

به نام خدا

پروژه شماره ۱ سیستم عامل

در این پروژه، از شما خواسته شده تا با استفاده از تابع **exec** در زبان جاوا و ایجاد پراسسهای متعدد، یک چارچوب برای محاسبات داده حجیم (**big-data**) پیاده سازی کنید. توضیح مربوط به داده حجیم و محاسبات آن در کلاس قبلا داده شده است، ولی برای بررسی بیشتر میتوانید به اینترنت رجوع کنید.

چارچوبهای داده حجیم محاسبات خود را بر روی کامپیوترهای متعدد و به صورت موازی (از طریق مکانیزمهایی مثل Map-Reduce) انجام می دهند و داده های ورودی نیز در فایل های متعددی توزیع شده اند، ولی در این پروژه محاسبات بر روی یک ماشین (منتها با ایجاد پراسسهای متعدد) صورت می گیرد و داده ورودی نیز از یک فایل متنی دریافت می شود. این فایل حاوی تعداد زیادی سطر می باشد که در هر سطر یک عدد صحیح نوشته شده است. اولین سطر این فایل نشان دهنده تعداد اعداد موجود در فایل می باشد.

هدف کلی برنامه محاسبه و نمایش میانگین اعداد و همینطور بزرگترین عدد نوشته شده در فایل می باشد. ابتدای اجرای برنامه، از طریق اجرای دستور شروع، آدرس فایل ورودی و تعداد پراسسهای کارگر از کاربر پرسیده می شوند (مثلا دستور زیر وارد می شود: `start 10 D:\input.txt`). به محض اجرای دستور شروع، پراسس اصلی همه پراسسهای کارگر را ایجاد می کند. هر پراسس مسئولیت پردازش بخشی از فایل را برعهده می گیرد و محاسبات مربوط به استخراج میانگین و بزرگترین نمره را تنها بر روی همان بخش انجام می دهد. به بیان دقیقتر، تعداد سطرهای فایل بر تعداد پراسسهای کارگر تقسیم می شود و هر پراسس فقط و فقط همان تعداد سطر را به صورت یکجا از فایل ورودی خوانده و سپس بر روی آن پردازش مربوط به محاسبه میانگین و حداکثر عدد را انجام می دهد. برای مثال، اگر تعداد اعداد موجود در فایل ۱۰۰۰ و تعداد پراسسها ۱۰ عدد باشد، آنگاه پراسس اول ۱۰۰ سطر اول را رسیدگی می کند، پراسس دوم ۱۰۰ سطر دوم و به همین ترتیب، پراسس دهم ۱۰۰ سطر آخر را. بعد از آنکه هر پراسس کار محاسبه میانگین و بزرگترین عدد را بر روی بخش خود انجام داد، نتیجه را از طریق نوشتن در فایل به اطلاع پراسس اصلی (که والد همه پراسسهای کارگر می باشد) می رساند. پراسس اصلی به محض اتمام پردازش متعلق به هر یک از پراسسهای کارگر نتایج را از فایل مربوطه خوانده و محاسبات مربوط به میانگین کل و بزرگترین عدد را انجام می دهد تا در انتها که همه پراسسهای کارگر کار خود را تمام کرده باشند، با اتمام کار، نتیجه مربوط به میانگین و بزرگترین عدد به همراه زمان لازم برای محاسبه آنها به کاربر نشان داده می شود. اگر در هر یک از بخشهای فایل، اشکالی در فایل ورودی وجود داشته باشد، پیغام خطا از طریق پراسس اصلی به کاربر ارسال می شود. در نهایت نیز نموداری از زمان لازم برای بدست آوردن خروجی هنگامی که تعداد پراسسهای کارگر را ۱ تا ۲۰ افزایش می دهیم ارائه دهید (فایل ورودی برای نمایش نمودار در گروه آپلود خواهد شد).