vantageo™

Enterprise Products & Solutions

Made in India for the world

www.vantageo.com

VANTAGEO at a glance

Founded in 2019
Leadership 2 decade + from leading product companies

Mainboard, SSD, DDR4 RAM, NVME are Made in India at global standards

Production Units at Delhi Gujarat & Thane Factory capacity 1,000 units PM

100% Bootstrapped with Positive Net worth

Ranked 14 in top 100 Supercomputer list (by CDAC)
1st Made in India

100+ Channel Partner across India

Global Certifications
VMware Mar 23
RHEL Apr 23
Microsoft Oct 23
Nvidia

280+ Workforce for R&D and After Sales Support across India with 24x7 TAC to L2 & L3

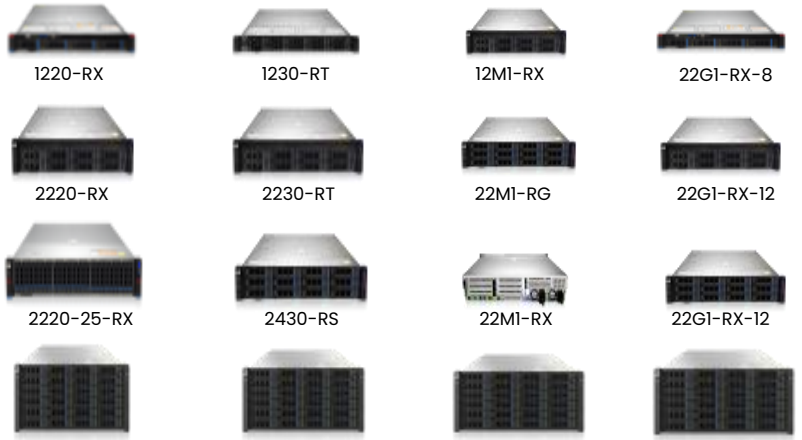
Open Source Engineered System
AI based
HPC system
Object Storage
OpenShift Platform

Successfully build and commissioned 1 Petaflop Super computer for Chandrayaan 3 mission

VANTAGEO Run-Rate Products

Enterprise Servers

Purley	Whitley	Genoa	Milan
1220-RX	1230-RT	12M1-RX	22G1-RX-8
2220-RX	2230-RT	22M1-RG	22G1-RX-12
2220-25-RX	2430-RS	22M1-RX	22G1-RX-12
2420-RX-36	2430-RX-36	14M1-RX-24	24G1-RX-24



High Density Storage

2130-RX-24-ST-2.5

2230-RS-24-ST

1430-RX-24ST-3.5

2430-RS-60-ST

2230-RS-90-ST-3.5

Edge Computing

2230-RX-T

22M1-RX-T

Intelligent Computing

HPC Nodes

2420-RX-GPU

2420-RX-GPU

2230-RS-GP

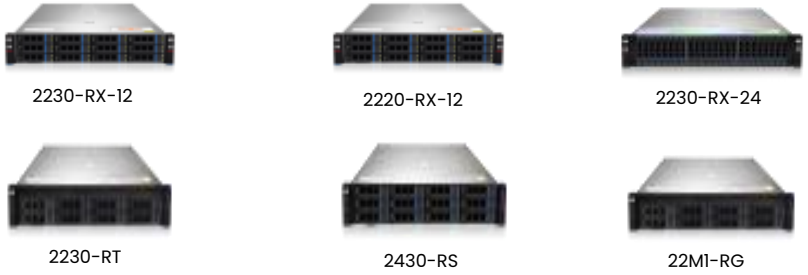
2540-RG-GPU

2240-RG-GPU

2530-WS

EXTEN-SCALE Series

Hyperscale-Hyper Converged - Compute Nodes



2230-RX-12	2220-RX-12	2230-RX-24
2230-RT	2430-RS	22M1-RG

Eminent Series Workstation

Content Creation

15E0-WS | 15G0-WS | 15M0-WS | 15D1-WS | 15R1-WS | 15M1-WS

Intensive Series Workstation

Machine Learning & Rendering

25M1-WG | 2530-WT | 2530-WS

Vantageo Solutions & Appliances

VANTAGEO HPC SOLUTIONS

General Purpose Computing

Scientific Research Supercomputing

Scientific Research Supercomputing

VANTAGEO AI SOLUTIONS

NVIDIA intel AMD

VANTAGEO TEQSO Cloud

TEQSO CLOUD

Hyperscaler on Premises

HYPERSTACK

Open Platform for HCI, SDN, SDS & Containers

VANTAGEO REDHAT OPENSHIFT – Certified Appliances

OPENSHIFT

VANTAGEO Storage Solutions

DATASCALE

VANTAGEO HPC Appliances Portfolio

Rack Servers

U 10 x 2.5”
•Dual Intel® Xeon Scalable Processors
•24 memory slots
•2 x 10GbE

2U 12 x 3.5”
Dual Intel® Xeon® Scalable Processors
12 memory slots
2 x GbE

2U 4 node, 24 x 2.5"
4 x AMD EPYC™ single socket
3 x16 PCIe Gen. 4 slots each node

Multi Nodes

Multi-node Server delivers the highest performance and efficiency in a **multi-node** design creating the flexibility to deploy independent workloads on a shared chassis infrastructure, including cooling and power

GPU Nodes

• 2U dual GPU server supports 8 hot swap 3.5” SATA drive bays
• Supports 2nd Gen. Intel® Xeon® Scalable processors (CascadeLakeRefresh/Cascade Lake/Skylake)
• Intel® C620 Series Chipset to provide 5+ years product life cycle
• Flexible I/O usage with Max I/O® to support 4 x PCIe Gen3 x16 slots

Storage Server

• 4U general purpose storage server supports 24 hot swap 3.5” SAS drive bays
• Supports 2nd Gen. Intel® Xeon® Scalable processors (Cascade Lake Refresh/Cascade Lake/Skylake)
• With Intel® Lewisburg C620 series Chipset to provide 5+ years product life cycle
• Onboard Baseboard Management Controller for system management and IPMI control

VANTAGEO HPC Architecture

Open Source Cluster Manager

Open Source Job Scheduler

HPC Applications

Operating System layer (Open source Linux & Commercial Linux)

Secondary Interconnect 10G / 25 / 40 / 100G Gigabit network

Primary Interconnect High speed 400 / 200 Gbps InfiniBand

Master Nodes

Compute Node

GPU Nodes

- **Parallel File System on Lustre**
- **OSS & MDS Nodes**
- **Hybrid Storage Capacity**

NVIDIA INTEL AMD

VANTAGEO NVIDIA AI Software Solutions Stack

NVIDIA AI Enterprise Software Stack

- AI and Data Science Tools and Frameworks
  - Tensor Flow
  - PyTorch
  - NVIDIA RAPIDS
  - NVIDIA TensorRT
  - NVIDIA Triton Inference Server
- Cloud-Native Deployment
  - NVIDIA TensorRT
  - NVIDIA Triton Inference Server
- Infrastructure Optimization
  - NVIDIA vGPU
  - NVIDIA Magnum IO
  - NVIDIA CUDA-X AI
- Cloud Orchestration
  - Linux Operating System
  - Containerization

NVIDIA SmartNIC/DPU
Network and Infrastructure Acceleration

NVIDIA GPU
Application Acceleration

Vantageo Appliance


2530-WS


2230-RS-GPU


2240-RG-GPU


2540-RG-GPU


2430-RS-60-ST

VANTAGEO AI 2450-RG-GPU Quick View

Most High-end CPU support

Under 35 degree air cool, with excellent thermal solution, 2540-RG-GPU series is able to support the high end CPU on Sapphire Rapid/Genoa platform.

700W H100 SXM5

With liquid cool support, 2540-RG-GPU is able to excel the best performance of HGX H100.

Front Serviceable GPU Tray

Easy access GPU tray for H100 SXM5 for maintenance

Advance Cooling

Partner with CoolIT, 2540-RG-GPU supports D2C liquid cool on CPU/GPU/NVLink to have better PUE and reduce energy consumption

Root Port Options

With different application requirement, offers either 4-root ports and 8-root ports

1:1 Balance Design

with RDMA/NVMe Direct Storage

With same number GPU & storage & PCIe slots under each PLX, it is perfect to use GPU RDMA & NVMe direct storage .

6x CRPS Redundant PSU

With option for CRPS 3600W PSU, it's able to reach N+N.

Standard	Support DP	Up to	Support	Support	Up to	Supports for
5U	**350W**	**32x 4800** MHZ	**8x HGX H100**	**Liquid Cool**	**8x U.2+4x M.2**	**12x**
Form Factor	TDP	DDR5 Memory	GPU	CPU/GPU/NVLink	Drives	Expansion Slots

Intel's Generative AI Silicon & Systems

For building and deploying production LLMs across the enterprise

For dedicated deep learning training and inference of LLMs

General purpose severs with CPU AI acceleration for fine tuning and running expert AI models with other applications on existing cloud native IT infrastructure and at the Edge

Purpose-built servers and clusters for running deep learning workloads of all sizes and across multi tenant in datacenters for fastest TTM and lowest latency

Core Processors

Intel® Core™ processor Meteor Lake

Xeon Workstations Servers and racks

4th Gen Intel® Xeon® SP

Gaudi®2 Server

Dual-socket Xeon SP with 8 Gaudi® 2 devices

1 MegaPOD

8 Gaudi®2 Servers + 3 400G leaf switches

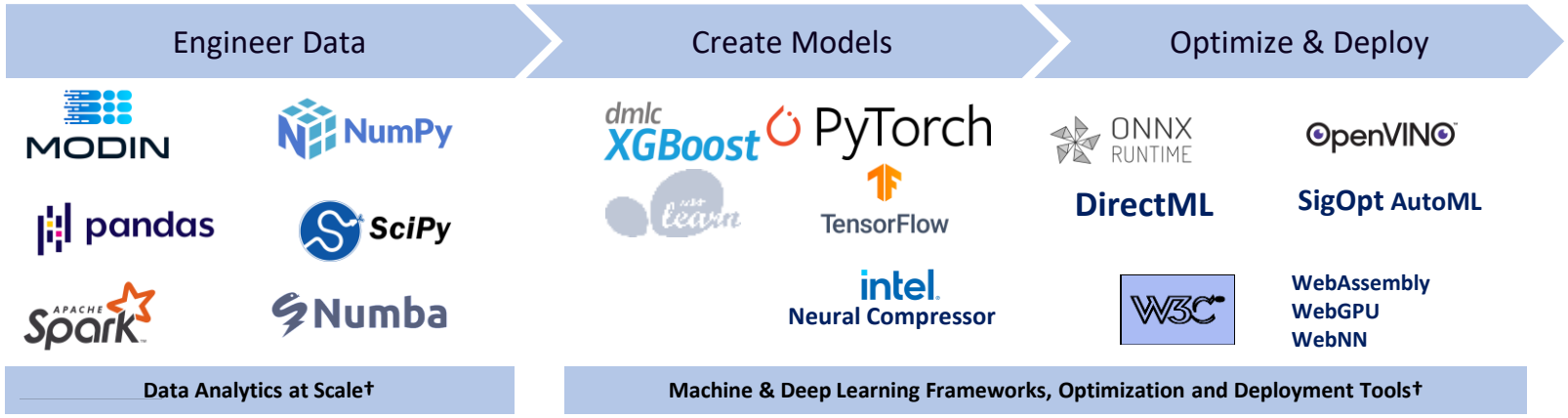
. . .

MegaPOD Cluster

Sized as needed

Training Parameters	N/A	~ 1B + E2E data Pipeline	~ 20B	10Bs Shared across models of many sizes	. . .	Large Scale Distributed Training 10Bs - 1T+ Largest foundational models
Fine-tuning Parameters	N/A	<~ 10B 1 - 8+ nodes	~ 70B			
Inference	~20 TOPS	< 20B **	~ 350B	10B-100Bs Shared across models of many sizes	. . .	Distributed Inference Shared across models of many sizes

Intel AI software portfolio



Accelerate end-to-end data science and AI

Intel® Developer Cloud and Intel® Developer Catalog

Try the latest Intel tools and hardware, and access optimized AI Models

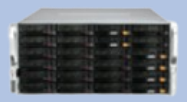
Full stack ML operating system

Intel® Geti™

Annotation/training/ optimization platform

Intel optimizations and fine-tuning recipes, optimized inference models, and model serving

VANTAGEO AI Appliances Portfolio

Model	Configuration	GPU Qty	Best Fit for
 2530-WS	1. Dual 3nd Generation Intel® Xeon® Scalable processors 2. 18 DDR4 ECC DIMM slots 3. PCIe - 4 PCIe 4.0 x16, 2 PCIe 4.0 x8	1 x RTX	AI data preparation, model design, preliminary model training and education . Computer Vision AI at edge
 2230-RS-GPU	1. Dual 3nd Generation Intel® Xeon® Scalable processors 2. 16 DDR4 ECC DIMM slots 3. 2x PCI-E Gen 4x8 / 4x PCI-E Gen4x16 4. NVMeSupport	2 x A100 / L40/L4	AI Training & AI Inference Computer Vision & Intelligent video Analytics Lost Prevention Smart Speech
 2240-RG-GPU	1. Dual 4th Generation Intel® Xeon® Scalable processors 2. 24 DDR5 ECC DIMM slots 3. 4 x FHFL PCIe Gen5 x16 or 8 x FHFL PCIe Gen5 x8 slots for GPUs 4. NVMeSupport 5. 2 x LP PCIe Gen5 x16 slots for add-on cards	4 x A100/ L40	AI Training & AI Inference NLP Speech AI Fraud Deduction Predictive Maintenance
 2540-RG-GPU	1. Dual 4th Generation Intel® Xeon® Scalable processors 2. 32 DDR5 ECC DIMM slots 3. 4 x FHFL PCIe Gen5 x16 or 8 x FHFL PCIe Gen5 x16 slots for GPUs 4. NVMe Support 5. 2 x LP PCIe Gen5 x16 slots for add-on cards	8 x H100	Large AI Training Modules & AI Inference ,NLP ,Speech AI, Fraud Deduction & Predictive Maintenance
 2430-RS-60-ST	1. Dual 3nd Generation Intel® Xeon® Scalable processors 2. 16 DDR4 ECC DIMM slots 3. 2x PCI-E Gen 4x8 / 4x PCI-E Gen4x16 4. NVMeSupport	-	Processing large data sets and Storing unstructured data

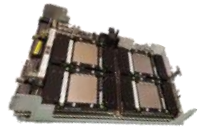
VANTAGEO Tested GPU Portfolio

Brand	GPU Model	GPU Memory	Max Power Consumption	Form-factor	2-way Bridge	Recommended Workloads
PCIe Adapter & SXM form-factor						
Nvidia	A2	16 GB GDDR6	60W	SW, HHHL or FHHL	n/a	AI Inferencing, Edge, VDI
Nvidia	L4	24 GB GDDR6	72W	SW, HHHL or FHHL	n/a	AI Inferencing, Edge, VDI
Nvidia	A16	64 GB GDDR6	250W	DW, FHFL	n/a	VDI
Nvidia	A40, L40	48 GB GDDR6	300W	DW, FHFL	Y, N	Performance graphics, Multi-workload
Nvidia	A30	24 GB HBM2	165W	DW, FHFL	Y	AI Inferencing, AI Training
Nvidia	A100	80 GB HBM2e	300W	DW, FHFL	Y	AI Training, HPC, AI Inferencing
Nvidia	H100	80 GB HBM2e	310 - 350W	DW, FHFL	Y	AI Training, HPC, AI Inferencing
AMD	MI210	64 GB HBM2e	300W	DW, FHFL	Y	HPC, AI Training
Intel	Flex 140	12 GB GDDR6	75W	SW, HHHL or FHHL	n/a	AI Inferencing, Edge, VDI
Intel	Max 1100*	48 GB HBM2e	300W	DW, FHFL	Y	HPC, AI Training
SXM / OAM form-factor						
Nvidia	HGX A100	80 GB HBM2	500W	SXM w/ NVLink	n/a	AI Training, HPC
Nvidia	HGX H100	80 GB or 94GB HBM3	700W	SXM w/ NVLink	n/a	AI Training, HPC
Intel	Max 1550	128 GB HBM2e	600W	OAM w/ XeLink	n/a	AI Training, HPC

PCIe with 2-way Bridge

PCIe Adapter

4-way SXM / OAM Baseboard



8-way SXM / OAM Baseboard

VANTAGEO Storage Solutions

BLOCK STORAGE

Efficient, fast and highly available block storage volumes designed for applications that benefit from fine tuning for performance, cost and capacity

FILE STORAGE

A simple, scalable, elastic file system for Linux-based workloads built to scale on demand, growing and shrinking automatically without disrupting applications

OBJECT STORAGE

Industry-leading S3-compatible cloud storage with unlimited scalability, data availability, security and easy-to-use management capabilities

Software-Defined Storage on Infrastructure Level

Efficient TCO

Achieve total cost of ownership up to 50% less than traditional converged infrastructure systems.

One Platform for Multiple Workloads

Build any service in a private or public cloud with S3 object storage, hot or cold block storage, or file storage. Deployable in hours that ensures quick time to market

Scalability and Flexibility

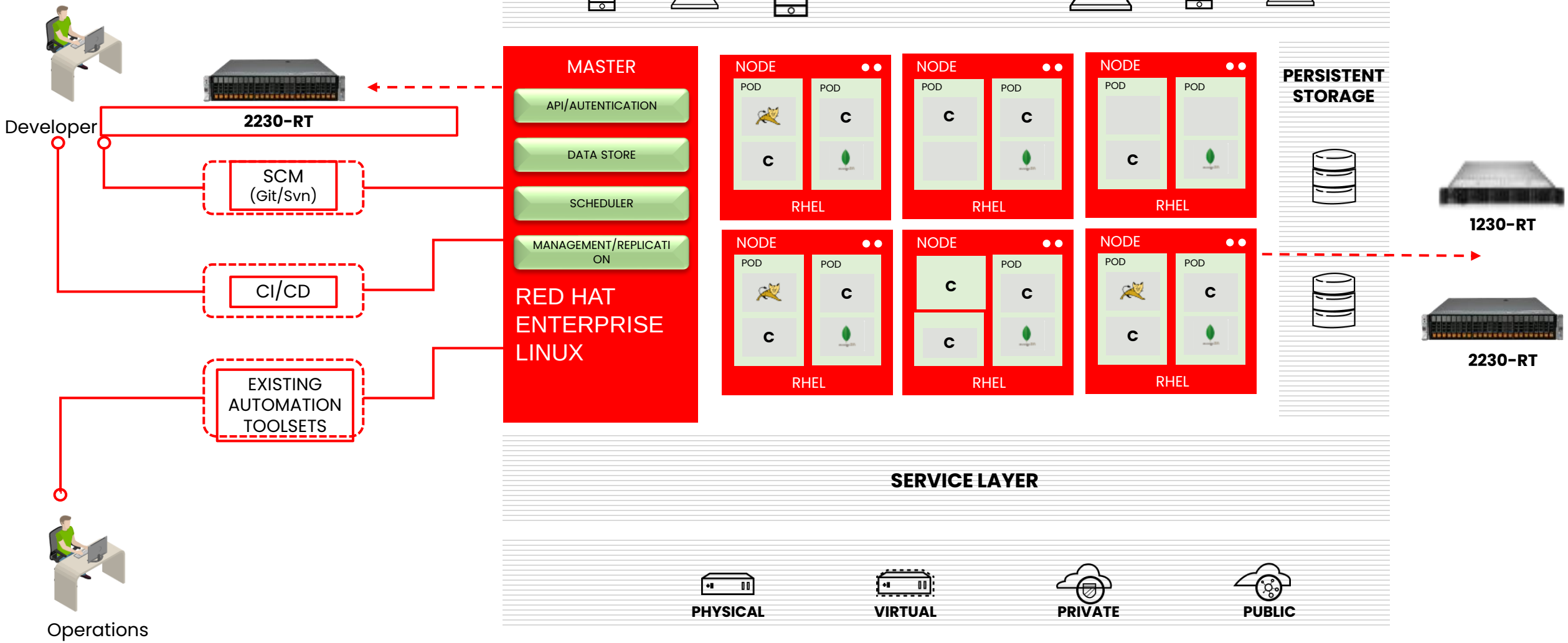
Grow your cluster in small steps – horizontally or vertically – without the need to buy a new chassis to increase your compute or storage capacity.

High Performance and Availability

- Run any applications or services in parallel on multiple nodes
- No single point of failure, ensuring data is always available
- Flexible redundancy schemes to fit both high performance and low redundancy overhead requirements
- Prevent data loss from any server or disk with clustered architecture by design

RedHat OpenShift Certified Appliances

vantageo™



VANTAGEO Success Story

Super Computer PARAM VIKRAM 1000

ISRO's one of the flagship research institute PRL (Physical research Labarotary). Vantageo systems performance benchmark nearly 1000 Terafops of compute & GPU workload of 7.6k+ CPU cores & 150k+ GPU cores.

https://www.prl.res.in/prl-eng/paramvikram1000/architecture

Vantageo is Ranked 14 is India top 100 Supercomputer List published by CDAC (Centre for Development of Advanced Computing (C-DAC) is the premier R&D organization of the Ministry of Electronics and Information Technology (MeitY)

https://topsc.cdacb.in/filterdetailstry?page=20&slug=July2023

VANTAGEO Support SLA Offered – 220 Locations in India

24 x7, 4 Hours CTR Covered Cities	
Mumbai / Thane / Navi-Mumbai	Delhi /Gurgaon / Noida
Bangalore	Jaipur
Hyderabad	Pune
Chennai	Ahmedabad
Kolkata	Cochin

Support powered by TECHSYS™

SUPPORT PARTICULARS	Standard	SILVER	GOLD	PLATINUM
Helpdesk Access	9x5	24x7	24x7	24x7
Onsite Support	Yes	Yes	Yes	Yes
Response Time	24 Hrs	8 Hrs	4 Hrs	4 Hrs
Part Shipping / Supply	Yes	Yes	Yes	Yes
Part Replacement Activity	Yes	Yes	Yes	Yes
Recovery Time SLA	Best Efforts	NBD	8 hrs	4 Hrs
BIOS Update support	No	Yes	Yes	Yes
Firmware Update Support	No	Yes	Yes	Yes
Integration Support - L3	No	No	Yes	Yes
Development Support - L4	No	No	Yes	Yes

THANK YOU!

Phone Numbers
Sales | (+91) 98339 86727

Support 1800 266 9898 (Toll Free)

Email ids
Sales | sales@vantageo.com
Service | support@vanatgeo.com