

General

1 What is Altos aiWorks?

Altos aiWorks provides integrated solutions including Altos Accelerator Resource Manager (AARM) management platform and various Altos BrainSphere™ AI systems from servers to workstations with Nvidia GPU. The solution helps students and developers deploy the environment quickly and allocate optimized CPU, GPU, memory, & storage resources automatically to confront the challenges of data science computing and AI tasks management.

2 How Altos aiWorks Optimize AI Resources?

It optimizes GPU resource allocation and reduces both the high complexity and learning threshold faced by developers in hardware resources and machine learning. Besides, when multiple users share hardware resources, an individual user can still maintain an independent development environment. It simplifies the complicated development processes and management; therefore, developers can focus more on their application developments.

3 Shall We Need an AI Expert to Plan AI Resources?

No need. Altos aiWorks already provides AI frameworks and AI development software stacks. The AI R&D supervisor or even the IT manager can allocate GPUs and deploy AI software stacks to students or developers within a one-click deployment.

4 What Kind of Operating System?

Linux Ubuntu 18.04 or above.

5 Is Altos aiWorks Docker-Based GPU Resource Management?

Yes, Altos aiWorks is a docker-based GPU resource allocation and AI development software stacks management.

6 What Kind of User Management dose Altos aiWorks Provide?

Altos aiWorks provides a 3-Tier Management Role such as the administrator, group manager, and user with different UI functions.

7 Does Altos aiWorks Provide any Kind of User Directory Access?

Altos aiWorks provides an LDAP (Lightweight Directory Access Protocol) to query and add accounts from Active Directory.

8 How User to Operate Altos aiWorks?

Users, who might be students or developers, can use their current desktop or laptop by using browsers to log in to Altos aiWorks to get optimized resources.

9 What Kind of Browsers Does Altos aiWorks Support to Access?

Microsoft Edge, Google Chrome, Firefox, Opera, & Safari.

10 Does Altos aiWorks Support Remote Access?

Yes, Altos aiWorks supports remote access to allow user login anywhere; but it depends on the university or enterprises' network security policies. Altos aiWorks will be under their security policies to provide services.

GPU

1 Does Altos aiWorks Support Multiple Graphics Cards in One System?

Yes. Altos aiWorks can work across multiple GPUs. This is not done automatically, however, so the Altos aiWorks has to be set-up for it.

2 What Kind of GPU Model Does Altos aiWorks Support?

- Data Center GPUs: NVIDIA A100 and A30 Tensor Core GPU, NVIDIA A40.
- Desktop Workstation GPU: NVIDIA RTX A6000.

3 Can Install Different Models of GPUs in One System?

No. Because each different model of GPU provides a different CUDA core, precision, and memory bandwidth, etc. Strongly recommend the same model of GPUs in one system.

4 Does Altos aiWorks Support One GPU Share to Multiple users?

Yes. Multiple users can share the same GPU card to maximize GPU efficiency by Altos aiWorks GPU sharing policy.

5 What Kind of GPU Sharing Policy Does Altos aiWorks has?

- **Multi-Instance GPU (MIG)** : Ideal for requiring more workloads such as AI model development and low-latency inference, the computing power required by users between dedicated and multi-process service sharing policy.
- **Multi-Process Service (MPS)** : It can be applied to AI teaching applications and teaching AI 101 does not require large computing power to run a small AI model.
- **Dedicated** : Applicable to users (professors, graduate students, or independent developers) who need computing power with independent GPU resources.

6 Based on GPU Sharing Policy, How Many Users Can Share One GPU?

- For Multi-Instance GPU (MIG) with NVIDIA A100 Tensor Core GPU is up to 7 users and A30 is up to 4 users.
- For Multi-Process Service (MPS) is up to 20 users and each user will get the same GPU memory which means more of GPU memory and each user will be getting more GPU memory, or otherwise instead.

7 Why Choose GPUs for Deep Learning and AI Training Instead of CPUs?

The major difference between GPUs and CPUs is that GPUs give over proportionally more transistors to arithmetic logic units and fewer to caches and flow control as compared to CPUs. GPUs are optimized for training artificial intelligence and deep learning models as they can process multiple computations simultaneously.

8 Can Altos aiWorks Monitor GPUs in real-time?

Yes. Altos aiWorks provides a visualizing resource metrics dashboard to show not only GPU, but CPU, memory, and AI job status by timeframe.

Software Stacks and Features

1 What Kind of Software and Framework Does Altos aiWorks Provide?

Altos aiWorks provides a bundle of AI packages: Jupyterlab, Tensorflow, R, Python, Keras, PyTorch, TensorboardX, Theano, Onnx, Horovod, OpenMpi, CUDA, and Cudnn.

2 Can Mount Other Software Stack, if Altos aiWorks Didn't Have One?

Yes. Group manager and user can mount their preferable development software stack.

3 Can Altos aiWorks Assist To Train An AI model, If Not An AI expert?

Yes. Altos aiWorks provides an AutoML appliance to assist a user who is not an AI expert to train an AI model. AutoML provides methods and processes to make Machine Learning available for non-Machine Learning experts, to improve the efficiency of Machine Learning, and to accelerate research on Machine Learning.

4 Does Altos aiWorks Support Any Appliance for Big Data?

Altos aiWorks integrates the most popular data science software which is MATLAB. Allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages.

5 Does Altos aiWorks Support Parallel Computing?

Yes. Altos aiWorks provides Spark and MPI to allow users to submit parallel computing jobs.

6 Does Altos aiWorks Support Inference?

Yes. Through the built-in NVIDIA Triton Inference Server, can deploy, run and scale trained AI models by developing teams uploaded, which can be verified by other users, saving the time of setting up the inference server.

7 Can Altos aiWorks Release the AI Resource Automatically?

Yes. Altos aiWorks can release all resources automatically when the user's request resources are idle based on the number of idle days and percentage of loading with CPU and GPU.

8 Does Altos aiWorks provide any message delivery?

Altos aiWorks provides an online bulletin to publish a message by grouping or all users; includes the automatically removing message by expiration date.

9 Does Altos aiWorks provide a Job Schedule?

Altos aiWorks provides 3 types of Jobs: a Single node job, and two parallel computing, Spark job, and MPI job for large computation and batch-oriented tasks each job has its own schedule to allow group managers and users can monitor job status; as administrator could delete their jobs when went wrong.

10 Does Altos aiWorks Support Cluster Topology?

Yes. Altos aiWorks provides multi-nodes as clusters to increase AI resource availability and prevent the loss of valuable time and information if one of the GPU nodes fails.

License & Warranty

1 What Kind of License Does Altos aiWorks Offer?

Altos aiWorks offers one perpetual license within one server node. It is not transferable to other servers if the current server, from a hardware point of view, is out of service.

2 What Kind of Warranty Does Altos aiWorks provide?

Altos aiWorks provides a First-Year Warranty to cover technical support, application and library updates, and core-base upgrades, if necessary.

3 Does any Service Support in Second-Year or More?

It can purchase an Extension Warranty to extend the original limited warranty of Altos aiWorks solution to provide extended coverage.