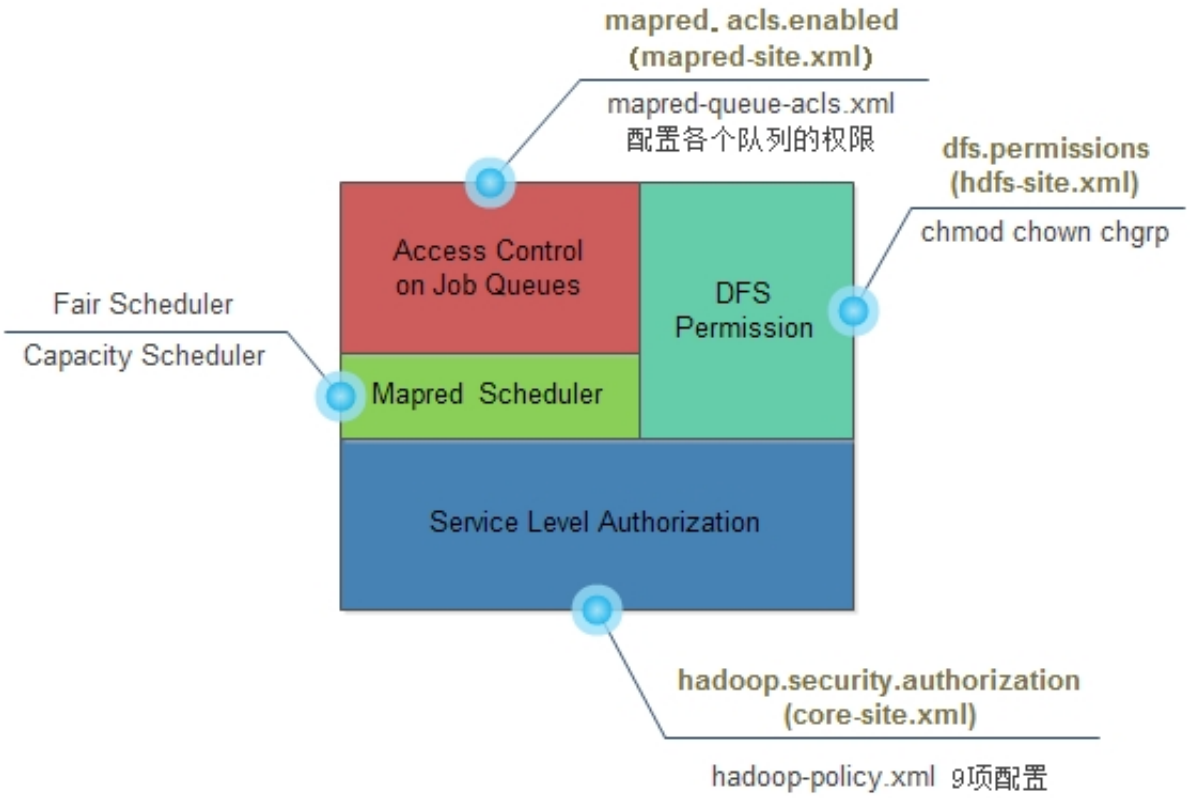
Hadoop服务层授权控制

Hadoop在服务层进行了授权（Service Level Authorization）控制，这是一种机制可以保证客户和Hadoop特定的服务进行链接，比如说我们可以控制哪个用户/哪些组可以提交Mapreduce任务。所有的这些配置可以在$HADOOP_CONF_DIR/hadoop-policy.xml中进行配置。它是最基础的访问控制，优先于文件权限和mapred队列权限验证。可以看看下图



Service Level Authorization

在默认情况下，service-level authorization功能是不启用的，如果你要开启service-level authorization，需要在$HADOOP_HOME/etc/hadoop/core-site.xml进行如下配置

```
<property>
  <name>hadoop.security.authorization</name>
  <value>true</value>
  <description>Is service-level authorization enabled?</description>
</property>
```

设置好后，就开启了Hadoop的service-level authorization，这个生效需要重新启动NameNode节点（其他的节点不需要重启）

在Hadoop 2.x版本中Service LevelAuthorization有10个可配置的属性，每个属性可指定拥有相应访问权限的用户或者用户组。每个属性描述如下：

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.
security.ha.service.protocol.acl	ACL for HAService protocol used by HAAdmin to manage the active and stand-by states of namenode.

$HADOOP_CONF_DIR/hadoop-policy.xml中的上述10个属性分别是用来控制不同服务的访问权限的。每个属性的值需要遵守一定的格式：每个可配置多个用户，用户之间用“,”分割；同时也可配置多个用户组，分组之间用“,”分割，用户和分组之间用空格分割，比如：Example: user1,user2 group1,group2。如果只有分组，需要在分组前面保留一个空格。如果这个值是*，这说明所有的用户都可以访问相应的服务。

下面举几个例子来说明：

（1）、如果只允许某个用户或者某个组内的所有用户提交Mapreduce任务，可以配置如下：

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

（2）、如果只运行属于某个组内的用户运行DataNode来和NameNode进行通信，可以使用下面配置实现

```
<property>
```

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

（3）、运行所有的用户访问集群上的HDFS，可以使用下面配置实现

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

对关于访问NameNode或者JobTracker的service-level authorization修改，不需要重启相应的守护进程，但是集群管理员需要通过下面的命令来加载已经修改的配置：如果修改了有关NameNode的服务配置，可以用下面的命令来动态加载

```
$ bin/hdfs dfsadmin -refreshServiceAcl
```

如果修改了有关JobTracker的服务配置，可以用下面的命令来动态加载

```
$ bin/hdfs mradmin -refreshServiceAcl
```

翻译自：《Service Level Authorization Guide》