

17-mapreduce-练习2-用户购买推荐

-----成都尚学堂-mr-zeng-----

-----以下是淘宝用户查看商品的记录(数据量极大-需要存储在hdfs用mapreduce分析)-----

ps : 三个用户 (用户名可能出现重复-》2, 3都是都叫'柳岩')

用户id 用户名 查看的商品

1 张伟 豆浆机
1 张伟 豆浆机
2 柳岩 电视
2 柳岩 电视
3 柳岩 电视
3 柳岩 衣服
1 张伟 耳机
2 柳岩 耳机
1 张伟 豆浆机
1 张伟 电视
2 柳岩 电视
1 张伟 耳机
3 柳岩 衣服
2 柳岩 衣服
3 柳岩 衣服
3 柳岩 衣服
1 张伟 耳机
1 张伟 豆浆机
2 柳岩 豆浆机

需求 : 找到每一个用户 最想购买的商品 (查看次数最多) 进行推荐

得到下面的数据

用户id 用户名 推荐的商品

1 张伟 豆浆机
2 柳岩 电视
3 柳岩 衣服

-----思路分析-----

提示 : 复杂的计算任务-可以分成多次mapreduce任务。

map输出reduce输出建议都text类型 (编码简单-其他类型转换为text即可)

第一步mapreduce得到

1 张伟 豆浆机 4
1 张伟 电视 1
1 张伟 耳机 3
2 柳岩 衣服 1
2 柳岩 耳机 1
2 柳岩 电视 3
3 柳岩 电视 1
3 柳岩 衣服 4

第二步mapreduce得到

1 张伟 豆浆机 4
2 柳岩 电视 3
3 柳岩 衣服 4

-----sql版实现-----

-----表结构和数据-----

```
DROP TABLE IF EXISTS `user_look_product`;
CREATE TABLE `user_look_product` (
  `user_id` varchar(255) default NULL,
  `user_name` varchar(255) default NULL,
  `look_product` varchar(255) default NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

```

INSERT INTO `user_look_product` VALUES ('1', '张伟', '豆浆机');
INSERT INTO `user_look_product` VALUES ('1', '张伟', '豆浆机');
INSERT INTO `user_look_product` VALUES ('2', '柳岩', '电视');
INSERT INTO `user_look_product` VALUES ('2', '柳岩', '电视');
INSERT INTO `user_look_product` VALUES ('3', '柳岩', '电视');
INSERT INTO `user_look_product` VALUES ('3', '柳岩', '衣服');
INSERT INTO `user_look_product` VALUES ('1', '张伟', '耳机');
INSERT INTO `user_look_product` VALUES ('2', '柳岩', '耳机');
INSERT INTO `user_look_product` VALUES ('1', '张伟', '豆浆机');
INSERT INTO `user_look_product` VALUES ('1', '张伟', '电视');
INSERT INTO `user_look_product` VALUES ('2', '柳岩', '电视');
INSERT INTO `user_look_product` VALUES ('1', '张伟', '耳机');
INSERT INTO `user_look_product` VALUES ('3', '柳岩', '衣服');
INSERT INTO `user_look_product` VALUES ('2', '柳岩', '衣服');
INSERT INTO `user_look_product` VALUES ('3', '柳岩', '衣服');
INSERT INTO `user_look_product` VALUES ('3', '柳岩', '衣服');
INSERT INTO `user_look_product` VALUES ('1', '张伟', '耳机');
INSERT INTO `user_look_product` VALUES ('1', '张伟', '豆浆机');

```

-----计算查询代码-----

```

select
u2.user_id,
u2.user_name ,
LEFT(GROUP_CONCAT(u2.look_product),POSITION(',', IN GROUP_CONCAT(u2.look_product))-1) want_product,
MAX(u2.count) look_count
FROM(
select u.user_id,u.user_name,u.look_product,count(*) count from user_look_product u
GROUP BY u.user_id,u.look_product
ORDER BY user_id,count desc
) u2
GROUP BY u2.user_id,u2.user_name;

```