



تمرین کامپیوتری سوم



سیستم‌های عامل - پاییز ۱۳۹۹

گزارش کار

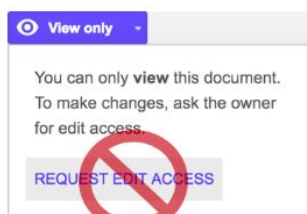
دانشکده مهندسی برق و کامپیوتر

نام و نام خانوادگی: علی حاجی زاده ایدلو

تاریخ: 24 آذر 1399

استاد:

دکتر مهدی کارگهی

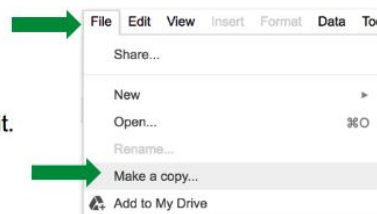


How to use this template:

This is a view-only file and cannot be edited.

Create your own copy of this template to edit.

In the menu, click **File > Make a copy...**



2

مقدمه

3

پیاده‌سازی سری

3

سوال اول

3

سوال دوم

3

جدول اول

3

پیاده‌سازی چندریسه‌ای

3

سوال سوم

4

سوال چهارم

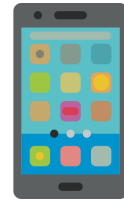
4

سوال پنجم

4

جدول دوم

مقدمه



در این تمرین به تحلیل داده‌هایی که از مشخصات و قیمت فروش گوشی‌های موبایل جمع‌آوری شده‌است پرداخته شده است. در ابتدا برنامه اقدام به خواندن و تجزیه مجموعه داده¹ی ارائه شده می‌کند و آنها را در حافظه خود ذخیره می‌کند. پس از استخراج داده‌ها و ویژگی‌های آنها، برنامه اقدام به نرمال‌سازی² داده‌ها و در نهایت اقدام به تعیین طبقه قیمتی گوشی‌ها می‌کند. این تمرین به دو روش این مسئله پیاده‌سازی شده است که در ادامه گزارش، نتایج حاصل آمده است.



¹ Dataset

² Data Normalization

پیاده‌سازی سری

سوال اول

چرا برای پیاده‌سازی یک برنامه بصورت چندریشه‌ای، بهتر است ابتدا این برنامه بصورت سری پیاده‌سازی شود؟
از آن جایی که دیدگاه در پیاده‌سازی سری ساده‌تر است، بهتر است ابتدا برنامه را به صورت سری پیاده‌سازی کنیم و پس از اطمینان یافتن از درستی برنامه، سعی کنیم مرحله به مرحله بخش‌هایی که شرایط موازی سازی دارند را موازی کنیم.

سوال دوم

با بررسی زمان اجرای بخش‌های مختلف برنامه، ³Hotspot های برنامه را مشخص کنید.

عملیات خواندن از فایل‌ها (io bound)

عملیات مربوط به محاسبه ی min و max هر ستون (cpu bound)

عملیات مربوط به محاسبه class تمامی موبایل‌ها (cpu bound)

جدول اول

زمان‌های اجرای ۶ اجرای متوالی از برنامه و میانگین آن‌ها را بازای ورودی نمونه‌ای که در شرح تمرین آمده است، در جدول زیر بیاورید.

میانگین	اجرای ششم	اجرای پنجم	اجرای چهارم	اجرای سوم	اجرای دوم	اجرای اول
0.032	0.030	0.033	0.034	0.032	0.030	0.032

³ توابعی که در برنامه‌تان بیشترین زمان اجراها را به خود اختصاص می‌دهند.

پیاده‌سازی چندریسه‌ای

سوال سوم

اگر هنگام موازی‌سازی برنامه به زمان اجرای بیشتری نسبت به حالت سری برخورد کنید، چه رویکردهایی را برای کاهش زمان اجرا و استفاده حداکثری از موازی‌سازی پیش می‌گیرید؟

باید میزان وابستگی داده‌ها (که موجب استفاده از lock و در نتیجه افزایش زمان می‌شود) را به حداقل رساند. همچنین باید حواسمان به سربار عملیات مربوط به ساخت ریس و... نیز باشد. تعداد ریس‌ها نیز نباید خیلی زیاد یا خیلی کم باشد.

سوال چهارم

در هنگام پیاده‌سازی این بخش، به چه چالش‌هایی برخورد کردید و بیان کنید که به چه صورت آن‌ها را رفع کردید. وابستگی داده‌ها: داده‌ها را به گونه‌ای به صورت گلوبال در نظر گرفتیم که هر ریس تنها با داده‌های مربوط به خودش کار داشته باشد (با استفاده از id).

دشواری دیباگ پیاده‌سازی موازی: با پیش روی مرحله به مرحله و چاپ اطلاعات مربوط به هر ریس، این مشکل تا حدودی تسهیل پیدا کرد.

سوال پنجم

با توجه به تجربه‌ای که در پیاده‌سازی این تمرین بدست آوردید، به نظر شما در چه مواقعی از قفل⁴ در یک طراحی چندریسه‌ای ضروری است؟ تاثیر استفاده از قفل‌ها را بر روی کارایی⁵ سامانه بیان کنید.

استفاده از قفل موجب افزایش زمان و کاهش کارایی می‌شود. بنابراین حتی الامکان باید از آن‌ها دوری کرد. بعضی مواقع به دلیل وابستگی ذاتی مراحل و داده‌ها به هم، چاره‌ای جز استفاده از قفل‌ها را نداریم. (مثلاً در این پروژه حتماً باید ابتدا مقادیر min و max هر ستون به دست آید تا بتوانیم کار را ادامه دهیم. و برای به دست آوردن این دو مقدار تمامی داده‌ها نیاز است. پس به ناچار مجبوریم مقداری معطل بمانیم.)

⁴ Lock

⁵ Performance

جدول دوم

زمان‌های اجرای ۶ اجرای متوالی از برنامه و میانگین آن‌ها را بازای ورودی نمونه‌ای که در شرح تمرین آمده است، در جدول زیر بیاورید.

میانگین	اجرای ششم	اجرای پنجم	اجرای چهارم	اجرای سوم	اجرای دوم	اجرای اول
0.020	0.018	0.021	0.018	0.022	0.021	0.021

میزان تسریع ($\frac{Serial\ Time}{Parallel\ Time}$) برنامه نسبت به حالت سری را در زیر بیاورید.

میزان تسریع	میانگین زمان اجرای موازی	میانگین زمان اجرای سری
1.6	0.020	0.032