Cloudera Data Science Workbench 1.10.4

# Cloudera Data Science Workbench Runtimes Release Notes

**Date published: 2020-07-16**
**Date modified: 2023-01-31**

# CLOUDΞRA

**https://docs.cloudera.com/**

# Legal Notice

# Contents

# ML Runtimes What's New

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2024.02.1.

> **Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

New Features

- Pandas is now included in the Python 3.8 and higher Runtimes.
- You can now see hidden files in the Jupyterlab file browser.

Improvements

- Multiple small improvements and bug fixes.
- Multiple CVE fixes.
- Added libraries to improve vision and audio processing model support.

**Related Information**

ML Runtimes Nvidia GPU Edition

ML Runtimes 2021.02

Upgrading R and Python Packages

ML Runtimes Environment Variables

## ML Runtimes Version 2023.12.1

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2023.12.1.

> **Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

New Runtimes

- Python 3.11 JupyterLab Standard
- Python 3.11 JupyterLab Nvidia GPU

> **Note:** Python 3.11 Runtimes are not compatible with Spark versions prior to Spark 3.4.

Improvements

- Python maintenance versions - Python maintenance versions have been upgraded to 3.8.18, 3.9.18, and 3.10.13.

Discontinued Runtimes

The following runtimes are no longer included.

- R 3.6 Workbench
- R 4.0 Workbench
- R 4.1 Workbench

Fixed Issue

Workbench editor for Python 3.8 and 3.9 now can parse multiline strings properly. (DSE-27222)

## ML Runtimes Version 2023.08.2

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2023.08.2.

**Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

Improvements

- CVE fixes - This maintenance release includes numerous improvements related to Common Vulnerability and Exposures (CVE).

## ML Runtimes Version 2023.08.1

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2023.08.1.

**Note:** New ML Runtime releases are automatically added to the deployment, if an internet connection is available.

New Features / Improvements

- New Runtimes - Python 3.10 Standard Workbench and R 4.3 Standard Workbench are included in this release.

Fixed Issues

- Chunker (DSE-27222) - Fixed an issue where the chunker failed to correctly parse multiline string inputs in classic Workbench Runtimes.
- Jupyter (DSE-28134) - `jupyter-kernel-gateway` has been upgraded to version 2.5.2. As a result, users cannot install incompatible versions of `jupyter-client` breaking their Python environments.
- Security - Various security fixes were implemented.

## ML Runtimes Version 2023.05.2

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2023.05.2.

**Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

Fixed Issues

- This maintenance release is to automatically update those ML Runtimes versions that were referenced with a wrong image identifier when ML Runtimes 2023.05.1 was released.

ML Runtime Releases included in CDSW versions

- Runtime Release 2023-05 included in CDSW 1.10.1, 1.10.2, 1.10.3
- Runtime Release 2022.11 included in CDSW 1.10.1, 1.10.2, 1.10.3
- Runtime Release 2022.04 included in CDSW 1.10.1, 1.10.2, 1.10.3
- Runtime Release 2021.12 included in CDSW 1.10.1, 1.10.2, 1.10.3
- Runtime Release 2021.09 included in CDSW 1.10.0
- Runtime Release 2021.06 included in CDSW 1.10.0
- Runtime Release 2021.02 included in CDSW 1.10.0
- Runtime Relase 2020.11 inclued in CDSW 1.9.x

# ML Runtimes Version 2023.05.1

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2023.05.1.

**Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

New features

- This release of ML Runtimes includes the following features that improve support for Large Language (LLMs) within CML:

  - Git Large File Storage (LFS) support is now included in all ML Runtimes images. Users can now easily download Large Language models via Git.

    LFS replaces large files such as audio samples, videos, datasets, and graphics with text pointers inside Git, while storing the file contents on a remote server like GitHub.com or GitHub Enterprise.
  - Upgraded CUDA to version 11.8.0 in JupyterLab Python CUDA runtimes. This change allows customers to use the latest pytorch version in CML. Affected Python versions are 3.7, 3.8, 3.9, and 3.10.
  - wget is now installed on all ML Runtimes. This enables users another way to download large language models in CML in addition to git-lfs.
- This release introduces a new Technical Preview Runtime, adding support for Anaconda, based on the JupyterLab editor. Conda environments can be created, modified, and activated from the Terminal without any limitations. For more information, see Using Conda Runtimes.
- New Python 3.10 Runtimes are available.
- Added the following extensions to our JupyterLab Runtimes (Python 3.8, 3.9, 3.10): Language Server Protocol (jupyterlab-lsp) and git (jupyterlab-git). Language Server Protocol integration provides coding assistance for JupyterLab (code navigation, hover suggestions, linters, autocompletion, rename).
- Added the git (jupyterlab-git) extension to our JupyterLab Runtimes (Python 3.7).
- Implya client is now updated to version 0.18.0 which has support for authentication via JWT token.

Version 2023.05.1

Fixed Issues

- DSE-26724 wget is now installed on all ML Runtimes. This enables users to download large language models in CML two ways: via Gi and via wget.
- DSE-25078 The OS package cmake is now installed for all ML Runtime images. This change enables users to install packages like nloptr.
- DSE-23762 Updated installed R package versions to make sparklyr R package usable in the same session as it was installed in.
- DSE-23751 Upgraded Jupyter to version 3.6.0 in all JupyterLab Runtimes.
- DSE-23613 Upgraded CUDA to version 11.8.0 in JupyterLab Python CUDA runtimes. This change was needed to allow customers to use the latest pytorch version in CML. Affected Python versions are 3.7, 3.8, 3.9, and 3.10.
- DSE-22425 Added OS packages to R Runtimes to support recent versions of the "devtools" package.
- DSE-20234 Jupyter session-specific data is no longer persisted in the Project directory. This prevents errors when starting multiple Jupyter sessions.

# ML Runtimes Version 2022.11.2

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2022.11.2.

**Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

Version 2022.11.2

Fixed Issues

- Multiple CVEs have been fixed.

# ML Runtimes Version 2022.11

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2022.11.

**Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

Version 2022.11

New features

N/A

Fixed Issues

- DSE-22433 Upgrade numpy version in Python runtimes to resolve issues.
- DSE-22374 Reduced log level to WARN log level can be manipulated through the HADOOP_ROOT_LOGGER environment variable. This is set to WARN,console in order to keep console logging, while reducing log level from INFO to WARN.
- DSE-22251 Add flextable dependencies to R Runtimes Added flextable dependencies (libfontconfig1-dev, libcairo2-dev) to R runtimes. Added two OS libraries so flextable can be installed and used in CML.

Known issues

- The CUDA version (11.4.1) being used to build Nvidia GPU Runtimes is not supported by newer Torch versions (1.13+).

  Workaround: Use Torch versions up to 1.12.1.

  Cloudera Bug: DSE-24038
- When installing sparklyr for R, the installation will succeed, but importing the library in the same session will fail. This happens, because earlier versions of dependent packages (vctrs, lifecycle, rlang, cli) are already loaded in the R session, therefore newer versions cannot be loaded.

  Workaround: Start a new session, import and use sparklyr there.

  Cloudera Bug: DSE-23762

# ML Runtimes Version 2022.04

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2022.04.

**Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

Version 2022.04

New features

- The Global Registry for npm is set to https://registry.npmjs.org/ in all of our runtimes. With this change, customers can use npm install in their Dockerfiles, when they build custom runtimes on top of ours.

> **Note:** RAPIDS Runtimes are now an optional extension of the Cloudera ML Runtime distribution. The RAPIDS images will no longer be distributed automatically. Administrators can Register them in the Runtime Catalog.

Fixed issues

- DSE-2064 Upgrade of Python Kernels across of all ML Runtimes:

  - Python 3.7.12 -> 3.7.13
  - Python 3.8.12 -> 3.8.13
  - Python 3.9.7 -> Python 3.9.11

# ML Runtimes Version 2021.12

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2021.12.

> **Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

Version 2021.12

New features

- New versions provided for all ML Runtimes.
- Added R 4.1 Runtimes
- RAPIDS base image was updated to use 21.12 RAPIDS release (DSE-19745)

> **Note:** Python 3.6 ML Runtimes are not longer provided due to Python 3.6 end of life

Fixed issues

- DSE-17817 psycopg2 now can be installed for Models.
- DSE-17547 Fixed a matplotlib warning in RAPIDS Runtimes.
- DSE-10146 Fixed issue causing R workbench to crash due to an illegal character.
- DSE-17657 Upgrade jupyterlab and notebook for CVE-2021-32797 and CVE-2021-32798.
- Various security and library version updates
- DSE-19805 Fixed CVE-2022-22817 critical vulnerability in Python Runtimes

# ML Runtimes Version 2021.12

Major features and updates for ML Runtimes.

This release is available with ML Runtimes version 2021.12.

> **Note:** New ML Runtime releases are automatically added to the deployment, if internet connection is available.

Version 2021.12

New features

- New versions provided for all ML Runtimes.
- Added R 4.1 Runtimes
- RAPIDS base image was updated to use 21.12 RAPIDS release (DSE-19745)

**Note:** Python 3.6 ML Runtimes are not longer provided due to Python 3.6 end of life

Fixed issues

- DSE-17817 psycopg2 now can be installed for Models.
- DSE-17547 Fixed a matplotlib warning in RAPIDS Runtimes.
- DSE-10146 Fixed issue causing R workbench to crash due to an illegal character.
- DSE-17657 Upgrade jupyterlab and notebook for CVE-2021-32797 and CVE-2021-32798.
- Various security and library version updates
- DSE-19805 Fixed CVE-2022-22817 critical vulnerability in Python Runtimes

# ML Runtimes Version 2021.09.02

Major features and updates for ML Runtimes.

Version 2021.09.02

Fixed issues

- DSE-19496 The log4j build dependency is removed for the Scala ML Runtime. Once this ML Runtimes version is deployed, the user interface will always point to this latest version of the 2021.09.02 Scala Runtime to run workloads using the Scala kernel (until a newer version is released). Users will not be able to select the originally released 2021.09 Scala Runtime.

# ML Runtimes Version 2021.09

Major features and updates for ML Runtimes.

Version 2021.09

New features

- Python 3.9 - ML Runtimes now support Python 3.9 kernels.
- Scala - Scala kernel is supported for running Sessions and Jobs.
- CUDA 11.4 - Nvidia GPU Editions are using CUDA 11.4.

Fixed issues

- DSE-17126 Starting a worker from Runtime session will create a worker with engine image

    This issue is resolved in ML Runtimes 2021.09 when used with the latest ML Workspace versions.

    For workers to function properly with ML Runtimes, please use ML Runtimes 2021.09 or later with CML Workspace version of 2.0.22 or later.

# ML Runtimes Version 2021.06

Major features and updates for ML Runtimes.

Version 2021.06

New features

- Impyla - Impyla is now preinstalled in standard Python Runtimes.

    Cloudera bug: DSE-14293
- Runtimes support for Models - Runtimes now support running models.

    Cloudera bug: DSE-14294

- JupyterLab shutdown_no_activity_timout configurable - JupyterLab shutdown_no_activity_timout is now configurable through environment variable.

  Cloudera bug: DSE-14817

Fixed issues

- DSE-15191 - Fixed installation issue with thrift-sasl python package.
- DSE-14886 - Fixed an issue preventing matplotlib installation into project.
- DSE-14737 - Fixed warnings appearing in the log due to using JupyterLab 3.
- DSE-15215 - Fixed an issue due to decorators that could not be used with Python 3.8 and Workbench editor.

# ML Runtimes Version 2021.04

Major features and updates for ML Runtimes.

Version 2021.04

New features

- RAPIDS Runtimes - The RAPIDS Edition Runtimes are built on top of community built RAPIDS docker images. The RAPIDS suite of software libraries relies on NVIDIA CUDA primitives for low-level compute optimization, but exposes that GPU parallelism and high-bandwidth memory speed through user-friendly Python interfaces.

Fixed issues

The following fixed issues relate only RAPIDS Runtimes.

- DSE-13743 - Idle JupyterLab sessions are now ended after around     IDLE_MAXIMUM_MINUTES. (See *ML Runtimes Environment Variables Environment Variables*).

  > **Note:** These sessions may run for an additional 5 minutes after  IDLE_MAXIMUM_MINUTES. (DSE-13743)

- DSE-14979 - matplotlib figures restore the styling used in engine:13 (MLRuntimes 2020.11 used the matplotlib defaults)
- DSE-12881 - For Python runtimes, py4j is now installed.
- Security fixes, python library version updates

# ML Runtimes Version 2021.02

Major features and updates for ML Runtimes.

Version 2021.02

New features

- Nvidia GPU support - Runtimes supporting Nvidia GPUs are available with preinstalled CUDA software for both Workbench and JupyterLab editors. (See *ML Runtimes Nvidia GPU Edition*.)
- Runtimes supporting R 3.6 and R 4.0 available for Workbench editor

  > **Note:** User R libraries are installed in /home/cdsw/.local/lib/R rather than /home/cdsw/R, which is where they were installed in engine:13. This change was made to support simultaneous use of R 3.6 and R 4.0. Because of this change you will need to reinstall user R packages when migrating an existing project from engine:13 to MLRuntimes 2021.02, or from R 3.6 to R 4.0

  Cloudera bug: DSE-3686

- Support for JupyterLab runtimes - The JupyterLab runtimes now use JupyterLab 3.0 (was 2.2) and are considered generally available using Standard Edition ML Runtimes. See JupyterLab blog for notes on this upgrade.

- Python runtimes include a C++ compiler - Python runtimes now include a C++ compiler (g++), which enables the installation and use of certain Python libraries such as impyla and pystan.

  Cloudera bug: DSE-14492

- Preinstalled Python libraries (See *ML Runtimes 2021.02*) are now installed in /usr/local/lib rather than /var/lib/ cdsw, which is where they were installed in engine:13 and runtimes 2020.11 (DSE-12177). This means that you can upgrade these packages more easily, but there are some packages you should not upgrade. (See *Upgrading R and Python Packages*)

- ML runtimes are supported on CDSW (version 1.9.0 or higher).

Fixed issues

- DSE-13743 - Idle JupyterLab sessions are now ended after around     IDLE_MAXIMUM_MINUTES. (See *ML Runtimes Environment Variables Environment Variables*).

  > **Note:**  These sessions may run for an additional 5 minutes after  IDLE_MAXIMUM_MINUTES. (DSE-13743)

- DSE-14979 - matplotlib figures restore the styling used in engine:13 (MLRuntimes 2020.11 used the matplotlib defaults)
- DSE-12881 - For Python runtimes, py4j is now installed.
- Security fixes, python library version updates

## ML Runtimes Version 2020.11

Release notes and fixed issues

Version 2020.11

New features

- New Runtimes versions added

  > **Note:**  Nvdia GPU Edition comes with CUDA 11.1 preinstalled.

Fixed issues

- DSE-13904 - Fixed an issue where CDSW workspace installation may take up to 10 minutes longer when the Autoscaling Group minimum is set to 0.
- DSE-13898 - Updated TGT image to fix an issue so that the freeIPA HA release retries on failure.
- CDPSDX-2673 - Added a Retry step to the login context to reduce the chance of PRE_AUTH failures.

# ML Runtimes Known Issues and Limitations

You might run into some known issues while using ML Runtimes.

### Packages may not load in a session

When installing R or Python packages in a Session, the kernel might not be able to load the package in the same session, if a previous version of the package or its newly installed dependencies have been loaded in the same Session.

Workaround: Start a new session, import and use the newly installed package there.

## Python Runtimes in CML fail to import the setuptools Python library and can fail installing some Python packages

Python Runtimes in CML fail to import the setuptools Python library and therefore can fail installing some Python packages when the library setuptools is present on the Runtime or is installed into the CML project with version 60.0.0 or higher.

Python 3.10 Runtimes from the 2023.05 Runtime release ship with a newer version of setuptools, so customers can run into this issue when they are using that Runtime. Also they can run into this issue when they are using Custom Runtimes that has a newer setuptools library version or when they install a new setuptools version into their project (regardless of what Runtime they use).

Workaround: Set the environmental variable SETUPTOOLS_USE_DISTUTILS=stdlib either on a project level under Project Settings -> Advanced or on a workspace level under Site Administration -> Runtime -> Environment variables.

## DSE-2804 PBJ CUDA Runtimes have wrong metadata version

PBJ NVidia GPU Runtimes Released in version 2023.05.1 had wrong runtime metadata version. Future releases from July, 2023 of CML Public Cloud might not be compatible with these Runtimes.

Workaround: Follow the instructions in TSB-678.

## ML Runtimes installation might pull wrong metadata

For ML Runtimes 2023.05.1 only. Check your installation to determine if your ML Runtimes 2023.05 installation pulled the wrong metadata:

1. Open Runtime Catalog.
2. Check if newer runtime variants have been downloaded, e.g. 'JupyterLab - Conda - Tech Preview'.
3. Open the version list for this Runtime.
4. Check if any of the versions' details page shows 'container.repository.cloudera.com...' under 'Runtime Image' section.
5. If the versions' Details page shows 'container.repository.cloudera.com..., complete the Workaround steps.

Workaround: This workaround applies for CML Public Cloud 2022.06.1 (CDSW-2.0.32) or CDSW 1.10.1 or higher versions.

Administrator should perform the following steps:

1. Go to the Runtime Catalog page.
2. On the right-hand side, open the version list of each Runtime variant.
3. Open the Details page for the 2023.05.1-b4 versions.
4. Use the three-dot menu and use the Set to Disabled function for those versions which have container.repository.cloudera.com... as Runtime Image.
5. Recheck within 24 hours that the runtimes with docker.repository.cloudera.com have been added.

Regular user should perform the following steps:

1. If the project does not have ML Runtimes 2023.05 after the previous actions has been taken by the Administrator, no action is required.
2. If the project does have ML Runtime 2023.05, after the Administrator disables the wrong one, it should display as 'Disabled'.
3. When the runtimes with the right metadata has been added automatically to the installation, use the Add Latest button to add the proper runtime(s).
4. Remove the disabled runtimes from the project.

Active workloads are not affected, as with the wrong versions no workload could be started.

### DSE-9818 JupyterLab Conda Tech Preview Runtime

**Sessions**

When starting a Notebook or a Console for a specific environment, the installed packages will be available and the interpreter used to evaluate the contents of the Notebook or Console will be the one installed in the environment. However, the Conda environment is not "activated" in these sessions, therefore commands like !which python will return with the base Python 3.10 interpreter on the Runtime. The recommended ways to modify a Conda environments or install packages are the following:

- conda commands must be used with the -n or --name argument to specify the environment, for example conda -n myenv install pandas
- When installing packages with pip, use the %pip magic to install packages in the active kernel's environment, for example %pip    install pandas

**Applications and Jobs**

To start an Application or Job, first create a launcher Python script containing the following line: ! source activate <conda_env_name> && python <job    / application script.py>

When starting the Application or Job, select the launcher script as the "Script".

**Models**

Models are currently not supported for the Conda Runtime.

**Spark**

Spark is not supported in JupyterLab Notebooks and Consoles.

Spark workloads are supported in activated Conda environments in JupyterLab Terminals, or in Jobs or Applications.

The CDSW libraries for Python and R are not available for the Conda Runtimes.

### Adding a new ML Runtimes when using a custom root certificate might generate error messages

When trying to add new ML Runtimes, a number of error messages might appear in various places when using a custom root certificate. For example, you might see: "Could not fetch the image metadata" or "certificate signed by unknown authority". This is caused by the runtime-puller pods not having access to the custom root certificate that is in use.

Workaround:

1. Create a directory at any location on the master node:

   For example:

   mkdir -p /certs/

2. Copy the full server certificate chain into this folder. It is usually easier to create a single file with all of your certificates (server, intermediate(s), root):

   ```
   # copy all certificates into a single file:
   cat server-cert.pem intermediate.pem root.pem > /certs/cert-chain.crt
   ```

3. (Optional) If you are using a custom docker registry that has its own certificate, you need to copy this certificate chain into this same file:

   ```
   cat docker-registry-cert.pem >> /certs/cert-chain.crt
   ```

4. Copy the global CA certificates into this new file:

   ```
   # cat /etc/ssl/certs/ca-bundle.crt >> /certs/cert-chain.crt
   ```

**5.** Edit your deployment of runtime manager and add the new mount.

Do not delete any existing objects.

#kubectl edit deployment    runtime-manager

**6.** Under VolumeMounts, add the following lines.

Note that the text is white-space sensitive - use spaces and not tabs.

```
- mountPath: /etc/ssl/certs/ca-certificates.crt
    name: mycert
    subPath: cert-chain.crt #this should match the new file name created in
  step 4
```

Under Volumes add the following text in the same edit:

```
- hostPath:
    path: /certs/  #this needs to match the folder created in step 1
    type: ""
name: mycert
```

**7.** Save your changes:

wq!

Once saved, you will receive the message "deployment.apps/runtime-manager edited" and the pod will be restarted with your new changes.

**8.** To persist these changes across cluster restarts, use the following Knowledge Base article to create a kubernetes patch file for the runtime-manager deployment: https://community.cloudera.com/t5/Customer/Patching-CDSW-Kubernetes-deployments/ta-p/90241

Cloudera Bug:  DSE-20530

### Spark Runtime Add-on required for Spark 2 integration with Scala Runtimes

Scala Runtimes on CML require Spark Runtime Addon to enable Spark2 integration. Spark3 is not supported with the Scala Runtime.

### DSE-17981 - Disable Scala runtimes in models, experiments and applications runtime selection

Scala Runtimes should not appear as an option for Models, Experiments, and Applications in the user interface. Currently Scala Runtimes only support Session and Jobs.