



# ساختمان داده‌ها و الگوریتم‌ها

نیم‌سال اول ۱۳۹۹ - ۱۴۰۰

مدرس: مسعود صدیقین

## فاز اول چالش

### مقدمه . جریان داده

«انتقال و تحلیل داده» لزوماً به معنای ذخیره‌سازی آن نیست. کم پیش نمی‌آید که حجم داده‌ای که با آن سر و کار دارید اصلاً با سخت‌افزار فیزیکی شما نخواند، اما لازم باشد که شما کاری روی آن انجام دهید؛ برای مثال همین سرورهای skyroom که کلاس شما روی آن برقرار می‌شود اگر می‌خواستند همه‌ی کلاس‌ها را ذخیره کنند باید هزینه‌ی زیادی روی سیستم‌های ذخیره اطلاعاتی انجام می‌دادند، تا جایی که حتی سرعت انتقال داده به شما - به دلیل عملیات‌های انجام شده - و پهنای باند شما نیز ممکن بود کاهش بیابد. با این حال، این سامانه کلاس‌ها را به شما انتقال می‌دهد و بعضاً امکاناتی برای بهبود تجربه‌ی کلاسی نیز در اختیار شما قرار می‌دهد.

در این چالش، شما با یک جریان داده مواجهید. با توجه به محدودیت‌های زمان اجرا، حجم این جریان خیلی زیاد نیست. اما حافظه‌ای که در اختیار دارید بسیار محدود است؛ و با حداقل حافظه باید تحلیل خود را انجام دهید.

### صورت . فرکانس

در این چالش، هدف شما پیدا کردن اعدادی است که فرکانسشان در یک جریان داده از یک  $k$  ثابت بیش‌تر باشد، یعنی اگر طول این جریان  $n$  باشد هر کدام حداقل  $nk$  باشد. اعداد شما طبیعی هستند هم‌چنین، شما اجازه دوبار عبور از داده‌ها را دارید، یعنی یک بار می‌توانید از داده‌ها عبور کنید، تحلیل کنید و سپس یک بار دیگر عبور کنید تا تحلیل خود را تکمیل کنید.

### ورودی

در خط اول ورودی،  $n$  و  $k$ ، طول جریان و فرکانس مورد نظر می‌آید.

در هر یک از  $n$  خط بعدی، یک عدد طبیعی مانند  $a$  می‌آید. در هر یک از  $n$  خط بعدی، باز هم یک عدد طبیعی می‌آید. تضمین می‌شود که عددی که در خط  $i$ ام می‌آید با عددی که در خط  $i + n$ ام می‌آید برابر است. ( $2 \leq i \leq n + 1$ )

$$a, n \leq 5 * 10^7$$

در این سوال، محدودیت زمانی شما ۱۰ ثانیه و محدودیت حافظه‌ی شما ۳۲KB است.

### خروجی .

تمامی اعدادی که فرکانس آن‌ها از  $k$  بیش‌تر است را در یک خط و با فاصله جدا کنید. ترتیب آن‌ها اهمیتی ندارد.

## نمره دهی .

شما به ازای هر تست کیس، تعدادی پنالتی دریافت می کنید که برابر با فاصله شما از جواب واقعی مسئله است. یعنی به ازای هر  $FalsePositive$  (جواب هایی که خروجی نمی دهید) و هر  $TrueNegative$  (جواب های غلطی که خروجی می دهید) یک پنالتی دریافت می کنید.

پنالتی های این فاز، ضریبی برابر نیمی از ضریب پنالتی های فاز دوم دارند.

نمره نهایی شما، از روی رتبه شما در تعداد پنالتی ها حساب می شود:

- اگر  $x$  جمع پنالتی های نفر دهم در دو فاز باشد، افرادی که حداکثر  $1/0.5x$  پنالتی داشته باشند، نمره کامل می گیرند.
- اگر  $y$  تعداد پنالتی های نفر سی ام در دو فاز باشد، افرادی که حداکثر  $1/0.5y$  پنالتی داشته باشند، دو سوم نمره را کسب می کنند
- همه افراد دیگری که در چالش مشارکت کرده باشند یک سوم نمره را کسب می کنند.
- به نفرات اول تا سوم نمره امتیازی تعلق می گیرد.