

تحلیل سیستم‌های داده‌های حجیم

مجموعه تمرین اول

سوال اول

فرض کنید که ورودی MapReduce شامل اعداد صحیح است (کلید آنها مهم نیست). ورودی تابع Map یک عدد صحیح مانند i است و در خروجی دوتایی‌هایی به صورت (p, i) تولید می‌کند که p مقسوم علیه اول i است. برای مثال:

$$\text{Map}(12) = [(2,12), (3,12)]$$

همچنین فرض کنید تابع Reduce کار جمع مقادیر را انجام می‌دهد. برای مثال:

$$\text{Reduce}(p, [i_1, i_2, \dots, i_k]) = (p, i_1 + i_2 + \dots + i_k)$$

برای ورودی 15,21,24,30,49 خروجی نهایی MapReduce را بدست آورید.

سوال دوم

برای ماتریس و بردار ذیل، با استفاده از روشی که در کلاس برای ضرب بردار در ماتریس معرفی کردیم، خروجی تابع Map را بدست آورید.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

سوال سوم

می‌خواهیم حاصل join دو رابطه R و S را بدست آوریم.

R		S	
A	B	B	C
0	1	0	1
1	2	1	2
2	3	2	3

تابع Map را روی این سطرهای این روابط اعمال کرده و سپس ورودی تابع Reduce را مشخص کنید.

سوال چهارم

فرض کنید می‌خواهیم الگوریتم یک مرحله‌ای MapReduce را برای محاسبه ضرب دو ماتریس M با X سطر و Y ستون و ماتریس N با Y سطر و Z ستون استفاده کنیم. جواب سوالات ذیل را به صورت توابعی از X و Y و Z بیان کنید.

الف) تعداد کلیدهای متفاوت تابع Map چه تعداد است؟ تعداد دوتایی‌های کلید-مقدار برای هر کلید چه تعداد است؟ تعداد کل دوتایی‌های کلید-مقدار چه تعداد است؟

ب) ورودی تابع Reduce چه تعداد کلید دارد؟ طول لیست مرتبط به هر کلید چه تعداد است؟

سوال پنجم

فایل notebook ذیل را باز کرده، خط به خط آنرا را در colab اجرا کرده و به پنج سوال مطرح شده در هر بخش پاسخ دهید.

<https://colab.research.google.com/drive/1YV2-SqDJTH-fDKsI4tBZYikKmuPacJE2?usp=sharing>