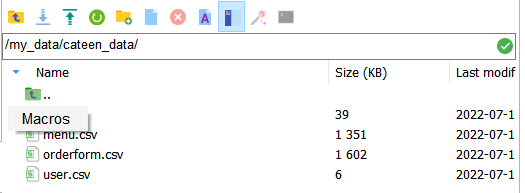
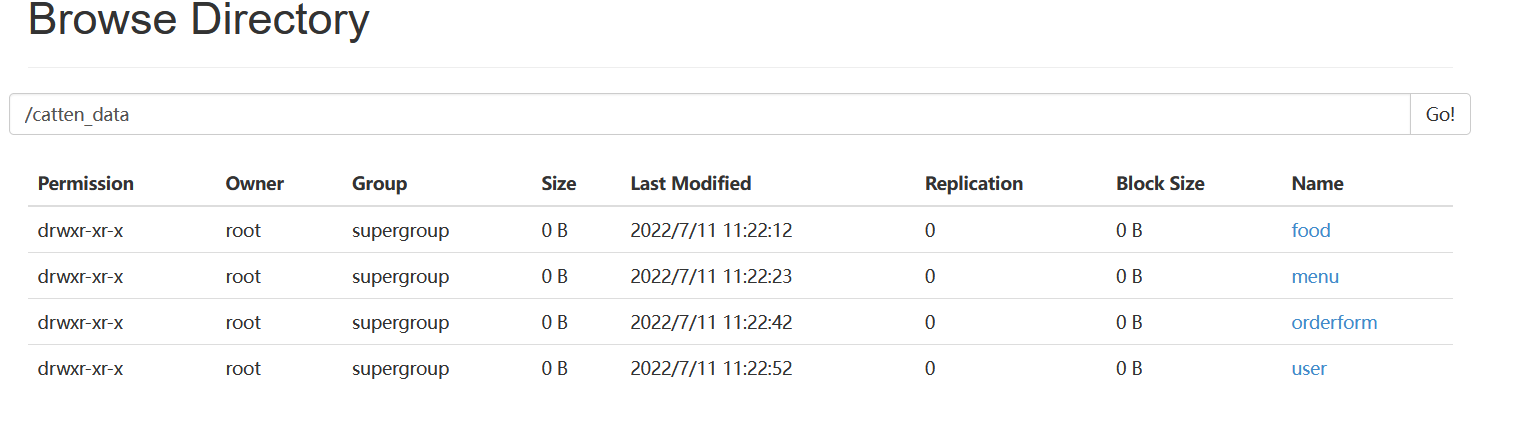
# 分布式数据处理文档

**陈鑫 2022-07-14**

## 1.上传到本地和hadoop上

注：删除了首行数据





## 2.上传到Hive数据仓库中

创建 catten\_data 数据库 ，然后依次创建四张表用来存储

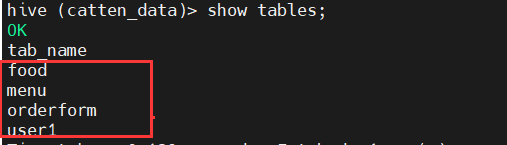
Create external table catten\_data.food(food\_id INT,name STRING,price FLOAT,caloire FLOAT,carbohydrate FLOAT,fat FLOAT,protein FLOAT,vitamin FLOAT) row format delimited fields terminated by ',' stored as textfile location '/catten\_data/food';

Create external table catten\_data.menu(menu\_id INT,food\_id INT,mark INT) row format delimited fields terminated by ',' stored as textfile location '/catten\_data/menu';

Create external table catten\_data.orderform(order\_id INT,user\_id INT,menu\_id INT,total\_money float,time datee) row format delimited fields terminated by ',' stored as textfile location '/catten\_data/orderform';

Create external table catten\_data.orderform(order\_id INT,user\_id INT,menu\_id INT,total\_money float,time date) row format delimited fields terminated by ',' stored as textfile location '/catten\_data/orderform';

Create external table catten\_data.user1(user\_id INT,gender INT,age INT) row format delimited fields terminated by ',' stored as textfile location '/catten\_data/user; (user不行，重名)



## 在Hive中进行数据操作，并存入表

**--> 先创建临时表用来存储最后结果**

#### 3.1 各个菜品每天的价格

create table catten\_data.table1(day string,name string,price float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table1 select \* from menu LEFT JOIN orderform ON menu.menu\_id=orderform.menu\_id LEFT JOIN food ON food.food\_id=menu.food\_id limit 10 where date\_format(orderform.time,'yyyy-HH-dd')='2022-06-12';

#### 3.2某天订单的平均评分

create table catten\_data.table2(day string,point float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table2 select avg(X.result1) as result from(select menu\_id,avg(mark) as result1 from menu where menu\_id in (select menu\_id from orderform where date(time)='2022-06-22') group by menu\_id) X;

#### 3.3每天的总交易额

create table catten\_data.table3(day string,total\_mone float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table3 select date(time)as time, sum(total\_money) from orderform group by date(time);

#### 3.4口碑最好的n道菜

create table catten\_data.table4(food\_id int,food\_name string,point float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table4 select food.food\_id,food.name,avg(mark) as avg\_mark

from menu,food where menu.food\_id=food.food\_id group by food\_id order by avg\_mark desc;

#### 3.5口碑最不好的n道菜

create table catten\_data.table5(food\_id int,food\_name string,point float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table5 select food.food\_id,food.name,avg(mark) as avg\_mark

from menu,food where menu.food\_id=food.food\_id group by food\_id order by avg\_mark asc;

#### 3.6最热门的n道菜

create table catten\_data.table6(food\_id int,food\_name string,num int) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table6 select food.food\_id,food.name,count(\*) as order\_num

from menu,food where menu.food\_id=food.food\_id group by food\_id order by order\_num desc;

#### 3.7最冷门的n道菜

create table catten\_data.table7(food\_id int,food\_name string,num int) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table7 select food.food\_id,food.name,count(\*) as order\_num

from menu,food where menu.food\_id=food.food\_id group by food\_id order by order\_num asc;

#### 3.8每天的各种营养平均值

create table catten\_data.table8(day string,avg\_calorie float,avg\_carbohydrate float, avg\_fat float, avg\_protein float,avg\_vitamin float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table8 select date(time) as time, sum(calorie)/count(distinct user\_id) as avg\_calorie, sum(carbohydrate)/count(distinct user\_id) as avg\_carbohydrate, sum(fat)/count(distinct user\_id) avg\_fat, sum(protein)/count(distinct user\_id) as avg\_protein, sum(vitamin)/count(distinct user\_id) as avg\_vitamin from menu,food,orderform where menu.menu\_id=orderform.menu\_id and food.food\_id= menu.food\_id group by date(time);

#### 3.9每天三个时间段的顾客数量

create table catten\_data.table9(morning\_num int,noon\_num int,evening\_num int) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table9 select date(time) as time,sum(if(hour(time)>=6 and hour(time)<10,1,0)) as morning\_num,sum(if(hour(time)>=10 and hour(time)<14,1,0)) as noon\_num,sum(if(hour(time)>=16 and hour(time)<21,1,0)) as evening\_num from orderform

group by date(time);

#### 3.10 \*某一时间段内的顾客数量

create table catten\_data.table10(morning\_num int,noon\_num int,evening\_num int) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table10 select concat( date\_format(time, '%Y-%m-%d %H:' ) , floor( date\_format(time, '%i' ) /#{n} ) ) AS timeperiod, count( \* ) as num FROM mybatis.orderform WHERE time between #{beginTime} and #{endTime} GROUP BY timeperiod;

#### 3.11获取某天每小时内的平均营养

create table catten\_data.table11(hour string,avg\_protein float,avg\_calorie float,avg\_carbohydrate float,avg\_fat float, avg\_vitamin float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table11 select hour(time) as hour,sum(protein)/count(distinct orderform.order\_id) as avg\_protein,sum(calorie)/count(distinct orderform.order\_id) as avg\_calorie,sum(carbohydrate)/count(distinct orderform.order\_id) as avg\_carbohydrate,sum(fat)/count(distinct orderform.order\_id) as avg\_fat,sum(vitamin)/count(distinct orderform.order\_id) as avg\_vitamin

from menu,orderform,food

where menu.menu\_id=orderform.menu\_id and food.food\_id= menu.food\_id and date(time)='2022-06-01' group by hour(time);

#### 3.12营养达标比例

create table catten\_data.table12(point float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table12 select sum(if(X.total\_calorie>=200 and X.total\_carbohydrate>=30 and X.total\_fat>=10 and X.total\_protein>=20 and total\_vitamin>=0.5,1,0))/count(\*) as p from (select sum(calorie) as total\_calorie,sum(carbohydrate) as total\_carbohydrate,sum(fat) as total\_fat,sum(protein) as total\_protein,sum(vitamin) as total\_vitamin from food,menu where food.food\_id=menu.food\_id group by menu.menu\_id) as X;

#### 3.13男女花钱比例

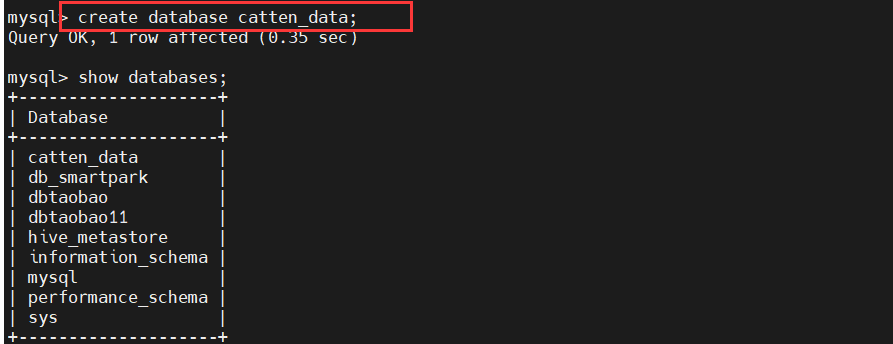
create table catten\_data.table13(sex int,ave\_money float) row format delimited fields terminated by ',' stored as textfile;

Insert overwrite table catten\_data.table13 SELECT gender as sex, avg(total\_money) as averageOrderMoney from mybatis.orderform NATURAL JOIN mybatis.user WHERE gender=sex;

## 传入MySQL

#### 在Mysql数据库中创建表

首先创建catten\_data数据库：



再创建对应表：

1. **各个菜品某天的评分**

create table catten\_data.table1(day varchar(40),name varchar(40),price float)engine=InnoDB default charset=utf8;

1. **某天订单的平均评分**

create table catten\_data.table2(day varchar(40),point float)engine=InnoDB default charset=utf8;

**（3）每天的总交易额**

create table catten\_data.table3(day varchar(40),total\_mone float)engine=InnoDB default charset=utf8;

**（4）口碑最好的n道菜**

create table catten\_data.table4(food\_id int,food\_name varchar(40),point float)engine=InnoDB default charset=utf8;

**（5）口碑最不好的n道菜**

create table catten\_data.table5(food\_id int,food\_name varchar(40),point float)engine=InnoDB default charset=utf8;

**（6）最热门的n道菜**

create table catten\_data.table6(food\_id int,food\_name varchar(40),num int)engine=InnoDB default charset=utf8;

**（7）最冷门的n道菜:**

create table catten\_data.table7(food\_id int,food\_name varchar(40),num int)engine=InnoDB default charset=utf8;

**（8）每天的各种营养平均值**

create table catten\_data.table8(day varchar(40),avg\_calorie float,avg\_carbohydrate float, avg\_fat float, avg\_protein float,avg\_vitamin float)engine=InnoDB default charset=utf8;

**（9）每天三个时间段的顾客数量**

create table catten\_data.table9(morning\_num int,noon\_num int,evening\_num int)engine=InnoDB default charset=utf8;

**（10）某一时间段内的顾客数量**

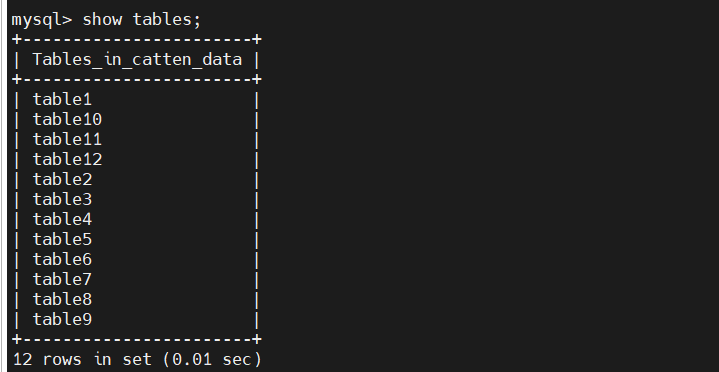
create table catten\_data.table10(morning\_num int,noon\_num int,evening\_num int)engine=InnoDB default charset=utf8;

**（11）获取某天每小时内的平均营养**

create table catten\_data.table11(hour varchar(40),avg\_protein float,avg\_calorie float,avg\_carbohydrate float,avg\_fat float, avg\_vitamin float)engine=InnoDB default charset=utf8;

**（12）营养达标比例**

create table catten\_data.table12(point float)engine=InnoDB default charset=utf8;



#### 传入Mysql数据库对应表

**（1）各个菜品某天的评分**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table1 --export-dir '/user/hive/warehouse/catten\_data.db/table1' --fields-terminated-by ',';

**（2）某天订单的平均评分**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table2 --export-dir '/user/hive/warehouse/catten\_data.db/table2' --fields-terminated-by ',';

1. **各个菜品某天的评分**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table3 --export-dir '/user/hive/warehouse/catten\_data.db/table3' --fields-terminated-by ',';

1. **某天订单的平均评分**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table4 --export-dir '/user/hive/warehouse/catten\_data.db/table4' --fields-terminated-by ',';

**（5）每天的总交易额**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table5 --export-dir '/user/hive/warehouse/catten\_data.db/table5' --fields-terminated-by ',';

**（6）口碑最好的n道菜**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table6 --export-dir '/user/hive/warehouse/catten\_data.db/table6' --fields-terminated-by ',';

**（7）口碑最不好的n道菜**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table7 --export-dir '/user/hive/warehouse/catten\_data.db/table7' --fields-terminated-by ',';

**（8）最热门的n道菜**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table8 --export-dir '/user/hive/warehouse/catten\_data.db/table8' --fields-terminated-by ',';

**（9）最冷门的n道菜:**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table9 --export-dir '/user/hive/warehouse/catten\_data.db/table9' --fields-terminated-by ',';

**（10）每天的各种营养平均值**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table10 --export-dir '/user/hive/warehouse/catten\_data.db/table10' --fields-terminated-by ',';

**（11）每天三个时间段的顾客数量**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table11 --export-dir '/user/hive/warehouse/catten\_data.db/table11' --fields-terminated-by ',';

**（12）某一时间段内的顾客数量**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table12 --export-dir '/user/hive/warehouse/catten\_data.db/table12' --fields-terminated-by ',';

**（11）获取某天每小时内的平均营养**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table1 --export-dir '/user/hive/warehouse/catten\_data.db/table1' --fields-terminated-by ',';

**（12）营养达标比例**

bin/sqoop export --connect jdbc:mysql://localhost:3306/catten\_data --username root --password Admin123. --table table1 --export-dir '/user/hive/warehouse/catten\_data.db/table1' --fields-terminated-by ',';