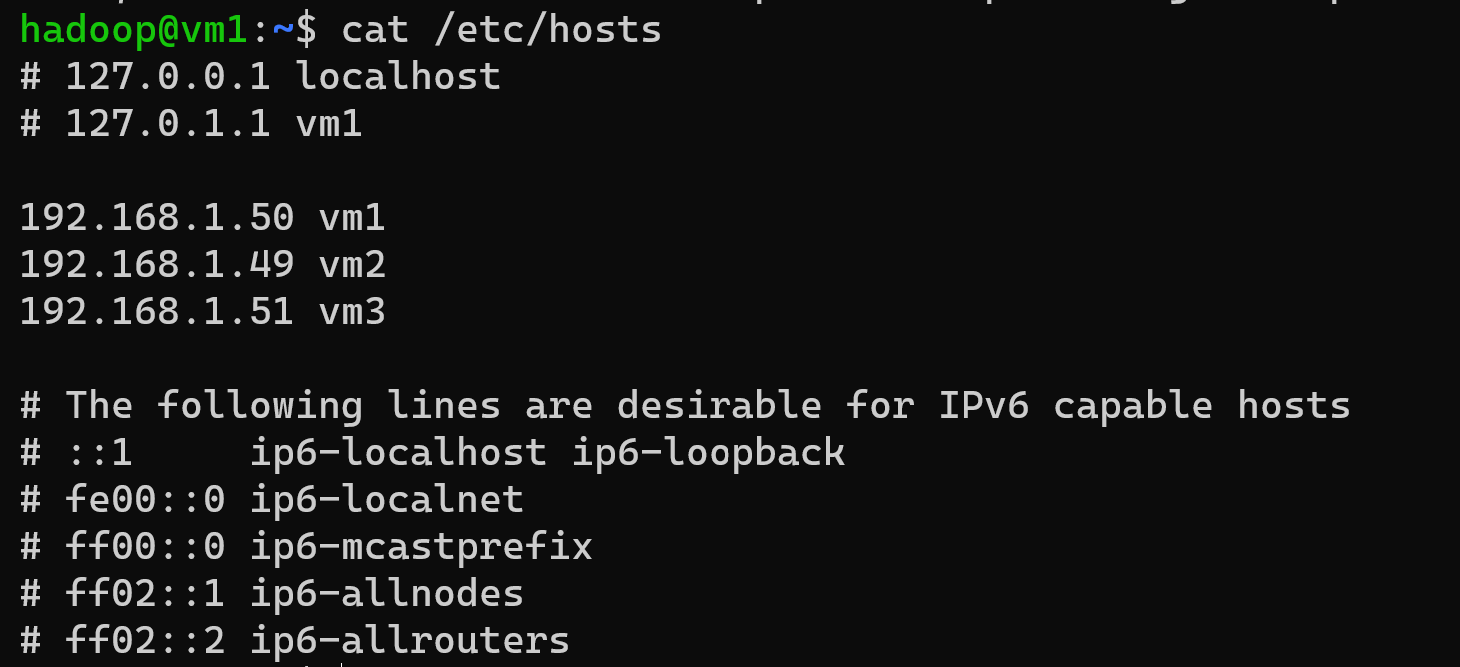
گزارش این تمرین به شرح زیر است:

## نصب و راه‌اندازی Hadoop cluster

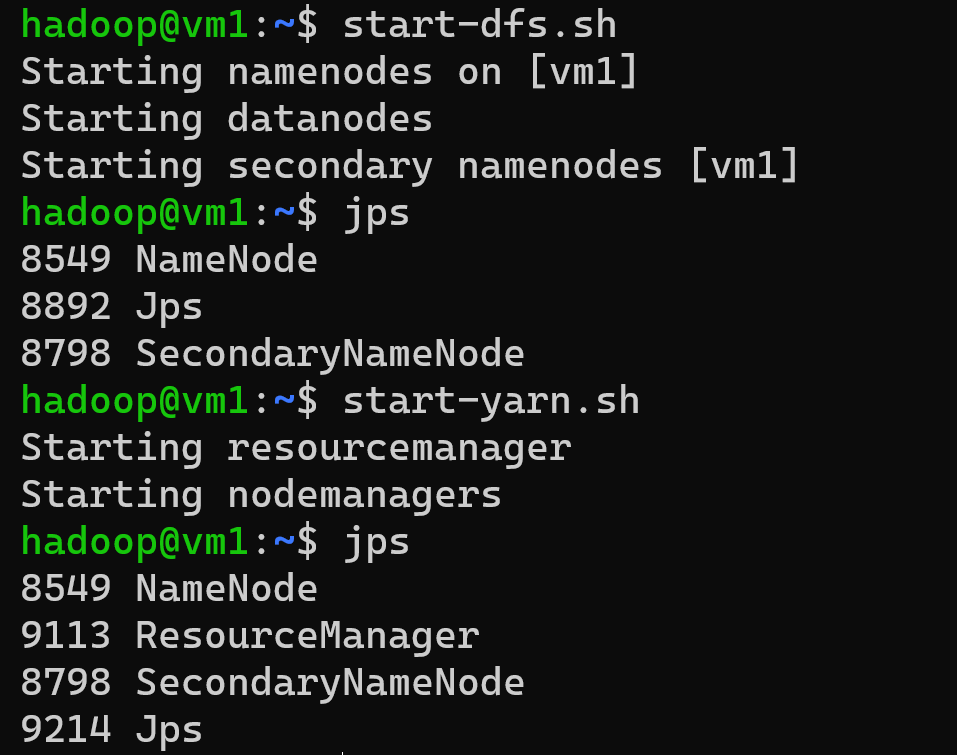
ابتدا یک ماشین ubuntu server ایجاد شد و در آن تلاش شد که دانلود و نصب Hadoop و java 8 صورت پذیرد تا دو ماشین را بتوانیم با کمک همین ماشین از طریق clone کردن بدست آوریم.

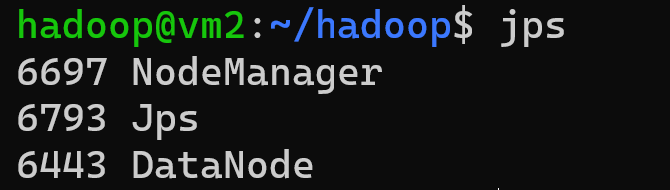
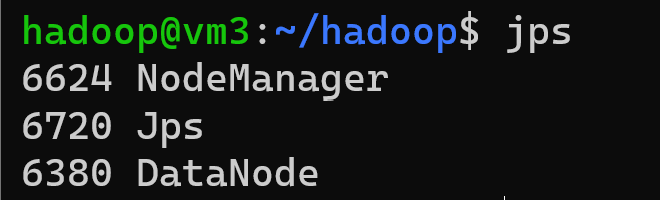
مراحل نصب همانند لینک موجود در تمرین و اسلایدهای درس انجام شد و ip ماشین‌ها نیز در قسمت hosts هر سه ماشین به صورت زیر قرار گرفت:

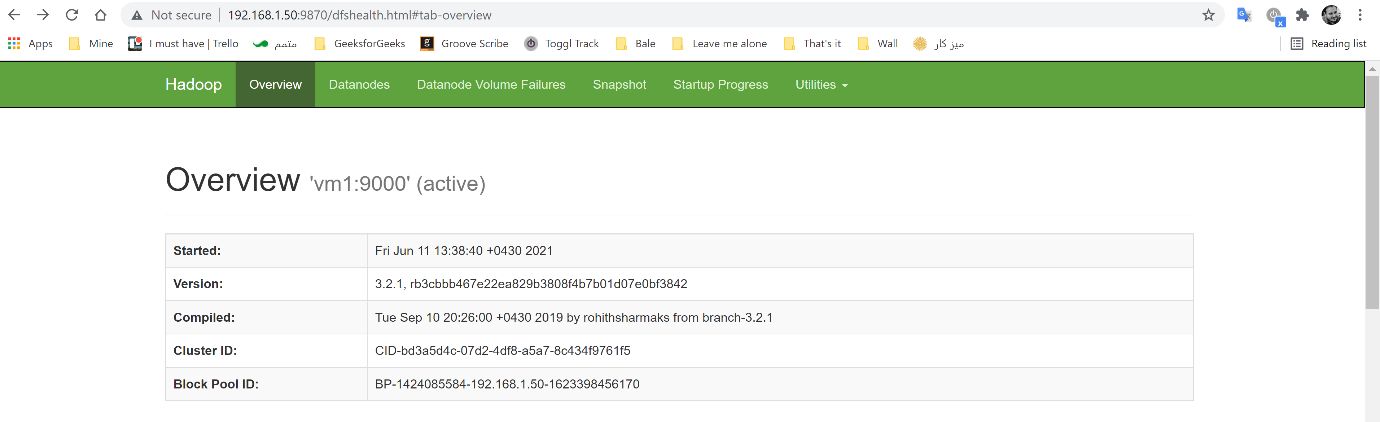
هم‌چنین public-key ماشین اصلی در دو ماشین دیگر قرار گرفت تا ارتباط بین ماشین‌ها نیز بدون مشکل انجام شود.

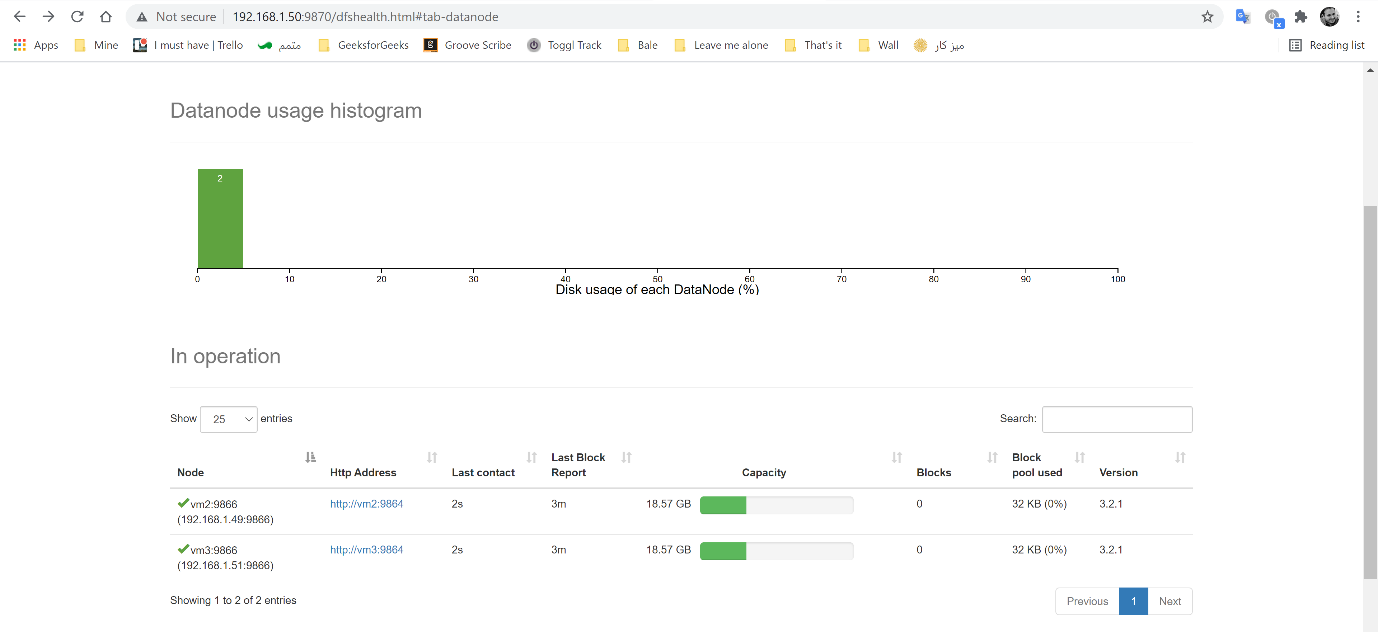
سپس تنظیمات خود Hadoop بر اساس لینک موجود در تمرین و اسلایدهای درس روی ماشین اصلی صورت گرفت و بر روی دو ماشین دیگر نیز کپی انجام شد.

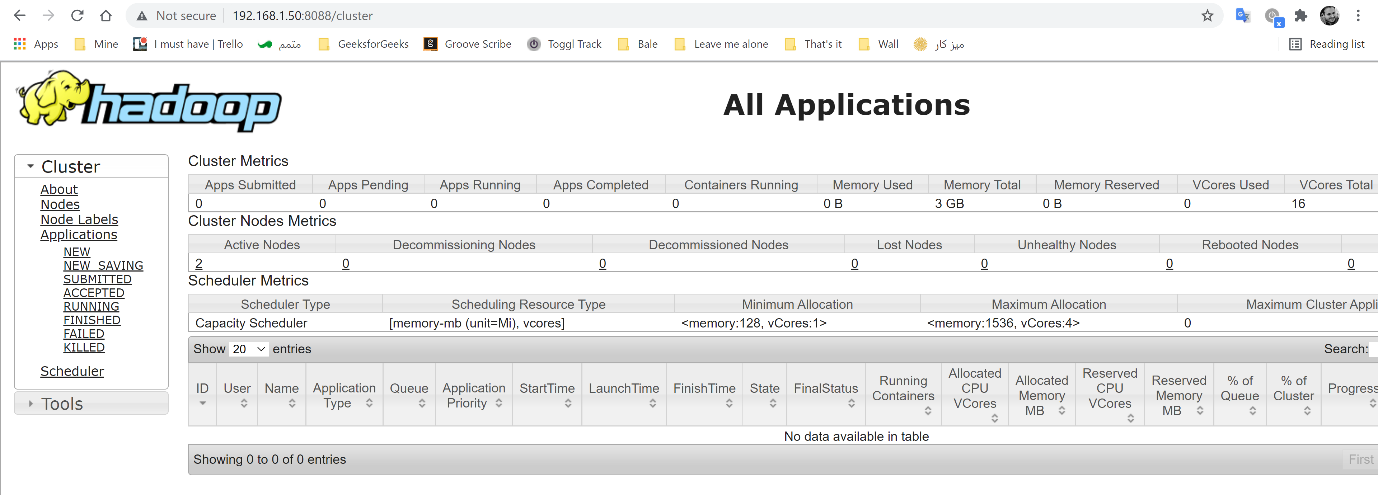
در این‌جا اسکریپت شروع Hadoop و Yarn اجرا شد و موارد مدنظر به درستی در حال اجرا بودند:





در زیر صحت اجرای این موارد در Web-GUI خود HDFS و Hadoop قابل مشاهده است:

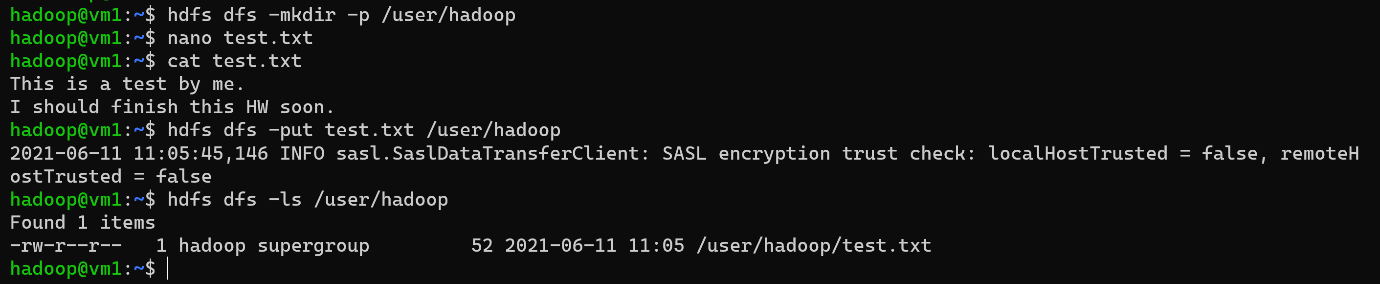


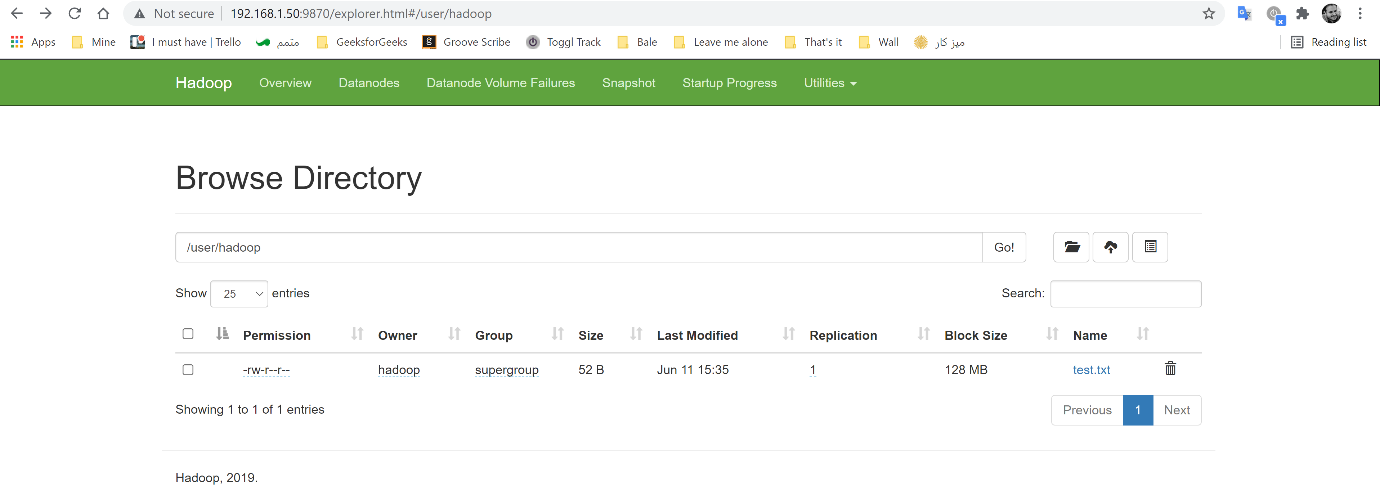


توضیحاتی که در مورد اطلاعاتی که WebGUI در اختیار ما قرار می‌دهد نیز به صورت زیر است:

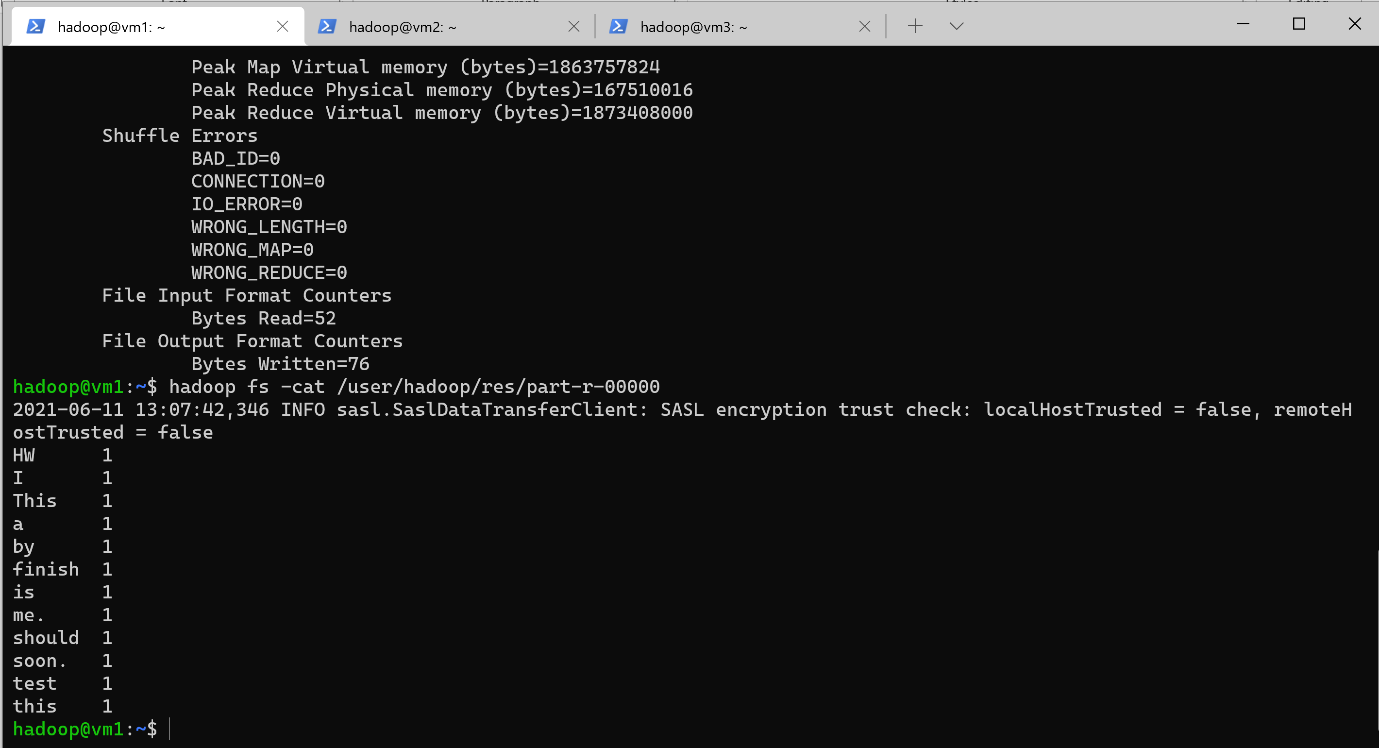
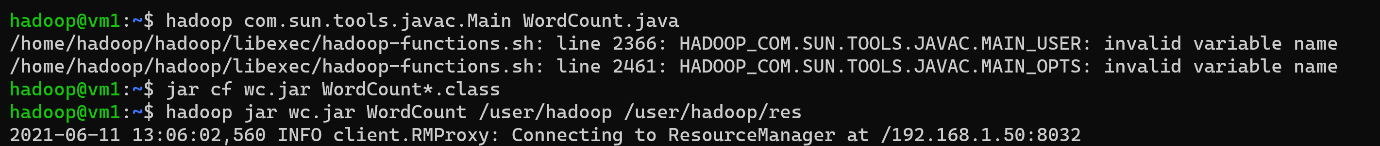
* در قسمت overview اطلاعاتی در مورد زمان شروع به کار سرور، نسخه‌ی آن، زمان کامپایل آن، آیدی cluster داده شده است.
* در قسمت summary نیز اطلاعاتی در مورد تعداد فایل‌های موجود، انواع مموری‌های مورد استفاده و... موجود است
* هم‌چنین در جدولی که در این قسمت وجود دارد اطلاعاتی در مورد کل فضای ذخیره‌سازی‌ای که در دسترس هست و میزانی از آن که توسط dfs مورد استفاده قراره گرفته و میزانی که آزاد است قابل مشاهده است.
* در پایین‌تر تعداد نودهایی که فعال هستند نیز قابل مشاهده است که با کلیک بر روی live node اطلاعات بیشتری در مورد هر کدام از این نودها مانند فضای ذخیره‌سازی آزاد در هر کدام، block pool استفاده شده در هر کدام از آن‌ها، آدرس ماشین آن‌ها و... قابل مشاهده است.

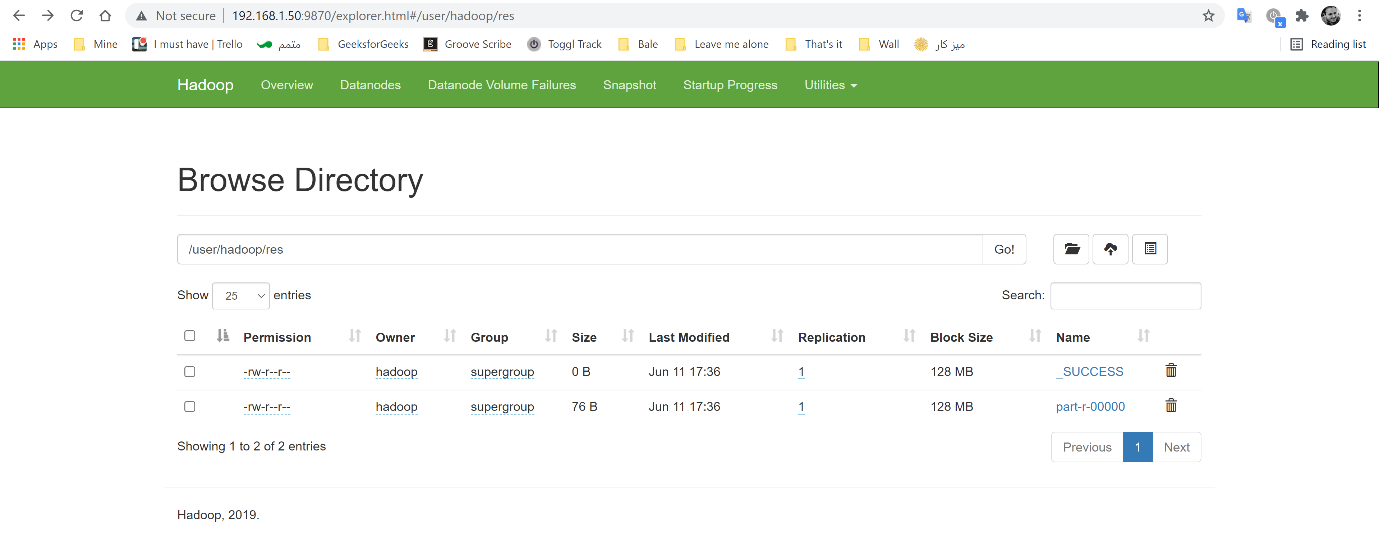
## بارگذاری فایل

همان‌طور که در تصویر زیر مشخص است، فایل مورد نظر پس از ساخته شدن و با موفقیت نیز بارگذاری شده است:

  
که از طریق WebGUI نیز قابل مشاهده است:

## تست WordCount

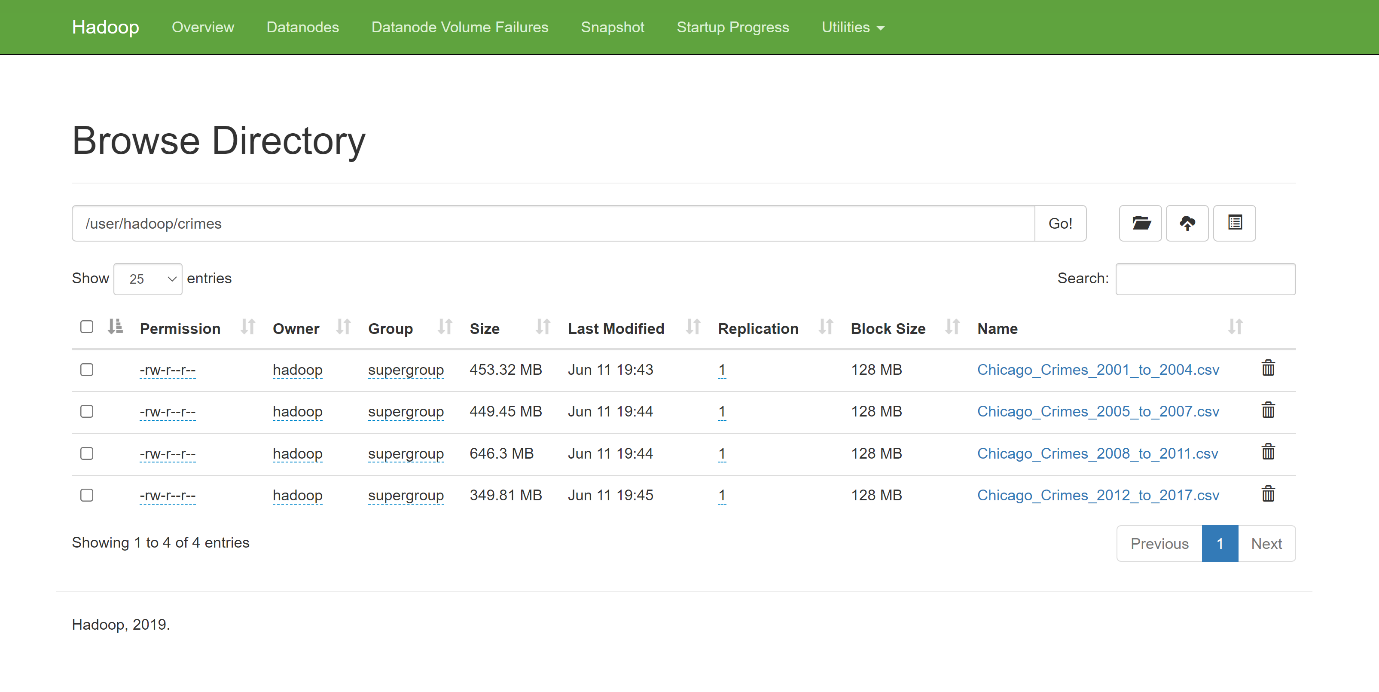
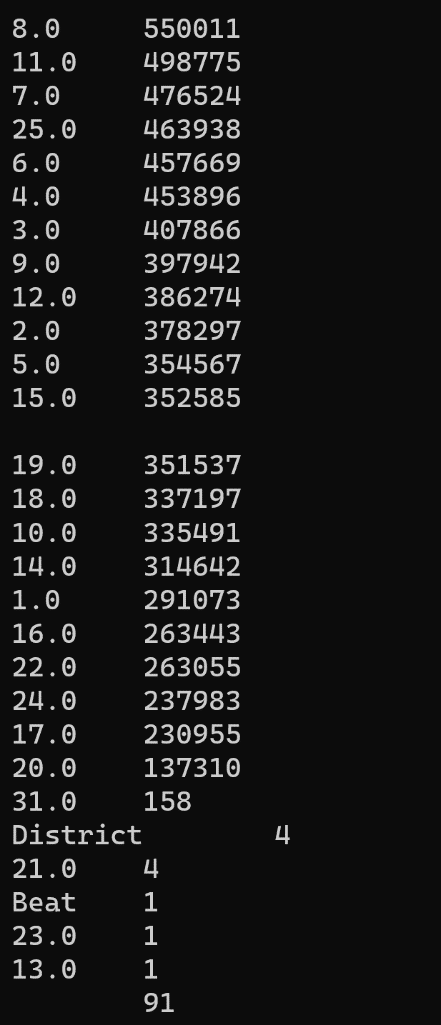
کد WordCount موجود در سایت Hadoop را درون ماشین اصلی قرار می‌دهیم و آن را کامپایل می‌کنیم و بر روی فایل تستی که در قسمت قبل ساختیم اجرا می‌کنیم:

  
در تصویر بالا نتیجه‌ی اجرای کد بر روی فایل تست نیز مشخص است.

در قسمت بعد به سراغ دیتاست‌های اصلی تمرین می‌رویم.

## اجرای MapReduce بر روی دیتاست جرم‌ها

دیتاست‌ها را در hdfs قرار می‌دهیم:

سپس کد WordCount را مقداری تغییر می‌دهیم تا بتواند ورودی جدیدی که از نوع csv است را پشتیبانی کند که به این منظور، tokenization بر اساس comma انجام شده است. اسم فایل مربوط به کد جدید نیز DistrictCount گذاشته شده است. پس از ساخت فایل اجرایی این کد، آن را بر روی دیتاست‌ها اجرا می‌کنیم و خروجی را در یک فایل csv ذخیره می‌کنیم.

که نتیجه‌ی اجرای این کد از طریق WebGUI نیز قابل مشاهده است:

