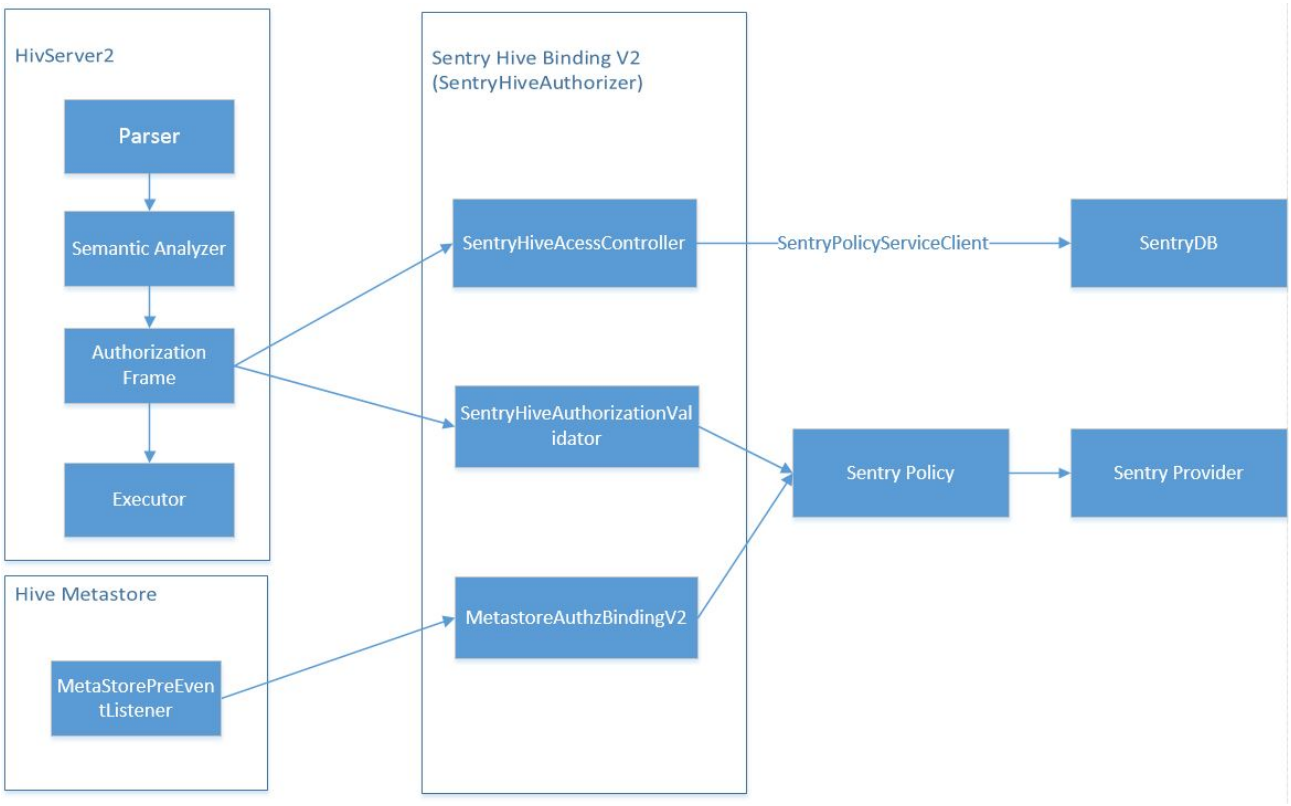
Sentry integration with Hive Authorization V2

Introduction of Sentry-Hive binding with V2

Currently Hive-Sentry binding with v1 grant/revoke privileges via hook DDLTask, and do authorization via HiveSemanticAnalyzerHook. Now hive has a pluggable authorization framework via exposing some interfaces HiveAccessController and HiveAuthorizationValidator. HiveAccessController is used to grant/revoke roles and privileges. HiveAuthorizationValidator is used to do fine-grained authorization.We add a new authorization V2 via implement Hive authorization framework.

Architecture diagram of Sentry-Hive binding with V2



Related configuration

Properties required on Hive to talk to Sentry policy store service (hive-site.xml):

Configuration Key	Configuration Value
hive.security.authorization.task.factory	org.apache.sentry.binding.hive.v2.SentryHiveAuthorizationTaskFactoryImplV2
hive.server2.session.hook	org.apache.sentry.binding.hive.v2.HiveAuthzBindingSessionHookV2
hive.server2.enable.doAs	false
hive.security.authorization.enabled	true
hive.security.authorization.manager	org.apache.sentry.binding.hive.v2.SentryAuthorizerFactory
hive.security.authenticator.manager	org.apache.hadoop.hive.ql.security.SessionStateUserAuthenticator

Properties required on Metastore to talk to Sentry policy store service: (hive-site.xml)

Configuration Key	Configuration Value
hive.metastore.rawstore.impl	org.apache.sentry.binding.hive.v2.metastore.AuthorizingObjectStoreV2
hive.metastore.pre.event.listeners	org.apache.sentry.binding.hive.v2.metastore.MetastoreAuthzBindingV2

Support Hive version

While we have some fixes at hive side, they are HIVE-11780, HIVE-11498, HIVE-11190, HIVE-11179. Especially HIVE-11179 which blocks the authorization of URI type privilege in Sentry. Hive 1.3.0 and Hive 2.0.0 can be adapted in theory, since 1.3.0 is not released, our E2E test works for Hive 2.0.0.

Testing

All the unit tests are passed after fixed some test failures which are caused by the difference output format of some certain commands between Hive 1.1.0 and Hive 2.0.0, such as SHOW INDEX ON [table] in “TestMetadataObjectRetrieval.java”, SHOW GRANT [type] [object], column type privilege are put into “[]”.

We can run independent unit test by

```
mvn test -Dtest=className#methodName \
    -DfailIfNoTests=false \
    -P-hive-authz1,hive-authz2,-datanucleus3,datanucleus4
```

To keep up with the version of datanucleus in Hive 2.0.0, we adopt 4.0.1 version in V2. Considering the version conflict, we remove “hive-authz1” and “datanucleus3” profile. We have done e2e tests in development environment and v2 nightly build. Due to lack of real cluster environment, we hope other committers could help verify it on real cluster.

Continuous integration

The nightly build in Apache Jenkins is

https://builds.apache.org/job/Sentry-jdk-1.7-v2

The pre-commit build in Apache Jenkins is ("Hive V2" component should be added at jira)

https://builds.apache.org/job/PreCommit-SENTRY-Build/

Plan to deprecate V1

Currently Sentry users are mainly using Hive 1.1.0, once most users have upgrade their Hive version to Hive 2.0.0, we will deprecate the v1 binding.

Moving V1 to V2

When we move V1 to V2, the users only need to update the configuration of hive-site.xml according to the “Related configuration” section.