

تحلیل اطلاعات بورس با استفاده از Stream Processing

قصد داریم در این پروژه با استفاده از Stream Processing و با استفاده از ابزار Apache Storm داده‌های بورسی را تحلیل نماییم. با توجه به اینکه به اطلاعات بورس دسترسی نداریم بنابراین از داده‌های تصادفی برای تحلیل استفاده میکنیم.

در این پروژه تعدادی نماد معاملاتی داریم. فرض میکنیم نمادهای معاملاتی ما از شناسه ۱ تا ۴۰ میباشند. قیمت هر سهم برابر با شناسه آن ضربدر ۱۰۰ میباشد. برای مثال قیمت نماد معاملاتی با شناسه ۳۷ برابر ۳۷۰۰ ریال میباشد. بازه قیمتی تا ۱ درصد قابل نوسان است. بنابراین نماد معاملاتی با شناسه ۳۷ در بازه ۳۶۶۳ ریال تا ۳۷۳۷ ریال قابل معامله میباشد. فرض کنید ۱۰۰۰۰ کد بورسی در بازه ۱ تا ۱۰۰۰۰ داریم.

فعالیت بورس ایران از ساعت ۸:۳۰ تا ۱۲:۳۰ میباشد. وظیفه شما این است که در ابتدا برنامه ای بنویسید که داده‌های خرید و فروش را به صورت تصادفی ایجاد نماید. هر کد بورسی باید حداقل در هر ۱ دقیقه یک سفارش خرید یا سفارش فروش داشته باشد.

رکوردهای خرید یا فروش به صورت زیر میتوانند باشند.

زمان، کدمعاملاتی، فروش یا خرید(فروش ۰ خرید ۱)، شناسه نماد معاملاتی، قیمت وارد شده، تعداد سهم وارد شده.

برای مثال یک رکورد ورودی میتواند به صورت زیر باشد.

۹۷، ۲۹۱۰، ۲۹، ۱، ۱۶۲، ۸:۳۲:۲۸

بنابراین کدمعاملاتی با شناسه ۱۶۲ در ساعت ۸:۳۲:۲۸ درخواست خرید ۹۷ سهم از سهم با شناسه ۲۹ با قیمت ۲۹۱۰ ریال را داده است.

نکته: تعداد سهام وارد شده را بین ۱ تا ۱۰۰ سهم در نظر بگیرید.

به دلیل اینکه هر کد بورسی حداقل باید در هر ۱ دقیقه یک خرید یا فروش را ثبت کند و تعداد کل کدهای بورسی ۱۰۰۰۰ است و بازه فعالیت ۴ ساعت یا ۲۴۰ دقیقه میباشد، بنابراین شما باید حداقل به تعداد ۲۴۰۰۰۰ رکورد تصادفی ایجاد نمایید.

نکته: حتما رکوردهای تصادفی را در به صورت یک فایل CSV ذخیره نمایید و برای صحت سنجی کد خود از آن استفاده نمایید. بنابراین مثال بالا در فرمت CSV به صورت زیر است.

8:32:28,162,1,29,2910,97

بعد از اینکه فایل با داده های تصادفی را ایجاد نمودید با استفاده از ابزار Storm این داده‌ها را پردازش نمایید. وظیفه شما این است که بعد از پردازش اطلاعات توسط ابزار Storm به ازای هر نماد تعداد سهام‌های معامله شده نمایش داده شود. در نهایت خروجی شما باید به صورت زیر باشد.

1,253

2,0

...

40,10

بنابراین خروجی شما شامل ۴۰ سطر است که در ابتدا شناسه نماد معاملاتی و بعد از آن تعداد سهام معامله شده آمده است.