User quotas in Admission Control (Preview)

Date published: 2025-04-23 Date modified: 2025-04-23

Legal Notice

© Cloudera Inc. 2025. 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

Legal Notice	2
Contents	3
User quotas in Admission Control	3
User quota configuration elements	4
Configuring user quotas in Admission Control	5
Example configuration rules	6
Limitation	7

User quotas in Admission Control

User quotas in Impala's Admission Control set per-user query limits to ensure fair resource distribution and prevent system overload.

This feature is Technical Preview except for workload aware autoscaling virtual warehouses where it is not supported.

Starting with Cloudera Data Warehouse Runtime 2025.0.19.0, user quotas introduce rules to restrict the number of queries a users/groups can run concurrently. These rules apply at both the pool and root levels and can be based on individual usernames, wildcard users, or user groups. Queries are counted against limits starting with admission control acceptance and continuing until they are released.

When a query exceeds the defined quota, it is rejected at submission time.

Note: The query counts are synchronized across coordinators through the Statestore, which may lead to over-admission.

User quota configuration elements

List of XML elements used to configure user quotas in Impala's Admission Control.

Element	Description
userQueryLimit	used to define a User or Wildcard Rule
groupQueryLimit	used to define a Group Rule
totalCount	used to define the number of queries that can run concurrently.
user	used to specify a username to define a User Rule, or, by using the wildcard '*', to define a Wildcard Rule.
group	in a Group rule, used to specify a group name that the rule applies to.

Rule precedence:

- User rules override group and wildcard rules.
- Group rules override wildcard rules.
- Pool-level rules are evaluated first; if passed, root-level rules are checked.
- If a user belongs to multiple groups, the least restrictive rule applies.

Configuring user quotas in Admission Control

Configure user quotas in Impala's Admission Control to limit concurrent queries for individual users or groups.

Before the task:

Ensure you have access to the fair-scheduler.xml file and necessary administrative privileges in Impala.

Steps

- Log in to the Cloudera web interface and navigate to the Cloudera Data Warehouse service.
- 2. In the Cloudera Data Warehouse service, click Virtual Warehouses in the left navigation panel.
- 3. Select the Impala Virtual Warehouse, click options for the warehouse
- 4. Click Edit and navigate to Impala Coordinator under the Configurations tab.
- 5. Select the fair-scheduler.xml under Configuration files.
- 6. Add or update <userQueryLimit> and groupQueryLimit> elements to define the
 quota rules. Example:

- 7. Save the file.
- 8. Restart the Impala coordinator to apply the changes.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Example configuration rules

Example rule definitions for user, group, and wildcard quotas.

The examples below are incomplete XML snippets used within Admission Control configuration files. They omit required elements such as <aclsubmitApps>

Example:

```
Unset
<queue name="group-set-small">
 <!-- Note: for brevity's sake, this example intentionality excludes other elements such
as weight, schedulingPolicy, and aclSubmitApps -->
 <!-- Any user can run 1 query in the small pool -->
 <userQueryLimit>
   <user>*</user>
   <totalCount>1</totalCount>
 </userQueryLimit>
 <!-- Members of the group 'it' can run 2 queries in the small pool -->
 <groupQueryLimit>
   <group>it</group>
   <totalCount>2</totalCount>
 </groupQueryLimit>
 <!-- The user 'fiona' can run 3 queries in the small pool -->
 <userQueryLimit>
   <user>fiona</user>
   <totalCount>3</totalCount>
 </userQueryLimit>
</queue>
```

In this example:

- Any user can run 1 query (* wildcard rule).
- Users in the it group can run 2 gueries.
- The user fiona can run 3 queries, overriding the it group rule.

User quotas help administrators ensure resource fairness and prevent system overloads by restricting query concurrency at both user and group levels.

For more configuration examples, see the <u>Impala documentation on user quotas</u>

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Limitation

This feature is not supported for virtual warehouses that use workload-aware autoscaling.

When an Impala virtual warehouse has been auto-suspended, you can submit as many queries as desired. All these queries will be queued with the message:

Query queued. Latest queuing reason: Waiting for executors to start. Only DDL queries and queries scheduled only on the coordinator (either NUM_NODES set to 1 or when small query optimization is triggered) can currently run.

Once the virtual warehouse resumes, all queued queries will be admitted and executed simultaneously. The limit defined in the

llama.am.throttling.maximum.queued.reservations setting in llama-site.xml still applies. These limitations are due to a known issue tracked at IMPALA-13965. There is currently no workaround for these issues.

Query counts are synchronized across the cluster by using the statestore. Because the statestore uses an eventually consistent model, the system may admit more queries than the configured limit when multiple queries start at the same time.