

In the name of Allah
Distributed Systems (Fall 99)
Instructor: Prof. Nasser Yazdani
Teaching Assistant: Reza Shojaei

این تمرین جهت آشنایی با Hadoop و Spark می‌باشد

1. تعریفی راجع به Hadoop و Spark بنویسید و تفاوت‌های آن دو را بیان نمایید سپس برای انجام سوالات بعدی آنها را بر روی Ubuntu نصب نمایید.
2. ابتدا HDFS را تعریف نمایید؟ اگر بخواهیم از پوشه روت hdfs یک ls بگیریم از چه دستوری استفاده می‌نماییم؟
3. دستورات لازم برای ساخت یک پوشه با نام myfolder در hdfs و لود کردن فایل mygraph¹ در آن پوشه را بنویسید.
4. در Hadoop با استفاده از زبان java و تکنیک MapReduce برنامه‌ای بنویسید که تمام گره‌هایی که تفاضل مجموع وزن‌های یال‌های ورودی از مجموع وزن‌های یال‌های خروجی، فرد می‌باشد را گزارش نماید. (پیشنهاد می‌گردد ابتدا مثال word count را برای Hadoop مطالعه فرمایید)
5. در spark با استفاده از زبان Scala سوال فوق را تکرار نمایید. (پیشنهاد می‌گردد ابتدا مثال word count را برای Spark مطالعه فرمایید)
6. چه تفاوت‌هایی میان نتایج سوال‌های چهارم و پنجم ملاحظه می‌نمایید؟ یافته‌های خود را شرح دهید.

¹ فایل mygraph از سه ستون تشکیل شده است که ستون اول نماینده گره مبدا، ستون دوم نماینده گره مقصد و ستون سوم، وزن آن یال را نشان می‌دهد