

به نام خدا



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)



DEPARTMENT OF COMPUTER  
ENGINEERING AND IT

دانشگاه صنعتی امیرکبیر  
دانشکده‌ی مهندسی کامپیوتر

تمرین سری سوم مبانی رایانش ابری

دکتر سید احمد جوادی

طرح سوال:

احمدی، لوكزاده

۱۴۰۱ دی

## توضیحات مهم:

- #StudentId\_HW3.zip حاوی پاسخ خود را در یک فایل فشرده با فرمت zip ذخیره کرده و با عنوان notebook بارگذاری نمایید (به عنوان مثال 99131000\_HW3.zip).
- مهلت انجام تمرین مطابق با دلاین در نظر گرفته شده در سیستم کورسز می‌باشد.
  - در این تمرین نیاز به فایل گزارش نمی‌باشد و تنها پاسخ را در فایل notebook وارد کنید.
  - مطابق قوانین دانشگاه هرگونه کپی برداری ممنوع می‌باشد و در صورت مشاهده نمره‌ی هر دو طرف صفر در نظر گرفته می‌شود.
  - در صورت داشتن هرگونه ابهام می‌توانید با تدریسیاران ارتباط برقرار کنید.

## آپاچی اسپارک

آپاچی اسپارک یک فریمورک متن باز است که در سال ۲۰۰۹ توانست با بهینه سازی ۱۰۰ برابر نسبت به پلتفرم روز آن زمان Hadoop جایگاه بالایی در میان علاقه‌مندان به مباحثه BigData باز کند. دلیل قدرت این فریمورک وجود هسته مرکزی با یک موتور پردازش توزیع یافته کلان داده است که می‌تواند به خوبی مقیاس بندی شود. همچنین یکی دیگر از مواردی که باعث افزایش کارایی اسپارک شده استفاده از مکانیسم حافظه نهان (cache) هنگام استفاده از داده‌های است که قرار است دوباره در برنامه استفاده شود. این کار باعث کاهش سربار ناشی از خواندن و نوشتن از دیسک می‌شود. با استفاده از مکانیسم حافظه نهان در اسپارک داده‌ها یک بار از دیسک خوانده می‌شوند و در حافظه اصلی کش می‌شوند و عملیات‌های متفاوت بر بروی آن اجرا می‌شود. در نتیجه استفاده از این روش نیز باعث کاهش چشمگیر سربار ناشی از ارتباط با دیسک در برنامه‌ها و بهبود کارایی می‌شود.

با توجه به قدرت این پلتفرم یکی از پتانسیل‌های استفاده از آن مبحث استریم دیتا می‌باشد. امروزه شرکت‌های تکنولوژی زیادی از آن با مقاصد تجاری استفاده می‌کنند. از طرف دیگر ویژگی ساده بودن یادگیری آن موجب شده تا این فریمورک به عنوان یک مهارت می‌باشد در رزومه توسعه دهنده‌گان قرار داشته باشد. در این تمرین تلاش شده تا از توانایی‌های ابتدایی این پلتفرم استفاده شود.



تاریخ سرزمین میانه به سه دوره زمانی تقسیم می‌شود: سال‌های لامپ‌ها، سال‌های درخت‌ها و سال‌های خورشید. خود سال‌های خورشید نیز به ۴ عصر تقسیم شده است.

### سال‌های لامپ‌ها(The Years of the Lamps)

این دوره با خلق آردا توسط والار به وجود آمد. پیش از این که خورشید به وجود آید، آردا گرما و نور حیاتی اش را از دو لامپ بزرگ دریافت می‌کرد: ایلوین (Iluin) و اورمال (Ormal). از این لامپ‌ها، اولین گیاهان و درختان توانستند رشد کنند و حیوان‌ها نیز در آردا به وجود آمدند. به این دوره زمانی بهار آردا هم می‌گویند. پایان سال‌های لامپ‌ها با تابود شدن دو لامپ توسط ملکور رقم خورد تا بهار آردا به انتهای بررسد. در این مقطع، آردا فقط توسط نور ستارگان روشن می‌شد و اکثر موجودات زنده به خواب یاوانا فرستاده شدند.

### سال‌های درخت‌ها(The Years of the Trees)

پس از نابودی ایلوین و اورمال، یاوانا دو درخت با نام‌های تلپریون (Telperion) و لارلین (Laurelin) به وجود آورد. این درخت‌ها به آمان، جایی که والار در آن زندگی می‌کردند، روشنایی می‌دادند اما نوری به سرزمین میانه نمی‌رسید. در انتهای سال‌های درخت‌ها، الفها در سرزمین میانه بیدار شدند. این اتفاق شروع دوران نخست بود که تا سال‌های خورشید ادامه داشت. در طی این دوره، ملکور در تالارهای ماندوس (Halls of Mandos) زندانی شده بود. سال‌های درخت‌ها پس از آزاد شدن ملکور توسط مانوه از زندان خود به پایان رسید. آنگولیانت دو درخت را نابود کرد و ملکور به سرزمین میانه گریخت.

### سال‌های خورشید(The Years of the Sun)

سال‌های خورشید وقتی آغاز شد که خورشید و ماه برای روشن کردن آردا خلق شدند. این دوره به ۴ عصر تقسیم می‌شود.  
دوران نخست

دوران نخست که با نام روزهای کهن هم شناخته می‌شود زمان متولد شدن نژادها و موجودات آردا بود. در این دوره زمانی نام ملکور به مورگوث تغییر پیدا کرد. پس از چندین نبرد بزرگ، سرزمین میانه برای مدتی طولانی به صلح و آرامش رفت که نام صلح طولانی را به خود گرفت. سپس مورگوث محاصره انگبند (Siege of Angband) را شکست و چندین امپاطوری الف سقوط

کردند. در نهایت مورگوث گیر افتاد و از آردا بیرون رانده شد. پایان دوران نخست زمانی بود که بلریاند (Beleriand) در جنگ خشم (War of Wrath) نابود شد.

### دوران دوم

دوران دوم طولانی‌ترین عصر محسوب می‌شود. این دوره زمانی آغازگر قدرت گرفتن سائورون بود. پس از شکست اربابش مورگوث، سائورون توانست از شرق به سرزمین میانه وارد شود. او قلعه تاریکش را با نام باراد دور (Barad-dûr) نزدیک کوهستان نابودی (Mount Doom) ساخت. همچنین در طی دوران دوم بود که حلقه‌های قدرت به وجود آمدند. الفها با دانشی که از سائورون به دست آورده بودند، ۱۹ حلقه ساختند. بیستمین حلقه توسط خود سائورون در آتش کوه نابودی به وجود آمد و همان حلقه یگانه برای فرمانروایی بر بقیه حلقه‌ها بود.

تقریباً در میانه دوران دوم، جنگ الفها و سائورون (War of the Elves and Sauron) شروع شد. در این جنگ سائورون علیه الفها در اریادور (Eriador) و متحданشان جنگید. پایان دوران دوم با شکست سائورون رقم خورد. ایسیلدور (Isildur) توانست حلقه قدرت را از دست ارباب تاریک بیرون بکشد و حالت فیزیکی او را از بین ببرد. در نهایت، حلقه یگانه برای بیش از ۲۰۰۰ سال گم شد.

### دوران سوم

در این دور، سائورون به موردور بازگشت و جادوگران به سرزمین میانه آمدند تا به مردم آزاد در جنگ علیه بازگشت ارباب تاریک کمک کنند. گاندalf با شک داشتن نسبت به این که حلقه یگانه در شایر در دست بیلو است، شروع به سر زدن مداوم به شایر کرد. وقتی حلقه یگانه به فرودو رسید، گاندalf او را به مأموریت نابود کردن این حلقه یک بار و برای همیشه فرستاد. این سرزمین میانه یا به انگلیسی میدل ارث (Middle-earth) که با نام‌های انور (Ennor) و اندور (Endor) هم شناخته می‌شود، خشکی و قاره اصلی آردا (Arda) است. سرزمین‌های مهم در سرزمین میانه موردور، شایر، روہان، روانیون، گاندور، آیرون هیلز می‌باشد.

شما در این دوره به عنوان یک تحلیل‌گر الف (Elf) در اواخر نبردها هستید. در این نبرد نژادهای اورک، نزگل، انسان، دورف، گابلین، الف (Elf) و هابیت‌ها گاها با یکدیگر متحد و برعلیه هم در حال نبرد هستند. فرمانده برای ارزیابی و تصمیم‌گیری نیاز دارد تا پیوسته گزارشی از تحلیل‌های جنگی داشته باشد. از شما خواسته شده تا با دیتابستی که درباره جرئیات نبردها ایجاد شده، گزارشی را برای او ایجاد کنید.

دیتابست اولیه در دو فایل فشرده شده json در اختیار شما قرارداده شده است که دیتابست اول شامل لیست نبردها و دیتابست دوم شامل سلاح‌هایی استفاده شده در این نبردها می‌باشد. دیتابست را لود کرده و به سوالاتی که در فایل jupyter notebook در پوشش shared\_dir از شما خواسته شده پاسخ دهید.

جزئیات فایل داکر کامپوز:

برای این تمرین یک فایل داکر کامپوز آماده شده است که برای شما پیش‌نیازهای زیرساختی تمرین را فراهم می‌کند. از جمله ۱ کلاستر hdfs برای ذخیره سازی فایل در فضای ابری به صورت توزیع شده که از یک کانتینر namenode و دو کانتینر datanode تشکیل شده است. ۲ کلاستر اسپارک که از یک مستر و دو ورکر تشکیل شده است که برای پردازش فایل‌ها به صورت موازی و توزیع شده از آن استفاده می‌شود. کلاستر اسپارک بالا امده از نوع مود standalone می‌باشد همچنین پکیج‌های کورد نیاز تمرین روی آنها از پیش نصب شده است. و ۳ یک فایل جوپیتر نوتبوک که در نتورک داکری همه این‌ها قرار گرفته تا به راحتی متصل باشد و این فایل باید توسط شما به زبان پایتون کامل گردد.

همچنین همه این کانتینرها در مسیر /shared\_dir با هم مشترک هستند و تغییرات در این مسیر باعث تغییرات در همه است. فایل تمرین notebook نیز در این مسیر وجود دارد همچنین فایلهای داده شده که باید از حالت فشرده خارج شوند و روی hdfs اضافه شوند نیز روی این مسیر هستند.

نکته:

توجه فرمایید که داکر کامپوز برای این تمرین یک interface مجازی شبکه با gateway با آیپی ۱۰.۲۰.۳۰.۱ می‌سازد و برای بقیه کانتینرها هم در ساختن مشخصی در همین بازه IP تعریف شده است که میتوانید در فایل کامپوز را مشاهده فرمایید.

برای کسانی که داکر روی ماشین مجازی نصب می‌کنند (برای مثال داکر دسکتاپ mac) این IP‌ها از بیرون فضای داکر قابل دسترسی نیستند اما چون پورت‌های مهم این کانتینرها forward شده است میتوانند به تمرین بپردازند.

برای بالا اوردن کانتینرها مورد نیاز و ادامه تمرین لطفاً فایل README.md را نیز ملاحظه فرمایید.