***1. 男性與女性的會員數量與總點播次數***

**input1 = load 'userdemand/userdemand.tsv' USING PigStorage ('\t') as (userid:chararray,time:chararray,artname:chararray,traname:chararray);**

**input2 = load 'userprofile/userprofile.tsv' USING PigStorage ('\t') as (userid:chararray,gender:chararray,age:int,country:chararray,**

**registered:chararray);**

**ri1 = FOREACH input1 GENERATE userid;**

**ri2 = FOREACH input2 GENERATE userid, gender;**

**gender2 = GROUP ri2 BY gender;**

**Cgender2 = FOREACH gender2 GENERATE group as gender,COUNT(ri2);**

**join12 = JOIN ri2 BY userid, ri1 BY userid;**

**gender1 = GROUP join12 BY gender;**

**Cgender1 = FOREACH gender1 GENERATE group as gender,COUNT(join12);**

**join\_gender = JOIN Cgender2 BY gender, Cgender1 BY gender;**

**outputtest1 = FOREACH join\_gender GENERATE $0,$1,$3;**

## ***2. 會員數量排行顯示前五名的國家，和這些國家的總點播次數***

**input1 = load 'userdemand/userdemand.tsv' USING PigStorage ('\t') as (userid:chararray,time:chararray,artname:chararray,traname:chararray);**

**input2 = load 'userprofile/userprofile.tsv' USING PigStorage ('\t') as (userid:chararray,gender:chararray,age:int,country:chararray,**

**registered:chararray);**

**ri1 = FOREACH input1 GENERATE userid;**

**ri2 = FOREACH input2 GENERATE userid, country;**

**country2 = GROUP ri2 BY country;**

**Ccountry2 = FOREACH country2 GENERATE group as country,COUNT(ri2);**

**join12 = JOIN ri2 BY userid, ri1 BY userid;**

**country1 = GROUP join12 BY country;**

**Ccountry1 = FOREACH country1 GENERATE group as country,COUNT(join12);**

**join\_country = JOIN Ccountry2 BY country, Ccountry1 BY country;**

**outputtest2 = FOREACH join\_country GENERATE $0,$1,$3;**

**LOCC = limit (ORDER outputtest2 BY $1 DESC) 5;**

## ***3. 那一天是會員透過我們的音樂網站聆聽音樂最多的一天 ?***

**input1 = load 'userdemand/userdemand.tsv' USING PigStorage ('\t') as (userid:chararray,time:chararray,artname:chararray,traname:chararray);**

**timesubstring = foreach input1 GENERATE userid, SUBSTRING (time, 0, 10) as stime;**

**countlisten = foreach (group timesubstring by stime) GENERATE group,COUNT(timesubstring) as countlisten;**

**top1 = order countlisten by countlisten desc;**

**lof = limit top1 1;**

## ***4. 會員年齡層分布，第一個欄位請依照指示輸出***

**input2 = load 'userprofile/userprofile.tsv' USING PigStorage ('\t') as (userid:chararray,gender:chararray,age:int,country:chararray,**

**registered:chararray);**

**ri2 = FOREACH input2 GENERATE $0,$2;**

**f19 = Filter ri2 by age <= 19;**

**f20\_39 = Filter ri2 by (age > 19) and (age <= 39);**

**f40\_64 = Filter ri2 by (age >= 40) and (age <= 64);**

**f65 = Filter ri2 by age >= 65;**

**ff19 = FOREACH (group f19 all) GENERATE '0-19',COUNT(f19);**

**ff20\_39 = FOREACH (group f20\_39 all) GENERATE '20-39',COUNT(f20\_39);**

**ff40\_64 = FOREACH (group f40\_64 all) GENERATE '40-64',COUNT(f40\_64);**

**ff65 = FOREACH (group f65 all) GENERATE '65--',COUNT(f65);**

**unionf = UNION ff19, ff20\_39, ff40\_64, ff65;**

**ouf = order unionf by $0 asc;**

## 

## 

## 

## ***5. 從過去所有資料中，找出會員在假日最常聽的三首歌曲***

**input1 = load 'userdemand/userdemand.tsv' USING PigStorage ('\t') as (userid:chararray,time:chararray,artname:chararray,traname:chararray);**

**calendar = load 'calendar/calendar.csv' USING PigStorage (',') as (yMd:chararray,dofw:chararray,wnofy:chararray);**

**holiday = filter calendar by (dofw == 'Sa') OR (dofw == 'Su');**

**timesubstring = foreach input1 GENERATE userid, SUBSTRING (time, 0, 10) as stime, artname, traname;**

**jointime = JOIN timesubstring BY stime, holiday BY yMd;**

**newtime = foreach jointime GENERATE $0 as userid, $1 as stime, $2 as artname, $3 as traname;**

**DH = DISTINCT newtime; DHG = group DH by (traname,artname);**

**FG = foreach DHG generate FLATTEN(group),COUNT(DH.userid) as num;**

**OFG = order FG by num desc,group::traname asc; LOFG = limit OFG 3;**