Service Level Authorization Guide

Table of contents

1. Purpose

This document describes how to configure and manage *Service Level Authorization* for Hadoop.

2. Prerequisites

Make sure Hadoop is installed, configured and setup correctly. For more information see:

- Single Node Setup for first-time users.
- Cluster Setup for large, distributed clusters.

3. Overview

Service Level Authorization is the initial authorization mechanism to ensure clients connecting to a particular Hadoop *service* have the necessary, pre-configured, permissions and are authorized to access the given service. For example, a MapReduce cluster can use this mechanism to allow a configured list of users/groups to submit jobs.

The `${HADOOP_CONF_DIR}/hadoop-policy.xml` configuration file is used to define the access control lists for various Hadoop services.

Service Level Authorization is performed much before to other access control checks such as file-permission checks, access control on job queues etc.

4. Configuration

This section describes how to configure service-level authorization via the configuration file `{HADOOP_CONF_DIR}/hadoop-policy.xml`.

4.1. Enable Service Level Authorization

By default, service-level authorization is disabled for Hadoop. To enable it set the configuration property `hadoop.security.authorization` to **true** in `${HADOOP_CONF_DIR}/core-site.xml`.

4.2. Hadoop Services and Configuration Properties

This section lists the various Hadoop services and their configuration knobs:

Property	Service
`security.client.protocol.acl`	ACL for ClientProtocol, which is used by user

	code via the DistributedFileSystem.
`security.client.datanode.protocol.acl`	ACL for ClientDatanodeProtocol, the client-to-datanode protocol for block recovery.
`security.datanode.protocol.acl`	ACL for DatanodeProtocol, which is used by datanodes to communicate with the namenode.
`security.inter.datanode.protocol.acl`	ACL for InterDatanodeProtocol, the inter-datanode protocol for updating generation timestamp.
`security.namenode.protocol.acl`	ACL for NamenodeProtocol, the protocol used by the secondary namenode to communicate with the namenode.
`security.inter.tracker.protocol.acl`	ACL for InterTrackerProtocol, used by the tasktrackers to communicate with the jobtracker.
`security.job.submission.protocol.acl`	ACL for JobSubmissionProtocol, used by job clients to communciate with the jobtracker for job submission, querying job status etc.
`security.task.umbilical.protocol.acl`	ACL for TaskUmbilicalProtocol, used by the map and reduce tasks to communicate with the parent tasktracker.
`security.refresh.policy.protocol.acl`	ACL for RefreshAuthorizationPolicyProtocol, used by the dfsadmin and mradmin commands to refresh the security policy in-effect.

4.3. Access Control Lists

`${HADOOP_CONF_DIR}/hadoop-policy.xml` defines an access control list for each Hadoop service. Every access control list has a simple format:

The list of users and groups are both comma separated list of names. The two lists are separated by a space.

Example: `user1,user2 group1,group2`.

Add a blank at the beginning of the line if only a list of groups is to be provided, equivalently a comman-separated list of users followed by a space or nothing implies only a set of given users.

A special value of ***** implies that all users are allowed to access the service.

4.4. Refreshing Service Level Authorization Configuration

The service-level authorization configuration for the NameNode and JobTracker can be changed without restarting either of the Hadoop master daemons. The cluster administrator can change `${HADOOP_CONF_DIR}/hadoop-policy.xml` on the master nodes and instruct the NameNode and JobTracker to reload their respective configurations via the *-refreshServiceAcl* switch to *dfsadmin* and *mradmin* commands respectively.

Refresh the service-level authorization configuration for the NameNode:

```
$ bin/hadoop dfsadmin -refreshServiceAcl
```

Refresh the service-level authorization configuration for the JobTracker:

```
$ bin/hadoop mradmin -refreshServiceAcl
```

Of course, one can use the `security.refresh.policy.protocol.acl` property in `${HADOOP_CONF_DIR}/hadoop-policy.xml` to restrict access to the ability to refresh the service-level authorization configuration to certain users/groups.

4.5. Examples

Allow only users `alice`, `bob` and users in the `mapreduce` group to submit jobs to the MapReduce cluster:

```
<property>
     <name>security.job.submission.protocol.acl</name>
     <value>alice,bob mapreduce</value>
</property>
```

Allow only DataNodes running as the users who belong to the group `datanodes` to communicate with the NameNode:

```
<property>
     <name>security.datanode.protocol.acl</name>
     <value>datanodes</value>
</property>
```

Allow any user to talk to the HDFS cluster as a DFSClient:

```
<property>
     <name>security.client.protocol.acl</name>
     <value>*</value>
</property>
```