

期末考试

期末考试——考试的和复习的永远没有交集~

今天终于考完了大学期间的最后一门专业必修考试，下面就都是水的考试了。在这里我想记一下这次考试的内容，以飨后面的几届hhh。

题型？题目难易度，分值分布？

题型是 5道2分的选择题，10道1分的判断题，10个2分的填空题，4道简答题总共30分，3道论述题总共30分。

题目的难易度总体不大，有 30 分是差不多可以直接从书上可以找到的记忆题，比较简单，死记硬背就行。有10分是编程题，会要求写一个迭代操作的伪代码，从 MapReduce, Spark, Flink 里面选两个实现，并且说明两种方法的优缺点，还会考一个填空说为了加快迭代速度 KMeans 聚类的中心点会作为_____ 这个空我填的广播变量（应该是对的吧）。还有 4 分是有点点超纲的，两道选择题，一道是考 Flink 是谁提出的，选 TU Berlin；另一道是靠以下哪个是协调服务系统，选 zookeeper。这两道题目说送分吧，也有点迷惑性。还有 40 分是比较难的记忆题，就是需要把书上的内容理解以后（尤其是把图都记住，这个很重要！）再进行判断 / 填空 / 解答 / 分析。有一道 12 分的让你画出来 Yarn 模式下 Spark client 和 cluster 的节点图，这个我只复习了 Standalone 模式没有太复习 Yarn，所以有点惨55，（可恶原来我 CoarseGrainedBackendExecutor 写的对的结果改成了 YarnChild 然后错了，血亏。）还有 16 分左右吧，是带有应用性质的，比如让你评价 Spark 和 Flink 两种检查点模式的区别，以及评价 Spark 相比于 MapReduce 的优势，还有给你5个节点让你实现HDFS和MapReduce，问怎么样让整个系统性能达到最优（这个是最后一题）。

我估分80分吧，如果宽一点的话。

如何备考？

我的建议：

- 课后习题一定要做，好好消化！好好做课后习题可以拿几乎一半的分。
- 老师上课提的问题一定要好好思考，多回答，这样既有利于提升平时分，也有利于记笔记增加印象，在备考的时候多问自己几个问题
- 书上教学的图一定要能记则记！虽然看起来这些名词很难记但是多消化多理解就可以了！而且整本书从HDFS到Flink 都是一以贯之的分布式架构。
- 看书的时候，按照书上p9的五个阶段思考，这个也很重要！一直要思考，分布式计算系统为何如此演进？HDFS相比于原有的存储系统、MapReduce 相比于 MPI 的优点、Spark 相比于 MapReduce 的优点等等。
- 有空也可以多看一下编程题