任务 3-3 认识Hive命令

【幻灯片 1-5】 任务3-3 认识Hive命令 - 任务描述

第三个任务——认识Hive命令。

要在Hive上操作数据，首先得知道怎么和Hive交互（jiāo hù）。

[슬라이드 4와 5같이 설명]

最常用的方式就是命令行界面，也叫CLI（Command Line Interface）。

Hive的数据保存在HDFS中，所以在Hive命令行里可以直接执行HDFS命令来查看文件。

甚至还可以在Hive里执行Linux系统命令！

【幻灯片 6】 3.3.3 Hive常用交互命令 - 第一代客户端

Hive有两代客户端。先看第一代客户端 Hive CLI。

启动很简单，在Linux终端输入hive就行了。

리눅스를 시동

```
-- 터미널 하나 open
~/hadoop.sh
```

선택 : 1번

```
-- 方式2: 查看帮助
hive --service cli -help
```

[다음 슬라이드 7]

```
-- 常用参数:
hive -e "select * from student;"     -- 直接执行SQL
hive -f /hivef.sql                    -- 执行SQL文件
hive --database test                  -- 指定数据库
hive -S -e "select * from student;"  -- 静默模式（不显示进度）
```

🔧 实践操作: 第一代客户端启动方式

```
-- 方式1: 直接启动
hive


-- 方式2: 查看帮助
hive --service cli --help


-- 常用参数:
hive -e "select * from student;"     -- 直接执行SQL
hive -f /hivef.sql                    -- 执行SQL文件
hive --database test                  -- 指定数据库
hive -S -e "select * from student;"  -- 静默模式（不显示进度）
```

[슬라이드 8을 보면서]

看幻灯片上的参数表：

-e 参数最常用，直接在命令行后面跟SQL语句。

-f 参数用来执行写好的SQL脚本（jiǎo běn）文件。

-S 参数是静默模式，不显示MapReduce的执行进度（jìn dù）。

【幻灯片 9】 3.3.3 第二代客户端 Beeline CLI

第二代客户端叫Beeline，是目前推荐（tuī jiàn）使用的客户端。

它通过JDBC协议连接HiveServer2服务。

实践操作: 第二代客户端连接方式

```
-- 方式1: 先进入beeline, 再连接
/usr/local/hive/bin/beeline
beeline> !connect jdbc:hive2://localhost:10000
-- 然后输入用户名和密码


-- 方式2: 启动时直接连接
beeline -u jdbc:hive2://localhost:10000 -n root -p
-- 然后输入密码
```

그러나 beeline은 사용하기에 불편. 그래서 alias를 만들었다.

Beeline의 alias는 hive-bee이다.
앞으로 우리는 실습하면서 hive-bee를 사용할 것이다.

然而，beeline使用起来不太方便，所以我创建了一个别名。
Beeline的别名是**hive-bee**。
接下来我们在实操中会使用hive-bee。

```
hadoop@hadoop01:~$ hive-bee
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.17.1.jar!/
org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-r
eload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Connecting to jdbc:hive2://localhost:10000
Connected to: Apache Hive (version 3.1.3)
Driver: Hive JDBC (version 3.1.3)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.3 by Apache Hive
0: jdbc:hive2://localhost:10000>
```

【幻灯片 10-13】 3.3.3 Hive其他操作命令

在Hive里还有很多实用的命令：

🔧 实践操作: Hive中的实用命令

```
-- 在Hive中查看HDFS文件
dfs -ls -R /;
dfs -ls /user/hive/warehouse;


-- 在Hive中查看HDFS文件内容
dfs -cat /user/hive/warehouse/test1/test1.txt;


-- 在Hive中执行Linux命令（前面加!号，感叹号（gǎn tàn hào））
!ls -la /usr/local/hive;
!pwd;


-- 查看历史命令
-- 在Linux中：cat ~/.hivehistory


-- 清屏：CTRL + L
-- 退出：CTRL + C 或者输入 !quit 或 !exit


-- Tab键自动补全：输入SELE后按Tab自动补全为SELECT


-- Hive脚本中添加注释：用 -- 开头
-- 这是一条注释
SELECT * FROM test1;
```

💡 *Beeline中退出用 !quit 或 !exit 或 Ctrl+D*

课程总结（kè chéng zǒng jié）

好，今天的课程到这里就结束了。我们来回顾（huí gù）一下今天学了什么：

任务3-1 认识数据类型：

- 原生类型：数值（INT/DOUBLE）、字符串（STRING）、日期（DATE）、布尔（BOOLEAN）
- 复杂类型：ARRAY数组、MAP映射、STRUCT结构体
- 类型转换：隐式转换（小→大自动转）、显式转换（CAST函数）

任务3-2 认识运算符：

- 关系运算符：=, <>, >, <, >=, <=, IS NULL, LIKE, RLIKE
- 算术运算符：+, -, *, /, %
- 逻辑运算符：AND, OR, NOT
- 复杂运算符：A[n], M['key'], S.x

任务3-3 认识Hive命令：

- 两代客户端：Hive CLI 和 Beeline CLI
- HDFS命令、Linux命令、Tab补全、历史命令

任务3-4 存储数据：

- 行式存储 vs 列式存储
- TextFile（默认）、SequenceFile、RCFile、ORC

大家课后要多实践（shí jiàn），把今天学的命令都敲（qiāo）一遍。有问题随时问！

谢谢大家！下课！