

Hive几种参数配置方法

写在前面的话

，学Hive这么久了，发现目前国内还没有一本完整的介绍Hive的书籍，而且互联网上面的资料很乱，于是我决定写一些关于《[Hive的那些事](#)》序列文章，分享给大家。我会在接下来的时间整理有关Hive的资料，如果对Hive的东西感兴趣，请关注本博客。<https://www.iteblog.com/archives/tag/hive-technology/>

Hive提供三种可以改变环境变量的方法，分别是：（1）、修改`#{HIVE_HOME}/conf/hive-site.xml`配置文件；（2）、命令行参数；（3）、在已经进入cli时进行参数声明。下面分别来介绍这几种设定。

方法一：

在Hive中，所有的默认配置都在`#{HIVE_HOME}/conf/hive-default.xml`文件中，如果需要对默认的配置进行修改，可以创建一个`hive-site.xml`文件，放在`#{HIVE_HOME}/conf`目录下。里面可以对一些配置进行个性化设定。在`hive-site.xml`的格式如下：

```
<configuration>
  <property>
    <name>hive.metastore.warehouse.dir</name>
    <value>/user/hive/warehouse</value>
    <description>location of
      default database for the warehouse</description>
  </property>
</configuration>
```

所有的配置都是放在`<configuration></configuration>`标签之间，一个`configuration`标签里面可以存在多个`<property></property>`标签。`<name>`标签里面就是我们想要设定属性的名称；`<value>`标签里面是我们想要设定的值；`<description>`标签是描述在这个属性的，可以不写。绝大多数配置都是在xml文件里面配置的，因为在这里做的配置都全局用户都生效，而且是永久的。用户自定义配置会覆盖默认配置。另外，Hive也会读入Hadoop的配置，因为Hive是作为Hadoop的客户端启动的，Hive的配置会覆盖Hadoop的配置。

方法二：

在启动Hive cli的时候进行配置，可以在命令行添加`-hiveconf param=value`来设定参数，例如：

```
[wyp@master ~]$ hive --hiveconf mapreduce.job.queueName=queue1
```

这样在Hive中所有MapReduce作业都提交到队列queue1中。这一设定对本次启动的会话有效，下次启动需要重新配置。

方法三：

在已经进入cli时进行参数声明，可以在HQL中使用SET关键字设定参数，例如：

```
hive> set mapreduce.job.queueName=queue1;
```

这样也能达到方法二的效果。这种配置也是对本次启动的会话有效，下次启动需要重新配置。在HQL中使用SET关键字还可以查看配置的值，如下：

```
hive> set mapreduce.job.queueName;  
mapreduce.job.queueName=queue1
```

我们可以得到mapreduce.job.queueName=queue1。如果set后面什么都不添加，这样可以查到Hive的所有属性配置，如下：

```
hive> set;  
datanucleus.autoCreateSchema=true  
datanucleus.autoStartMechanismMode=checked  
datanucleus.cache.level2=false  
datanucleus.cache.level2.type=none  
datanucleus.connectionPoolingType=DBCP  
datanucleus.identifierFactory=datanucleus  
datanucleus.plugin.pluginRegistryBundleCheck=LOG  
datanucleus.storeManagerType=rdbms  
datanucleus.transactionIsolation=read-committed  
datanucleus.validateColumns=false  
datanucleus.validateConstraints=false  
datanucleus.validateTables=false
```

.....

上述三种设定方式的优先级依次递增。即参数声明覆盖命令行参数，命令行参数覆盖配置文件设定。

注意：

某些系统级的参数，例如log4j相关的设定，必须用前两种方式设定，因为那些参数的读取在会话建立以前已经完成

了。所以在HQL中设定是无效的。这个特列可以参见本博客[《Hive日志调试》](#)进行了解。

本博客文章除特别声明，全部都是原创！

原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。

本文链接: [【】（）](#)