Empowering AI Transformation with AiHPC

Welcome to the AiHPC Innovation, a transformative solution designed to revolutionize the way enterprises tackle the challenges of large-scale data processing, analysis, and workflow management. Our cutting-edge platform and services can help you unlock new possibilities in your business endeavors.

AiHPC
ENABLER OF FUTURE INNO

Executive Summary

Comprehensive GPU Platform for Secure, Governed Enterprise LLM Deployments

Infrastructure Management

- Centralized GPU control with container-level allocation
- On demand HPC application deployment

AI Capabilities

- Supporting models from LLM, NLP to OCR
- Fine-tuning and optimization pipelines

Enterprise Framework

- Authentication integration
- Role-based access controls
- Resources monitoring

√ Maximize GPU Utilization √ Deploy Multiple LLM Instances
√ Ensure Data Governance √ Scale with Demand

Navigating the AI Revolution in Enterprise Environment

Potential we have now and the challenges for the industries

The trend of AI cannot be reversed - Enterprises Embracing AI for Transformation

$154.9B
To 2030

25.7%
CAGR

The enterprise AI market is growing rapidly.

67%
professionals regularly use AI tools

Consumers using AI have built a strong foundation for enterprise AI adoption

96%
Reduction in doc processing time[2]

AI-driven automation delivers dramatic substantial cost reductions and efficiency gains.

70.0%
Healthcare providers plan to invest in AI technologies[3]

AI enhances productivity, decision-making, and competitive advantage.

Enterprise AI Adoption Barriers: Critical Requirements Gap

An On-Premises solution is needed:
Critical data and information must be stored in-house, driven by regulatory requirements and data sovereignty needs..

Data Security & Privacy Concerns:
Proper AuthN and AuthZ are critical in enterprise environments, requiring seamless integration with existing security systems.

Scalability and resource allocation:
Enterprises need efficient scalability and resource allocation for multiple AI models within limited infrastructure.

Accountability and monitoring:
Enterprise-level audit mechanisms and performance monitoring are essential in AI governance.

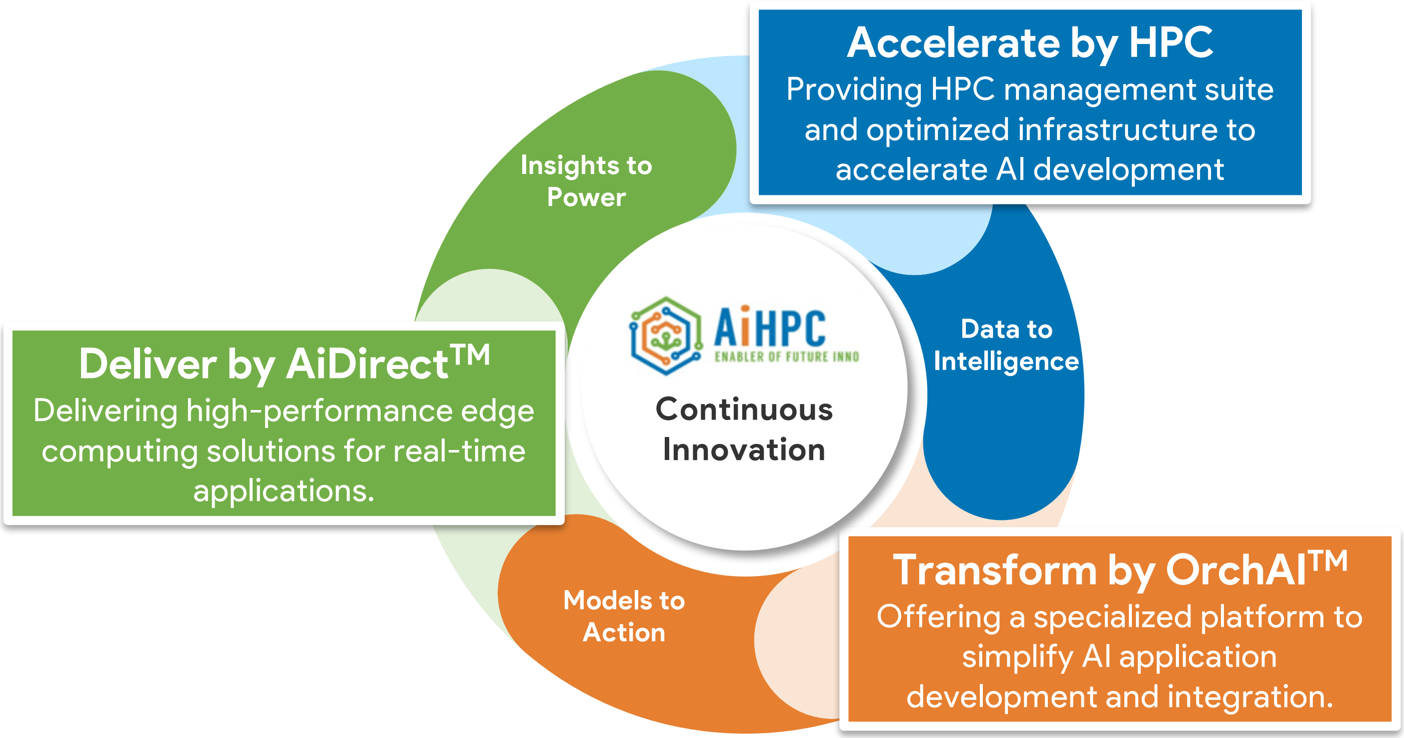
Limited Solutions and talents:
Shortage of mature, turnkey AI solutions for private infrastructure, compounded by the scarcity of specialized AI talent.

How can organizations effectively overcome the technological, regulatory, and resource barriers to adopt and scale AI securely and efficiently within their own on-premises or hybrid environments?

1.

AiHPC Delivers E2E Solutions to Accelerate AI Adoption

Our Integrated Approach Overcomes Barriers and Drives Outcomes



Accelerate by HPC
Providing HPC management suite and optimized infrastructure to accelerate AI development

Deliver by AiDirect^{TM}
Delivering high-performance edge computing solutions for real-time applications.

Transform by OrchAI^{TM}
Offering a specialized platform to simplify AI application development and integration.

End-to-End Service: Streamlined process from scratch to deployment, reducing complexity on GTM

Optimization infrastructure: Tailored solutions that meet the specific needs of each organization.

Cost Efficiency: Optimizing resource utilization and offering affordable solutions compared to competitors

Compliance and Security: Built-in frameworks for regulatory compliance and robust data security measures.

Our E2E Solution: Lowers implementation costs, increases R&D efficiency, overcomes technical barriers, and addresses the AI talent shortage.

End-to-end AI Infrastructure Solution: Simplifying Complex HPC for Enterprise AI

Supercharge Data Processing Capabilities with Our HPC Solutions

AiHPC Portal

Managing AI workloads is complex → Simplified management

- Enterprise-Grade Security
- Interactive Job Control
- Dynamic Resource Allocation
- Smart Usage Analytics

AiHPC Meter

Performance optimization is challenging → Automated optimization

- Framework-specific metrics
- Multi-aspect monitoring
- Real-time performance insights
- Optimization recommendations

AiHPC Service

Support needs are diverse → Complete support

- Infrastructure provisioning
- Framework deployment
- Technical consultation
- Ongoing maintenance & support

Powered by:

Selected stories:

- Deployment of AiHPC Portal at **100+ nodes** HPC environment for interactive jobs and billing system.
- Designed and implemented HPC architecture for **Genomic Research**
- Provided AiHPC Portal for enhanced resource and workflow management.
- Designed and developed **AI software-orientated GPU card performance evaluation** under the liquid cooling environment with AiHPC Meter suite.

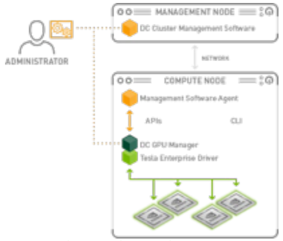
We empower AI availability with highly flexible, cost-efficient, user-friendly enterprise HPC solutions.

Technology: Enterprise-Grade Infrastructure Management

What We Believe is Maximizing Hardware Investment with Granular Control

On-Demand Resource Allocation

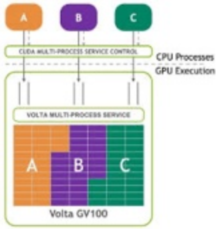
Dynamic allocation of GPU resources based on workload demands



Efficient containerization enables a 300% improvement in GPU utilization rates[1]

Fine-grained GPU Partitioning

Centralized administration with API integration


A
B
C
CUDA MULTI-PROCESS SERVICE CONTROL
CPU Processes
GPU Execution
VOLTA MULTI-PROCESS SERVICE
A
B
C
Volta GV100

Dynamic GPU sharing shows 4.5x improvement in specific research environments.[2]

Enterprise Integration

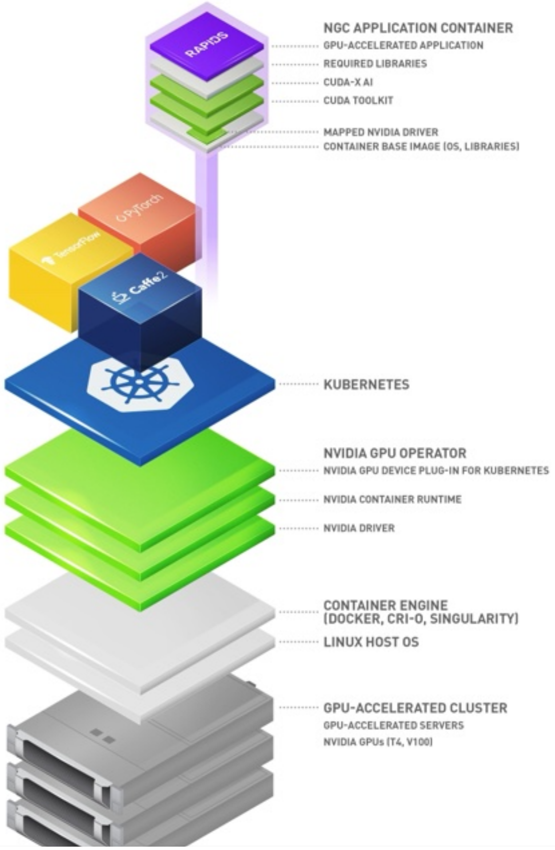
Role-based access control and seamless Integration with existing identity management systems

- Supports LDAP, OAuth, and other enterprise authentication standards
- Automated user provisioning and de-provisioning workflows

Monitoring and Tracking

Comprehensive usage analytics and reporting Cost allocation and chargeback capabilities

- Real-time resource utilization visualizations and historical performan trending, and predictive analytics

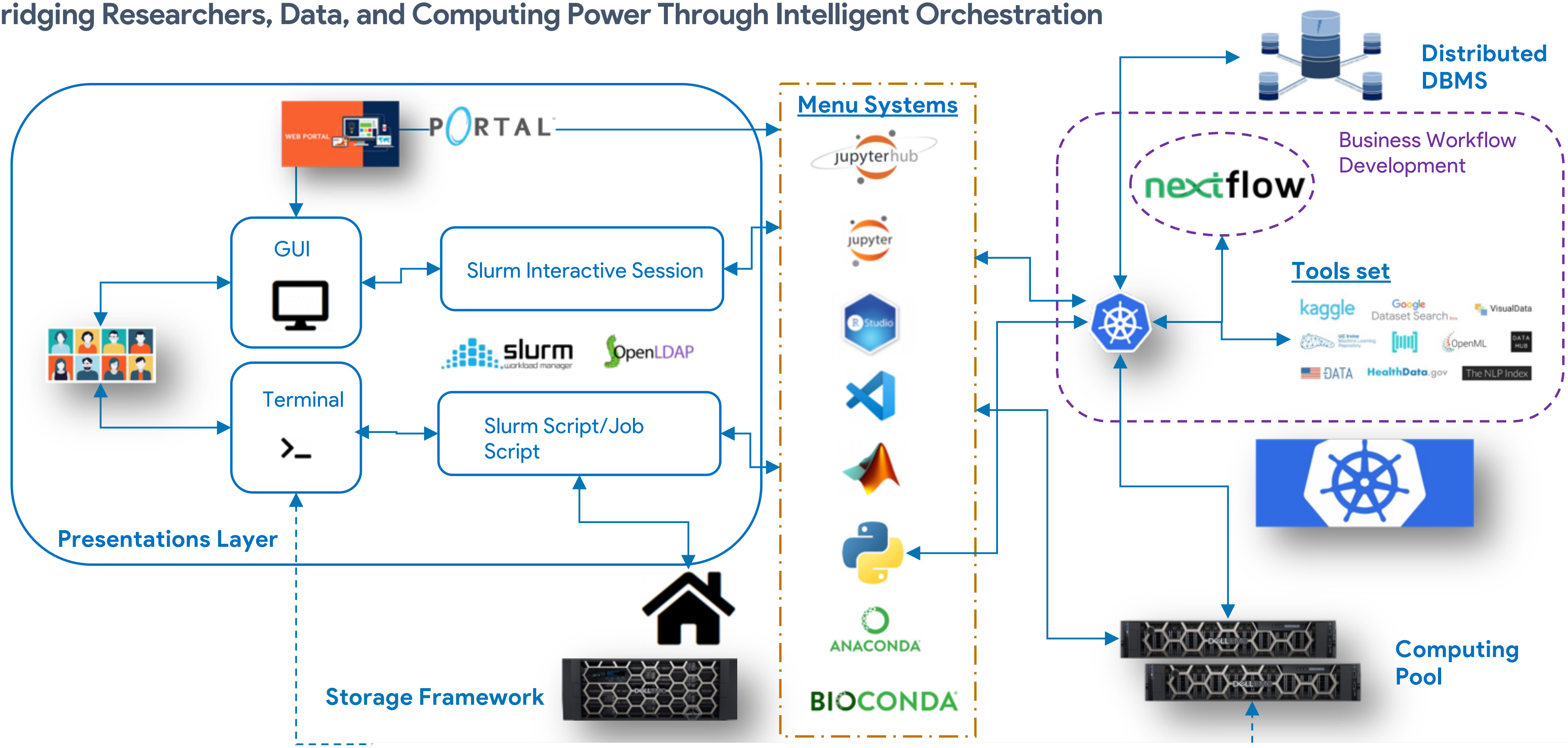


A well-architected HPC infrastructure is the foundation determining the success of the entire AI journey.

[1] NVIDIA. (2023). Maximizing GPU utilization for AI workloads with MIG technology.
[2] Zhang, Y., et al. (2021). Dynamic GPU sharing for deep learning in Kubernetes.

AiHPC Portal to Manage AI Data Flow and Computing for Research Institute

Bridging Researchers, Data, and Computing Power Through Intelligent Orchestration



PORTAL

GUI

Terminal

Slurm Interactive Session

Slurm Script/Job Script

Presentations Layer

Storage Framework

Menu Systems

Distributed DBMS

Business Workflow Development

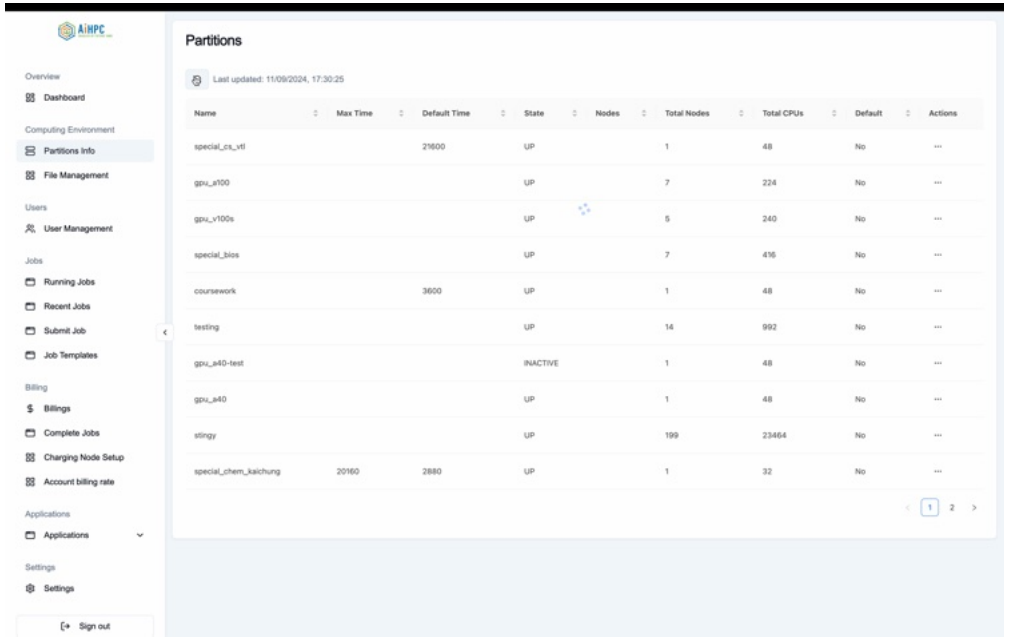
Tools set

Computing Pool

By unifying access, computation, and storage, the AiHPC Portal creates the one-stop solution optimizing institutional infrastructure investments.

AiHPC Portal in Action: Streamlining Job Management and Workflow Management

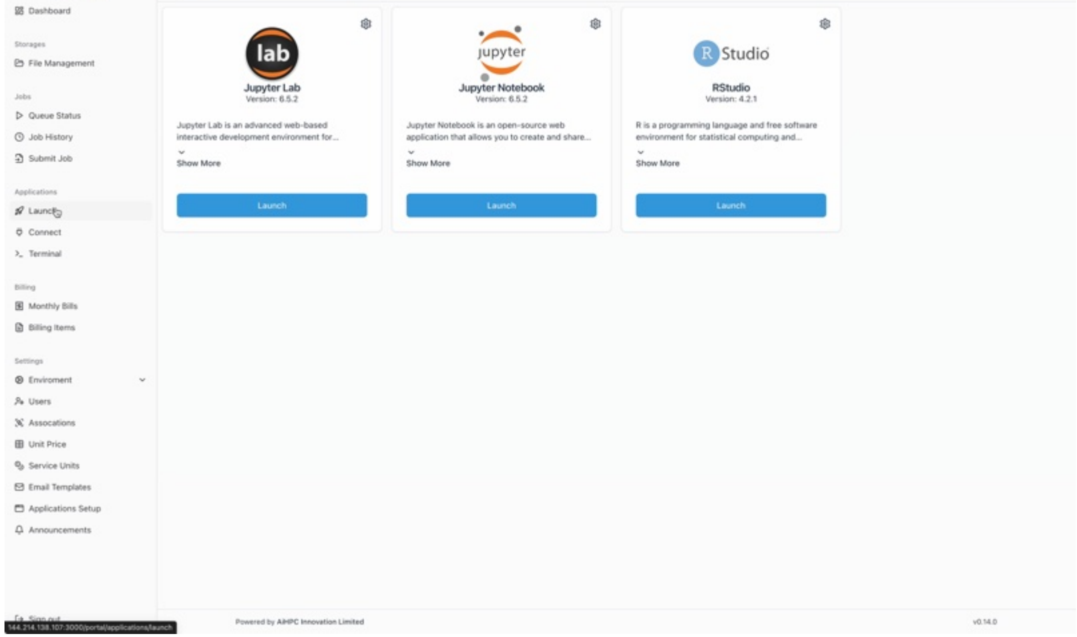
Demonstration of the HPC management with a GUI platform



Partitions

Last updated: 11/09/2024, 17:30:25

Name	Max Time	Default Time	State	Nodes	Total Nodes	Total CPUs	Default	Actions
special_cs_vtl		21600	UP		1	48	No	···
gpu_a100			UP		7	224	No	···
gpu_v100s			UP		5	240	No	···
special_bios			UP		7	416	No	···
coursework		3600	UP		1	48	No	···
testing			UP		14	992	No	···
gpu_a40-test			INACTIVE		1	48	No	···
gpu_a40			UP		1	48	No	···
stingy			UP		199	23464	No	···
special_chem_kaichung	20160	2880	UP		1	32	No	···



Jupyter Lab
Version: 6.5.2

Jupyter Lab is an advanced web-based interactive development environment for...

Show More

Launch

Jupyter Notebook
Version: 6.5.2

Jupyter Notebook is an open-source web application that allows you to create and share...

Show More

Launch

RStudio
Version: 4.2.1

R is a programming language and free software environment for statistical computing and...

Show More

Launch

Powered by AiHPC Innovation Limited

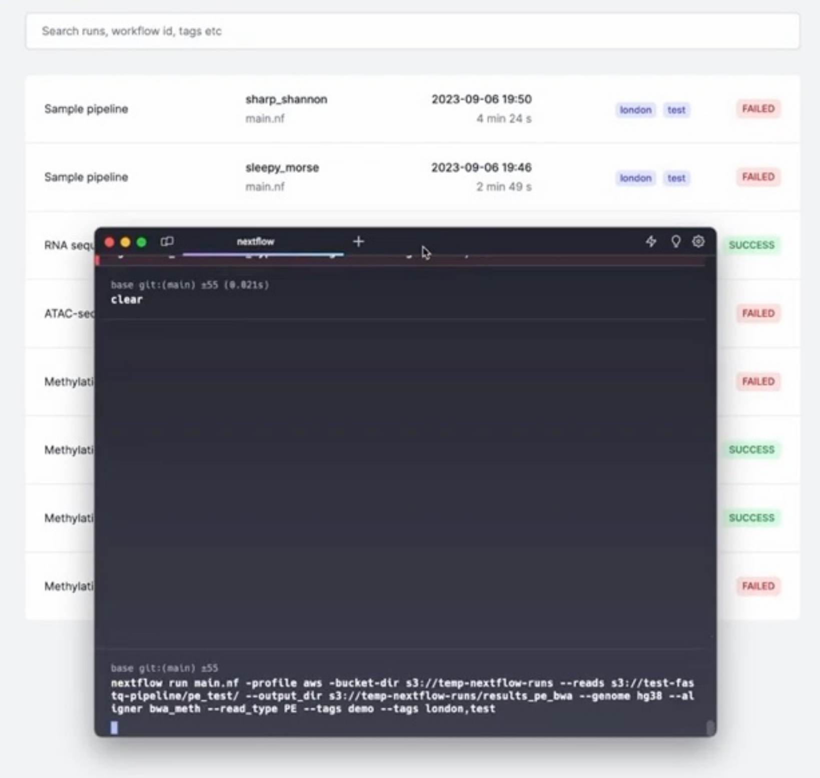
v0.14.0

Demonstrates comprehensive HPC job monitoring with real-time tracking of computing resources, partition utilization, and historical job retrieval capabilities.

Showcases script-free launching of scientific applications, including Jupyter Lab, Notebook, and RStudio, enabling researchers to instantly access computing resources without technical barriers.

AiHPC Portal in Action: Streamlining Job Management and Workflow Management

Demonstration of the HPC management with a GUI platform



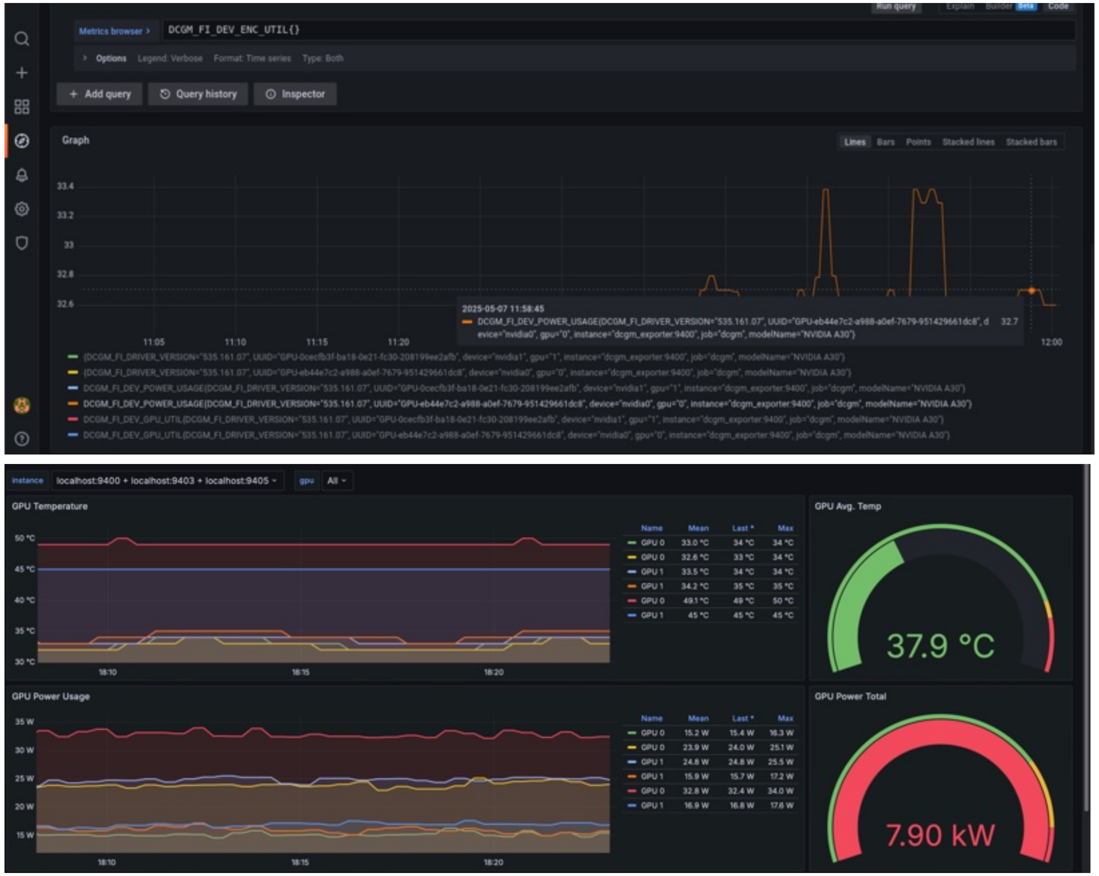
Search runs, workflow id, tags etc

Sample pipeline	sharp_shannon main.nf	2023-09-06 19:50 4 min 24 s	london test	FAILED
Sample pipeline	sleepy_morse main.nf	2023-09-06 19:46 2 min 49 s	london test	FAILED
RNA sequ				SUCCESS
ATAC-sec				FAILED
Methylati				FAILED
Methylati				SUCCESS
Methylati				SUCCESS
Methylati				FAILED

```
base git:(main) ±55 (0.021s)
clear
```

```
base git:(main) ±55
nextflow run main.nf -profile aws -bucket-dir s3://temp-nextflow-runs --reads s3://test-fas
tq-pipeline/pe_test/ --output_dir s3://temp-nextflow-runs/results_pe_bwa --genome hg38 --al
igner bwa_meth --read_type PE --tags demo --tags london,test
```

Example usage: The Portal provided comprehensive workflow monitoring with real-time status tracking, resource utilization metrics, and process-level performance analysis for computational biology applications.



Run query
Explain
Builder Beta
Code
Metrics browser ›
DCGM_FI_DEV_ENC_UTIL{}
› Options Legend: Verbose Format: Time series Type: Both
+ Add query
Query history
Inspector
Graph
Lines
Bars
Points
Stacked lines
Stacked bars
33.4
33.2
33
32.8
32.6
11:05
11:10
11:15
11:20
12:00
2025-05-07 11:58:45
DCGM_FI_DEV_POWER_USAGE{DCGM_FI_DRIVER_VERSION="535.161.07", UUID="GPU-eb44e7c2-a988-a0ef-7679-951429661dc8", device="nvidia0", gpu="0", instance="dcgm_exporter:9400", job="dcgm", modelName="NVIDIA A30"} 32.7
{DCGM_FI_DRIVER_VERSION="535.161.07", UUID="GPU-0cecfb3f-ba18-0e21-fc30-208199ee2afb", device="nvidia1", gpu="1", instance="dcgm_exporter:9400", job="dcgm", modelName="NVIDIA A30"}
{DCGM_FI_DRIVER_VERSION="535.161.07", UUID="GPU-eb44e7c2-a988-a0ef-7679-951429661dc8", device="nvidia0", gpu="0", instance="dcgm_exporter:9400", job="dcgm", modelName="NVIDIA A30"}
DCGM_FI_DEV_POWER_USAGE{DCGM_FI_DRIVER_VERSION="535.161.07", UUID="GPU-0cecfb3f-ba18-0e21-fc30-208199ee2afb", device="nvidia1", gpu="1", instance="dcgm_exporter:9400", job="dcgm", modelName="NVIDIA A30"}
DCGM_FI_DEV_POWER_USAGE{DCGM_FI_DRIVER_VERSION="535.161.07", UUID="GPU-eb44e7c2-a988-a0ef-7679-951429661dc8", device="nvidia0", gpu="0", instance="dcgm_exporter:9400", job="dcgm", modelName="NVIDIA A30"}
DCGM_FI_DEV_GPU_UTIL{DCGM_FI_DRIVER_VERSION="535.161.07", UUID="GPU-0cecfb3f-ba18-0e21-fc30-208199ee2afb", device="nvidia1", gpu="1", instance="dcgm_exporter:9400", job="dcgm", modelName="NVIDIA A30"}
DCGM_FI_DEV_GPU_UTIL{DCGM_FI_DRIVER_VERSION="535.161.07", UUID="GPU-eb44e7c2-a988-a0ef-7679-951429661dc8", device="nvidia0", gpu="0", instance="dcgm_exporter:9400", job="dcgm", modelName="NVIDIA A30"}
instance
localhost:9400 + localhost:9403 + localhost:9405
gpu
All
GPU Temperature
50 °C
45 °C
40 °C
35 °C
30 °C
18:10
18:15
18:20
GPU Avg. Temp
37.9 °C
GPU Power Usage
35 W
30 W
25 W
20 W
15 W
18:10
18:15
18:20
GPU Power Total
7.90 kW

Name	Mean	Last *	Max
GPU 0	33.0 °C	34 °C	34 °C
GPU 0	32.6 °C	33 °C	34 °C
GPU 1	33.5 °C	34 °C	34 °C
GPU 1	34.2 °C	35 °C	35 °C
GPU 0	49.1 °C	49 °C	50 °C
GPU 1	45 °C	45 °C	45 °C

Name	Mean	Last *	Max
GPU 0	15.2 W	15.4 W	16.3 W
GPU 0	23.9 W	24.0 W	25.1 W
GPU 1	24.8 W	24.8 W	25.5 W
GPU 1	15.9 W	15.7 W	17.2 W
GPU 0	32.8 W	32.4 W	34.0 W
GPU 1	16.9 W	16.8 W	17.6 W

Comprehensive hardware monitoring dashboards display real-time GPU performance metrics including temperature, power consumption, and utilization trends for optimizing computational resource efficiency.

Transform AI Integration with OrchAI™ Platform

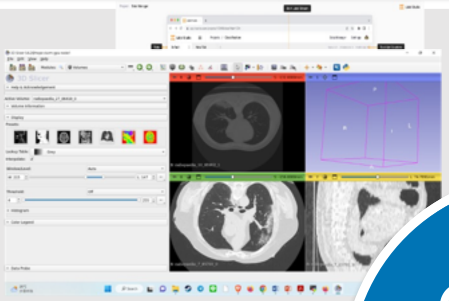
Unified Orchestration for Enterprise AI Operations

Data Integration → Model Selection → Monitoring → Optimization

Quality Local Data Hub

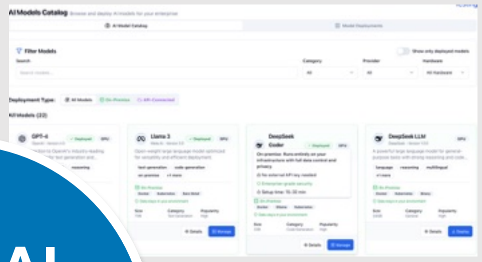
Connect & Transform

- Data Connectors
- ETL Pipeline
- Data Validation
- Real-time Sync
- Security Assurance



One System, multiple AI

Store & Utilize



- Multi-model Registry
- Version Control
- RAG Integration
- Model Marketplace

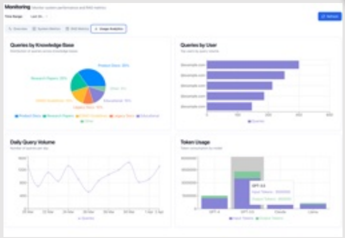
OrchAI

One Stop Enterprise AI Platform

Enterprise Security and Monitoring

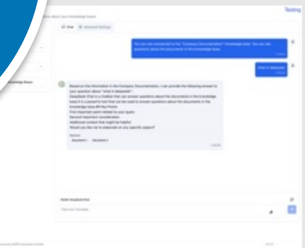
Review & Improve

- Performance Metrics
- Cost Analysis
- Usage Tracking
- Real-time Alerts



Interaction with your AI

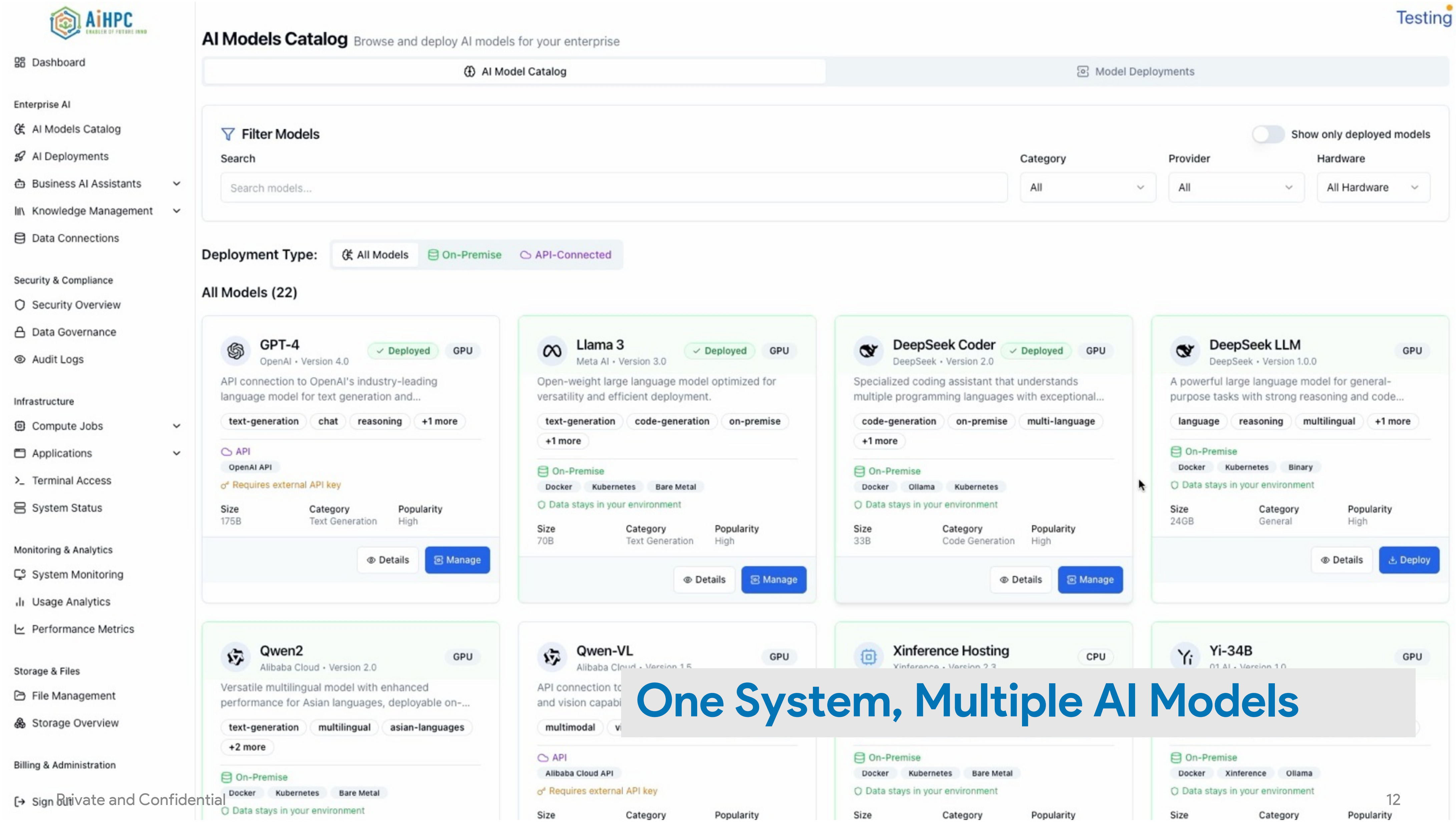
Continues & Reproduce



- Workflow Engine
- Task Scheduling
- Auto-scaling
- Pipeline Management

✓ **Rapid AI Deployment -** From concept to production in days ✓ **Data Security:** Your data stays within your control
✓ **Scalable Architecture:** Grows with your enterprise needs ✓ **Full Transparency:** Track every AI decision and interaction

One System, Multiple AI Models



AiHPC

Testing

- Dashboard

Enterprise AI

- AI Models Catalog
- AI Deployments
- Business AI Assistants
- Knowledge Management
- Data Connections

Security & Compliance

- Security Overview
- Data Governance
- Audit Logs

Infrastructure

- Compute Jobs
- Applications
- Terminal Access
- System Status

Monitoring & Analytics

- System Monitoring
- Usage Analytics
- Performance Metrics

Storage & Files

- File Management
- Storage Overview

Billing & Administration

- Sign Out

AI Models Catalog

Browse and deploy AI models for your enterprise

AI Model Catalog | Model Deployments

Filter Models

Show only deployed models

Search	Category	Provider	Hardware
Search models...	All	All	All Hardware

Deployment Type: All Models | On-Premise | API-Connected

All Models (22)

GPT-4 ✓ Deployed GPU
OpenAI • Version 4.0
API connection to OpenAI's industry-leading language model for text generation and...
text-generation | chat | reasoning | +1 more
API
OpenAI API
Requires external API key

Size	Category	Popularity
175B	Text Generation	High

Details | Manage

Llama 3 ✓ Deployed GPU
Meta AI • Version 3.0
Open-weight large language model optimized for versatility and efficient deployment.
text-generation | code-generation | on-premise | +1 more
On-Premise
Docker | Kubernetes | Bare Metal
Data stays in your environment

Size	Category	Popularity
70B	Text Generation	High

Details | Manage

DeepSeek Coder ✓ Deployed GPU
DeepSeek • Version 2.0
Specialized coding assistant that understands multiple programming languages with exceptional...
code-generation | on-premise | multi-language | +1 more
On-Premise
Docker | Ollama | Kubernetes
Data stays in your environment

Size	Category	Popularity
33B	Code Generation	High

Details | Manage

DeepSeek LLM GPU
DeepSeek • Version 1.0.0
A powerful large language model for general-purpose tasks with strong reasoning and code...
language | reasoning | multilingual | +1 more
On-Premise
Docker | Kubernetes | Binary
Data stays in your environment

Size	Category	Popularity
24GB	General	High

Details | Deploy

Qwen2 GPU
Alibaba Cloud • Version 2.0
Versatile multilingual model with enhanced performance for Asian languages, deployable on-...
text-generation | multilingual | asian-languages | +2 more
On-Premise
Docker | Kubernetes | Bare Metal
Data stays in your environment

Qwen-VL GPU
Alibaba Cloud • Version 1.5
API connection tc
and vision capabi
multimodal | vi
API
Alibaba Cloud API
Requires external API key

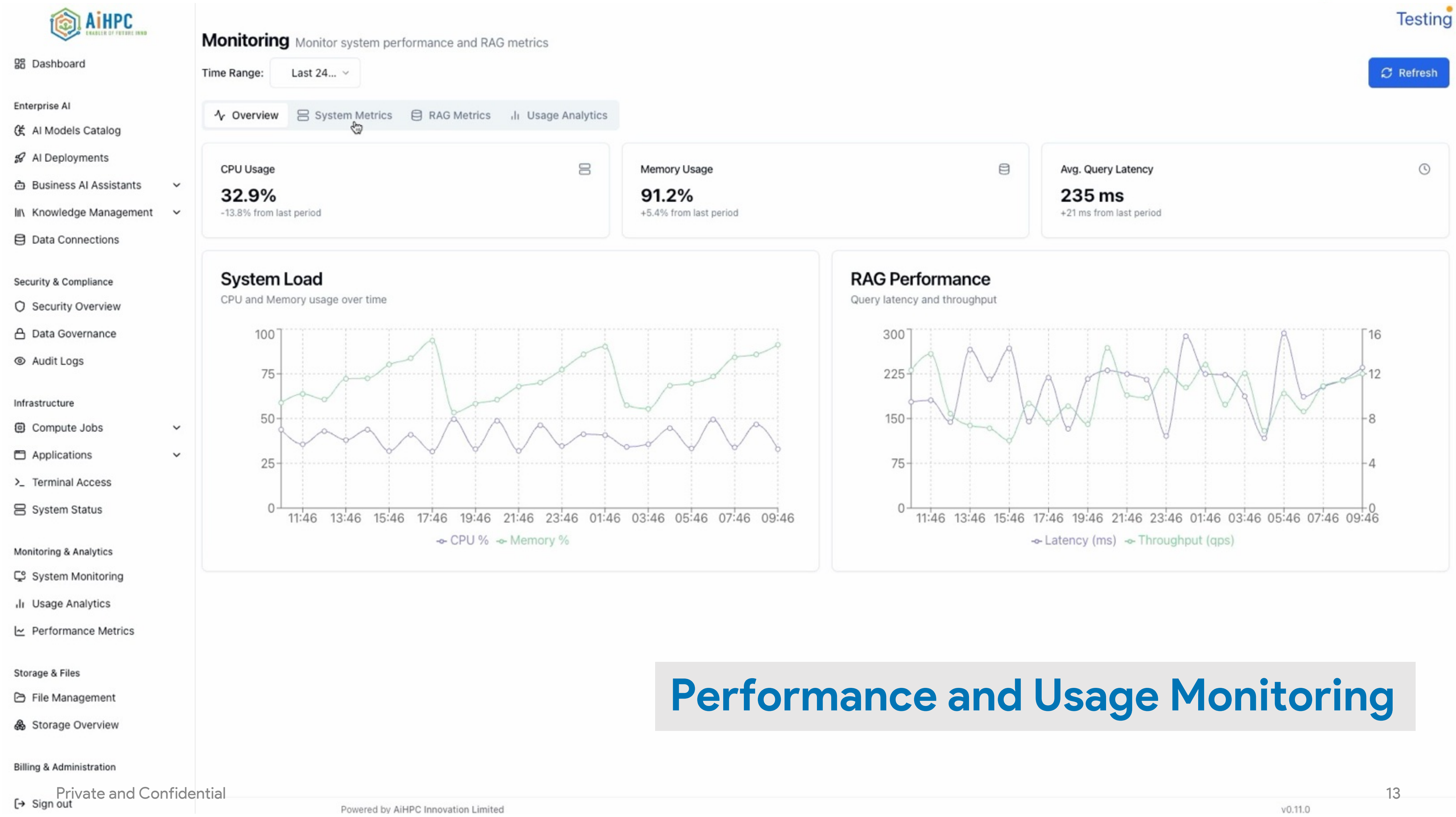
Size	Category	Popularity

Xinference Hosting CPU
Xinference • Version 2.3
On-Premise
Docker | Kubernetes | Bare Metal
Data stays in your environment

Size	Category	Popularity

Yi-34B GPU
01.AI • Version 1.0
On-Premise
Docker | Xinference | Ollama
Data stays in your environment

Size	Category	Popularity



Monitoring Monitor system performance and RAG metrics

Time Range: Last 24...

CPU Usage
32.9%
-13.8% from last period

Memory Usage
91.2%
+5.4% from last period

Avg. Query Latency
235 ms
+21 ms from last period

System Load
CPU and Memory usage over time

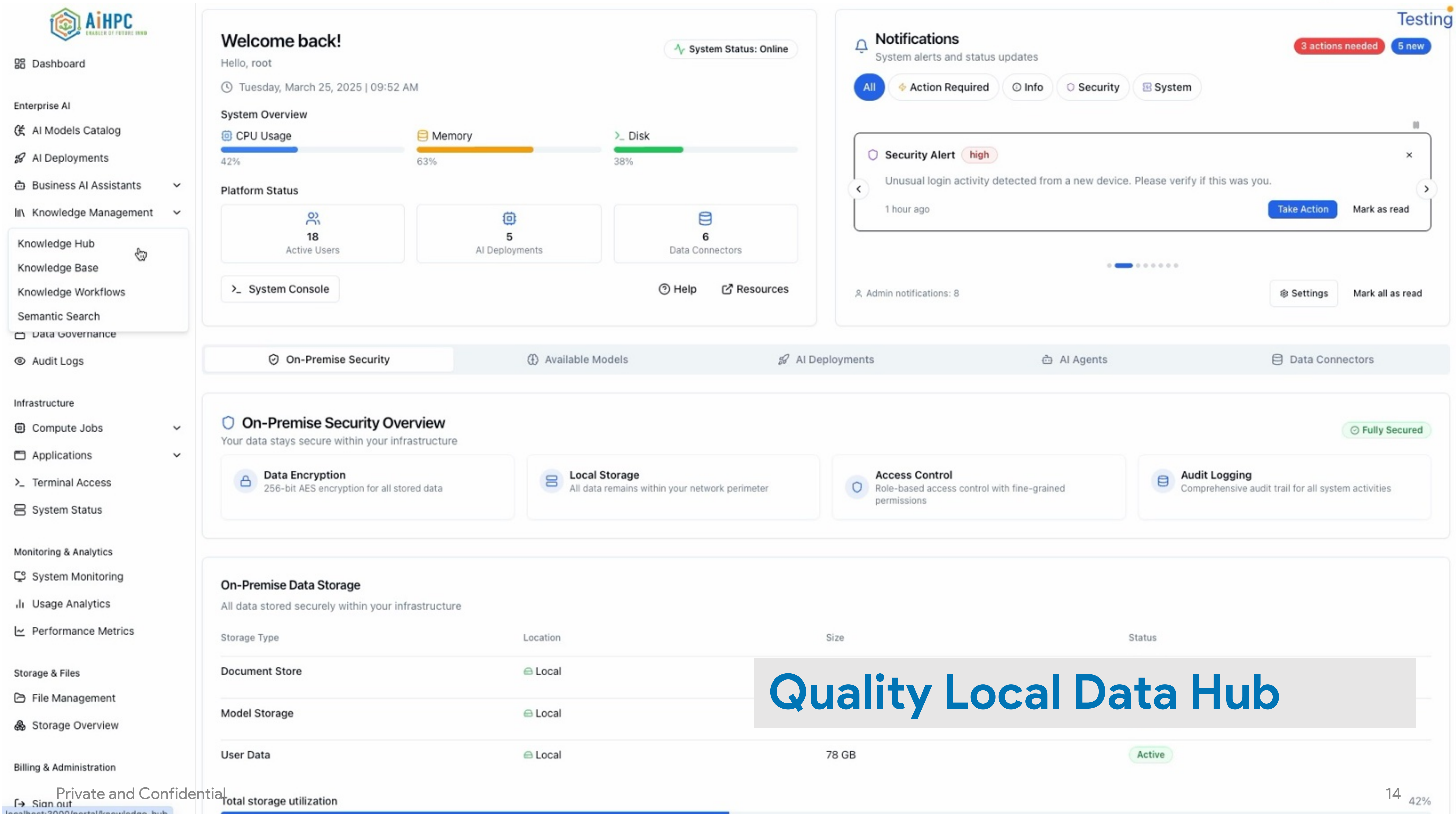
RAG Performance
Query latency and throughput

Powered by AiHPC Innovation Limited

v0.11.0

Performance and Usage Monitoring

Quality Local Data Hub



AiHPC
ENABLER OF FUTURE INNO

Dashboard

Enterprise AI

- AI Models Catalog
- AI Deployments
- Business AI Assistants
- Knowledge Management
  - Knowledge Hub
  - Knowledge Base
  - Knowledge Workflows
  - Semantic Search
- Data Governance
- Audit Logs

Infrastructure

- Compute Jobs
- Applications
- Terminal Access
- System Status

Monitoring & Analytics

- System Monitoring
- Usage Analytics
- Performance Metrics

Storage & Files

- File Management
- Storage Overview

Billing & Administration

- Sign out

localhost:3000/portal/knowledge-hub

Welcome back!

System Status: Online

Hello, root

Tuesday, March 25, 2025 | 09:52 AM

System Overview

CPU Usage	Memory	Disk
42%	63%	38%

Platform Status

18	5	6
Active Users	AI Deployments	Data Connectors

System Console · Help · Resources

Notifications

Testing

System alerts and status updates

3 actions needed · 5 new

All · Action Required · Info · Security · System

Security Alert · high

Unusual login activity detected from a new device. Please verify if this was you.

1 hour ago

Take Action · Mark as read

Admin notifications: 8

Settings · Mark all as read

On-Premise Security · Available Models · AI Deployments · AI Agents · Data Connectors

On-Premise Security Overview

Your data stays secure within your infrastructure

Fully Secured

- **Data Encryption** 256-bit AES encryption for all stored data
- **Local Storage** All data remains within your network perimeter
- **Access Control** Role-based access control with fine-grained permissions
- **Audit Logging** Comprehensive audit trail for all system activities

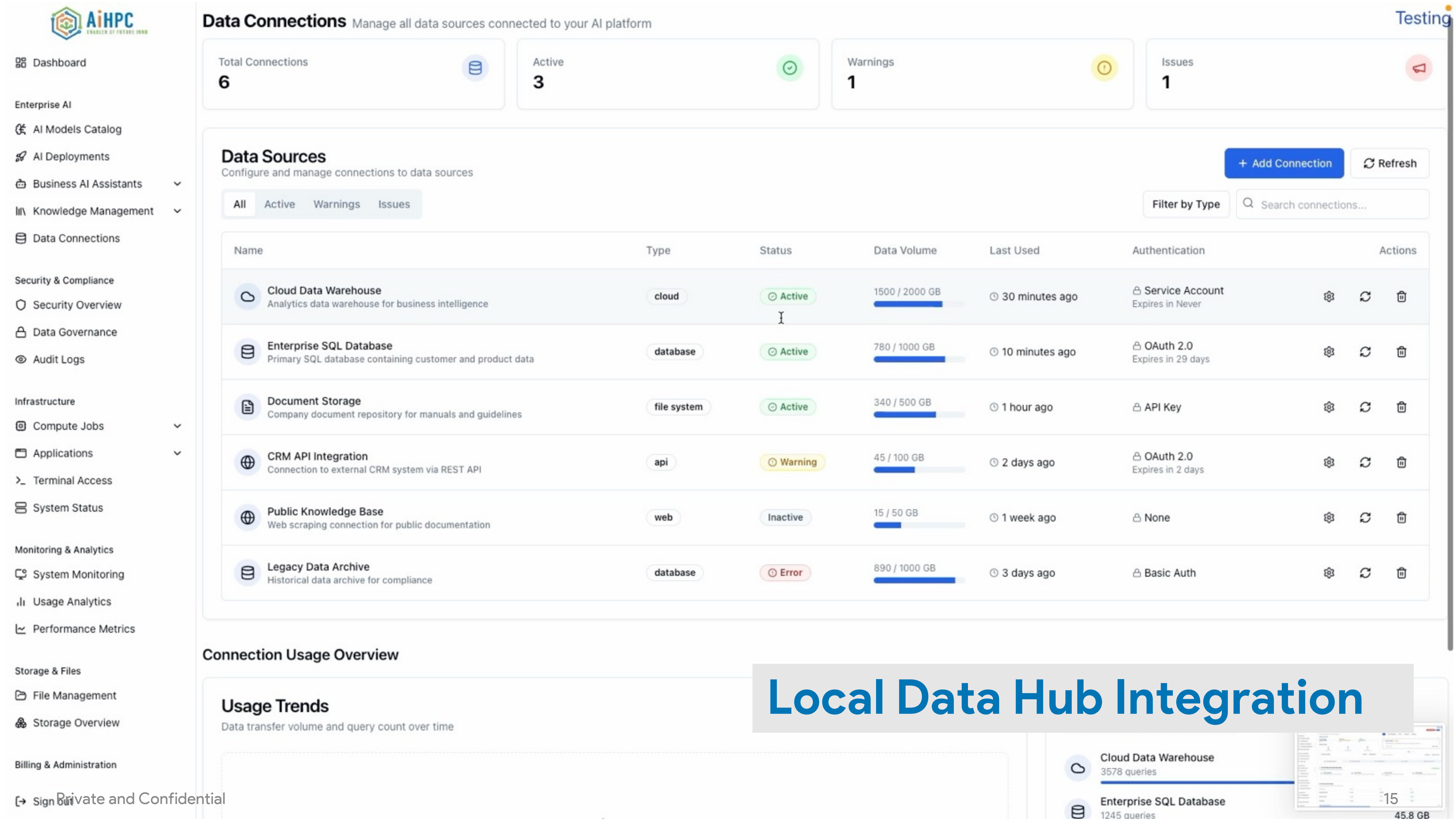
On-Premise Data Storage

All data stored securely within your infrastructure

Storage Type	Location	Size	Status
Document Store	Local		
Model Storage	Local		
User Data	Local	78 GB	Active

Total storage utilization 42%

Local Data Hub Integration



Data Connections

Manage all data sources connected to your AI platform

Total Connections	Active	Warnings	Issues
6	3	1	1

Data Sources

Configure and manage connections to data sources

Name	Type	Status	Data Volume	Last Used	Authentication	Actions
Cloud Data Warehouse Analytics data warehouse for business intelligence	cloud	Active	1500 / 2000 GB	30 minutes ago	Service Account Expires in Never	
Enterprise SQL Database Primary SQL database containing customer and product data	database	Active	780 / 1000 GB	10 minutes ago	OAuth 2.0 Expires in 29 days	
Document Storage Company document repository for manuals and guidelines	file system	Active	340 / 500 GB	1 hour ago	API Key	
CRM API Integration Connection to external CRM system via REST API	api	Warning	45 / 100 GB	2 days ago	OAuth 2.0 Expires in 2 days	
Public Knowledge Base Web scraping connection for public documentation	web	Inactive	15 / 50 GB	1 week ago	None	
Legacy Data Archive Historical data archive for compliance	database	Error	890 / 1000 GB	3 days ago	Basic Auth	

Connection Usage Overview

Usage Trends

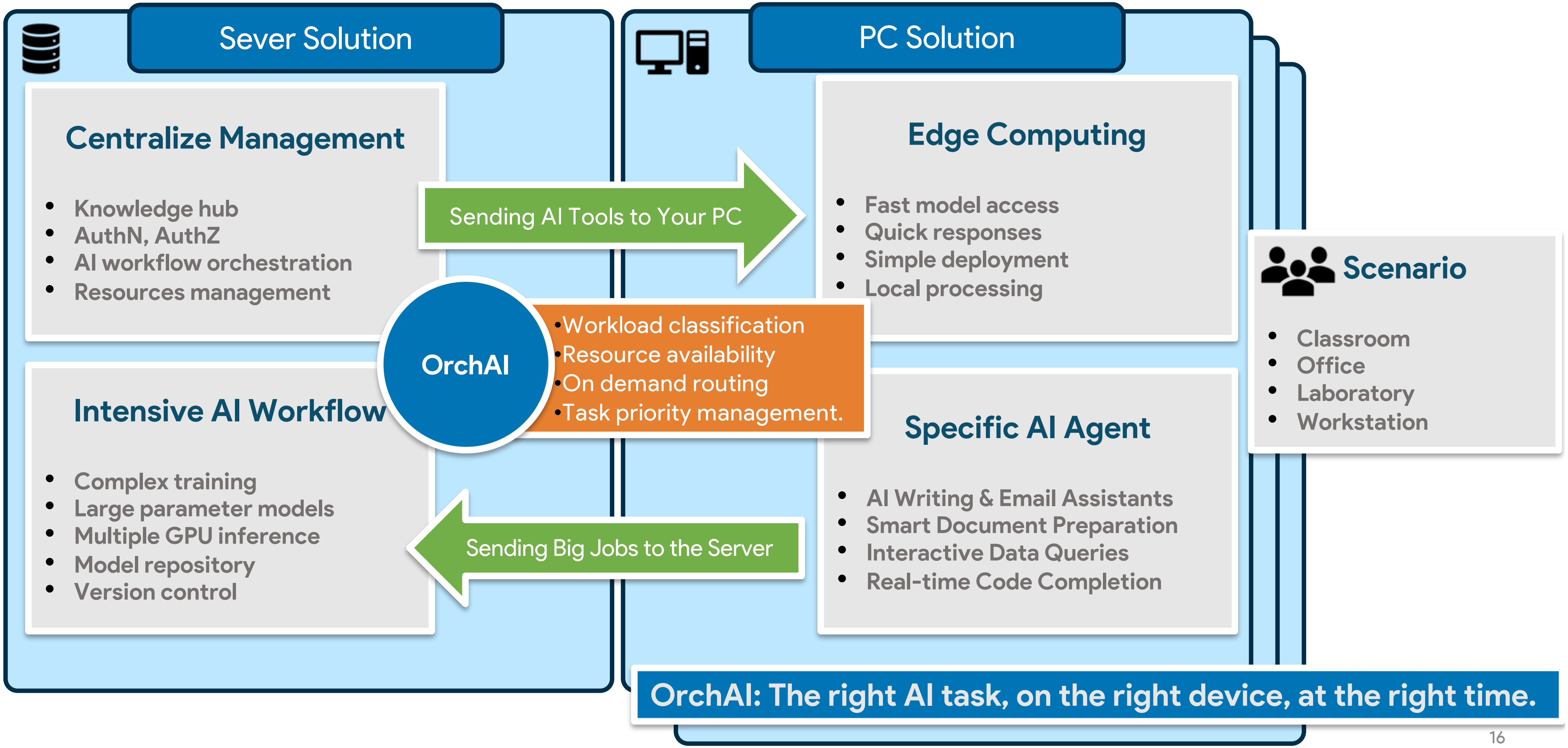
Data transfer volume and query count over time

Cloud Data Warehouse
3578 queries

Enterprise SQL Database
1245 queries
45.8 GB

OrchAI Hybrid Architecture: Intelligent Workload Distribution

Utilizing the resources to gain the greatest user experience in AI adoption



Sever Solution

Centralize Management

- Knowledge hub
- AuthN, AuthZ
- AI workflow orchestration
- Resources management

Intensive AI Workflow

- Complex training
- Large parameter models
- Multiple GPU inference
- Model repository
- Version control

Sending AI Tools to Your PC

Sending Big Jobs to the Server

OrchAI

- Workload classification
- Resource availability
- On demand routing
- Task priority management.

PC Solution

Edge Computing

- Fast model access
- Quick responses
- Simple deployment
- Local processing

Specific AI Agent

- AI Writing & Email Assistants
- Smart Document Preparation
- Interactive Data Queries
- Real-time Code Completion

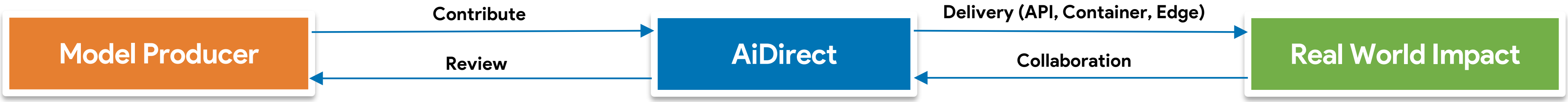
Scenario

- Classroom
- Office
- Laboratory
- Workstation

OrchAI: The right AI task, on the right device, at the right time.

The AiDirect^{TM} – Deliver AI Models to Real-World Applications

The Professional AI Channel for Enterprise Innovation



AIDirect is a **domain specific AI model distribution platform** designed to deliver AI models into real-world applications.

Similar to how **LinkedIn** caters to **professional networking** compared to **Facebook**, **AiDirect** focuses on professional, secure, and compliant AI deployment, whereas platforms like Poe cater to general AI usage.

vs

AiDirect

Professional-Grade AI Model Library:

- Access a **curated selection** of AI models developed for professional applications in different industries.
- Models are vetted for **compliance, performance, and reliability** in operational settings.

Seamless Integration into Real-World Applications:

- Efficient deployment AI models with **AI Agent** available.
- Supports various integration methods, including **APIs, containers, and edge deployment**.

Compliance and Security:

- Data Protection: Comply with relevant industry regulations (e.g., GDPR, CCPA, and others) and data privacy mandates.
- Security Protocols: Implements robust security measures to protect sensitive enterprise data and system integrity.

Collaborative Platform:

Enables collaboration between AI developers, domain experts, and organizations to foster innovation and address real-world business challenges.

Intelligence Decision Support Systems:
Augment expert decision-making with AI-driven insights for improved accuracy and efficiency across various functions..

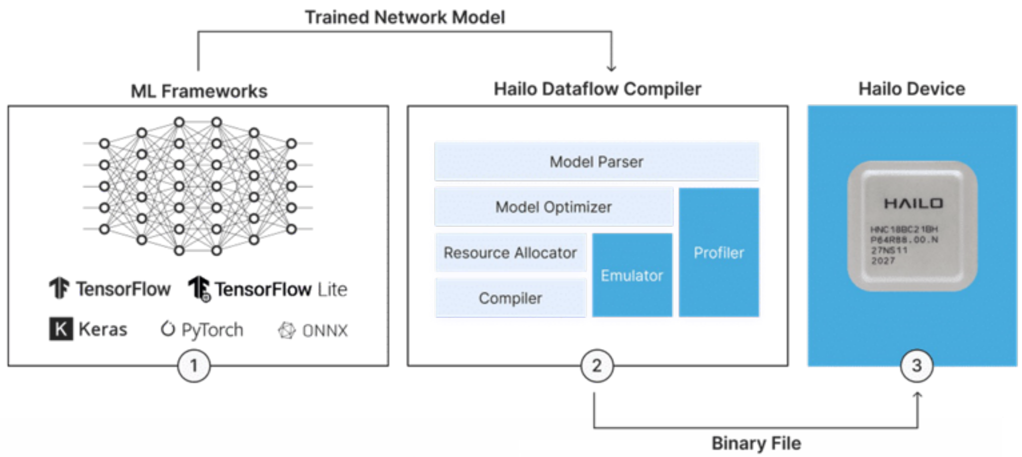
Medical Imaging Analysis:
Enhances imaging technologies, providing faster and more precise interpretations of scans.

Streamlined Processes:
Automation scheduling, billing, and other administrative tasks, increasing efficiency.

Accelerating AI Deployment with Purpose-Built Edge Computing

Bringing AI to the Point of Care with Enterprise-Grade Performance

- **Customized Edge Devices –** e.g. optimized for medical imaging workloads
- **Real-Time Analytics -** 95% faster than traditional cloud processing
- **Secure Deployment –** on-premises and e2e encryption
- **Remote Management -** for enterprise-wide monitoring



Traditional AI Deployment vs. Edge Approach:

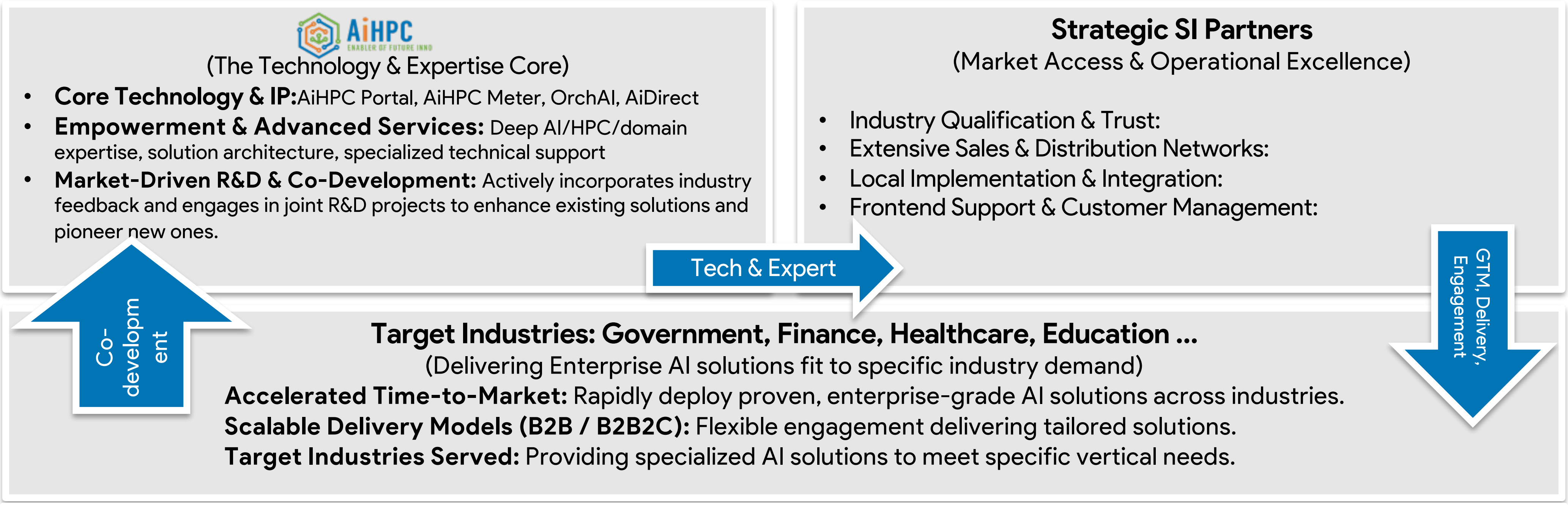
	Traditional	AiDirect
Processing Time	Minutes	Milliseconds
Data Privacy	Network Risk	On-Premise
IT Complexity	High	Simplified
Deployment Cost	Per-Use	Fixed Cost

✓ **Future-Proof Investment:** Scale as AI models evolve ✓ **Reduced Bandwidth Costs:** Process locally, only send results to cloud ✓ **No Vendor Lock-in:** Standard ML framework compatibility ✓ **Simplified Compliance:** Streamlined data governance and auditing

Our Dynamic AI Ecosystem: Driving Innovation & Rapid Industry Adoption

Making advanced AI accessible and rapidly deployable across industries.

HPC Consultancy → AI Implementation → System Integration → Frontend Support Services



AiHPC

(The Technology & Expertise Core)

- **Core Technology & IP:** AiHPC Portal, AiHPC Meter, OrchAI, AiDirect
- **Empowerment & Advanced Services:** Deep AI/HPC/domain expertise, solution architecture, specialized technical support
- **Market-Driven R&D & Co-Development:** Actively incorporates industry feedback and engages in joint R&D projects to enhance existing solutions and pioneer new ones.

Strategic SI Partners

(Market Access & Operational Excellence)

- Industry Qualification & Trust:
- Extensive Sales & Distribution Networks:
- Local Implementation & Integration:
- Frontend Support & Customer Management:

Target Industries: Government, Finance, Healthcare, Education ...

(Delivering Enterprise AI solutions fit to specific industry demand)

Accelerated Time-to-Market: Rapidly deploy proven, enterprise-grade AI solutions across industries.
Scalable Delivery Models (B2B / B2B2C): Flexible engagement delivering tailored solutions.
Target Industries Served: Providing specialized AI solutions to meet specific vertical needs.

AiHPC's Technical Innovation + SI's Market Expertise = Accelerated Customer Success

AiHPC's core technology, amplified by SI partnerships and enriched by industry co-development, delivers continuously evolving AI solutions.

Founding Members

Pioneering Enterprise AI with Deep Technical and Market Expertise

Sam Chu, Ph.D.

Chief Executive Officer, Founder

Sam Chu science and technology entrepreneur, serving as CIO for Sanomics Limited and ACT Genomics, overseeing a clinical trial on early-stage cancer detection using plasma DNA tests. Ensuring efficient data integration, storage, and analysis, they have also played a key role in developing the tele-oncology platform, Aurora Tele-Oncology, enabling remote cancer care during the pandemic.

Sammy Tang

Chief Technology Officer, Co-Founder

Sammy Tang is a highly skilled professional with expertise in mathematical modeling, scientific problem-solving, HPC research, and clinical solution deployment. With a track record of managing engineering support science projects in areas such as data centers, air pollution studies, and ocean modeling, they have made significant contributions to research and development in various fields

Mark Chang, D.Phil.

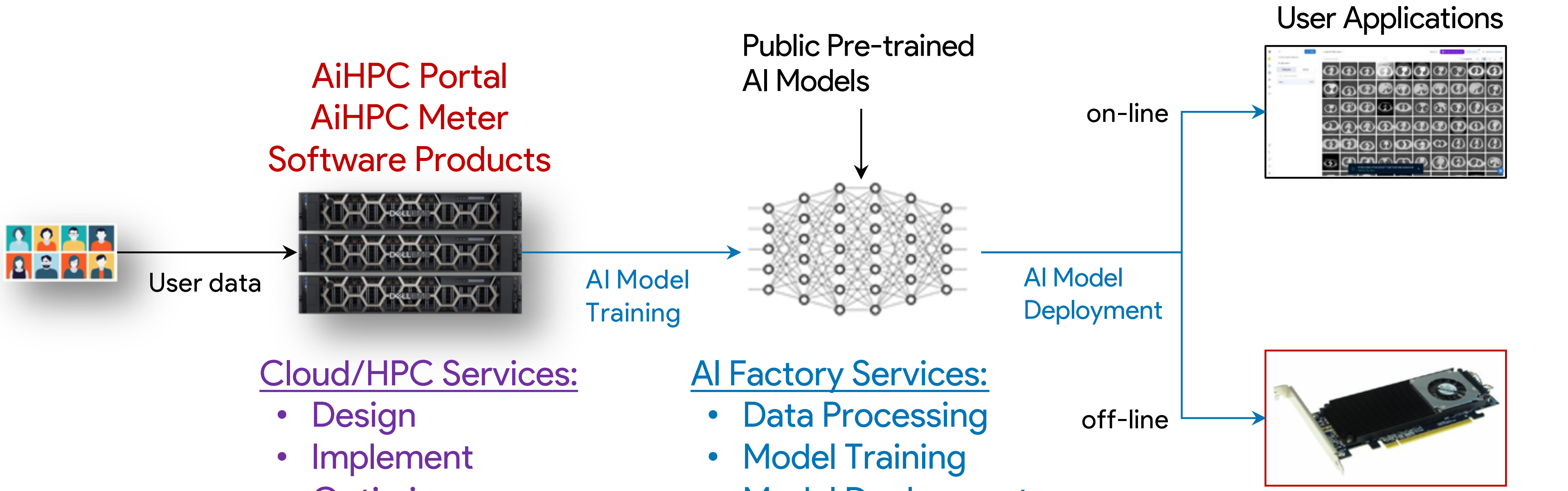
Chief Strategy Officer, Co-Founder

Mark Chang is a highly accomplished individual with a strong background in healthcare, technology, and government initiatives in Taiwan. With roles such as Taiwan Managing Partner, Assistant Professor, Committee Member, and Chief Information Officer, he contributes significantly to the advancement of healthcare and technology integration in Taiwan..

Hong Kong • Taiwan • Singapore

Providing local support with global standards across key Asian technology and healthcare hubs

Summary: Total Solution of Cloud/HPC/AI



Edge Inferencing Cards

<u>Cloud/HPC Services:</u>

- Design
- Implement
- Optimize
- Maintenance

<u>AI Factory Services:</u>

- Data Processing
- Model Training
- Model Deployment
  - User Application
  - PCIe Card

<u>Domain:</u>
Healthcare/Manufacturing/
Retail/Security/Finance/...