Apache Zeppelin 3

Configuring Apache Zeppelin

Date of Publish: 2018-04-01

http://docs.hortonworks.com

Contents

Configure Livy on an Ambari-Managed Cluster

About this task

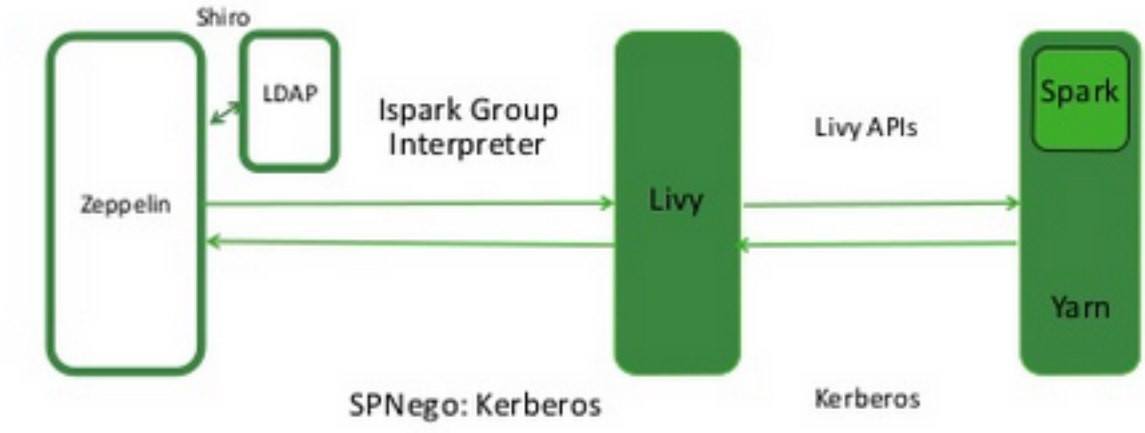
This section describes how to configure Livy on an Ambari-managed cluster.

Livy is a proxy service for Apache Spark; it offers the following capabilities:

- Zeppelin users can launch a Spark session on a cluster, submit code, and retrieve job results, all over a secure connection.
- When Zeppelin runs with authentication enabled, Livy propagates user information when a session is created. Livy user impersonation offers an extended multi-tenant experience, allowing users to share RDDs and cluster resources. Multiple users can access their own private data and session, and collaborate on a notebook.

Note: Livy supports Kerberos, but does not require it.

The following graphic shows process communication among Zeppelin, Livy, and Spark:



On an Ambari-managed cluster, Livy is installed with Spark.

The following sections describe several optional configuration steps.

Kerberos-enabled clusters

Ensure that access is enabled only for groups and hosts where Livy runs.

Check the Livy host URL

1. Navigate to the Interpreter configuration page in the Zeppelin Web UI.
2. In the livy interpreter section, make sure that the zeppelin.livy.url property contains the full Livy host name – replace localhost if necessary.
3. Scroll down and click Save.

Note:

On an Ambari-managed cluster you can find the Livy host from the Ambari dashboard by selecting Spark2 > Summary > Livy for Spark2 Server.

Configure Livy impersonation

1. On the Ambari dashboard, select Spark2 > Configs.
2. Click Custom livy2-conf.

3. Ensure that livy.superusers is listed – if not, add the property.
4. Set livy.superusers to the user account associated with Zeppelin,zeppelin.livy.principal.

   For example, if zeppelin.livy.principal is zeppelin-sr1@example.com, set livy.superusers to the same account, zeppelin-sr1@example.com.

Configure Livy user access control

You can use the livy.server.access-control.enabled property to configure Livy user access.

When this property is set to false, only the session owner and the superuser can access (both view and modify) a given session. Users cannot access sessions that belong to other users. ACLs are disabled, and any user can send any request to Livy.

When this property is set to true, ACLs are enabled, and the following properties are used to control user access:

- livy.server.access-control.allowed-users – A comma-separated list of users who are allowed to access Livy.
- livy.server.access-control.view-users – A comma-separated list of users with permission to view other users' infomation, such as submitted session state and statement results.
- livy.server.access-control.modify-users – A comma-separated list of users with permission to modify the sessions of other users, such as submitting statements and deleting the session.

Specify a timeout value for Livy sessions.

By default, Livy preserves cluster resources by recycling sessions after one hour of session inactivity. When a Livy session times out, the Livy interpreter must be restarted.

To specify a larger or smaller value using Ambari, select Spark2 > Configs > Advanced livy2-conf, then use thelivy.server.session.timeout property to specify the timeout in milliseconds (the default value is 3600000, or one hour).

Restart the Livy interpreter after changing settings.

If you change any Livy interpreter settings, restart the Livy interpreter. Navigate to the Interpreter configuration page in the Zeppelin Web UI. Locate the Livy interpreter, then click restart.

Verify that the Livy server is running

To verify that the Livy server is running, access the Livy web UI in a browser window. The default port is 8998:

```
http://<livy-hostname>:8998/
```

Configure User Impersonation for Access to Hive

This section describes how to configure Apache Zeppelin user impersonation for Apache Hive.

About this task

User impersonation runs Hive queries under the user ID associated with the Zeppelin session.

Non-Kerberos Cluster

If Kerberos is not enabled on the cluster, complete the following steps:

1. In the Zeppelin UI, navigate to the %jdbc section of the Interpreter page.
2. Click edit, then add a hive.proxy.user.property property and set its value tohive.server2.proxy.user.
3. Click Save, then click restart to restart the JDBC interpreter.

Kerberos-enabled Cluster

If Kerberos is enabled on the cluster, enable user impersonation as follows:

To configure the %jdbc interpreter, complete the following steps:

1. In Hive configuration settings, set hive.server2.enable.doAs to true.
2. In the Zeppelin UI, navigate to the %jdbc section of the Interpreter page.
3. Enable authentication via the Shiro configuration: specify authorization type, keytab, and principal.
   a. Set zeppelin.jdbc.auth.type to KERBEROS.
   b. Set zeppelin.jdbc.principal to the value of the principal.
   c. Set zeppelin.jdbc.keytab.location to the keytab location.
4. Set hive.url to the URL for HiveServer2. (On an Ambari-managed cluster you can find the URL under Hive > HiveServer2 JDBC URL.) Here is the general format:

   ```
   jdbc:hive2://HiveHost:10001/default;principal=hive/
   _HOST@HOST1.COM;hive.server2.proxy.user=testuser
   ```

   The JDBC interpreter adds the user ID as a proxy user, and sends the string to HiveServer2; for example:

   ```
   jdbc:hive2://
   dkhdp253.dk:2181,dkhdp252.dk:2181,dkhdp251.dk:2181/;serviceDiscoveryMode=zooKeeper;zo
   ```

5. Add a hive.proxy.user.property property and set its value tohive.server2.proxy.user.
6. Click Save, then click restart to restart the interpreter.

For information about authenticating Zeppelin users through Active Directory or LDAP, see "Configuring Zeppelin Security" in this guide.

Configure User Impersonation for Access to Phoenix

This section describes how to configure Apache Zeppelin user impersonation for Apache Phoenix.

About this task

User impersonation runs Phoenix queries under the user ID associated with the Zeppelin session.

To enable user impersonation for Phoenix, complete the following steps in the Ambari dashboard:

Procedure

1. In the HBase configuration settings, enable phoenix sql.
2. In Advanced HBase settings, set the following properties:

   ```
   hbase.thrift.support.proxyuser=true
   hbase.regionserver.thrift.http=true
   ```

3. In HDFS configuration settings, set the following properties:

   ```
   hadoop.proxyuser.hbase.groups=*
   hadoop.proxyuser.hbase.hosts=*
   hadoop.proxyuser.zeppelin.groups=*
   hadoop.proxyuser.zeppelin.hosts=*
   ```

4. Make sure that the user has access to HBase. You can verify this from the HBase shell with the user_permissions command.