به نام خداوند جان و خرد

تمرین تحویلی سه

درس رایانش ابری

امیرفاضل کوزه گر کالجی

9931099

مسئله دایکسترا

در این مسئه از سه تابع بهره می بریم. Mapper, reducer, Dijkstra

تابع map، به ازای همسایه های نود ابتدایی، فاصله جدیدی را جساب کرده و آن را به همراه آن گره ذخیره می‌کند.

تابع reducer یک شی از فاصله هایی که پتانسیل کمترین فاصله بودن را دارند تهیه می کند.

و در نهایت تابع Dijkstra بررسی و مقایسه نهایی را انجام می‌دهد و مقادیر نتیجه را مشخص می‌کند

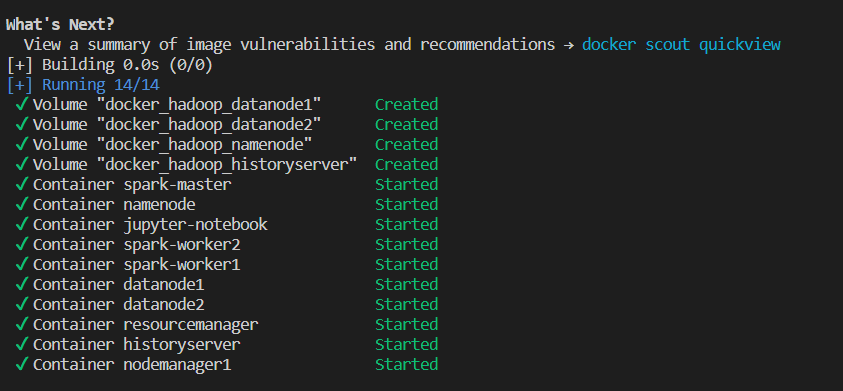
مسئله پیج رنک

تابع مپ در این مسئله، بر تمامی نود های موجود در گراف پیمایش می کند. به ازای هر نود یک مقدار پیج رنکی که باید contribute بکند را محاسبه کرده و آن را بر تعداد گره های همسایه تقسیم می کند. در نهایت لیستی از همکاری ها را برخواهد گرداند.

تابع reducer، با پیمایش بر لیست همکاری ها، مقدار پیج رنک های بعدی را حساب می کند.

اسپارک

با اجرای run\_cluster.sh کانتینر های خواسته شده را می‌سازیم.



در ادامه به توضیح کانتینر ها به صورت مختصر خواهیم پرداخت:

* Resource manager

این کامپوننت، وظیفه مدیریت منابع، تخصیص دادن آنها و نظار بر cpu سیستم را بر عهده دارد.

* History server

این کامپوننت، وظیفه ردیابی جاب های پایان یافته در محیط هدوپ را برعهده دارد. و در مورد آنها جزئیاتی را درمورد تاریخچه شان نگهداری می‌کند.

* Data node

این کامپوننت به طور کلی، مسئول نگهداری داده های هدوپ در سیستم های توزیع شده می‌باشد.

* Nodemanager

این کامپوننت بخشی از YARN می‌باشد . این کامپوننت بر هر ماشین کلاستر اجرا می شود و وظیفه مدیریت کانتینر ها در آن ماشین را بر عهده دارد.

* Namenode

این کامپوننت همانند مرکز نگهداری متادیتا در هدوپ می باشد. داده هایی مانند name space و control access ها توسط این کامپوننت نگهداری می‌شود.

Notebook questions:

* DAG scheduler

یکی از مولفه های مهم اپاچی اسپارک می باشد. مخفف directed acyclic graph می باشد.

هدف مهم آن، تبدیل روند های فرایندی داده کاربر به مراحلی است که بتوان به صورت توزیع شده آن را اجرا کرد.

وقتی یک جاب در اسپارک ثبت می شود، این مولفه آن را به تسک های گوچک و قابل مدیریت تری تقسیم می کند.

DAG به گرافی از RDD ها مبدل می‌شود و عملیات روی آنها اعمال می شود.

پس از اینکه مراحل یا stage ها تعریف شدند، DAG scheduler عملگر های درون سیستم را به یکدیگر به مانند یک پایپلاین وصل می‌کند و یک پلن اجرایی را آماده می‌کند. سپس تمامی این تسک ها را به task scheduler می‌سپارد.

* نحوه عملکرد شافلینگ

در آپاچی اسپارک، عمل shuffling به عنوان روشی برای توزیع مجدد داده ها استفاده می شود.

این توزیع مجدد به گونه ای است که داده ها را در گره های مختلف توزیع می کند.