README

1. Spark 源码层级

以下均在 `spark-2.2.0` 目录下，以**SPARK_SRCD**简称，请自行 `cd` 进入。
重点关注如下子目录：

- `SPARK_SRCD/common`
- `SPARK_SRCD/core`

2. SPARK_SRCD/common

此目录下多为java底层框架的实现，供上层scala使用。

进入更深层级目录：

2.1 SPARK_SRCD/common/network-common

- `SPARK_SRCD/common/network-common/src/main/java/org/apache/spark/network/buffer`

其中的 `NioManagedBuffer.java` 以及 `ManagedBuffer.java` 需要关注。

2.2 SPARK_SRCD/common/network-shuffle

- `SPARK_SRCD/common/network-shuffle/src/main/java/org/apache/spark/network/shuffle`

其中的 `BlockFetchingListener.java` 以及 `ShuffleClient.java` 需要关注。

3. SPARK_SRCD/core

进入更深的目录： `SPARK_SRCD/core/src/main/scala/org/apache/spark` ，以下以**SPARK_CORE**简称。

3.1 SPARK_CORE

- `SparkConf.scala`
- `SparkContext.scala`

3.2 SPARK_CORE/deploy

- `SparkSubmit.scala`

3.3 SPARK_CORE/network

- `BlockTransferService.scala` 我在这里做了一些小修改，主要是看能否 `println` 出中间数据的基本信息。

新增代码如下：

```
println("****************************** [Fetch Data] ******************************\
n" +
        "DataSize:  ${data.size.toInt}\n" +
        "Host:      $host\n" +
        "Port:      $port\n" +
        "ExecId:    $execId\n" +
        "BlockId:   $blockId\n")
```

3.4 SPARK_CORE/scheduler

- `DAGScheduler.scala`
- `MapStatus.scala`
- `schedul[able, er, ing].scala`
- `Shuffle*.scala`
- `Task*.scala`

3.5 SPARK_CORE/shuffle

- `Shuffle*.scala`

3.6 SPARK_CORE/storage

- `BlockManager*.scala`