

大数据数据库系统

6.9 hiveserver2与case when

◆ 主要内容

hiveserver2

Case when语句

Cast函数

unix_timestamp函数

hiveserver2

◆ Hive的连接方法

之前介绍的通过hive shell的连接方式

Hiveserver2方式

将hive的服务变成服务端，可以通过客户端进行连接

hiveserver2

◆ Hiveserver2方式

进入hive/bin/目录执行hiveserver2

```
[hpsk@bigdata-training01 hive-0.13.1-bin]$ ll bin/  
total 32  
-rwxr-xr-x 1 hpsk hpsk 881 Jan 30 2014 beeline  
drwxrwxr-x 3 hpsk hpsk 4096 May 23 11:11 ext  
-rwxr-xr-x 1 hpsk hpsk 7164 May 13 2014 hive  
-rwxr-xr-x 1 hpsk hpsk 1900 Jan 30 2014 hive-config.sh  
-rwxr-xr-x 1 hpsk hpsk 885 Jan 30 2014 hiveserver2  
-rwxr-xr-x 1 hpsk hpsk 832 Jan 30 2014 metatool  
-rwxr-xr-x 1 hpsk hpsk 884 Jan 30 2014 schematool  
[hpsk@bigdata-training01 hive-0.13.1-bin]$
```

可以看到bin目录下有个hiveserver2，启动hiveserver2启动后，hive就以服务器的形式提供服务

hiveserver2

◆ Hiveserver2的配置

在hive /conf所在目录下的hive-default.xml中可以查看到一些有关Hiveserver2的配置项

1、确认hiveserver2服务端口号配置

```
<property>
  <name>hive.server2.thrift.port</name>
  <value>10000</value>
  <description>Port number of HiveServer2 Thrift interface.
  Can be overridden by setting $HIVE_SERVER2_THRIFT_PORT</description>
</property>
```

提供给客户端连接的端口，默认端口号为10000

2、hadoop的配置文件core-site.xml增加如下配置，重启hdfs。将“xxx”替换成自己的用户名

```
<property>
```

```
    <name>hadoop.proxyuser.xxx.hosts</name>
```

```
    <value>*</value>
```

```
</property>
```

```
<property>
```

```
    <name>hadoop.proxyuser.xxx.groups</name>
```

```
    <value>*</value>
```

```
</property>
```

“*” 表示可通过超级代理“xxx”操作hadoop的用户、用户组和主机

hiveserver2

3、修改HDFS的文件系统权限，执行命令行，不设置会报错：

```
bin/hdfs dfs -chmod -R 777 /tmp/
```

2、3两步完成后一定要重启hadoop，否则不会生效，运行“start-all.sh”

4、初始化元数据，执行命令行：

```
schematool -dbType mysql -initSchema
```

5、启动Hive 的 Metastore Server服务进程，执行如下命令：

```
hive --service metastore &
```

hiveserver2

查看启动metastore的进程号：

```
ps -aux | grep 'metastore'
```

- 如果成功启动了metastore，则会显示进程和进程号；
- 如果启动metastore提示已经有metastore在运行则可以使用kill命令杀死进程
- kill -9 进程号

至此，hiveserver2的服务器端就已经配置完成

hiveserver2

◆ 启动hiveserver2服务

进入/hive/bin/目录，执行hiveserver2，就完成了hiveserver2的启动

```
[hpsk@bigdata-training01 hive-0.13.1-bin]$ bin/hiveserver2
Starting HiveServer2
```

◆ 客户端连接到hiveserver2

进入/hive/bin/目录下可以看到beeline（hive的标志就是个蜜蜂）

```
[hpsk@bigdata-training01 hive-0.13.1-bin]$ ll bin/
total 32
-rwxr-xr-x 1 hpsk hpsk 881 Jan 30 2014 beeline
drwxrwxr-x 3 hpsk hpsk 4096 May 23 11:11 ext
-rwxr-xr-x 1 hpsk hpsk 7164 May 13 2014 hive
-rwxr-xr-x 1 hpsk hpsk 1900 Jan 30 2014 hive-config.sh
-rwxr-xr-x 1 hpsk hpsk 885 Jan 30 2014 hiveserver2
-rwxr-xr-x 1 hpsk hpsk 832 Jan 30 2014 metatool
-rwxr-xr-x 1 hpsk hpsk 884 Jan 30 2014 schematool
[hpsk@bigdata-training01 hive-0.13.1-bin]$
```

hiveserver2

◆ 通过Beeline客户端连接到hiveserver2

查看帮助信息： bin/beeline -h

```
[hpsk@bigdata-training01 hive-0.13.1-bin]$ bin/beeline -h
Usage: java org.apache.hive.cli.beeline.BeeLine
  -u <database url>      the JDBC URL to connect to
  -n <username>          the username to connect as
  -p <password>          the password to connect as
  -d <driver class>      the driver class to use
  -e <query>             query that should be executed
  -f <file>              script file that should be executed
  --hiveconf property=value Use value for given property
  --hivevar name=value    hive variable name and value
                          This is Hive specific settings in which variables
                          can be set at session level and referenced in Hive
                          commands or queries.
```

-u <database url> 指定要连接hiveserver2的URL地址 (jdbc方式连接)

-n <username> 指定以哪个用户连接到hiveserver2

-p <password> 输入用户密码

hiveserver2

- ◆ 如果要配置连接hiveserver2的用户名密码，修改如下配置项

```
1 <property>
2   <name>hive.server2.thrift.client.user</name>
3   <value>data1</value>
4   <description>Username to use against thrift client</description>
5 </property>
6 <property>
7   <name>hive.server2.thrift.client.password</name>
8   <value>123456</value>
9   <description>Password to use against thrift client</description>
10 </property>
```

hiveserver2

◆ 连接方法一：

1、查询hiveserver2的主机的hostname

```
[hpsk@bigdata-training01 hive-0.13.1-bin]$ hostname  
bigdata-training01.hpsk.com
```

2、新打开一个终端，输入

bin/beeline -u jdbc:hive2://**bigdata-training01.hpsk.com**:10000 -n hpsk -p hpsk

端口号

密码

主机的hostname

用户名

```
[hpsk@bigdata-training01 hive-0.13.1-bin]$ bin/beeline -u jdbc:hive2://bigdata-training01.hpsk.com:10000 -n hpsk -p hpsk  
scan complete in 3ms  
Connecting to jdbc:hive2://bigdata-training01.hpsk.com:10000  
Connected to: Apache Hive (version 0.13.1)  
Driver: Hive JDBC (version 0.13.1)  
Transaction isolation: TRANSACTION_REPEATABLE_READ  
Beeline version 0.13.1 by Apache Hive  
0: jdbc:hive2://bigdata-training01.hpsk.com:1>
```


连接成功后输入“show databases”进行测试

```
0: jdbc:hive2://bigdata-training01.hpsk.com:1> show databases;
```

database_name
default
emp_test
student
tmp2
tmp3

```
5 rows selected (0.801 seconds)
```

```
0: jdbc:hive2://bigdata-training01.hpsk.com:1>
```

查询emp表的内容

```
0: jdbc:hive2://bigdata-training01.hpsk.com:1> use emp_test;
```

```
No rows affected (0.034 seconds)
```

```
0: jdbc:hive2://bigdata-training01.hpsk.com:1> select * from emp;
```

emp.empno	emp.ename	emp.job	emp.mgr	emp.hiredate	emp.sal	emp.comm	emp.deptn
7369	SMITH	CLERK	7902	1980-12-17	800.0	NULL	20
7499	ALLEN	SALESMAN	7698	1981-2-20	1600.0	300.0	30
7521	WARD	SALESMAN	7698	1981-2-22	1250.0	500.0	30
7566	JONES	MANAGER	7839	1981-4-2	2975.0	NULL	20
7654	MARTIN	SALESMAN	7698	1981-9-28	1250.0	1400.0	30
7698	BLAKE	MANAGER	7839	1981-5-1	2850.0	NULL	30
7782	CLARK	MANAGER	7839	1981-6-9	2450.0	NULL	10
7788	SCOTT	ANALYST	7566	1987-4-19	3000.0	NULL	20
7839	KING	PRESIDENT	NULL	1981-11-17	5000.0	NULL	10
7844	TURNER	SALESMAN	7698	1981-9-8	1500.0	0.0	30
7876	ADAMS	CLERK	7788	1987-5-23	1100.0	NULL	20
7900	JAMES	CLERK	7698	1981-12-3	950.0	NULL	30
7902	FORD	ANALYST	7566	1981-12-3	3000.0	NULL	20
7934	MILLER	CLERK	7782	1982-1-23	1300.0	NULL	10

```
14 rows selected (0.72 seconds)
```

```
0: jdbc:hive2://bigdata-training01.hpsk.com:1> █
```

使用!quit命令退出

```
0: jdbc:hive2://bigdata-training01.hpsk.com:1> !quit  
Closing: 0: jdbc:hive2://bigdata-training01.hpsk.com:10000  
[hpsk@bigdata-training01 hive-0.13.1-bin]$ █
```

hiveserver2

此时我们回到hiveserver2服务器端看下终端显示的内容

```
[hpsk@bigdata-training01 hive-0.13.1-bin]$ bin/hiveserver2
Starting HiveServer2
OK
OK
OK
```

因为我们刚才在客户端执行了3条hive指令，所以这里显示了3个ok，
如果要退出hiveserver2服务器端，按ctrl+c就可以退出至linux命令行

hiveserver2

◆ 连接方法二

直接在/hive/bin目录下执行beeline，不加参数

```
[hpsk@bigdata-training01 hive-0.13.1-bin]$ bin/beeline  
Beeline version 0.13.1 by Apache Hive  
beeline>
```

此时会直接进入beeline命令行，输入help命令查看帮助，可以看到有一个!connect命令

```
beeline> help  
!all          Execute the specified SQL against all the current connections  
!autocommit   Set autocommit mode on or off  
!batch        Start or execute a batch of statements  
!brief        Set verbose mode off  
!call         Execute a callable statement  
!close        Close the current connection to the database  
!closeall     Close all current open connections  
!columns      List all the columns for the specified table  
!commit       Commit the current transaction (if autocommit is off)  
!connect      Open a new connection to the database  
!dbinfo       Give metadata information about the database  
!describe     Describe a table
```

hiveserver2

输入!connect命令查看!connect命令的格式

```
beeline> !connect  
Usage: connect <url> <username> <password> [driver]
```

输入!connect jdbc:hive2://主机名:10000开始连接

```
beeline> !connect jdbc:hive2://bigdata-training01.hpsk.com:10000  
scan complete in 2ms  
Connecting to jdbc:hive2://bigdata-training01.hpsk.com:10000
```

连接成功后会提示你输入用户名和密码

```
Connecting to jdbc:hive2://bigdata-training01.hpsk.com:10000  
Enter username for jdbc:hive2://bigdata-training01.hpsk.com:10000: hpsk  
Enter password for jdbc:hive2://bigdata-training01.hpsk.com:10000: ****  
Connected to: Apache Hive (version 0.13.1)  
Driver: Hive JDBC (version 0.13.1)  
Transaction isolation: TRANSACTION_REPEATABLE_READ  
) jdbc:hive2://bigdata-training01.hpsk.com:1> █
```

输入正确的用户名和密码后，进入

Case when语句

◆ Case when...then语句的使用

第一种格式:

```
case col  
  
when value1 then value2  
  
when value3 then value4  
  
else value5  
  
End
```

第二种格式:

```
Case  
  
when col=value1 then value2  
  
when col=value3 then value4  
  
else value5  
  
end
```

col表示字段名

如果col的字段值为value1，则返回value2；如果col的字段值为value3，则返回value4；否则返回，则返回value5

例：在emp表中的deptno字段总共有三种，即10，20，30，根据每个员工的部门编号，显示字符串“your part is 10” 或“your part is 20” 或“your part is 30”

写法一：

```
select ename ,  
case deptno  
when 10 then 'your part is 10'  
when 20 then 'your part is 20'  
else 'your part is 30'  
end part  
from emp;
```

写法二：

```
select ename ,  
case  
when deptno='10' then 'your part is 10'  
when deptno='20' then 'your part is 20'  
else 'your part is 30'  
end part  
from emp;
```

```
OK  
ename      part  
SMITH      your part is 20  
ALLEN      your part is 30  
WARD       your part is 30  
JONES      your part is 20  
MARTIN     your part is 30  
BLAKE      your part is 30  
CLARK      your part is 10  
SCOTT      your part is 20  
KING       your part is 10  
TURNER     your part is 30  
ADAMS      your part is 20  
JAMES      your part is 30  
FORD       your part is 20  
MILLER     your part is 10  
Time taken: 29.433 seconds, Fetched: 14 row(s)  
hive (emp_test)> █
```

准备数据：

/opt/module/hive/data/emp_sex.txt

name	dept_id	sex
悟空	A	男
大海	A	男
宋宋	B	男
凤姐	A	女
婷婷	B	女
婷婷	B	女

创建 hive 表并导入数据

```
create table emp_sex(  
name string,  
dept_id string,  
sex string)  
row format delimited fields terminated  
by "\t";  
load data local inpath  
'/opt/module/hive/data/emp_sex.txt'  
into table emp_sex;
```

需求：求出不同部门男女各多少人

```
select
dept_id,
sum(case sex when '男' then 1 else 0 end) male_count,
sum(case sex when '女' then 1 else 0 end) female_count
from emp_sex
group by dept_id;
```

```
dept_id malecount      femalecount
A        2            1
B        1            2
Time taken: 31.056 seconds, Fetched: 2 row(s)
hive (default)> █
```

Cast命令

◆ Cast命令

用于hive中的数据类型转换

格式: `cast (col as type)`

Col表示字段名

Type表示希望转换成的类型

Cast命令

◆ Cast命令

官方文档

Recall our `employees` table uses a `FLOAT` for the `salary` column. Now, imagine for a moment that `STRING` was used for that column instead. How could we work with the values as `FLOATS`?

The following example casts the values to `FLOAT` before performing a comparison:

```
SELECT name, salary FROM employees  
WHERE cast(salary AS FLOAT) < 100000.0;
```

假定salary原来是string类型，无法在where中进行“<”比较，因此使用cast函数将其转化为float类型，

例：将emp表中的sal转化为string类型

1、我们先查看下原先emp表的sal字段是什么类型的：

用Desc formatted emp可以看到是double类型

```
hive (emp_test)> desc formatted emp;  
OK  
col_name      data_type      comment  
# col_name      data_type      comment  
  
empno          int  
ename          string  
job            string  
mgr            int  
hiredate       string  
sal            double  
comm           double  
deptno         int
```

2、 create table caststr as select empno,ename,cast(sal as string) salary
from emp;

3、使用Desc formatted caststr 查看sal的类型

```
hive (emp_test)> desc formatted caststr;  
OK  
col_name      data_type      comment  
# col_name      data_type      comment  
  
empno          int  
ename          string  
salary         string
```

Cast命令

注意：Hive 的部分数据类型是可以进行隐式转换的

例如某表达式使用 INT 类型，TINYINT 会自动转换为 INT 类型，但是 Hive 不会进行反向转化，例如，某表达式使用 TINYINT 类型，INT 不会自动转换为 TINYINT 类型，它会返回错误，除非使用 CAST 操作。

Cast命令

◆ 隐式类型转换规则：

- (1) 任何整数类型都可以隐式地转换为一个范围更广的类型，如 TINYINT 可以转换成INT，INT 可以转换成 BIGINT。
- (2) 所有整数类型、FLOAT 和 **STRING 类型**都可以隐式地转换成 DOUBLE。
- (3) TINYINT、SMALLINT、INT 都可以转换为 FLOAT。
- (4) BOOLEAN 类型不可以转换为任何其它的类型。

上面的STRING类型指的是可以转换成double类型的数字字符串，比如字符串“32.6”，如果是“abc”是不能隐式转换的

Cast命令

可以隐式转换的STRING类型指的是可以转换成double类型的数字字符串，比如字符串“32.6”，如果是“abc”是不能隐式转换成double，也无法使用cast强制转换成int，float类型

例1: CAST('1' AS INT)将把STRING '1' 转换成INT 1;

CAST('X' AS INT)则会转换失败，返回空值 NULL

例2: 输入下述语句，结果是什么？

```
select '1'+2, cast('1'as int) + 2;
```

_c0	_c1
3.0	3

为什么？

unix_timestamp命令

◆ unix_timestamp命令的使用

功能：将标准时间格式转换为Unix_timestamp，一般用于时间差值的计算

参数的格式：yyyy-MM-dd HH:mm:ss

这种格式存在的问题，如果要求两个时间的时间差比较麻烦，如计算会话时长，进入页面的时间与离开页面的时间的差值，得到在页面的停留时间

UNIX时间：记录从1970年1月1日0点0分0秒到目前为止的秒数

1、unix_timestamp() 获取当前时间戳

2、unix_timestamp(string timestamp)

输入的时间戳格式必须为'yyyy-MM-dd HH:mm:ss',如不符合则返回
null

select unix_timestamp('2019-08-15 16:40:00') 返回1565858400

select unix_timestamp('2019-08-15') 返回null

3、unix_timestamp(string date,string pattern)

将指定时间字符串格式字符串转化成unix时间戳,如不符合则返回null

select unix_timestamp('2019-08-15','yyyy-MM-dd') 返回1565798400

select unix_timestamp('2019-08-15 16:40:00','yyyy-MM-dd HH:mm:ss')

返回1565858400

select unix_timestamp('2019-08-15','yyyy-MM-dd HH:mm:ss') 返回null

◆ 百度“unixtimestamp”，可以找到站长工具

当前位置: [站长工具](#) > [Unix时间戳](#)

广告 国际短信推广, 灰鸽子值得您

[Unicode编码](#)

[UTF-8编码](#)

[URL编码/解码](#)

[Unix时间戳](#)

[Ascii/Native编码互转](#)

[Hex编码/解码](#)

[Html编码/解码](#)

现在的Unix时间戳(Unix timestamp)是:

1610100510

开始

停止

刷新

Unix时间戳 (Unix timestamp)

1610100459

秒

转换

时间 (年/月/日 时:分:秒)

转换成Unix时间戳

秒

时间

2021

年

月

日

时

分

秒

转换成Unix时间戳

秒

例：将“2017-01-01 00:00:00”转换为unix_timestamp格式

```
select unix_timestamp('2017-01-01 00:00:00') from emp;
```

```
OK
_c0
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
1483200000
Time taken: 18.169 seconds, Fetched: 14 row(s)
hive (emp_test)>
```

使用站长工具进行转换，两者得到的时间一样

The screenshot shows a web-based time conversion tool with three main sections:

- Top Section:** A label "Unix时间戳 (Unix timestamp)" is followed by a text input field containing "1495763267". To its right is a dropdown menu set to "秒". A button labeled "转换成北京时间" is followed by a text output field showing "2017/5/26 9:47:47".
- Middle Section:** A label "北京时间 (年/月/日 时:分:秒)" is followed by an empty text input field. To its right is a button labeled "转换成Unix时间戳", followed by another empty text input field and a dropdown menu set to "秒".
- Bottom Section:** A label "北京时间" is followed by a series of input fields for date and time: "2017" (year), "1" (month), "1" (day), "0" (hour), "0" (minute), and "0" (second). To the right of these fields is a button labeled "转换Unix时间戳", followed by a text output field containing "1483200000" and a dropdown menu set to "秒".

from_unixtime命令

◆ from_unixtime命令的使用

与unix_timestamp功能相反

格式：from_unixtime(bigint unixtime,string format)

将时间戳秒数转化为UTC时间，并用字符串表示，可通过format规定的时间格式，指定输出的时间格式

例：

```
select from_unixtime(1565858389,'yyyy-MM-dd HH:mm:ss');
```

返回 2019-08-15 16:39:49

```
select from_unixtime(1565858389,'yyyy-MM-dd');
```

返回2019-08-15

```
select from_unixtime(unix_timestamp(),'yyyy-MM-dd HH:mm:ss')
```

获取当前时间

总结

- ◆ Hive的连接方法与hiveserver2

Hiveserver2建立服务器，如何配置

beeline连接hiveserver2两种方法

总结

◆ Case when语句的使用

第一种格式:

```
case col  
  
when value1 then value2  
  
when value3 then value4  
  
else value5  
  
End
```

第二种格式:

```
Case  
  
when col=value1 then value2  
  
when col=value3 then value4  
  
else value5  
  
end
```

总结

◆ Cast语句

强制转换类型

隐式转换规则

◆ unix_timestamp函数

◆ from_unixtime函数的使用