



پروژه پایانی

مربیان دوره کارآموزی

میثم رستم‌نژاد (mymrs@aut.ac.ir)

فرزین حیدری (farzin.heidary@gmail.com)

مقدمه

در این پروژه قصد داریم تا مفاهیم آموخته شده در دوره کارآموزی کرونایی رو با هم مرور کنیم و همچنین، به جمع‌بندی اون‌ها بپردازیم. این پروژه از چندین قسمت اصلی تشکیل شده که در این سند قصد داریم تا هر بخش رو به صورت جداگانه توضیح بدیم و ارتباطش رو با مازول‌های دیگه بررسی کنیم. مهلت این پروژه تا ساعت ۲۳:۵۹ تاریخ ۱۳ تیرماه ۱۳۹۹ خواهد بود.

توضیح کلی پروژه

هدف از این پروژه، تحلیل فروش محصولات مختلف در استان‌ها و شهرهای مختلف می‌باشد. ورودی برنامه وارد یک برنامه به زبان C خواهد شد. این برنامه وظیفه دارد تا اطلاعات محصولات مختلف را برای یک شهر جمع‌بندی کند. در واقع، هر شهر دارای فروشگاه‌های مختلفی می‌باشد که اطلاعات فروش محصولات خود را در این سامانه وارد می‌کنند. برنامه نوشته شده به زبان C این اطلاعات را برای هر شهر جمع‌بندی می‌کند و در داخل پایگاه داده ذخیره می‌کند.

در مرحله بعد، برنامه نوشته شده به زبان پایتون این اطلاعات را از پایگاه داده خوانده و تحلیل‌های مختلفی برای آن ایجاد می‌کند که در ادامه به بررسی بیش‌تر این قسمت‌ها خواهیم پرداخت. پیش از آن که به سراغ تفصیل بخش‌های مختلف این پروژه بپردازیم، نکاتی که در این پروژه ملزم به رعایت آن‌ها هستید عبارتند از:

- تمامی قسمت‌های پروژه می‌بایست در داخل یک ریپازیتوری^۱ قرار داده شود.
- از کامیت‌هایی که شامل تغییرات کلی و زیاد در برنامه هستند، پرهیز کنید و سعی کنید که هر ویژگی جدیدی که به سیستم اضافه می‌شود را با کامیت به گیت پروژه منتقل کنید.
- تمامی قسمت‌های پروژه می‌بایست مستندسازی شوند.
- تحلیل‌هایی که روی داده‌ها انجام می‌شود می‌بایست با ذکر دلیل و آوردن داده باشد.
- پیاده‌سازی این سیستم را به صورت کامل در پروژه توضیح دهید.
- نحوه نصب، راه‌اندازی و استفاده از پروژه را در فایل README.md در گیت پروژه قرار دهید.

^۱ repository

۱ ورودی برنامه

ورودی برنامه به شکل فایل خواهد بود. ابتدا با استفاده از یک برنامه بش اسکرپت اطلاعات را از یک سرور که در روزهای پایانی پروژه در اختیارتان قرار خواهد گرفت، دانلود کرده و در مسیر /tmp/final_project/ ذخیره نمایید. فرمت ورودی اطلاعات به شکل زیر خواهد بود:

```
time, province, city, market_id, product_id, price, quantity, has_sold
```

هر خط فایل‌های ورودی به این شکل خواهد بود. با توجه به اینکه ممکن است اطلاعات از منابع مختلفی وارد برنامه شود، در هر لحظه ممکن است تعدادی فایل در دایرکتوری ذکر شده وجود داشته باشد؛ شما باید تمامی این فایل‌ها را بخوانید و آن‌ها را از این مسیر پاک کنید تا اطلاعات زمان بعدی در آن قرار گیرند.

برنامه C

برنامه‌ای که به زبان C نوشته خواهد شد، وظیفه دارد تا اطلاعات مختلفی را درون پایگاه داده قرار دهد. این اطلاعات به صورت زیر خواهد بود (این داده‌ها را در جداول جداگانه قرار دهید):

۱. کل اطلاعات به دست آمده از فایل‌ها را داخل پایگاه داده قرار دهید (اسم جدول: fp_stores_data).
۲. اطلاعات تجمیع شده بر اساس شهر؛ یعنی اطلاعات را از فایل‌ها بخواند، بر اساس زمان تجمیع کند و داخل پایگاه داده قرار دهد. واضح است که مقدار کل محصول (quantity) و مقدار به فروش رفته (has_sold) باید جمع زده شوند. میانگین وزن‌دار قیمت را داخل پایگاه داده قرار دهید (اسم جدول: fp_city_aggregation).
۳. اطلاعات را بر اساس فروشگاه تجمیع کنید که در کل چقدر محصول و به چه میزان به فروش رفته است؛ یعنی تعداد کل محصولاتی که یک فروشگاه فروخته است و کل پولی که از این فروش به دست آورده را داخل پایگاه داده ذخیره کنید (اسم جدول: fp_store_aggregation).

بنابراین، وظیفه اصلی برنامه C، خواندن اطلاعات از فایل، تجمیع آن‌ها بر اساس معیارهای مختلف و ذخیره آن‌ها در پایگاه داده می‌باشد.

قسمت امتیازی: برنامه C را به شکل یک سرویس اجرا کنید و سعی کنید در مراحل مختلف برنامه لاگ تولید کنید (به عنوان مثال، زمانی که به پایگاه داده متصل می‌شوید، لاگ اتصال موفقیت‌آمیز یا ناموفق به پایگاه داده تولید کنید). نحوه تعریف یک سرویس در سیستم عامل لینوکس را در گزارش خود ذکر کنید. می‌توانید از کتابخانه‌های تولید لاگ برای زبان C استفاده کنید که لازم است نحوه نصب و استفاده از آن را در گزارش و گیت ذکر کنید.

پایگاه داده

جداول مربوط به این پروژه که در قسمت قبلی ذکر شدند را در یک پایگاه داده به نام fpdb ذخیره نمایید. در این پروژه، از postgresql استفاده کنید.

برنامه پایتون

این برنامه وظیفه دارد تا هر نیم ساعت اجرا شود و داده‌های موردنیاز برای تحلیل را تولید نماید. در واقع، داده‌هایی که توسط برنامه C جمع شده بودند، در این برنامه برای نیم ساعت گذشته به شکل‌های مختلف جمع خواهند شد که این داده‌ها را از پایگاه داده استخراج خواهد کرد. در این سند یک نمونه از این تحلیل‌ها را توضیح خواهیم داد و پنج تحلیل دیگر را بر عهده خودتان می‌گذاریم که بر اساس داده‌ها و نیازی که به تحلیل سیستم‌های فروش احساس می‌کنید، پنج تحلیل با ذکر مسئله، نحوه استخراج داده‌ها و توضیح کامل اطلاعات به دست آمده از تحلیل در گزارش خود ذکر کنید. مسئله: بیش‌ترین محصولی که در هر استان مورد استقبال مشتریان قرار گرفته است، چیست و به چه میزان (تعداد و قیمت) خریداری شده است.

برای این مسئله و پنج مسئله دیگری که ذکر می‌کنید، نحوه استخراج داده‌ها از پایگاه داده و توضیح کاملی از اطلاعاتی که از این داده‌ها به دست می‌آورد را در گزارش به صورت مفصل بیان نمایید. سعی کنید در تحلیل‌های خود، از تمامی جداول پایگاه داده استفاده نمایید.

شاد و پیروز باشید