Storage Based Authorization in the Metastore Server

The metastore server security feature with storage based authorization was added to Hive in release 0.10. This feature was introduced previously in HCatalog.

For additional information about storage based authorization in the metastore server, see the HCatalog document Storage Based Authorization.

For an overview of Hive authorization models and other security options, see the Authorization document.

The Need for Metastore Server Security

When multiple clients access the same metastore in a backing database, such as MySQL, the database connection credentials may be visible in the hive-site.xml configuration file. A malicious or incompetent user could cause serious damage to metadata even though the underlying data is protected by HDFS access controls.

Also, when a Hive metastore server uses Thrift to communicate with clients and has a backing database for metadata storage and persistence, the authentication and authorization done on the client side cannot guarantee security on the metastore side. To provide security for metadata, Hive release 0.10 added authorization capability to the metastore. (See HIVE-3705.)

Storage Based Authorization

When metastore server security is configured to use Storage Based Authorization, it uses the file system permissions for folders corresponding to the different metadata objects as the source of truth for the authorization policy. Use of Storage Based Authorization in metastore is recommended. See details in the HCatalog Storage Based Authorization document.

Starting in Hive 0.14, storage based authorization authorizes read privilege on database and tables. The get_database api call needs database directory read privilege. The get_table_* calls that fetch table information and get_partition_* calls to list the partitions of a table require read privilege on the table directory. It is enabled by default with storage based authorization. See hive.security.metastore.authorization.auth.reads in the next section on configuration.

Configuration Parameters for Metastore Security

To enable Hive metastore server security, set these parameters in hive-site.xml:

- hive.metastore.pre.event.listeners
  This turns on metastore-side security.

- hive.security.metastore.authorization.manager
  This tells Hive which metastore-side authorization provider to use. The default setting uses DefaultHiveMetastoreAuthorizationProvider, which implements the standard Hive grant/revoke model. To use an HDFS permission-based model (recommended) to do your authorization, use StorageBasedAuthorizationProvider as instructed above.

Versions 0.10.0 and 0.12.0

The StorageBasedAuthorizationProvider was introduced in Hive 0.10.0, running on the metastore side only (HIVE-3705). Starting in Hive 0.12.0 it also runs on the client side (HIVE-5048 and HIVE-5402).
Set to org.apache.hadoop.hive.ql.security.HadoopDefaultMetastoreAuthenticator.

- hive.security.metastore.authorization.auth.reads
  When this is set to true, Hive metastore authorization also checks for read access. It is set to true by default. Read authorization checks were introduced in Hive 0.14.0.

Sample hive-site.xml: Default Settings

The snippet below shows the keys as they are in a default state in hive-site.xml (metastore-side security set up to use the default authorization /authentication, but disabled). Please edit in information as above to get the desired authorization behaviour:

```xml
<property>
    <name>hive.security.metastore.authorization.manager</name>
    <value>org.apache.hadoop.hive.ql.security.authorization.DefaultHiveMetastoreAuthorizationProvider</value>
    <description>authorization manager class name to be used in the metastore for authorization. The user defined authorization class should implement interface org.apache.hadoop.hive.ql.security.authorization.HiveMetastoreAuthorizationProvider.</description>
</property>

<property>
    <name>hive.security.metastore.authenticator.manager</name>
    <value>org.apache.hadoop.hive.ql.security.HadoopDefaultMetastoreAuthenticator</value>
    <description>authenticator manager class name to be used in the metastore for authentication. The user defined authenticator should implement interface org.apache.hadoop.hive.ql.security.HiveAuthenticationProvider.</description>
</property>

<property>
    <name>hive.metastore.pre.event.listeners</name>
    <value>org.apache.hadoop.hive.ql.security.authorization.AuthorizationPreEventListener</value>
    <description>pre-event listener classes to be loaded on the metastore side to run code whenever databases, tables, and partitions are created, altered, or dropped. Set to org.apache.hadoop.hive.ql.security.authorization.AuthorizationPreEventListener if metastore-side authorization is desired.</description>
</property>
```