

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

[Spark\(98\)](#)
[SparkMllib\(50\)](#)
[Spark SQL\(36\)](#)
[Scala\(33\)](#)
[Hadoop\(31\)](#)
[Spark源码\(26\)](#)
[Java\(23\)](#)
[HBase\(15\)](#)
[MySQL\(11\)](#)

Hive为什么要启用Metastore?

转载来自：

https://blog.csdn.net/qg_40990732/article/details/80914873

<https://blog.csdn.net/tp15868352616/article/details/80891057>

相关概念

1.Metadata概念：

元数据包含用Hive创建的database、table等的元信息。元数据存储的关系型数据库中。如Derby、MySQL等。

2.Metastore作用：

客户端连接metastore服务，metastore再去连接MySQL数据库来存取元数据。有了metastore服务，就可以有多个客户端同时连接，而且这些客户端不需要知道MySQL数据库的用户名和密码，只需要连接metastore 服务即可。

```
bin/hin --访问--> metaStore server --访问-->MySQL
beeline --访问-->hiveServer2 --访问--> metaStore server --访问--> MySQL
```

3.Metastore 有3中开启方式：

--1-->默认开启方式：

没有配置metaStore的时候,每当开启bin/hive;或者开启hiveServer2的时候,都会在内部启动一个metastore
嵌入式服务;资源比较浪费,如果开启多个窗口,就会存在多个metastore server。

--2-->local mataStore(本地)

当metaStore和装载元数据的数据库(MySQL)存在同一机器上时配置是此模式,
开启metastore服务就只需要开启一次就好,避免资源浪费!

--3-->Remote Metastore(远程)

当metaStore和装载元数据的数据库(MySQL)不存在同一机器上时配置是此模式,
开启metastore服务就只需要开启一次就好,避免资源浪费!

Metastore三种配置方式

由于元数据不断地修改、更新，所以Hive元数据不适合存储在HDFS中，一般存在RDBMS中。

1、内嵌模式（Embedded）

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)
- 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)

- hive服务和metastore服务运行在同一个进程中， derby服务也运行在该进程中.内嵌模式使用的是内嵌的Derby数据库来存储元数据，也不需要额外起Metastore服务。
- 这个是默认的，配置简单，但是一次只能一个客户端连接（这句话说实在有点坑，其实就是你启动一个hive服务会内嵌一个metastore服务，然后在启动一个又会内嵌一个metastore服务，并不是说你的客户端只能启动一个hive，是能启动多个，但是每个都有metastore，浪费资源），适用于用来实验，不适用于生产环境。

2、本地模式（Local） :本地安装mysql 替代derby存储元数据

- 不再使用内嵌的Derby作为元数据的存储介质，而是使用其他数据库比如MySQL来存储元数据。hive服务和metastore服务运行在同一个进程中， mysql是单独的进程，可以同一台机器，也可以在远程机器上。（我之前有种方式是：只在接口机配置hive，并配置mysql数据库，用户和密码等；但是集群不配置hive，不起hive任何服务，就属于这种情况）
- 这种方式是一个多用户的模式，运行多个用户client连接到一个数据库中。这种方式一般作为公司内部同时使用Hive。每一个用户必须要有对MySQL的访问权利，即每一个客户端使用者需要知道MySQL的用户名和密码才行。
- 服务端配置如下：



```
<configuration>
<property>
  <name>hive.metastore.warehouse.dir</name>
  <value>/user/hive_remote/warehouse</value>
</property>

<property>
  <name>hive.metastore.local</name>
  <value>true</value>
</property>

<property>
  <name>javax.jdo.option.ConnectionURL</name>
  <value>jdbc:mysql://localhost/hive_remote?createDatabaseIfNotExist=true</value>
</property>
或者 jdbc:mysql://ip:3306/hive?characterEncoding=UTF-8（推荐这个。不会web，不然两个参数都给加上。。。）

<property>
  <name>javax.jdo.option.ConnectionDriverName</name>
  <value>com.mysql.jdbc.Driver</value>
</property>

<property>
  <name>javax.jdo.option.ConnectionUserName</name>
  <value>hive</value>
</property>
```

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 workflows 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) (26742)
- 2. mysql数据库的初始化及相关配置(17752)

```
<property>
  <name>javax.jdo.option.ConnectionPassword</name>
  <value>password</value>
</property>
</configuration>
```

3、远程模式（Remote）：远程安装mysql 替代derby存储元数据

- Hive服务和metastore在不同的进程内，可能是不同的机器（集群实例来说metastore有3个

角色类型	状态	主机
Hive Metastore Server	已启动	[redacted]
Hive Metastore Server	已启动	[redacted]
Hive Metastore Server	已启动	[redacted]

，启动在三台机器，但是都是指向的一台主机的mysql，当然mysql也能配置成主备集群的模式），该模式需要将hive.metastore.local设置为false（但是在0.10，0.11或者之后的HIVE版本hive.metastore.local属性不再使用。），将hive.metastore.uris设置为metastore服务器URL，

- 远程元存储需要单独起metastore服务，然后每个客户端都在配置文件里配置连接到该metastore服务。将metadata作为一个单独的服务进行启动。各种客户端通过beeline来连接，连接之前无需知道数据库的密码。

- 仅连接远程的mysql并不能称之为“远程模式”，是否远程指的是metastore和hive服务是否在同一进程内。

- 服务端配置如下：

```
<configuration>

<property>
  <name>hive.metastore.warehouse.dir</name>
  <value>/user/hive/warehouse</value>
</property>

<property>
  <name>javax.jdo.option.ConnectionURL</name>
  <value>jdbc:mysql://192.168.1.214:3306/hive_remote?createDatabaseIfNotExist=true</value>
</property>

<property>
  <name>javax.jdo.option.ConnectionDriverName</name>
  <value>com.mysql.jdbc.Driver</value>
</property>
```

- 3. mysql给root开启远程访问权限，修改root密码(17714)
- 4. Spark中的Spark Shuffle详解(15182)
- 5. Spark的Java API例子详解(13041)

评论排行榜

- 1. Kylin构建cube时状态一直处于pending(4)
- 2. Spark RDD的依赖解读(2)
- 3. Spark将计算结果写入到Mysql中(2)
- 4. 在Spark中使用Kryo序列化(1)
- 5. spark性能调优：资源优化(1)

推荐排行榜

- 1. Spark中的Spark Shuffle详解(6)
- 2. 彻底搞懂spark的shuffle过程（shuffle write）(4)
- 3. Spark DataFrame vector 类型存储到Hive表(1)
- 4. Hadoop在win7下部署的问题(1)
- 5. 【漫画解读】HDFS存储原理(转载)(1)

```
<property>
  <name>javax.jdo.option.ConnectionUserName</name>
  <value>hive</value>
</property>
```

```
<property>
  <name>javax.jdo.option.ConnectionPassword</name>
  <value>password</value>
</property>
```

```
<property>
  <name>hive.metastore.local</name>
  <value>false</value>
</property>
```

但是在0.10 , 0.11或者之后的HIVE版本 hive.metastore.local 属性不再使用。

```
<property>
  <name>hive.metastore.uris</name>
  <value>thrift://192.168.1.188:9083</value>
</property>
```

```
</configuration>
```



附：远程模式下，hive客户端配置：

(1) hive-site.xml



```
<configuration>
<property>
  <name>hive.metastore.uris</name>
  <value>thrift://hive服务端ip:9083</value>
</property>
</configuration>
```



(2) 配置日志存放位置：配置文件为hive-log4j.properties.template

修改如下：

```
# Define some default values that can be overridden by system properties
hive.log.threshold=ALL
hive.root.logger=WARN,DRFA
hive.log.dir=/opt/beh/log/hive
hive.log.file=hive.log
```

https://blog.csdn.net/qq_35440040

```
hive.log.dir=/opt/beh/log/hive
```

或者修改为:

```
hive.log.dir=/opt/beh/logs/hive/${user.name}
```

这样对于同一台机器不同的用户可以把日志动态写入各自用户目录下,前提是/opt/beh/log/hive权限足够: rwx

重点: hive-log4j.properties.template要修改为hive-log4j.properties不然无法识别。

(3) 服务端metastore 启动方式:

第一种: hive --service metastore -p 9083 &

第二种: 在hive-site.xml中配置hive.metastore.uris, 指定端口, 然后直接 hive --service metastore

(4) 对比测试: (客户端主机2*6C, 128Gmem)

=====我之前是本地模式, 集群不配置hive, 不启动hive metastore, 直接在客户端配置链接数据库, 使用用户名和密码, 然后启动hive执行命令查看如下:

简单的showdatabase 执行时间为: 3.29s, 费时

```
WARNING: hive-cli is deprecated and migration to be
hive> show databases;
OK
default
Time taken: 3.29 seconds, Fetched: 1 row(s)
hive>
> https://blog.csdn.net/qq_35440040
>
```

cpu损耗较大: 178%

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
23173	hadoop	20	0	2405036	391832	28308	S	178.0	0.8	17.69	java -Dhttp0:17b69gjsdn.net/qq_35440040

=====后来配置了远程模式, 在集群配置了hive, 并后台运行了metastore, 然后配置客户端连接metastore后, 测试如下:

简单的showdatabase 执行时间为: 0.8s, 明显快了很多

```
Logging initialized using configuration in file:/c
WARNING: Hive CLI is deprecated and migration to E
hive> show databases;
OK
default
Time taken: 0.806 seconds, Fetched: 1 row(s)
hive> show databases;
OK
default
Time taken: 0.01 seconds, Fetched: 1 row(s)
hive> https://blog.csdn.net/qg\_35440040
```

cpu损耗: 74.8% 少了一倍

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
28269	hadoop	20	0	2425572	363268	28180	S	74.8	0.3	0:08.33	java -Djava.library

总结: 为什么集群要启用metastore, 因为启用后, 直接bin/hive启动进程, 只会有hive交互式客户端一个服务, 不会再有metastore同时存在在该进程中, 资源占用率下降, 查询速度更快。

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

[+加关注](#)

0

0

« 上一篇: [Hive中MetaServer与HiveServer2的应用](#)

» 下一篇: [SparkML之推荐引擎 \(一\) ---电影推荐](#)

posted @ 2019-03-29 15:37 大葱拌豆腐 阅读(236) 评论(0) 编辑 收藏

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论, 请 [登录](#) 或 [注册](#), [访问](#) 网站首页。

- 【推荐】超50万行VC++源码: 大型组态工控、电力仿真CAD与GIS源码库
- 【推荐】腾讯云热门云产品限时秒杀, 爆款1核2G云服务器99元/年!
- 【推荐】阿里云双11返场来袭, 热门产品低至一折等你来抢!
- 【推荐】天翼云双十一翼降到底, 云主机11.11元起, 抽奖送大礼
- 【推荐】流程自动化专家UiBot, 体系化教程成就高薪RPA工程师
- 【活动】京东云服务器_云主机低于1折, 低价高性能产品备战双11
- 【优惠】七牛云采购嘉年华, 云存储、CDN等云产品低至1折

相关博文:

- [Hive安装配置指北 \(含Hive Metastore详解\)](#)
- [Hive 学习笔记 \(启动方式, 内置服务\)](#)
- [hive+mysql安装、hive.metastore.local属性的真正用途](#)
- [Hive的三种安装方式 \(内嵌模式, 本地模式远程模式\)](#)
- [Hive教程之metastore的三种模式](#)
- » [更多推荐...](#)

[斩获阿里offer的必看12篇面试合辑](#)

最新 IT 新闻:

- [贺建奎论文手稿意外曝光! HIV抗性未被验证, 且造成意外突变](#)
- [高通骁龙865仍由台积电代工 采用与苹果A13相同工艺](#)
- [贺建奎“基因编辑婴儿”论文手稿曝光](#)
- [政务装上AI大脑, 平安手里的『王牌』是什么?](#)
- [一个月开工三个超大云计算中心 百度欲将AI进行到底](#)
- » [更多新闻...](#)

历史上的今天:

[2019-03-29 Hive中MetaServer与HiveServer2的应用](#)