Cloudera Stream Processing 3.0.0

Cloudera Data Flow for Data Hub Release

Notes

Date published: 2019-12-16 Date modified: 2019-12-16



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

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

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

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

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

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

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

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

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

Contents

What's new in Cloudera Data Flow for Data Hub	
What's new in Apache Kafka	4
What's new in Streams Messaging Manager	5
What's new in Schema Registry	5
Component support	5
Known issues	6
	7
Fixed issues	-
Fixed issues Fixed issues in Apache Kafka	
Fixed issues in Apache Kafka Fixed issues in Streams Messaging Manager	
Fixed issues in Apache Kafka Fixed issues in Streams Messaging Manager Fixed issues in Schema Registry	

What's new in Cloudera Data Flow for Data Hub

This section lists major features and updates for Cloudera Data Flow for Data Hub.



Cloudera Data Flow for Data Hub is available for Technical Preview. Cloudera encourages you to explore these technical preview features in non-production environments and provide feedback on your experiences through the Cloudera Support Portal.

December 18, 2019

This release introduces the technical preview release of Streams Messaging cluster definitions for installation using Data Hub. The Streams Messaging templates include Kafka, Schema Registry, Streams Messaging Manager, and ZooKeeper. There are two template options, depending on your operational objectives:

- Streams Messaging Heavy Duty
- Streams Messaging Light Duty

Streams Messaging provides advanced messaging and real-time processing on streaming data using Apache Kafka, centralized schema management using Schema Registry, as well as management and monitoring capabilities powered by Streams Messaging Manager.

These templates set up fault-tolerant standalone deployments of Apache Kafka and supporting Cloudera components (Schema Registry and Streams Messaging Manager), which can be used for Kafka workloads in the cloud or as a disaster recovery instance for on-prem Kafka clusters.

See the following sections for information about highlighted new features added to the Streams Messaging cluster definition components.

What's new in Apache Kafka

This topic lists new features for Apache Kafka in this release of Cloudera Data Flow for Data Hub.

Rebase on Apache Kafka 2.3.0

Kafka available with this version of Cloudera Data Flow for Data Hub is based on Apache Kafka 2.3.0. For more information, see Apache Kafka Notable Changes and Apache Kafka Release Notes in the upstream documentation.

Access to Kafka Metadata in Zookeeper is restricted by default

The Enable Zookeeper ACL (zookeeper.set.acl) property is now directly configurable in Cloudera Manager and is enabled by default. As a result of this change, access to Kafka metadata stored in Zookeeper is restricted by default. The data is still world readable, however, administrative operations, for example topic creation, deletion, any configuration changes and so on, can only be performed by authorized users. For more information, see Restrict access to Kafka metadata in Zookeeper and Unlock Kafka metadata in Zookeeper.

Ranger authorization support

Ranger support for Kafka is added. You can now use Ranger to provide authorization for Kafka. For more information, see Using Ranger to Provide Authorization in CDP as well as the documentation on Kafka Authorization with Ranger.

4

The resource-based Ranger service used by Kafka is user configurable

The resource-based Ranger service used by Kafka for authorization can now be manually configured with the Ranger service name for this Kafka cluster property in Cloudera Manager. In addition, if a resource-based service is set in Kafka that does not yet exist in Ranger, it will be automatically created after the Kafka service is restarted. The name of the newly created service is based on the value of the Ranger service name for this Kafka cluster property. For more information, see Configure the resource-based Ranger service used for authorization.

PAM authentication support

You can now configure Kafka to authenticate clients using PAM. For more information, see PAM Authentication.

LDAP authentication support

You can now configure Kafka to authenticate clients using LDAP. For more information, see LDAP Authentication.

New metric for monitoring garbage collector runs

A new metric called kafka_jvm_gc_runs is added to the Kafka service. This metric enables users to monitor the number of garbage collector runs performed on each broker.

What's new in Streams Messaging Manager

This topic lists new features for Streams Messaging Manager in this release of Cloudera Data Flow for Data Hub.

- Trusted Proxy support for Kafka SMM REST agent. (CSP-108)
- Consistent SSL configuration settings for SMM UI and Server. (CSP-305)
- The SMM installation parcel contains Forever and Node dependencies. (CSP-310)
- SMM supports Cloudera Runtime and Cloudera Manager 7.x. (CSP-392)
- You can configure SMM Ranger authorization using Cloudera Manager. (CSP-393)
- SMM keystore and truststore passwords have been removed from configuration files. (CSP-399)
- The Schema Registry API URL is automatically configured for SMM. (CSP-650)
- Nickname fields have been added to API Operations annotations in NotificationsQueryResource. (CSP-672)

What's new in Schema Registry

This topic lists new features for Schema Registry in this release of Cloudera Data Flow for Data Hub.

- Schema Registry supports Avro 1.9. (CSP-132)
- Schema Registry supports auto-TLS. (CSP-264)
- Schema Registry introduces support for deleting Schema. (CSP-379)
- Schema Registry supports Oracle database connections support TLS-MA. (CSP-380)
- Schema Registry supports Kafka in CDP Data Hub. (CSP-395)
- The Schema Registry client allows JAAS dynamic configuration. (CSP-589)

Component support

Cloudera Data Flow in Data Hub includes the following components.

- Schema Registry 0.8.1
- Apache Kafka 2.3.0
- Streams Messaging Manager 2.0.1

Known issues

A summary of known issues for this release.

Topics created with the kafka-topics tool are only accessible by the user who created them when the deprecated --zookeeper option is used

By default all created topics are secured. However, when topic creation and deletion is done with the kafka-topics tool using the --zookeeper option, the tool talks directly to Zookeeper. Because security is the responsibility of ZooKeeper authorization and authentication, Kafka cannot prevent users from making ZooKeeper changes. As a result, if the --zookeeper option is used, only the user who created the topic will be able to carry out administrative actions on it. In this scenario Kafka will not have permissions to perform tasks on topics created this way.

Use kafka-topics with the --bootstrap-server option that does not require direct access to Zookeeper.

Certain Kafka command line tools require direct access to Zookeeper

The following command line tools talk directly to ZooKeeper and therefore are not secured via Kafka:

- kafka-configs
- kafka-reassign-partitions

None.

The offsets.topic.replication.factor property must be less than or equal to the number of live brokers

The offsets.topic.replication.factor broker configuration is now enforced upon auto topic creation. Internal auto topic creation will fail with a GROUP_COORDINATOR_NOT_AVAILABLE error until the cluster size meets this replication factor requirement.

None.

Requests fail when sending to a nonexistent topic with auto.create.topics.enable set to true

The first few produce requests fail when sending to a nonexistent topic with auto.create.topics.e nable set to true.

Increase the number of retries in the producer configuration setting retries.

Custom Kerberos principal names cannot be used for kerberized ZooKeeper and Kafka instances

When using ZooKeeper authentication and a custom Kerberos principal, Kerberos-enabled Kafka does not start. You must disable ZooKeeper authentication for Kafka or use the default Kerberos principals for ZooKeeper and Kafka.

None.

Performance degradation when SSL Is enabled

In some configuration scenarios, significant performance degradation can occur when SSL is enabled. The impact varies depending on your CPU, JVM version, Kafka configuration, and message size. Consumers are typically more affected than producers.

Configure brokers and clients with ssl.secure.random.implementation =SHA1PRNG. It often reduces this degradation drastically, but its effect is CPU and JVM dependent.

Apache JIRA: KAFKA-2561

OPSAPS-43236: Kafka garbage collection logs are written to the process directory

By default Kafka garbage collection logs are written to the agent process directory. Changing the default path for these log files is currently unsupported.

None.

Fixed issues

Summary of fixed issues for this release.

Fixed issues in Apache Kafka

This section lists the issues that have been fixed since the previous release.

December 19, 2019

There are no fixed issues since the Apache Kafka 2.3.0 release.

Fixed issues in Streams Messaging Manager

This section lists the issues fixed since the previous release

December 19, 2019

CSP-303: Hide replication graphs in the Topics Profile page when SRM is disabled.

CSP-548: SMM with Ranger should work without Knox.

CSP-502: SMM Server stops after it has been started my CM.

CSP-457: When SMM is configured with Kerberos, it also requires TLS.

CSP-361: Consumer Group Lag is inactive after it has been disconnected for more than 24 hours.

CSP-622: SMM does not start because it is missing the hdfs-site.xml file.

CSP-674: The SMM UI icon uses a direct link to Schema Registry in CDP.

CSP-708: SMM does not start with a minimum 1G maximum heap size.

CSP-735: SMM explorer experiences the following error when it is used with Schema Registry: Failed to fetch schema versions for topic.

CSP-702: Realign checksum for SQL migration scripts before spinning up SMM or Schema Registry.

CSP-758: SMM does not inherit Kafka Ranger service.

CSP-499: SMM does not start with SSL.

CSP-580: SMM test cases are failing after upgrading to Kafka version 2.3.

CSP-738: Update Kafka version for SMM.

CSP-513: SMM Alerts module throws error when using Open JDK.

CSP-366: SMM logs show CM authentication multiple times.

CSP-530: OOM in SMM log, Failed to fetch on SMM UI.

CSP-706: SMM uses RangerKafkaAuthorizer when Ranger is not installed.

CSP-490: Update Kafka dependency version in SMM.

CSP-640: Expose Oracle TLS configurations for Schema Registry and SMM in CM.

CSP-541: NPE thrown when CM is on HTTPS, but SMM is configured to the HTTP.

CSP-695: SMM service does not start (conflicting slf4j bindings).

CSP-753: SMM-UI should send request to fetch the avro schema based on user Input.

CSP-558: ConsumerGroupFetcherTask failed to close when shutting down SMM.

CSP-760: Remove kafka-clients dependency from notifier-api.

CSP-792: SMM Monitoring interceptor cannot create metrics topics with SSL enabled.

Fixed issues in Schema Registry

This section lists the issues that have been fixed since the previous version. **CSP-559: Schema Registry UI occasionally hangs on first use.**

CSP-591: Schema Registry Client always requires a key store.

CSP-640: Expose Oracle TLS configurations for Schema Registry and SMM in CM.

CSP-694: Schema Registry UI API URL is incorrect.

CSP-702: Realign checksum for SQL migration scripts before spinning up SMM or Schema Registry.

CSP-763: SchemaRegistryClient class does not support Knox proxy or the SSL cert file for HTTPS.

CSP-768: SchemaRegistryClient class should support Knox proxy.

CSP-771: The Hikari version has been updated.

CSP-773: Schema Registry REST API is not accessible through Knox proxy with basic authorization.

CSP-774: Error when creating a schema in Schema Registry with field that is nullable but with default value.

Unsupported features

Some features exist within this release of Cloudera Data Flow for Data Hub components, but are not supported by Cloudera.

The following Kafka features are not supported in Cloudera Data Platform:

- Only Java based clients are supported. Clients developed with C, C++, Python, .NET and other languages are currently not supported.
- Kafka Connect is not supported. NiFi and Sqoop are proven solutions for batch and real time data loading that complement Kafka's message broker capability. For more information, see Cloudera Flow Management.
- The Kafka default authorizer is not supported. This includes setting ACLs and all related APIs, broker functionality, and command-line tools.