**实 验 报 告 册**

2023学年 春季 学期

专业班级： 计科6班

学 号： 2020204331

姓 名： 王博文

指导教师： 纪俊

青岛大学计算机科学技术学院

|  |  |  |  |
| --- | --- | --- | --- |
| 实验项目 | Hive数据库操作实例 | 日 期 期 | 2023.6.5 |
| 实验环境 | LINUX | 成绩评定 |  |
| 实验目的和要求：  **搭建**hive  **并实现wordcount计算单词个数的操作**  **参考网站：**[Hive3.1.2安装指南\_厦大数据库实验室博客 (xmu.edu.cn)](https://dblab.xmu.edu.cn/blog/2440/)  [Ubuntu安装MySQL及常用操作\_厦大数据库实验室博客 (xmu.edu.cn)](https://dblab.xmu.edu.cn/blog/1002/) | | | |
| 实验内容或结果：   1. 安装hive，我安装的是mysql-5.7，hive-3.1.2   1. 下载并解压Hive安装包 首先需要下载Hive安装包文件， [Hive官网下载地址](http://www.apache.org/dyn/closer.cgi/hive/) 也可以直接[点击这里从百度云盘下载软件](https://pan.baidu.com/s/1lLjW9cmS1tkBHkrqtpkjWw)（提取码：ziyu）。进入百度网盘后，进入“软件”目录，找到apache-hive-3.1.2-bin.tar.gz文件，下载到本地。  sudo tar -zxvf ./apache-hive-3.1.2-bin.tar.gz -C /usr/local # 解压到/usr/local中  cd /usr/local/  sudo mv apache-hive-3.1.2-bin hive # 将文件夹名改为hive  sudo chown -R dblab:dblab hive # 修改文件权限  Shell 命令  注意，上面的dblab:dblab是用户组和用户名，如果你当前使用用户名hadoop登录了Linux系统，则把dblab替换成hadoop。 2. 配置环境变量 为了方便使用，我们把hive命令加入到环境变量中去， 请使用vim编辑器打开.bashrc文件，命令如下：  vim ~/.bashrc  Shell 命令  在该文件最前面一行添加如下内容：  export HIVE\_HOME=/usr/local/hive  export PATH=$PATH:$HIVE\_HOME/bin  export HADOOP\_HOME=/usr/local/hadoop  HADOOP\_HOME需要被配置成你机器上Hadoop的安装路径，比如这里是安装在/usr/local./hadoop目录。 保存退出后，运行如下命令使配置立即生效：  source ~/.bashrc  Shell 命令  3. 修改/usr/local/hive/conf下的hive-site.xml 执行如下命令：  cd /usr/local/hive/conf  mv hive-default.xml.template hive-default.xml  Shell 命令  上面命令是将hive-default.xml.template重命名为hive-default.xml； 然后，使用vim编辑器新建一个配置文件hive-site.xml，命令如下：  cd /usr/local/hive/conf  vim hive-site.xml  Shell 命令  在hive-site.xml中添加如下配置信息：  <?xml version="1.0" encoding="UTF-8" standalone="no"?>  <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>  <configuration>  <property>  <name>javax.jdo.option.ConnectionURL</name>  <value>jdbc:mysql://localhost:3306/hive?createDatabaseIfNotExist=true</value>  <description>JDBC connect string for a JDBC metastore</description>  </property>  <property>  <name>javax.jdo.option.ConnectionDriverName</name>  <value>com.mysql.jdbc.Driver</value>  <description>Driver class name for a JDBC metastore</description>  </property>  <property>  <name>javax.jdo.option.ConnectionUserName</name>  <value>hive</value>  <description>username to use against metastore database</description>  </property>  <property>  <name>javax.jdo.option.ConnectionPassword</name>  <value>hive</value>  <description>password to use against metastore database</description>  </property>  </configuration>  然后，按键盘上的“ESC”键退出vim编辑状态，再输入:wq，保存并退出vim编辑器。  二、安装并配置mysql  这里我们采用MySQL数据库保存Hive的元数据，而不是采用Hive自带的derby来存储元数据。 1.Ubuntu下mysql的安装请参考：[Ubuntu安装MySQL](https://dblab.xmu.edu.cn/blog/install-mysql/) 2.下载mysql jdbc 包,[下载地址](http://www.mysql.com/downloads/connector/j/)  tar -zxvf mysql-connector-java-5.1.40.tar.gz #解压  cp mysql-connector-java-5.1.40/mysql-connector-java-5.1.40-bin.jar /usr/local/hive/lib #将mysql-connector-java-5.1.40-bin.jar拷贝到/usr/local/hive/lib目录下  Shell 命令  3. 启动并登陆mysql shell  service mysql start #启动mysql服务  mysql -u root -p #登陆shell界面  Shell 命令  4. 新建hive数据库。  mysql> create database hive; #这个hive数据库与hive-site.xml中localhost:3306/hive的hive对应，用来保存hive元数据  mysql  5. 配置mysql允许hive接入：  mysql> grant all on \*.\* to hive@localhost identified by 'hive'; #将所有数据库的所有表的所有权限赋给hive用户，后面的hive是配置hive-site.xml中配置的连接密码  mysql> flush privileges; #刷新mysql系统权限关系表  mysql  6. 启动hive 启动hive之前，请先启动hadoop集群。  start-dfs.sh #启动Hadoop的HDFS  hive #启动hive  Shell 命令  注意，我们这里已经配置了PATH，所以，不要把start-all.sh和hive命令的路径加上。如果没有配置PATH，请加上路径才能运行命令，比如，本教程Hadoop安装目录是“/usr/local/hadoop”，Hive的安装目录是“/usr/local/hive”，因此，启动hadoop和hive，也可以使用下面带路径的方式：  cd /usr/local/hadoop #进入Hadoop安装目录  ./sbin/start-dfs.sh  cd /usr/local/hive  ./bin/hive  cd /usr/local/hive  ./bin/schematool -dbType mysql -initSchema  Shell 命令  执行后，再启动Hive，应该就正常了。  启动进入Hive的交互式执行环境以后，会出现如下命令提示符：  hive>  hive  可以在里面输入SQL语句，如果要退出Hive交互式执行环境，可以输入如下命令：  hive>exit;   1. 由于hive是部署在hdfs上的组件，故打开hive之前要打开hdfs和yarn框架来调度资源，而hive产生的元数据存储在mysql中，故打开hdfs和yarn之后，且在打开hive之前要打开mysql。 2. Z)5V69UO1OCLBMVRIKFVH%C创建名为hh的数据库并使用   4)]JRHCPH1YFCSD$`J[VF(3   1. 创建名为docs的表格，运行结果如下：   IMG_256   1. 将圣经这个文件中的数据装载进docs表中 2. 最后一步，将各词汇装进word\_count表中，以空格划分   create table word\_count as  select word, count(1) as count from  (select explode(split(line,' '))as word from docs) w  group by word  order by word;  IMG_256   1. 执行完成后，用select语句查看结果如下   select \* from word\_count;  IMG_256  IMG_256 | | | |
| 思考或体会：  这个hive的问题也还是不少的  1.在刚开始装mysql时，用的是教程里的操作，但我的ubuntu是20.4的，执行指令sudo apt-get update #更新软件源  sudo apt-get install mysql-server后，我看到我下的版本是8.0，但当时也就没多想，以为一般软件都能向下兼容，所以就继续做了下去。  教程中的有一步要改变编码方式，进入mysql -u root -p（mysql shell）内部，但需要密码。装8.0时不会像5.7一样弹出来问你密码，他的密码时随机的，所以需要进去查看密码然后更改他，这是我完成此目的的网站**https://blog.csdn.net/m0\_70885101/article/details/127414184**  2.换成了自己的密码后就继续按照教程走，把mysql jdbc解压，直到做这步  发现怎样都运行失败，因此我往回倒退，一步一步寻找问题，最终发现是版本的问题，因为是用指令下载的，所以就上网搜索  **彻底删除mysql**  **https://blog.csdn.net/sixiangfy/article/details/128374171**  最终把mysql彻底删除，然后重新安装5.7版本的mysql  **<https://blog.csdn.net/liuhuango123/article/details/128264867>** | | | |
| 备注：  无 | | | |