

# Hive简介

- **Hive由来：**

Hive是著名的社交网站Facebook研发的，在Facebook中有很大的一个hadoop集群（有庞大的数据存储和需要统计（计算）的需求）

计算：MapReduce就能解决这些问题

公司的情况是运维人员（管服务器，管数据库（DBA）、安全等等）大于开发人员（编码、设计）

DBA:主管数据库的设计、安全和优化——他们写的sql质量一般优于开发人员

让所有的DBA都学习MapReduce，写一些自己想要的需求，但MapReduce在写之前，至少要会一门编程语言（java），难度太大，所以Facebook让开发人员写一个工具，将sql转换MapReduce在hadoop中运行！

- **所以Hive到底是个什么东西？**

Hive是基于Hadoop的一个数据仓库工具，可以将**结构化的数据**文件映射为一张数据库表，并提供类SQL查询功能。

其本质是将SQL转换为MapReduce的任务进行运算，底层由HDFS来提供数据的存储，说白了hive可以理解为一个将SQL转换为MapReduce的任务的工具，甚至更进一步可以说hive就是一个MapReduce的客户端。

- **为什么要去使用Hive？**

**优点：**

- \* 采用类SQL语法去操作数据，提供快速开发的能力。
- \* 避免了去写MapReduce，减少开发人员的学习成本。
- \* 功能扩展很方便。

**缺点：**

hive的缺点（局限性）

(1) hive的HQL表达能力有限

1)迭代式算法无法表达，比如pagerank

2)数据挖掘方面，比如kmeans

(2) hive的效率比较低

1)hive自动生成的mapreduce作业，通常情况下不够智能化

2)hive调优比较困难，粒度较粗

3)hive可控性差

# 安装Hive

1 官网下载: <https://hive.apache.org/downloads.html>

2 上传安装包到安装目录, 解压安装 `tar -zxvf 安装包`

3 安装mysql, 因为Hive的元数据我们使用mysql管理

第一步: 在线安装mysql相关的软件包

```
yum install mysql mysql-server mysql-devel
```

第二步: 启动mysql的服务

```
/etc/init.d/mysqld start
```

第三步: 通过mysql安装自带脚本进行设置

```
/usr/bin/mysql_secure_installation
```

第四步: 进入mysql的客户端然后进行授权

```
grant all privileges on *.* to 'root'@'%' identified by '123456' with grant option;  
flush privileges;
```

4 修改hive配置文件

1) 修改hive-env.sh (本来没有这个文件, 需要去安装目录下conf文件夹下复制hive-env.sh.template一份) 如:

```
1 cd /export/servers/apache-hive-2.1.1-bin/conf  
2 #复制一份改名  
3 cp hive-env.sh.template hive-env.sh
```

2) 修改hive-site.xml (这个文件本身没有, 需要在配置文件夹conf下新建一个) 如:

```
1 cd /export/servers/apache-hive-2.1.1-bin/conf  
2 vim hive-site.xml
```

具体配置内容:

```
1 <?xml version="1.0" encoding="UTF-8" standalone="no"?>  
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>  
3 <configuration>  
4 <!-- 配置连接数据库用户名 -->  
5 <property>  
6     <name>javax.jdo.option.ConnectionUserName</name>  
7     <value>root</value>  
8 </property>
```

```
9      <!-- 配置连接数据库密码 -- >
10      <property>
11          <name>javax.jdo.option.ConnectionPassword</name>
12          <value>123456</value>
13      </property>
14      <!-- 配置连接数据库地址 -- >
15      <property>
16          <name>javax.jdo.option.ConnectionURL</name>
17          <value>jdbc:mysql://node03:3306/hive?createDatabaseIfNotExist=true&useS
18      </property>
19      <!-- 配置连接数据库驱动 -- >
20      <property>
21          <name>javax.jdo.option.ConnectionDriverName</name>
22          <value>com.mysql.jdbc.Driver</value>
23      </property>
24      <property>
25          <name>hive.metastore.schema.verification</name>
26          <value>>false</value>
27      </property>
28      <property>
29          <name>datanucleus.schema.autoCreateAll</name>
30          <value>true</value>
31      </property>
32      <property>
33          <name>hive.server2.thrift.bind.host</name>
34          <value>node03</value>
35      </property>
36  </configuration>
```