Machine Learning 1.5.1

Release Notes

CLOUDERA

Legal Notice

Contents

What's New

CDP Private Cloud 1.5.1 includes the following features for Cloudera Machine Learning.

New features

CDSW to CML Migration GA

> CML installations on CDSW can now be migrated to CML workspaces.

Backup/Restore on PVC TP

> Backup and Restore of workspaces is now available on Private Cloud, for both OCP and ECS clusters. This feature allows taking backups of workspaces and rolling back to earlier versions of a workspace.

Model Registry

> The Model Registry stores and manages machine learning models and associated metadata, such as the model's version, dependencies, and performance. The registry enables MLOps and facilitates the development, deployment, and maintenance of machine learning models in a production environment.

Disable Engines

> Starting with version 1.5.1, Legacy Engines are disabled by default on the workspace level. On upgraded workspaces, selecting this option disables Legacy Engines and changes all existing project types to ML Runtime.

Migration Preflight checklist

> Added the capability to validate the pre-reqs checklist before migration to ensure prerequisites are met before proceeding with migration.

Iceberg v2 Support

> Iceberg v2 is now supported in a new Runtime Addon. See the CDE Spark documentation for support details.

Spark 3 pushdown

> Spark 3 pushdown to CDP Base functionality is supported.

Ozone addon integration

> Ozone object storage is now supported as a backend storage provider.

Free IPA certification

> All CML functionality is certified on clusters using Free IPA, for both OCP and ECS clusters.

Related Information

Known Issues and Limitations

Known Issues

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

(DSE-30602) CML Deamonsets are scheduled on Master/Infra nodes (non-worker nodes)

CML workspace installation fails because the LiveLog Publisher pod cannot get scheduled on master or infra nodes.

CML daemonset pods like LiveLog Publisher can be only scheduled on nodes labelled `worker`. They are not supported on master or infra nodes.

A LiveLog daemonset can be scheduled on master or infra nodes as well as on worker nodes if the custom labels are different than the expected `worker` label. Use this workaround to make sure LiveLog deamonsets get scheduled on worker nodes only.

```
oc patch daemonset livelog-publisher -p '{"spec": {"template":
       {"spec": {"nodeSelector": {"type": "<custom-label>"}}}}}'
```

From Private Cloud version 1.5.1, the CML Workspace installation timeout is increased to 60 mins. Hence, it is strictly advisable to perform the above work-around within 60 mins of initiating CML workspace creation.

(DSE-28621) db-0 pod crash-loops with error

When upgrading an ML workspace from version 1.4.1 to 1.5.1, the following errors may occur:

- CDSW_PG FATAL: lock file "postmaster.pid" already exists
- CDSW_PG HINT: Is another postmaster (PID 7) running in data directory

This happens only when the db-migrate-xxx job is still in the Running state, and is trying to access to the embedded database, and fails with the error:

```
/usr/pgsql-12/bin/pg_isready -h db.host.local -p 5432
db.host.local:5432 - no response
shell-logger Migrations.Run Debug 'Database not ready'
sleep 2
```

Workaround: Follow these steps to resolve these issues.

1. Delete the db-migrate-xxx job from jobs. The db pod will recover automatically within 5 to 10 minutes.

   ```
   kubectl delete jobs db-migrate-2.0.39-b73-xxx -n <ML-Workspace-Namespace>
   ```

2. If the db pod fails to come up, delete the postmaster.pid file by mounting the database to a temporary pod and delete the lock file (postmaster.pid).
3. In ML Workspaces, retry the upgrade.

Pods fail to come up on single node

After upgrading a workspace, if the pods fail with a FailedMount error, then restart the instance manager engine. The replica pods, Longhorn manager, and volumes will be attached.

If the pod-evaluator pod fails with the following error:

```
Warning  FailedScheduling  34m   yunikorn  0/1 nodes are available: 1 node(s
) didn't have free ports for the requested pod ports.
```

Delete the previous instance of the running pod-evaluator. This error is caused by a port conflict on the single node.

(DSE-26696) Atlas integration does not work in Private Cloud

If the integration is not working, it may be due to the new Oracle JDK version 1.8.351 which disables weaker encryption types. If customers upgrade to this version, Kerberos keytab files may be prevented from being picked up, due to the weaker cryptography used for keytab files.

To correct this problem, you need to enable the enable_weak_crypto option in Cloudera Manager:

1. In Cloudera Manager, go to Administration > Settings.
2. In the search box, start typing Advanced Configuration Snippet (Safety Valve) for [libdefaults] section of krb5.conf file until you see the corresponding text box.
3. In the text box, enter enable_weak_crypto = true
4. Click Save Changes.

The setting takes effect immediately.

Do not use backtick characters in environment variable names

Avoid using backtick characters (`) in environment variable names, as this will cause sessions to fail with exit code 2.

Model Registry is not supported on R models

Model Registry is not supported on R models.

After upgrading a workspace, the pod evaluator is in CrashLoopBackOff state

This typically happens after a workspace upgrade. The solution is to upgrade the workspace as soon as the Control Plane is upgraded.

Migration from CDSW fails if Enable Model Metrics is not selected

CDSW to CML migration fails if you do not select the Enable Model Metrics option.

Workaround: In Cloudera Machine Learning Workspaces, Provision Workspace (CDSW Migration), select the Enable Model Metrics option:
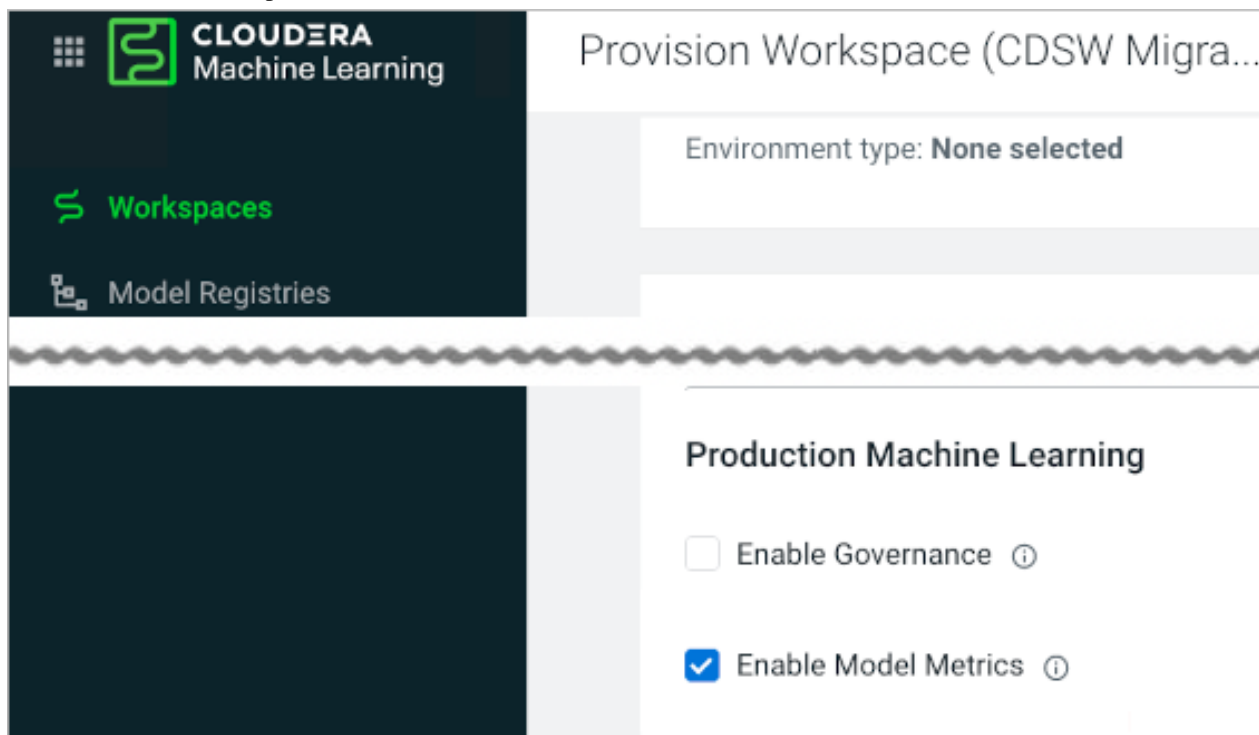
DSE-24690 CDSW to CML migration fails with the Cloudera default docker repository

During CDP Private Cloud installation, you are prompted to configure the Docker repository that Cloudera uses to deliver CDP Private Cloud Services. Choosing Use Cloudera's default Docker Repository can cause the Cloudera Data Science Workbench (CDSW) to CML migration to fail.

Workaround: During installation, in Configure Docker Repository, choose either one of the following options:

• Use an embedded Docker Repository
• Use a custom Docker Repository

The mlflow.log_model registered model files might not be available on NFS Server

When using mlflow.log_model, registered model files might not be available on the NFS server due to NFS server settings or network connections. This could cause the model to remain in the registering status.

Workaround:

- Re-register the model. It will register as an additional version, but it should correct the problem.
- Add the ARTIFACET_SYNC_PERIOD environment variable to ds-vfs Kubernetes deployment and set it to an integer value. This will set the model registry retry operation to twice the number of seconds specified by the artifact sync period integer value. If the ARTIFACET_SYNC_PERIOD is set to 30 seconds then model registry will retry for 60 seconds. The default value is 10 and model registry retries for 20 seconds. For example: -name: ARTIFACT_SYNC_PERIOD value: "30".

DSE-23329: Spark API fails when reading or writing to Ozone filesystem

CML with Spark as a runtime addon does not work with the Ozone file system.

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.

Workarounds:

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.

DSE-20923: User search requires lower case

When searching for a user name in Site Administration Users , you must enter only lower-case letters for the name you are searching for. Lower-case letters will match upper-case letters in the target name.

DSE-19036: Model and experiment building fails after ECS upgrade

After ECS Control Plane upgrade, your ML workspace may fail to perform Model and Experiment building with a message like: Cannot connect to the Docker daemon at   unix:///var/run/docker.sock. Is the docker daemon running?

To resolve this problem, perform these steps:

1. Access Cloudera Manager
2. Navigate to the Embedded Cluster ECS Web UI: Clusters Your Embedded Cluster ECS Web UI ECS Web UI
3. Select the namespace of your ML workspace on the top left dropdown.
4. In Workloads Deployments , locate s2i-builder in the list.
5. In the Action menu for s2i-builder, select Restart.

DSE-13117: Container Image Registries assuming mutual TLS for authentication are not supported

If Private Cloud images are hosted in an image registry assuming mutual TLS for authentication, this will cause Model deployments and Experiments to fail. Mutual TLS registries are not supported.

DSE-12541: Self Signed Certificates for Container Registry cause Models and Experiments to fail

If you are using self-signed or Private CA signed certificates for Container image registry authentication, model deployments and experiments will fail with an error similar to: Error initializing source docker://<registryIP>:5000/

alpine:latest: error   pinging docker registry <registryIP>:5000: Get https://<registryIP>:5000/v2/: x509:   certificate signed by unknown authority

As a workaround, create a ConfigMap in the namespace where the CML workspace is installed.

1. Create a ConfigMap as shown in this example. Here, <namespace> indicates the workspace where the CML workspace is installed.

```
kind: ConfigMap
apiVersion: v1
metadata:
  name: <external-registry-name>
  Namespace: <namespace>
data:
  registry.crt: |
    -----BEGIN CERTIFICATE-----
    < certificate content goes here >
    -----END CERTIFICATE-----
    -----BEGIN CERTIFICATE-----
    < certificate content goes here >
    -----END CERTIFICATE-----
```

2. Mount the ConfigMap to the s2i-builder deployment as shown here. Add the following mountPath in the volu meMounts section for the s2i-builder pod:

```
- mountPath: /etc/docker/certs.d/<registry>[:portnum]
  name: external-registry
```

Under the volumes section, add the ConfigMap reference:

```
- configMap:
    defaultMode: 420
    name: <external-registry-name>
  name: external-registry
```

3. Run the following command and check the output. Note in particular the mountPath and configMap specifications at the end.

```
# kubectl get deployment s2i-builder -n <namespace> -o yaml

apiVersion: extensions/v1beta1
kind: Deployment
metadata:
  [...]
  name: s2i-builder
  namespace: <namespace/workspacename>
  [...]
spec:
  [...]
  template:
    [...]
    spec:
    [...]
    containers:
    - name: s2i-builder
    [...]
    volumeMounts:
          [...]
          - mountPath:/etc/docker/certs.d/<registry>[:portnum]
            name: external-registry
        [...]
  volumes:
    [...]
```

```
    - configMap:
       defaultMode: 420
       name: <external-registry-name>
    name: external-registry
status:
[...]
```

DSE-12367: s2i-queue pod goes into CrashLoop Failure causing ML workspace installation to fail

CML workspace install can fail because the s2i-queue pod may be stuck in a CrashLoop Failure. The error in the logs might look similar to: Failed to create thread: Resource temporarily unavailable (11) /usr/lib/rabbitmq/bin/rabbitmq-server: line 182: 45   Aborted (core dumped) start_rabbitmq_server "$@" Only root or rabbitmq can run   rabbitmq-server

To fix this, apply the following workaround: # kubectl set env   statefulset/s2i-queue RABBITMQ_IO_THREAD _POOL_SIZE="50" -n <namespace>

DSE-12329: Email invitation feature

The feature to invite new users by email does not work in Public or Private cloud, but it still appears in the UI.

DSE-12289: Airgap support: Proxies are not supported in CML Private Cloud 1.0

Use of a proxy server, for example for external internet connectivity for an airgap cluster, is not supported. Transparent proxies, however, should work normally.

DSE-12238: Create Project request takes longer than timeout

If a Create Project request takes longer than a certain timeout, a second request might be submitted. If this happens, multiple projects with similar names might be created.

As a workaround, create an empty project, create a session inside the project, then git clone your project inside a workbench terminal. Additionally, you can upload a zip file or a folder using the file preview table.

If multiple forks are created, delete the extra ones.

DSE-12090: User displays as unknown in Event History

In the Event History on the workspace Events tab, a user may display as unknown if they are authenticated by LDAP.

Fix: The user needs to be assigned the IamViewer role to view these details.

DSE-11979: Certificate failure when pulling images from the S2I container registry

During Model or Experiment deployment, a certificate failure similar Failed to pull   image x509: certificate signed by unknown authority can occur. This is due to a Red Hat issue with OpenShift Container Platform 4.3.x where the image registry cluster operator configuration must be set to Managed.

To set the configuration, first apply a patch using this command: # oc patch   configs.imageregistry.operator.openshif t.io cluster --patch   '{"spec":{"managementState":"Managed"}}'

Next, run the following command: # oc get config cluster -o yaml

The managementState is now set to Managed.

DSE-11870: Hung File, Stale File, and Fork issues with NFS

Hung File Operations: Certain file operations, such as stat(2) or stat(1) might stop responding, and if the file operation was performed through the CML web UI, the web operation might timeout. This indicates an NFS server

that is not reachable for some reason. The error might manifest itself on the web UI when you try to open an ML project as an HTTP error, code 500. Check the logs for error messages similar to the following:

```
 2020-07-13 22:42:23.914 1 ERROR AppServer.Lib.Utils Finish grpc, failed dat
a =
        [{"rpc":"1","service":"2","reqId":"3","err":"4"},"stat","VFS","18a07
980-c55a-11ea-9bb9-a35829b422d9",{"message":"
        5","stack":"6","code":4,"metadata":"7","details":"8","futureStack"
:"6"},"4
        DEADLINE_EXCEEDED: Deadline Exceeded","Error: 4 DEADLINE_EXCEEDED:
Deadline Exceeded\n at
        Object.exports.createStatusError (/home/cdswint/services
        /web/node_modules/grpc/src/common.js:91:15)\n at Object.onReceiveSta
tus
        (/home/cdswint/services/web/node_modules/grpc/src/client_interceptor
s.js:1209:28)\n at
        InterceptingListener._callNext (/home/cdswint/services/web/
        node_modules/grpc/src/client_interceptors.js:568:42)\n at
        InterceptingListener.onReceiveStatus
        (/home/cdswint/services/web/node_modules/grpc/src/client_intercept
ors.js:618:8)\n at
        callback (/home/cdswint/services/web/n
        ode_modules/grpc/src/client_interceptors.js:847:24)",{"_internal_r
epr":"9","flags":0},"Deadline
        Exceeded",{}]
```

Solution: Check your NFS server and make sure it is running. You will need to restart the NFS clients in your ML workspace's namespace. These are the "ds-vfs" and "s2i-client" pods. Simply delete the Kubernetes pods whose names start with "ds-vfs" and "s2i-client".

Stale File Handles: When opening a project from the ML web UI, an error message like "NFS: Stale file handle" shows up on the UI.

Solution: This is indicative of an NFS server and a client being out of sync, probably caused by a server restart along with file system content change on the server that the client is not aware of. You should restart NFS client pods in your ML workspace's namespace. The are the "ds-vfs", "s2i-client", and any user sessions that are affected by the "Stale file handle" error.

Project Fork Creating Multiple Copies: When creating a new project from an existing project using the "Fork" feature, you might see the operation seemingly fail on the UI, but it still ends up creating multiple copies of the source project.

Solution: This issue happens when forking a project takes longer than the idle connection timeout set on the external load balancer, as well as in HA Proxy policy settings on OpenShift. Increase the idle connection timeout to at least 5 minutes. Depending on the performance of the NFS server, a higher timeout may be necessary.

DSE-11837: Timeout limitation for Project API

If you create a project in the UI using git clone, you may get the error message Whoops, there was an unexpected error. If you create a project using the API, a timeout may occur.

Prerequisites for CML in Private Cloud:

- Set any external load balancer server timeout to 5 min.

For a TLS Enabled Workspace:

- Set the annotation haproxy.router.openshift.io/timeout=300 on each route in a deployed CML workspace namespace:

```
oc annotate route --all=true --overwrite=true -n
            <cml-namespace> haproxy.router.openshift.io/timeout=300s
```

For non-TLS Enabled Workspaces, this setting is made automatically.

Workaround: Even though an error message displays, project creation still occurs. Check the Projects page after a few minutes; project creation should be complete.

DSE-9549: TLS enabled workspaces require manual configuration

To provision a TLS-enabled workspace, the customer needs to perform several manual steps. This procedure is described in Deploy an ML Workspace with Support for TLS.

OPSAPS-58019: CML workspace installation failure due to includedir in krb5.conf file

If the /etc/krb5.conf on the Cloudera Manager host contains include or includedir directives, Kerberos-related failures may occur.

As a workaround, comment out the include and includedir lines in /etc/krb5.conf on the Cloudera Manager host. If configuration in those files and directories are needed, add them directly to /etc/krb5.conf.

DSE-21768: Spark3 runtime addons are not supported with Python 3.8 Runtimes

If the /etc/krb5.conf on the Cloudera Manager host contains include or includedir directives, Kerberos-related failures may occur.

Spark3 runtime addons are currently not supported with Python 3.8 Runtimes.

Related Information
Known Issues and Limitations