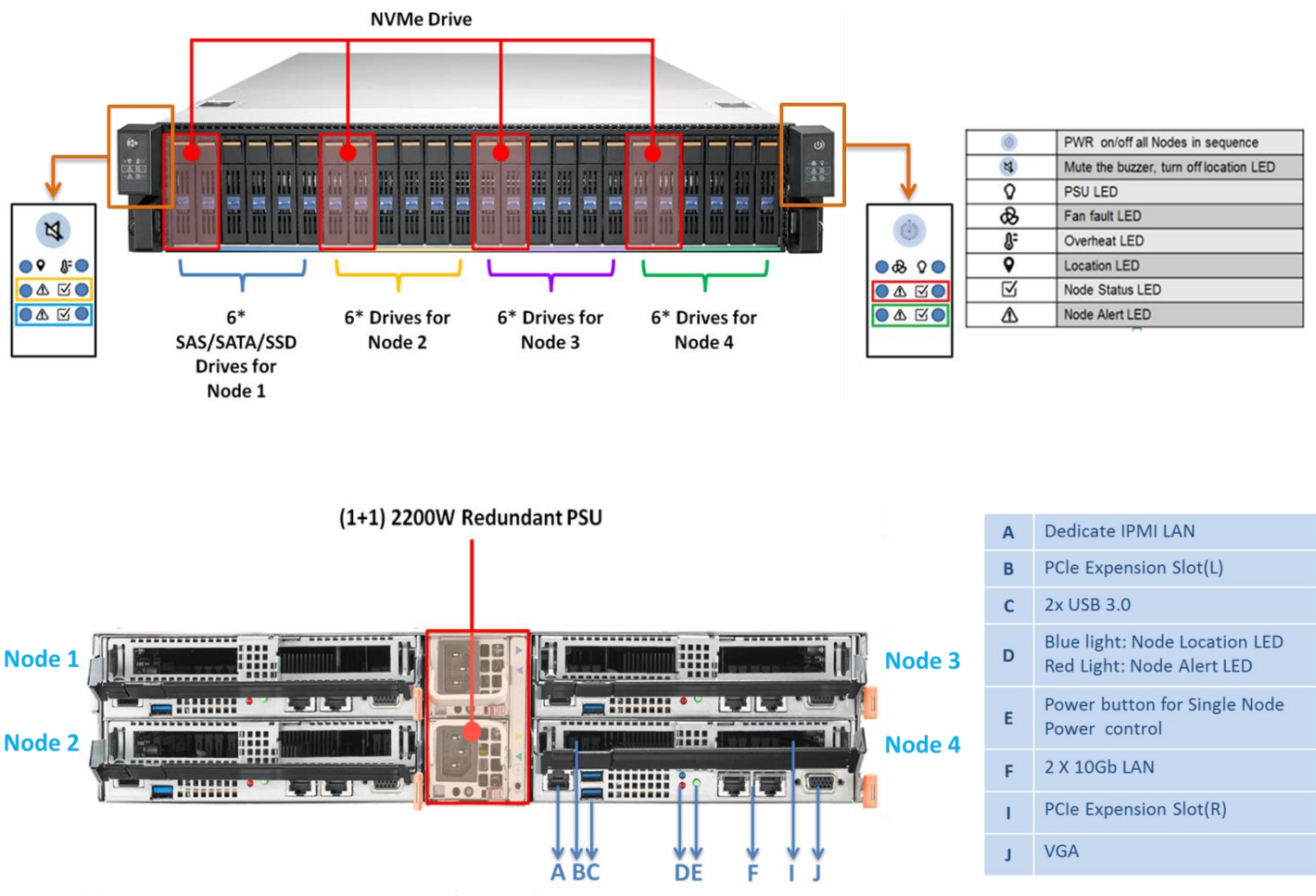|	
A	Dedicate IPMI LAN
B	PCIe Expension Slot(L)
C	2x USB 3.0
D	Blue light: Node Location LED Red Light: Node Alert LED
E	Power button for Single Node Power control
F	2 X 10Gb LAN
I	PCIe Expension Slot(R)
J	VGA

Figure3. Advantech SKY-5240 Server Front View and Rear View

For more information of Advantech SKY-5240 verified Intel Select Solutions for HPC & AI Converged Clusters, please go to the following website or contact Advantech directly:

- **Advantech's Intel Select Solutions for HPC & AI Converged Clusters**

- **SKY-5240**

- **Advantech Website**

Advantech Contact Information

Email: solution.iiot.ana@advantech.com

Hotline: Europe: 00-800-24-26-80-80 | USA: 1-800-205-7940.

For more regional numbers: http://www.advantech.com/contact/