



تمرین کامپیوتری سوم



سیستم‌های عامل - بهار 1400

گزارش کار

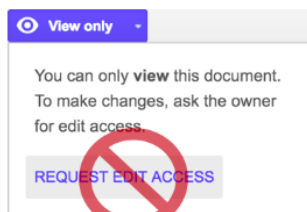
دانشکده مهندسی برق و کامپیوتر

نام و نام خانوادگی: فاطمه داودی نیا

تاریخ: 1400/3/10

استاد:

دکتر مهدی کارگهی

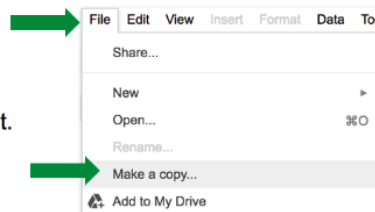


How to use this template:

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 شما به تحلیل داده‌هایی که از مشخصات و قیمت فروش خانه‌ها جمع‌آوری شده‌است پرداخته شده است. در ابتدا برنامه اقدام به خواندن و تجزیه مجموعه داده¹ی ارائه شده کرده و آنها را در حافظه خود ذخیره می‌کند. پس از استخراج داده‌ها و ویژگی‌های آنها، برنامه اقدام به برچسب گذاری ستون قیمت داده‌ها، بدست آوردن میانگین و انحراف معیار داده‌ها و در نهایت اقدام به تعیین رده قیمتی خانه‌ها می‌کند. این تمرین به دو روش این مسئله پیاده‌سازی شده است که در ادامه گزارش، نتایج حاصل آمده است.



¹ Dataset

پیاده‌سازی سری

سوال اول

چرا برای پیاده‌سازی یک برنامه بصورت چندریسه‌ای، بهتر است ابتدا این برنامه بصورت سری پیاده‌سازی شود؟
زیرا اشکال زدایی در حالت سری راحتتر از اشکال زدایی در حالت چند ریه ای است. پس برنامه اول بصورت سری پیاده سازی میشود و پس از اطمینان از درستی منطق و کارکرد برنامه، اقدام به پیاده سازی چند ریه ای میکنیم.

سوال دوم

با بررسی زمان اجرای بخش‌های مختلف برنامه، Hotspot² های برنامه را مشخص کنید.
بخش خواندن از فایلها، قسمت مربوط به محاسبه میانگین هر ستون، بخش پیشبینی قیمت خانه ها و بخش محاسبه تعداد پیشبینی های درست

جدول اول

زمان‌های اجرای ۶ اجرای متوالی از برنامه و میانگین آن‌ها را به ازای 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	0.039	0.022	0.022	0.023	0.021	0.069

² 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 تلاش کرد که وابستگی داده‌ها کم باشد تا استفاده از lock و آزاد کردن آن زیاد از حد رخ ندهد. همچنین تعداد داده‌ها باید به اندازه‌ای زیاد باشد که سربار مربوط به ساختن ریس‌ها در برابر پردازش داده‌ها بصورت سری مقدار ناچیزی باشد. علاوه بر این تعداد ریس‌ها نباید خیلی کم یا خیلی زیاد باشد.

سوال چهارم

در هنگام پیاده‌سازی این بخش، به چه چالش‌هایی برخورد کردید و بیان کنید که به چه صورت آن‌ها را رفع کردید.

مهمترین چالش رفع وابستگی داده‌ها بود که برای این کار از تعریف بعضی از داده‌ها بصورت global کمک گرفتیم.

سوال پنجم

با توجه به تجربه‌ای که در پیاده‌سازی این تمرین بدست آوردید، به نظر شما در چه مواقعی استفاده از مکانیزم‌های همگام‌سازی مانند قفل³ در یک طراحی چندریسه‌ای ضروری است؟ در این پروژه از چه مکانیزم‌هایی استفاده کرده‌اید؟ آیا استفاده از این مکانیزم‌ها می‌تواند اثر مخربی داشته باشد؟

بطور کلی مواقعی که نمیتوان وابستگی بین داده‌ها را از بین برد برای جلوگیری از برخورد و تضاد محاسبات از قفل استفاده میشود.

در این پروژه بدلیل وابستگی ذاتی بعضی مراحل به یکدیگر (برای مثال نیازمندی به میانگین برای محاسبه انحراف معیار) ناگزیر از قفل استفاده کرده ایم تا تضادی در محاسبات رخ ندهد اما بطور کلی باید در نظر داشت که استفاده از قفل موجب افزایش زمان و کاهش کارایی می‌شود.

³ Lock

جدول دوم

زمان‌های اجرای ۶ اجرای متوالی از برنامه و میانگین آن‌ها را به ازای 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
0.047	0.041	0.046	0.058	0.035	0.055	0.052

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

میزان تسریع	میانگین زمان اجرای موازی	میانگین زمان اجرای سری
0.68	0.047	0.032

در اینجا زمان اجرا سریعتر نشده است زیرا تعداد داده ها زیاد نیست و زمان ایجاد ریشه ها تقریباً برابر با همان زمان پردازش داده هاست. (با حجم داده بیشتر تست شد و زمان اجرا در حالت موازی تقریباً 5 برابر بزرگتر از حالت سری بوده است)