Going Beyond Accomplishing Your Dreams
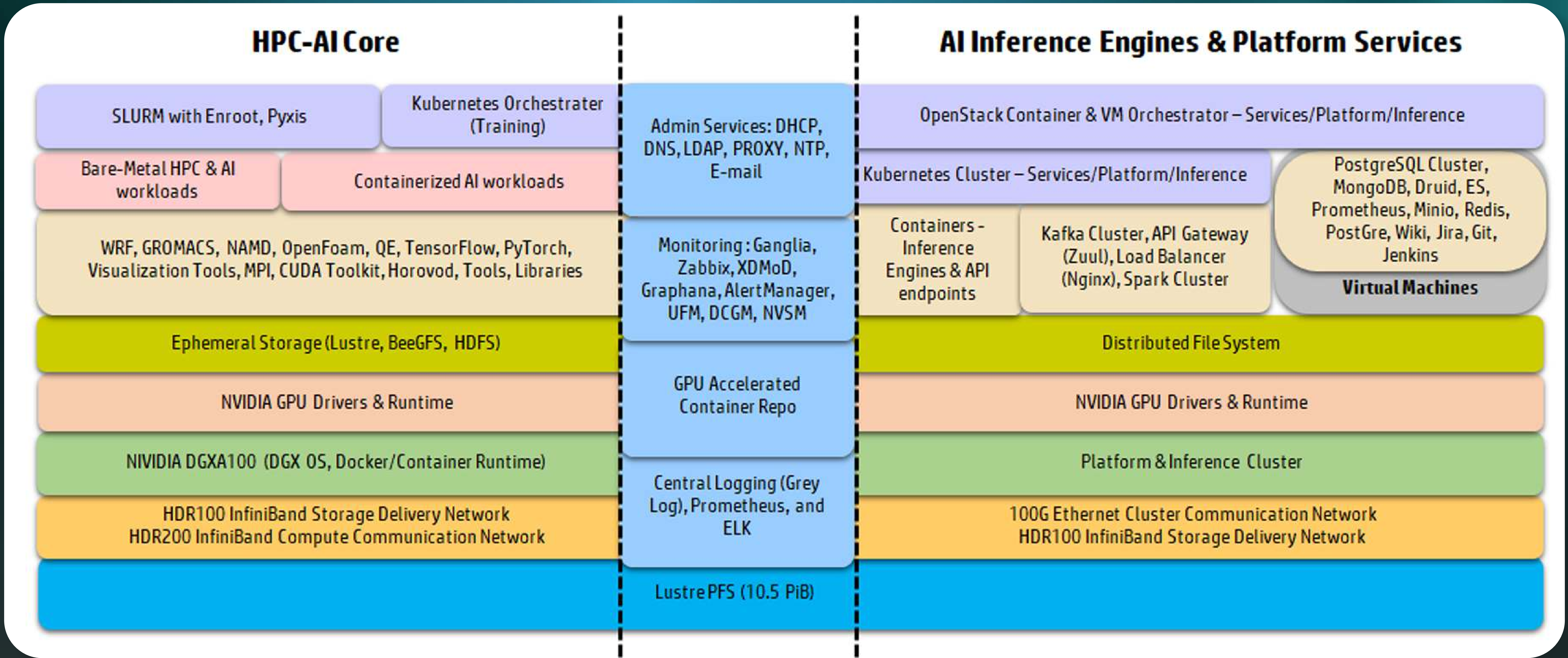
PARAM Siddhi AI software stack diagram

HPC-AI Core			AI Inference Engines & Platform Services		
SLURM with Enroot, Pyxis	Kubernetes Orchestrater (Training)	Admin Services: DHCP, DNS, LDAP, PROXY, NTP, E-mail	OpenStack Container & VM Orchestrator – Services/Platform/Inference		
Bare-Metal HPC & AI workloads	Containerized AI workloads		Kubernetes Cluster – Services/Platform/Inference		PostgreSQL Cluster, MongoDB, Druid, ES, Prometheus, Minio, Redis, PostGre, Wiki, Jira, Git, Jenkins
WRF, GROMACS, NAMD, OpenFoam, QE, TensorFlow, PyTorch, Visualization Tools, MPI, CUDA Toolkit, Horovod, Tools, Libraries		Monitoring : Ganglia, Zabbix, XDMoD, Graphana, AlertManager, UFM, DCGM, NVSM	Containers - Inference Engines & API endpoints	Kafka Cluster, API Gateway (Zuul), Load Balancer (Nginx), Spark Cluster	**Virtual Machines**
Ephemeral Storage (Lustre, BeeGFS, HDFS)			Distributed File System		
NVIDIA GPU Drivers & Runtime		GPU Accelerated Container Repo	NVIDIA GPU Drivers & Runtime		
NIVIDIA DGXA100 (DGX OS, Docker/Container Runtime)			Platform & Inference Cluster		
HDR100 InfiniBand Storage Delivery Network HDR200 InfiniBand Compute Communication Network		Central Logging (Grey Log), Prometheus, and ELK	100G Ethernet Cluster Communication Network HDR100 InfiniBand Storage Delivery Network		
Lustre PFS (10.5 PiB)					

Technical Specifications

- Access to world class resources: ML/DL Software Stack along with inferencing development kit, frameworks, libraries, cloud management software, containerized applications
- Resource partitioning and dynamically provisioned application execution environment
- Provision to support plethora of HPC/AI adaptable workloads and ML/DL frameworks based on user choices
- Energy-saving, high teraflops per watt per server rack space
- Low latency high bandwidth network
- Stream processing capabilities from edge using KAFKA and SPARK working in unison with batch queuing and scheduling orchestrator
- Resource consumption and service delivery over NKN with failover leased line for business continuity
- Multi-tenant highly secured and isolated data protection environment
- Service delivery models analogous to cloud computing paradigm

Bengaluru | Chennai | Hyderabad | Kolkata | Mohali | Mumbai
New Delhi | Noida | North East (Silchar) | Patna | Pune (HO) | Thiruvananthapuram

बेंगलुरू | चेन्नई | हैदराबाद | कोलकाता | मोहाली | मुंबई
नई दिल्ली | नॉएडा | नॉर्थ ईस्ट (सिलचर) | पटना | पुणे (मुख्यालय) | तिरुवनंतपुरम