

Machine Learning 1.5.2

Release Notes

Date published: 2020-07-16

Date modified: 2023-10-31

CLOUDERA

<https://docs.cloudera.com/>

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

What's New.....	4
Known Issues.....	4
Fixed Issues.....	6
Behavioral Changes.....	7

What's New

CDP Private Cloud 1.5.2 includes the following features for Cloudera Machine Learning.

New features

Support for Buildkit on ECS

Switching from docker to BuildKit on RHEL 8.x and CentOS 8.x for building models and experiments images.

Quota Management at workspace level (Tech Preview)

Enables you to control how resources are allocated within your CML workspace.

Parallel Rsync in CDSW to CML migration tool

Added parallel rsync to CDSW to CML migration tool.

Related Information

[Known Issues and Limitations](#)

Known Issues

You might run into some known issues while using Cloudera Machine Learning on Private Cloud.

On-premise CDSW incompatibility with Python 3-based Cloudera Manager

Not all versions of the on-premise Cloudera Data Science Workbench (CDSW) are compatible with Python 3-based Cloudera Manager (CM). As Private Cloud Data Services (PvC DS) 1.5.2 requires CM 7.11.3 which is Python 3-based, it is recommended not to upgrade an older version of CM to 7.11.3 (or other Python 3-based CMs) that has a running instance of on-premise CDSW.

Upgrade CDSW to CDSW 1.10.5, as it is Python 3-based and works with Python 3 CM. Please refer to CDSW 1.10.5 documentation for more details.

DSE-32887: Runtime Updates can not be enabled from the Site administration UI on PVC

With this bug present on PVC 1.5.2 and 1.5.3, Runtime repositories added on the Site Administration>Runtimes page will not be scanned for new Runtimes. Adding Runtime repositories to CML was a new feature in 1.5.2. Unfortunately with this bug, the feature is completely unusable.

Manually enable Runtime updates with an API call using the legacy API key of the admin user. That will let CML check the registered runtime registries for new runtimes every 24 hours. The syntax of the API call is as follows:

```
curl 'https://<CML_URL>/api/v1/site/config' \  
-X 'PATCH' \  
-u <ADMIN_APIv1_KEY>: \  
-H 'content-type: application/json' \  
-d '{"enable_runtime_updates":true}'
```

DSE-33636: Workloads unable to start up after changing default hadoopCLI addon

Changing the default Hadoop CLI Runtime Addon causes jobs, models, and application workloads to be unable to start up.

1. Open affected workload settings.
2. Update the workload (this updates the Hadoop CLI Addon associated with the workload to the default one.)

3. For Jobs: update.
4. For Applications: update and restart.
5. For Models: deploy a new build.

Please see [Machine Learning Site administration](#)>Disable Addons for related tasks.

ENGSEC-22794: CML Sessions fail randomly with status 34

CML Sessions fail randomly with status 34 for fresh CML PVC 1.5.2 workspace installs with Base cluster timezone set to a non-UTC timezone.

Edit the CML workspace web deployment to force the timezone to UTC:

```
kubectl edit deployments web -n mlx

- command:
  - /usr/bin/bash
  - -c
  - TZ=UTC /home/cdswint/services/web/pod-bootstrap.sh 2>&1 | tee /var/fluentbit/log/web.log
```

Note the TZ=UTC before pod-bootstrap.sh command

OPSX-4603: Buildkit in ECS in CML PrivateCloud

Issue: This release introduces BuildKit in ECS for building images of models and experiments. BuildKit is a replacement for Docker, which was previously used to build images of CML's models and experiments in ECS. Buildkit will be only for OS RHEL8.x and CentOS 8.x.

Buildkit in CML Private Cloud 1.5.2 is a Tech Preview feature. Hence, having Docker installed on the nodes/hosts is still mandatory for models and experiments to work smoothly. Upcoming release will be completely eliminating the dependency of Docker on the nodes.

None.

DSE-32285: Migration: Migrated models are failing due to image pull errors

Issue: After CDSW to CML migration (on-premises) via full-fledged migration tool, migrated models on CML Private Cloud Workspace fails on initial deployment. This is because the initial model deployment tries to pull images from on-premises's registry.

: Re-deploy the migrated model. As this involves the build and deploy process, the image will be built, pushed to the CML Private Cloud Workspace's configured registry, and then the same image will be consumed for further usage.

DSE-28768: Spark Pushdown is not working with Scala 2.11 runtime

Issue: Scala and R are not supported for Spark Pushdown.

None.

DSE-32304 On CML Private Cloud ECS terminal and ssh connections can terminate

Issue: In Private Cloud ECS, CML Terminal and SSH connections can terminate after an uncertain amount of time, usually after 4-10 minutes. This issue affects customers using local IDEs to work with CML, as well as any customer application using a websocket connection.

None.

Related Information

[Known Issues and Limitations](#)

Fixed Issues

This section lists issues fixed in this release of Cloudera Machine Learning on Private Cloud.

DSE-28066: Port conflict on Single Node

CML UI shows timeout error during upgrade on a single node cluster because of an underlying port conflict error: 1 node(s) didn't have free ports for the requested pod ports.

This issue has been fixed.

DSE-27283: Pod evaluator should not use hostNetworking on Azure and Private Cloud

CML workspace Pod Evaluator used host networking which could cause HA issues and deployment failures owing to port conflicts.

This issue has been fixed.

DSE-28005: Environment variables hidden on Application>Settings page

Previously, environment variables were not visible or modifiable on Application's settings page.

This issue has been fixed.

DSE-32672: Timezone issue for start, stop and schedule session activities

Earlier, the saved date value for sessions was not set to the UTC timezone. This issue caused the sessions to terminate for Private Cloud Data Services if the web pod was running in a timezone different from UTC.

This issue has been fixed.

DSE-34100: Migration readiness check fails on BuildKit-enabled clusters

Earlier, the migration readiness checks failed on clusters having BuildKit instead of Docker.

This issue has been fixed.

DSE-33099: CML team members' type value could not be configured

Earlier, the team members' type value could not be edited, and team members with the value none could not contribute to synchronized team activities. All team members need to have write permissions so that the roles can be set from the UI.

This issue has been fixed.

DSE-31914: The model metrics ingress configuration does not support TLS

The model metrics ingress configuration did not have a template for specifying the TLS section.

This issue has been fixed.

DSE-32250: The spark executor resource usage is not reported correctly

Earlier, when you exported the Spark executor usage data from the Site Admin > Usage tab, the exported data in the report was the same as the session data the executor was launched in. This included incorrect CPU, memory, create time, and stop time details of the Spark executors.

This issue has been fixed.

DSE-33098: Unable to synchronize teams in CML without having an MLAdmin user in that team

Earlier, CML required that each team had one user with the MLAdmin role, by default, to manage the team, even if the role was not necessary for the team later. It is now possible to manage the team with the MLUser role without having the MLAdmin role in the team.

This issue has been fixed.

DSE-33765: CML workspace upgrade fails due to a wrong logic for getting helm values

Earlier it was wrongly assumed that the PersistentVolumeClaim (PVC) requests are returned from K8 in the same order. This wrong logic could lead to the CML upgrade being broken. To avoid this, the returned list shall rather be searched for the particular volume instead.

This issue has been fixed.

DSE-34088: Unable to view Spark UI executor logs in CML

The names for Spark executor pods changed in the dashboard pods DB table.

This issue has been fixed.

DSE-34112: Migration readiness timeout logic is required to avoid the migration being stuck

A timeout value is implemented to exit the migration readiness workflow.

This issue has been fixed.

Behavioral Changes

Learn about behavioral changes in Cloudera Machine Learning.

The minimum length for environment name has changed

When creating new CML Workspaces, the minimum for environment name length has changed from 4 to 5 characters. Workspaces may fail with the error Invalid Argument: Environment name must only include lowercase letters, numbers, and hyphens and must be between 5 and 28 characters.

commitProtocolClass optimised for S3 blobstore

CML Spark session generates a default spark-defaults.conf in /etc/spark/conf with the following commitProtocolClass optimised for Cloud blobstore, which however does not support dynamicPartitionOverwrite mode:

```
spark.sql.sources.commitProtocolClass=org.apache.spark.internal.io.cloud.PathOutputCommitProtocol
```

- Create a file named spark-default.conf in the Project and add the following configurations:

```
spark.sql.parquet.output.committer.class=org.apache.parquet.hadoop.ParquetOutputCommitter
```

```
spark.sql.sources.commitProtocolClass=org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol
```