

## پروژه ارزیابی فنی گروه مهندسی داده شرکت فناوری ارتباطات و اطلاعات مهیمین

با سلام و احترام

پروژه پیش رو جهت ارزیابی فنی گام دوم تیم مهندسی سامانه راهکارهای مخابراتی مهیمین میباشد که در راستای افزایش شناخت متقابل اولیه قبل از شروع همکاری طراحی شده است.

این ارزیابی شامل 2 بخش میباشد که بخش اول بر روی دانش برنامه نویسی و همچنین آشنایی با ابزارهای Redis، Docker، Postgresql، و Apache Kafka تمرکز دارد و در بخش دوم تلاش شده است تا مسئله ای در حوزه داده با ابزار Apache Spark مورد تحلیل قرار گیرد.

**1.** در بخش اول میبایست یک اپلیکیشن برای ذخیره و دریافت کد کشور شهرها به کمک FastAPI ایجاد نمایید. دیتابیس اصلی این اپلیکیشن Postgresql می باشد و در کنار آن از Redis بعنوان Caching tier برای cache کردن ریکوئست های اخیر استفاده می شود. ریکوئست های دریافت کد کشور نیز در ابزاری به نام Apache Kafka لاگ می شود.

### مرحله اول:

لطفا با استفاده از Docker یک دیتابیس Postgresql ایجاد کرده و سپس با استفاده از زبان برنامه نویسی پایتون داده های موجود در پوشه Cities را با استفاده از API مطرح شده در مرحله دوم در این دیتابیس ذخیره کنید.

### مرحله دوم:

با استفاده از FastAPI یک API پیاده سازی بفرمایید که با دریافت City و CountryCode یک رکورد جدید در دیتابیس Postgresql ایجاد کند. اگر City در دیتابیس موجود بود CountryCode آن را با مقدار داده شده در ریکوئست آپدیت کند.

### مرحله سوم:

با استفاده از Docker یک دیتابیس Redis ایجاد نمایید. از این دیتابیس به عنوان LRU cache ریکوئست ها به API مرحله بعد استفاده خواهد شد.

### مرحله چهارم:

ابزار Apache Kafka را نصب نمایید و یک topic برای ذخیره log اپلیکیشن ایجاد کنید.

## مرحله پنجم:

یک API پیاده سازی بفرمایید که با دریافت City، رشته CountryCode آن را برگرداند. بدین صورت که:

- کلمه دریافتی را در Redis جستجو کند و در صورت وجود آن را برگرداند و cache اپلیکیشن را آپدیت کند.
- در صورتی که کلمه در Redis یافت نشد از PostgreSQL نتیجه را بازگرداند و cache اپلیکیشن را آپدیت کند.
- به ازای هر ریکوئست باید یک message در کافکا جهت log کردن موارد زیر منتشر کند:
  - مدت زمان پاسخ به ریکوئست
  - Cache hit/miss آن ریکوئست
  - درصد cache hit از ابتدای اجرای اپلیکیشن تا آن لحظه
- همچنین Redis باید تنها ۱۰ ریکوئست آخر را cache کند و هر ریکوئست cache شده بعد از گذشتن ۱۰ دقیقه expire شود.

2. در این بخش انتظار میرود با استفاده از **Apache Spark Structured Streaming** گزارش های تحلیلی درخواست شده را ایجاد کنید و خروجی تحلیل های مورد نظر را در MinIO بصورت فایل ذخیره بفرمایید. مراحل ایجاد این گزارش به شرح زیر می باشد:

- داده های موجود در پوشه REF\_SMS شامل یک سورس داده و یک جدول رفرنس جهت تغییر یکی از ستون های موجود در گزارش و ایجاد مقادیر معنی دار میباشد.
  - گزارش های درخواستی زیر را در قالب فایل CSV بدست آورید:
    - 1- میزان درآمد روزانه به تومان چقدر می باشد (میزان درآمد از ستون Debit\_amount\_42 بر حسب میلی ریال بدست می آید)
    - 2- میزان درآمد با ریزدانی 15 دقیقه ای به ازای هر paytype چقدر می باشد (بر حسب تومان)
- برای مثال:

RECORD_DATE	Pay type	revenue
12:00:00 2021/06/22	0	25000
12:00:00 2021/06/22	1	15000
12:15:00 2021/06/22	0	21000

3- حداکثر و حداقل درآمد بر حسب تومان با ریزدانی 15 دقیقه به ازای paytype.

مثلا در تاریخ/زمان 2021/06/22 12:00:00 پیک درآمدی چند بوده است.  
4- میزان درآمد و تعداد رکورد با ریزدانی 15 دقیقه ای به ازای هر paytype چقدر می باشد(بر حسب تومان). با استفاده از فایل رفرنس وضعیت paytype مشخص شود.

برای مثال:

RECORD_DATE	Pay type	Record_Count	revenue
12:00:00 2021/06/22	Prepaid	20	25000
12:00:00 2021/06/22	Postpaid	30	15000
12:15:00 2021/06/22	Prepaid	16	21000

- مدت زمان انجام پروژه تعریف شده 5 روز میباشد و توانایی سرچ و حل مسئله شما مهمترین فاکتور در ارزیابی شماست. لطفا در صورت وجود هرگونه ابهام و پرسش با ایمیل [a.karimian@mohaymen.ir](mailto:a.karimian@mohaymen.ir) و [s.mobasher@mohaymen.ir](mailto:s.mobasher@mohaymen.ir) در ارتباط باشید.