# 下载指定位置的数据

下载附件中的数据文件【任务2清洗后的数据.zip】，解压，使用SFTP相关工具将数据传至slave2服务器的/home目录下

# 使用Hadoopshell命令创建目录/data/

Hadoop fs –mkdir /data

# 上传数据至HDFS的/data目录下

使用命令：hadoop fs –put /home/任务2清洗后的数据 /data/

# Hive创建数据库

Create database db\_phone\_raw\_2;

# Hive创建表

## 原始数据表

Create table tbl\_phone\_data\_2(fld\_Brand\_Name string, fld\_Phone\_name string, fld\_sales\_time string, fld\_sales\_channel string, fld\_user\_score string) row format delimited fields terminated by '|';

## 手机型号销售时间统计表

Create table tbl\_sales\_time\_count(fld\_phone\_name string, fld\_sales\_time string，fld\_sale\_count int);

## 手机销量渠道统计表

Create table tbl\_sales\_channel\_count(fld\_phone\_name string,fld\_sales\_channel string ,fld\_sale\_count int);

## 手机用户评分统计表

Create table tbl\_user\_score\_count(fld\_phone\_name string, fld\_sale\_count int,fld\_user\_score string);

# 数据加载命令

Load data inpath '/Clean\_Data/\*' into table tbl\_phone\_data\_2;

# 查询分析

## 手机型号销售时间统计表

insert overwrite table tbl\_sales\_time\_count select fld\_phone\_name, fld\_sales\_time,count(1) as fld\_sale\_count from tbl\_phone\_data\_2 group by fld\_phone\_name, fld\_sales\_time;

## 手机销量渠道统计表

insert overwrite table tbl\_sales\_channel\_count select fld\_phone\_name, fld\_sales\_channel,count(1) as fld\_sale\_count from tbl\_phone\_data\_2 group by fld\_phone\_name, fld\_sales\_channel;

## 手机销量前十的手机评分

insert overwrite table tbl\_user\_score\_count select \* from (select fld\_phone\_name, count(1) as fld\_sale\_count,avg(fld\_user\_score) from fld\_sale\_count group by fld\_phone\_name) a order by fld\_sale\_count desc limit 10;