

# HDFS脚本语法修正



为什么脚本递归删除/user/hadoop/zjhshagou时不询问我是否删除/user/hadoop目录的内容？



在你的描述中，脚本递归删除`/user/hadoop/zjhshagou`时，并未询问是否删除`/user/hadoop`目录的内容。这是因为递归删除是从指定的目录`/user/hadoop/zjhshagou`开始的，并且**不会影响上级目录（即`/user/hadoop`）的内容**，这是一种正常行为。

## 原因分析：

### 1. HDFS 删除逻辑的作用范围

- `hdfs dfs -rm -r /user/hadoop/zjhshagou`` 的作用范围仅限于删除`/user/hadoop/zjhshagou`目录及其内容。
- 它不会递归删除其父目录`/user/hadoop`的内容。也就是说，删除子目录时，上层目录（如`/user/hadoop`）是独立存在的，不会被影响。

### 2. 脚本逻辑未包含对上级目录的递归删除

- 在你的脚本中，删除的范围是用户传递的路径`\$HDFS\_PATH`。例如，传递的路径是`/user/hadoop/zjhshagou`，脚本仅删除该路径及其内容，而不会再向上操作`/user/hadoop`。
- **HDFS 的命令设计**：`hdfs dfs -rm -r`` 仅会删除指定路径及以下的文件或子目录，而不会触碰父目录。

### 3. 设计上的合理性

- 在正常情况下，删除`/user/hadoop/zjhshagou`不应该涉及到`/user/hadoop`，因为`/user/hadoop`可能包含其他子目录（例如`/user/hadoop/another\_dir`）。如果脚本不加区分地删除上级目录的内容，就会导致严重的数据丢失。

### 4. 脚本未包含上级目录删除逻辑

- 如果你希望在删除子目录`/user/hadoop/zjhshagou`时，询问是否删除`/user/hadoop`的内容，你需要显式在脚本中添加逻辑。但通常这不是推荐的行为，因为上级目录往往包含其他重要内容。

## 正常的行为解释：

- 当你运行脚本删除`/user/hadoop/zjhshagou`时，删除的范围是`/user/hadoop/zjhshagou`及其子内容。
- `/user/hadoop`` 是该目录的父级，脚本逻辑不会操作上级目录，因为：
  1. 父目录和子目录是分开的实体。
  2. 删除子目录不会自动触发对父目录的操作，除非你明确编写逻辑要求删除父目录。

## 如果你想递归删除上级目录的内容：

假设你希望在删除某个子目录（如`/user/hadoop/zjhshagou`）后，询问是否删除其上级目录（如`/user/hadoop`），可以在脚本中手动添加逻辑来实现。但你需要小心处理，以免误删数据。

以下是一个改进版本，添加逻辑在删除 `/user/hadoop/zjhshagou` 后，询问用户是否删除 `/user/hadoop`：

## 改进后的脚本：

```
bash

#!/bin/bash

HDFS_PATH=$1

# 检查路径是否存在，不存在则退出
echo "检查路径: $HDFS_PATH"
hdfs dfs -test -d "$HDFS_PATH"
if [ $? -ne 0 ]; then
    echo "路径 $HDFS_PATH 不存在"
    exit 1
else
    echo "路径已存在: $HDFS_PATH"
fi

# 删除指定路径及其内容
delete_directory() {
    local path=$1
    # 检查路径是否为空
    files=$(hdfs dfs -ls "$path" | wc -l)

    if [ "$files" -gt 1 ]; then
        # 目录非空，询问是否删除
        echo "目录 $path 非空，是否删除该目录及其所有内容？ (Y/N)"
        read -p "请输入 y 或 n : " USER_CHOICE
        if [[ "$USER_CHOICE" = "Y" || "$USER_CHOICE" = "y" ]]; then
            echo "正在删除目录 $path 下的所有内容..."
            hdfs dfs -rm -r "$path"
            if [ $? -eq 0 ]; then
                echo "目录 $path 及其内容删除成功"
            else
                echo "目录 $path 删除失败！"
                exit 1
            fi
        else
            echo "目录 $path 内容未删除"
            return # 如果用户选择不删除，退出函数
        fi
    else
        # 如果目录为空，直接删除
        echo "目录 $path 为空，直接删除"
        hdfs dfs -rmdir "$path"
        if [ $? -eq 0 ]; then
            echo "目录 $path 删除成功"
        else
            echo "目录 $path 删除失败！"
            exit 1
        fi
    fi
}

# 调用函数删除子目录
delete_directory "$HDFS_PATH"

# 检查上级目录是否删除
PARENT_PATH=$(dirname "$HDFS_PATH")
echo "是否删除上级目录 $PARENT_PATH 的内容？ (Y/N)"
read -p "请输入 y 或 n : " PARENT_CHOICE
if [[ "$PARENT_CHOICE" = "Y" || "$PARENT_CHOICE" = "y" ]]; then
```

```
delete_directory "$PARENT_PATH"
else
    echo "上级目录 $PARENT_PATH 保留"
fi
```

## 改进点说明：

### 1. 新增上级目录逻辑：

- 使用 ``dirname "$HDFS_PATH"`` 获取上级目录路径。
- 在删除子目录后，询问用户是否删除上级目录的内容。

### 2. 独立的删除函数：

- ``delete_directory`` 函数可以被复用，删除指定路径及其内容。
- 上级目录删除逻辑和子目录删除逻辑分开处理，防止误操作。

### 3. 保护机制：

- 每次删除前都询问用户是否确认，防止误删。
- 如果用户选择不删除上级目录，脚本会直接退出，而不会误删其他内容。

## 执行逻辑：

### 1. 假设路径为 ``/user/hadoop/zjhshagou``：

- 脚本会检查 ``/user/hadoop/zjhshagou`` 是否存在。
- 如果存在，询问用户是否删除其内容及子目录。

### 2. 删除 ``/user/hadoop/zjhshagou`` 后：

- 脚本会自动获取其父目录 ``/user/hadoop``。
- 询问用户是否删除 ``/user/hadoop`` 的内容。

### 3. 用户可以选择删除 ``/user/hadoop`` 或保留。

## 结果：

- 默认情况下，脚本只删除子目录内容，不会自动删除上级目录内容。
- 如果用户需要，也可以通过脚本删除上级目录的内容，确保逻辑清晰且数据安全。

希望这能解决你的疑问！如果有进一步需求，随时告诉我。