

# Hive运维管理

[www.huawei.com](http://www.huawei.com)





# 目标

- 学完本课程后，您将能够：
  - 掌握**Hive**常用维护命令/参数
  - 掌握**Hive**性能调优命令/参数
  - 熟悉**Hive**常见问题及定位



# 目录

- 1、服务检查与日志
- 2、维护命令
- 3、性能调优命令与参数
- 4、常见问题与定位

# 目录介绍

Hive组件目录:

`{${BIGDATA_HOME}}/FusionInsight_V100R002C60/FusionInsight-Hive-1.3.0/hive-1.3.0`

目录/文件	目录/文件说明
<b>bin</b>	<b>Hive</b> 运行脚本目录
<b>RELEASE_NOTES.txt</b>	该版本变更信息
<b>conf</b>	<b>Hive</b> 配置文件目录
<b>examples</b>	<b>Hive web</b> 文件目录
<b>hcatalog</b>	<b>Hcatlog</b> 相关文件目录
<b>lib</b>	<b>Hive</b> 依赖包目录
<b>scripts</b>	<b>Hive</b> 脚本
<b>LICENSE</b>	<b>licence</b>

# 日志介绍

## 日志存储路径：

Hive相关日志的默认存储路径为/var/log/Bigdata/hive/角色名

- **HiveServer**： /var/log/Bigdata/hive/hiveserver（运行日志），  
/var/log/Bigdata/audit/hive/hiveserver（审计日志）
- **MetaStore**： /var/log/Bigdata/hive/metastore（运行日志），  
/var/log/Bigdata/audit/hive/metastore（审计日志）
- **WebHCat**： /var/log/Bigdata/Hive/webhcat（运行日志），  
/var/log/Bigdata/audit/hive/webhcat（审计日志）

## 日志归档规则：

**Hive**的日志启动了自动压缩归档功能，缺省情况下，当日志大小超过**20MB**的时候（此日志文件大小可进行配置，详情请参见“配置日志级别与文件大小”），会自动压缩，压缩后的日志文件名规则为：“<原有日志名>-<yyyy-mm-dd\_hh-mm-ss>.[编号].log.zip”。最多保留最近的**20**个压缩文件，压缩文件保留个数可以在**FusionInsight Manager**界面中配置。

# 日志介绍

日志类型	日志文件名	描述
运行日志	<b>hive-&lt;SSH_USER&gt;-&lt;pid&gt;-gc.log</b>	<b>Hive</b> 服务垃圾回收日志。
	<b>checkServiceDetail.log</b>	<b>Hive</b> 服务启动是否成功的检查日志。
	<b>hive.log</b>	<b>Hive</b> 系统日志，记录 <b>Hive</b> 系统运行时所产生的大部分日志。
审计日志	<b>Hive-audit.log</b>	<b>Hive</b> 安全审计日志。

# 日志格式

日志类型	组件	格式
运行日志	HiveServer	<yyyy-MM-dd HH:mm:ss,SSS> <Log Level> <产生该日志的线程名字> <log中的message> <日志事件的发生位置>
	MetaStore	<yyyy-MM-dd HH:mm:ss,SSS> <Log Level> <产生该日志的线程名字> <log中的message> <日志事件的发生位置>
	WebHCat	<yyyy-MM-dd HH:mm:ss,SSS> <Log Level> <产生该日志的线程名字> <log中的message> <日志事件的发生位置>

# 日志格式

日志类型	组件	格式
审计日志	HiveServer	<yyyy-MM-dd HH:mm:ss,SSS> <Log Level> <产生该日志的线程名字> <log中的message> <日志事件的发生位置>
	MetaStore	<yyyy-MM-dd HH:mm:ss,SSS> <Log Level> <产生该日志的线程名字> <log中的message> <日志事件的发生位置>
	WebHCat	<yyyy-MM-dd HH:mm:ss,SSS> <Log Level> <产生该日志的线程名字> <log中的message> <日志事件的发生位置>



# 日志级别介绍

## 日志级别

Hive中提供了多种的日志级别。日志级别优先级从高到低分别是**FATAL**、**ERROR**、**WARN**、**INFO**、**DEBUG**。程序会打印高于或等于所设置级别的日志，设置的日志等级越高，打印出来的日志就越少。

级别	描述
<b>FATAL</b>	<b>FATAL</b> 表示当前事件处理出现严重错误信息，可能导致系统崩溃。
<b>ERROR</b>	<b>ERROR</b> 表示当前事件处理出现错误信息，系统运行出错。
<b>WARN</b>	<b>WARN</b> 表示当前事件处理存在异常信息，但认为是正常范围，不会导致系统出错。
<b>INFO</b>	<b>INFO</b> 表示记录系统及各事件正常运行状态信息。
<b>DEBUG</b>	<b>DEBUG</b> 表示记录系统及系统的调试信息。

# Hive日志收集

- FusionInsight HD提供了简便的日志收集方法，在 系统设置->日志下载 界面，可以设置并下载在一定时间段内的相关日志。界面如下图：

系统设置 > 日志下载

💡 下载日志占用系统资源，请在空闲时间下载日志。

▶ 帮助：主机匹配表达式输入请参考下面主机列表里的匹配方式。

\* 服务：  ... ❌ 服务不能为空

主机：

\* 开始时间：

\* 结束时间：

下载 取消



# 目录

- 1、服务检查与日志
- 2、维护命令**
- 3、性能调优命令与参数
- 4、常见问题与定位

# Hive维护命令

- 检查hive使用的hdfs大小

**hdfs dfs -du -h /user/hive/warehouse**

- 检查是否可以连上hive服务

**beeline -e "show database"**



# 目录

- 1、服务检查与日志
- 2、维护命令
- 3、性能调优命令与参数**
- 4、常见问题与定位

# Hive性能调优命令与参数

参数名称	参数说明
<b>hive.execution.engine</b>	当存在多轮任务且数据量较小时，选择 <b>spark</b> 引擎以获得更短的执行时间
<b>hive.exec.compress.intermediate</b>	是否压缩中间结果，默认 <b>false</b>
<b>hive.intermediate.compression.codec</b>	当压缩中间结果时，选择的压缩格式， <b>Snappy</b> 压缩性能较好，建议使用
<b>hive.exec.reducers.bytes.per.reducer</b>	每多少字节分一个 <b>reduce</b> ，默认 <b>256M</b> 。可根据集群实际情况合理设置

# Hive性能调优命令与参数

参数名称	参数说明
<b>hive.auto.convert.join.noconditionaltask.size</b>	当其它表的文件总大小不大于该值时， <b>common join</b> 转成 <b>map join</b> 。默认 <b>10M</b> ，可根据集群情况适当调大
<b>hive.optimize.index.filter</b>	做查询优化时，是否使用索引信息。当输入文件是 <b>ORC</b> 时，可开启此参数。
<b>hive.optimize.bucketmapjoin.sortedmerge</b>	是否自动将 <b>mapjoin</b> 转化成 <b>sorted bucket merge map join</b> ，多表 <b>join</b> 场景，表都做了分桶+ <b>join</b> 字段排序时，可尝试此参数
<b>hive.optimize.sort.dynamic.partition</b>	是否开启分区字段排序，默认 <b>false</b> 。适用于插入分区表，当分区数较多场景



# 目录

- 1、服务检查与日志
- 2、维护命令
- 3、性能调优命令与参数
- 4、常见问题与定位**



# Hive常见问题1

## Beeline下执行set命令报错：

现象描述	在beeline下执行set命令时提示如下错误 > set mapred.job.queue.name=QueueA; Error: Error while processing statement: Cannot modify mapred.job.queue.name at runtime. It is not in list of params that are allowed to be modified at runtime (state=42000,code=1)
原因分析	为了安全考虑FusionInsight中加入白名单机制，即只有加入白名单的参数才可以通过set命令修改，否则会报上述错误
解决方案	在OM页面点击 Hive>>configuration>>All>>HiveServer>>Security, 找到配置项 hive.security.authorization.sqlstd.confwhitelist, 将需要set的参数加入到配置项然后重启hiveserver即可。

# Hive常见问题 2

现象描述	<b>load data local inpath '/opt/testdata/test.txt' overwrite into table test;</b> <b>Error: Error while compiling statement: FAILED: SemanticException Line 1:23 Invalid path</b> <b>''/opt/testdata/test.txt': No files matching path</b> <b>file:/opt/testdata/test.txt (state=42000,code=40000)</b>
原因分析	<b>HiveServer</b> 进程是以操作系统上的 <b>omm</b> 用户启动的，所以要求 <b>omm</b> 用户对此文件有读权限，对此文件的目录有读、执行权限；同时本地文件应放在 <b>hiveserver</b> 所在节点上。
解决方案	按照以上权限要求修改权限即可。

# Hive常见问题 3

## 移动数据到Hive路径时发生错误

现象描述	当执行 <code>dfs put /opt/kv1.txt /tmp/kv1.txt</code> 类似的上传命令时报 <b>Permission denied. Principal [name=admin, type=USER] does not have following privileges on Object [type=COMMAND_PARAMS,name=[put, /opt/kv1.txt, /tmp/kv1.txt]] for operation DFS : [ADMIN PRIVILEGE] (state=,code=1</b>
原因分析	当前登录的用户不具备操作此命令的权限。
解决方案	如果登录的当前用户具有 <b>admin</b> 角色，请用 <b>set role admin</b> 来切换成 <b>admin</b> 角色操作。如果不具备 <b>admin</b> 角色，在 <b>OM</b> 管理界面 <b>System</b> 下 <b>Role Management</b> 中给对应的角色赋予对应的权限



## 思考题

1. **Hive**的安装目录是什么?
2. **Hive**告警有哪些级别?
3. 如何进行**Hive** 日志收集?



## 习题

- 判断题
  1. **Hive**日志级别更改无需重启服务。 ( )
  2. **Hive**审计日志收集需要到每个**hive**节点上手动下载。 ( )
- 多选题
  1. 下列哪些服务，是**Hive**依赖的服务(包含强依赖和弱依赖)? ( )
    - A.YARN
    - B.Spark
    - C.HDFS
    - D.HBase



## 本章总结

- 本章学习了**Hive**常用维护命令/参数以及性能调优命令/参数，最后通过案例的方式展示了**Hive**常见问题的处理。

# Thank you

[www.huawei.com](http://www.huawei.com)