Cloudera Data Science Workbench

Cloudera Data Science Workbench Release Notes

Date published: 2020-02-28
Date modified: 2020-12-15

CLOUDƎRA

https://docs.cloudera.com/

Legal Notice

Contents

What's New

Major features and updates plus known issues for Cloudera Data Science Workbench 1.9.0.

New features

- ML Runtimes - As an alternative to the existing Engines in CML, ML Runtimes are more lightweight than the current monolithic Engines. By specifying the desired Editor, Kernel, Edition, and Version, a streamlined Runtime will be used to run the user's code in Sessions, Jobs, Experiments, Models, or Applications.
- Support for ML Ops Governance - MLOps Governance is now supported on CDP Private Base.
- Updated Data Science Project dashboard - The Data Scientist Project dashboard has been updated.
- LDAP Group Sync - You can now create a team that synchronizes with an LDAP group to more easily manage user access to a team.
- Support for applications with public, unauthenticated access - Applications can now be configured for public, unauthenticated access.
- Support for SLES12 SP5, CentOS/RHEL 7.8 - SLES12 SP5, CentOS/RHEL 7.8 are now supported operating systems.
- ML Runtimes supports JupyterLab (Technical Preview) - JupyterLab is now an available ML Runtime, in technical preview.
- Documentation new look and feel -

  - Navigation Bar - Subjects are now grouped under major categories.

    You'll notice that we've added some major categories and grouped subjects under these categories. For example, there are now Concept, Installation & Upgrade, and How To categories among others. Relevant subjects are grouped under these categories. Under Concepts you'll find Overview and Architecture information. Under How To, you'll find the Quick Start and User Guides, along with information on how to set up and configure Projects, Editors, Model, etc.

  - Introduction of foyer pages

    Foyer pages provide additional information about the topics grouped under that subject. Foyer pages are only used for subjects that contain a lot of major topics. For example, Site Administration provides a foyer page that lists and defines the major subjects contained in Site Administration, including Monitoring, Quotas, and Data Collection.

  - New links are necessary

    As part of restructuring the documentation, we separated many large topics into smaller, more concise topics. As a result, topic titles have changed and any links you've saved to previous CDSW releases will not work for CDSW 1.9.0.

Fixed issues

- The issue with Restart is now fixed.

  If multiple users clicked Restart multiple times around the same time, it was possible to have multiple engines running per applications. This resulted in a round-robin like behavior.

  Also, in this situation, if a user stopped the application, even though the application said it stopped, the application was still serving on the provided subdomain, because of the zombie engines created in the process above.

  Cloudera Bug: DSE-12732

- Cloudera Data Science Workbench Administrator can now view all projects on Projects page.

  Cloudera Bug: DSE-12057

- You can now start, stop, and delete applications stuck in a stopped state after upgrading to CDSW 1.8.0.

  Cloudera Bug: DSE-12374

- The Export Session List option on the Administration/Usage page now generates a list that is sortable by date. The Created At column displays a date value that can be used for sorting.

  Cloudera Bug: DSE-9661

- The Export PDF function is improved and will display charts, tables, and maps. However, some content still might not display well, such as tables with many columns and dynamic content, video, or iframe content containing dynamic content.

  Cloudera Bug: DSE-11685

- Models and Applications can now use Cloudera Manager and CDSW proxy settings properly. This issue also caused problems building certain Models and Applications and requring proxy settings to install additional packages (for example, pip). This issue is also fixed.

  Cloudera Bug: DSE-5467

- Previously CDSW hosts required that IPv6 support was enabled. This issue is now fixed. CDSW will function with IPv6 networking disabled or without IPv6 kernel support.

  Cloudera Bug: DSE-13465

For a list of supported CDSW versions, see Support Lifecycle Policy

Cloudera Data Science Workbench 1.8.1

Major features and updates plus known issues for Cloudera Data Science Workbench 1.8.1.

New Features and Changes in Cloudera Data Science Workbench 1.8.1

This release fixed various bugs that are listed in Issues Fixed in Cloudera Data Science Workbench 1.8.1.

Issues Fixed in Cloudera Data Science Workbench 1.8.1

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an upgrade issue impacting database upgrade part when exceeding maxBuffer length can prevent cluster from coming up.

  Cloudera Bug: DSE-10120

- Fixed an issue where Applications can stick in a stopped state after upgrading to CDSW 1.8.0.

  Cloudera Bug: DSE-12374

- Fixed an issue where an Application could be stuck with status starting and a user might not be able to restart or delete it.

  Cloudera Bug: DSE-12732

- Fixed an issue handling large responses from Model.

  Cloudera Bug: DSE-12310

- Fixed Admin Usage page refresh issue.

  Cloudera Bug: DSE-12370

- Fixed issue displaying latest engine image in case engine9 or earlier versions are also present.

  Cloudera Bug: DSE-12372

- Resolved issue with python scripts not being able to interact with git while running as Jobs.

  Cloudera Bug: DSE-12624

- The Options tab on the Project Settings page is now only viewable to the Administrator. This is because the Private and Team visibility options on the Options page are only viewable by the Administrator.

  Cloudera Bug: DSE-12027

- Fixed an issue that prevented Terminal Access in case of Sessions on a cluster behind firewall and/or with proxy settings.

  Cloudera Bug: DSE-12867

Cloudera Data Science Workbench 1.8.0

Major features and updates plus known issues for Cloudera Data Science Workbench 1.8.0.

New Features and Changes in Cloudera Data Science Workbench 1.8.0

Major features and updates for Cloudera Data Science Workbench.

Production Machine Learning

> Machine Learning (ML) lifecycle focused functionality enabling ML Engineers and Data Scientists to cut time to production for ML models from weeks to minutes and scale ML use cases without compromising enterprise security, maintainability, and governance standards.

Ability to monitor model metrics

> CDSW allows you to track individual model predictions and analyze the metrics using custom code. Monitoring models' functional and business performance requires specialized tooling and CDSW now includes native functionality to enable the storage and access of custom and arbitrary model metrics. Included as well is the ability to track individual predictions to ground truth ensuring models are performing optimally and compliantly.

Improved model security

> Model REST endpoints now have additional security features that allow user-level access control to prevent unauthorized users from accessing the endpoints. This enables models to be served in a production ML environment without compromising security.

Quota management is generally available

> Resource quotas provide constraints that limit the aggregate CPU, memory, and GPU resource consumption per user. Default quotas can be configured for a Workspace, which can be overridden on a per-user basis with Custom quotas. Quota management is now generally available.

New Web Browser Timeout Settings for Admin and User

> The default Admin user timeout is one hour, regardless of activity. For users, the default timeout is 24 hours. Five minutes before the timeout is reached, a message displays, asking if the user wants to remain logged in. If the user clicks Dismiss or does not see the reminder, they have to log back in when their session expires. The timeout settings can be changed in Admin Security .

Updated Session Start view

> The user experience around starting a new CDSW session has been enhanced.

Ability to add CDSW session metadata information

> Administrators can make it mandatory for the users to enter additional metadata before starting a CDSW session. Administrators can configure up to three fields to collect such additional metadata information for compliance purposes and can make these mandatory to be filled in. This metadata is captured and stored in the user_events database table and in the audit log as key-value pairs.

Base Engine 13 support

CDSW supports and deploys Base engine 13 that ships with both Python versions 2.7.18 and 3.6.10, and R version 3.6.3.

Ability to use custom command-line arguments for sessions and jobs

CDSW jobs run scripts, and the scripts need you to pass arguments while they are being executed. You can set these command-line arguments in the Engine's Command Line Arguments field when you create a job. These command-line arguments are shared by CDSW sessions, jobs, and experiments in a project. The arguments can be accessed inside your scripts the same way CLI arguments are normally used.

CDSW displays engine total resource consumption

The Engine Profile now reflects the total resource consumption required to run engines. This means that apart from the "effective" resources used for computation, the additionally required "sidecar" resources are also counted, and enforced when using quotas.

Syntax highlighting support for additional programming languages

CDSW now supports syntax highlighting for the following languages when you are previewing a file: XML, CSS, JS, HTML, CoffeeScript, SQL, and CSV.

Ability to configure Windows line endings for Git integration

Previously, when CDSW Brackets editor interacted with files having Windows line endings, the Brackets editor converted all the Windows line endings to Linux line endings distorting the git log. The behavior to use Windows line endings is now configurable by the Administrator on a site level.

Cross-Origin Resource Sharing (CORS) is disabled by default

CORS is now disabled by default. You can enable CORS if you have web applications on different domains that need cross-domain communication with the CDSW API.

OS support changes

CentOS 7.2 is no longer supported.

Issues Fixed in Cloudera Data Science Workbench 1.8.0

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Improved platform stability and performance

  - Fixed a number of potential security vulnerabilities in the base operating systems of our Kubernetes pods
  - Upgraded Kubernetes to 1.14
  - Fixed the Sweet32 Birthday attack vulnerability within the Kubernetes infrastructure that supports CDSW

  Cloudera Bug: DSE-9061

- A number of CDSW pods now run as non-root users

  Cloudera Bug: DSE-10493, DSE-7357

- Changed how the web sessions are handled

  - CDSW provides an option to refresh the session five minutes before your web session expires so that your active users can continue their session without the cookie expiring.

  Cloudera Bug: DSE-11773

  - Administrators can now disable the ability to have multiple concurrent sessions. This prevents users from logging in to the same account with multiple browsers or computers. By default, users will still be able to have concurrent web browser sessions. This may change in a future release.

  Cloudera Bug: DSE-8109

- Accessibility fixes

  - The CDSW UI has been polished with many accessibility fixes. The main application strives to be WCAG2AA compliant with the exception of the workbench editor.

    Cloudera Bug: DSE-4490
- Re-starting Cloudera Data Science Workbench does not automatically restart active models

  These models must be manually restarted so they can serve requests again.

  Cloudera Bug: DSE-4950
- Fixed the CDSW restart issue on multi-node deployments.

  Cloudera Bug: DSE-9587, DSE-9663
- Cloudera Manager now generates an "https" link for the CDSW web UI instead of "http" if you have enabled TLS for your CDSW cluster.

  Cloudera Bug: DSE-8698
- (If quotas are enabled) Experiments no longer remain stuck in the Scheduled state due to lack of resources.

  You no longer need to manually reschedule the experiment.

  Cloudera Bug: DSE-8736
- Job notification emails no longer fail intermittently when attachments are included.

  Cloudera Bug: DSE-9469, DSE-8806

Cloudera Data Science Workbench 1.7.2

Major features and updates plus known issues for Cloudera Data Science Workbench 1.7.2.

New Features and Changes in Cloudera Data Science Workbench 1.7.2

Major features and updates for Cloudera Data Science Workbench.

Added support for CDP Data Center 7.0

Cloudera Data Platform (CDP) Data Center is the on-premises version of Cloudera Data Platform. This new product combines the best of Cloudera Enterprise Data Hub and Hortonworks Data Platform Enterprise along with new features and enhancements across the stack. CDP Data Center is comprised of a variety of components such as Apache HDFS, Apache Hive 3, Apache HBase, and Apache Impala, along with many other components for specialized workloads. For details, see CDP Data Center Overview.

You can use Cloudera Manager to install CDSW 1.7.2 as a parcel on CDP Data Center 7.0. RPM package installs are not supported.

Environment variables set at the Site Admin-level and project-level are now passed to models and experiments during the container build process.

Previously (CDSW 1.7.1 and lower), the environment variables set at the site admin level and project level did not automatically get pulled into the builds created for models and experiments. They needed to be explicitly coded into the cdsw-build.sh file. With CDSW 1.7.2 and higher, experiments and models will automatically inherit these admin and project-level environment variables.

Note that custom mounts or environment variables configured in cdsw.conf (such as NO_PROXY, HTTP(S)_PROXY, etc.) are still not passed to the container builds for experiments and models (even though they are applied to sessions, jobs, and deployed models/experiments).

WARNING: This change affects access to Analytical Applications

The expected host IP of CDSW_PUBLIC_PORT has been changed from 0.0.0.0 to localhost (127.0.0.1). This will affect the ability of analytical applications to connect for users who are not authenticated by CDSW. Existing applications listening on "0.0.0.0:CDSW_PUBLIC_PORT" must be migrated to "localhost:CDSW_PUBLIC_PORT".

Issues Fixed in Cloudera Data Science Workbench 1.7.2

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue on multi-node CDSW 1.7.1 deployments where the CDSW Web UI would not automatically come up after upgrading to CDSW 1.7.1

  Cloudera Bug: DSE-9587
- Fixed an issue where environmental variables set at the site admin level and at the project level would not get passed down to experiments and models at container build time.

  Cloudera Bug: DSE-6708
- Fixed an issue where new users could not log in when the Require invitation to sign up checkbox was enabled.

  Additionally, the Test LDAP Configuration form did not return any error message if the user being tested wasn't already synced to the local CDSW database.

  Cloudera Bug: DSE-3829
- Fixed an issue where license files could not be uploaded to CDSW through the UI.

  Cloudera Bug: DSE-9874, DSE-8865

Cloudera Data Science Workbench 1.7.1

Major features and updates plus known issues for Cloudera Data Science Workbench 1.7.1.

Note: Cloudera Data Science Workbench 1.7.1 is the next official release after Cloudera Data Science Workbench 1.6.x. Version 1.7.0 is no longer publicly available.

This section lists the release notes for Cloudera Data Science Workbench 1.7.1.

New Features and Changes in Cloudera Data Science Workbench 1.7.1

Major features and updates for Cloudera Data Science Workbench.

Supported upgrade paths to CDSW 1.7.1

Cloudera Data Science Workbench only supports upgrades to version 1.7.1 from version 1.5.x and 1.6.x. If you are using an earlier version of CDSW, you must first upgrade to version 1.5.x or 1.6.x, and then upgrade to version 1.7.1.

Analytical Applications

Cloudera Data Science Workbench now gives data scientists a way to create long-running standalone ML web applications/dashboards that can easily be shared with other business stakeholders. Applications can range from single visualizations embedded in reports, to rich dashboard solutions such as Tableau.

Applications stand alongside other existing forms of workloads in CDSW (sessions, jobs, experiments, models). For details, see Analytical Applications.

Monitoring CDSW with Grafana

CDSW now leverages Prometheus and Grafana to provide a dashboard that allows you to monitor how CPU, memory, storage, and other resources are being consumed by CDSW deployment. For details, see Cluster Monitoring with Grafana.

Feature flag overrides

This is a new property available in the CDSW service in Cloudera Manager. It can be used to enable/disable experimental features (such as quotas) and disable metric collection in diagnostic bundles.

Quotas

CDSW site administrators can now enable CPU, GPU, and memory usage quotas per user. You can set default quotas for each user on the deployment as well as overriding custom quotas for specific users. For details, see Configuring Quotas.

> ⚠️ **Important:** The Quotas feature is in Technical Preview. It is not enabled by default. To enable quotas, use the Feature flag overrides property in Cloudera Manager.

Usage Metrics Collection

By default, CDSW 1.7.1 now gathers highly redacted information on which feature is being used on your deployment. When you create a diagnostic bundle, this information is packed alongside the diagnostic information.

You can use the Feature flag overrides property in Cloudera Manager to disable collection of usage metrics.

Engine Upgrade 1.7.1

Cloudera Data Science Workbench 1.7.1 (and later) ships version 10 of the base engine image.

Cloudera Data Science Workbench 1.7.1 (and later) ships version 10 of the base engine image which includes the following versions of R and Python:

- R - 3.5.1
- Python - 2.7.17, 3.6.9

Pre-installed Packages in Engine 10

For details about the packages included in the base engine, see Cloudera Data Science workbench Engine Versions and Packaging.

Issues Fixed in Cloudera Data Science Workbench 1.7.1

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue where Cloudera Manager's 5 minute timeout for generating support bundles would lead to CDSW data and metrics missing from the bundles.

  Cloudera Bug: DSE-3160
- Fixed an issue where scheduled jobs would not start if scheduled in timezones other than UTC.

  Cloudera Bug: DSE-8563
- Fixed an issue where inactive Jupyter sessions in the Workbench would behave in ways that were inconsistent with the rest of the application.

  Cloudera Bug: DSE-7867
- Fixed an issue in version 1.6.1 where deployments using a custom Certificate Authority (signed by either their organisation's internal CA or a non-default CA) would have to explicitly set the REQUESTS_CA_BUNDLE environmental variable to force Python to use the system truststore.

  Cloudera Bug: DSE-7441

- Fixed UI issues where the application did not open project files consistently or as expected.

  Cloudera Bug: DSE-6274
- Fixed an issue where CSV files with Chinese characters could not be previewed in the Workbench or Files view.

  Cloudera Bug: DSE-4892
- Fixed an issue where CDSW would retain session data for a long time which led to the var/lib/cdsw mount filling up with old data that affected application performance.

  Cloudera Bug: DSE-3170
- CDSW now clears all iptables rules on application restart.

  Cloudera Bug: DSE-5095

Cloudera Data Science Workbench 1.6.1

Major features and updates plus known issues for Cloudera Data Science Workbench 1.6.1.

New Features and Changes in Cloudera Data Science Workbench 1.6.1

Major features and updates for Cloudera Data Science Workbench.

Security

- SAML Identity Provider Logout: With version 1.6.1, a user clicking the Sign Out button on CDSW can also be logged out of their identity provider.

  For instructions on how to configure this, see SAML Identity Provider Configuration.
- Root CA configuration: Added a new field called Root CA configuration to the Admin Security page. Organizations that use an internal custom Certificate Authority can use this field to paste in the contents of their internal CA's root certificate file.

  For instructions, see Configuring Custom Root CA Certificate.

IPv6 Requirement

Cloudera Data Science Workbench 1.6.x requires you to enable IPv6 on all CDSW gateway hosts.

Resource Usage captured in User Events

You can now query the user events table for resources used by each session, job, experiment, and model.

Editors (Windows only)

For the Windows cdswctl client, CDSW now automatically adds the .exe extension to the file-name. The format of the downloaded file has also changed from tar.gz to zip.

Kubernetes

Kubernetes has been upgraded to version 1.13.9.

Engine Upgrade 1.6.1

Cloudera Data Science Workbench 1.6.1 (and later) ships version 10 of the base engine image.

Cloudera Data Science Workbench 1.6.1 (and later) ships version 10 of the base engine image which includes the following versions of R and Python:

- R - 3.5.1
- Python - 2.7.17, 3.6.9

Engine 10 uses Ubuntu 18.04 as its base operating system. This is an upgrade from Engine 8 which used Ubuntu 16.04.

Issues Fixed in Cloudera Data Science Workbench 1.6.1

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue where deployments using a custom Certificate Authority (signed by either their organisation's internal CA or a non-default CA) would see HTTP Error 500 when attempting to launch the Terminal or Jupyter Notebook sessions from the Workbench

  With version 1.6.1, if you are using a custom CA, site administrators can go to the Admin Security page and paste your internal CA's root certificate file contents into the Root CA configuration field.

  The contents of this field are then inserted into the engine's root certificate store every time a session (or any workload) is launched. This allows processes inside the engine to communicate with the ingress controller.

  Cloudera Bug: DSE-7237, DSE-7173

- Fixed an issue where the Admin License page was displaying an incorrect number of CDSW users on a deployment. This count has now been updated to reflect only the total number of active users.

  Cloudera Bug: DSE-6350

- Fixed an issue where setting HADOOP_USER_NAME to the CDSW username had certain unintended consequences. This fix now sets HADOOP_USER_NAME to the first part of the Kerberos principal in kerberized environments. In non-kerberized environments, it is still set to the CDSW username.

  Cloudera Bug: DSE-6928

- Fixed an issue where sessions on non-kerberized environments would throw the following error even though no principal was provided: Kerberos principal   provided, but no krb5.conf and cluster is not Kerberized.

  Cloudera Bug: DSE-7236

- Fixed an issue in version 1.6.0 where GPUs were not being recognized on air-gapped environments.

  Cloudera Bug: DSE-7138

- Fixed an issue where the transition to/from Daylight Savings Time would cause scheduled CDSW jobs to fail.

  Cloudera Bug: DSE-3399

- Fixed an issue in CDSW 1.6.0 where CDSW sessions would fail to launch or the web application crashes due to a Node.js issue.

  Cloudera Bug: DSE-7238

Cloudera Data Science Workbench 1.6.0

Major features and updates plus known issues for Cloudera Data Science Workbench 1.6.0.

New Features and Changes in Cloudera Data Science Workbench 1.6.0

Major features and updates for Cloudera Data Science Workbench.

Bring Your Own Editor

> You can now take advantage of all the benefits of Cloudera Data Science Workbench while using an editor you are familiar with. This feature supports third-party IDEs that run on your local machine like PyCharm and browser-based IDEs such as Jupyter. Base Image v8 ships with Jupyter preconfigured and can be selected from the Start Session menu.

> For details, see Editors.

Multiple Cloudera Data Science Workbench Deployments

You can now have multiple Cloudera Data Science Workbench CSD deployments associated with one instance of Cloudera Manager.

For details, see Multiple Cloudera Data Science Workbench Deployments.

Audits

Cloudera Data Science Workbench logs specific events, such as user logins and sharing, that you can view by querying a database. For more information, see Monitoring User Events and Tracked User Events.

Expanded Support for Distributed Machine Learning

Cloudera Data Science Workbench 1.6 (and higher) allows you to run distributed workloads with frameworks such as TensorFlowOnSpark, H2O, XGBoost, and so on. This is similar to what you can already do with Spark workloads that run on the attached CDH/HDP cluster. For details, see Running Distributed ML Workloads on YARN.

cdswctl CLI Client

The cdswctl client provides an additional way to interact with your Cloudera Data Science Workbench deployment to perform certain actions. For example, you can use the cdswctl client to start an SSH-endpoint on your local machine and then connect a local IDE, such as PyCharm, to Cloudera Data Science Workbench.

You can download cdswctlfrom the Cloudera Data Science Workbench web UI and use it from your local machine. Note that this client differs from the cdsw CLI tool used to run commands such as cdsw status, which exists within the Cloudera Data Science Workbench deployment.

For details, see cdswctl Command Line Interface Client.

Status and Validate Commands

The CDSW service in Cloudera Manager now includes two new commands that can be used to assess the status of your Cloudera Data Science Workbench deployment: Status and Validate. They are the equivalent of the cdsw   status and cdsw validate commands that are available via the CLI.

For details, see Checking the Status of the CDSW Service.

Experiments

- If your cluster has been equipped with GPUs, you can now use GPUs to run experiments on Cloudera Data Science Workbench.
- Tracked experiment files now refresh and appear automatically on the Overview page for a run of an experiment. Previously, you had to manually refresh the page after an experiment completes.

Command Line Interface (CLI) Changes - RPM Deployments only

- The cdsw reset command has been removed and replaced by the cdsw stop command.
- The cdsw init command has been removed and replaced by the cdsw start command.

For details on how these commands behave on the master and worker hosts, refer to the Cloudera Data Science Workbench Command Line Reference .

Kubernetes and Weave

Kubernetes has been upgraded to version 1.11.7. Weave Net has been upgraded to version 2.5.1. This upgrade resolves Weave issue #2934.

Loggs

- Staging Directory

  You can now configure the temporary directory that Cloudera Data Science Workbench uses to stage logs when collecting a diagnostic bundle. Old logs in the directory are deleted when a new diagnostic bundle is collected or when the size grows larger than 10 MB.

- Logs tab

  Running sessions now display a Logs tab. This tab displays engine logs and, if applicable, Spark logs for the running session. Previously, if you wanted to access these logs, that required logging into the Cloudera Data Science Workbench host(s) and the Spark server.

  For details, see Diagnostic Bundles.

Operating System

Cloudera Data Science Workbench 1.6 supports RHEL and CentOS 7.6.

Workload Scheduling Changes

- Starting with version 1.6, Cloudera Data Science Workbench allows you to specify a list of CDSW gateway hosts that are labeled as Auxiliary Nodes. These hosts will be deprioritized during workload scheduling. That is, they will be chosen to run workloads that can't be scheduled on any other hosts. For example, sessions with very large resource requests, or when the other hosts are fully utilized.

  For details, see Customize Workoad Scheduling.
- Reserve Master Host

  Cloudera Data Science Workbench 1.4.3 introduced a new feature that allowed you to reserve the CDSW Master host for running internal application components. Starting with version 1.6, this feature can be enabled on CSD-based deployments using the Reserve Master Host property in Cloudera Manager. Safety valves are no longer needed.

  For details, see Reserving the Master Host for Internal CDSW Components.

Security

- FreeIPA Support

  In addition to MIT Kerberos and Active Directory, Cloudera Data Science Workbench now also supports FreeIPA as an identity management system. For details, see Configure FreeIPA.
- New User Role - Operator

  Version 1.6 includes a new access role called Operator. When a user is assigned the Operator role on a project, they will be able to start and stop pre-existing jobs and will have view-only access to project code, data, and results.
- Restricting User-Controlled Kubernetes Pods

  Cloudera Data Science Workbench 1.6 includes three new properties that allow you to control the permissions granted to user-controlled Kubernetes pods. An example of a user-controlled pod is the engine pod, which provides the environment for sessions, jobs, etc. These pods are launched in a per-user Kubernetes namespace. Since the user has the ability to launch arbitrary pods, these settings restrict what those pods can do.

  For details, see Retricting User-Controlled Kubernetes Pods.
- LDAP/SAML Configuration Changes

  Previously, if you wanted to grant the site administrator role to users of an LDAP/SAML group, that group had to be listed under 2 properties: LDAP/SAML Full Administrator Groups and LDAP/SAML User Groups. If a group was only listed under LDAP/SAML Full Administrator Groups, and not under LDAP/SAML User Groups, users of that group would not be able to log in to CDSW.

  With version 1.6, you do not need to list the admin groups under both properties. Users belonging to groups listed under LDAP/SAML Full Administrator Groups will be able to log in and have site administrator access to Cloudera Data Science Workbench as expected.

- Project and Team Creation

  Site administrators can now restrict whether or not users can create projects or teams with the following properties on the Settings page:

  - Allow users to create projects
  - Allow users to create teams

  For details, see User Access to Features.
- Session Tokens

  The method by which the Cloudera Data Science Workbench web UI session tokens are stored has been hardened. Users must log out of the Cloudera Data Science Workbench web UI and back in after upgrading to version 1.6.0.
- Sharing

  Site administrators can now control whether consoles can be shared with the Allow console output sharing property on the Admin Security page. Disable this property to remove the Share button from the project workspace and workbench UI as well as disable access to all shared console outputs across the deployment. Note that re-enabling this property does not automatically grant access to previously shared consoles. You will need to manually share each console again.
- TLS/SSL

  Cloudera Data Science Workbench now defaults to using TLS 1.2. The default cipher suites have also been upgraded to Mozilla's Modern cipher suites.

IPv6 Requirement

Cloudera Data Science Workbench 1.6.x requires you to enable IPv6 on all CDSW gateway hosts. For instructions, refer the workaround provided in Known Issue: CDSW cannot start sessions due to connection errors.

Spark UI

The Spark UI is now available as a tab within running sessions that use Spark.

Engine Upgrade 1.6.0

Cloudera Data Science Workbench 1.6.0 (and later) ships version 8 of the base engine image.

Cloudera Data Science Workbench 1.6.0 (and later) ships version 8 of the base engine image which includes the following versions of R and Python:

- R - 3.5.1
- Python - 2.7.11, 3.6.8

Pre-installed Packages in Engine 8

For details about the packages included in the base engine, see Cloudera Data Science Workbench Engine Versions and Packaging.

(For Upgrades Only) Move Existing Projects to the Latest Base Engine Images

Make sure you test and upgrade existing projects to Base Image v8 (Project Settings Engine) to take advantage of the latest fixes. There are two reasons to do this:

- Container Security

  Security best practices dictate that engine containers should not run as the root user. Engines (v7 and lower) briefly initialize as the root user and then run as the cdsw user. Engines v8 (and higher) now follow the best practice and run only as the cdsw user. For more details, see Restricting User-Created Pods.

- CDH 6 Compatibility

  The base engine image you use must be compatible with the version of CDH you are running. This is especially important if you are running workloads on Spark. Older base engines (v6 and lower) cannot support the latest versions of CDH 6. If you want to run Spark workloads on CDH 6, you must upgrade your projects to base engine 7 (or higher).

Incompatible Changes in Cloudera Data Science Workbench 1.6.0

Cloudera Data Science Workbench 1.6.0 no longer supports some of the software and components supported in previous releases.

SLES 12 SP2, SP3 are not supported with Cloudera Data Science Workbench 1.6.0

SLES 12 SP2 and SP3 have reached the end of general support with SUSE and will not be supported with Cloudera Data Science Workbench 1.6.0 (and higher).

GPU Setup Changes

- nvidia-docker1 is no longer supported.
- The NVIDIA Library Path property is no longer available.

  Cloudera Data Science Workbench 1.6 ships with nvidia-docker2 installed by default. The path to the NVIDIA library volumes is also set automatically when GPUs are enabled. Review the revised GPU setup steps in Enabling Cloudera Data Science Workbench to use GPUs.

The CDSW_PUBLIC_PORT environment variable has been deprecated and will be removed in a future release. Use CDSW_APP_PORT or CDSW_READONLY_PORT environment variables instead.

For details, see Engine Environment Variables.

Issues Fixed in Cloudera Data Science Workbench 1.6.0

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue where you had to include pd.options.display.html.table_schema = True to show a horizontal scroll bar for Pandas Dataframe if there were too many columns. You no longer have to include the property.

  Cloudera Bug: DSE-3562
- Fixed an issue where the built-in Workbench editor did not properly recognize imported code that uses tabs instead of spaces. This also resolves navigation issues that occurred within the editor when working with imported code that uses tabs.

  Cloudera Bug: DSE-2976, DSE-3221
- Fixed an issue where an email with attachments triggered by a job fail to send if the attachment is over 4 MB.

  Cloudera Bug: DSE-5980, DSE-6003
- Fixed an issue where large R scripts do not respond when run in the built-in Workbench editor.

  Cloudera Bug: DSE-2817
- Fixed an issue where .md files were not rendered in Markdown. Previously, only README.md was rendered correctly.

  Cloudera Bug: DSE-3315
- Fixed an issue with predict.py, the model training script in the Python template project.

  Cloudera Bug: DSE-5314
- Fixed an issue where logs generated by the Cloudera Data Science Workbench diagnostic bundle were occupying too much space the /var/log/cdsw directory. The size of the generated bundle has been reduced and you can now configure a temporary staging directory to be used when a diagnostic bundle is generated.

  Cloudera Bug: DSE-5921

- The cdsw-build.sh script used with models and experiments now runs as the cdsw user.

  Cloudera Bug: DSE-4340

- The changes to GPU support in version 1.6 have also fixed an issue where GPUs were not automatically detected after a machine reboot.

  Cloudera Bug: DSE-2847

- Fixed an issue where iFrame visualizations would not render in the Workbench due to the new HTTP security headers added in version 1.4.x.

  Cloudera Bug: DSE-5274

Known Issues and Limitations in Cloudera Data Science Workbench 1.6.0

You might run into some known issues while using Cloudera Data Science Workbench.

For a complete of known issues and limitations, see, Cloudera Data Science Workbench 1.8.0.x

Cloudera Data Science Workbench 1.5.0

Major features and updates plus known issues for Cloudera Data Science Workbench 1.5.0.

New Features and Changes in Cloudera Data Science Workbench 1.5.0

Major features and updates for Cloudera Data Science Workbench.

Cloudera Enterprise 6.1 Support

Cloudera Data Science Workbench is now supported with Cloudera Manager 6.1.x (and higher) and CDH 6.1.x (and higher). For details, see Cloudera Manager and CDH Requirements.

Cloudera Data Science Workbench on Hortonworks Data Platform (HDP)

Cloudera Data Science Workbench can now be deployed on HDP 2.6.5 and HDP 3.1.0. For an architecture overview and installation instructions, see Deploying Cloudera Data Science Workbench 1.8.0 on Hortonworks Data Platform.

Security Enhancements

- Allow Site Administrators to Enable/Disable Project Uploads and Downloads - By default, all Cloudera Data Science Workbench users are allowed to upload and download files to/from a project. Version 1.5 introduces a new feature flag that allows site administrators to hide the UI features that let users upload and download project files.

  Note that this feature flag only removes the relevant features from the Cloudera Data Science Workbench UI. It does not disable the ability to upload and download files through the backend web API.

  For details on how to enable this feature, see Disabling Project File Uploads and Downloads.

OpenJDK Support

Cloudera Data Science Workbench now supports Open JDK 8 on Cloudera Enterprise 5.16.1 (and higher). For details, see Product Compatibility Matrix - Supported JDK Versions.

Engines

- Base engine upgraded with a new version of R - 3.5.1 (Base Image v7)

- Debugging Improvements - Previously, engines and their associated logs were deleted immediately after an exit or a crash. With version 1.5, engines are now available for about 5 minutes after they have ended to allow you to collect the relevant logs.

  Additionally, when an engine exits with a non-zero status code, the last 50 lines from the engine's logs are now printed to the Workbench console. Note that a non-zero exit code and the presence of engine logs in the Workbench does not always imply a problem with the code. Events such as session timeouts and out-of-memory issues are also assigned non-zero exit codes and will display engine logs.

Installation and Upgrade

- New Configuration Parameters - Version 1.5 includes three new configuration parameters that can be used to specify the type of distribution you are running, the directory for the installed packages/parcels, and the path where Anaconda is installed (for HDP only).

  - DISTRO
  - DISTRO_DIR
  - ANACONDA_DIR

  Details and sample values for these properties have been added to the relevant installation topics for CDH and HDP.
- DOCKER_TMPDIR changed to /var/lib/cdsw/tmp/docker - Previously the Cloudera Data Science Workbench installer would temporarily decompress the base engine image file to the /var/lib/docker/tmp directory. Starting with version 1.5, the installer will use the /var/lib/cdsw/tmp/docker directory instead. Make sure you have an Application block device mounted to /var/lib/cdsw as recommended so that installation/upgrade can proceed without issues.
- Improved Validation Checks - Improved the validation checks run by the installer and the error messages that are displayed during the installation process. Cloudera Data Science Workbench now:

  - Checks that space is available on the root directory, the Application Block Device and the Docker Block Device(s).
  - Checks that DNS forward and reverse lookup works for the Cloudera Data Science Workbench Domain and Master IP address provided.
  - Displays better error messages for the cdsw status and cdsw validate commands for easier debugging.

Command Line

- cdsw logs - Previously, the cdsw logs command generated two log bundles - one in plaintext and one with sensitive information redacted. With version 1.5, the command now generates only a single bundle that has all the sensitive information redacted by default.

  To turn off redaction of log files for internal use, you can use the new --skip-redaction option as follows:

```
cdsw logs --skip-redaction
```

Networking

- Cloudera Data Science Workbench now uses DNS hostnames (not IP addresses) for internal communication between components. As a result, the wildcard DNS hostname configured for Cloudera Data Science Workbench must now be resolvable from both, the CDSW cluster, and your browser.
- Cloudera Data Science Workbench now enables IPv4 forwarding (net.ipv4.conf.default.forwarding) during the installation process.

Engine Upgrade 1.5.0

Cloudera Data Science Workbench 1.5.0 (and later) ships version 7 of the base engine image.

Cloudera Data Science Workbench 1.5.0 (and later) ships version 7 of the base engine image which includes the following versions of R and Python:

- R - 3.5.1
- Python - 2.7.11, 3.6.1

Pre-installed Packages in Engine 7

For details about the packages included in the base engine, see Cloudera Data Science Workbench Engine Versions and Packaging .

Upgrade Projects to Use the Latest Base Engine Images - Make sure you test and upgrade existing projects to Base Image v7 (Project Settings Engine) to take advantage of the latest fixes.

Note: This is a required step if you are upgrading to using Cloudera Data Science Workbench on CDH 6.

The base engine image you use must be compatible with the version of CDH you are running. This is especially important if you are running workloads on Spark. Older base engines (v6 and lower) cannot support the latest versions of CDH 6. That is because these engines were configured to point to the Spark 2 parcel. However, on CDH 6 clusters, Spark is now packaged as a part of CDH 6 and the separate add-on Spark 2 parcel is no longer supported. If you want to run Spark workloads on CDH 6, you must upgrade your projects to base engine 7 (or higher).

Table 1: CDSW Base Engine Compatibility for Spark Workloads on CDH 5 and CDH 6

Base Engine Versions	CDH 5	CDH 6
Base engines 6 (and lower)	Yes	No
Base engines 7 (and higher)	Yes	Yes

Incompatible Changes in Cloudera Data Science Workbench 1.5.0

Cloudera Data Science Workbench 1.5.0 no longer supports some of the software and components supported in previous releases.

Deprecated Property - CDH Parcel Directory

The CDH parcel directory property is no longer available in the Site Administration panel at Admin Engines . Depending on your deployment, use one of the following ways to configure this property:

- CSD deployments: If you are using the default parcel directory, /opt/cloudera/parcels, no action is required. If you want to use a custom location for the parcel directory, configure this in Cloudera Manager as documented here.
- RPM deployments: If you are using the default parcel directory, /opt/cloudera/parcels, no action is required. If you want to specify a custom location for the parcel directory, configure the DIST RO_DIR property in the cdsw.conf file on both master and worker hosts. Run cdsw restart after you make this change.

Issues Fixed in Cloudera Data Science Workbench 1.5.0

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue with RPM installations where NO_PROXY settings were being ignored.

    Cloudera Bug: DSE-4444

- Fixed an issue where CDSW would not start because of IP issues with web pods. Version 1.5 fixes this by enabling IPv4 forwarding at startup.

  Cloudera Bug: DSE-4609

- Fixed an issue where engines would get deleted immediately after an exit/crash and engine logs did not persist which made it difficult to debug issues with crashes or auto-restarts.

  Cloudera Bug: DSE-4008, DSE-4417

- Fixed intermittent issues with starting and stopping Cloudera Data Science Workbench on CSD deployments.

  Cloudera Bug: DSE-4426, DSE-4829

- Fixed an issue where Cloudera Data Science Workbench was reporting incorrect file sizes for files larger than 2 MB.

  Cloudera Bug: DSE-4531, DSE-4532

- Fixed an issue where the Run New Experiment dialog box did not include the file selector and the Script name had to be typed in manually.

  Cloudera Bug: DSE-3650

- Fixed an issue where underlying Kubernetes processes were running out of resources leading to Out of Memory (OOM) errors. Cloudera Data Science Workbench now reserves compute resources for Kubernetes components.

  Cloudera Bug: DSE-4896, DSE-5001

- Fixed an issue where the PYSPARK3_PYTHON environment variable was not working as expected for Python 3 workloads.

  Cloudera Bug: DSE-4329

- Fixed an issue where Docker commands would fail on Cloudera Data Science Workbench engines that are not available locally (such as custom engine images) when an HTTP/HTTPS proxy was in use.

  Cloudera Bug: DSE-4427

- Fixed an issue where installation of the XML package would fail in the R kernel.

  Cloudera Bug: DSE-2201

Known Issues and Limitations in Cloudera Data Science Workbench 1.5.0

You might run into some known issues while using Cloudera Data Science Workbench.

For a complete list of the current known issues and limitations in Cloudera Data Science Workbench 1.5.x, see Cloudera Data Science Workbench 1.5.0.x.

Cloudera Data Science Workbench 1.4.3

Major features and updates plus known issues for Cloudera Data Science Workbench 1.4.3.

New Features and Changes in Cloudera Data Science Workbench 1.4.3

Major features and updates for Cloudera Data Science Workbench.

Reserve Master Host for Internal Application Components

Cloudera Data Science Workbench now allows you to reserve the master host for running internal application components and services such as Livelog, the PostgreSQL database, and so on, while user workloads run exclusively on worker hosts.

By default, the master host runs both, user workloads as well as the application's internal services. However, depending on the size of your CDSW deployment and the number of workloads running

at any given time, it's possible that user workloads might dominate resources on the master host. Enabling this feature will ensure that CDSW's application components always have access to the resources they need on the master host and are not adversely affected by user workloads.

> **Note:** This feature only applies to deployments with more than one Cloudera Data Science Workbench host. Enabling this feature on single-host deployments will leave Cloudera Data Science Workbench incapable of scheduling any workloads.

For details on how to enable this feature, see Reserving the Master Host for Internal CDSW Components.

Allow Only Session Creators to Execute Commands in Active Sessions

By default, project contributors, project administrators, and site administrators have the ability to execute commands within your actively running sessions in the Workbench. Cloudera Data Science Workbench 1.4.3 introduces a new feature that allows site administrators to restrict this ability. When this feature is enabled, only the user that creates the session will be able to execute commands in that session. No other users, regardless of their permissions in the team or as project collaborators/administrators, will be able to execute commands on active sessions that were not created by them.

For details on how to enable this feature, see Restricting Collaborator and Administrator Access to Active Sessions.

Issues Fixed in Cloudera Data Science Workbench 1.4.3

The current release of Cloudera Data Science Workbench includes fixes for bugs.

TSB-349: SQL Injection Vulnerability in Cloudera Data Science Workbench

An SQL injection vulnerability was found in Cloudera Data Science Workbench. This would allow any authenticated user to run arbitrary queries against CDSW's internal database. The database contains user contact information, bcrypt-hashed CDSW passwords (in the case of local authentication), API keys, and stored Kerberos keytabs.

Products affected: Cloudera Data Science Workbench (CDSW)

Releases affected: CDSW 1.4.0, 1.4.1, 1.4.2

Users affected: All

Date/time of detection: 2018-10-18

Detected by: Anna Szabo-Simon (Cloudera)

Severity (Low/Medium/High): Critical (9.9): CVSS:3.0/AV:N/AC:L/PR:L/UI:N/S:C/C:H/I:H/A:H

Impact: An authenticated CDSW user can arbitrarily access and modify the CDSW internal database. This allows privilege escalation in CDSW, Kubernetes, and the Linux host; creation, deletion, modification, and exfiltration of data, code, and credentials; denial of service; and data loss.

CVE: CVE-2018-20091

Immediate action required:

1. Strongly consider performing a backup before beginning. We advise you to have a backup before performing any upgrade and before beginning this remediation work.
2. Upgrade to Cloudera Data Science Workbench 1.4.3 (or higher).

3. In an abundance of caution Cloudera recommends that you revoke credentials and secrets stored by CDSW. To revoke these credentials:

   a. Change the password for any account with a keytab or kerberos credential that has been stored in CDSW. This includes the Kerberos principals for the associated CDH cluster if entered on the CDSW "Hadoop Authentication" user settings page.

   b. With Cloudera Data Science Workbench 1.4.3 running, run the following remediation script on each CDSW host, including the master and all workers: Remediation Script for TSB-349

   Note: Cloudera Data Science Workbench will become unavailable during this time.

   c. The script performs the following actions:

      1. If using local user authentication, logs out every user and resets their CDSW password.
      2. Regenerates or deletes various keys for every user.
      3. Resets secrets used for internal communications.

   d. Fully stop and start Cloudera Data Science Workbench (a restart is not sufficient).

      • For CSD-based deployments, restart the CDSW service in Cloudera Manager.

        OR
      • For RPM-based deployments, run   cdsw stop   followed by   cdsw start   on the CDSW master host.

   e. If using internal TLS termination: revoke and regenerate the CDSW TLS certificate and key.

   f. For each user, revoke the previous CDSW-generated SSH public key for git integration on the git side (the private key in CDSW has already been deleted). A new SSH key pair has already been generated and should be installed in the old key's place.

   g. Revoke and regenerate any credential stored within a CDSW project, including any passwords stored in projects' environment variables.

4. Verify all CDSW settings to ensure they are unchanged (e.g. SMTP server, authentication settings, custom docker images, host mounts, etc).

5. Treat all CDSW hosts as potentially compromised with root access. Remediate per your policy.

Addressed in release/refresh/patch: Cloudera Data Science Workbench 1.4.3

For the latest update on this issue see the corresponding Knowledge article:

TSB 2019-349: CDSW SQL Injection Vulnerability

TSB-350: Risk of Data Loss During Cloudera Data Science Workbench (CDSW) Shutdown and Restart

Stopping Cloudera Data Science Workbench involves unmounting the NFS volumes that store CDSW project directories and then cleaning up a folder where CDSW stores its temporary state. However, due to a race condition, this NFS unmount process can take too long or fail altogether. If this happens, any CDSW projects that remain mounted will be deleted.

TSB-2018-346 was released in the time-frame of CDSW 1.4.2 to fix this issue, but it only turned out to be a partial fix. With CDSW 1.4.3, we have fixed the issue permanently. However, the script that was provided with TSB-2018-346 still ensures that data loss is prevented and must be used to shutdown/restart all the affected CDSW released listed below. The same script is also available under the Immediate Action Required section below.

Products affected: Cloudera Data Science Workbench

Releases affected: Cloudera Data Science Workbench versions

• 1.0.x
• 1.1.x
• 1.2.x
• 1.3.0, 1.3.1
• 1.4.0, 1.4.1, 1.4.2

Users affected: This potentially affects all CDSW users.

Detected by: Nehmé Tohmé (Cloudera)

Severity (Low/Medium/High): High

Impact: Potential data loss.

CVE: N/A

Immediate action required: If you are running any of the affected Cloudera Data Science Workbench versions, you must run the following script on the CDSW master host every time before you stop or restart Cloudera Data Science Workbench. Failure to do so can result in data loss.

This script should also be run before initiating a Cloudera Data Science Workbench upgrade. As always, we recommend creating a full backup prior to beginning an upgrade.

cdsw_protect_stop_restart.sh - Available for download at: cdsw_protect_stop_restart.sh.

```bash
#!/bin/bash
set -e
cat << EXPLANATION

This script is a workaround for Cloudera TSB-346 and TSB-350. It protects
your
CDSW projects from a rare race condition that can result in data loss.
Run this script before stopping the CDSW service, irrespective of whether
the stop precedes a restart, upgrade, or any other task.

Run this script only on the master node of your CDSW cluster.

You will be asked to specify a target folder on the master node where the
script will save a backup of all your project files. Make sure the target
folder has enough free space to accommodate all of your project files. To
determine how much space is required, run 'du -hs /var/lib/cdsw/current/pro
jects'
on the CDSW master node.
This script will first back up your project files to the specified target
folder. It will then temporarily move your project files aside to protect
against the data loss condition. At that point, it is safe to stop the CDSW
service. After CDSW has stopped, the script will move the project files back
into place.

Note: This workaround is not required for CDSW 1.4.3 and higher.


EXPLANATION

read -p "Enter target folder for backups: " backup_target

echo "Backing up to $backup_target..."
rsync -azp /var/lib/cdsw/current/projects "$backup_target"

read -n 1 -p "Backup complete. Press enter when you are ready to stop CDSW:
 "
echo "Deleting all Kubernetes resources..."
kubectl delete configmaps,deployments,daemonsets,replicasets,services,ingre
ss,secrets,persistentvolumes,persistentvolumeclaims,jobs --all
kubectl delete pods --all

echo "Temporarily saving project files to /var/lib/cdsw/current/projects_t
mp..."
mkdir /var/lib/cdsw/current/projects_tmp
mv /var/lib/cdsw/current/projects/* /var/lib/cdsw/current/projects_tmp
```

```
echo -e "Please stop the CDSW service."
read -n 1 -p "Press enter when CDSW has stopped: "
echo "Moving projects back into place..."
mv /var/lib/cdsw/current/projects_tmp/* /var/lib/cdsw/current/projects
rm -rf /var/lib/cdsw/current/projects_tmp
echo -e "Done. You may now upgrade or start the CDSW service."
echo -e "When CDSW is running, if desired, you may delete the backup data at
 $backup_target"
```

Addressed in release/refresh/patch: This issue is fixed in Cloudera Data Science Workbench 1.4.3.

Note that you are required to run the workaround script above when you upgrade from an affected version to a release with the fix. This helps guard against data loss when the affected version needs to be shut down during the upgrade process.

TSB-351: Unauthorized Project Access in Cloudera Data Science Workbench

Malicious CDSW users can bypass project permission checks and gain read-write access to any project folder in CDSW.

Products affected: Cloudera Data Science Workbench

Releases affected: Cloudera Data Science Workbench 1.4.0, 1.4.1, 1.4.2

Users affected: All CDSW Users

Date/time of detection: 10/29/2018

Detected by: Che-Yuan Liang (Cloudera)

Severity (Low/Medium/High): High (8.3: CVSS :3.0/AV:N/AC:L/PR:L/UI:N/S:U/C:H/I:H/A:L)

Impact: Project data can be read or written (changed, destroyed) by any Cloudera Data Science Workbench user.

CVE: CVE-2018-20090

Immediate action required:

Upgrade to a version of Cloudera Data Science Workbench with the fix (version 1.4.3, 1.5.0, or higher).

Addressed in release/refresh/patch: Cloudera Data Science Workbench 1.4.3 (and higher)

For the latest update on this issue see the corresponding Knowledge article:

TSB 2019-351: Unauthorized Project Access in Cloudera Data Science Workbench

Other Notable Fixed Issues in Cloudera Data Science Workbench 1.4.3

• Fixed an issue where malicious Cloudera Data Science Workbench users were able to bypass project permission checks and gain read-write access to any project folder in Cloudera Data Science Workbench.

  Cloudera Bug: DSE-5138
• Fixed an issue where Cloudera Data Science Workbench would become unresponsive because the web application was making too many simultaneous requests to the Kubernetes API server. CDSW now caches calls to the API and refreshes the cache periodically.

  Cloudera Bug: DSE-5265, DSE-5269
• Fixed an issue where Cloudera Data Science Workbench workloads would intermittently crash with Exit Code 2: Misuse of Shell builtins.

  Cloudera Bug: DSE-4709
• Fixed an issue where Cloudera Data Science Workbench would not start when internal TLS termination was enabled and the TLS private key/certificate pair in use did not include a trailing newline character.

  Cloudera Bug: DSE-4853

Known Issues and Limitations in Cloudera Data Science Workbench 1.4.3

You might run into some known issues while using Cloudera Data Science Workbench.

For a complete list of the current known issues and limitations in Cloudera Data Science Workbench 1.4.x, see restart the Cloudera Data Science Workbench 1.4.x.

Cloudera Data Science Workbench 1.4.2

Major features and updates plus known issues for Cloudera Data Science Workbench 1.4.2.

Note: Cloudera Data Science Workbench 1.4.2 is the next official maintenance release after Cloudera Data Science Workbench 1.4.0. Version 1.4.1 is no longer publicly available.

New Features and Changes in Cloudera Data Science Workbench 1.4.2

Major features and updates for Cloudera Data Science Workbench.

Operating System

Added support for RHEL / CentOS / Oracle Linux RHCK 7.5.

Engines

- Mounts - By default, host mounts (specified at Admin > Engines > Mounts) are loaded into engine containers with read-only permissions. With version 1.4.2, a new checkbox allows you to make these mounted directories available in engine containers with read-write permissions instead.
- Engine upgrade (Base Image v6)

Models

In Cloudera Data Science Workbench 1.4.0, model request sizes were limited to 100 KB. In version 1.4.2, this limit has now been increased to 5 MB. To take advantage of this higher threshold, you will need to upgrade to Cloudera Data Science Workbench 1.4.2 and rebuild your existing models.

Security

Added three new properties to the Admin > Security page that allow you to customize HTTP headers accepted by Cloudera Data Science Workbench.

- Enable HTTP security headers
- Enable cross-origin resource sharing (CORS)
- Enable HTTP Strict Transport Security (HSTS)

For details, see Configuring HTTP Headers for Cloudera Data Science Workbench.

Engine Upgrade 1.4.2

Cloudera Data Science Workbench 1.4.2 ships version 6 of the base engine image.

Cloudera Data Science Workbench 1.5.0 (and later) ships version 7 of the base engine image which includes the following versions of R and Python:

- R - 3.4.1
- Python - 2.7.11, 3.6.1

Pre-installed Packages in Engine 7

For details about the packages included in the base engine, see Cloudera Data Science Workbench Engine Versions and Packaging.

Additionally, Cloudera Data Science Workbench will now alert you when a new engine version is available. Make sure you test and upgrade existing projects to Base Image v6 (Project Settings > Engine) to take advantage of the latest fixes.

Issues Fixed in Cloudera Data Science Workbench 1.4.2

The current release of Cloudera Data Science Workbench includes fixes for bugs.

TSB-346: Risk of Data Loss During Cloudera Data Science Workbench (CDSW) Shutdown and Restart

Stopping Cloudera Data Science Workbench involves unmounting the NFS volumes that store CDSW project directories and then cleaning up a folder where the kubelet stores its temporary state. However, due to a race condition, this NFS unmount process can take too long or fail altogether. If this happens, CDSW projects that remain mounted will be deleted by the cleanup step.

Products affected: Cloudera Data Science Workbench

Releases affected: Cloudera Data Science Workbench versions -

- 1.0.x
- 1.1.x
- 1.2.x
- 1.3.0, 1.3.1
- 1.4.0, 1.4.1

Users affected: This potentially affects all CDSW users.

Detected by: Nehmé Tohmé (Cloudera)

Severity (Low/Medium/High): High

Impact: If the NFS unmount fails during shutdown, data loss can occur. All CDSW project files might be deleted.

CVE: N/A

Immediate action required: If you are running any of the affected Cloudera Data Science Workbench versions, you must run the following script on the CDSW master host every time before you stop or restart Cloudera Data Science Workbench. Failure to do so can result in data loss.

This script should also be run before initiating a Cloudera Data Science Workbench upgrade. As always, we recommend creating a full backup prior to beginning an upgrade.

cdsw_protect_stop_restart.sh - Available for download at: cdsw_protect_stop_restart.sh.

```
#!/bin/bash
set -e
cat << EXPLANATION

This script is a workaround for Cloudera TSB-346. It protects your
CDSW projects from a rare race condition that can result in data loss.
Run this script before stopping the CDSW service, irrespective of whether
the stop precedes a restart, upgrade, or any other task.

Run this script only on the master node of your CDSW cluster.

You will be asked to specify a target folder on the master node where the
script will save a backup of all your project files. Make sure the target
folder has enough free space to accommodate all of your project files. To
```

```
determine how much space is required, run 'du -hs /var/lib/cdsw/current/pro
jects'
on the CDSW master node.
This script will first back up your project files to the specified target
folder. It will then temporarily move your project files aside to protect
against the data loss condition. At that point, it is safe to stop the CDSW
service. After CDSW has stopped, the script will move the project files back
into place.

Note: This workaround is not required for CDSW 1.4.2 and higher.


EXPLANATION

read -p "Enter target folder for backups: " backup_target

echo "Backing up to $backup_target..."
rsync -azp /var/lib/cdsw/current/projects "$backup_target"

read -n 1 -p "Backup complete. Press enter when you are ready to stop CDSW:
 "
echo "Deleting all Kubernetes resources..."
kubectl delete configmaps,deployments,daemonsets,replicasets,services,ingre
ss,secrets,persistentvolumes,persistentvolumeclaims,jobs --all
kubectl delete pods --all

echo "Temporarily saving project files to /var/lib/cdsw/current/projects_t
mp..."
mkdir /var/lib/cdsw/current/projects_tmp
mv /var/lib/cdsw/current/projects/* /var/lib/cdsw/current/projects_tmp

echo -e "Please stop the CDSW service."
read -n 1 -p "Press enter when CDSW has stopped: "
echo "Moving projects back into place..."
mv /var/lib/cdsw/current/projects_tmp/* /var/lib/cdsw/current/projects
rm -rf /var/lib/cdsw/current/projects_tmp
echo -e "Done. You may now upgrade or start the CDSW service."
echo -e "When CDSW is running, if desired, you may delete the backup data at
 $backup_target"
```

Addressed in release/refresh/patch: This issue is fixed in Cloudera Data Science Workbench 1.4.2.

Note that you are required to run the workaround script above when you upgrade from an affected version to a release with the fix. This helps guard against data loss when the affected version needs to be shut down during the upgrade process.

For the latest update on this issue see the corresponding Knowledge article:

TSB 2018-346: Risk of Data Loss During Cloudera Data Science Workbench (CDSW) Shutdown and Restart

TSB-328: Unauthenticated User Enumeration in Cloudera Data Science Workbench

Unauthenticated users can get a list of user accounts of Cloudera Data Science Workbench.

Products affected: Cloudera Data Science Workbench

Releases affected: Cloudera Data Science Workbench 1.4.0 (and lower)

Users affected: All users of Cloudera Data Science Workbench 1.4.0 (and lower)

Date/time of detection: June 11, 2018

Severity (Low/Medium/High): 5.3 (Medium) CVSS :3.0/AV:N/AC:L/PR:N/UI:N/S:U/C:L/I:N/A:N

Impact: Unauthenticated user enumeration in Cloudera Data Science Workbench.

CVE: CVE-2018-15665

Immediate action required: Upgrade to the latest version of Cloudera Data Science Workbench (1.4.2 or higher).

Note that Cloudera Data Science Workbench 1.4.1 is no longer publicly available due to TSB 2018-346: Risk of Data Loss During Cloudera Data Science Workbench (CDSW) Shutdown and Restart.

Addressed in release/refresh/patch: Cloudera Data Science Workbench 1.4.2 (and higher)

For the latest update on this issue see the corresponding Knowledge article:

TSB 2018-318: Unauthenticated User Enumeration in Cloudera Data Science Workbench

Other Notable Fixed Issues in Cloudera Data Science Workbench 1.4.2

- Fixed an issue where attempting to fork a large project would result in unexpected 'out of memory' errors.

  Cloudera Bug: DSE-4464
- Fixed an issue in version 1.4.0 where Cloudera Data Science Workbench workloads would intermittently get stuck in the Scheduling state due to a Red Hat kernel slab leak.

  Cloudera Bug: DSE-4098
- Fixed an issue in version 1.4.0 where the Hadoop username on non-kerberized clusters defaulted to cdsw. This was a known issue and has been fixed in version 1.4.2. The Hadoop username will now once again default to your Cloudera Data Science Workbench username.

  Cloudera Bug: DSE-4240
- Fixed an issue in version 1.4.0 where creating a project using Git via SSH did not work.

  Cloudera Bug: DSE-4278
- Fixed an issue in version 1.4.0 where environmental variables set in the Admin panel were not being propagated to projects (experiments, sessions, jobs) as expected.

  Cloudera Bug: DSE-4422
- Fixed an issue in version 1.4.0 where Cloudera Data Science Workbench would not start when external TLS termination was enabled.

  Cloudera Bug: DSE-4640
- Fixed an issue in version 1.4.0 where HTTP/HTTPS proxy settings in Cloudera Manager were erroneously escaped when propagated to Cloudera Data Science Workbench engines.

  Cloudera Bug: DSE-4421
- Fixed an issue in version 1.4.0 where SSH tunnels did not work as expected.

  Cloudera Bug: DSE-4741
- Fixed an issue in version 1.4.0 where copying multiple files into a folder resulted in unexpected behavior such as overwritten files and incorrect UI messages.

  Cloudera Bug: DSE-4831
- Fixed an issue in version 1.4.0 where workers (in engines) and collection of usage metrics failed on TLS-enabled clusters.

  Cloudera Bug: DSE-4293, DSE-4572
- Fixed an issue in version 1.4.0 where the Files New Folder dialog box did not work.

  Cloudera Bug: DSE-4807
- Fixed an issue in version 1.4.0 where deleting an experiment did not work from certain dashboards. Consequently, deleting the parent project would also fail in such cases.

  Cloudera Bug: DSE-4227

Known Issues and Limitations in Cloudera Data Science Workbench 1.4.2

You might run into some known issues while using Cloudera Data Science Workbench.

For a complete list of the current known issues and limitations in Cloudera Data Science Workbench 1.4.x, see Cloudera Data Science Workbench 1.4.x.

Cloudera Data Science Workbench 1.4.0

Major features and updates plus known issues for Cloudera Data Science Workbench 1.4.0.

New Features in Cloudera Data Science Workbench 1.4.0

Major features and updates for Cloudera Data Science Workbench.

Models and Experiments

Cloudera Data Science Workbench 1.4 extends the machine learning platform experience from research to production. Now you can use Cloudera Data Science Workbench to build, train, and deploy models in a unified workflow.

- Experiments - Train and compare versioned, reproducible models
- Models - Deploy and manage models as REST APIs to serve predictions

External Authentication

LDAP/SAML users can now restrict access to Cloudera Data Science Workbench to specific LDAP/SAML groups. Additionally, you can now specify groups that should automatically be granted site administrator privileges when they log in to *Cloudera Data Science Workbench. For details, see Configuring External Authentication with LDAP and SAML.*

Cloudera Data Science Workbench now supports multiple identity provider signing certificates for SAML 2.0 authentication.

Cloudera Data Science Workbench now supports SAML 2.0 Errata 05 E43 for SAML 2.0 authentication.

Projects and Workbench

- Site administrators can now disable individual built-in template projects by using a checkbox in the Project Templates table at Admin Settings . Only enabled project templates will be displayed in the dropdown menu when creating a new project.
- The default .gitignore file that is created with each new project has been updated to:

```
R
node_modules
*.pyc
.*
!.gitignore
```

- Added support for multiple Terminal windows within a single session.

Networking

- Cloudera Data Science Workbench now supports DNS resolution of localhost to non-local IP address (not 127.0.0.1).

- Cloudera Data Science Workbench now appends the following default values to the NO_P ROXY parameter if any of the following properties are configured: HTTP_PROXY, HTTPS_PR OXY, or ALL_PROXY.

```
"127.0.0.1,localhost,100.66.0.1,100.66.0.2,100.66.0.3,
100.66.0.4,100.66.0.5,100.66.0.6,100.66.0.7,100.66.0.8,100.66
.0.9,
100.66.0.10,100.66.0.11,100.66.0.12,100.66.0.13,100.66.0.14,
100.66.0.15,100.66.0.16,100.66.0.17,100.66.0.18,100.66.0.19,
100.66.0.20,100.66.0.21,100.66.0.22,100.66.0.23,100.66.0.24,
100.66.0.25,100.66.0.26,100.66.0.27,100.66.0.28,100.66.0.29,
100.66.0.30,100.66.0.31,100.66.0.32,100.66.0.33,100.66.0.34,
100.66.0.35,100.66.0.36,100.66.0.37,100.66.0.38,100.66.0.39,
100.66.0.40,100.66.0.41,100.66.0.42,100.66.0.43,100.66.0.44,
100.66.0.45,100.66.0.46,100.66.0.47,100.66.0.48,100.66.0.49,
100.66.0.50,100.77.0.10,100.77.0.128,100.77.0.129,100.77.0.130
,
100.77.0.131,100.77.0.132,100.77.0.133,100.77.0.134,100.77.0.1
35,
100.77.0.136,100.77.0.137,100.77.0.138,100.77.0.139"
```

Installation Validation Checks

Improved validation checks run during the installation process. Cloudera Data Science Workbench now:

- Verifies that the wildcard DNS subdomain has been configured.
- Verifies that resolv.conf is not pointing to 127.0.0.1.
- Validates iptables chains to ensure there are no custom rules being set.
- Throws a warning if you are using a self-signed TLS certificate, an expired certificate, or if the certificate is not valid for the wildcard domain used for Cloudera Data Science Workbench.

Command Line

Added a verbose option to the cdsw status command.

```
cdsw status [-v|--verbose]
```

Kubernetes has been upgraded to version 1.8.12.

Engine Upgrade 1.4.0

Cloudera Data Science Workbench 1.4.0 ships version 5 of the base engine image.

Cloudera Data Science Workbench 1.4.0 (and later) ships version 5 of the base engine image which includes the following versions of R and Python:

- R - 3.4.1
- Python - 2.7.11, 3.6.1

Pre-installed Packages in Engine 7

For details about the packages included in the base engine, see Cloudera Data Science Workbench Engine Versions and Packaging.

Additionally, Cloudera Data Science Workbench will now alert you when a new engine version is available. Make sure you test and upgrade existing projects to Base Image v5 (Project Settings > Engine) to take advantage of the latest fixes.

Incompatible Changes in Cloudera Data Science Workbench 1.4.0

Cloudera Data Science Workbench 1.4.0 no longer supports some of the software and components supported in previous releases.

Host Mounts are now Read-Only in Engines - Previously, mounts (specified at Admin Engines Mounts) were loaded into engine containers with read-write permissions.

Starting with version 1.4.0, mount points are now loaded into engines with read-only permissions.

Issues Fixed in Cloudera Data Science Workbench 1.4.0

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue where Git would timeout when cloning a project took too long. The timeout has now been increased to 60 seconds when creating a new project from Git.

  Cloudera Bug: DSE-3363
- Fixed an issue where manual parcel deployments could not detect parcel hash files with a .sha1 extension.

  Cloudera Bug: DSE-3301
- Fixed several usability issues (file create, save, and so on) with Internet Explorer 11.

  Cloudera Bug: DSE-3426, DSE-3434
- Fixed an issue where CSD installations would fail to recognize Oracle Linux 7.3 as a supported operating system.

  Cloudera Bug: DSE-3257
- Fixed an issue where Cloudera Data Science Workbench would not respond with 100 percent CPU utilization.

  Cloudera Bug: DSE-3450
- Fixed a SAML 2.0 configuration issue where uploading the identity provider metadata XML file did not update identity provider signing certificate and/or SSO URL on Cloudera Data Science Workbench correctly.

  Cloudera Bug: DSE-3076
- Fixed an issue with SAML 2.0 authentication where the identity provider's signature was not being validated correctly.

  Cloudera Bug: DSE-3694
- Fixed the Save As functionality in the project Workbench.

  Cloudera Bug: DSE-3870
- Fixed an issue where if a user had some files opened in the Workbench in a previous session, and those files no longer existed in the project filesystem, a File Not Found error would occur when opening the Workbench.

  Cloudera Bug: DSE-3835

Known Issues and Limitations in Cloudera Data Science Workbench 1.4.0

You might run into some known issues while using Cloudera Data Science Workbench.

For a complete list of the current known issues and limitations in Cloudera Data Science Workbench 1.4.x, see Cloudera Data Science Workbench 1.4.0.x.

Cloudera Data Science Workbench 1.3.1

Major features and updates plus known issues for Cloudera Data Science Workbench 1.3.1.

New Features in Cloudera Data Science Workbench 1.3.1

Major features and updates for Cloudera Data Science Workbench.

- Operating System: Added support for RHEL / CentOS / Oracle Linux RHCK 7.5.
- SAML
  - Cloudera Data Science Workbench now supports multiple identity provider signing certificates for SAML 2.0 authentication.
  - Cloudera Data Science Workbench now supports SAML 2.0 Errata 05 E43 for SAML 2.0 authentication.

Issues Fixed in Cloudera Data Science Workbench 1.3.1

The current release of Cloudera Data Science Workbench includes fixes for bugs.

Remote Command Execution and Information Disclosure in Cloudera Data Science Workbench

A configuration issue in Kubernetes used by Cloudera Data Science Workbench can allow remote command execution and privilege escalation in CDSW. A separate information permissions issue can cause the LDAP bind password to be exposed to authenticated CDSW users when LDAP bind search is enabled.

Products affected: Cloudera Data Science Workbench

Releases affected: Cloudera Data Science Workbench 1.3.0 (and lower)

Users affected: All users of Cloudera Data Science Workbench 1.3.0 (and lower)

Date/time of detection: May 16, 2018

Severity (Low/Medium/High): High

Impact: Remote command execution and information disclosure

CVE: CVE-2018-11215

Immediate action required: Upgrade to the latest version of Cloudera Data Science Workbench (1.3.1 or higher) and change the LDAP bind password if previously configured in Cloudera Data Science Workbench.

Addressed in release/refresh/patch: Cloudera Data Science Workbench 1.3.1 (and higher)

For the latest update on this issue see the corresponding Knowledge Base article:

TSB: 2018-313: Remote Command Execution and Information

Other Notable Fixed Issues in Cloudera Data Science Workbench 1.3.1

- Fixed an issue where CSD installations would fail to recognize Oracle Linux 7.3 as a supported operating system.

  Cloudera Bug: DSE-3257
- Fixed several usability issues (file create, save, and so on) with Internet Explorer 11.

  Cloudera Bug: DSE-3426, DSE-3434
- Fixed a SAML 2.0 configuration issue where uploading the identity provider metadata XML file did not update identity provider signing certificate and/or SSO URL on Cloudera Data Science Workbench correctly.

  Cloudera Bug: DSE-3265
- Fixed an issue where the owner of a console output could not view their own shared consoles from sessions /job runs when sharing with Specific user/team.

  Cloudera Bug: DSE-3143
- Fixed issue with missing connectors in Jobs dependency chart.

  Cloudera Bug: DSE-3185

Known Issues and Limitations in Cloudera Data Science Workbench 1.3.1

You might run into some known issues while using Cloudera Data Science Workbench.

For a list of the current known issues and limitations, refer the documentation for version 1.3.x at Cloudera Data Science Workbench 1.3.x.

Cloudera Data Science Workbench 1.3.0

Major features and updates plus known issues for Cloudera Data Science Workbench 1.3.0.

New Features and Changes in Cloudera Data Science Workbench 1.3.0

Major features and updates for Cloudera Data Science Workbench.

- Added support for SUSE Linux Enterprise Server 12 SP3.
- Site administrators can now add template projects that are customized for their organization's use-cases.
- Version 1.3 introduces a new environment variable for Python 3 sessions called PYSPARK3_PYTHON. Python 2 sessions will continue to use the default PYSPARK_PYTHON variable. This will allow you to configure distinct variables for Python 2 and Python 3 applications.
- In the Cloudera Manager CDSW service, the Wildcard DNS Domain property has been renamed to Cloudera Data Science Workbench Domain.
- Output for the cdsw version command now includes the type of deployment you are running – RPM or CSD.
- Added log4j and spark-defaults sample configuration to the PySpark and Scala template projects.

Issues Fixed in Cloudera Data Science Workbench 1.3.0

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue where the cdsw status command failed to run all the required system checks.

  Cloudera Bug: DSE-3070
- Session lists now include additional metadata to help distinguish between different sessions.

  Cloudera Bug: DSE-2814
- Pre-install validation checks have been improved to detect issues with iptables modules and Java settings.

  Cloudera Bug: DSE-2293
- Fixed an issue with the cdsw status command output when TLS is enabled.

  Cloudera Bug: DSE-3182
- CDS 2.2 Release 2 fixes the issue where a PySpark application could only be run once per active Workbench session.

  Cloudera Bug: CDH-58475
- Fixed an issue that prevented Bokeh plots from rendering.

  Cloudera Bug: DSE-3134
- Fixed an issue in Cloudera Data Science Workbench 1.2.2 that prevented WebSocket re-connections and caused the console to stop responding.

  Cloudera Bug: DSE-3085
- Improved CDSW service restart performance for CSD deployments.

  Cloudera Bug: DSE-2937

Incompatible Changes in Cloudera Data Science Workbench 1.3.0

Cloudera Data Science Workbench 1.3.0 no longer supports some of the software and components supported in previous releases.

Deploying Cloudera Data Science Workbench with Cloudera Director 2.7

While this is not a Cloudera Data Science Workbench change, you should note that Cloudera Director 2.7 includes a new instance-level setting that sets the mountAllUnmountedDisks property to false:

```
normalizationConfig {
 mountAllUnmountedDisks: false
}
```

This means Cloudera Director 2.7 (and higher) users no longer need to set lp.normalization.mountAllUnmountedDisksRequired to false in the Cloudera Director server's application.properties file. Note that Cloudera Director 2.6 still requires this setting.

Known Issues and Limitations in Cloudera Data Science Workbench 1.3.0

You might run into some known issues while using Cloudera Data Science Workbench.

For a list of the current known issues and limitations, refer the documentation for version 1.3.x at Cloudera Data Science Workbench 1.3.x.

Cloudera Data Science Workbench 1.2.2

Major features and updates plus known issues for Cloudera Data Science Workbench 1.2.2.

This section lists the release notes for Cloudera Data Science Workbench 1.2.2. The documentation for version 1.2.x can be found at Cloudera Data Science Workbench 1.2.x.

New Features and Changes in Cloudera Data Science Workbench 1.2.2

Major features and updates for Cloudera Data Science Workbench.

- Added support for SUSE Linux Enterprise Server 12 SP2.
- Added support for multi-homed networks.
- Cloudera Director now allows you to deploy CSD-based Cloudera Data Science Workbench 1.2.x deployments on AWS. For more specifics on supported platforms, see Cloudera Altus Director Support (AWS and Azure Only).
- Added a new environment variable called MAX_TEXT_LENGTH that allows you to set the maximum number of characters that can be displayed in a single text cell. By default, this value is set to 800,000 and any more characters will be truncated.

Engine Upgrade 1.2.2

Cloudera Data Science Workbench 1.2.2 ships version 4 of the base engine image.

Cloudera Data Science Workbench 1.2.2 (and later) ships version 4 of the base engine image which includes bug fixes related to Python development and Kerberos authentication. Engine 4 ships the following versions of R and Python:

- R - 3.4.1
- Python - 2.7.11, 3.6.1

For details about the packages included in the base engine, see Cloudera Data Science Workbench Engine Versions and Packaging.

Make sure you upgrade existing projects to Base Image v4 (Project Settings > Engine) to take advantage of these fixes.

The new engine also changes how you configure and use Conda in Python sessions and extended engines. For more details, see Using Conda with Cloudera Data Science Workbench.

Issues Fixed In Cloudera Data Science Workbench 1.2.2

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue where Conda environmental variables were not being propagated to the Terminal correctly.

  Cloudera Bug: DSE-2256
- Fixed an issue where GPUs were not being detected by Cloudera Data Science Workbench due to incorrect mount settings.

  Cloudera Bug: DSE-2957
- Fixed an issue where jobs were failing due to Kerberos TGT renewal issues.

  Cloudera Bug: DSE-1007
- Fixed an issue on Internet Explorer 10 and 11 where the browser would fail to render console output after launching too many interactive sessions.

  Cloudera Bug: DSE-2998, DSE-2979
- Cloudera Data Science Workbench now correctly renders HTML that contains iFrames with the srcdoc attribute.

  Cloudera Bug: DSE-2034
- Fixed an issue where logging in using LDAP/Active Directory would sometimes crash the Cloudera Data Science Workbench web application.

  Cloudera Bug: DSE-2672
- The file tree in the Workbench now refreshes correctly when switching between sessions or launching a new session.

  Cloudera Bug: DSE-2829
- Fixed a file descriptors leak that would cause the "Failed to get Kubernetes client configuration" error in Cloudera Manager.

  Cloudera Bug: DSE-2910
- Fixed an issue where the host-controller process was consuming too much CPU. This was occurring due to a bug in the Kubernetes client-go library.

  Cloudera Bug: DSE-2993

Known Issues and Limitations in Cloudera Data Science Workbench 1.2.2

You might run into some known issues while using Cloudera Data Science Workbench.

For a list of known issues and limitations, refer the documentation for version 1.2.x at Cloudera Data Science Workbench 1.2.x.

Cloudera Data Science Workbench 1.2.1

Major features and updates plus known issues for Cloudera Data Science Workbench 12.1.

The documentation for version 1.2.x can be found at Cloudera Data Science Workbench 1.2.x.

Issues Fixed In Cloudera Data Science Workbench 1.2.1

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- The Master Node IPv4 Address parameter has been added to Cloudera Manager's Add Service wizard and is now a required parameter for installation on AWS. This should fix any related installation issues for deployments on AWS.

  Cloudera Bug: DSE-2879
- Fixed an issue with CSD-based deployments where certain operations would fail because the Prepare Node command was not installing all the required packages during First Run of the service. To see the updated list of packages that are now being installed by the Prepare Node command, refer the CSD Install Guide.

  Cloudera Bug: DSE-2869
- Fixed an issue where the LD_LIBRARY_PATH environmental variable was not getting propagated to CUDA engines.

  Cloudera Bug: DSE-2828
- Fixed an issue where stopping Cloudera Data Science Workbench on worker hosts resulted in the application hanging indefinitely.

  Cloudera Bug: DSE-2880

Incompatible Changes in Cloudera Data Science Workbench 1.2.1

Upgrading from Cloudera Data Science Workbench 1.2.0 to 1.2.1 on CSD-based deployments

After upgrading from Cloudera Data Science Workbench 1.2.0 to 1.2.1 on a CSD-based deployment, CLI commands might not work as expected due to missing binaries in the environment. Note that this issue does not affect fresh installs.

Known Issues and Limitations in Cloudera Data Science Workbench 1.2.1

You might run into some known issues while using Cloudera Data Science Workbench.

For a list of known issues and limitations, refer the documentation for version 1.2.x at Cloudera Data Science Workbench 1.2.x.

Cloudera Data Science Workbench 1.2.0

Major features and updates plus known issues for Cloudera Data Science Workbench 1.2.0.

The documentation for version 1.2.x can be found at Cloudera Data Science Workbench 1.2.x.

New Features and Changes in Cloudera Data Science Workbench 1.2.0

Major features and updates for Cloudera Data Science Workbench.

- Cloudera Data Science Workbench is now available as an add-on service for Cloudera Manager. To this end, Cloudera Data Science Workbench is now distributed in a parcel that integrates with Cloudera Manager using a Custom Service Descriptor (CSD). You can use Cloudera Manager to install, upgrade, and monitor Cloudera Data Science Workbench. Diagnostics data bundles can be generated and submitted to Cloudera through Cloudera Manager.

- Cloudera Data Science Workbench now enables secure sharing of job and session consoles. Additionally, site administrators can disable anonymous sharing from the Site Administrator dashboard (Admin Security). See Sharing Job and Session Console Outputs
- The Admin Usage page now includes graphs for monitoring usage activity such as number of CPUs or GPUs used, memory usage, and total session runs, over customizable periods of time.
- Cloudera Data Science Workbench now lets you configure session, job, and idle timeouts. These can be configured using environmental variables either for the entire deployment or per-project.
- The cdsw enable and disable commands are no longer needed. The master host will now automatically detect the IP addresses of worker hosts joining or leaving Cloudera Data Science Workbench. See the revisedCloudera Data Science Workbench Command Line Reference.
- The Kudu Python client is now included in the Cloudera Data Science Workbench base engine image.
- Interactive session names can now be modified by project contributors and admins. By default, session names are set to 'Untitled Session'.
- All-numeric usernames are now accepted.
- Kubernetes has been upgraded to version 1.6.11.

Engine Upgrade

- Cloudera Data Science Workbench 1.2.0 ships version 3 of the base engine image which includes matplotlib improvements and the Kudu client libraries. Engine 3 ships the following versions of R and Python:

  - R - 3.4.1
  - Python - 2.7.11, 3.6.1

  Make sure you upgrade existing projects to Base Image v3 (Project Settings Engine) to take advantage of the new features and bug fixes included in the new engine.

Engine Upgrade 1.2.0

Cloudera Data Science Workbench 1.2.0 ships version 3 of the base engine image.

Cloudera Data Science Workbench 1.2.0 ships version 3 of the base engine image which includes matplotlib improvements and the Kudu client libraries. Engine 3 ships the following versions of R and Python:

- R - 3.4.1
- Python - 2.7.11, 3.6.1

Pre-installed Packages in Engine 7

Make sure you upgrade existing projects to Base Image v3 (Project Settings > Engine) to take advantage of the new features and bug fixes included in the new engine.

Issues Fixed in Cloudera Data Science Workbench 1.2.0

The current release of Cloudera Data Science Workbench includes fixes for bugs.

Privilege Escalation and Database Exposure in Cloudera Data Science Workbench

Several web application vulnerabilities allowed malicious authenticated Cloudera Data Science Workbench (CDSW) users to escalate privileges in CDSW. In combination, such users could exploit these vulnerabilities to gain root access to CDSW hosts, gain access to the CDSW database which includes Kerberos keytabs of CDSW users and bcrypt hashed passwords, and obtain other privileged information such as session tokens, invitations tokens, and environmental variables.

Products affected: Cloudera Data Science Workbench

Releases affected: Cloudera Data Science Workbench 1.0.0, 1.0.1, 1.1.0, 1.1.1

Users affected: All users of Cloudera Data Science Workbench 1.0.0, 1.0.1, 1.1.0, 1.1.1

Date/time of detection: September 1, 2017

Detected by: NCC Group

Severity (Low/Medium/High): High

Impact: Privilege escalation and database exposure.

CVE: CVE-2017-15536

Addressed in release/refresh/patch: Cloudera Data Science Workbench 1.2.0 or higher.

Immediate action required: Upgrade to the latest version of Cloudera Data Science Workbench.

Other Notable Fixed Issues in Cloudera Data Science Workbench 1.2.0

- Fixed an issue where the Workbench editor screen jumps unexpectedly when typing or scrolling.
- Fixed auto-scroll behavior in the Workbench console. This was a browser compatibility issue that affected Chrome and Firefox, but not Safari.
- Fixed an issue where if a user logged out of Cloudera Data Science Workbench, and logged back in as a different user, they may see a SecurityError message in the Workbench.
- Fixed an issue that was preventing site administrators from uploading the SAML metadata file.
- Fixed several issues related to plotting with matplotlib. If you have previously used any workarounds for plotting, you might consider removing them now.
- Engines now use the same build of Kerberos utilities (ktutil, kinit, and klist) as the rest of Cloudera Data Science Workbench. This will improve logs obtained from kinit and make debugging Kerberos issues easier.
- KRB5_TRACE is now included in the error logs obtained when you kinit.
- Fixed an issue that was affecting health checks in deployments using AWS elastic load balancing.

Incompatible Changes in Cloudera Data Science Workbench 1.2.0

DSE-2948 Proxy Configuration Change: If you are using a proxy server, you must ensure that the IP addresses for the web and Livelog services are skipped from the proxy.

Depending on your deployment (parcel or package), append the following IP addresses to either the No Proxy property in the Cloudera Manager CDSW service, or to the NO_PROXY parameter in cdsw.conf.

```
100.77.0.129
100.77.0.130
```

These have also been added to the installation instructions.

Known Issues and Limitations in Cloudera Data Science Workbench 1.2.0

You might run into some known issues while using Cloudera Data Science Workbench.

For a list of known issues and limitations, refer the documentation for version 1.2.x at Cloudera Data Science Workbench 1.2.x.

Cloudera Data Science Workbench 1.1.1

Major features and updates plus known issues for Cloudera Data Science Workbench 1.1.1.

The documentation for version 1.1.x can be found at Cloudera Data Science Workbench 1.1.x.

New Features in Cloudera Data Science Workbench 1.1.1

Major features and updates for Cloudera Data Science Workbench.

- Keytab Authentication - With version 1.1.1, you can now authenticate yourself to the CDH cluster by uploading your Kerberos keytab to Cloudera Data Science Workbench. To use this feature, go to the top-right dropdown menu, click Account settings Hadoop Authentication , enter your Kerberos principal and click Upload Keytab.

Issues Fixed In Cloudera Data Science Workbench 1.1.1

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed an issue with airgapped installations where the installer could not pull the alpine 3.4 image into the airgapped environment.
- Fixed an issue where Cloudera Data Science Workbench would fail to log a command trace when the Kerberos process exits.
- Fixed authentication issues with older versions of MIT KDC.

Known Issues and Limitations in Cloudera Data Science Workbench 1.1.1

You might run into some known issues while using Cloudera Data Science Workbench.

For a list of known issues and limitations, refer the documentation for version 1.1.x at Cloudera Data Science Workbench 1.1.x.

Cloudera Data Science Workbench 1.1.0

Major features and updates plus known issues for Cloudera Data Science Workbench 1.1.0.

The documentation for version 1.1.x can be found at Cloudera Data Science Workbench 1.1.x.

New Features and Changes in Cloudera Data Science Workbench 1.1.0

Major features and updates for Cloudera Data Science Workbench.

- Added support for RHEL/CentOS 7.3 and Oracle Linux 7.3.
- Cloudera Data Science Workbench now allows you to run GPU-based workloads. For more details, see Using NVIDIA GPUs for Cloudera Data Science Workbench Projects.
- For Cloudera Manager and CDH clusters that are not connected to the Internet, Cloudera Data Science Workbench now supports fully offline installations. See the installation guide for more details.
- Web UIs for processing frameworks such as Spark 2, Tensorflow, and Shiny, are now embedded in Cloudera Data Science Workbench and can be accessed directly from active sessions and jobs. For more details, see Accessing Web User Interfaces from Cloudera Data Science Workbench.
- Added support for a Jobs REST API that lets you orchestrate jobs from 3rd party workflow tools. See Cloudera Data Science Workbench Jobs API.
- DataFrames are now scrollable in the workbench session output pane. For examples, see the section on grid displays.
- Added support for rich visualizations in Scala engine using Jupyter jvm-repr. For an example, see HTML Visualizations - Scala.
- JAVA_HOME is now set in cdsw.conf, and not from the Site Administrator dashboard (Admin Engines).

Engine Upgrade 1.1.0

Cloudera Data Science Workbench 1.1.0 ships version 2 of the base engine image.

Cloudera Data Science Workbench 1.1.0 ships version 2 of the base engine image that includes new versions of Pandas, seaborn, and assorted bug fixes. Engine 2 ships the following versions of R and Python:

- R - 3.3.0
- Python - 2.7.11, 3.6.1

Make sure you upgrade existing projects to Base Image v2 (Project Settings > Engine) to take advantage of the new features and bug fixes included in the new engine.

Issues Fixed in Cloudera Data Science Workbench 1.1.0

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Improved support for dynamic data visualizations in Python, including Bokeh.
- Fixed issues with the Python template project. The project now supports offline mode and will therefore work on airgapped clusters.
- Fixed issues related to cached responses in Internet Explorer 11.
- Fixed issues with Java symlinks outside of JAVA_HOME.
- The cdsw status command can now be run on worker hosts.
- Removed unauthenticated localhost access to Kubernetes.
- Fixed Kerberos authentication issues with specific enc-types and Active Directory.
- Removed restrictions on usernames with special characters for better compatibility with external authentication systems such as Active Directory.
- Fixed issues with LDAP configuration validation that caused application crashes.
- Improved LDAP test configuration form to avoid confusion on parameters being sent.

Incompatible Changes in Cloudera Data Science Workbench 1.1.0

- DSE-2088 Upgrading from version 1.0.x to 1.1.x

  During the upgrade process, you will encounter incompatibilities between the two versions of cdsw.conf. This is because even though you are installing the latest RPM, your previous configuration settings in cdsw.conf will remain unchanged. Depending on the release you are upgrading from, you will need to modify cdsw.conf to ensure it passes the validation checks run by the 1.1.x release.

  Key changes to note:

  - JAVA_HOME is now a required parameter. Make sure you add JAVA_HOME to cdsw.conf before you start Cloudera Data Science Workbench.
  - Previous versions allowed MASTER_IP to be set to a DNS hostname. If you are still using a DNS hostname, switch to an IP address.
- DSE-2133 Python engine updated in version 1.1.x

  Version 1.1.x includes an updated base engine image for Python which no longer uses the deprecated pylab mode in Jupyter to import the numpy and matplotlib functions into the global scope. With version 1.1.x, engines will now use built-in functions like any rather than the pylab counterpart, numpy.any. As a result of this change, you might see certain behavioral changes and differences in results between the two versions.

  Also note that Python projects originally created with engine 1 will be running pandas version 0.19, and will not auto-upgrade to version 0.20 by simply selecting engine 2. You will also need to manually install version 0.20.1 of pandas when you launch a project session.

Known Issues and Limitations in Cloudera Data Science Workbench 1.1.0

You might run into some known issues while using Cloudera Data Science Workbench.

For a list of known issues and limitations, refer the documentation for version 1.1.x at Cloudera Data Science Workbench 1.1.x.

Cloudera Data Science Workbench 1.0.1

Major features and updates plus known issues for Cloudera Data Science Workbench 1.0.1.

The documentation for version 1.0.x can be found at Cloudera Data Science Workbench 1.0.x.

Issues Fixed in Cloudera Data Science Workbench 1.0.1

The current release of Cloudera Data Science Workbench includes fixes for bugs.

- Fixed a random port conflict that could prevent Scala engines from running.
- Improved formatting of validation, and visibility of some errors.
- Fixed an issue with Firefox that was resulting in duplicate jobs on job creation.
- Removed the Mathjax external dependency on CDN.
- Improved PATH and JAVA_HOME handling that previously broke Hadoop CLIs.
- Fixed an issue with Java security policy files that caused Kerberos issues.
- Fixed an issue that caused git clone to fail on some repositories.
- Fixed an issue where updating LDAP admin settings deactivated the local fallback login.
- Fixed an issue where bad LDAP configuration crashed the application.
- Fixed an issue where job environmental variable settings did not persist.

Known Issues and Limitations in Cloudera Data Science Workbench 1.0.x

You might run into some known issues while using Cloudera Data Science Workbench.

For a list of known issues and limitations, refer the documentation for version 1.0.x at Cloudera Data Science Workbench 1.0.x.

Cloudera Data Science Workbench 1.0.0

Major features and updates plus known issues for Cloudera Data Science Workbench 1.0.0.

Version 1.0 represents the first generally available (GA) release of Cloudera Data Science Workbench. For information about the main features and benefits of Cloudera Data Science Workbench, as well as an architectural overview of the product, see Cloudera Data Science Workbench Overview.

Known Issues and Limitations in Cloudera Data Science Workbench 1.9.0

This topic lists the current known issues and limitations in Cloudera Data Science Workbench 1.9.0.

- There is an issue with installing and upgrading CDSW in an airgapped environment where docker service has no access to the k8s docker repository or has no proxy set. .

  Workaround:

  The following steps must be executed after deploying the CDSW parcel and DOCKER role is started but the MASTER, APPLICATION, and WORKER roles are not yet started. In case of an RPM install, this needs to be done after starting the dockerd service but before starting the CDSW master node.

  1. Pull the missing etcd image to local docker repo in a node that has access to the k8s.gcr.io docker repository.

     ```
     # docker pull k8s.gcr.io/etcd:3.3.15-0
     ```

  2. Save the missing etcd image.

     ```
     # docker save k8s.gcr.io/etcd:3.3.15-0 -o etcd.tar
     ```

  3. Copy tarball to the CDSW master node.

     ```
     # scp etcd.tar <user>@<node>:<location>/etcd.tar
     ```

  4. SSH in to the CDSW master node and load the image to the local docker repo.

     ```
     # docker load -i <location>/etcd.tar
     ```

For previous versions, see:

- Known Issues in Cloudera Data Science Workbench 1.8.x
- Known Issues in Cloudera Data Science Workbench 1.7.x
- Known Issues in Cloudera Data Science Workbench 1.6.x
- Known Issues in Cloudera Data Science Workbench 1.5.x
- Known Issues in Cloudera Data Science Workbench 1.4.x
- Known Issues in Cloudera Data Science Workbench 1.3.x
- Known Issues in Cloudera Data Science Workbench 1.2.x
- Known Issues in Cloudera Data Science Workbench 1.1.x.

Installation

Known installation issues.

- Installation error highlights a wrong path. When there is insufficient space left for DOCKER_TMPDIR (configurable parameter, set to /var/lib/cdsw/docker-tmp by default), the error message refers to a wrong directory: "Please make sure that you have sufficient space under /var/lib/docker/tmp " This message will be fixed in CDSW 1.10.

  Due to storage requirements on the worker nodes to handle docker images, this directory should have at least 20GB allocated. On the master node this should be 1TB.

  Cloudera Bug: DSE-11086/DSE-10767

- Installation error highlights a wrong path. When there is insufficient space left for DOCKER_TMPDIR (configurable parameter, set to /var/lib/cdsw/docker-tmp by default), the error message refers to a wrong directory: "Please make sure that you have sufficient space under /var/lib/docker/tmp " This message will be fixed in CDSW 1.10.

  Due to storage requirements on the worker nodes to handle docker images, this directory should have at least 20GB allocated. On the master node this should be 1TB.

  Cloudera Bug: DSE-10767

- During the Cloudera Data Science Workbench startup process, you might see certain timeout issues.

```
Pods not ready in cluster default ['role/<pod_name>'].
```

  This is due to an issue with some pods taking longer to start up and other dependent processes timing out. Restart the CDSW service to get past this issue.

  Cloudera Bug: DSE-6855
- Cloudera Data Science Workbench cannot be managed by Apache Ambari.

Upgrades

Please read the following upgrade issues before you being the upgrade process.

symlinks do not work in CDSW 1.9.x and CDSW 1.10.x

"If /var/lib/cdsw is a symlink to another location, CDSW might fail to start. If this occurs, you will see various errors in the ds-vfs pod about MountVolume.Setup fail, or No Such Project"   errors: mount.nfs: mounting 192.168.225. 133:/var/lib/cdsw/current/projects failed,   reason given by server: No such file or directory.

Workaround:

Avoid using a symlink for /var/lib/cdsw.

Cloudera Bug: DSE-22863

Upgrades supported from CDSW 1.7.x (and higher) to CDSW 1.9.x

Upgrades supported from CDSW 1.7.x (and higher) to CDSW 1.9.x Cloudera Data Science Workbench only supports upgrades to version 1.9.0 from version 1.7.x and 1.8.x. If you are using an earlier version, you must first upgrade to version 1.7.x or 1.8.x, and then upgrade to version 1.9.0.

Issue with Upgrading CDSW in an airgapped environment

There is an issue with upgrading CDSW in an airgapped environment where docker service has no access to the k8s docker repository or has no proxy set. .

Workaround:

Upgrade to CDSW 1.9.1.

Alternatively, if you choose to continue using CDSW 1.9.0, you can complete the following steps. The following steps must be executed after deploying the CDSW parcel and DOCKER role is started but the MASTER, APPLICATION, and WORKER roles are not yet started. In case of an RPM install, this needs to be done after starting the dockerd service but before starting the CDSW master node.

1. Pull the missing etcd image to local docker repo in a node that has access to the k8s.gcr.io docker repository.

```
# docker pull k8s.gcr.io/etcd:3.3.15-0
```

2. Save the missing etcd image.

```
# docker save k8s.gcr.io/etcd:3.3.15-0 -o etcd.tar
```

3. Copy tarball to the CDSW master node.

```
# scp etcd.tar <user>@<node>:<location>/etcd.tar
```

4. SSH in to the CDSW master node and load the image to the local docker repo.

```
# docker load -i <location>/etcd.tar
```

Now you can start the rest of the roles for parcel installs and bring up cdsw cluster in case of an rpm install.

Cloudera Bug: DSE-13956

CDH Integration

Known CDH integration issues.

CDH client configuration changes require a full Cloudera Data Science Workbench restart

Cloudera Data Science Workbench does not automatically detect configuration changes on the CDH cluster. Therefore, any changes made to CDH services, ranging from updates to service configuration properties to complete CDH or CDS parcel upgrades, must be followed by a full reset of Cloudera Data Science Workbench.

Workaround: Depending on your deployment, use one of the following sets of steps to perform a full reset of Cloudera Data Science Workbench. Note that this reset does not impact your data in any way.

- CSD Deployments - To reset Cloudera Data Science Workbench using Cloudera Manager:

    1. Log into the Cloudera Manager Admin Console.
    2. On the Cloudera Manager homepage, click ⏷ to the right of the CDSW service and select Restart. Confirm your choice on the next screen and wait for the action to complete.

    OR

- RPM Deployments - Run the following steps on the Cloudera Data Science Workbench master host:

```
cdsw stop
cdsw start
```

Cloudera Manager Integration

Known Cloudera Manager integration issues.

CSD distribution/activation fails on mixed-OS clusters when there are third-party parcels running on OSs that are not supported by Cloudera Data Science Workbench

For example, adding a new CDSW gateway host on a RHEL 6 cluster running RHEL-6 compatible parcels will fail. This is because Cloudera Manager will not allow distribution of the RHEL 6 parcels on the new host which will likely be running a CDSW-compatible operating system such as RHEL 7.

Workaround: To ensure adding a new CDSW gateway host is successful, you must create a copy of the 'incompatible' third-party parcel files and give them the corresponding RHEL 7 names so that Cloudera Manager allows them to be distributed on the new gateway host. Use the following sample instructions to do so:

1. SSH to the Cloudera Manager Server host.
2. Navigate to the directory that contains all the parcels. By default, this is /opt/cloudera/parcel-repo.

```
cd /opt/cloudera/parcel-repo
```

3. Make a copy of the incompatible third-party parcel with the new name. For example, if you have a RHEL 6 parcel that cannot be distributed on a RHEL 7 CDSW host:

```
cp <PARCELNAME.cdh5.x.x.p0.123>-el6.parcel <PARCELNAME.cdh5.x.x.p0.123>-
el7.parcel
```

4. Repeat the previous step for parcel's SHA file.

```
cp <PARCELNAME.cdh5.x.x.p0.123>-el6.parcel.
sha <PARCELNAME.cdh5.x.x.p0.123>-el7.parcel.sha
```

5. Update the new files' owner and permissions to match those of existing parcels in the /opt/cloudera/parcel-repo directory.

```
chown cloudera-scm:cloudera-scm <PARCELNAME.cdh5.x.x.p0.123>-el7.parcel
chown cloudera-scm:cloudera-scm <PARCELNAME.cdh5.x.x.p0.123>-el7.parcel.
sha
chmod 640 <PARCELNAME.cdh5.x.x.p0.123>-el7.parcel
chmod 640 <PARCELNAME.cdh5.x.x.p0.123>-el7.parcel.sha
```

You should now be able to add new gateway hosts for Cloudera Data Science Workbench to your cluster.

Cloudera Bug: OPSAPS-42130, OPSAPS-31880

CDSW Service health status after a restart does not match the actual state of the application

After a restart, the Cloudera Data Science Workbench service in Cloudera Manager will display Good health even though the Cloudera Data Science Workbench web application might need a few more minutes to get ready to serve requests.

Apache Spark

Known issues with Apache Spark.

Monitoring Spark Application

To monitor spark_on_yarn applications invoked from CDSW, an embedded Spark UI is displayed right next to the session/job. This was achieved by disabling RM proxy. However, with this change, attempts to access the same Spark application using the RM UI will result in Error 500 (connection refused).

Workaround: If the Administrator wants to troubleshoot a running spark-on-yarn application invoked by an end-user from the workbench, the user must share their session using the Share button on the right side of the console. An alternate workaround which will not provide realtime updates is to access the Spark Application UI from the Spark History Server UI > Incomplete Applications.

Cloudera Bug: DSE-4979

Spark UI does not work on HDP and CDP

The Spark UI in CDSW does not work on HDP and CDP clusters.

Scala sessions can fail if dependencies take longer than 15 minutes

If the dependencies in spark-defaults.conf (spark.jars, spark.packages, etc) take longer than 15 minutes to resolve, then scala sessions will fail the first time.

Workaround: Use one of the following workarounds:

- Restart the session.
- Mount the Spark dependency directory from the CDSW host machines.

On TLS-enabled CDSW deployments, the embedded Spark UI does not work

If you have a TLS-enabled CDSW deployment, the embedded Spark UI tab does not render as expected.

Workaround: To work around this issue, launch the Spark UI in a separate tab and append '/jobs' after the URL. For example, if your engineID is tb0z9ydiua5q9v2d and the DOMAIN is example.com then view the Spark UI at: https://spark-tb0z9ydiua5q9v2d.example.com/jobs/

Alternative workaround: To view running Spark jobs, navigate to Spark History Server UI Show Incomplete Applications Application ID

Affected Versions: This issue affects CDSW 1.6.x and CDSW 1.7.x on the following platforms:

- CDH 5: CDS 2.4 release 2 (and lower)
- CDH 6: Versions of Spark that ship with CDH 6.0.x, CDH 6.1.x, CDH 6.2.1 (and lower), CDH 6.3.2 (and lower)

Solution: Upgrade to CDSW version 1.7.1 or higher, and either:

- CDH version 6.4.0, 6.2.2, 6.3.3 or higher
- CDH 5 with Spark 2.4 release 3

Spark lineage collection is not supported with Cloudera Data Science Workbench

Lineage collection is enabled by default in Spark 2.3. This feature does not work with Cloudera Data Science Workbench because the lineage log directory is not automatically mounted into CDSW engines when a session/job is started.

Affected Versions: CDS 2.3 release 2 (and higher) Powered By Apache Spark

With Spark 2.3 release 3 (or higher), if Spark cannot find the lineage log directory, it will automatically disable lineage collection for that application. Spark jobs will continue to execute in Cloudera Data Science Workbench, but lineage information will not be collected.

With Spark 2.3 release 2, Spark jobs will fail in Cloudera Data Science Workbench. Either upgrade to Spark 2.3 release 3 which includes a partial fix (as described above) or use one of the following workarounds to disable Spark lineage:

Workaround 1: Disable Spark Lineage Per-Project in Cloudera Data Science Workbench

To do this, set spark.lineage.enabled to false in a spark-defaults.conf file in your Cloudera Data Science Workbench project. This will need to be done individually for each project as required.

Workaround 2: Disable Spark Lineage for the Cluster

1. Log in to Cloudera Manager and go to the Spark 2 service.
2. Click Configuration.
3. Search for the Enable Lineage Collection property and uncheck the checkbox to disable lineage collection.
4. Click Save Changes.
5. Go back to the Cloudera Manager homepage and restart the CDSW service for this change to go into effect.

Cloudera Bug: DSE-3720, CDH-67643

Crashes and stops responding

Known issues with crashes and unresponsive behaviours.

- The CDSW terminal remains active even after the CDSW web session times out.

  Cloudera Bug: DSE-12064
- The CDSW web UI might freeze while updating the job schedule to Dependant if there is no other job to be depended on.

- Third-party security and orchestration software (such as McAfee, Tanium, Symantec) can sometimes prevent CDSW from restarting correctly

  Workaround: Disable all third-party security agents on CDSW hosts, start or restart CDSW, and then when CDSW is running, re-enable the security software.

  Cloudera Bug: DSE-8550, DSE-14287
- High I/O utilization on the application block device can cause the application to stall or become unresponsive. Users should read and write data directly from HDFS rather than staging it in their project directories.
- Installing ipywidgets or a Jupyter notebook into a project can cause Python engines to stop responding due to an unexpected configuration. The issue can be resolved by deleting the installed libraries from the R engine terminal.

Third-party Editors

Known issues with third-party editors.

- You cannot disable file upload and download when using the Jupyter Notebook.

  Cloudera Bug: DSE-12065
- Logs generated by a browser IDE do not appear within the IDE. They are displayed in the Logs tab for the session.

  Cloudera Bug: DSE-6570
- Sessions with Browser IDEs running do not adhere to the limit set in IDLE_MAXIMUM_MINUTES. Session logs show the warning message that states that the idle session will timeout, but the timeout does not occur. The session continues to run and consume resources until the timeout set in SESSION_MAXIMUM_MINUTES is reached. Ensure that you manually stop a session after you are finished, so that the resources are available to other users.

  Cloudera Bug: DSE-6651
- Sessions with Browser IDEs running time out with no warning after the time limit set in SESSION_MAXIMUM_ MINUTES is reached, regardless of whether or not the session is idle. Periodically stop the browser IDE and session manually to avoid reaching SESSION_MAXIMUM_MINUTES.

  Cloudera Bug: DSE-6652
- The lack of a ROOT CA certificate can cause issues with terminals and the Jupyter editor after upgrading CDSW.

  Problem: After upgrading from CDSW version 1.5 to version 1.7.1, the terminal does not open for any kernel, and the Jupyter notebook does not work.

  Workaround: In CDSW, go to Admin Security , and paste the internal CA root certificate file contents directly into the Root CA configuration field. You should be able to launch a new session and start the terminal or launch the Jupyter editor. It is not necessary to restart CDSW. This procedure is described at Configuring Custom Root CA Certificate

Cloudera Data Science Workbench Engines

Cloudera Data Science Workbench engines are responsible for running R, Python, and Scala code written by users and intermediating access to the CDH cluster.

You can think of an engine as a virtual machine, customized to have all the necessary dependencies to access the CDH cluster while keeping each project's environment entirely isolated. To ensure that every engine has access to the parcels and client configuration managed by the Cloudera Manager Agent, a number of folders are mounted from the host into the container environment. This includes the parcel path -/opt/cloudera, client configuration, as well as the host's JAVA_HOME. For more details on basic concepts and terminology related to engines in Cloudera Data Science Workbench, see Cloudera Data Science Workbench Engines.

Known Issues

- Apache Phoenix requires additional configuration to run commands successfully from within Cloudera Data Science Workbench engines (sessions, jobs, experiments, models).

  Workaround

  Explicitly set HBASE_CONF_PATH to a valid path before running Phoenix commands from engines.

  ```
  export HBASE_CONF_PATH=/usr/hdp/hbase/<hdp_version>/0/
  ```

Runtimes

Known ML Runtimes issues.

- Code completion in the Workbench editor does not work when using Runtimes. It does work in the console on the right hand side, and it does work in both the editor and console when using Legacy Engines.
- Limitations for Spark support for ML Runtimes:

  - ML Runtimes running Python 3.8 kernel does not support running Spark.
  - ML Runtimes on CDH 7.x does not support running Spark.

    Cloudera Bug: DSE-13916

- ML Runtimes do not support the CLI interface (cdswctl command) for the current release.

  Cloudera Bug: DSE-13699

- In order to use Spark with ML Runtimes on Cloudera Data Science Workbench, prior to using ML Runtimes the first time, you must install py4j. As part of the Session, run the following:

  ```
  run pip install py4j
  ```

- Jupyter Notebook sessions in legacy engine:8-engine:13 do not exit after IDLE_MAXIMUM_MINUTES of inactivity. They will run until   SESSION_MAXIMUM_MINUTES (which is seven days by default). .

  Workaround:You can change the configuration of your cluster to apply the fix for this issue. Change the editor command for Jupyter Notebook in every engine that uses it to the following:

  ```
  NOTEBOOK_TIMEOUT_SECONDS=$(python3 -c "print(${IDLE_MAXIMUM_MINUTES}*60)
  ") /usr/local/bin/jupyter notebook --no-browser --ip=127.0.0.1 --port=${
  CDSW_APP_PORT} --NotebookApp.token= --NotebookApp.allow_remote_access=True
   --NotebookApp.quit_button=False --log-level=ERROR --NotebookApp.shutdown_
  no_activity_timeout=300 --MappingKernelManager.cull_idle_timeout=${NOTEB
  OOK_TIMEOUT_SECONDS} -- TerminalManager.cull_inactive_timeout=${NOTEBOOK
  _TIMEOUT_SECONDS} --MappingKernelManager.cull_interval=60 --TerminalMana
  ger.cull_interval=60 --MappingKernelManager.cull_connected=True
  ```

  This does the following:

  - Kills each running notebook after IDLE_MAXIMUM_MINUTES of inactivity
  - Kills the CDSW/CML session in which Jupyter is running after 5 minutes with no notebooks

  Cloudera Bug: DSE-13741, DSE-6651

Custom Legacy Engine Images

Known custom legacy engine images issues.

- Cloudera Data Science Workbench only supports customized engines that are based on the Cloudera Data Science Workbench base image.

- The contents of certain pre-existing standard directories such as /home/cdsw, /tmp, /opt/cloudera, and so on, cannot be modified while creating customized engines. This means any files saved in these directories will not be accessible from sessions that are running on customized engines.

  Workaround: Create a new custom directory in the Dockerfile used to create the customized engine, and save your files to that directory. Or, create a new custom directory on all the Cloudera Data Science Workbench gateway hosts and save your files to those directories. Then, mount this directory to the custom engine.

Experiments

Starting with version 1.4, Cloudera Data Science Workbench allows data scientists to run batch experiments that track different versions of code, input parameters, and output (both metrics and files).

Demo: Watch the following video for a quick demonstration of the steps described in this topic: Experiments with Cloudera Data Science Workbench

GPU Support

Known issues with GPU support.

Only CUDA-enabled NVIDIA GPU hardware is supported

Cloudera Data Science Workbench only supports CUDA-enabled NVIDIA GPU cards.

Heterogeneous GPU hardware is not supported

You must use the same GPU hardware across a single Cloudera Data Science Workbench deployment.

Known Issues with Tensorflow 2.4 and CUDA 11.2

During GPU set up, the dynamic library libculsolver.so.10 is not read. For more information, see https://github.com/tensorflow/tensorflow/issues/44777.

Workaround: Enter the following commands when starting a session:

```
!ln -s /usr/local/cuda-11.1/targets/x86_64-linux/lib/libcusolver.so.11.0.1.1
05 /usr/local/cuda-11.1/targets/x86_64-linux/lib/libcusolver.so.10
!ln -s /usr/lib/x86_64-linux-gnu/libcuda.so.460.73.01 /usr/lib/x86_64-linux-
gnu/libcuda.so.1
```

Jobs

Known issues with jobs.

- When you start a job that has a dependent job, CDSW does not log the job start event for the dependent job in the user_events table.

  Cloudera Bug: DSE-11855
- Cloudera Data Science Workbench does not support changing your API key, or having multiple API keys.
- Currently, you cannot use the Jobs API to create a job, stop a job, or get the status of a job.

Models

Known issues for models.

- Model request timeout

  You can set the model request timeout duration to a custom value. The default value is 30 seconds. The timeout can be changed if model requests might take more than 30 seconds.

  To set the timeout value:

  1. As an Admin user, open a CLI.
  2. At the prompt, execute the following command. Substitute <value> with the number of seconds to set.

  ```
  kubectl set env deployment model-proxy MODEL_REQUEST_TIMEOUT_SECONDS=<va
  lue> -n mlx
  ```

  This edits the kubeconfig file and sets a new value for the timeout duration.

- • (If quotas are enabled) Models that are stuck in the Scheduled state due to lack of resources do not automatically start even if you free up existing resources.

  Workaround: Stop the Model that is stuck in the Scheduled state. Then manually reschedule that Model.

  Cloudera Bug: DSE-6886

- Unable to create a model with the name of a deleted Model.

  Workaround: For now, Models shall have unique names across the lifespan of the cluster installation.

  Cloudera Bug: DSE-4237

- Re-deploying or re-building models results in model downtime (usually brief).
- Model deployment will fail if your project filesystem is too large for the Git snapshot process. As a general rule, any project files (code, generated model artifacts, dependencies, etc.) larger than 50 MB must be part of your project's .gitignore file so that they are not included in snapshots for model builds.
- JSON requests made to active models should not be more than 5 MB in size. This is because JSON is not suitable for very large requests and has high overhead for binary objects such as images or video. Call the model with a reference to the image or video, such as a URL, instead of the object itself.
- Any external connections, for example, a database connection or a Spark context, must be managed by the model's code. Models that require such connections are responsible for their own setup, teardown, and refresh.
- Model logs and statistics are only preserved so long as the individual replica is active. Cloudera Data Science Workbench may restart a replica at any time it is deemed necessary (such as bad input to the model).
- The use_model_metrics.py file which is available within the CDSW Templates misses the code for setting the user_api_key and is not up-to-date. Use the following code instead:

```
import cdsw
import time
from sklearn import datasets
import numpy as np
# This script demonstrates the usage of several model metrics-
# related functions:
# - call_model: Calls a model deployed on CDSW as an HTTP endpoint.
# - read_metrics: Reads metrics tracked for all model predictions
#   made within a time window. This is useful for  doing analytics
#   on the tracked metrics.
# - track_delayed_metrics: Adds metrics for a given prediction
#   retrospectively, after the prediction has already been made.
#   Common examples of such metrics are ground truth and various
#   per-prediction accuracy metrics.
# - track_aggregate_metrics: Adds metrics for a set or batch of
#   predictions within a given time window, not an individual
#   prediction. Common examples of such metrics are mean or
#   median accuracy, and various measures of drift.

# This script can be used in a local development mode, or in
# deployment mode. To use it in deployment mode, please:
# - Set dev = False
```

```
# - Create a model deployment from the function 'predict' in
#   predict_with_metrics.py
# - Obtain the model deployment's CRN from the model's overview
#   page and the model's access key from its settings page and
#   paste them below.
# - If you selected "Enable Authentication" when creating the
#   model, then create a model API key from your user settings
#   page and paste it below as well.

dev = True

# Conditionally import the predict function only if we are in
# dev mode
try:
    if dev:
        raise RuntimeError("In dev mode")
except:
    from predict_with_metrics import predict

if dev:
    model_deployment_crn=cdsw.dev_model_deployment_crn # update modelDep
loymentCrn
    model_access_key=None
else:
    # The model deployment CRN can be obtained from the model overview
    # page.
    model_deployment_crn=None
    if model_deployment_crn is None:
        raise ValueError("Please set a valid model deployment Crn")

    # The model access key can be obtained from the model settings page.
    model_access_key=None
    if model_access_key is None:
        raise ValueError("Please set the model's access key")

    # You can create a models API key from your user settings page.
    # Not required if you did not select "Enable Authentication"
    # when deploying the model. In that case, anyone with the
    # model's access key can call the model.
    user_api_key = None

# First, we use the call_model function to make predictions for
# the held-out portion of the dataset in order to populate the
# metrics database.
iris = datasets.load_iris()
test_size = 20
# This is the input data for which we want to make predictions.
# Ground truth is generally not yet known at prediction time.
score_x = iris.data[:test_size, 2].reshape(-1, 1) # Petal length
# Record the current time so we can retrieve the metrics
# tracked for these calls.
start_timestamp_ms=int(round(time.time() * 1000))

uuids = []
predictions = []
for i in range(len(score_x)):
    if model_access_key is not None:
        output = cdsw.call_model(model_access_key, {"petal_length": score
_x[i][0]}, api_key=user_api_key)["response"]
    else:
        output = predict({"petal_length": score_x[i][0]})
    # Record the UUID of each prediction for correlation with ground tr
uth.
    uuids.append(output["uuid"])
```

```
        predictions.append(output["prediction"])

    # Record the current time.
    end_timestamp_ms=int(round(time.time() * 1000))

    # We can now use the read_metrics function to read the metrics we just
    # generated into the current session, by querying by time window.
    data = cdsw.read_metrics(model_deployment_crn=model_deployment_crn,
                start_timestamp_ms=start_timestamp_ms,
                end_timestamp_ms=end_timestamp_ms, dev=dev)
    data = data['metrics']

    # Now, ground truth is known and we want to track the true value
    # corresponding to each prediction above.
    score_y = iris.data[:test_size, 3].reshape(-1, 1) # Observed petal width
    # Track the true values alongside the corresponding predictions using
    # track_delayed_metrics. At the same time, calculate the mean absolute
    # prediction error.
    mean_absolute_error = 0
    n = len(score_y)
    for i in range(n):
        ground_truth = score_x[i][0]
        cdsw.track_delayed_metrics({"actual_result":ground_truth}, uuids[i],
    dev=dev)

        absolute_error = np.abs(ground_truth - predictions[i])
        mean_absolute_error += absolute_error / n

    # Use the track_aggregate_metrics function to record the mean absolute
    # error within the time window where we made the model calls above.
    cdsw.track_aggregate_metrics(
        {"mean_absolute_error": mean_absolute_error},
        start_timestamp_ms,
        end_timestamp_ms,
        model_deployment_crn=model_deployment_crn,
        dev=dev
    )
```

- Limitations

    - Scala models are not supported.
    - Spawning worker threads is not supported with models.
    - Models deployed using Cloudera Data Science Workbench are not highly-available.
    - Dynamic scaling and auto-scaling are not currently supported. To change the number of replicas in service, you will have to re-deploy the build.

Applications

Known issues with applications.

- Analytical Application settings do not take into account Application specific environmental variables, therefore this option is currently not visible in the UI.
- The subdomain names on the **Application** page has some constraints, but the CDSW UI may not display a complete error message when these are violated.

    Workaround: Use the characters from the set of ASCII letters, digits, and hyphens (a-z, 0-9, -) to form the subdomain name. Ensure that the subdomain name does not start or end with a hyphen.

    Cloudera Bug: DSE-11883

Platform

Known issues with platforms.

- In the monitoring view of the Grafana the UI, the values displayed in the Used CPU Usage and the Total CPU Usage are the same when both the Node and the Pod filters are set to All.

  Workaround: Use a specific filter other than All to overcome this issue.

  Cloudera Bug: DSE-12072
- Deleting the projects through the UI might leave residual files on disk.

  Workaround: Please contact Cloudera support to remove residual files in case of disk usage issues.

  Cloudera Bug: DSE-393

Networking

Known issues with networking.

- Custom /etc/hosts entries on Cloudera Data Science Workbench hosts do not propagate to sessions and jobs running in containers.

  Cloudera Bug: DSE-2598
- Initialisation of Cloudera Data Science Workbench (cdsw init) will fail if localhost does not resolve to 127.0.0.1.
- Cloudera Data Science Workbench does not support DNS servers running on

  127.0.0.1:53 . This IP address resolves to the container localhost within Cloudera Data Science Workbench containers. As a workaround, use either a non-loopback address or a remote DNS server.
- Kubernetes throws the following error when /etc/resolv.conf lists more than three domains:

```
Resolv.conf file '/etc/resolv.conf' contains search line consisting of m
ore than 3 domains!
```

  Due to a limitation in the libc resolver, only two DNS servers are supported in /etc/resolv.conf. Kubernetes uses one additional entry for the cluster DNS.

Security

Known issues with security.

Working in the terminal or an editor should not count as idle session

If a user opens a workbench and is either working exclusively in the terminal or just editing files, Cloudera Data Science Workbench counts that time as idle time and the user gets kicked out after the configured max idle timeout.

Workaround:

- Increase the idle session timeout by adding a new environmental variable IDLE_MAXIMUM_MINUTES. Click CDSW > Project > Settings > Environmental variables.Note: This approach would require you to keep your containers running.

  You can set the value of the variables IDLE_MAXIMUM_MINUTES or SESSION_MAXIMUM_MINUTES to their maximum allowed value, which is 35000 (~3 weeks).
- Alternatively, run a simple script inside CDSW session to keep the session alive. Opening the Cloudera Data Science Workbench and create a file as shown here (assuming Python project), and then run it in the Workbench.

```
import time
time.sleep(10000)
```

Cloudera Bug: DSE-3080

SSH access to Cloudera Data Science Workbench hosts must be disabled

The container runtime and application data storage is not fully secure from untrusted users who have SSH access to the gateway hosts. Therefore, SSH access to the gateway hosts for untrusted users should be disabled for security and resource utilization reasons.

TLS/SSL

- Self-signed certificates where the Certificate Authority is not part of the user's trust store are not supported for TLS termination.
- Cloudera Data Science Workbench does not support the use of encrypted private keys for TLS.

Cloudera Bug: DSE-1708

- A "certificate has expired" error displays when you log in to the Cloudera Data Science Workbench web UI. This issue can occur if Cloudera Data Science Workbench exceeds 365 days of continuous uptime because the internal certificate for Kubernetes expires after 1 year.

Workaround: Restart the Cloudera Data Science Workbench deployment.

  - For CSD installations, restart the Cloudera Data Science Workbench service in Cloudera Manager.
  - For RPM installations, run the following command on the Master host:

```
cdsw restart
```

Kerberos

- Using Kerberos plugin modules in krb5.conf is not supported.
- Modifying the default_ccache_name parameter in krb5.conf does not work in Cloudera Data Science Workbench. Only the default path for this parameter, /tmp/krb5cc_${uid}, is supported.
- PowerBroker-equipped Active Directory is not supported.

Cloudera Bug: DSE-1838

- When you upload a Kerberos keytab to authenticate yourself to the CDH cluster, Cloudera Data Science Workbench might display a fleeting error message ('cancelled') in the bottom right corner of the screen, even if authentication was successful. This error message can be ignored.

Cloudera Bug: DSE-2344

Usability

Known issues with usability.

- An application can become stuck with the status 'starting' and the user might not be able to find the 'restart' or 'delete' action button.This could be caused by a failure in the selected application script.The exact errors can be found in the logs of the application in the individual description page.

Workaround: Stop the application from the UI and then Restart the application once the issue in the script is addressed.

- The Options tab on the Project Settings page is now only viewable to the Administrator. This is because the Private and Team visibility options on the Options page are only viewable by the Administrator.
- Environment variables with the dollar ($) character are not parsed correctly by CDSW. For example, if you set PASSWORD="pass$123" in the project environment variables, and then try to read it using the echo command, you see the following output: pass23

Workaround: Use one of the following commands to print the $ sign:

```
echo 24 | xxd -r -p
```

```
or
echo JAo= | base64 -d
```

Insert the value of the environment variable by wrapping it in the command substitution using $() or ``. For example, if you want to set the environment variable to ABC$123, specify:

```
ABC$(echo 24 | xxd -r -p)123
or
ABC`echo 24 | xxd -r -p`123
```

- In some cases, the application switcher (grid icon) does not show any other applications, such as Hue or Ranger.

  Cloudera Bug: DSE-865

- Scala sessions do not respond when running large scripts (longer than 100 lines) in the Workbench editor.

  Workaround 1:

  Execute the script in manually-selected chunks. For example, highlight the first 50 lines and select Run Run Line(s) .

  Workaround 2:

  Restructure your code by moving content into imported functions so as to bring the size down to under 100 lines.
- The R engine is unable to display multi-byte characters in plots. Examples of multi-byte characters include languages such as Korean, Japanese, and Chinese.

  Workaround: Use the   showtext   R package to support more fonts and characters. For example, to display Korean characters:

  ```
  install.packages('showtext')
  library(showtext)
  font_add_google("Noto Sans KR", "noto")
  showtext_auto()
  ```

  Cloudera Bug: DSE-7308

- In a scenario where 100s of users are logged in and creating processes, the nproc and nofile limits of the system may be reached. Use ulimits or other methods to increase the maximum number of processes and open files that can be created by a user on the system.
- When rebooting, Cloudera Data Science Workbench hosts can take a significant amount of time (about 30 minutes) to become ready.
- Long-running operations such as fork and clone can time out when projects are large or connections outlast the HTTP timeouts of reverse proxies.
- The Scala kernel does not support auto-complete features in the editor.
- Scala and R code can sometimes indent incorrectly in the workbench editor.

  Cloudera Bug: DSE-1218

General

Known installation issues.

- During the Cloudera Data Science Workbench startup process, you might see certain timeout issues.

  ```
  Pods not ready in cluster default ['role/<pod_name>'].
  ```

  This is due to an issue with some pods taking longer to start up and other dependent processes timing out. Restart the CDSW service to get past this issue.

  Cloudera Bug: DSE-6855

- Cloudera Data Science Workbench restricts a maximum of 50 concurrent Sessions, Jobs, Models, Applications, and Spark Executors that can run parallel by single users. You can use the Workers API to create additional pods which count toward the 50 limit.
- Cloudera Data Science Workbench cannot be managed by Apache Ambari.
- Apache Phoenix requires additional configuration to run commands successfully from within Cloudera Data Science Workbench engines (sessions, jobs, experiments, models). Workaround: Explicitly set HBASE_CONF_PATH to a valid path before running Phoenix commands from engines.

```
export HBASE_CONF_PATH=/usr/hdp/hbase/<hdp_version>/0/
```