Hadoop使用LDAP

# 1.HDFS LDAP源码解析

HDFS实现了一个类似POSIX模型的文件和文件夹权限管理模型。每一个文件和文件夹都有一个所有组和一个组，文件或文件夹通过权限区分是所有者还是组成员或其他用户。每一个客户端访问HDFS都有两个表示部分包括用户名和组列表，当客户端进程接入文件或者文件夹时，HDFS必须要做一个权限检查。如果是用户名和文件的所有者可以使用所有者权限，如果组一致，那么可以使用组的权限，否则使用其他权限。

在访问过程中，首先确定用户名，simple模式下使用命令whoami，使用Kerberized操作，使用客户端的Kerberos证书来确定用户。然后通过group mapping服务（组映射）获取用户组列表。组映射的方式通过hadoop.security.group.mapping来确定，默认使用JniBasedUnixGroupsMappingWithFallback。如果没有JNI接口，那么使用ShellBasedUnix-

GroupsMapping，这个组映射通过bash -c group或者net group（windows）来获取用户的用户组列表。如果使用的是LDAP，那么配置LdapGroupsMapping。类图如下：



当前支持三种方式进行User->Groups映射，ShellBasedUnixGroupsMapping：从Unix系统中直接获取Groups。LdapGroupsMapping，从LDAP Server中获取映射关系。

LdapGroupsMapping#doGetGroups的流程如下：

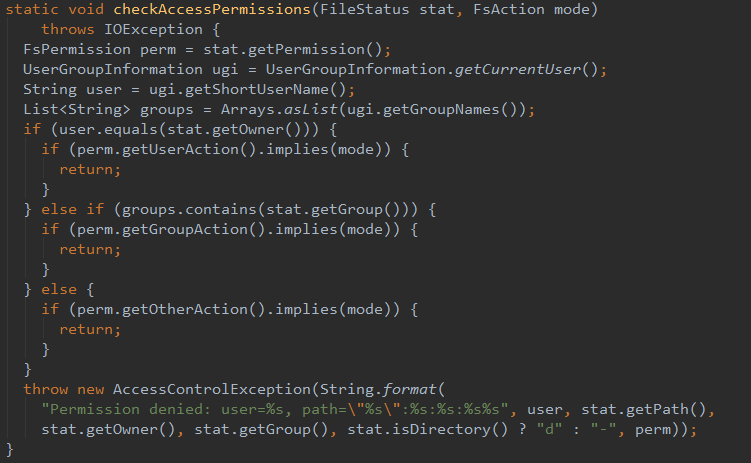
* 1. 根据用户名，从baseDN中查询User，Filter条件：

(&(objectClass=person)(cn ={0}))

* 1. 根据上一步搜索结果，获取User的DN,userDN，
  2. 在userDN下，搜索其对应的用户组，Filter条件：

(&(objectClass=groupOfNames)(member ={0}))

LdapGroupsMapping在Groups中使用，UserGroupInformation中进行用户及用户组的管理，在访问文件系统时，FileSystem首先要进行checkAccessPermission，源码如下：



根据用户名和用户组来确定是否有权限访问该文件。

# 2.HDFS LDAP配置与操作

相关参数及配置:前缀是hadoop.security.group.mapping.ldap

|  |  |  |
| --- | --- | --- |
| 参数名 | 描述 | 默认值 |
| url | LDAP server的Url |  |
| ssl | 连接LDAP是否使用SSL |  |
| bind.user | 绑定LDAP用户 |  |
| bind.passwd | 用户对应的密码 |  |
| base | LDAP搜索Base DN | NULL |
| search.filter.user | 搜索相关Users的过滤器 | &(objectclass=person)  (cn={0}) |
| group | 搜索相关group的过滤器 | objectClass=groupOfNames |
| Search.attr.member | 决定LDAP组成员的attribute | member |
| Search.attr.group.name | 标识group名字的attribute | cn |

配置core-site.xml，如下所示：

<property>

<name>hadoop.security.group.mapping</name>

<value>org.apache.hadoop.security.LdapGroupsMapping</value>

</property>

<property>

<name>hadoop.security.group.mapping.ldap.bind.user</name>

<value>cn=Manager,dc=hadoop,dc=org</value>

</property>

<property>

<name>hadoop.security.group.mapping.ldap.bind.password</name>

<value>123123</value>

</property>

<property>

<name>hadoop.security.group.mapping.ldap.url</name>

<value>ldap://10.210.91.6:389/dc=hadoop,dc=org</value>

</property>

<property>

<name>hadoop.security.group.mapping.ldap.base</name>

<value></value>

</property>

<property>

<name>hadoop.security.group.mapping.ldap.search.filter.user</name>

<value>(&;(objectclass=person)(cn={0}))</value>

</property>

<property>

<name>hadoop.security.group.mapping.ldap.search.filter.group</name>

<value>(objectclass=groupOfNames)</value>

</property>

<property>

<name>hadoop.security.group.mapping.ldap.search.attr.member</name>

<value>member</value>

</property>

<property>

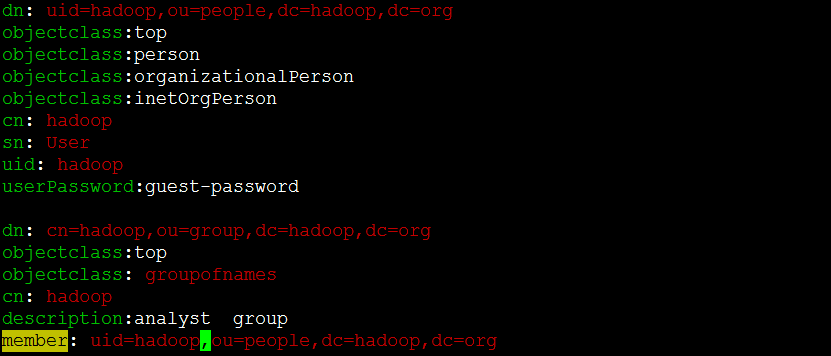
<name>hadoop.security.group.mapping.ldap.search.attr.group.name</name>

<value>cn</value>

</property>

# 3.用户及用户组的添加

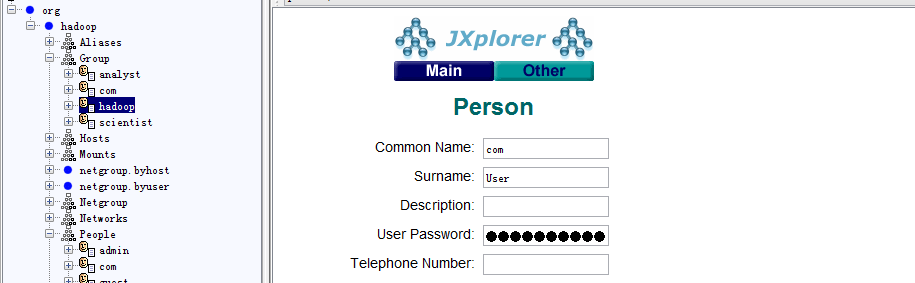
将用户及用户组写入LDAP，ldif格式如下：



文件为Users.ldif，上面文件为添加的hadoop:hadoop（用户及用户组），执行下面的命令：

$ldapadd -x -D "cn=Manager,dc=hadoop,dc=org" -W -f Users.ldif

通过JXplorer(OpenLDAP UI)，可以查看已添加的用户和用户组，如下图所示：



通过下面的命令查看Hadoop正确使用LDAP：

$hdfs groups

该命令获取当前用户的用户组，结果如下：



获取当前用户com的用户组为com。

Reference: http://zh.hortonworks.com/blog/hadoop-groupmapping-ldap-integration/