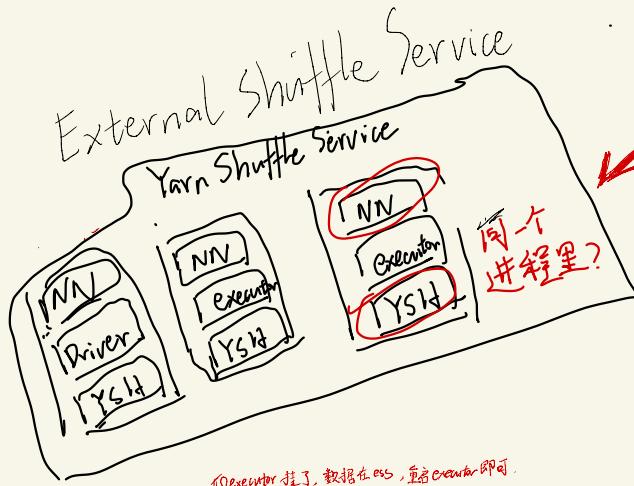


# Spark ESS

Executor 除了执行 task 外，还需负责写 shuffle 数、为其它 executor 提供 shuffle 数据。当 GC 繁忙时，会影响 Task 的执行

why



- ESS 的好处：
- ① Executor 挂了，数据在 ESS，重启 executor 即可。
  - ② Executor 可释放、节省资源。

Impl

Yarn Shuffle Service

K8s Shuffle Service

Magnet

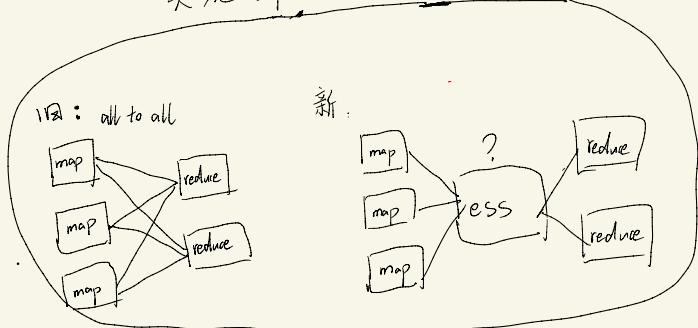
Firestorm 2021.11 腾讯开源

link:

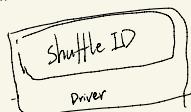
Spark 3.2.0

Spark - 30602

实现了 push-base shuffle service.



maanet 架构

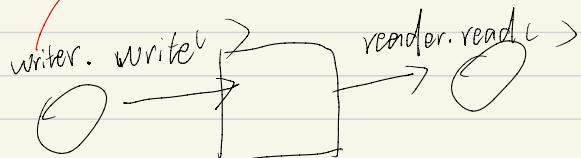


2021.11 腾讯 fire storm

2个问题

数据丢失

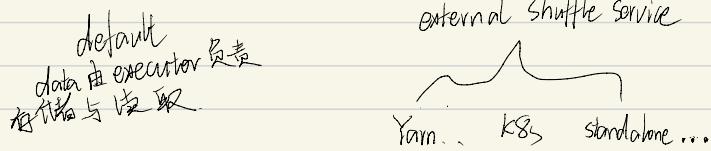
数据不完整



1. 为什么需要 ESS.

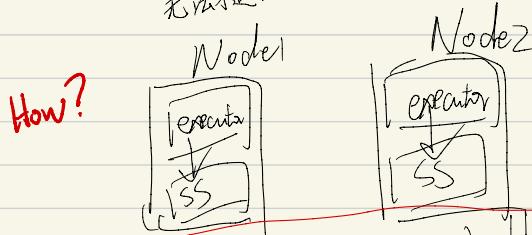
2. 如何启用 Yarn ShuffleService.

3. Yarn ShuffleService 与 Yarn 的关系



why?

executor 除了要执行 Task, 还要写 shuffle 数  
还要读取 shuffle 数据. 当 GC 严重时,  
无法提供数据. Task 运行受影响



Executor 挂了 不该影响 shuffle.

Yarn-site.yaml

① yarn.nodemanager.aux-service = spark-shuffle

② yarn.nodemanager.aux-service.spark-shuffle-class

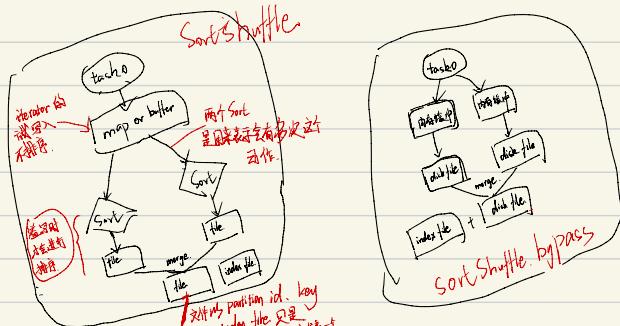
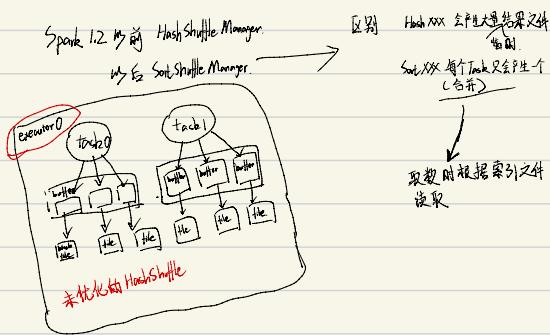
= YarnShuffleService

YarnShuffleService 运行在  
NodeManager 中

① 为什么 1.2 版本用  
SortShuffle Manager

② 优化前后的 map 阶段  
所产生的中间结果文件  
分别是多少？举个例子

③ sort shuffle 实现的一些细节  
例如：输出的数据文件  
索引文件是怎组提的？  
什么时候对数据进行排序的？



文件数，假设 Executor 为 5 个，map 为 10，reduce 为 5

Hash 优化前  $map \times reduce = 30$  个

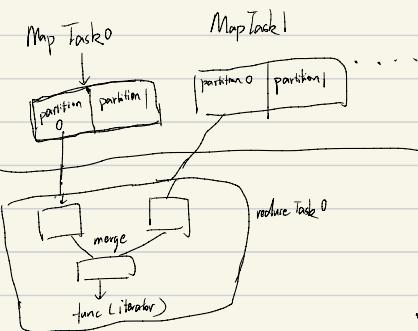
Hash 优化后  $executor \times reduce = 15$  个

SortShuffle  $map = 10$  个

了解 HashShuffle Manager 与 SortShuffle Manager 的区别，以及 SortShuffle Manager 在输出中间结果文件时的一些细节。

--	--

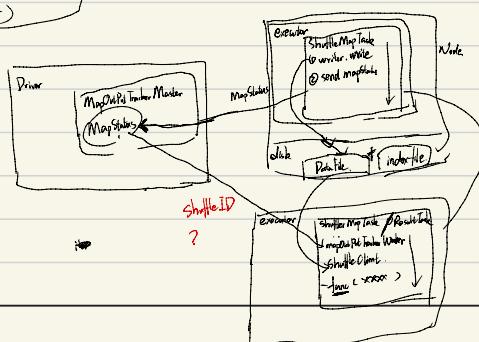
# Shuffle Read / shuffle 寻址



shuffle RDD. iterator( )

shuffleManager.getReader( )

reader.read( )



MapOutPutTrackerWorker → MapOutPut