

公告

昵称： 大葱拌豆腐
园龄： 3年8个月
粉丝： 108
关注： 6
[+加关注](#)

<	2019年12月						>
日	一	二	三	四	五	六	
1	2	3	4	5	6	7	
8	9	10	11	12	13	14	
15	16	17	18	19	20	21	
22	23	24	25	26	27	28	
29	30	31	1	2	3	4	
5	6	7	8	9	10	11	

搜索

找找看

谷歌搜索

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

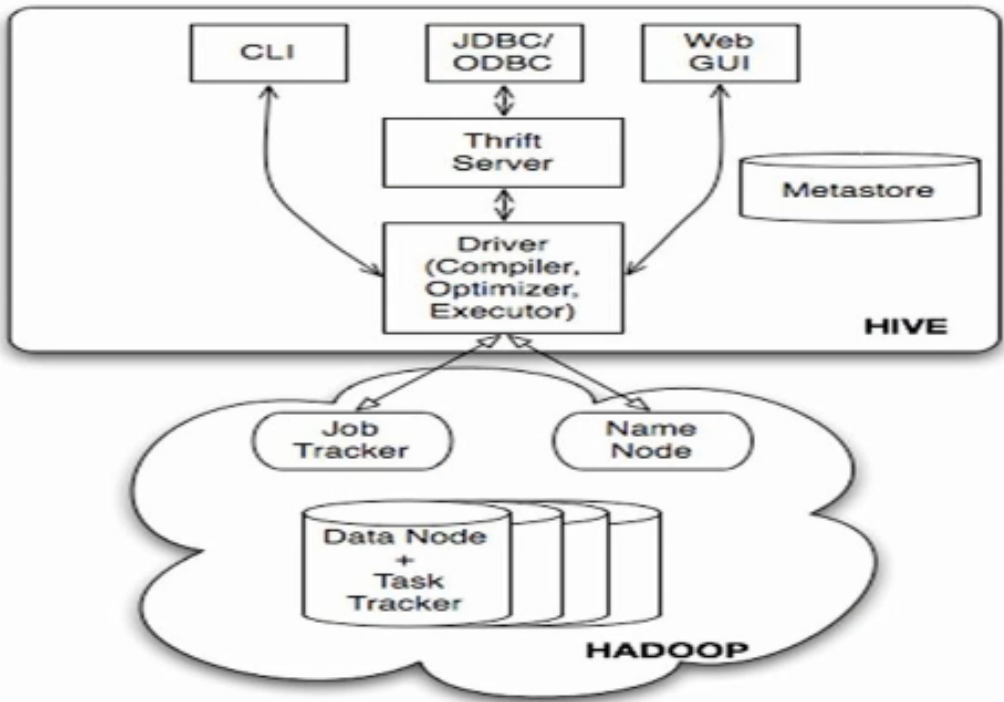
我的标签

随笔 - 409 文章 - 0 评论 - 19

Hive中MetaServer与HiveServer2的应用

在hive中有metaServer与hiveServer2两种服务，看了好多文章说这两个的区别，文章内容有对有错，不够全面，故在这里好好总结一下。

首先，下面这个hive构架图，我们一定不陌生，它反应出hive有哪些组件结构



当然下面的图是hadoop1的部分，现在JobTracker是Yarn了
上面的部分是访问Hive的三个入口，

- 1: 直接Cli
- 2: 通过JDBC

Spark(98)

SparkMllib(50)

Spark SQL(36)

Scala(33)

Hadoop(31)

Spark源码(26)

Java(23)

HBase(15)

MySQL(11)

Problem Log(10)

更多

随笔档案

2019年11月(2)

2019年9月(1)

2019年8月(1)

2019年7月(9)

2019年6月(16)

2019年5月(5)

2019年4月(7)

2019年3月(20)

2019年2月(3)

2019年1月(7)

2018年12月(9)

2018年11月(18)

2018年10月(20)

2018年9月(8)

2018年8月(3)

2018年7月(22)

2018年6月(53)

2018年4月(9)

3: webUI

当我们要连接Hive进行操作时，首先必须是安装了，安装hive很简单，直接在conf/hive-site配置存放Hive元数据的连接信息，通常是用mysql，如下：



```
<property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mysql://localhost:3306/hive?
createDatabaseIfNotExist=true</value>
</property>
<property>
    <name>javax.jdo.option.ConnectionDriverName</name>
    <value>com.mysql.jdbc.Driver</value>
</property>
<property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>root</value>
</property>
<property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>root</value>
</property>
```



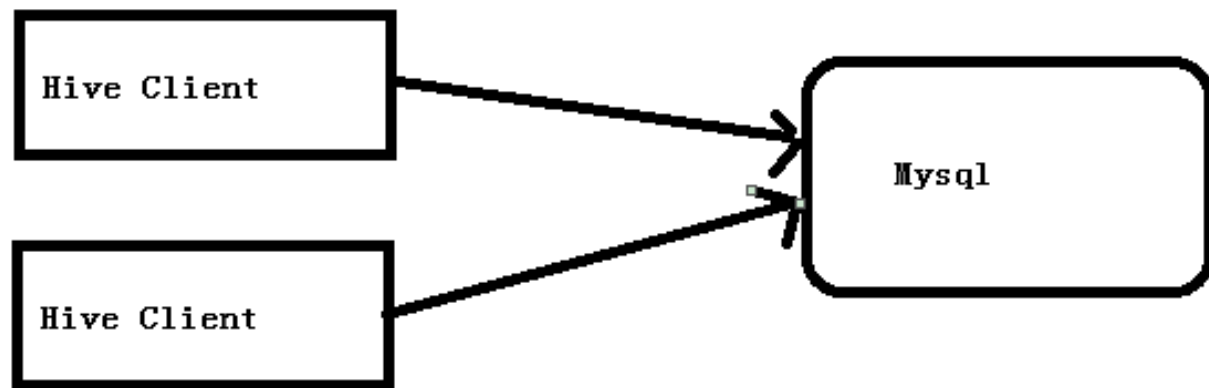
这样配置好后，在hive的lib里加入Mysql JDBC 驱动后，我们就可以通过 bin/hive的方式进行hive客户端，请求数据。如下图：

2018年3月(2)
2018年2月(8)
2018年1月(22)
2017年12月(6)
2017年11月(13)
2017年10月(4)
2017年9月(5)
2017年8月(2)
2017年7月(6)
2017年6月(8)
2017年5月(13)
2017年4月(14)
2017年3月(11)
2017年2月(2)
2017年1月(7)
2016年12月(19)
2016年11月(8)
2016年10月(17)
2016年8月(5)
2016年7月(13)
2016年6月(1)
2016年5月(7)
2016年4月(3)

最新评论

1. Re:Spark源码分析之Sort-Based Shuffle读写流程

sortshuffle时，默认情况下的排序只是 partitionId的排序？而不是每个 partition内部的数据根据key来排序？也



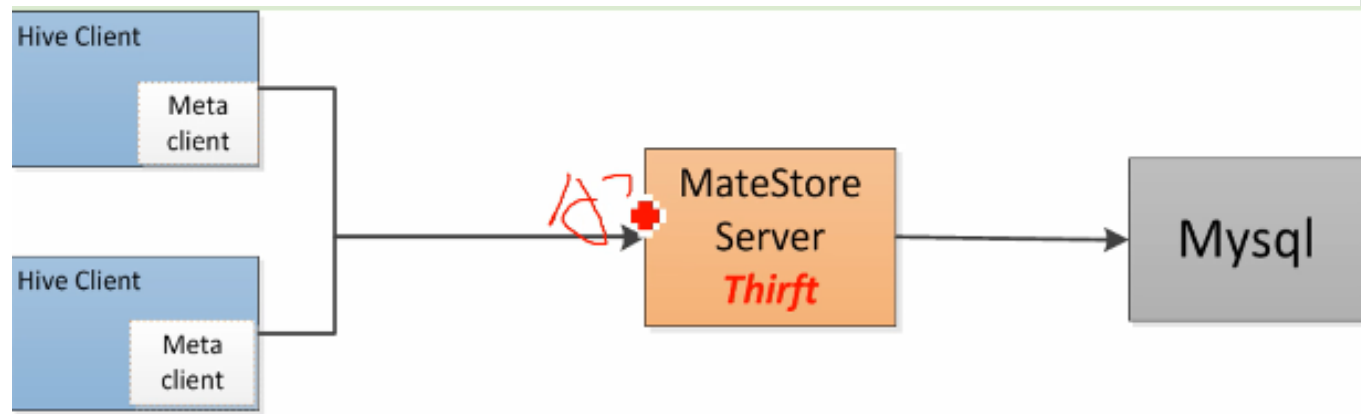
这样在大多数公司，特别是小公司小集群里基本是这样操作的，当然我们公司也是这样的，这样本身没有错误。

但我们看一下上面的配置，这样是不是把连接数据库的信息全泄露了，你可能会说，大家都是同公司的大数据部分，还有各种权限，泄露也无所谓，但我们想一下，在一个大的公司里，大数据平台是几个部门共用的，这样会有多台hive cli连接mysql，这样泄露mysql的信息的风险还是挺大的。

另一方面，比如我们元数据Mysql库无法启动，我们要替换从库时，或者我们要移植元数据到另一台机器上时，那么我们这么多的hive cli是不是全要修改一遍....

那么这时我们Hive的另一个组件出场了，它就是MetaStoreServer

```
启动: nohup $HIVE_HOME/bin/hive -metastore &
```



不是所有数据的全排序？那如果是
partitionId的排序，有什么...
--天地枯荣

2. Re:Apache Spark 2.0三种API的传说：RDD、DataFrame和Dataset
好
--cn_world

3. Re:Spark2.0机器学习系列之3：决策树
请问一下model.toDebugString 打印出来的数据怎么分析啊 看不懂这些数据,大神可以指教一下吗?谢谢
--爱吃橙子的老郭

4. Re:Yarn之ResourceManager详细分析
您这个黑名单和白名单是不是整反了?
--老衲不吃肉

5. Re:Yarn框架和工作流程研究
博主，我看你YARN这部分讲的很棒，有推荐的学习资料嘛？我在官网上没看到这么详细的介绍呢
--SnailDove

阅读排行榜

- 1. 彻底搞懂spark的shuffle过程 (shuffle write) (26741)
- 2. mysql数据库的初始化及相关配置(17751)
- 3. mysql给root开启远程访问权限，修改root密码(17714)

如图，我们在Hive cli与mysql中间是不是启动一个MetaStoreServer，
这样我们的Hive cli就不需要连接Mysql，直接连接这个MetaStoreServer不就行了吗，
在hive-site.xml只要简单的配置一下：

```
<property>
  <name>hive.metastore.uris</name>
  <value>thrift://xxxxxx:9083</value>
  <description>
</property>
```

这样我就通过metaserver取得了元数据的信息对吧
当然上面的图只是一个MetaStoreServer，存在单点问题，但我们完全可以配置两个或者多个MetaStoreServer，就实现了负载均衡与容错的功能了，如下面的配置

```
<property>
  <name>hive.metastore.uris</name>
  <value>thrift://dw1:9083,thrift://dw2:9083</value>
  <description>A comma separated list of metastore uris on which metastore
service is running
</description>
</property>
```

接下来我们就要谈谈另一个有用的组件HiveServer2了

标签: HIVE

好文要顶

关注我

收藏该文

大葱拌豆腐
关注 - 6
粉丝 - 108