



سند پروژه سحابی نو

سلام

به دوره فشرده سحابی نو خوش اومدین.

توی این دوره قراره سیستمی طراحی کنیم که مجموعه‌ای از داده‌های مرتبط با آمار کلی همچون امتیاز، رتبه‌بندی و همچنین نقدهای چندین Application که بر روی Google Playstore قرار دارند را دریافت کرده و یک سری پردازش روی اون‌ها انجام بده.

در گام اول با زبان پایتون و کتابخانه‌های موجود برای خزش Playstore داده‌ها رو دریافت می‌کنیم، اون‌ها رو برای استفاده خودمون و دیگران در یک صف توزیع شده می‌فرستیم و با داده‌هایی که از این صف خونده میشن یک پایگاه داده رابطه‌ای می‌سازیم.

داده‌های ذخیره شده در پایگاه داده رو برای استفاده‌هایی همچون تحلیل تغییرات آمار و اطلاعات در گذر زمان آماده می‌کنیم و در نهایت با استفاده از ابزار هوش تجاری metabase کمی مصورسازی و تحلیل داده انجام می‌دیم.

هدف پروژه

در نظر داشته باشید این پروژه در وهله‌ی نخست، کارکرد آموزش و یادگیری دارد. اما برای معنابخشی و خلق ارزش این پروژه می‌توان سوالاتی مطرح کرد و به آن‌ها پاسخ داد. این تحلیل‌ها می‌تواند به بیزینس‌ها برای ارزیابی تأثیرات تصمیمات خود کمک کند یا حتی تصمیماتی جدید برای آن‌ها بسازد. از این رو، در طول دوره در پی آن هستیم که پس از توسعه‌ی مؤلفه‌ها و اتصال آن‌ها به یکدیگر، با استفاده از ابزارها و تحلیل داده‌ها بتوانیم ارزشی خلق کنیم یا سؤالی بپرسیم و پاسخ دهیم. در طول دوره برای این خلق ارزش برای شما پیشنهاد خواهیم داشت؛ اما چه بهتر که شما نیز پس از آشنایی با داده، پیشنهادهای خود را نیز مطرح کنید!

این پروژه از زیر سیستم‌های زیر تشکیل شده

- زیر سیستم ذخیره لیست اپلیکیشن‌ها
 - این بخش باید به گونه‌ای توسعه داده شود که با استفاده از یک api بتوان در لیست اپلیکیشن‌هایی که در خزش استفاده می‌شود تغییر ایجاد کرد و لیست اپلیکیشن‌ها برای استفاده خزشگر ذخیره شود
 - Apiهایی که باید توسعه داده شود به شکل زیر است (پیاده‌سازی با fast api انجام شود، پیاده‌سازی با django و مستندسازی با swagger امتیازی است)
 - - یک api برای افزودن اپلیکیشن جدید (Create)
 - یک api برای دریافت لیست اپلیکیشن‌های ثبت شده (Read)
 - یک api برای به‌روزرسانی اطلاعات یک اپلیکیشن از پیش ذخیره شده (Update)

- یک api برای غیرفعال کردن یک اپلیکیشن ذخیره شده (Delete)
- برای ذخیره لیست می‌توان از هر روش ذخیره‌سازی استفاده کرد

• زیر سیستم خزش playstore

- استخراج آمار کلی

برای هر یک از موارد موجود در فهرست application ها قرار است که هر یک ساعت یک‌بار اطلاعات زیر استخراج شود و با کلید مشخص در صف پیام درج گردد. (پیاده‌سازی با redis انجام شود، پیاده‌سازی با کافکا امتیازی است)

▪ حداقل تعداد نصب (minInstalls)

▪ امتیاز (score)

▪ رتبه‌بندی (ratings)

▪ تعداد نقدها (reviews)

▪ آخرین به‌روزرسانی (updated)

▪ نسخه (version)

▪ پشتیبانی از تبلیغ (adSupported)

- استخراج نقدها و آمار مربوط به آن‌ها

برای هر یک از موارد موجود در فهرست application ها قرار است که هر یک ساعت یک‌بار اطلاعات ۱۰۰۰ نقد آخر حاوی ستون‌های زیر استخراج شود و با کلید مشخص در صف پیام (پیاده‌سازی با redis) درج کند.

▪ شناسه نقد (reviewId)

▪ زمان (at)

▪ نام کاربر (userName)

▪ تعداد موافق‌ها (thumbsUpCount)

▪ امتیاز (score)

▪ متن (content)

توجه شود که داده‌ها باید با توجه به فیلد reviewId در پایگاه داده پس از هر بار دریافت اطلاعات جدید به‌روزرسانی شوند.

نکته: ممکن است در دریافت اطلاعات از playstore با مشکلاتی همچون ایجاد محدودیت در نرخ دریافت اطلاعات مواجه شوید که باید راهکارهای مواجهه با این مشکلات را پیدا کرده و موانع را رفع کنید.

• زیرسیستم ذخیره‌کردن داده‌ها

این مؤلفه به طور مستمر داده‌های تولید شده مربوط به هر Application که توسط خزشگر در صف پیام درج شده‌اند را خوانده و در پایگاه داده postgresql با timestamp زمان خزش درج می‌کند

- تحلیل داده
 ۱. با استفاده از ابزار هوش تجاری metabase که به پایگاه داده postgres متصل شده، نمودارهای مناسب برای نمایش و تحلیل داده‌های زیر ترسیم شوند.
 ۲. نمایش نمودار تغییرات امتیاز برنامه‌ها در هر دسته اپلیکیشن بر اساس امتیاز کلی و بر اساس امتیاز داخل نقدها
 ۳. تحلیل صعود یا نزول تعداد نصب‌ها و بررسی Trendها
 ۴. بالاترین نقدهای مؤثر به ازای هر برنامه
 ۵. در این بخش سعی کنید سؤالاتی اضافه کنید که می‌توان با توجه به داده پرسید و در داده به دنبال پاسخ آن‌ها گشت

نکات نهایی

- دقت کنید که یکی از اهداف انجام این پروژه آشنایی شما با ابزارهایی مثل redis, postgresql و metabase و همچنین محک‌زدن توانایی شما در برنامه‌نویسی پایتون و معماری نرم‌افزار است؛ اما در طراحی جدول‌های پایگاه داده و فرمت داده‌ها در صف پیام و به‌طور کلی در جزئیات پیاده‌سازی اختیار با شماست.
- برای کدهای خود unit test بنویسید.
- داگری کردن تمام مؤلفه‌های پروژه الزامی است.
- ممکن است با توجه به مهلت تحویل، وقت کم بیاورید؛ بنابراین سعی کنید زمان خود را مدیریت کرده و در یک نقطه گیر نکنید.
- بیشتر ابهاماتی که با خواندن این سند برای شما ایجاد شده عمدی هستند، هر چند که طبق برنامه اعلام شده می‌تونید با mentor خودتون ارتباط بگیرید و با اون‌ها مشورت کنید.
- لازم است برای نصب و راه‌اندازی پروژه، به وسیله bash scripting خودکارسازی انجام گرفته و نحوه‌ی استفاده از آن به عنوان راهنمای راه‌اندازی تمامی قسمت‌ها مستند شود. (استفاده از Ansible و docker compose اختیاری است)
- امتیاز بیشتر:
 - شما می‌توانید با استفاده از کتابخانه‌های تحلیل احساس موجود، پیام‌ها را به سه دسته مثبت، خنثی و منفی تقسیم‌بندی کنید و نمودارهای مربوطه را استخراج کنید.



پیوست فهرست Applicationها:

(می‌توانید به دسته‌های این لیست اضافه کنید و همچنین می‌توانید به برنامه‌های هر دسته نیز بیافزایید اما مواردی که ذکر شده‌اند لازم هستند.)

• پیام‌رسان

- telegram
- whatsapp

• اپراتورها

- Myirancell
- Mymci
- MyRightel

• مالی

- seke
- ۷۲۴
- top

• تماشای ویدئو

- Namava
- Telewebion
- tamashakhane

• بازی کلمه‌ای

- fandogh
- Amirza
- hadsekalamat

• شبکه‌های اجتماعی

- instagram
- facebook
- tiktok