Fixstars AI Booster

Quick Start Guide

Introduction

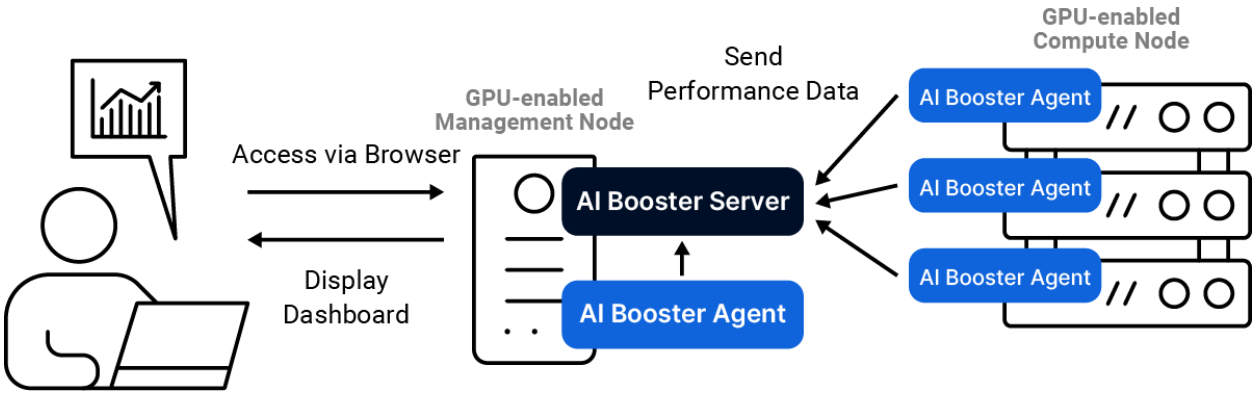
Fixstars AI Booster consists of two components:

- PO: Performance Observation Dashboard
- PI: Performance Improvement Framework

PO provides visibility into the performance of your systems. Simply install Fixstars AI Booster on one or more nodes and open the dashboard in your browser to view information about the hardware and software running on those nodes.

PI is an AI performance improvement framework that can be applied manually by users. This package includes a tuning tool for hyperparameters related to distributed learning of AI. After installation,Please refer to the documents and source code located at `/opt/aibooster`.

System Configuration



The Fixstars AI Booster dashboard consists of multiple Docker containers running on a Linux node. There are two types of containers depending on their roles:

- Agent: Collects various performance information (telemetry)
- Server: Data storage and visualization

The Agent containers are assumed to be running all the time on all nodes to be monitored. These containers monitor the hardware and system status of the nodes at regular points and collect metrics on the performance of programs running on them. They have the following functions, and some containers require operation in privileged mode (container startup with administrator privileges).

- Node Exporter: Collects CPU and I/O related metrics
- DCGM Exporter: Collects GPU metrics
- PCM Exporter: Collects metrics specific to the Intel CPU/Memory Subsystem
- eBPF Profiler: Collects program execution status

The Server containers are assumed to run on a Linux node connected to the same network as the Compute Node on which the Agent runs. They can be placed on a dedicated management node or coexist with one of the compute nodes on which the Agent is installed. The containers included in the Server are as follows:

- ClickHouse: Stores data
- Grafana: Provides visualization capabilities
- Nginx: Acts as a reverse proxy

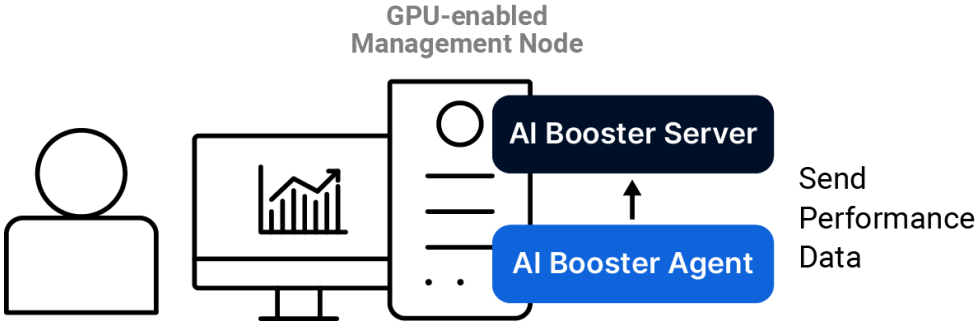
In addition, the following ports must be open on the nodes where the Server containers are running:

Port number	Expected access source	Purpose
3000	User's PC	Access to the Performance Observation Dashboard
9000	Nodes on which the Agent runs	Metrics Collection

Typical Configuration

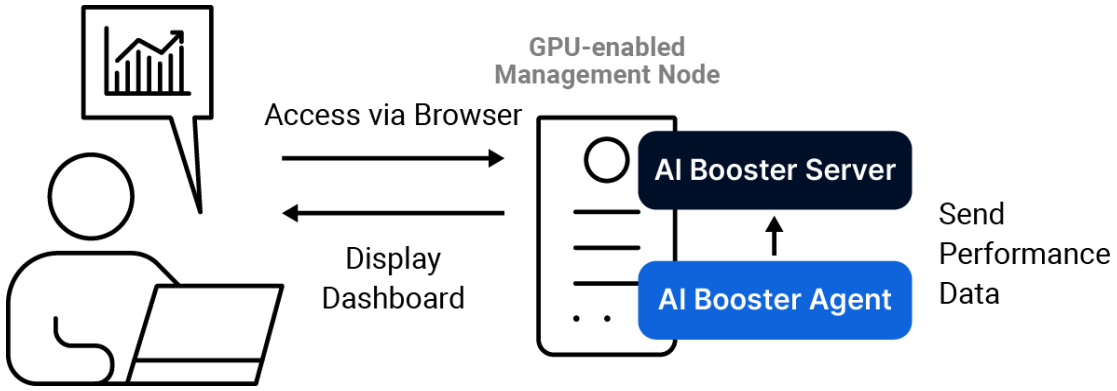
Single node configuration 1 - Minimum configuration for verification

Both the AI Booster Server and the AI Booster Agent are installed on a single workstation server equipped with a GPU.Connect a monitor and check performance information on the fly from the dashboard.This is the shortest route when you want to try it out on an offline verification machine or benchmark machine. No network settings are required.



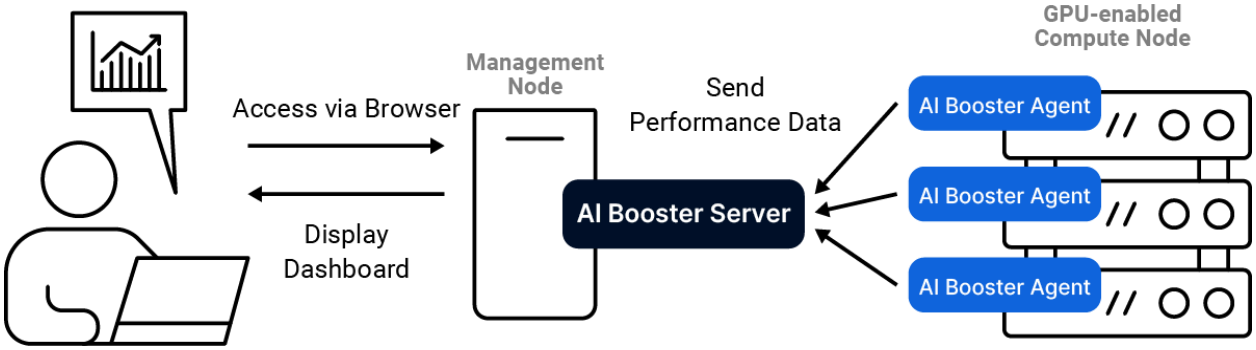
Single-node configuration 2 - Configuration for testing by multiple users

Both the AI Booster Server and the AI Booster Agent are installed on a single workstation server equipped with a GPU. Users view the dashboard provided by the server through a browser on their personal PC via TCP port 3000. This setup is ideal for small-scale PoC projects where multiple users need to access the dashboard simultaneously.



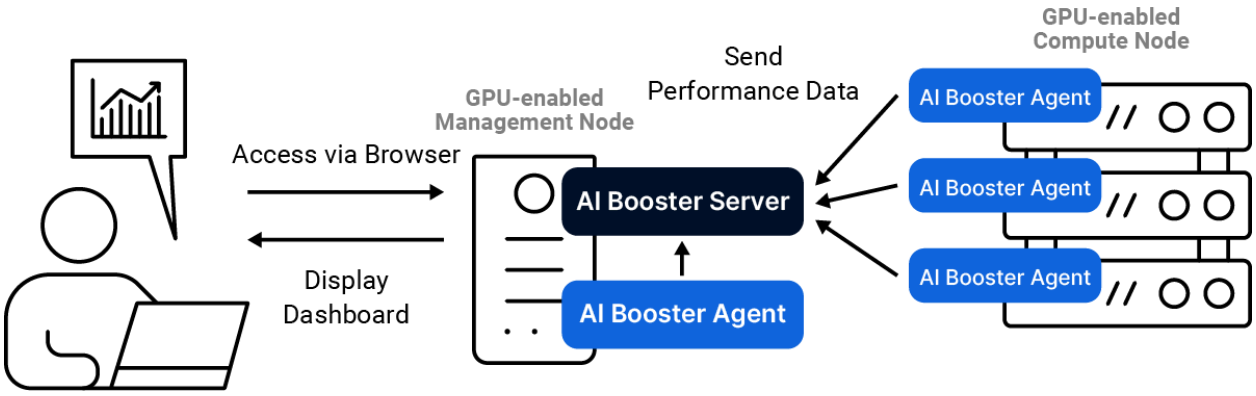
Multi-node Configuration 1 - Production Configuration for GPU Cluster Servers

The AI Booster Server is installed on the management node, and the AI Booster Agent is installed on each GPU compute node. Users access the dashboard provided by the management node from their personal PCs using a browser via TCP port 3000. This is the recommended configuration for most GPU cluster server systems.



Multi-node Configuration 2 - Production Configuration for GPU Cluster Servers

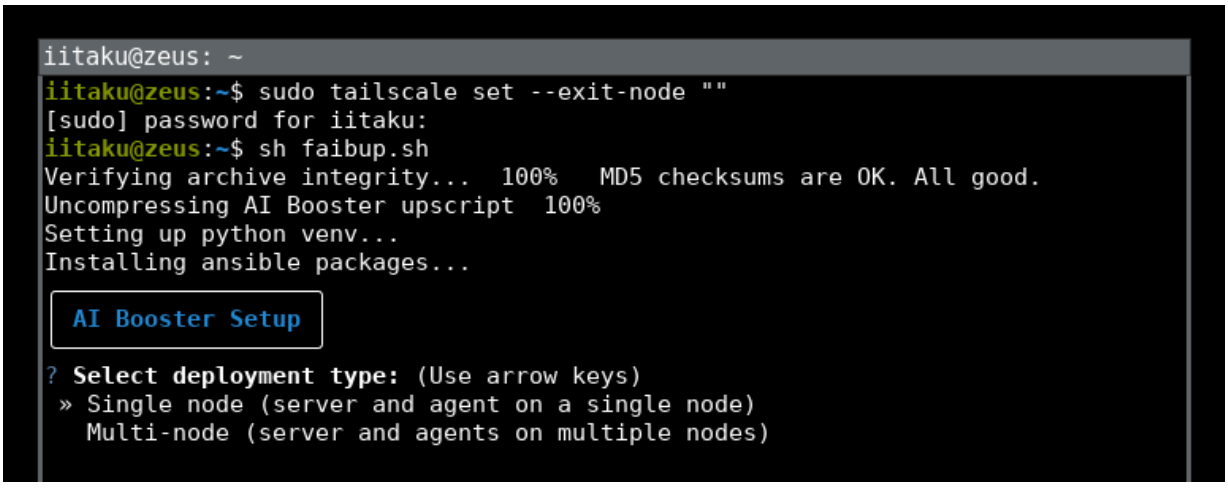
If there is no specific management node, select a node equipped with a GPU and install the AI Booster Server and the AI Booster Agent for that node on it. Install only the Agent on the remaining GPU-equipped nodes. Users can view the dashboard provided by the GPU-equipped node on which the AI Booster Server is installed via a browser on their personal PC, via TCP port 3000.



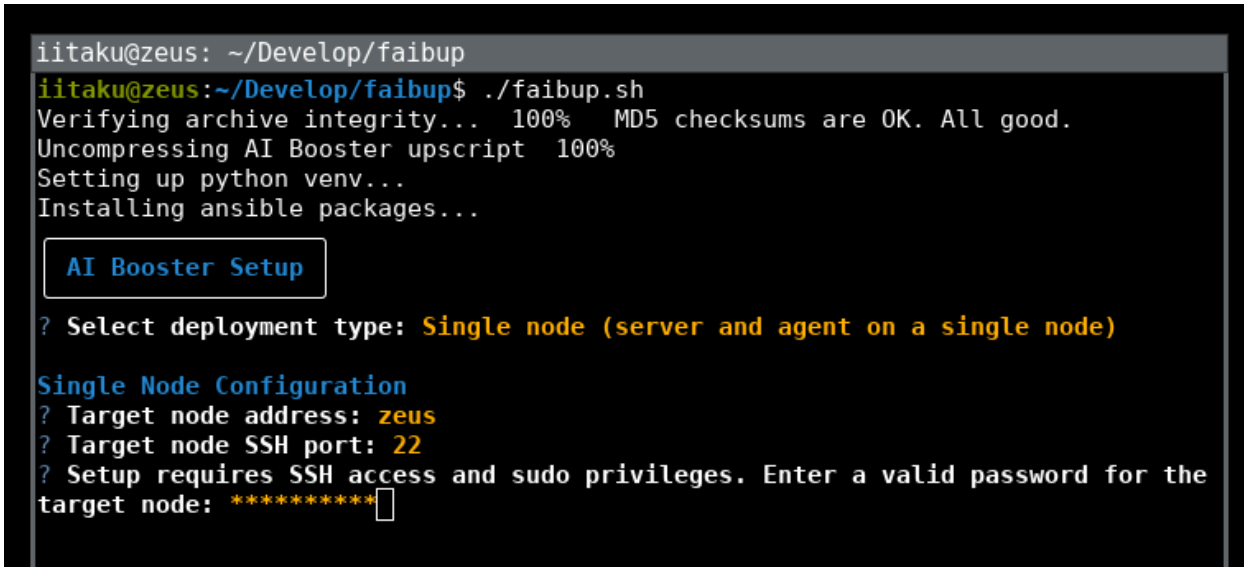
Installation Guide

Run the command provided in the invitation email on one of the Linux nodes where you want to install Fixstars AI Booster.

After launching the setup script, you will first be asked to choose whether to deploy as a single node or multi-node configuration.



Single node configuration



1. **Target node address input**

Target node address: Enter the IP address or resolvable hostname of the target node in.

2. **Target node SSH port input**

Target node SSH port: Enter the SSH port of the target node.

3. **SSH and gaining sudo privileges**

To install various components, you need to connect to the target node via SSH as the current user and obtain administrator privileges. Even if you have set up a public key and do not need a password for SSH access, you need to enter a password to elevate to administrator privileges.

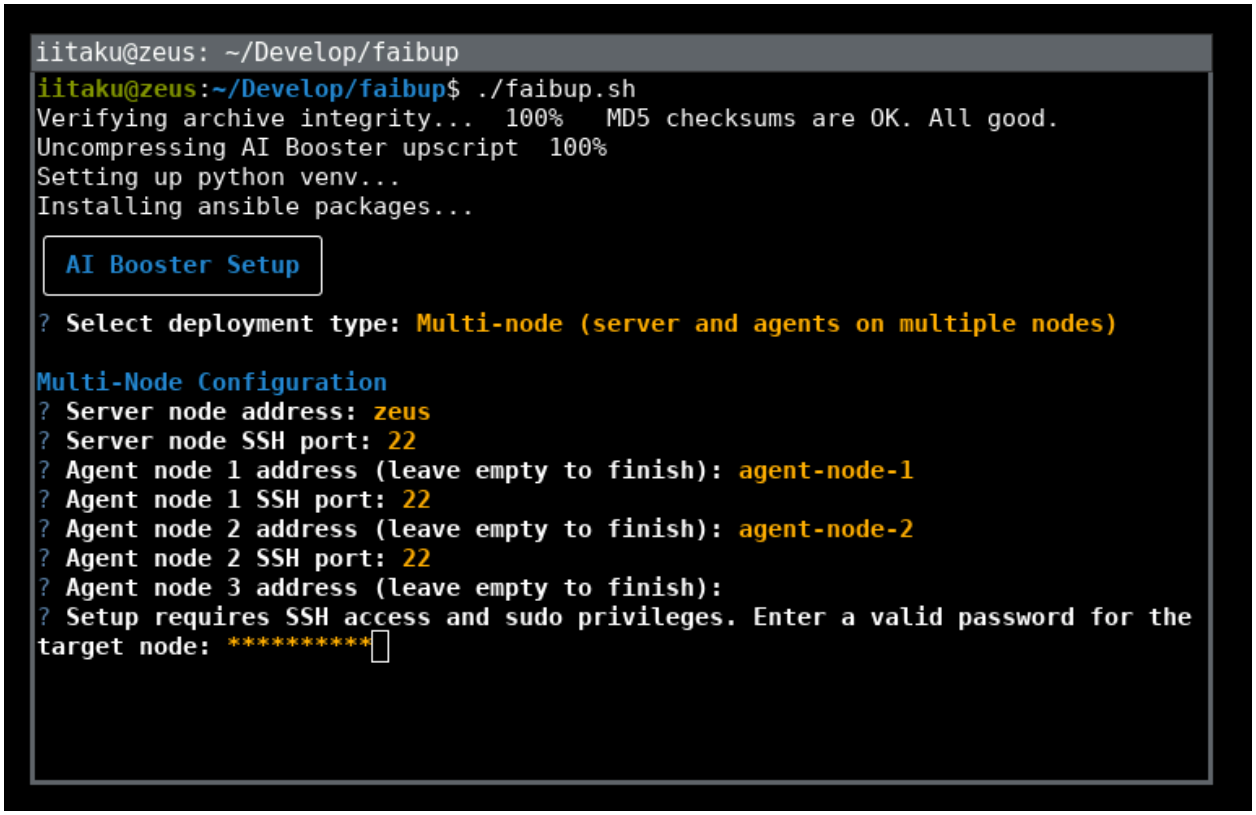
4. **Installation completed**

When the message "AI Booster setup completed successfully!" is displayed, the setup has finished successfully. Open the URL shown at the end in your browser and verify that the dashboard appears. The initial password is set as **admin/admin**.

5. **(If data does not appear)**

Please follow the firewall restriction removal instructions described later.

Multi-node configuration



1. **Server Node address input**

`Server node address:` Enter the IP address or resolvable host name of the node running the AI Booster Server in the field.

2. **Registering the Agent Node**

`Agent node address (leave empty to finish):` will be repeatedly prompted; enter the addresses of each Compute Node on which you want to install the AI Booster Agent, one by one. After entering all the addresses, leave the line blank and press Enter to confirm.

6. **SSH and gaining sudo privileges**

To install various components, you need to connect to the target node via SSH as the current user and obtain administrator privileges. Even if you have set up a public key and do not need a password for SSH access, you need to enter a password to elevate to administrator privileges.

7. **Installation completed**

When the message "AI Booster setup completed successfully!" is displayed, the setup has finished successfully. Open the URL shown at the end in your browser and verify that the dashboard appears. The initial password is set as **admin/admin**.

8. **(If data does not appear)**

Please follow the firewall restriction removal instructions described later.

Firewall restriction removal

1. **Removing restrictions on Server Node**

The Server component uses TCP ports 3000 and 9000 for communication. If firewall restrictions exist, please allow communication on these ports.

For example, using ufw:

```shell
sudo ufw allow 3000

sudo ufw allow 9000
```

2. **Removing restrictions on Agent Nodes and Single Node**

The Agent component uses TCP port 9100 for communication. If firewall restrictions exist, please allow communication on this port.

For example, using ufw:

```shell
sudo ufw allow 9100
```

Uninstallation Guide

Fixstars AI Booster is managed internally with Docker Compose. Run the steps below **on every Linux node** from which you want to remove AI Booster.

1. **Stopping and removing Docker containers**

```shell
# Single-node configuration
docker compose -p "local" down

# Multi-node configuration / Agent node
docker compose -p "agent" down

# Multi-node configuration / Server node
docker compose -p "server" down
```

2. **Removing Docker volumes**

Run the following commands to list the volumes that will be removed, and review the output.

If the list contains any unintended volumes (anything other than grafana-data or clickhouse-data), select those volumes carefully and delete them individually.

```shell
# Single-node configuration
docker volume ls -q --filter name=^local_.+-data$

# Multi-node configuration / Server node
docker volume ls -q --filter name=^server_.+-data$
```

To actually delete the volumes, run the commands below.

Executing these commands deletes all data collected by AI Booster.

```shell
# Single-node configuration
docker volume rm $(docker volume ls -q --filter name=^local_.+-data$)

# Multi-node configuration / Server node
```

```
docker volume rm $(docker volume ls -q --filter name=^server_.+-data$)
```

3. **Deleting the folder that contains the Docker Compose YAML**

Remove the repository, documents, and Docker Compose definition files:

```shell
# All nodes
sudo rm -r /opt/aibooster
```

4. **Removing Docker images**

To completely remove the Docker images downloaded by AI Booster, execute the commands below.

```shell
# Single-node configuration

docker images | grep -E \
    'public.ecr.aws.*aibooster|prom/node-exporter|ghcr.io/intel/pcm|nginx' \
      | awk '{print $3}' | xargs docker rmi

# Multi-node configuration / Agent node

ssh <Agent-node-address> -p <Agent-SSH-port> \
  docker images | grep -E \
    'public.ecr.aws.*aibooster|prom/node-exporter|ghcr.io/intel/pcm' \
      | awk '{print $3}' | xargs docker rmi

# Multi-node configuration / Server node

ssh <Server-node-address> -p <Server-SSH-port> \
  docker images | grep -E \
    'public.ecr.aws.*aibooster|nginx' \
      | awk '{print $3}' | xargs docker rmi
```