



Lenovo Intelligent Computing Orchestration (LiCO)

Product Guide

Lenovo Intelligent Computing Orchestration (LiCO) is a software solution that simplifies the use of clustered computing resources for Artificial Intelligence (AI) model development and training, and HPC workloads. LiCO interfaces with an open-source software orchestration stack, enabling the convergence of AI onto an HPC cluster.

The unified platform simplifies interaction with the underlying compute resources, enabling customers to take advantage of popular open-source cluster tools while reducing the effort and complexity of using it for HPC and AI.

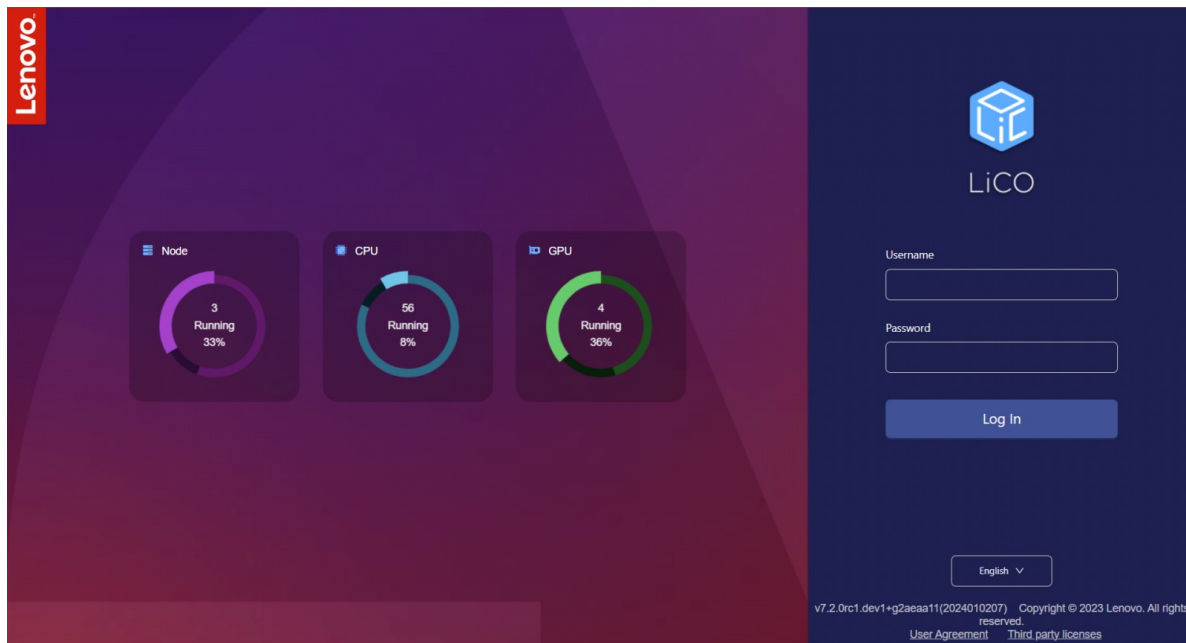


Figure 1. LiCO 7.2.1 login

Did You Know?

LiCO enables a single cluster to be used for multiple AI workloads simultaneously, with multiple users accessing the available cluster resources at the same time. Running more workloads can increase utilization of cluster resources, driving more user productivity and value from the environment.

What's new in LiCO 8.0.0

Lenovo recently announced LiCO Version 8.0.0 with the following updates:

- Support Lenovo ThinkSystem SR680a V3 and SR685a V3 servers
- Support NVIDIA H200 HGX 8 GPU board, H200 HGX 4-GPU board
- Support Rocky/RHEL v9.4
- Support OpenHPC v3.1
- Support Virtual GPUs (NVIDIA AI Enterprise)
- Support managing multi-cluster with same schedulers (only for SLURM)
- Optimize Monitoring performance (Replace icinga2 with prometheus)
- UI Improvements
- New job template
- DeepSpeed (only for SLURM)
- Support EAR 5.0

Part numbers

The following table lists the ordering information for LiCO.

Table 1. LiCO HPC/AI version ordering information

Description	LFO	Software CTO	Feature code
Lenovo HPC AI LiCO Software 90 Day Evaluation License	7S090004WW	7S09CTO2WW	B1YC
Lenovo HPC AI LiCO Webportal w/1 yr S&S	7S09002BWW	7S09CTO6WW	S93A
Lenovo HPC AI LiCO Webportal w/3 yr S&S	7S09002CWW	7S09CTO6WW	S93B
Lenovo HPC AI LiCO Webportal w/5 yr S&S	7S09002DWW	7S09CTO6WW	S93C
Lenovo HPC AI LiCO customization Service per man day*	7S09004HWW	7S09CTO6WW	SBF2

*LiCO customization service

Lenovo Intelligent Computing Orchestration (LiCO) customization services enable customers to request customized features tailored for their own needs. The service is evaluated and quoted in the form of man-days, based on the actual order list.

HPC solution sellers need to provide pre-sales support, collaborate with HPC architects, and communicate with LiCO's R&D team (Ding Hong dinghong1@lenovo.com) for workload evaluations. A quote is provided by Lenovo based on the output SOW and analysis of the workload. After the sellers place an order, they will email the R&D team to request implementation. The LiCO R&D team will deliver the work based on the order content and SOW.

Features for LiCO users

Topics in this section:

- [Benefits to users](#)
- [Features for users](#)
- [Lenovo Accelerated AI](#)
- [Cloud Tools](#)
- [Workflow](#)
- [Admin](#)

Benefits to users

LiCO provides users the following benefits:

- A web-based portal to deploy, monitor and manage AI development and training jobs on a distributed cluster
- Container-based deployment of supported AI frameworks for easy software stack configuration
- Direct browser access to Jupyter Notebook instances running on the cluster
- Standard and customized job templates to provide an intuitive starting point for less experienced users
- Lenovo Accelerated AI pre-defined training and inference templates for many common AI use cases
- Lenovo end-to-end workflow for Image Classification, Object Detection, Instance Segmentation, Image GAN, Text Classification, Seq2seq and Memory Network
- Workflow to define multiple job submissions as an automated workflow to deploy in a single action
- TensorBoard visualization tools integrated into the interface (TensorFlow-based)
- Management of private space on shared storage through the GUI
- Monitoring of job progress and log access

Features for users

Those designated as LiCO users have access to dashboards related primarily to HPC and AI development and training tasks. Users can submit jobs to the cluster, and monitor their results through the dashboards. The following menus are available to users:

- **Home menu for users** – provides an overview of the resources available in the cluster. Jobs and job status are also given, indicating the runtime for the current job, and the order of jobs deployed. Users may click on jobs to access the associated logs and job files. The figure below displays the home menu.

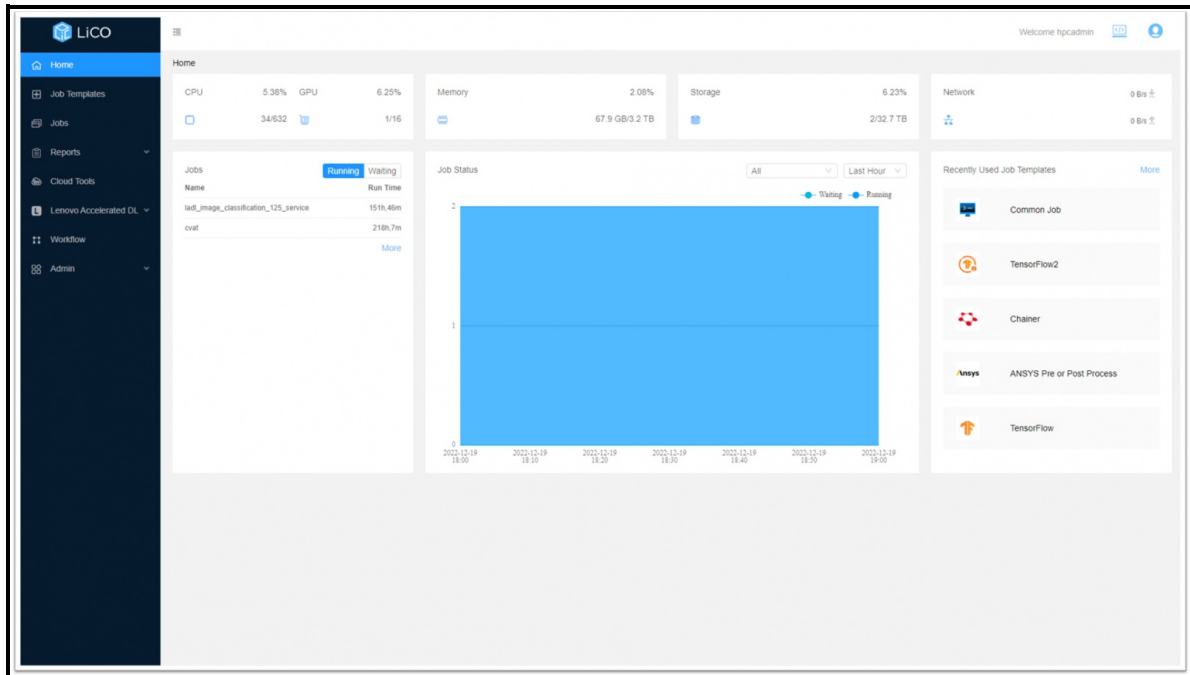


Figure 2. User Home Menu

- **Job Templates** – allows users to set up a job and submit it to the cluster. The user first picks a job template. After selecting the template, the user gives the job a name and inputs the relevant parameters, chooses the resources to be requested on the cluster and submits it. Users can take advantage of Lenovo Accelerated AI templates, industry-standard AI templates, submit generic jobs via the Common Job template, as well as create their own templates requesting specified parameters.

Job Templates available in LICO:

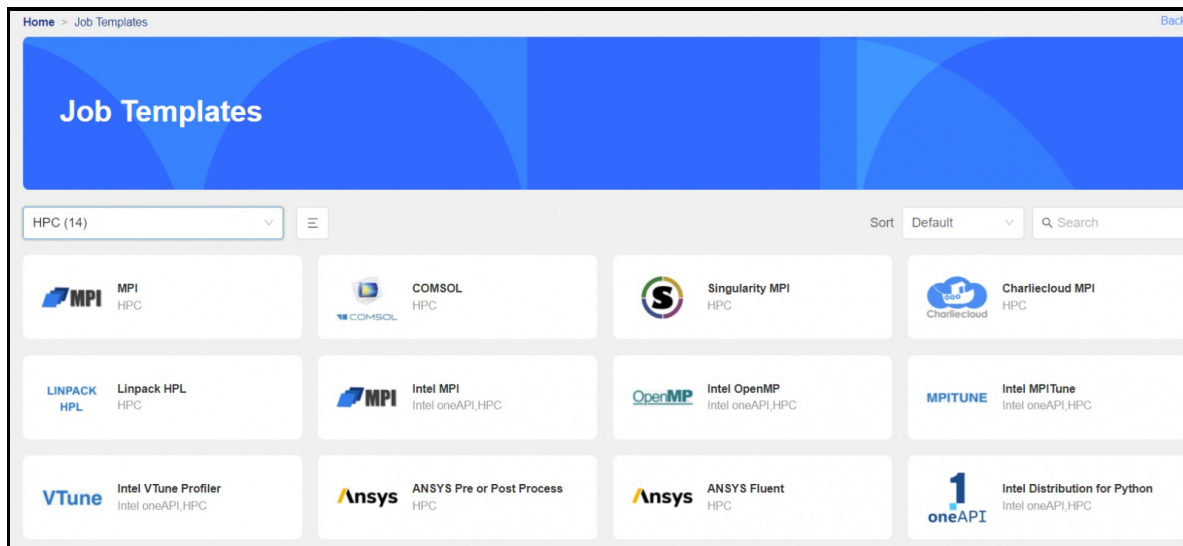


Figure 3. HPC job templates available in LiCO

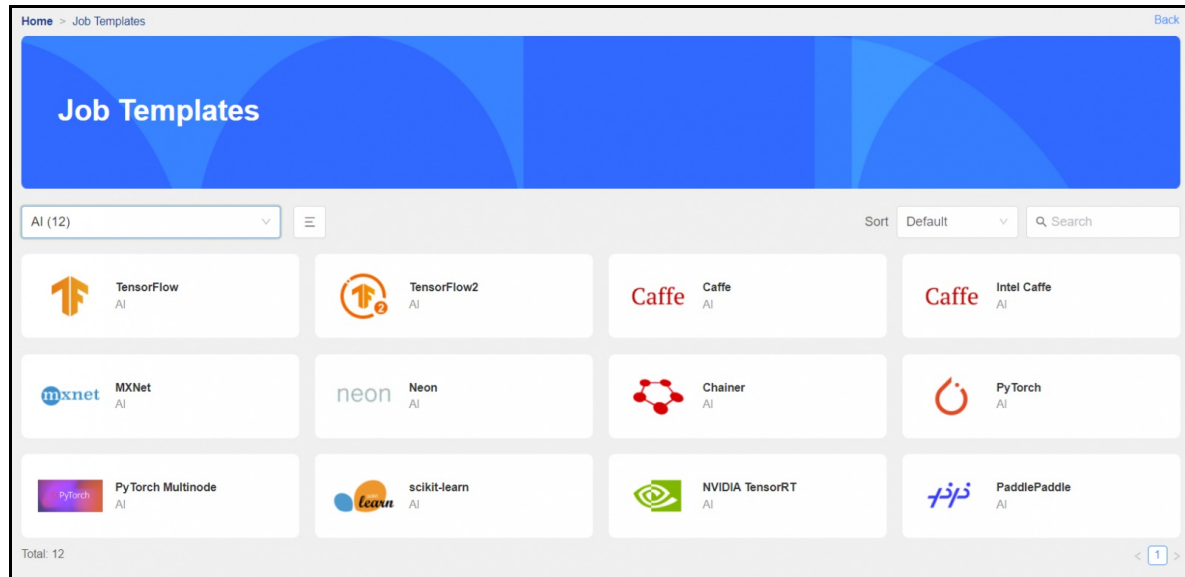


Figure 4. AI job templates available in LiCO

The figure below displays a job template for training with TensorFlow on a single node.

The screenshot shows the configuration page for the 'TensorFlow Single Node' job template. The page has a header with 'Home > Submit Job > TensorFlow Single Node' and a 'Back' link. Below the header, there's a title 'TensorFlow Single Node' and a brief description: 'TensorFlow is an open source software library for numerical computation using data flow graphs commonly used for deep learning. Use this job template to submit a single node CPU&GPU job.' The main content area is divided into sections: 'Template Information' with fields for 'Job Name' and 'Workspace' (with a 'Browse' button); 'Template Parameters' with fields for 'Runtime' (set to 'OS Default'), 'Container image' (set to 'tensorflow-cpu'), 'Program (py or sh)' (with a 'Browse' button), and 'Program args'; 'Resource Options'; and 'Notification Options'. A 'Submit' button is at the bottom left. On the right side, there's a sidebar with links: 'Template Information', 'Template Parameters', 'Resource Options', and 'Notification Options'.

Figure 5. AI Job Template

LiCO also provides TensorBoard monitoring when running certain TensorFlow workloads, as shown in the following figure.

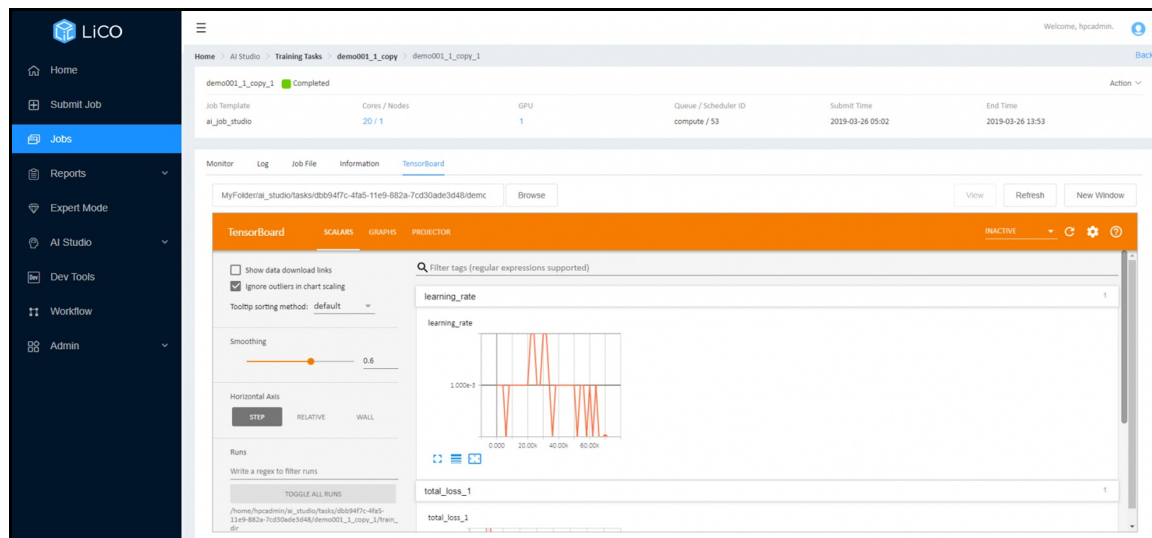


Figure 6. LiCO and TensorBoard monitoring

- **Jobs menu** – displays a dashboard listing jobs and their statuses. In addition, users can select the job and see results and logs pertaining to the job in progress (or after completion). Tags and comments can be added to completed jobs for easier filtering.
- **Reports** – displays a dashboard for obtaining reports on expenses. Expense Reports is supported currently, where the job and storage billing statistics are displayed.
- **Cloud Tools menu** – enables users to create, run and view Jupyter Notebook and Jupyter Lab instances on the cluster from LiCO for model experimentation and development. Users will be able to launch a CVAT labelling environment, Tiger VNC and the RStudio development environment. See the section for more information.
- **Lenovo Accelerated DL** – provides users with the ability to label data, optimize hyperparameters, as well as test and publish trained models from within an end-to-end workflow in LiCO. LiCO supports Text Classification, Image Classification, Object Detection, and Instance Segmentation workflows. See the section for more information.
- **Workflow menu** – allows users to create multi-step jobs that execute as a single action. Workflows can contain serially-executed steps as well as multiple jobs to execute in parallel within a step to take full advantage of cluster resources. See the section for more information.
- **Admin menu** – allows users to access a number of capabilities not directly associated with deploying workloads to the cluster, including access to shared storage space on the cluster through a drag-and-drop interface and access to provision API and git interfaces. See the section for more information.

Lenovo Accelerated AI

Lenovo Accelerated AI provides a set of templates that aim to make AI training and inference simpler, more accessible, and faster to implement. The Accelerated AI templates differ from the other templates in LiCO in that they do not require the user to input a program; rather, they simply require a workspace (with associated directories) and a labelled dataset.

Lenovo Accelerated DL is based on the LeTrain project. LeTrain is a distributed training engine based on TensorFlow and optimized by Lenovo. Its goal is to make distributed training as easy as single GPU training and achieve linear scaling performance.

Lenovo Accelerated DL provides an end-to-end workflow for Text Classification, Image Classification, Object Detection, and Instance Segmentation, with training based on Lenovo Accelerated AI pre-defined models. A user can import an unprocessed, unlabeled data set of images, label them, train multiple instances with a grid of parameter values, test the output models for validation, and publish to a git repository for use in an application environment. Additionally, users can initiate the workflow steps from a REST API call to take advantage of LiCO as part of a DevOps toolchain.

Following is the workflow illustrating the main features of Lenovo Accelerated DL:

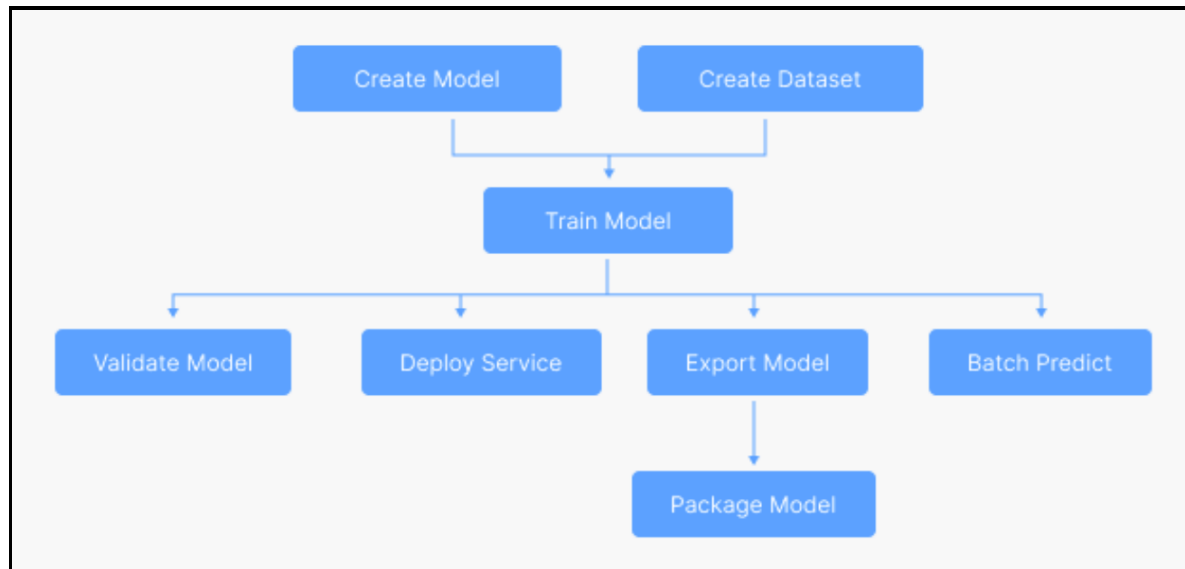


Figure 7. Lenovo Accelerated DL main features workflow

The following use cases are supported with Lenovo Accelerated AI templates:

- Image Classification
- Object Detection
- Instance Segmentation
- Medical Image Segmentation
- Seq2Seq
- Memory Network
- Image GAN
- Text Classification

The following figure displays the Lenovo Accelerated AI templates.

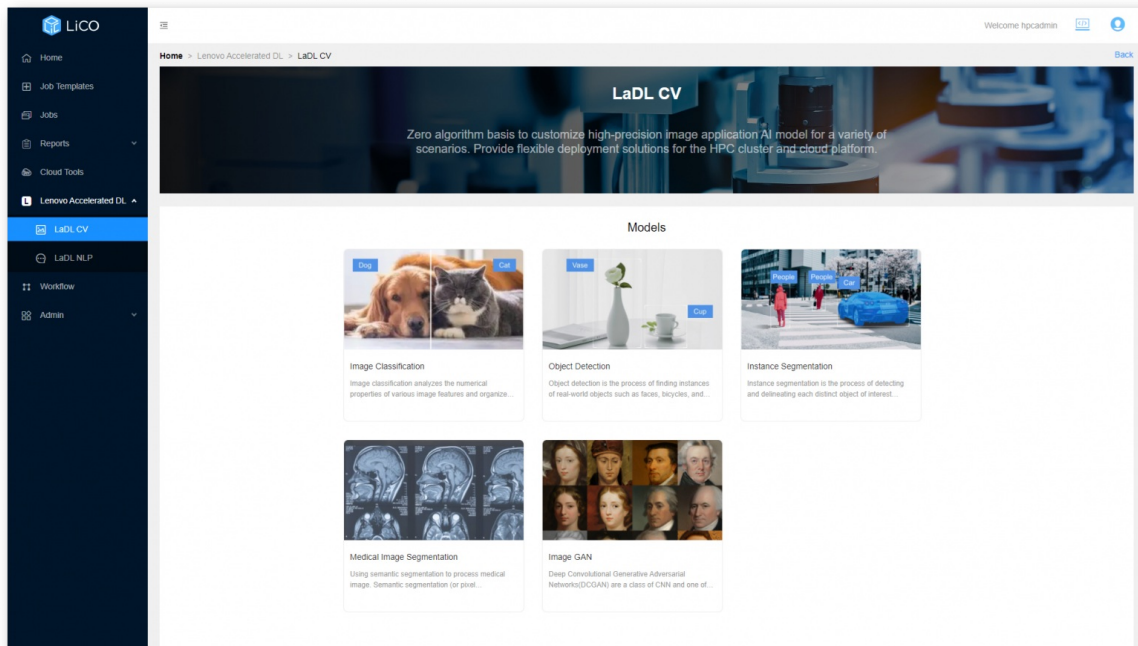


Figure 8. Lenovo Accelerated Deep Learning (DL) computer vision (CV) templates

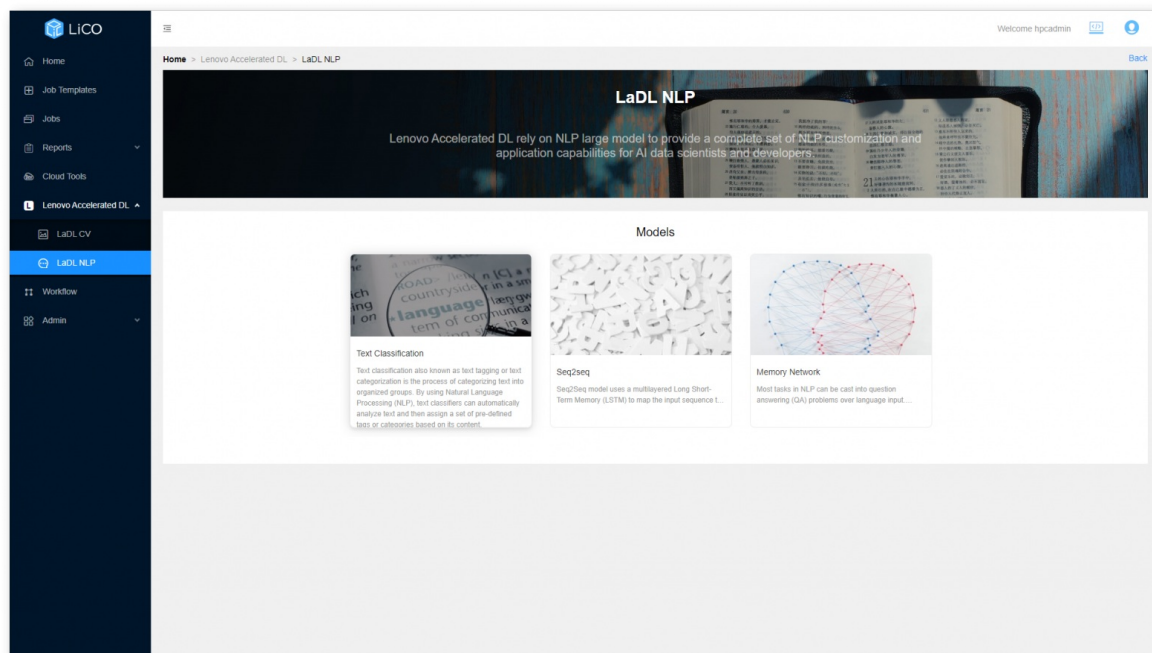


Figure 9. Lenovo Accelerated Deep Learning (DL) natural language processing (NLP) templates

Each Lenovo Accelerated AI use-case is supported by both a training and inference template. The training templates provide parameter inputs such as batch size and learning rate. These parameter fields are pre-populated with default values, but are tunable by those with data science knowledge. The templates also provide visual analytics with TensorBoard; the TensorBoard graphs continually update in-flight as the job runs, and the final statistics are available after the job has completed.

In LiCO the Image Classification and Object Detection templates include the ability to select a topology based on the characteristics of a target inference device, such as an IoT Device, Edge Server, or Data Center server.

The following figure displays the embedded TensorBoard interface for a job. TensorBoard provides visualizations for TensorFlow jobs running in LiCO, whether through Lenovo Accelerated AI templates or the standard TensorFlow AI templates.

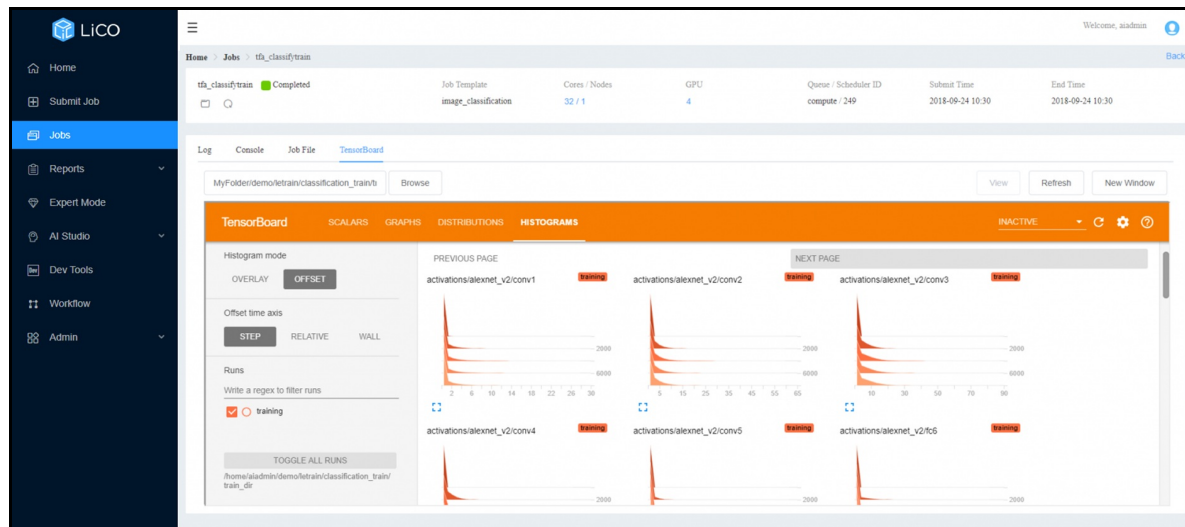


Figure 10: TensorBoard in LiCO

LiCO also provides inference templates which allow users to predict with new data based on models that have been trained with Lenovo Accelerated AI templates. For the inference templates, users only need to provide a workspace, an input directory (the location of the data on which inference will be performed), an output directory, and the location of the trained model. The job will run, and upon completion, the output directory will contain the analyzed data. For visual templates such as Object Detection, images can be previewed directly from within LiCO's Manage Files interface.

The following two figures display an input file to the Object Detection inference template, as well as the corresponding output.



Figure 11: JPG file containing image of cat for input into inference job

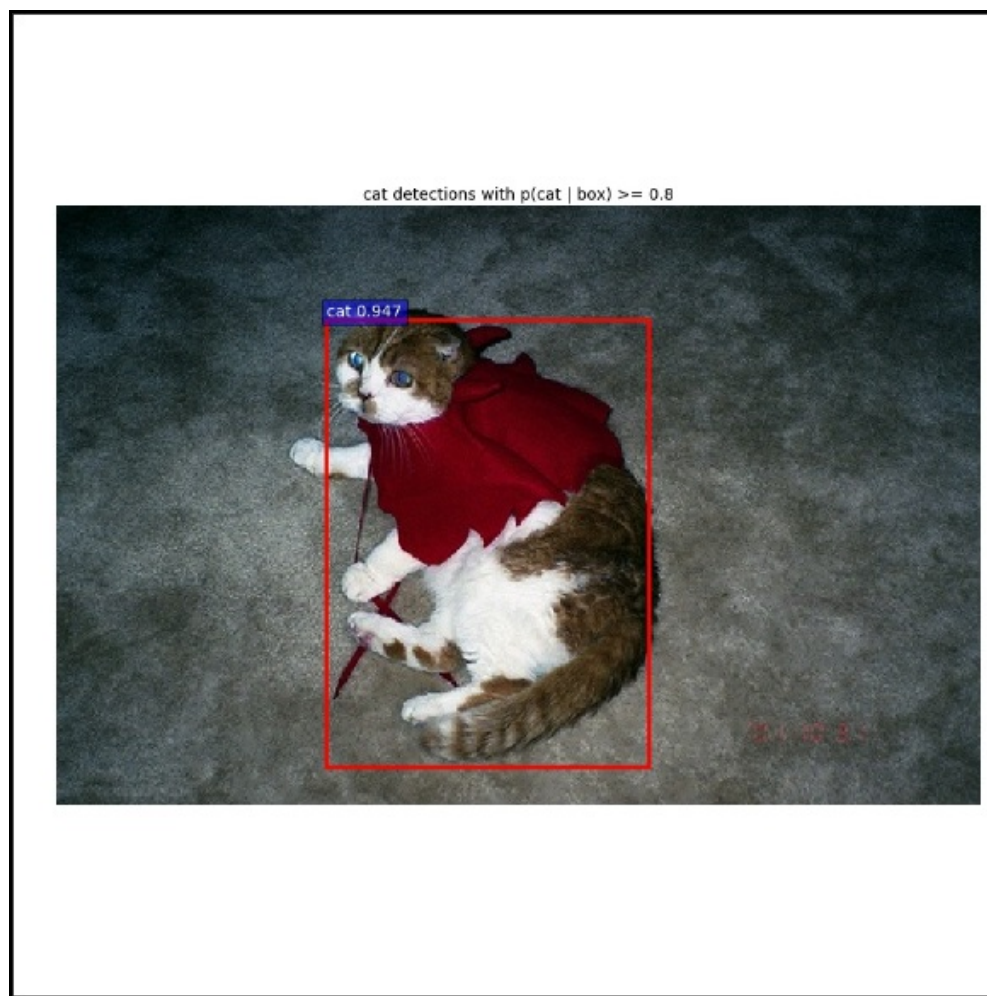


Figure 12: LiCO output displaying the section of the JPG containing the cat image

Cloud Tools

LiCO includes the capability to create and deploy instances of Jupyter, RStudio Server and TigerVNC on the cluster. Users may create multiple instances, to customize for different software environments and projects. At the launch of an instance, the user can define the amount of compute resource requirements needed (CPU and GPU) to better optimize the performance of the task and optimize resource usage on the cluster.

Once a Jupyter, TigerVNC or an RStudio Server instance is created, the user can deploy it to the cluster and use the environment directly from their browser in a new tab. The user can leverage the interface directly to upload, download and run code as they normally would, utilizing the shared storage space used for LiCO.

Note: RStudio Server does not support the Chinese version.

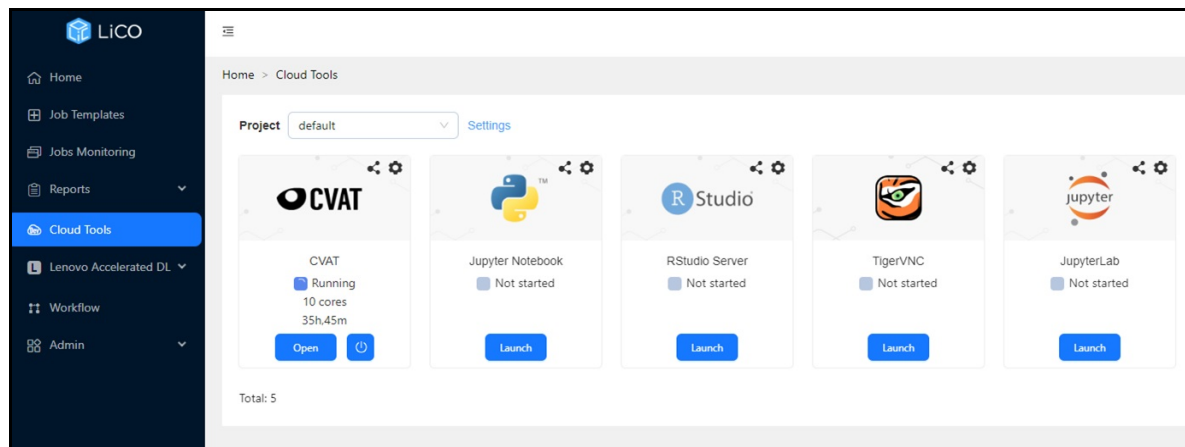


Figure 13. Cloud Tools menu

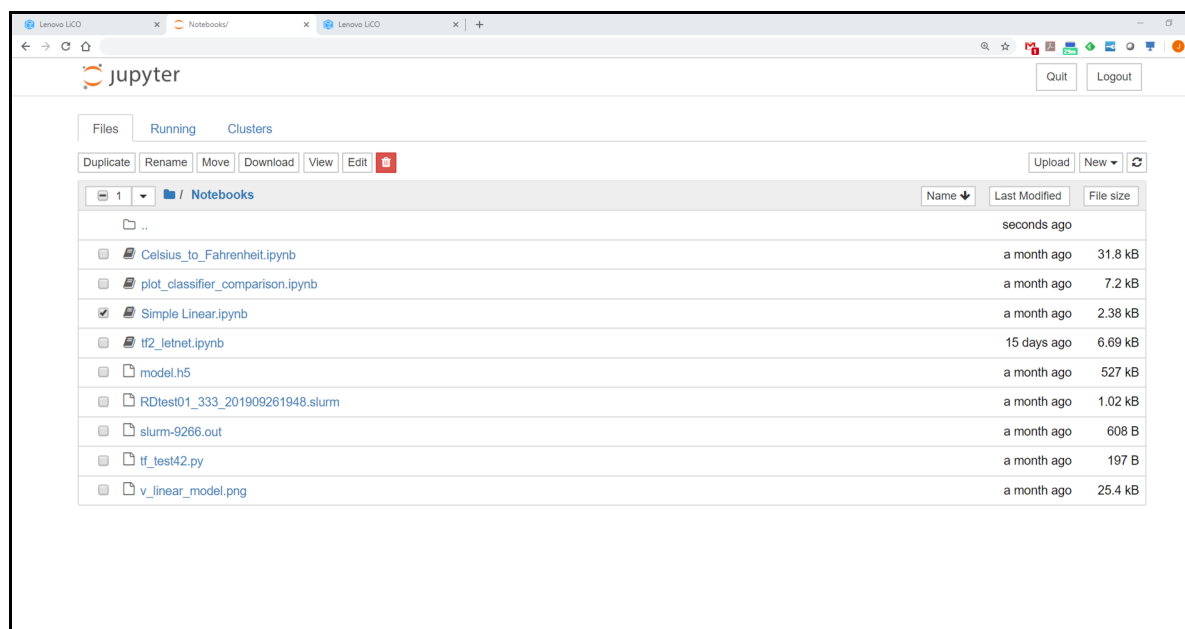


Figure 14. Jupyter instance accessible in new browser tab

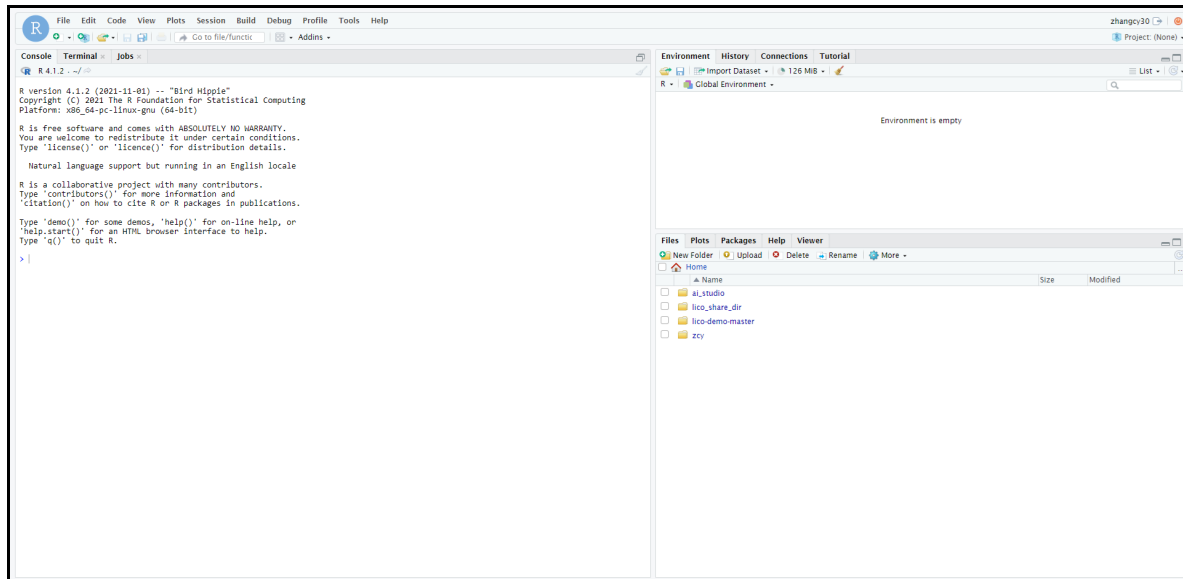


Figure 15. Integrated RStudio Server environment

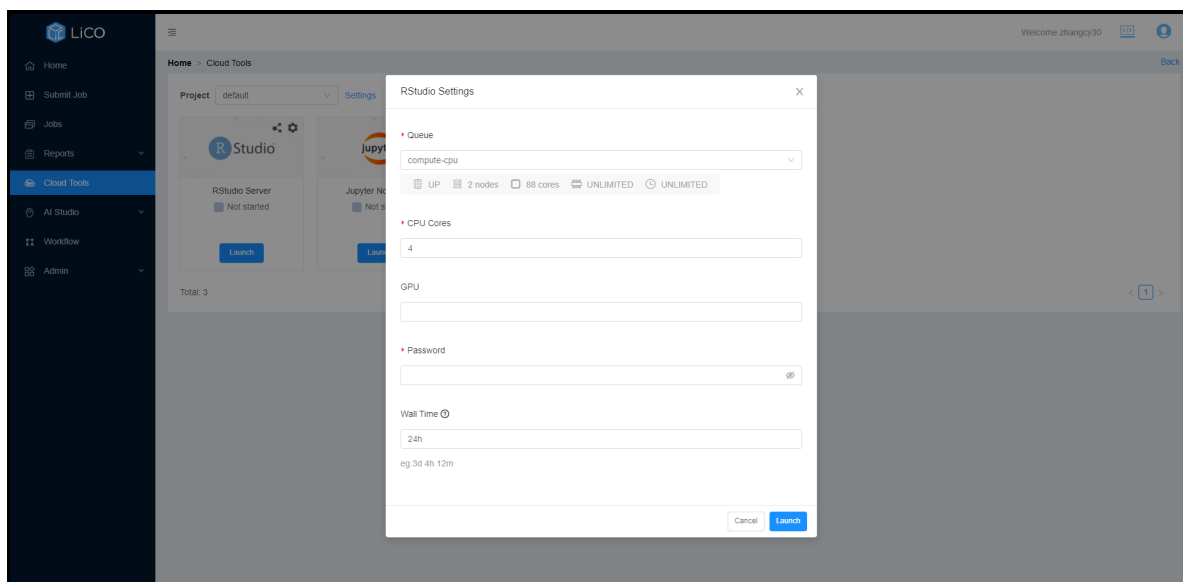


Figure 16. Settings definition for an RStudio Server instance

LiCO includes the capability to launch a CVAT labelling environment for image annotation. Users may create and edit multiple CVAT instances for different projects, login to the CVAT web panel to label images and export the labelling as a dataset. The dataset which is created through CVAT can be managed through the dataset management page in LiCO.

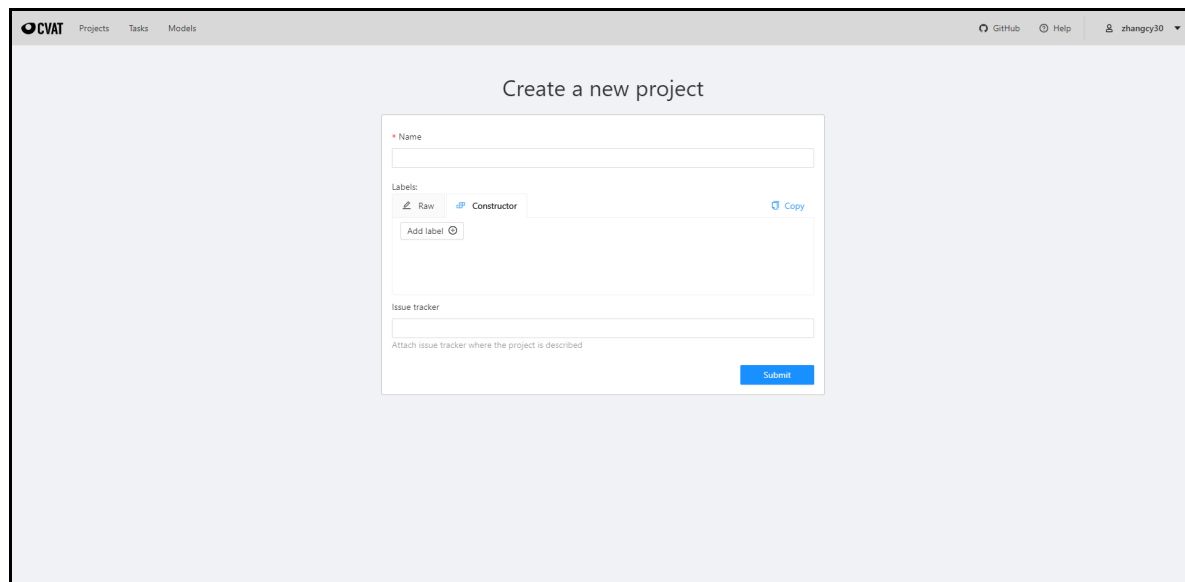


Figure 17. CVAT instance accessible in a new browser tab

Opened instances of RStudio Server, Jupyter Notebook, Jupyter Lab, TigerVNC and CVAT can be shared using the online platform URL of that instance.

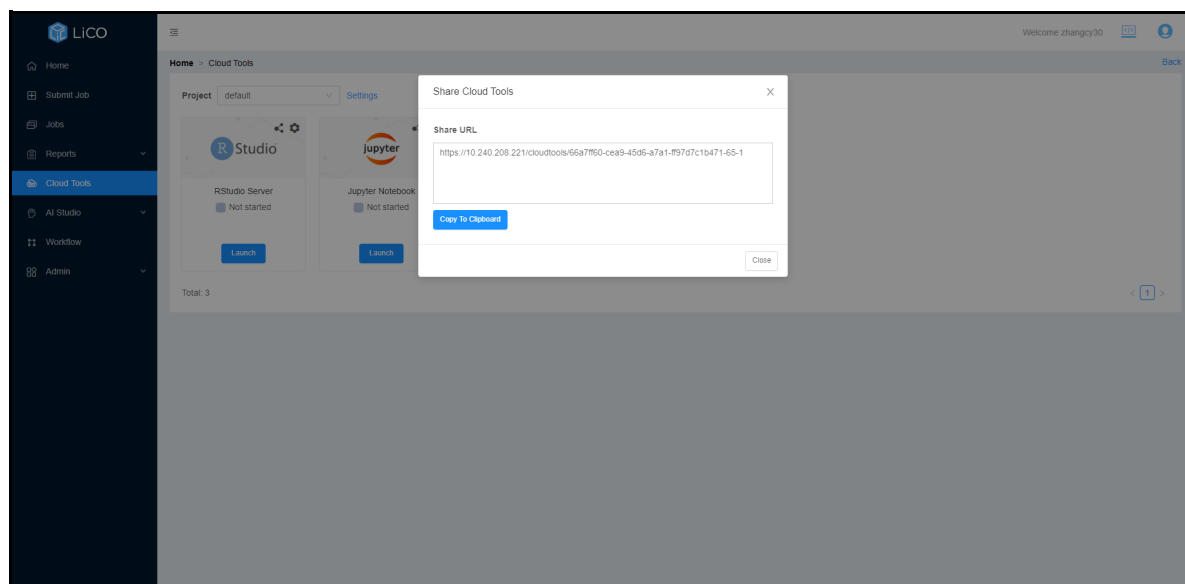


Figure 18. User can share the Cloud Tools with other or non-hpc users

Workflow

LiCO provides the ability to define multiple job submissions into a single execution action, called a Workflow. Steps are created to execute job submissions in serial, and within each step multiple job submissions may be executed in parallel. Workflow uses LiCO job submission templates to define the jobs for each step, and any template available including custom templates can be used in a workflow.

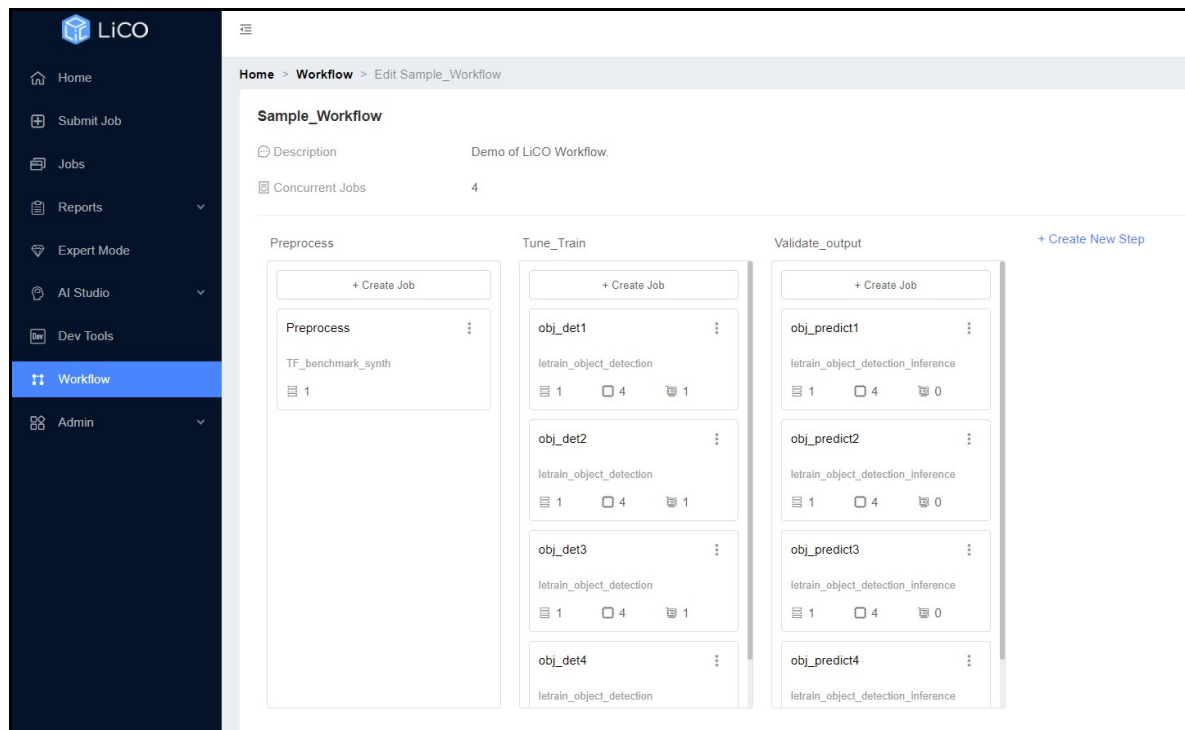


Figure 19. Defining a workflow in LiCO

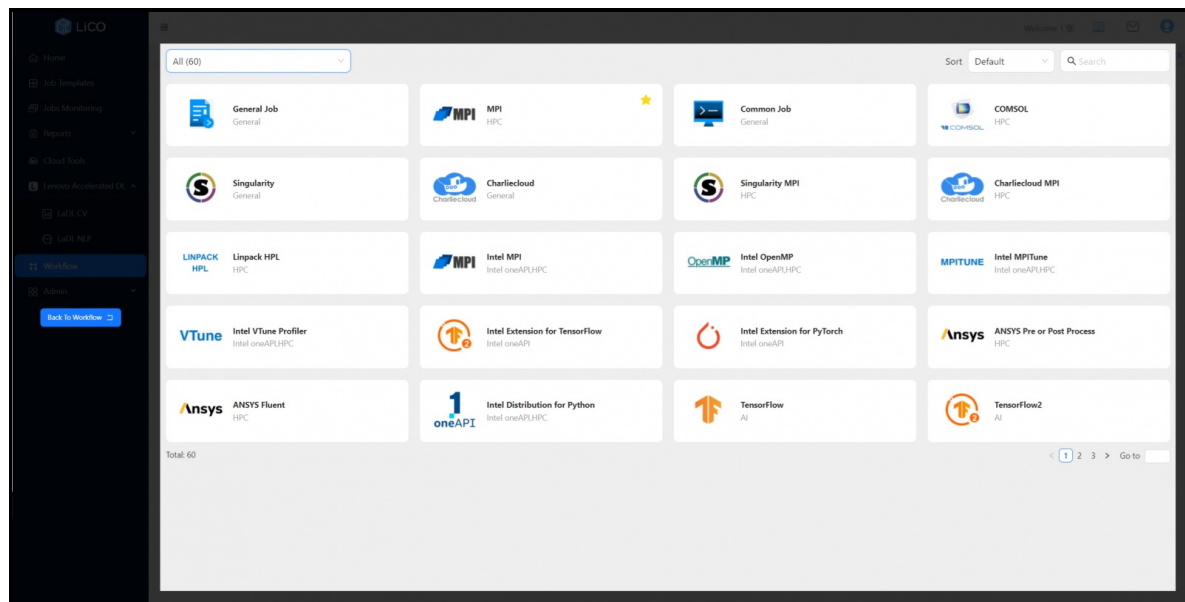


Figure 20. Adding a template to a workflow in LiCO

LiCO workflows allow users to automate the deployment of multiple jobs that may be required for a project, so the user can execute and monitor as a single action. Workflows can be easily copied and edited, allowing users to quickly customize existing workflows for multiple projects.

Admin

The Admin tab for the user provides access to container and VNC management.

The Admin tab also enables users to publish a trained model to a git repository or as a docker container image.

LiCO can bill users for jobs and storage instances. Users can download their daily and monthly bills generated automatically on the system.

Some open application programming interfaces (APIs) are available in the API key sub-tab of the Admin tab.

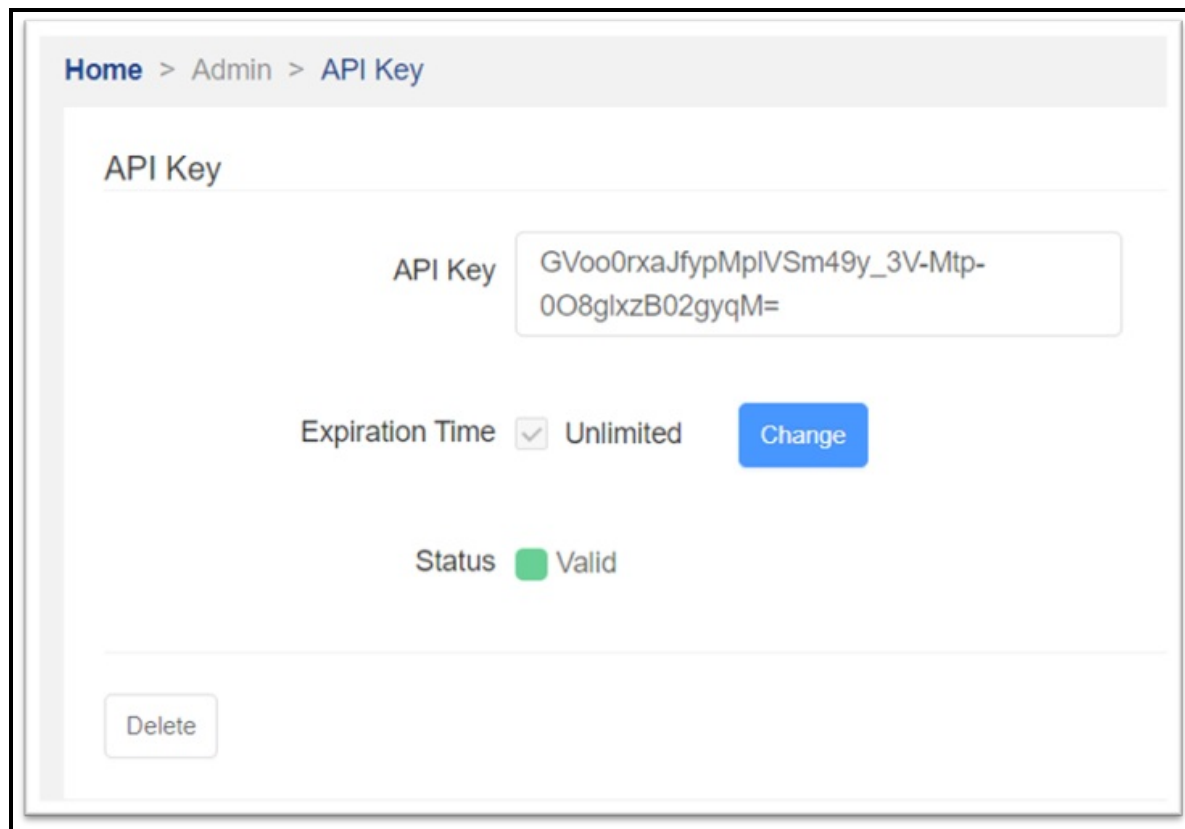


Figure 21. API key page

Additional features for LiCO HPC/AI Users

In addition to the user features above, the LiCO HPC/AI version contains a number of features to simplify HPC workload deployment with a minimal learning curve for users vs. console-based scripting and execution. HPC users can submit jobs easily through standard or custom templates, utilize containers, pre-define runtime modules and environment variables for submission, and since LiCO 6.3 take advantage of advanced features such as Energy Aware Runtime and Intel oneAPI tools and optimizations all from within the LiCO interface.

Topics in this section:

- [Energy Aware Runtime](#)
- [Intel oneAPI](#)
- [HPC Runtime Module Management](#)
- [Container-based HPC workload deployment](#)
- [Singularity Container Image Management](#)
- [Reports](#)
- [System tools](#)

Energy Aware Runtime

Energy Aware Runtime (EAR) is software technology designed to provide a solution for running MPI applications with higher energy efficiency. Developed in collaboration with Barcelona Supercomputing Center as part of the BSC-Lenovo Cooperation project, EAR is supported for use with the SLURM scheduler through a SPANK plugin. LiCO exposes EAR deployment options within the standard MPI template, allowing users to take advantage of the capability for MPI workloads.

Once the workload has been profiled through a learning phase, EAR will minimize CPU frequency to reduce energy consumption while maintaining a set threshold of performance. This is particularly helpful where MPI applications may not take significant advantage of higher clock frequencies, so the frequency can be reduced to save energy while maintaining expected performance.

Users can select EAR options at job submission in the standard MPI template, either to run the default set by the administrator, minimum time to solution, or minimum energy. Administrators can set the policies and thresholds for EAR usage within the LiCO Administrator portal, as well as which users are authorized to use EAR.

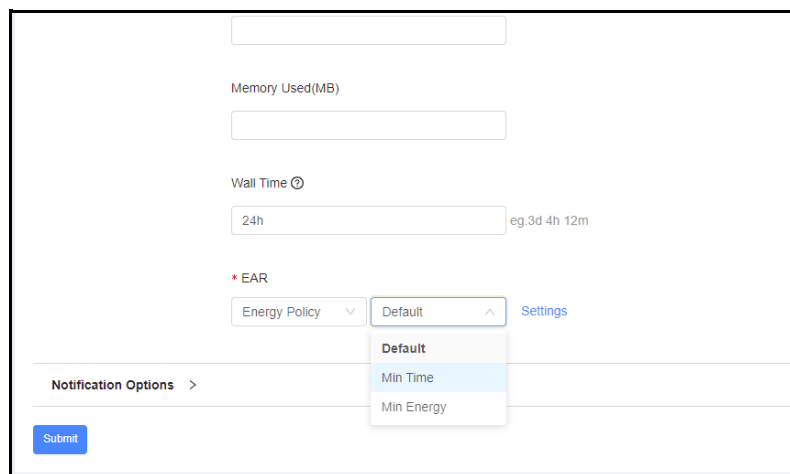
The image shows a web form for submitting an MPI job. It includes input fields for 'Memory Used(MB)' and 'Wall Time' (with a hint 'eg. 3d 4h 12m'). Below these, there is a section for 'EAR' (Energy Aware Runtime) with a dropdown menu for 'Energy Policy'. The dropdown is open, showing three options: 'Default' (selected), 'Min Time', and 'Min Energy'. To the right of the dropdown is a 'Settings' link. At the bottom left, there is a 'Notification Options' link and a 'Submit' button.

Figure 22. Selection of Energy Policy in MPI template

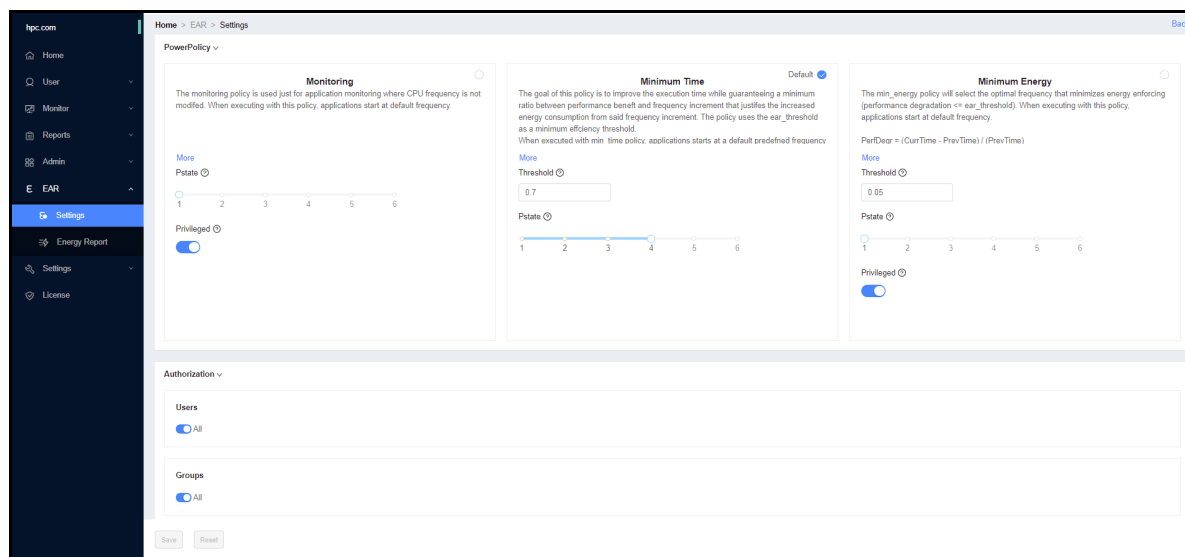


Figure 23. Administrator portal EAR power policy management

The software technology for EAR is supported separately by Energy Aware Solutions S.L. For more information see <https://www.eas4dc.com>.

Intel oneAPI

Intel oneAPI is an open-source and standard programming model designed for all industries, which provides the uniform service for the developers of CPU, GPU, and FPGA accelerators. Based on industrial standard and existing programming model of developers, oneAPI open standard can be widely used in varied structures and hardware from different suppliers. The use of Intel oneAPI improves the performance of MPI, OpenMP, TensorFlow, Pytorch, and other programs.

LiCO features templates base on Intel oneAPI – that are optimized to run on Intel processors – developed, tested and validated in collaboration with Intel.

Note: This function is unavailable if Intel oneAPI is not installed.

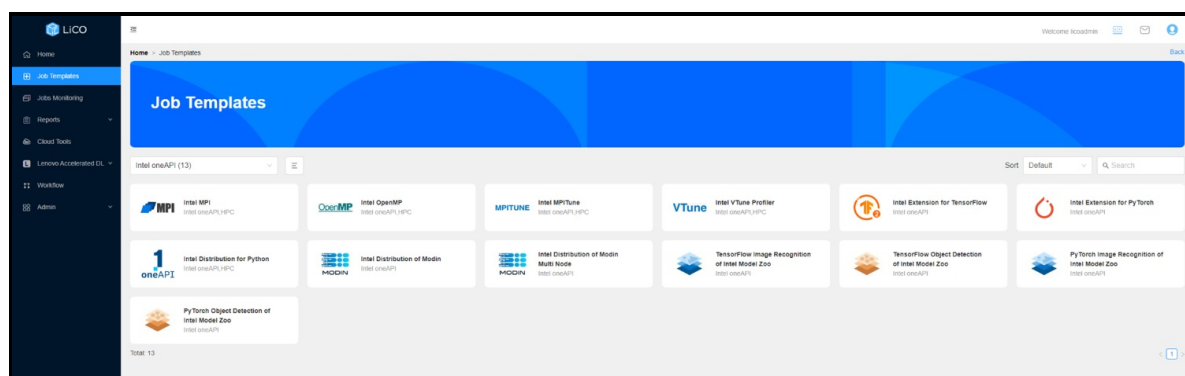


Figure 24. LiCO Templates for leveraging Intel oneAPI technology

Intel Neural Compressor

Intel Neural Compressor performs model compression to reduce the model size and increase the speed of deep learning inference for deployment on CPUs or GPUs. This open-source Python library automates popular model compression technologies, such as quantization, pruning, and knowledge distillation across multiple deep learning frameworks.

The Python library is integrated in LiCO and used for exporting an image classification model. Intel Neural Compressor FP32, BF16 and INT8 models can be selected.

Figure 25. Export an Image Classification – Intel Neural Compressor job

Intel MPI

Intel MPI Library is a multifabric message-passing library that implements the open-source MPICH specification. Use the library to create, maintain, and test advanced, complex applications that perform better on high-performance computing (HPC) clusters based on Intel processors.

Intel OpenMP

Using the OpenMP pragmas requires an OpenMP-compatible compiler and thread-safe libraries. A perfect choice is the Intel C++ Compiler version 7.0 or newer. (The Intel Fortran compiler also supports OpenMP.) Adding the following command-line option to the compiler instructs it to pay attention to the OpenMP pragmas and to insert threads.

Intel MPITune

The MPITune utility allows users to automatically adjust Intel MPI Library parameters, such as collective operation algorithms, to their cluster configuration or application. The tuner iteratively launches a benchmarking application with different configurations to measure performance and stores the results of each launch. Based on these results, the tuner generates optimal values for the parameters being tuned.

Intel VTune Profiler

Intel VTune Profiler optimizes application performance, system performance, and system configuration for HPC, cloud, IoT, media, storage, and more. Intel VTune Profiler, provided with Intel snapshot performance analyzer, enables users to analyze the serial and multi-threaded applications in hardware platforms (CPU, GPU, FPGA), and analyze the local and remote targets.

In LiCO, users can submit an Intel VTune Profiler job and administrators can perform a platform analysis.

For an Intel MPI, Intel OpenMP and Intel Distribution for Python jobs users can select the VTune Analysis Type.

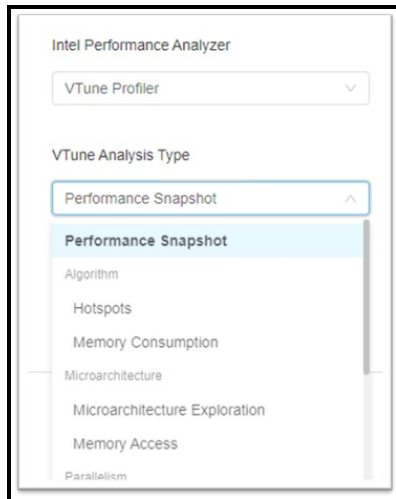


Figure 26. Intel VTune Profiler integration

Intel Distribution for GDB

The Intel Distribution for GDB application debugger is a companion tool to Intel compilers and libraries. It delivers a unified debugging experience that allows users to efficiently and simultaneously debug cross-platform parallel and threaded applications developed in C, C++, SYCL, OpenMP, or Fortran.

When submitting an Intel MPI or OpenMP job in LiCO, while setting the Template Parameters, users can set the Remotely Debug option to Intel Distribution for GDB. With this setting the running program can be debugged.

Intel Extension for TensorFlow

Intel® Extension for TensorFlow* is a heterogeneous, high performance deep learning extension plugin based on TensorFlow [PluggableDevice](#) interface, aiming to bring Intel CPU or GPU devices into [TensorFlow](#) open source community for AI workload acceleration. It allows users to flexibly plug an XPU into TensorFlow on-demand, exposing the computing power inside Intel's hardware.

The Intel Extension for TensorFlow job template in LiCO supports running programs on one or more nodes using CPUs or Intel GPUs. For distributed training, using CPU for training currently only supports PS Worker distributed architecture, and using Intel GPU for training currently only supports Horovod distributed training framework.

Intel Extension for PyTorch

Intel® Extension for PyTorch* extends PyTorch* with up-to-date features optimizations for an extra performance boost on Intel hardware. Optimizations take advantage of AVX-512 Vector Neural Network Instructions (AVX512 VNNI) and Intel® Advanced Matrix Extensions (Intel® AMX) on Intel CPUs as well as Intel X^e Matrix Extensions (XMX) AI engines on Intel discrete GPUs. Moreover, through PyTorch* xpu device, Intel® Extension for PyTorch* provides easy GPU acceleration for Intel discrete GPUs with PyTorch*.

LiCO supports users running Intel Extension for PyTorch program on HPC clusters. To use Intel GPU for your pytorch program, If your program contains code like `torch.device("cpu")`, you should make the corresponding simple modifications before running the job.

Intel Distribution for Python

The Intel Distribution for Python achieve fast math-intensive workload performance without code changes for data science and machine learning problems. Intel Distribution for Python is included as part of the Intel oneAPI AI Analytics Toolkit, which provides accelerated machine learning and data analytics pipelines with optimized deep-learning frameworks and high-performing Python libraries.

Intel Distribution of Modin

The Intel Distribution of Modin is a performant, parallel, and distributed dataframe system that is designed around enabling data scientists to be more productive with the tools that they love. This library is fully compatible with the pandas API. It is powered by OmniSci in the back end and provides accelerated analytics on Intel platforms.

In LiCO, Intel Distribution of Modin templates are available for a single node or multi nodes.

- Intel Distribution of Modin Single Node
- Intel Distribution of Modin Multi Node

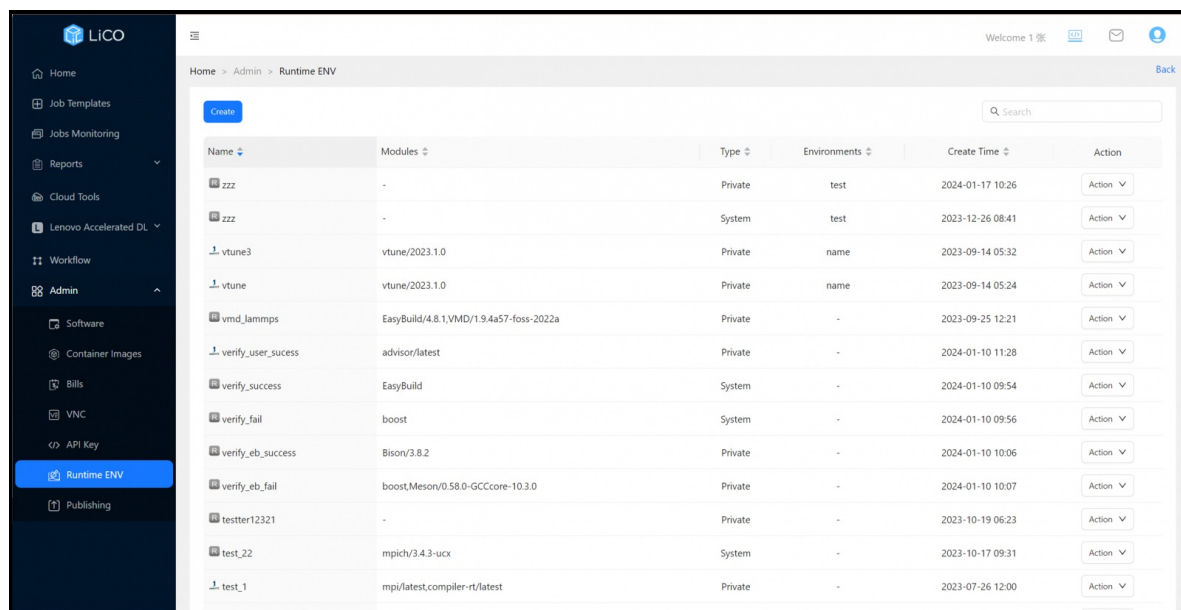
Model Zoo for Intel Architecture

Model Zoo for Intel Architecture contains Intel optimizations for running deep learning workloads on Intel Xeon Scalable processors. In LiCO Image recognition and Object Detection jobs are available with both TensorFlow and PyTorch. Multiple models can be selected when the user is submitting the job.

- TensorFlow Image Recognition of Intel Model Zoo
- TensorFlow Object Detection of Intel Model Zoo
- PyTorch Image Recognition of Intel Model Zoo
- PyTorch Object Detection of Intel Model Zoo

HPC Runtime Module Management

LiCO HPC/AI version allows the user to pre-define modules and environmental variables to load at the time of job execution through Job submission templates. These user-defined modules eliminate the step of needing to manually load required modules before job submission, further simplifying the process of running HPC workloads on the cluster. Through the Runtime interface, users can choose from the modules available on the system, define their loading order, and specify environmental variables for repeatable, reliable job deployment.



The screenshot shows the LiCO web interface for HPC Runtime Module Management. The left sidebar contains navigation links: Home, Job Templates, Jobs Monitoring, Reports, Cloud Tools, Lenovo Accelerated DL, Workflow, and Admin. The Admin section is expanded, showing Software, Container Images, Bills, VNC, API Key, Runtime ENV (selected), and Publishing. The main content area is titled 'Runtime ENV' and contains a table of modules. The table has columns for Name, Modules, Type, Environments, Create Time, and Action. A 'Create' button and a search bar are at the top of the table.

Name	Modules	Type	Environments	Create Time	Action
zzz	-	Private	test	2024-01-17 10:26	Action
zzz	-	System	test	2023-12-26 08:41	Action
vtune3	vtune/2023.1.0	Private	name	2023-09-14 05:32	Action
vtune	vtune/2023.1.0	Private	name	2023-09-14 05:24	Action
vmd_lammps	EasyBuild/4.8.1,VMD/1.9.4a57-foss-2022a	Private	-	2023-09-25 12:21	Action
verify_user_success	advisor/latest	Private	-	2024-01-10 11:28	Action
verify_success	EasyBuild	System	-	2024-01-10 09:54	Action
verify_fail	boost	System	-	2024-01-10 09:56	Action
verify_eb_success	Bison/3.8.2	Private	-	2024-01-10 10:06	Action
verify_eb_fail	boost,Meson/0.58.0-GCCcore-10.3.0	Private	-	2024-01-10 10:07	Action
tester12321	-	Private	-	2023-10-19 06:23	Action
test_22	mpich/3.4.3-ucx	System	-	2023-10-17 09:31	Action
test_1	mpi/latest,compiler-rt/latest	Private	-	2023-07-26 12:00	Action

Figure 27. HPC runtime module list

Home > Submit Job > MPI

MPI

Message Passing Interface (MPI) is a communication standard used in parallel computing and HPC applications. Use this template to submit your MPI application.

Template Information ▾

* Job Name

* Workspace

Template Parameters ▾

Runtime

OS Default ^

OS Default

test

runtime

MPI environment

Figure 28. MPI job template with custom module setup

Container-based HPC workload deployment

Additional standard templates are provided to support deployment of containerized HPC workloads through Singularity or CharlieCloud. These templates simplify deploying containers for HPC workloads by eliminating the need to create custom runtimes and custom templates for these workloads unless needed for more granularity.

In addition to providing a certain number of basic container images, LiCO also allows users to upload customized container images. LiCO 5.2.0 and later versions support running jobs on NGC images.

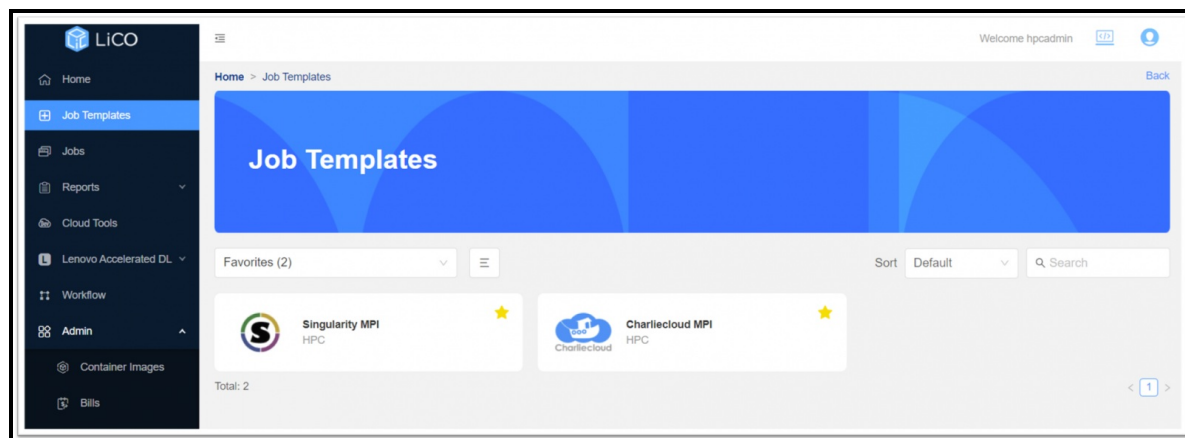
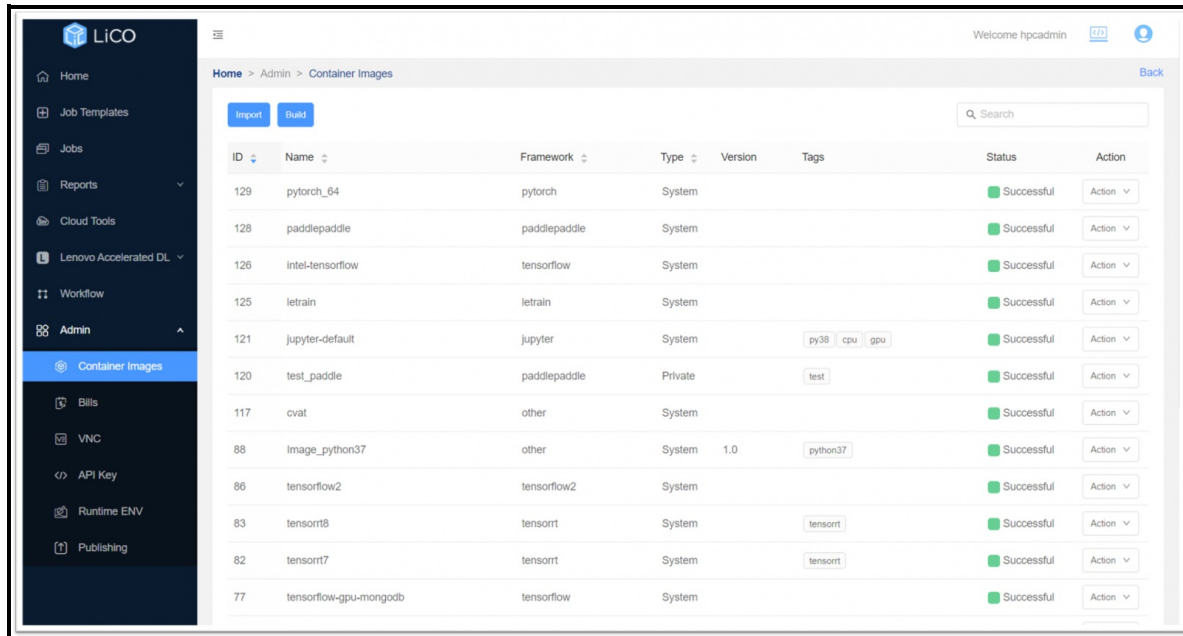


Figure 29. CharlieCloud and Singularity standard job templates

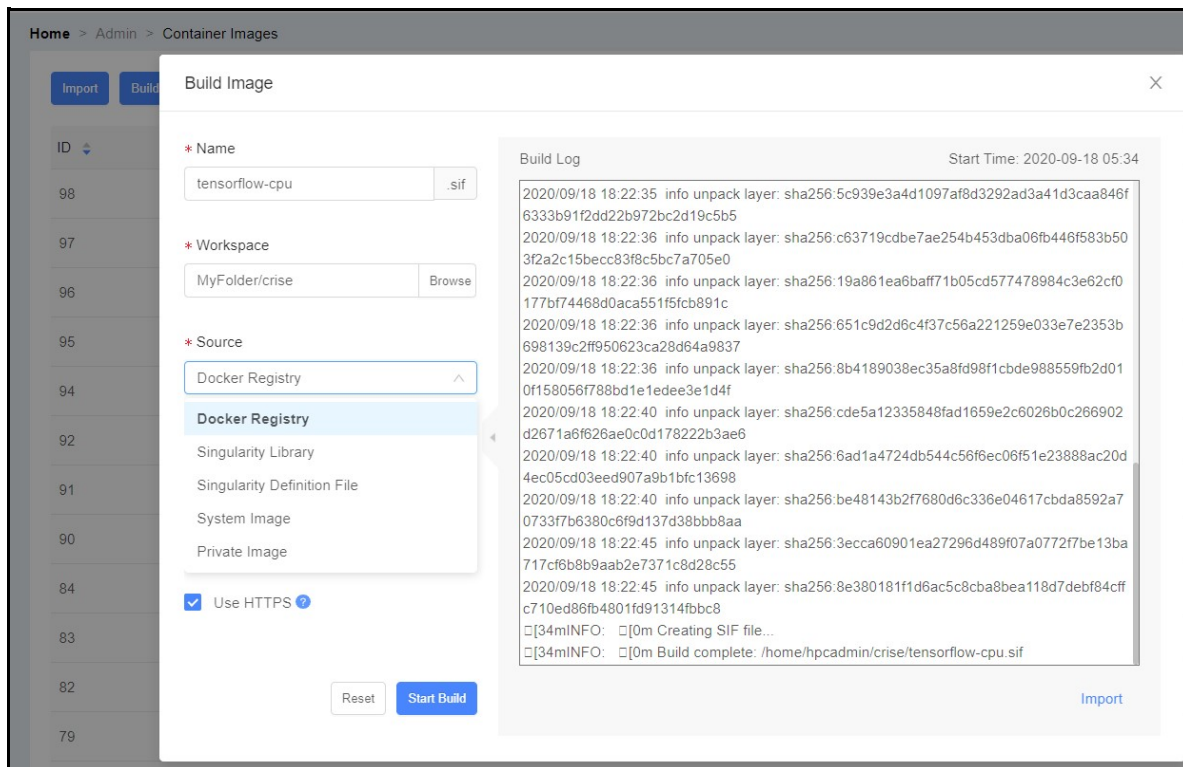
Singularity Container Image Management

LiCO HPC/AI version provides both users and administrators with the ability to build, upload and manage application environment images through Singularity containers. These images can support users with AI frameworks and HPC workloads, as well as others. Singularity containers may be built from Docker containers, imported from NVIDIA GPU Cloud (NGC), or other image repositories such as the Intel Container Portal. Containers created by administrators are available to all users, and users can create container images for their individual use as well. Users looking to deploy a custom image can also create a custom template that will deploy the container and run workloads in that environment.



ID	Name	Framework	Type	Version	Tags	Status	Action
129	pytorch_64	pytorch	System			Successful	Action
128	paddlepaddle	paddlepaddle	System			Successful	Action
126	intel-tensorflow	tensorflow	System			Successful	Action
125	letrain	letrain	System			Successful	Action
121	jupyter-default	jupyter	System		py38 cpu gpu	Successful	Action
120	test_paddle	paddlepaddle	Private		test	Successful	Action
117	cvat	other	System			Successful	Action
88	Image_python37	other	System	1.0	python37	Successful	Action
86	tensorflow2	tensorflow2	System			Successful	Action
83	tensoirt8	tensoirt	System		tensoirt	Successful	Action
82	tensoirt7	tensoirt	System		tensoirt	Successful	Action
77	tensorflow-gpu-mongodb	tensorflow	System			Successful	Action

Figure 30. Singularity container management through the Administrator portal



Build Image

* Name

tensorflow-cpu

.sif

* Workspace

MyFolder/crise

Browse

* Source

Docker Registry

Docker Registry

Singularity Library

Singularity Definition File

System Image

Private Image

☒ Use HTTPS

Reset

Start Build

Build Log

Start Time: 2020-09-18 05:34

```

2020/09/18 18:22:35 info unpack layer: sha256:5c939e3a4d1097af8d3292ad3a41d3caa846f
6333b91f2dd22b972bc2d19c5b5
2020/09/18 18:22:36 info unpack layer: sha256:c63719cde7ae254b453dba06fb446f583b50
3f2a2c15becc83f8c5bc7a705e0
2020/09/18 18:22:36 info unpack layer: sha256:19a861ea6baff71b05cd577478984c3e62cf0
177bf74468d0aca551f5fcb891c
2020/09/18 18:22:36 info unpack layer: sha256:651c9d2d6c4f37c56a221259e0337e2353b
698139c2ff950623ca28d64a9837
2020/09/18 18:22:36 info unpack layer: sha256:8b4189038ec35a8fd98f1cbde988559fb2d01
0f158056f788bd1e1edee3e1d4f
2020/09/18 18:22:40 info unpack layer: sha256:cde5a12335848fad1659e2c6026b0c266902
d2671a9f626ae0c0d178222b3ae6
2020/09/18 18:22:40 info unpack layer: sha256:6ad1a4724b544c56f6ec06f51e23888ac20d
4ec05cd03eed907a9b1bfc13698
2020/09/18 18:22:40 info unpack layer: sha256:be48143b2f7680d6c336e04617cbda8592a7
0733f7b6380c6f9d137d38bbb8aa
2020/09/18 18:22:45 info unpack layer: sha256:3ecca60901ea27296d489f07a0772f7be13ba
717cf6b8b9aab2e7371c8d28c55
2020/09/18 18:22:45 info unpack layer: sha256:8e380181fd6ac5c8cba8bea118d7defb84cff
c710ed86fb4801fd91314fbbc8
[34m]INFO: [0m Creating SIF file...
[34m]INFO: [0m Build complete: /home/hpcadmin/crise/tensorflow-cpu.sif

```

Import

Figure 31. Singularity container building within LiCO

Lenovo Intelligent Computing Orchestration (LiCO)

22

Reports

LiCO HPC/AI version provides expanded billing capabilities and provides the user access to monitor charges incurred for a date range via the Expense Reports subtab. Users can also download daily or monthly billing reports as a .xlsx file from the Admin tab.

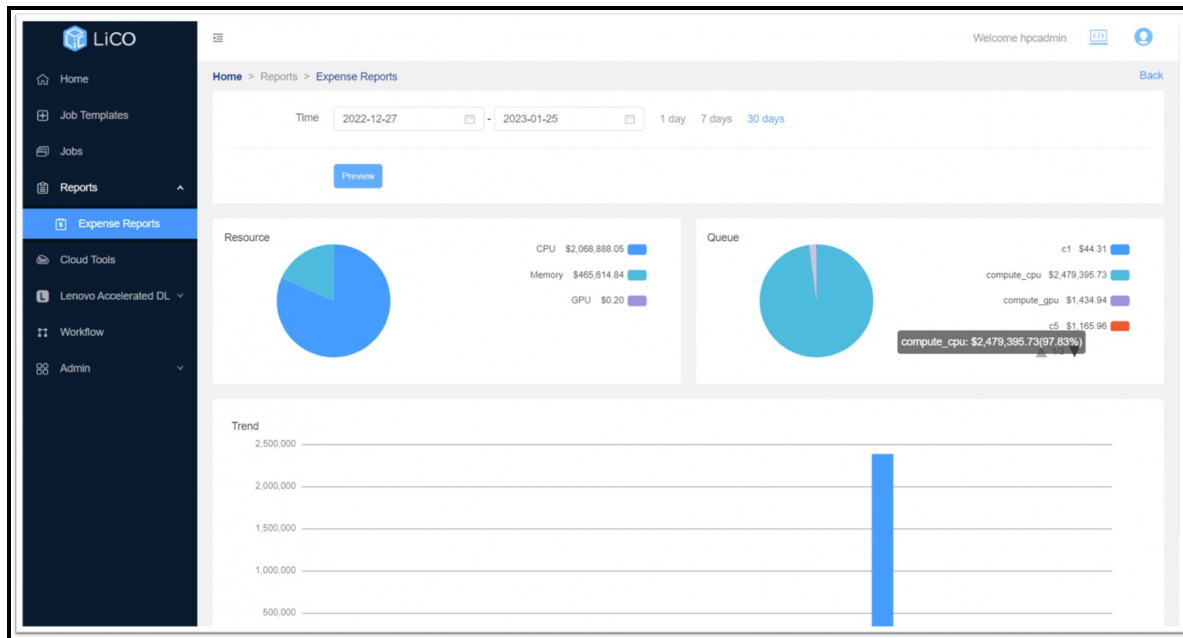


Figure 32. LiCO User view of Expense Reports

System tools

The system tools option for the user provides access to their storage space on the cluster. The user can upload, download, cut/copy/paste, preview and edit files on the cluster storage space from within the LiCO portal. The text editor within LiCO allows syntax-aware display and editing based on the file extension. Multiple files editor option is available.

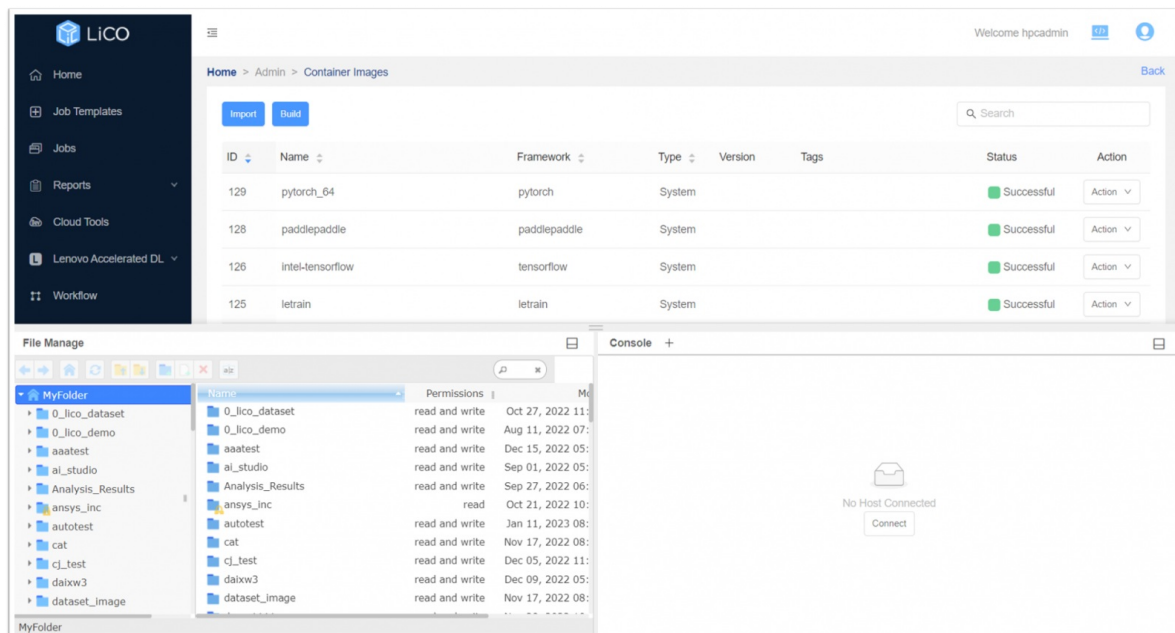


Figure 33. Cluster storage access

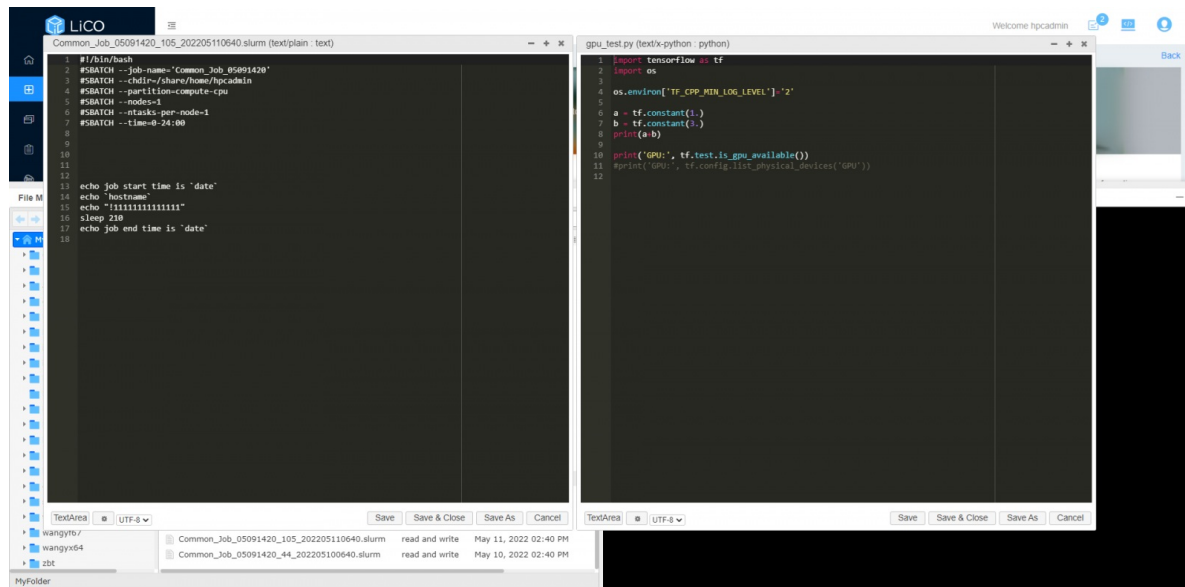


Figure 34. Text file editor

Features for LiCO Administrators

Topics in this section:

- [Features for LiCO HPC/AI version Administrators](#)
- [Features for LiCO Operators](#)
- [LiCO Deployment](#)
- [LiCO Upgrade](#)

Features for LiCO HPC/AI version Administrators

For cluster administrators, LiCO provides a sophisticated monitoring solution, built on OpenHPC tooling. The following menus are available to administrators:

- **Home menu for administrators** – provides dashboards giving a global overview of the health of the cluster. Utilization is given for the CPUs, GPUs, memory, storage, and network. Node status is given, indicating which nodes are being used for I/O, compute, login, and management. Job status is also given, indicating runtime for the current job, and the order of jobs in the queue. The Home menu is shown in the following figure.

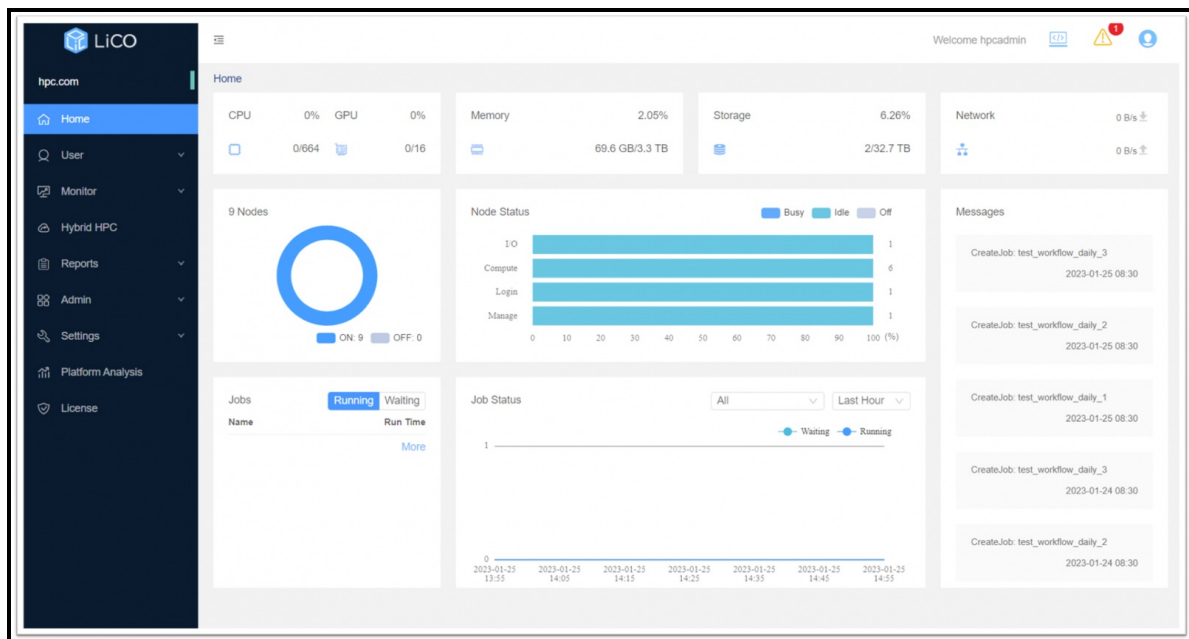


Figure 35. LiCO HPC/AI Administrator Home Menu

- **User Management menu** – provides dashboards to control user groups and users, determining permissions and access levels (based on LDAP) for the organization. Administrators can also control and provision billing groups for accurate accounting.
- **Monitor menu** – provides dashboards for interactive monitoring and reporting on cluster nodes, including a list of the nodes, or a physical look at the node topology. Administrators may also use the Monitor menu to drill down to the component level, examining statistics on cluster CPUs, GPUs, networking, jobs, and operations. Administrators can access alerts that indicate when these statistics reach unwanted values (for instance, GPU temperature reaching critical levels). These alerts are created using the Setting menu. Additionally, a large screen view is available to display a high-level summary of cluster status, and a cluster view was added since LiCO 6.2 for a focused view of compute resource utilization across the cluster. The figures below display the component and alert dashboards.

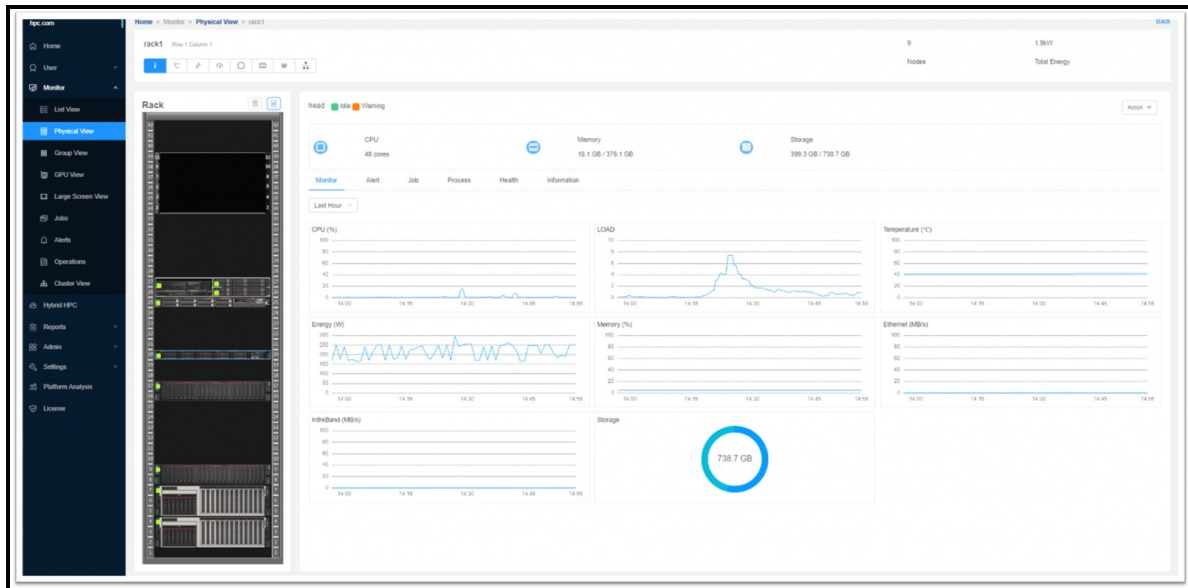


Figure 36. LiCO HPC/AI Administrator Component dashboard

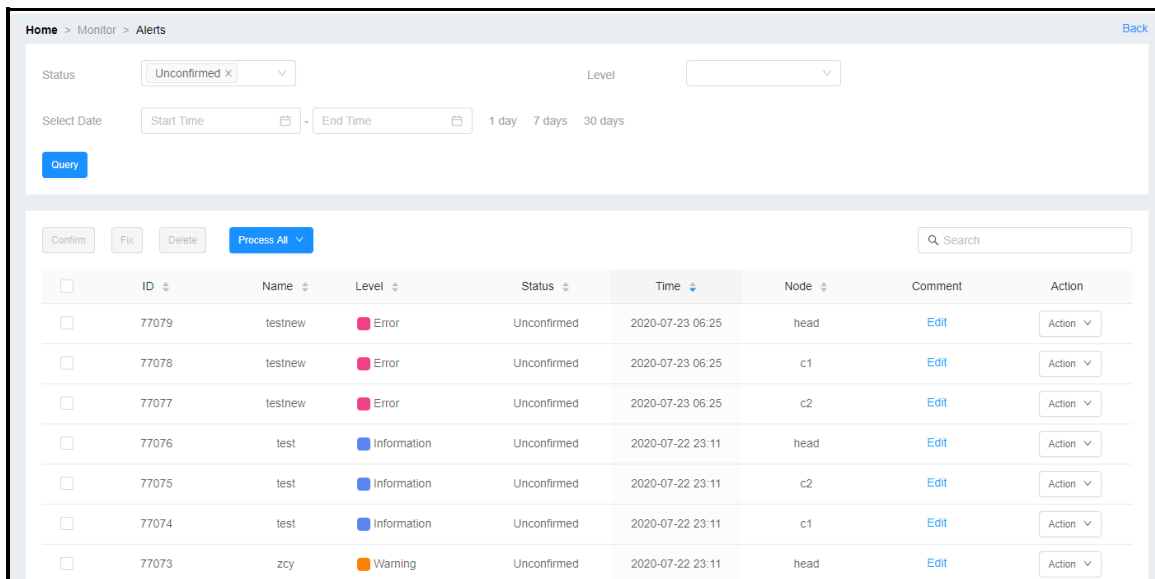


Figure 37. LiCO HPC/AI Administrator Alert dashboard

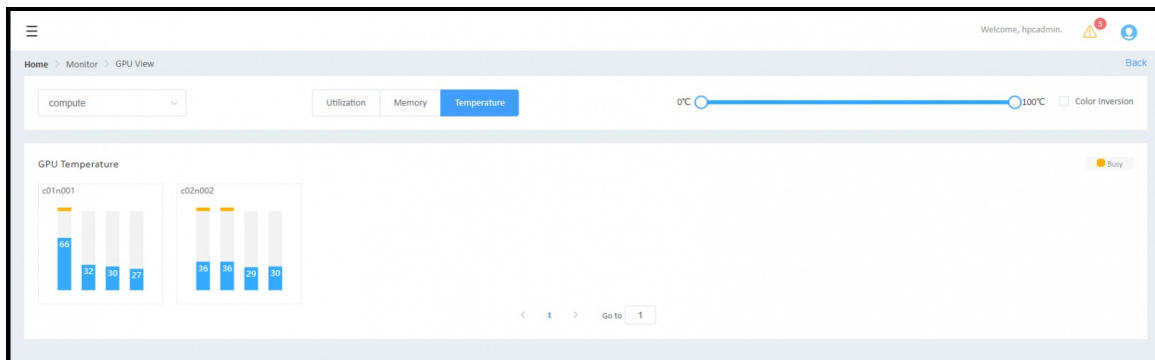


Figure 38. LiCO HPC/AI GPU View dashboard

- **Hybrid HPC** - Bursting on cloud is now possible. LiCO 7.0 has support for hybrid cloud integration leveraging Microsoft Azure. That will allow our customers to add Microsoft Azure access as

Supercomputer resources that will be included in the scheduling considerations. The Hybrid HPC enable users to leverage the public cloud to dynamically scale the computing resources of the local HPC cluster.

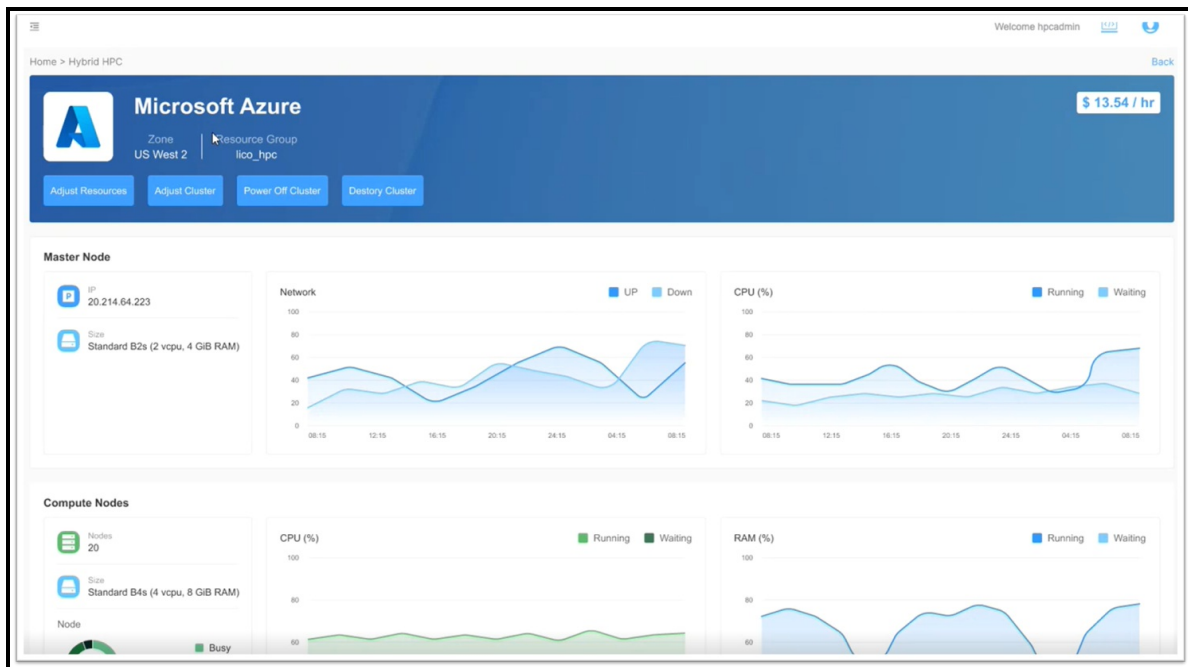


Figure 39. LiCO Hybrid HPC for Microsoft Azure

- LiCO version 7.2 introduces enhanced cloud integration capabilities through the incorporation of Covalent, an open-source Pythonic workflow orchestration platform. This integration allows LiCO to seamlessly operate in a Hybrid Cloud environment, supporting popular cloud providers such as AWS, Azure, and Google Cloud.

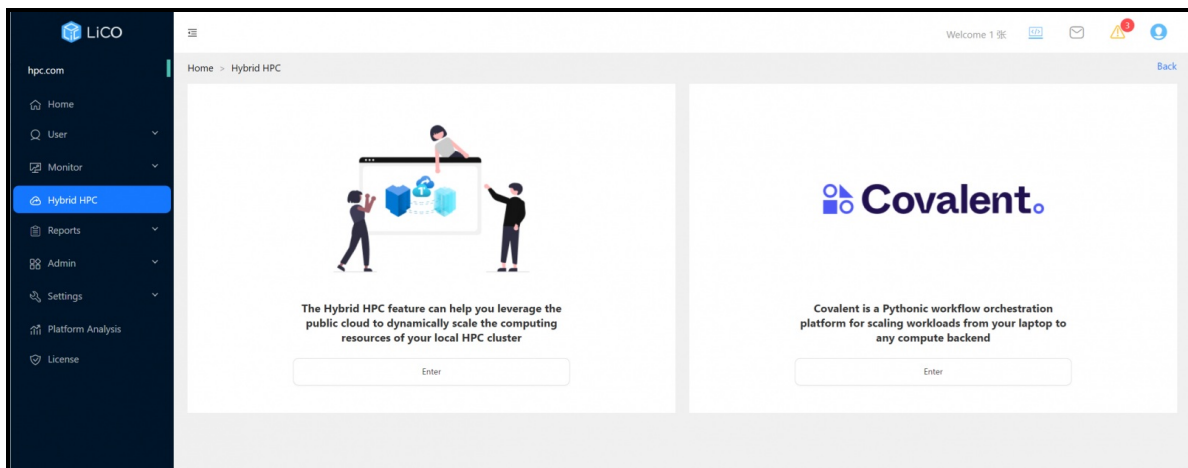


Figure 40. LiCO Hybrid HPC

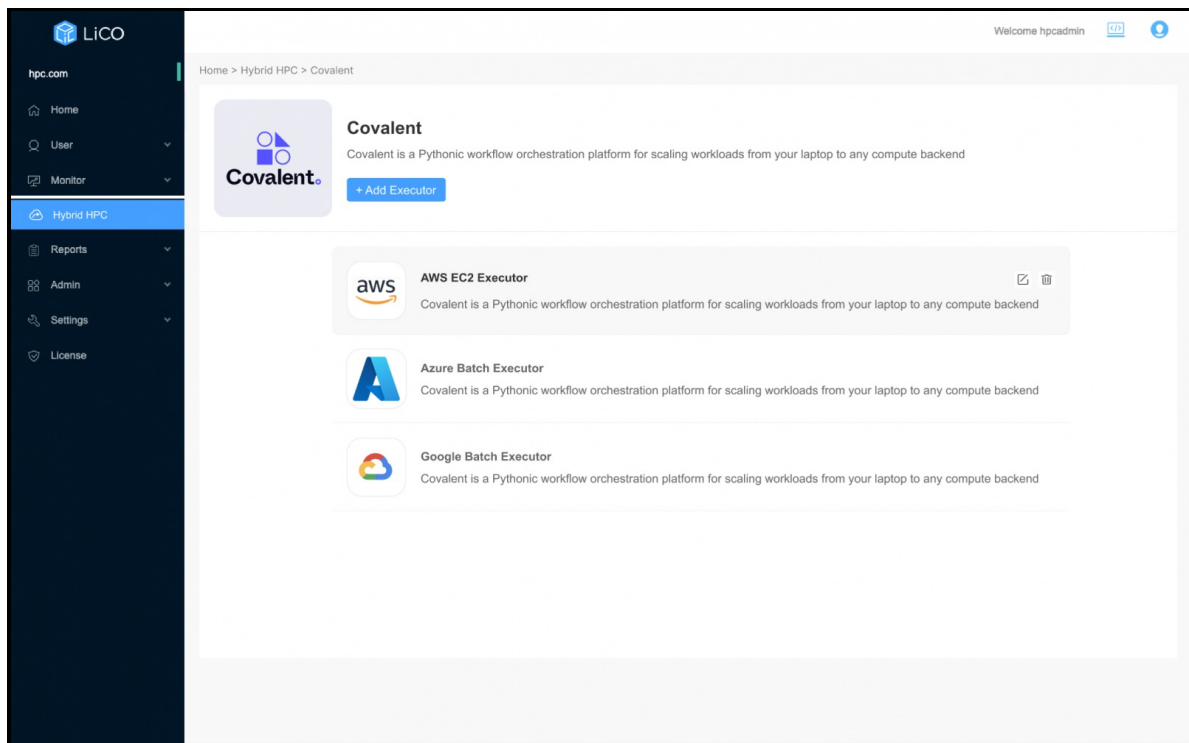


Figure 41. LiCO integration with Covalent for Multi-Cloud support

- **Reports menu** – allows administrators the ability to generate reports on jobs, cluster utilization, alerts, and view current charges and cluster utilization.
- **Admin menu** – Provides the administrator with the capability to create Singularity images for use by all users, generate billing spreadsheets, examine processes and assets, monitor VNC sessions, and download web logs. The administrators can also publish announcements to users using the Notice function.
- **Settings menu** – allows administrators to set up automated notifications and alerts. Administrators may enable the notifications to reach users and interested parties via email, SMS, and WeChat. Administrators may also enable notifications and alerts via uploaded scripts. The Settings menu also allows administrators to create and modify queues. These queues allow administrators to subdivide hardware based on different types or needs. For example, one queue may contain systems that are exclusively machines with GPUs, while another queue may contain systems that only contain CPUs. This allows the user running the job to select the queue that is more applicable to their requirement. Within the Settings menu, administrators can also set the status of queues, bringing them up or down, draining them, or marking them inactive. Administrators can also limit which queues are available to users by user group.

Starting with LiCO 7.1 administrators can configure Quality of Service (QOS) for each job submitted to Slurm from the Scheduler sub tab. Administrators can edit and delete a limitation, create a new limitation and then associate it with the corresponding Billing Group. This function requires to configure the SLURM account on the cluster.

- **Platform Analysis menu** – allows the administrator to analyze and optimize program performance. Administrators can determine the cause of poor performance by finding software and hardware performance bottlenecks and identifying program hotspots. After that, developers can optimize programs according to the causes.

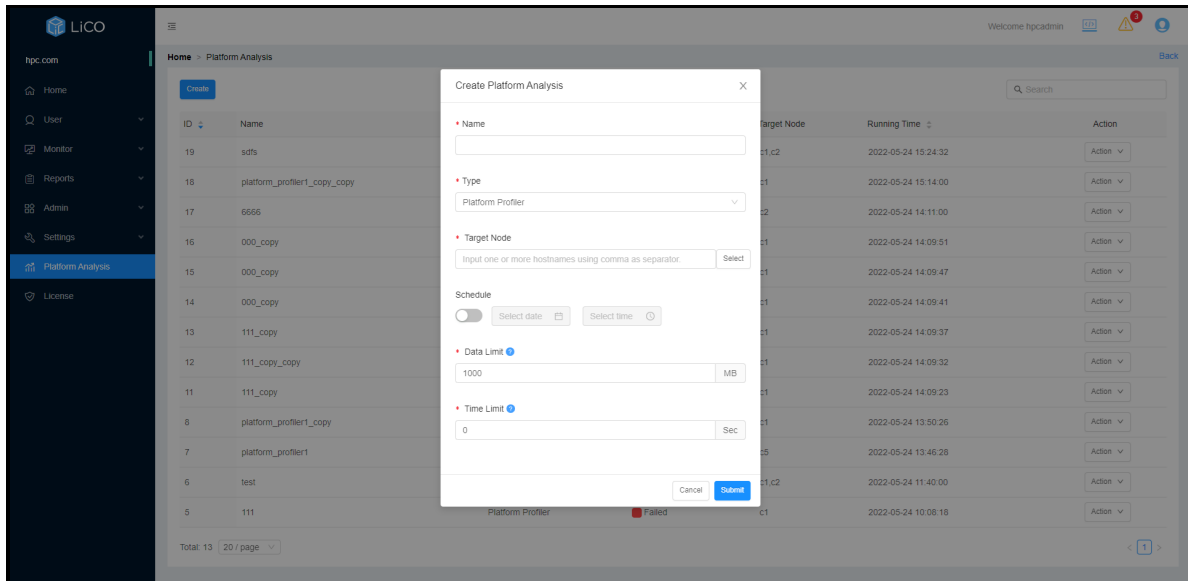


Figure 42. Platform analysis tools for HPC cluster administrator and end user

- **License menu** – displays the software licenses active in LiCO including the number of licensed processing entitlements and the expiration date of the license.

Features for LiCO Operators

For the purpose of monitoring clusters but not overseeing user access, LiCO provides the Operator designation. LiCO Operators have access to a subset of the dashboards provided to Administrators; namely, the dashboards contained in the Home, Monitor, and Reports menus:

- **Home menu for operators** – provides dashboards giving a global overview of the health of the cluster. Utilization is given for the CPUs, GPUs, memory, storage, and network. Node status is given, indicating which nodes are being used for I/O, compute, login, and management. Job status is also given, indicating runtime for the current job, and the order of jobs in the queue.
- **Monitor menu** – Dashboard that enables interactive monitoring and reporting on cluster nodes, including a list of the nodes, or a physical look at the node topology. Operators may also use the Monitor menu to drill down to the component level, examining statistics on cluster CPUs, GPUs, jobs, and operations. Operators can access alarms that indicate when these statistics reach unwanted values (for instance, GPU temperature reaching critical levels.) These alarms are created by Administrators using the Settings menu (for more information on the Settings menu, see the [Features for LiCO Administrators](#) section.)
- **Reports menu** – allows operators the ability to generate reports on jobs, alerts, or actions for a given time interval. Operators may export these reports as a spreadsheet, in a PDF, or in HTML.

LiCO Deployment

Docker containerized deployment

LiCO can be deployed in Docker containers with all the supported operating systems.

In LiCO we offer containerized deployment as a preferred method. This approach involves running the HPC cluster drivers, monitoring software, scheduler and applications on the host operating system, while the container instance contains the LiCO web portal, including the back-end LiCO service.

The container version of our product supports all the schedulers supported in the deployment.

Benefits of Containerized Deployment

- **Simplify the deployment process:** Containerized deployment enables LiCO to work efficiently with the

latest OpenHPC package, streamlining updates and maintaining compatibility with the latest features while the number of packages to be deployed is minimal.

- **Simplified Test Systems:** Containerized deployment reduces the complexity of maintaining and testing multiple systems for different OS versions, leading to a more efficient and cost-effective deployment process.
- **Reduced Operating System Dependency:** Containerization reduces the dependency on specific operating systems, enhancing portability and making it easier to deploy LiCO across diverse environments.

Supported Operating Systems

The container version of our product supports the following operating systems:

- Rocky Linux 8 (Default Base OS)
- RHEL (Red Hat Enterprise Linux)
- Ubuntu (Only in Container Mode)
- SUSE (Only in Container Mode)

By default, the container version of our product uses Rocky Linux 8 as the base operating system for container images.

Customers with specific security considerations for the operating system can opt for host OS deployment. For Rocky and RHEL operating systems, we offer support for both host OS deployment and container deployment. Customers can choose the mode that best suits their requirements.

For any further assistance or inquiries, please review the Deploy LiCO in container section from the installation guides available at <https://support.lenovo.com/us/en/solutions/HT507011>

LiCO Upgrade

Below you can find the instructions for upgrading LiCO (License and Configuration Optimizer) from version 7.x (where "x" represents the source version) to version 7.y (where "y" represents the latest available version) on different hosting environments.

This upgrade guide below outlines the steps for transitioning from LiCO version 7.x to version 7.y on different hosting environments. Follow the instructions carefully to ensure a successful upgrade while minimizing disruptions to your LiCO system. If you encounter any issues during the upgrade process, consult the official LiCO documentation or seek assistance from L3 support.

This guide is intended for system administrators, developers, or individuals responsible for managing LiCO installations.

Versions:

Source LiCO version: 7.x

Target LiCO version: 7.y

Upgrade Scenarios:

- LiCO v7.x on Host > v7.y on Host:

In this scenario, you are upgrading LiCO on the host machine directly from version 7.x to version 7.y. No manual changes are required, and you can use an auto-upgrade script or guide provided by the LiCO team.

- LiCO v7.x on Host > v7.y on Container:

For this upgrade, you will be migrating from the host-based installation of LiCO version 7.x to a containerized version 7.y. The process involves removing LiCO packages, retaining configuration files and database files, installing Docker, building or downloading the LiCO v7.y container image, and following the installation guide to configure LiCO v7.y.

- LiCO v7.x on Container > v7.y on Container:

If you already have LiCO version 7.x running in a container and want to upgrade to a newer version 7.y, you need to replace the LiCO container image with the one corresponding to version 7.y. Afterward, follow the installation guide to configure LiCO v7.y.

- LiCO v7.x on Container > v7.y on Host:

This upgrade scenario is not recommended and should only be attempted with the assistance of L3 support. The process involves migrating from LiCO version 7.x running in a container to a host-based installation of LiCO version 7.y.

Important Notes:

Always back up your configuration files and database files before proceeding with any upgrade.

Make sure to follow the official LiCO installation and upgrade guides provided by the vendor for each specific version.

LiCO GUI Installer is a tool that simplifies HPC cluster deployment and LiCO setup. It runs on the management node and it can use Confluent to deploy the OS on the compute nodes.

The user can define the following node types:

- head node (currently only a single head node is supported. (This is the same machine on which the installer runs)
- login nodes - one or more
- compute nodes - one or more

The compute nodes that have at least 1 GPU defined in the config file are treated as GPU nodes and NVIDIA drivers will be installed on these.

You can download LiCO Installation GUI from [here](#), and follow this [guide](#) to deploy HPC cluster and LiCO easily.



Figure 43. LiCO GUI Installer

Diskless installation

A diskless boot system (otherwise known as a PXE boot setup) is a computer system without hard drives. Instead, each computer uses network-attached storage drives on a server to store data.

LiCO support the option to have a diskless installation. You can follow this [guide](#) to deploy HPC cluster with the diskless option.

Subscription and Support

LiCO HPC/AI is enabled through a per-CPU and per-GPU subscription and support entitlement model, which once entitled for the all the processors contained within the cluster, gives the customer access to LiCO package updates and Lenovo support for the length of the acquired term.

The entitlement must be applied to all systems being managed by LiCO, including the login, management and all the compute nodes. Each CPU and each GPU requires an entitlement.

Lenovo will provide interoperability support for all software tools defined as validated with LiCO, and development support (Level 3) for specific Lenovo-supported tools only. Open source and supported-vendor bugs/issues will be logged and tracked with their respective communities or companies if desired, with no guarantee from Lenovo for bug fixes. Full support details are provided at the support links below for each respective version of LiCO. Additional support options may be available; please contact your Lenovo sales representative for more information.

LiCO can be acquired as part of a Lenovo Scalable Infrastructure (LeSI) solution or for “roll your own” (RYO) solutions outside of the LeSI framework, and LiCO software package updates are provided directly through the Lenovo Electronic Delivery system. More information on LeSI is available in the LeSI product guide, available from <https://lenovopress.com/lp0900>.

- Lenovo provides support in English globally and in Chinese for China (24x7)
- Support response times are as follows:
 - Severity 1 issues response is 1 business day
 - Other issues: 3 business days

LiCO has 1-year lifecycle for each release, customer should upgrade to latest version if out of support.

The following table lists end of support for LiCO versions.

Table 3. End of support list

Version	Date (mm/dd/yyyy)
LiCO 8.0.0	3/31/2026
LiCO 7.2.2	12/31/2025
LiCO 7.2.1	7/22/2025
LiCO 7.2	2/06/2025
LiCO 7.1	6/28/2024
LiCO 7.0	12/20/2023
LiCO 6.4	6/24/2023
LiCO 6.3.1	3/29/2023
LiCO 6.3	12/15/2022
LiCO 6.2	6/2/2022
LiCO 6.1	12/15/2021
LiCO 6.0	8/3/2021
LiCO 5.5.0	4/15/2021
LiCO 5.4.0	11/5/2020
LiCO 5.3.1	6/18/2020
LiCO 5.3.0	4/12/2020
LiCO 5.2.1	1/9/2020
LiCO 5.2.0	11/21/2019
LiCO 5.1.0	5/3/2019

Validated software components

LiCO's software packages are dependent on a number of software components that need to be installed prior to LiCO in order to function properly. Each LiCO software release is validated against a defined configuration of software tools and Lenovo systems, to make deployment more straightforward and enable support. Other management tools, hardware systems and configurations outside the defined stack may be compatible with LiCO, though not formally supported; to determine compatibility with other solutions, please check with your Lenovo sales representative.

The following software components are validated by Lenovo as part of the overall LiCO software solution entitlement:

LiCO HPC/AI version support

- Lenovo Development Support (L1-L3)
 - Graphical User Interface: LiCO
 - System Management & Provisioning: Confluent
- Lenovo LiCO HPC/AI Configuration Support (L1 only)
 - Job Scheduling & Orchestration: SLURM, OpenPBS, Torque/Maui (HPC only)
 - System Monitoring: Icinga v2
 - Container Support (AI): Singularity, CharlieCloud, NGC
 - AI Frameworks (AI): Caffe, Intel-Caffe, TensorFlow, MxNet, Neon, Chainer, Pytorch, Scikit-learn, PaddlePaddle, NVIDIA TensorRT, PyTorch Lightning, ONNX
 - Visualization: Grafana, TensorBoard

The following software components are validated for compatibility with LiCO HPC/AI:

- Supported by their respective software provider
 - Operating System: RHEL 8.8, Rocky Linux 8.8, SUSE SLES 15 SP3, CentOS 7.9, Ubuntu 22.04 LTS
 - File Systems: IBM Spectrum Scale, Lustre GPFS
 - Job Scheduling & Orchestration: IBM Spectrum LSF v10, Altair PBS Pro
 - Development Tools: GNU compilers, Intel Cluster Toolkit

Validated hardware components

Supported GPUs

- NVIDIA L40, NVIDIA H100, NVIDIA A100, NVIDIA A30, NVIDIA A40, NVIDIA T4, NVIDIA V100, NVIDIA RTX8000, NVIDIA RTX6000, RTX6000 Ada
- Intel Flex Series 140 GPU, Intel Flex Series 170 GPU
- NVIDIA H100 Multi-Instance GPU (MIG), NVIDIA A100 Multi-Instance GPU (MIG), NVIDIA HGX H200

Note: Subject to specific ThinkSystem platform support, not all GPUs available on all systems

Supported Networks

- Intel OmniPath 100
- Mellanox Infiniband (FDR, EDR, HDR, NDR)
- Gb Ethernet (1, 10, 25, 40,50, 100)

Supported servers

LiCO seamlessly integrates with both Lenovo servers and workstations, offering robust support for Lenovo hardware within the cluster. Additionally, LiCO extends its compatibility beyond Lenovo infrastructure, providing full support for **non-Lenovo hardware** within the cluster environment. This versatility ensures optimal performance and flexibility, allowing organizations to leverage LiCO's capabilities across a diverse range of hardware configurations for efficient and scalable computing orchestration.

The following Lenovo systems are supported to run with LiCO HPC/AI. This systems must run one of the supported operating systems as well as the validated software stack, as described in the Validated Software Components section.

- **ThinkSystem SR680a V3** - The Lenovo ThinkSystem SR680a V3 is a powerful 8U server that features two 5th Gen Intel Xeon Scalable processors and eight high-performance GPUs. This air-cooled server is the ultimate Generative AI server, with advanced GPU-to-GPU communications and high-speed PCIe 5.0 connectivity between the GPUs, CPUs, and networking. It was designed fully in-house by Lenovo from ground up to provide best-in-class modularity, thermal performance and reliability.
- **ThinkSystem SR685a V3** - The Lenovo ThinkSystem SR685a V3 is a powerful 8U server that features two AMD EPYC 9004 processors and eight high-performance GPUs. This air-cooled server is the ultimate Generative AI server, with advanced GPU-to-GPU communications and high-speed PCIe 5.0 connectivity between the GPUs, CPUs, and networking. It was designed fully in-house by Lenovo from ground up to provide best-in-class modularity, thermal performance and reliability.
- **ThinkSystem SR780a V3** - The Lenovo ThinkSystem SR780a V3 is a powerful liquid-cooled 5U server that features two 5th Gen Intel Xeon Scalable processors and eight high-performance GPUs. This server is the ultimate Generative AI server, with advanced GPU-to-GPU communications and high-speed PCIe 5.0 connectivity between the GPUs, CPUs, and networking. It was designed fully in-house by Lenovo from ground up to provide best-in-class modularity, thermal performance and reliability.
- **ThinkSystem SD650-N V3** – The ThinkSystem SD650-N V3 Neptune DWC server is the next-generation high-performance server based on the fifth generation Lenovo Neptune™ direct water cooling platform. With two 5th Gen Intel Xeon Scalable or Intel Max Series CPUs, along with four NVIDIA H100 SXM5 GPUs, the ThinkSystem SD650-N V3 server features the latest technology from Intel and NVIDIA. For more information, see the [SD650-N V3 product guide](#).
- **ThinkSystem SD665-N V3** – The ThinkSystem SD665-N V3 Neptune DWC server is the next-generation high-performance server based on the fifth generation Lenovo Neptune™ direct water cooling platform. With one or two fourth-generation AMD EPYC processors and four powerful NVIDIA H100 Tensor Core GPUs, the ThinkSystem SD665-N V3 server features the latest technology from AMD and NVIDIA. For more information, see the [SD665-N V3 product guide](#).
- **ThinkSystem SR860 V3** – The Lenovo ThinkSystem SR860 V3 is a 4-socket server that features a 4U rack design with support for up to eight high-performance GPUs. The server offers technology advances, including 4th Gen Intel Xeon Scalable processors, 4800 MHz DDR5 memory, and PCIe 5.0. For more information, see the [SR860 V3 product guide](#).
- **ThinkSystem SR850 V3** – The Lenovo ThinkSystem SR850 V3 is a 4-socket server that is densely packed into a 2U rack design. The server offers technology advances, including 4th Gen Intel Xeon Scalable processors, 4800 MHz DDR5 memory, and PCIe Gen 5. For more information, see the [SR850 V3 product guide](#).
- **ThinkSystem SR675 V3** – The Lenovo ThinkSystem SR675 V3 is a versatile GPU-rich 3U rack server that supports eight double-wide GPUs including the new NVIDIA H100 and L40 Tensor Core GPUs, or the NVIDIA HGX H100 4-GPU offering with NVLink and Lenovo Neptune hybrid liquid-to-air cooling. The server is based on the new AMD EPYC 9004 Series processors (formerly codenamed "Genoa"). For more information, see the [SR675 V3 product guide](#).
- **ThinkSystem SD665 V3** – The ThinkSystem SD665 V3 Neptune DWC server is the next-generation high-performance server based on the fifth generation Lenovo Neptune™ direct water cooling platform. For more information, see the [SD665 V3 product guide](#).
- **ThinkSystem SR655 V3** – The Lenovo ThinkSystem SR655 V3 is a 1-socket 2U server that features the AMD EPYC 9004 "Genoa" family of processors. With up to 96 cores per processor and support for the new PCIe 5.0 standard for I/O, the SR655 V3 offers the ultimate in one-socket server performance in a 2U form factor. For more information, see the [SR655 V3 product guide](#).

- **ThinkSystem SR635 V3** – The Lenovo ThinkSystem SR635 V3 is a 1-socket 1U server that features the AMD EPYC 9004 "Genoa" family of processors. With up to 96 processor cores and support for the new PCIe 5.0 standard for I/O, the SR635 V3 offers the ultimate in one-socket server performance in a 1U form factor. For more information, see the [SR635 V3 product guide](#).
- **ThinkSystem SD650-I V3** – The ThinkSystem SD650-I V3 Neptune DWC server is the next-generation high-performance server based on the fifth generation Lenovo Neptune™ direct water cooling platform. For more information, see the [SD650-I V3 product guide](#).
- **ThinkSystem SD650 V3** – The ThinkSystem SD650 V3 Neptune DWC server is the next-generation high-performance server based on the fifth generation Lenovo Neptune™ direct water cooling platform. For more information, see the [SD650 V3 product guide](#).
- **ThinkSystem SR650 V3** – The Lenovo ThinkSystem SR650 V3 is an ideal 2-socket 2U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. For more information, see the [SR650 V3 product guide](#).
- **ThinkSystem SR630 V3** – The Lenovo ThinkSystem SR630 V3 is an ideal 2-socket 1U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. For more information, see the [SR630 V3 product guide](#).
- **ThinkSystem SR665 V3** – The Lenovo ThinkSystem SR665 V3 is a 2-socket 2U server that features the AMD EPYC 9004 "Genoa" family of processors. With up to 96 cores per processor and support for the new PCIe 5.0 standard for I/O, the SR665 V3 offers the ultimate in two-socket server performance in a 2U form factor. For more information, see the [SR665 V3 product guide](#).
- **ThinkSystem SR645 V3** – The Lenovo ThinkSystem SR645 V3 is a 2-socket 1U server that features the AMD EPYC 9004 "Genoa" family of processors. With up to 96 cores per processor and support for the new PCIe 5.0 standard for I/O, the SR645 V3 offers the ultimate in two-socket server performance in a 1U form factor. For more information, see the [SR645 V3 product guide](#).
- **ThinkSystem SR670 V2** – The Lenovo ThinkSystem SR670 V2 is a versatile GPU-rich 3U rack server that supports eight double-wide GPUs including the new NVIDIA A100 and A40 Tensor Core GPUs, or the NVIDIA HGX A100 4-GPU offering with NVLink and Lenovo Neptune hybrid liquid-to-air cooling. The server is based on the new third-generation Intel Xeon Scalable processor family (formerly codenamed "Ice Lake"). The server delivers optimal performance for Artificial Intelligence (AI), High Performance Computing (HPC) and graphical workloads across an array of industries. For more information, see the [SR670 V2 product guide](#).
- **ThinkSystem SD650 V2** – The ThinkSystem SD650 V2 server is the next-generation high-performance server based on Lenovo's fourth generation Lenovo Neptune™ direct water cooling platform. With two third-generation Intel Xeon Scalable processors, the ThinkSystem SD650 V2 server combines the latest Intel processors and Lenovo's market-leading water cooling solution, which results in extreme performance in an extreme dense packaging, supporting your application From Exascale to Everscale™. For more information, see the [SD650 V2 product guide](#).
- **ThinkSystem SD650-N V2** – The ThinkSystem SD650-N V2 server is the next-generation high-performance GPU-rich server based on Lenovo's fourth generation Lenovo Neptune™ direct water cooling platform. With four NVIDIA A100 SXM4 GPUs and two third-generation Intel Xeon Scalable processors, the ThinkSystem SD650-N V2 server combines advanced NVIDIA acceleration technology with the latest Intel processors and Lenovo's market-leading water cooling solution, which results in extreme performance in an extreme dense packaging supporting your accelerated application From Exascale to Everscale™. For more information, see the [SD650-N V2 product guide](#).
- **ThinkSystem SR650 V2** – The Lenovo ThinkSystem SR650 V2 is an ideal 2-socket 2U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. The SR650 V2 is a very configuration-rich offering, supporting 28 different drive bay configurations in the front, middle and rear of the server and 5 different slot configurations at the rear of the server. This level of flexibility ensures that you can configure the server to meet the needs of your workload. For more information, see the [SR650 V2 product guide](#).
- **ThinkSystem SR630 V2** – The Lenovo ThinkSystem SR630 V2 is an ideal 2-socket 1U rack server

designed to take full advantage of the features of the 3rd generation Intel Xeon Scalable processors, such as the full performance of 270W 40-core processors, support for 3200 MHz memory and PCIe Gen 4.0 support. The server also offers onboard NVMe PCIe ports that allow direct connections to 12x NVMe SSDs, which results in faster access to store and access data to handle a wide range of workloads. For more information, see the [SR630 V2 product guide](#).

- **ThinkSystem SD530** – The Lenovo ThinkSystem SD530 is an ultra-dense and economical two-socket server in a 0.5U rack form factor. With up to four SD530 server nodes installed in the ThinkSystem D2 enclosure, and the ability to cable and manage up to four D2 enclosures as one asset, you have an ideal high-density 2U four-node (2U4N) platform for enterprise and cloud workloads. The SD530 also supports a number of high-end GPU options with the optional GPU tray installed, making it an ideal solution for AI Training workloads. For more information, see the [SD530 product guide](#).
- **ThinkSystem SD650** – The Lenovo ThinkSystem SD650 direct water cooled server is an open, flexible and simple data center solution for users of technical computing, grid deployments, analytics workloads, and large-scale cloud and virtualization infrastructures. The direct water cooled solution is designed to operate by using warm water, up to 50°C (122°F). Chillers are not needed for most customers, meaning even greater savings and a lower total cost of ownership. The ThinkSystem SD650 is designed to optimize density and performance within typical data center infrastructure limits, being available in a 6U rack mount unit that fits in a standard 19-inch rack and houses up to 12 water-cooled servers in 6 trays. For more information, see the [SD650 product guide](#).
- **ThinkSystem SR630** – Lenovo ThinkSystem SR630 is an ideal 2-socket 1U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. The SR630 server is designed to handle a wide range of workloads, such as databases, virtualization and cloud computing, virtual desktop infrastructure (VDI), infrastructure security, systems management, enterprise applications, collaboration/email, streaming media, web, and HPC. For more information, see the [SR630 product guide](#).
- **ThinkSystem SR650** – The Lenovo ThinkSystem SR650 is an ideal 2-socket 2U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. The SR650 server is designed to handle a wide range of workloads, such as databases, virtualization and cloud computing, virtual desktop infrastructure (VDI), enterprise applications, collaboration/email, and business analytics and big data. For more information, see the [SR650 product guide](#).
- **ThinkSystem SR670** – The Lenovo ThinkSystem SR670 is a purpose-built 2 socket 2U accelerated server, supporting up to 8 single-wide or 4 double-wide GPUs and designed for optimal performance required by both Artificial Intelligence and High Performance Computing workloads. Supporting the latest NVIDIA GPUs and Intel Xeon Scalable processors, the SR670 supports hybrid clusters for organizations that may want to consolidate infrastructure, improving performance and compute power, while maintaining optimal TCO. For more information, see the [SR670 product guide](#).
- **ThinkSystem SR950** – The Lenovo ThinkSystem SR950 is Lenovo's flagship server, suitable for mission-critical applications that need the most processing power possible in a single server. The powerful 4U ThinkSystem SR950 can expand from two to as many as eight Intel Xeon Scalable Family processors. The modular design of SR950 speeds upgrades and servicing with easy front or rear access to all major subsystems that ensures maximum performance and maximum server uptime. For more information, see the [SR950 product guide](#).
- **ThinkSystem SR655** – The Lenovo ThinkSystem SR655 is a 1-socket 2U server that features the AMD EPYC 7002 "Rome" family of processors. With up to 64 cores per processor and support for the new PCIe 4.0 standard for I/O, the SR655 offers the ultimate in single-socket server performance. ThinkSystem SR655 is a multi-GPU optimized rack server, providing support for up to 6 low-profile GPUs or 3 double-wide GPUs. For more information, see the [SR655 product guide](#).
- **ThinkSystem SR635** – The Lenovo ThinkSystem SR635 is a 1-socket 1U server that features the AMD EPYC 7002 "Rome" family of processors. With up to 64 cores per processor and support for the new PCIe 4.0 standard for I/O, the SR635 offers the ultimate in single-socket server performance. For more information, see the [SR635 product guide](#).
- **ThinkSystem SR645** – The Lenovo ThinkSystem SR645 is a 2-socket 1U server that features the AMD EPYC 7002 "Rome" family of processors. With up to 64 cores per processor and support for the new

PCIe 4.0 standard for I/O, the SR645 offers the ultimate in two-socket server performance in a space-saving 1U form factor. For more information, see the [SR645 product guide](#).

- **ThinkSystem SR665** – The Lenovo ThinkSystem SR665 is a 2-socket 2U server that features the AMD EPYC 7002 "Rome" family of processors. With support for up to 8 single-wide or 3 double-wide GPUs, up to 64 cores per processor and support for the new PCIe 4.0 standard for I/O, the SR665 offers the ultimate in two-socket server performance in a 2U form factor. ThinkSystem SR665 is a multi-GPU optimized rack server, providing support for up to 8 low-profile GPUs or 3 double-wide GPUs. For more information, see the [SR665 product guide](#).
- **ThinkSystem SR850** – The Lenovo ThinkSystem SR850 is a 4-socket server that features a streamlined 2U rack design that is optimized for price and performance, with best-in-class flexibility and expandability. The SR850 now supports second-generation Intel Xeon Scalable Family processors, up to a total of four, each with up to 28 cores. The ThinkSystem SR850's agile design provides rapid upgrades for processors and memory, and its large, flexible storage capacity helps to keep pace with data growth. For more information, see the [SR850 product guide](#).

China only:

- **ThinkServer SR660 V2** - The Lenovo ThinkServer SR660 V2 is an ideal 2-socket 2U rack server for SMB, large enterprises and cloud service provider that need industry-leading performance and flexibility for future growth. The SR660 V2 is based on the new 3rd generation Intel Xeon Scalable processor, with the new Intel Optane Persistent Memory 200 Series, the low latency NVMe SSD and the powerful GPU to support most customers workload such as databases, virtualization and cloud computing, virtual desktop infrastructure (VDI), infrastructure security, systems management, enterprise applications, collaboration/email, streaming media, web, and HPC. For more information, see the [SR660 V2 product guide](#).
- **ThinkServer SR590 V2** – The Lenovo ThinkServer SR590 V2 is an ideal 2-socket 2U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. The SR590 V2 is based on the 3rd generation Intel Xeon Scalable processor family (formerly codenamed "Ice Lake") and the Intel Optane Persistent Memory 200 Series. For more information, see the [SR590 V2 product guide](#).
- **WenTian WR5220 G3** – Lenovo WenTian WR5220 G3 is designed for customers of large, SMB enterprises and cloud service providers. It is 2-socket 2U rack server with excellent performance and high scalability. It is based on the 4th or 5th generation Intel Xeon Scalable processor family (codenamed "Sapphire Rapids", "Emerald Rapids") which can reach up to 385W TDP*, it can also support high-performance and high-frequency DDR5 memory, low latency NVMe SSD, and strong GPU performance to meet the most of customer workloads, such as databases, virtualization and cloud computing, AI, high-performance computing, virtual desktop infrastructure, infrastructure security, system management, enterprise applications, collaboration/email, streaming media, etc. For more information, see the [WR5220 G3 product guide](#).
- **Lenovo WenTian G7780 G3**
- **Lenovo WA5480 G**

Workstations:

- **ThinkStation P620** – The ThinkStation P620 workstation tower is equipped with abundant storage and memory capacity, numerous expansion slots, enterprise-class AMD Ryzen PRO manageability, and security features. With unprecedented visual computing powered by NVIDIA® professional graphics support, this eminently configurable workstation is equipped with up to two NVIDIA® RTX™ A6000 graphics cards with NVLink.
- **ThinkStation PX** - The Lenovo ThinkStation PX breaks the mold when it comes to workstations. Co-designed with Aston Martin, masters of racing car design and aerodynamics, the PX introduces an iconic, modular build that offers more accessibility than ever before. It's packed with ultra-high-end features, including the latest Intel® Xeon® scalable dual processors and support for up to four optional AI-enabled NVIDIA® RTX™ discrete graphics cards, plus heaps of high-speed DDR5 memory and PCIe storage.

Additional Lenovo ThinkSystem and System x servers and workstations may be compatible with LiCO. Contact your Lenovo sales representative for more information.

LiCO Implementation services

Customers who do not have the cluster management software stack required to run with LiCO may engage Lenovo Professional Services to install LiCO and the necessary open-source software. Lenovo Professional Services can provide comprehensive installation and configuration of the software stack, including operation verification, as well as post-installation documentation for reference. Contact your Lenovo sales representative for more information.

Client PC requirements

A web browser is used to access LiCO's monitoring dashboards. To fully utilize LiCO's monitoring and visualization capabilities, the client PC should meet the following specifications:

- Hardware: CPU of 2.0 GHz or above and 8 GB or more of RAM
- Display resolution: 1280 x 800 or higher
- Browser: Chrome (v62.0 or higher) or Firefox (v56.0 or higher) is recommended

Seller training courses

The following sales training courses are offered for employees and partners (login required). Courses are listed in date order.

1. VTT HPC: LiCO-Computing Orchestration for AI and HPC 2024-07-30 | 92 minutes | Employees Only

****NOTE:** To download the attached PPT, Launch the course, exit the player to return to this screen, then scroll down to find the PPT to download.**

Please view this session as Ana Irimiea, AI Systems and Solutions Product Manager at ISG, speaks with us about LiCO version 7.2.1.

She will talk about:

- Overview of LiCO
- Administrator and user capabilities
- Deployment options
- Ordering LiCO
- Roadmap

Tags: High-Performance Computing (HPC), Sales, Technical Sales

Published: 2024-07-30

Length: 92 minutes

Start the training:
Employee link: Grow@Lenovo

Course code: DVHPC214

2. **Enterprise Deployment of AI and Phases of Model Development**

2024-05-23 | 12 minutes | Employees and Partners

Lenovo Senior AI Data Scientist Dr David Ellison whiteboards the concepts of using data from multiple sources to derive customer benefits through Artificial Intelligence and LiCO (Lenovo Intelligent Computing Orchestration) software.

By the end of this training, you should be able to:

- Describe enterprise deployment of Artificial Intelligence
- Explain the process of model development
- State the purpose of LiCO (Lenovo Intelligent Computing Orchestration) software
- List three examples of Artificial Intelligence solutions

Tags: Artificial Intelligence (AI), Data & Analytics, Technical Sales

Published: 2024-05-23

Length: 12 minutes

Start the training:

Employee link: [Grow@Lenovo](#)

Partner link: [Lenovo 360 Learning Center](#)

Course code: DAIS101

3. **Selling Lenovo Intelligent Computing Orchestration**

2021-08-25 | 18 minutes | Employees and Partners

The goal of this course is to help ISG and Business Partner sellers understand Lenovo Intelligent Computing Orchestration (LiCO) software. Learn when and how to propose LiCO in order to continue the conversation with the customer and making a sale.

Tags: Artificial Intelligence (AI)

Published: 2021-08-25

Length: 18 minutes

Start the training:

Employee link: [Grow@Lenovo](#)

Partner link: [Lenovo 360 Learning Center](#)

Course code: DAIS202

Related links

For more information, see the following resources:

- LiCO website:
<https://www.lenovo.com/us/en/data-center/software/lico/>
- LiCO HPC/AI (Host) Support website:
<https://support.lenovo.com/us/en/solutions/HT507011>
- Technical LiCO Documentation:
<https://hpc.lenovo.com/users/lico/>
- Lenovo HPC & AI Software Stack Product Guide
<https://lenovopress.lenovo.com/lp1651-lenovo-hpc-ai-software-stack>
- Lenovo DCSC configurator:
<https://dcsc.lenovo.com>
- Lenovo AI website:
<https://www.lenovo.com/us/en/data-center/solutions/analytics-ai/>
- Lenovo HPC website:
<https://www.lenovo.com/us/en/data-center/solutions/hpc/>
- OpenHPC User Resources:
<https://github.com/openhpc/ohpc/wiki/User-Resources>
- Intel oneAPI:
<https://software.intel.com/content/www/us/en/develop/tools.html>
- Altair PBS Professional Documentation:
<https://www.altair.com/pbs-professional/>
- Lenovo Compute Orchestration in HPC Data Centers with Slurm
<https://lenovopress.lenovo.com/lp1701-lenovo-compute-orchestration-in-hpc-data-centers-with-slurm>
- Optimizing Power and Energy in HPC Data Centers with Energy Aware Runtime
<https://lenovopress.lenovo.com/lp1646.pdf>
- OpenLiCO
<https://github.com/lenovo/openlico>

Related product families

Product families related to this document are the following:

- [Artificial Intelligence](#)
- [High Performance Computing](#)

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

© Copyright Lenovo 2025. All rights reserved.

This document, LP0858, was created or updated on April 10, 2025.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
<https://lenovopress.lenovo.com/LP0858>
- Send your comments in an e-mail to:
comments@lenovopress.com

This document is available online at <https://lenovopress.lenovo.com/LP0858>.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at <https://www.lenovo.com/us/en/legal/copytrade/>.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Lenovo®

From Exascale to Everyscale®

Neptune®

System x®

ThinkServer®

ThinkStation®

ThinkSystem®

The following terms are trademarks of other companies:

AMD, AMD EPYC™, and AMD Ryzen™ are trademarks of Advanced Micro Devices, Inc.

Intel®, Intel Optane®, VTune®, and Xeon® are trademarks of Intel Corporation or its subsidiaries.

Linux® is the trademark of Linus Torvalds in the U.S. and other countries.

Microsoft® and Azure® are trademarks of Microsoft Corporation in the United States, other countries, or both.

IBM®, IBM Spectrum®, and LSF® are trademarks of IBM in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.