

# Micro-Stack: Open-Source AI Infrastructure Suite

Three Open-Source Tools for Private AI Infrastructure Setup + One RAG Demo App

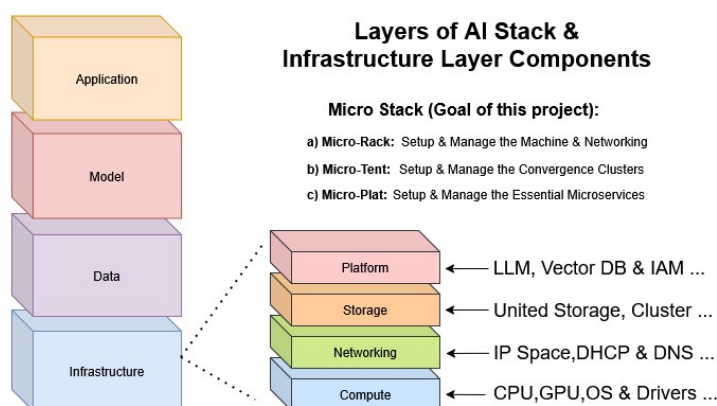
(\* This documentation was created with the help of AI. If you have any feedback or suggestions, feel free to chat with our robot assistant in the bottom right corner of the page. \*)

## I. Introduction

Micro-Stack is an open-source suite of tools designed to simplify building private AI infrastructures from the ground up. It enables AI researchers, developers, and organizations to deploy scalable, no-code AI stacks without deep technical expertise. By automating infrastructure setup, Micro-Stack accelerates experimentation and reduces barriers to AI innovation.

Designed for AI practitioners, system administrators, and teams needing reliable, open-source infrastructure, Micro-Stack supports everything from small-scale testing to enterprise-grade deployments. Whether you're prototyping a new AI model or scaling production workloads, Micro-Stack provides the building blocks for efficient, cost-effective AI development.

Key benefits include reduced setup time, seamless integration between components, and support for modern technologies like containerization and distributed storage. Micro-Stack empowers users to focus on AI innovation rather than infrastructure complexities.



1. **Micro-Rack:** A PXE boot environment for automated Linux OS installations on bare metal computers. It enables rapid deployment of high-performance computing resources, saving time on large-scale setups. Ideal for AI researchers needing compute clusters without manual configuration—it supports diverse hardware and scales to thousands of devices.
2. **Micro-Tent:** A hyper-converged infrastructure (HCI) deployment tool that automatically discovers and manages network resources. It builds resilient HCI clusters with storage,

compute, and networking. Perfect for creating stable foundations for AI workloadsâ€”leverages open-source technologies like Ceph and Docker Swarm for flexibility and high availability.

3. **Micro-Plat:** A performance-focused RAG AI platform deployment environment. It identifies HCI clusters from Micro-Tent and sets up AI platforms for large user bases. Streamlines RAG (Retrieval-Augmented Generation) applicationsâ€”automates scaling and management to handle demanding AI tasks efficiently.
4. **Micro-Play:** A ready-to-use RAG AI chatbot boilerplate. It runs on Micro-Plat platforms, providing a demo interface for document upload and AI chat. Simplifies AI app developmentâ€”includes pre-built components for quick experimentation and prototyping.

## II. Architecture

Micro-Stack follows a layered architecture that builds AI-ready infrastructure step-by-step. This modular design allows users to start small and scale up as needed, with each layer providing specific functionality while integrating seamlessly with others.

The stack operates in four main layers:

1. **Hardware Provisioning (Micro-Rack):** Handles bare metal setup using PXE booting for automated OS installations across multiple devices.
2. **Cluster Formation (Micro-Tent):** Discovers resources and creates hyper-converged infrastructure (HCI) clusters with storage, compute, and networking.
3. **AI Platform Deployment (Micro-Plat):** Builds on HCI clusters to deploy scalable AI platforms optimized for RAG applications.
4. **Application Hosting (Micro-Play):** Provides demo AI chatbots and tools that run on the deployed platforms.

This layered approach ensures reliability and scalability, allowing components to be updated independently while maintaining overall system integrity.

Key concepts include:

- **PXE (Preboot Execution Environment):** Enables network-based OS installations without local media.
- **HCI (Hyper-Converged Infrastructure):** Combines storage, compute, and networking in a single platform for simplified management.
- **RAG (Retrieval-Augmented Generation):** Enhances AI responses by retrieving relevant information from knowledge bases.

For deeper technical details, refer to the separate Technical Documentation.

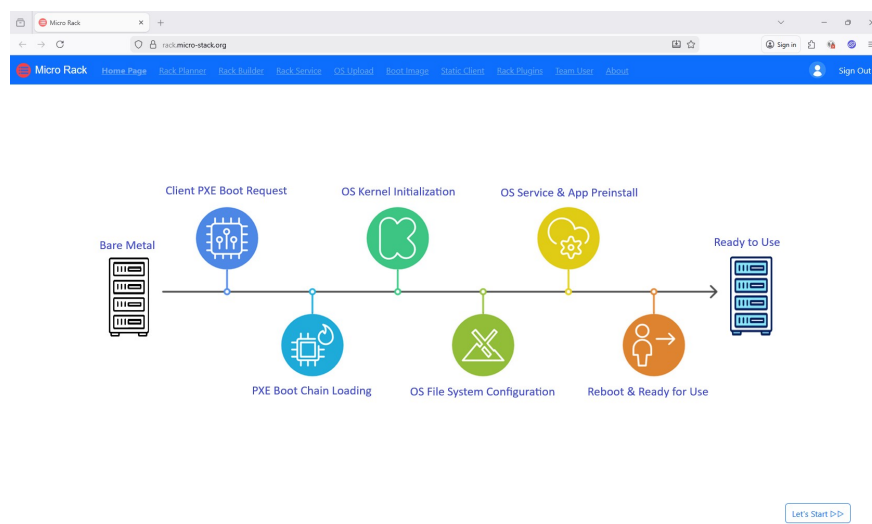
## III. Guides

This section provides step-by-step guides for using each Micro-Stack component. Start with Micro-Rack for hardware setup, then progress through Micro-Tent for clustering, Micro-Plat for AI platform deployment, and Micro-Play for applications. Each guide includes workflows, tips, and screenshots.

to help you get started quickly.

## Micro-Rack

Micro-Rack provides a user-friendly interface for managing PXE server operations. Use the demo credentials (username: demo, password: demo) to explore. It supports booting from thousands of devices. The [source code repository](#) has more details.



**Workflow:** Plan network settings â†’ Upload OS images â†’ Configure clients â†’ Boot devices.

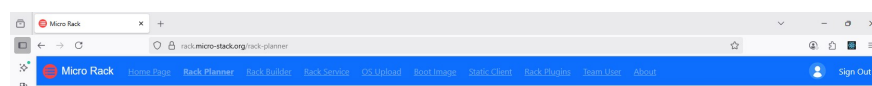
### Steps:

1. [Plan Network Subnet](#)
2. [Setup PXE Environment](#)
3. [Check PXE Health](#)
4. [Upload OS ISOs](#)
5. [Manage OS Images](#)
6. [Manage Static IPs](#)
7. [Manage Plugins](#)
8. [Manage Team](#)
9. [Boot Clients](#)

**Tip:** Ensure your network supports PXE booting. Start with demo mode to explore safely.

## 1. Plan Network Subnet

Begin by accessing the Planner interface from the home page. Here, you can configure network subnets to define the IP ranges for your PXE boot environment. Ensure that the subnets align with your network topology to avoid conflicts. This step is crucial for establishing the foundation of your automated OS installation setup. Once configured, proceed to the next steps for uploading images and booting devices.



The screenshot shows the 'PXE Environment Planner' web interface. It contains three main sections: 'Network Interface Assignment', 'DHCP Configuration Options', and 'PXE Root & Auto Conf Folder'. The 'Network Interface Assignment' section has fields for 'Network Interface Binding' (set to 'enp6s21'), 'Interface Mac IP Address' (set to 'bc24:11:54:37:7f'), and 'Pxe Service IP Binding' (set to '172.16.80.198/24'). The 'DHCP Configuration Options' section has fields for 'Net Address' (172.16.80.0), 'Net Mask' (255.255.255.0), 'IP Pool Begin' (172.16.80.101), 'IP Pool End' (172.16.80.190), 'Router/Gateway Address' (172.16.80.1), 'Local DNS Server' (172.16.80.1), and 'Local Domain Name' (example.lan). The 'PXE Root & Auto Conf Folder' section has fields for 'PXE Root Folder' (/pxeboot) and 'OS Auto-setup Conf Folder' (/pxeboot/os-install). There are 'Reset' and 'Save' buttons at the bottom left, and a 'Rack Builder >>' button at the bottom right.

## 2. Setup PXE Environment

Next, utilize the Builder tool to apply the configurations set in the Planner. This involves setting up the PXE server with the necessary boot files and network settings. The Builder ensures that all components are properly integrated for seamless operation. You can monitor the setup process through the interface. Once completed, your PXE environment will be ready for OS installations.

The screenshot shows the 'PXE Environment Builder' web interface. It contains three main sections: 'Network Interface Assignment', 'Local Network Services', and 'PXE Boot & OS Image Server'. The 'Network Interface Assignment' section has fields for 'Network Interface Binding' (set to 'enp6s21'), 'Interface Mac IP Address' (set to 'bc24:11:54:37:7f'), and 'Pxe Service IP Binding' (set to '172.16.80.198/24'). The 'Local Network Services' section has fields for 'DHCP Server' (dnsmasq), 'DHCP Proxy' (None), 'DNS Server' (dnsmasq), and 'NTP Server' (ntp). The 'PXE Boot & OS Image Server' section has fields for 'PXE Boot Loader' (ipxe), 'OS Image TFTP Server' (dnsmasq), 'OS Image HTTP Server' (nginx), and 'OS Image NFS Server' (nfs-kernel-server). There is a 'Dependent Software Packages' section with checkboxes for 'genisoimage', 'cpio', 'zstd', and '7zip'. There are 'Reset', 'Save', and 'Apply' buttons at the bottom left, and '<< Rack Planner' and 'Rack Service >>' buttons at the bottom right.

## 3. Check PXE Health

Access the Health section to monitor the overall system status and run diagnostics. This helps identify any issues with the PXE setup or network connectivity. Regular checks ensure optimal performance and reliability. Use the diagnostics tools to troubleshoot problems early. Maintaining system health is key to successful deployments.

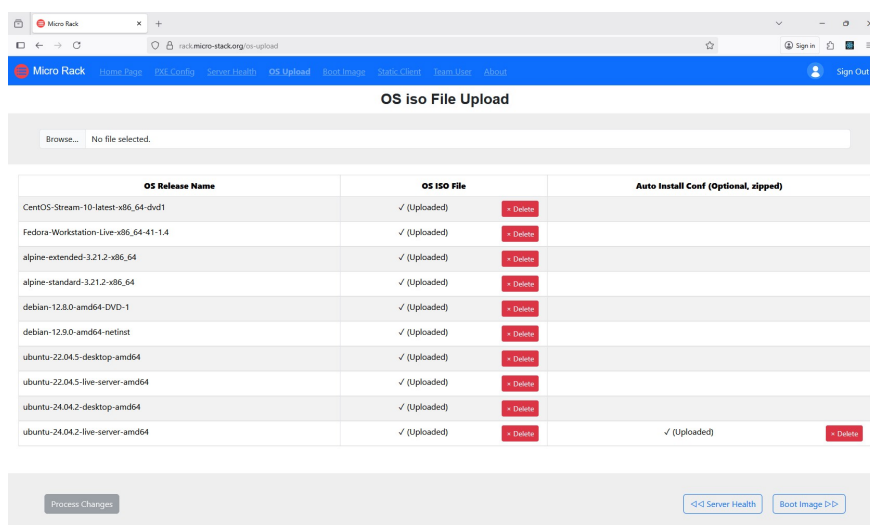
The screenshot shows the 'PXE Service Status' web interface, which displays a table of service status information.

Service Name	Service Protocol	Service Port	Service Provider	Service Status
BOOT_LOADER (*)	(n/a)	(n/a)	ipxe	✓ (Running)
DHCP_SERVER (*)	UDP	67	dnsmasq	✓ (Running)
DHCP_PROXY ()	UDP	67	(None)	(Not found)
TFTP_SERVER (*)	UDP	69	dnsmasq	✓ (Running)
HTTP_SERVER (*)	TCP	80	nginx	✓ (Running)
NFS_SERVER	TCP	2049	nfs-kernel-server	(Not found)
DNS_SERVER	TCP	53	dnsmasq	✓ (Running)



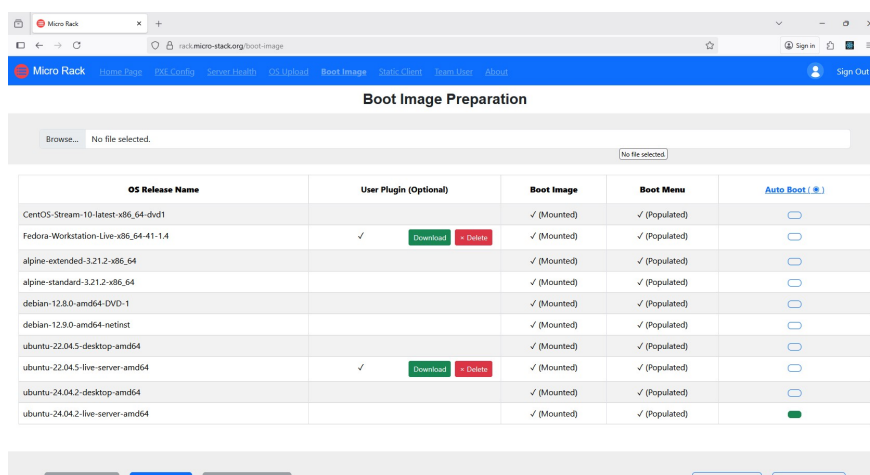
## 4. Upload OS ISOs

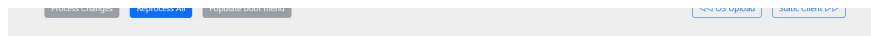
In this step, upload the ISO files for the operating systems you want to deploy. Ensure the ISOs are compatible with your hardware and PXE environment. The upload process may take time depending on file size and network speed. Once uploaded, the ISOs are stored and ready for use in installations. Proper ISO management is essential for efficient OS deployment.



## 5. Manage OS Images

After uploading ISOs, manage the OS images through the dedicated interface. This includes organizing, updating, or removing images as needed. Image management allows you to maintain a library of available OS versions. Ensure images are optimized for your deployment scenarios. Effective image management streamlines the boot process.





## 6. Manage Static IPs

Configure static IP addresses for client machines to ensure consistent network identification. This step involves assigning IPs within the defined subnets. Static IPs prevent conflicts and improve reliability in automated setups. Use the interface to input IP details for each client. Proper IP configuration is crucial for PXE booting.

The screenshot shows the 'Host Static IP Assignment' page. It features a table with columns: mac\_address, ipv4\_address, ipv6\_address, hostname, lease\_time, is\_active, pingable, and Action. The table lists 20 hosts with their respective MAC addresses and assigned static IP addresses. The 'Action' column contains icons for editing and deleting each row. Below the table, there are buttons for 'Apply Static IP Settings' and 'Ping All Static Host IPs'. A pagination bar at the bottom indicates '1 of 21' items and a 'Show 15' dropdown.

mac_address	ipv4_address	ipv6_address	hostname	lease_time	is_active	pingable	Action
BC24115EC162	172.16.60.101		host-060101	24h	true	false	[edit] [delete]
BC241157A856E	172.16.60.102		host-060102	24h	true	false	[edit] [delete]
BC241161D253	172.16.60.103		host-060103	24h	true	false	[edit] [delete]
BC2411529C353	172.16.60.104		host-060104	24h	true	false	[edit] [delete]
BC241150404D	172.16.60.105		host-060105	24h	true	false	[edit] [delete]
BC241152DE4F7	172.16.60.106		host-060106	24h	true	false	[edit] [delete]
BC24115190D30	172.16.60.107		host-060107	24h	true	false	[edit] [delete]
BC2411559484	172.16.60.108		host-060108	24h	true	false	[edit] [delete]
BC24115A5836	172.16.60.109		host-060109	24h	true	false	[edit] [delete]
BC24116F040C	172.16.60.101		host-080101	24h	true	true	[edit] [delete]
BC24115F1509	172.16.60.102		host-080102	24h	true	true	[edit] [delete]
BC24115D0631A	172.16.60.103		host-080103	24h	true	true	[edit] [delete]
BC241157D5606	172.16.60.104		host-080104	24h	true	true	[edit] [delete]
BC24115F814B	172.16.60.105		host-080105	24h	true	true	[edit] [delete]
BC24115B3F5D0	172.16.60.106		host-080106	24h	true	true	[edit] [delete]

## 7. Manage Plugins

Enhance your setup with plugins for additional customization options. Plugins can add features like custom scripts or integrations. Manage plugins through the interface to install, update, or remove them. Choose plugins that align with your deployment needs. Plugin management allows for tailored PXE environments.

The screenshot shows the 'Rack Plugin Manager' page. It displays a table of installed plugins categorized by 'Plugin Category'. The categories shown are 'pxe', 'dhcp', 'proxy', 'tftp', and 'http'. Each category has a list of plugins with columns for 'Plugin Name', 'Plugin Type', 'User Uploaded Plugin', 'Plugin Mounted', and 'Plugin Loadable'. The 'pxe' category includes 'grub', 'ipxe', and 'syllinux'. The 'tftp' category includes 'chromasq' and 'tftp-lpa'. The 'http' category is currently empty.

Plugin Name	Plugin Type	User Uploaded Plugin	Plugin Mounted	Plugin Loadable
grub	pxe		✓	
ipxe	pxe		✓	✓
syllinux	pxe		✓	
dhcp				
proxy				
tftp				
chromasq	tftp		✓	✓
tftp-lpa	tftp		✓	
http				

## 8. Manage Team

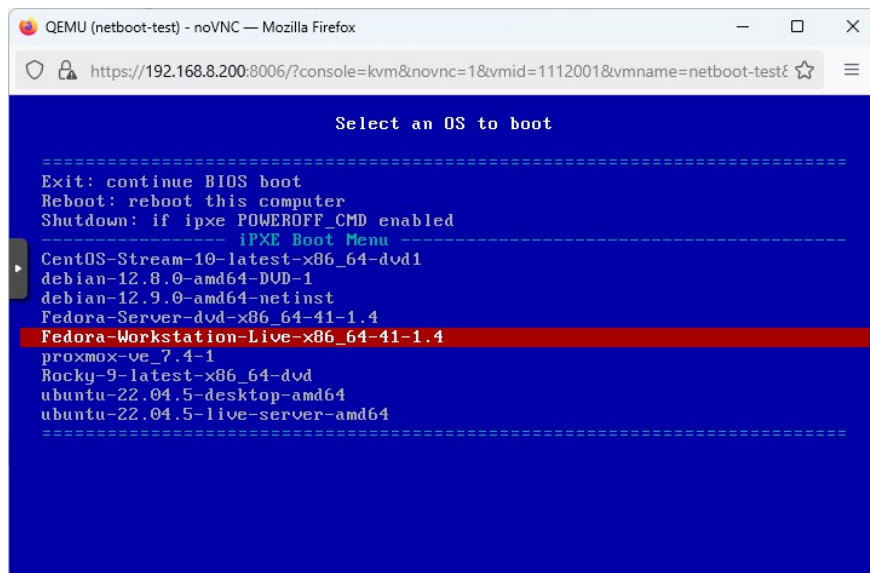
Set up team users and profiles for collaborative access to the system. This involves creating accounts with appropriate permissions. User management ensures secure and controlled access. Assign roles based on responsibilities. Proper team management facilitates efficient teamwork.

Login Name (▲▼)	Full Name (▲▼)	Email (▲▼)	Role (▲▼)	Created Date (▲▼)	Active (▲▼)	Modifiers
John	John Doe	scineros@fb.com	Viewer	2024-05-20	Yes	<a href="#">Edit</a> <a href="#">Delete</a>
abjork	Angela Bjork	abjork@helms.com	Admin	2024-05-22	No	<a href="#">Edit</a> <a href="#">Delete</a>
mkrisin	Maria Krisin	mkristin@airatlanta.com	User	2024-05-22	Yes	<a href="#">Edit</a> <a href="#">Delete</a>
amargret	Anna Margret	amargret@geko.com	User	2024-05-22	No	<a href="#">Edit</a> <a href="#">Delete</a>
admin	John Doe	admin@micro-rack.org	Admin	2025-03-07	Yes	<a href="#">Edit</a> <a href="#">Delete</a>
john	John Smith	jsmith@homelab.lan	Viewer	2025-03-19	Yes	<a href="#">Edit</a> <a href="#">Delete</a>
jing	Jing Zhong	zhong@homelab.lan	Admin	2025-03-19	No	<a href="#">Edit</a> <a href="#">Delete</a>
tester	tester	tester@homelab.lan	Admin	2025-03-27	Yes	<a href="#">Edit</a> <a href="#">Delete</a>
user	user	user@homelab.lan	User	2025-04-01	Yes	<a href="#">Edit</a> <a href="#">Delete</a>
viewer	viewer	viewer@homelab.lan	Viewer	2025-04-01	Yes	<a href="#">Edit</a> <a href="#">Delete</a>

[Add Team Member](#) [Home Page](#) [About](#)

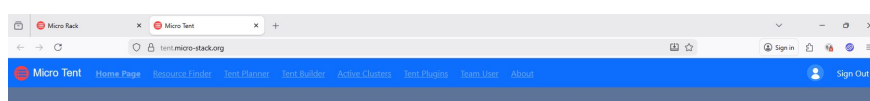
## 9. Boot Clients

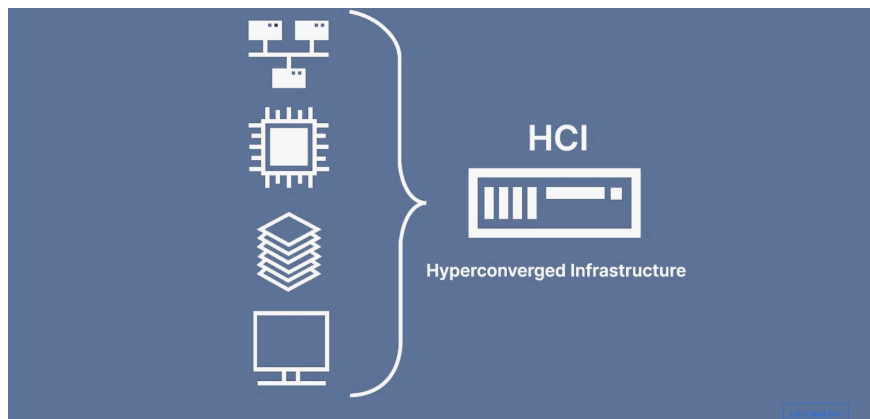
Finally, initiate PXE booting on client computers. Ensure clients are configured to boot from the network. Monitor the boot process for successful OS installation. This step completes the automated deployment. Troubleshoot any issues that arise during booting.



## Micro-Tent

Micro-Tent automates HCI setup across your network. Access the [live demo](#) with demo/demo. It discovers resources and builds clusters. See the [GitHub repo](#) for more.





**Workflow:** Discover resources → Plan clusters → Build HCI → Manage clusters.

### Steps:

1. [Discover Resources](#)
2. [Plan Resources](#)
3. [Build Clusters](#)
4. [Check Clusters](#)
5. [Manage Plugins](#)
6. [Manage Team](#)

**Tip:** Ensure devices are on the same network. Clusters from Micro-Tent can be used in Micro-Plat.

## 1. Discover Resources

Utilize the Resource Finder tool to scan your network for available devices. This automated discovery helps identify potential resources for clustering. Ensure your network is properly configured for scanning. The tool provides a list of discovered devices. Use this information to plan your cluster formation.

Host number	IPv4 address	Proposed hostname	Ping-able	SSH-able	Toggle all (in/out)	Already in use
1	172.16.80.101	host-080101	✓	✓	⛔	✓
2	172.16.80.102	host-080102	✓	✓	⛔	✓
3	172.16.80.103	host-080103	✓	✓	⛔	✓
4	172.16.80.104	host-080104	✓	✓	⛔	✓
5	172.16.80.105	host-080105	✓	✓	⛔	✓
6	172.16.80.106	host-080106	✓	✓	⛔	✓
7	172.16.80.107	host-080107	✓	✓	⛔	✓
8	172.16.80.108	host-080108	✓	✓	⛔	✓
9	172.16.80.109	host-080109	✓	✓	⛔	✓

## 2. Plan Resources

Access the Resource Planner to allocate discovered resources for your clusters. This involves selecting devices and assigning roles. Proper allocation ensures balanced and efficient clusters. Review the planner's recommendations. Confirm your selections before proceeding.

The screenshot shows the 'Resource Planner' interface in the Micro Tent application. At the top, there are dropdown menus for 'Interface' (set to 'enp6s21'), 'Subnet' (set to '172.16.80.198/24'), and 'Cluster' (set to 'hci-cluster-2 (list with free hosts)'). An 'Action' button labeled 'Edit Cluster Stub' is to the right. Below these is a table with columns: 'Host number', 'IPv4 address', 'Host name', 'Cluster Member (All In/out)', 'Storage node (All In/out)', 'Log', 'node', 'Compute type', and 'Balancer Node (All In/out)'. The table lists three hosts with their respective IP addresses and host names. At the bottom, there are buttons for 'Save Allocation', '<< Resource Finder', and 'Cluster Builder >>'. A dropdown menu is open for the 'Cluster' selection, showing options: 'hci-cluster-2 (list with free hosts)', '(All free hosts in subnet)', 'hci-cluster-1 (list with free hosts)', 'hci-cluster-2 (list with free hosts)', and 'hci-cluster-3 (list with free hosts)'.

Host number	IPv4 address	Host name	Cluster Member (All In/out)	Storage node (All In/out)	Log	node	Compute type	Balancer Node (All In/out)
1	172.16.80.104	host-080104			/dev/sdb	Hard Disk		(CPU On)
2	172.16.80.105	host-080105			/dev/sdb	Hard Disk		(CPU On)
3	172.16.80.106	host-080106			/dev/sdb	Hard Disk		(CPU On)

### 3. Build Clusters

Use the Cluster Builder tool to create hyper-converged clusters from allocated resources. This step integrates storage, compute, and networking. Monitor the building process for any issues. The tool automates much of the configuration. Successful building results in operational clusters.

The screenshot shows the 'Cluster Builder' interface. It has dropdown menus for 'Interface' (enp6s21), 'Subnet' (172.16.80.198/24), and 'Cluster' (hci-cluster-1). An 'Action' button 'Assign Components to Selected HCI' is present. Below is a table with columns: 'Host number', 'IPv4 address', 'Host name', 'Storage Role', 'Compute Role', and 'Balancer Role'. The table shows three hosts with roles assigned: Host 1 is Control Plane, Host 2 is Worker, and Host 3 is Worker. At the bottom are buttons for 'Save Config', 'Build Cluster', 'Setup Dashboard', '<< Resource Planner', and 'Active Clusters >>'. A dropdown menu is open for the 'Cluster' selection, showing options: 'hci-cluster-1', '(All free hosts in subnet)', 'hci-cluster-2 (list with free hosts)', 'hci-cluster-2 (list with free hosts)', and 'hci-cluster-3 (list with free hosts)'.

Host number	IPv4 address	Host name	Storage Role	Compute Role	Balancer Role
1	172.16.80.101	host-080101	Hybrid	Control Plane	(None)
2	172.16.80.102	host-080102	Hybrid	Worker	(None)
3	172.16.80.103	host-080103	Hybrid	Worker	(None)

### 4. Check Clusters

Monitor your active clusters through the dedicated interface. Check status indicators and performance metrics. This helps maintain optimal operation. Address any alerts or issues promptly. Regular monitoring ensures reliability.

The screenshot shows the 'Active HCI Clusters' interface. It features a table with columns: 'Admin Subnet', 'Cluster Name', 'Cluster Virt-IP', 'Total Members', 'Compute Cluster', 'Compute Nodes', 'Storage Cluster', 'Storage Nodes', and 'Cluster Status'. The table is currently empty.

Admin Subnet	Cluster Name	Cluster Virt-IP	Total Members	Compute Cluster	Compute Nodes	Storage Cluster	Storage Nodes	Cluster Status
--------------	--------------	-----------------	---------------	-----------------	---------------	-----------------	---------------	----------------

172.16.80.198/24	hci-cluster-1	172.16.80.201/24	3	Kubernetes	3	Ceph	3	✓ (Running)
172.16.80.198/24	hci-cluster-2	172.16.80.202/24	3	Swarm	3	Gluster	3	✓ (Running)
172.16.80.198/24	hci-cluster-3	172.16.80.203/24	3	OpenStack	3	(None)	3	✓ (Running)

[Page Refresh](#)[Cluster Builder](#)

## 5. Manage Plugins

Manage plugins to customize your cluster environment. Plugins can add features or integrations. Install or remove plugins as needed. Ensure compatibility with your setup. Plugin management enhances functionality.

Micro Tent

Home PageResource ExplorerResource PlannerCluster BuilderActive ClustersTent PluginsTeam UsersAbout

Sign InSign Out

Tent Plugin Manager

Browse...

storage.Test.zip

Plugin Name	Plugin Type	User Uploaded Plugin	Plugin Mounted	Option Loaded
Test	storage	storage.Test.zip <a href="#">Delete</a>	✓	✓
Dedicated-KeepAlived-LB	balancer		✓	✓
Overlapped-KeepAlived-HA	balancer		✓	✓
Kubernetes	compute		✓	✓
OpenStack	compute		✓	✓
Swarm	compute		✓	✓
Ceph	dashboard		✓	✓
Portainer	dashboard		✓	✓
Ceph	storage		✓	✓
Gluster	storage		✓	✓

Process Changes

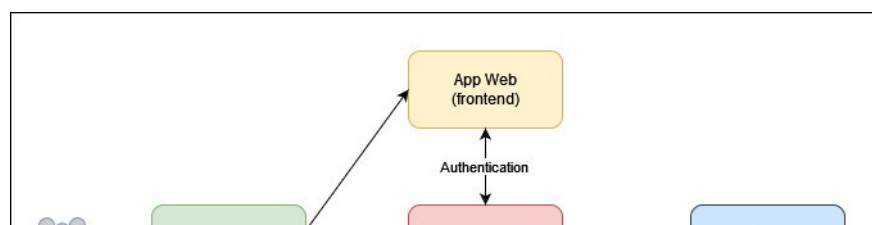
[Home Page](#)[Team Users](#)

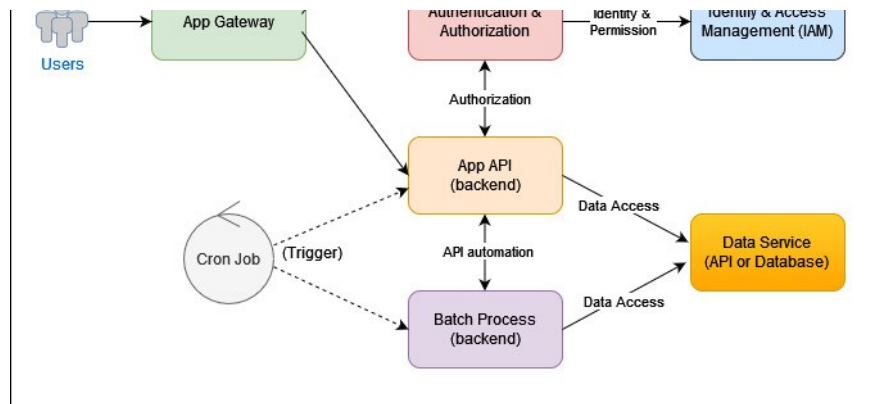
## 6. Manage Team

[\(Please refer to the Micro-Rack Manage Team step.\)](#)

## Micro-Plat

Micro-Plat deploys RAG AI platforms on HCI clusters. Try the [demo](#) with demo/demo. It auto-discovers clusters from Micro-Tent. Check the [repo](#).





**Workflow:** Discover clusters â†’ Plan platforms â†’ Build AI setups â†’ Manage platforms.

### Steps:

1. [Manage Platforms](#)
2. [Plan Platforms](#)
3. [Build Platforms](#)
4. [Check Platforms](#)
5. [Manage Plugins](#)
6. [Manage Team](#)

**Tip:** Platforms support large user bases. Use for deploying RAG applications efficiently.

## 1. Manage Platforms

Use the Plat Manager to create new platforms or select existing ones. This interface allows you to define platform configurations. Ensure your selections match your requirements. The manager provides options for customization. Proceed to planning once selected.

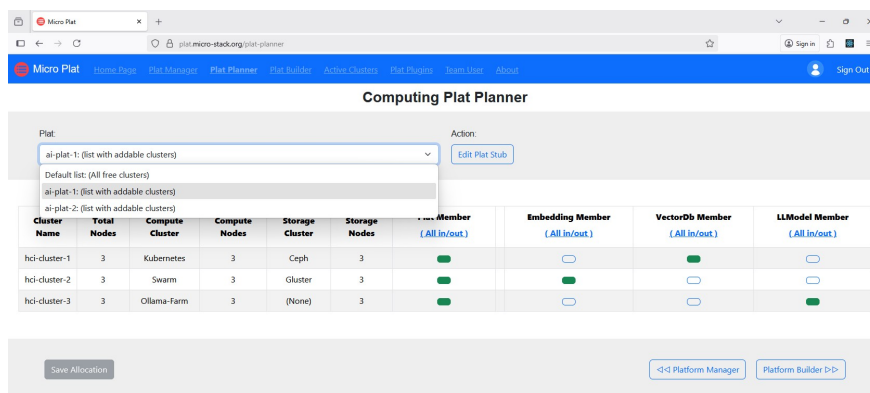
Plat Name (▲▼)	Plat Note (▲▼)	Plat vIP (▲▼)	Plat Type (▲▼)	Created At (▲▼)	Active (▲▼)	Modifiers
ai-plat-1	ai platform 1	http://172.16.60.201:1443		2025-08-15	Yes	<a href="#">Edit</a> <a href="#">Delete</a>
ai-plat-2	ai platform 2	http://172.16.60.202:10443		2025-08-16	No	<a href="#">Edit</a> <a href="#">Delete</a>

[Add A New Plat](#) [Platform Planner >>](#)

## 2. Plan Platforms

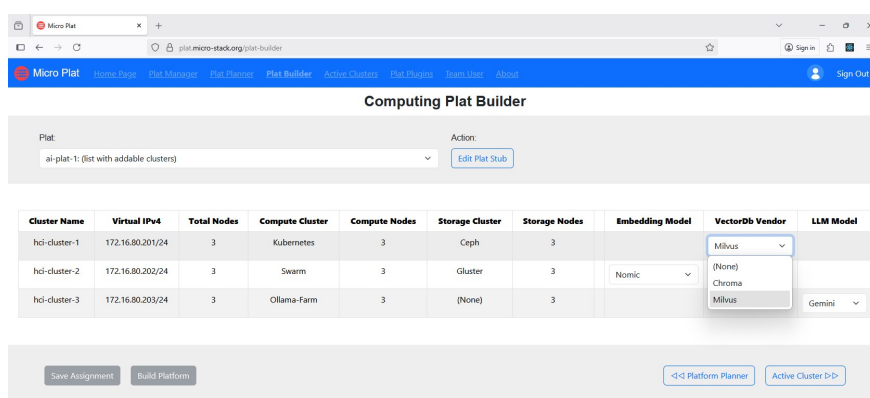
Plan resources with the Plat Planner to allocate necessary components. This involves spec.,

compute, storage, and networking needs. Review the planner's suggestions for optimization. Adjust allocations as needed. Proper planning ensures efficient platform deployment.



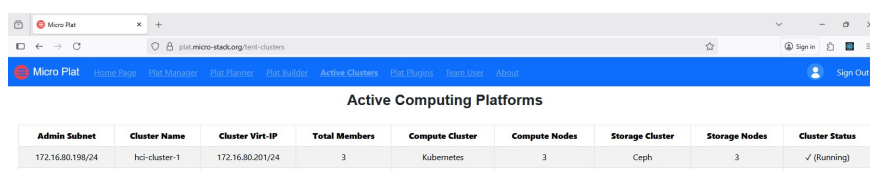
### 3. Build Platforms

Build platforms via the Plat Builder to assemble the configured resources. This step deploys the platform on the selected clusters. Monitor the build process for completion. The builder automates the setup. Successful builds result in ready-to-use platforms.



### 4. Check Platforms

View Active Plats to monitor running platforms and their status. This interface shows performance and health metrics. Check for any issues or updates. Use this to manage active deployments. Regular viewing ensures ongoing reliability.



172.16.80.198/24	hc-cluster-2	172.16.80.202/24	3	swarm	3	cluster	3	✓ (running)
172.16.80.198/24	hc-cluster-3	172.16.80.203/24	3	Ollama-Farm	3	(None)	3	✓ (Running)

Page Refresh

&lt; Platform Builder

## 5. Manage Plugins

Manage plugins for AI customizations to enhance platform functionality. Plugins can add AI models or integrations. Install or configure plugins as required. Ensure compatibility with your AI applications. Plugin management optimizes AI performance.

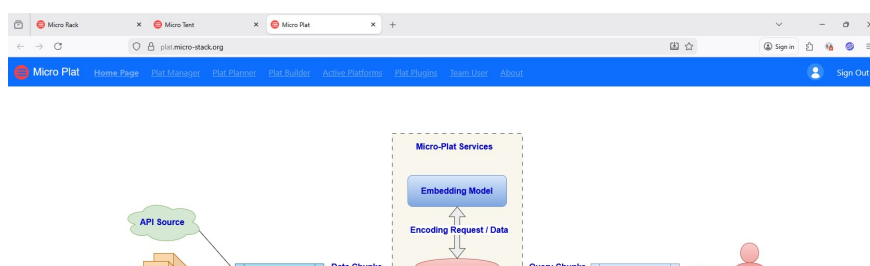
Plugin Name	Plugin Type	User Uploaded Plugin	Plugin Mounted	Option Loaded
Test	embedding	embedding.test.zip <span>Delete</span>	✓	✓
NVIDIA	embedding		✓	✓
Nomic	embedding		✓	✓
GPT-4o	llm		✓	✓
Gemini	llm		✓	✓
Chroma	vectordb		✓	✓
Milvus	vectordb		✓	✓

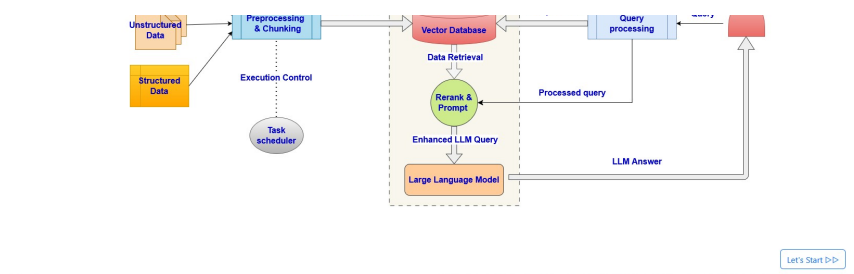
## 6. Manage Team

(Please refer to the [Micro-Rack Manage Team step.](#))

## Micro-Play

Micro-Play is a demo RAG AI chatbot running on Micro-Plat platforms. Explore the [live app](#) to see RAG in action. It uses small AI models for quick demos. See the [repo](#).





**Workflow:** Upload documents â†’ Interact with chat.

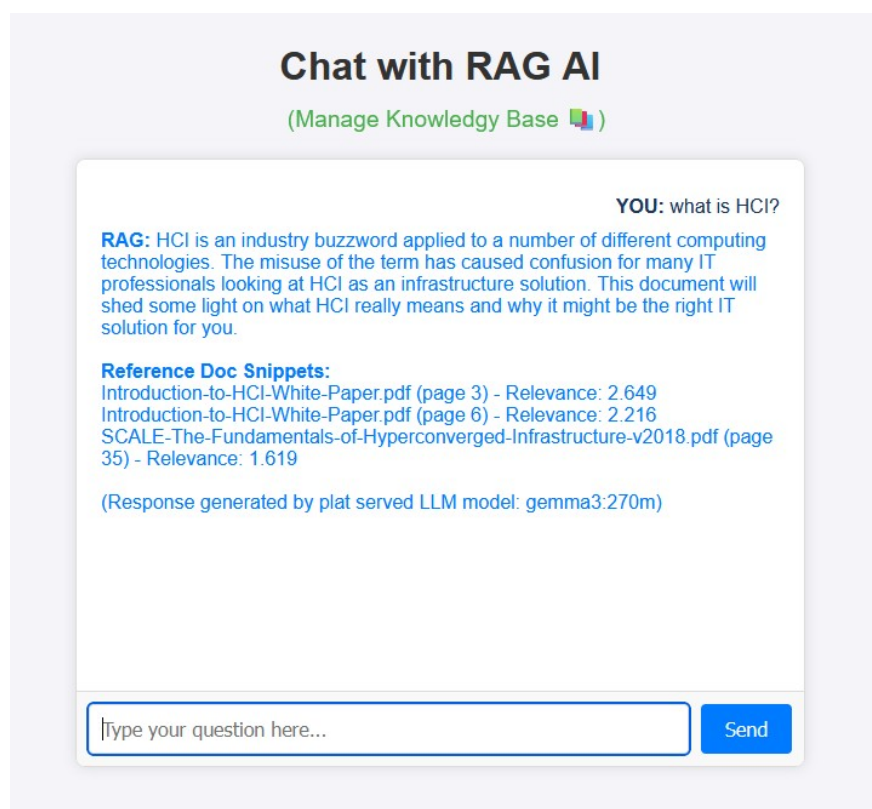
### Steps:

1. [Chat with RAG AI](#)
2. [Manage Documents](#)

**Tip:** This is a demoâ€”use it to understand RAG pipelines before building your own apps on Micro-Plat.

## 1. Chat with RAG AI

Use the Web Chat interface for interactive AI conversations. Engage with the RAG AI model through natural language. Ask questions and receive responses based on uploaded documents. Explore different topics and queries. This demo showcases AI capabilities.



## 2. Manage Documents

Access the Doc Admin to upload and manage documents for the knowledge base. Add files to enhance AI responses. Organize and update documents as needed. Ensure documents are relevant and accurate. Proper management improves AI performance.

### Upload Knowledge Base Files

Choose a File

Upload

No file selected

### Uploaded Document Files

ebook-hyperconverged-infrastructure.pdf	Delete
Introduction-to-HCI-White-Paper.pdf	Delete
SCALE-The-Fundamentals-of-Hyperconverged-Infrastructure-v2018.pdf	Delete

Index Knowledge Base

Back to Chatting Page