1) ویژگیهای اصلی OLAP که آن را از OLTP متمایز میکند در یک پاراگراف توضیح دهید.
2) فایلهای london12.csv و countries_by_continent.csv داده شده اند. با استفاده از داده های موجود در این دو فایل جدولی به نام london12 در دیتابیس olapdb درmysql ایجاد کنید که دارای ستونهای زیر باشد:

(id, continent, country, gender, agegroup, sport, gold, silver, bronze)

برای پیدا کردن قاره هر کشور میتوانید از فایل countries_by_continent.csv استفاده کنید. توجه داشته باشید که نام کشورها ممکن است در دو فایل csv دقیقا با هم یکسان نباشد. در مورد ستون agegroup به صورت زیر عمل کنید:

Age range	Age Group
Age < 20	A
20 <= Age < 25	B
25 <= Age < 30	C
else	D

3) درmysql برای یک ستون از یک جدول میتوان یک یا چند index ایجاد کرد. مزیت استفاده از index آن است که سرعت جستجو را در اکثر موارد سریعتر میکند. برنامه ای بنویسید که تمامی index های تکی و ترکیبی ممکن را برای ستونهای continent, country, gender, agegroup, sport ایجاد کند. برای نوع index از ساختار BTree استفاده کنید.
4) نمودارهای زیر را رسم کنید:
   a. نمودار ستونی 10 کشور برتر از نظر تعداد شرکت کنندگان
   b. نمودار دایره ای توزیع مدال در قاره های مختلف
   c. نمودار ستونی 10 کشور برتر از نظر نسبت مدال به تعداد شرکت کنندگان در میان تمامی کشورهایی که حداقل 30 شرکت کننده داشته اند. برای هر کشور نسبت تعداد مدالهای طلا، نقره و برنز را در ستونهای جداگانه رسم نمایید.
5) میخواهیم با استفاده از عملیات slice, dice و drilldown و با استفاده از مقادیر continent, agegroup, sport, gender تمامی query های ممکن را تولید کنیم. یک query ترکیبی از عملیات ذکر شده میباشد. تعداد کل query های ممکن را محاسبه کنید و محاسبات خود را توضیح دهید.
   توجه: برای پاسخگویی به این سوال نیازی به در نظر گرفتن اجتماع مقادیر در عملیات dice نمیباشد.
6) برنامه ای بنویسید که تمامی query های عنوان شده در سوال قبل را تولید و از میان آنها query های زیر را گزارش کند:
   a. منجر به دسته بندی از مقادیر شود که بیشترین مقدار انحراف معیار نسبی را برای تعداد مدال بدست آمده داشته باشد. در پاسخگویی به این بخش موارد زیر را در نظر بگیرید:
      i. تنها query هایی را در نظر بگیرید که خروجی آنها یک sub-cube با حداقل 100 رکورد و 20 مدال باشد
      ii. انحراف معیار نسبی به صورت $CV = \sigma/\mu$ محاسبه میشود.
      iii. به عنوان مثال query زیر را در نظر بگیرید:

Slice = “agegroup=C”
Drilldown= “continent, gender”

این query منجر به جدول نتیجه ای به صورت زیر میشود:

Continent	Gender	Medal Count
Africa	F	9
Africa	M	5
Asia	F	63
Asia	M	77
Europe	F	114
Europe	M	158
North America	F	77
North America	M	69
Oceania	F	22
Oceania	M	28
South America	F	19
South America	M	11

این یک query مجاز میباشد چون بدیهتا خروجی آن یک sub-cube با حداقل 100 رکورد میباشد (کلیه ورزشکاران رده سنی C بیش از 100 رکورد میباشد). همچنین مجموع تعداد مدال های این sub-cube بیش از 20 میباشد. انحراف معیار نسبی برای این sub-cube به صورت انحراف معیار اعداد موجود در ستون Medal Count تقسیم بر میانگین اعداد موجود در ستون Medal Count محاسبه میشود که برابر با 0.8770 میباشد.

b. بیشترین نسبت کل مدالهای بدست آمده به تعداد کل رکورد را داشته باشد. در پاسخگویی به این بخش محدودیت زیر را نیز در نظر بگیرید:
   i. نتیجه query باید یک sub-cube باشد که حداقل 10 رکورد را پوشش دهد.
c. کوئری که در حال توسعه ترین sub-cube را نتیجه میدهد. در حال توسعه ترین sub-cube به صورت زیر تعریف میشود:
   i. 90 درصد یا بیشتر مدالها نقره و برنز باشند
   ii. بیشترین تعداد مدال را داشته باشد
d. کوئری که توسعه یافته ترین sub-cube را نتیجه میدهد. توسعه یافته ترین sub-cube به صورت زیر تعریف میشود:
   i. بیش از 50 درصد مدالها طلا باشند
   ii. بیشترین تعداد مدال را داشته باشد

نکات کلی:

- برای انجام تمرین از زبان برنامه نویسی پایتون استفاده نمایید.
- برای پیاده سازی query ها میتوانید از کتابخانه cubes و یا زبان sql استفاده نمایید.
- مهلت تحویل تمرین 4 آبان 98 میباشد.
- به هیچ عنوان تمدید نخواهد شد.
- نمره این تمرین 4 درصد از نمره کل میباشد.
- به ازای هر روز تاخیر 10 درصد از نمره کاسته خواهد شد.
- گزارش خود را در قالب pdf به نام "id_cubes.pdf" به همراه فایلهای py. در قالب یک فایل zip. به نام "id_cubes.zip" آپلود کنید. مثلا اگر شماره دانشجویی شما 9613190 میباشد فایل را 9613190_cubes.zip نامگذاری کنید.