**一、存储性能需求估算**

**用户量预估**

假定给教育部做系统，直接推算为 1000 万用户规模。

关键行为：考试

**用户行为建模和性能估算**

假设每门学科每年 2 次考试，每个学生平均一学期 20 门课，考试采取机考的方式，每门考试的答案 20 判断题、20 选择题、4 道大题（200 字以内），考试结果永久保存，在校学生能够看到自己曾经的考试结果，则考试结果记录的存储量为：

* 在校学生：1000 万 \* 20（课）\* 2（考试次数）\* 1000（答案） \* 2（学期） \* 3（只有前三年考试） = 2.4T
* 离校学生：每年 250 万，存储量为 0.6T

假设学校的考试都安排在某一个月内，考试的时候请求试卷，提交答案，中间答题过程浏览器本地完成，由于考试集中在上午 4 小时和下午 4 小时，且请求试卷集中在考试考试的前 1 分钟，提交答案集中在考试结束前的 30 分钟，因此估算如下：

* 请求试卷：1000 万\*20（课）/20(周末不考试) /4(每天 4 堂考试)/1 分钟(60s)，约等于 5w/s
* 提交试卷：1000 万\*20（课）/20(周末不考试) /4(每天 4 堂考试)/30 分钟 = 1700/s

假设试卷有 20 判断题（50 字节）、20 选择题(200 字节)、4 道大题(400 字节)来计算，估算一张试卷的存储如下：

* 20 \* 50 + 20 \* 200 + 4 \* 400 按照 10k 计算
* 考试中涉及图片的，使用图片存储，试卷本身存储图片的 URL

假设大学一学期开设 500 门课程，一年消耗的存储：

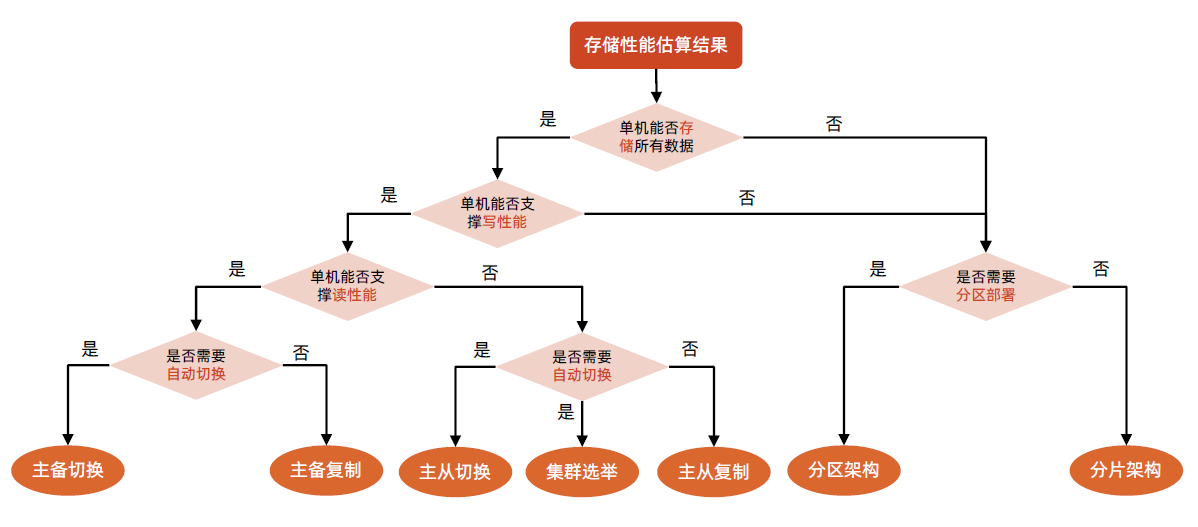
* 2（学期）\* 500（课程数） \* 10k 按照 10M 计算

**存储性能需求汇总**

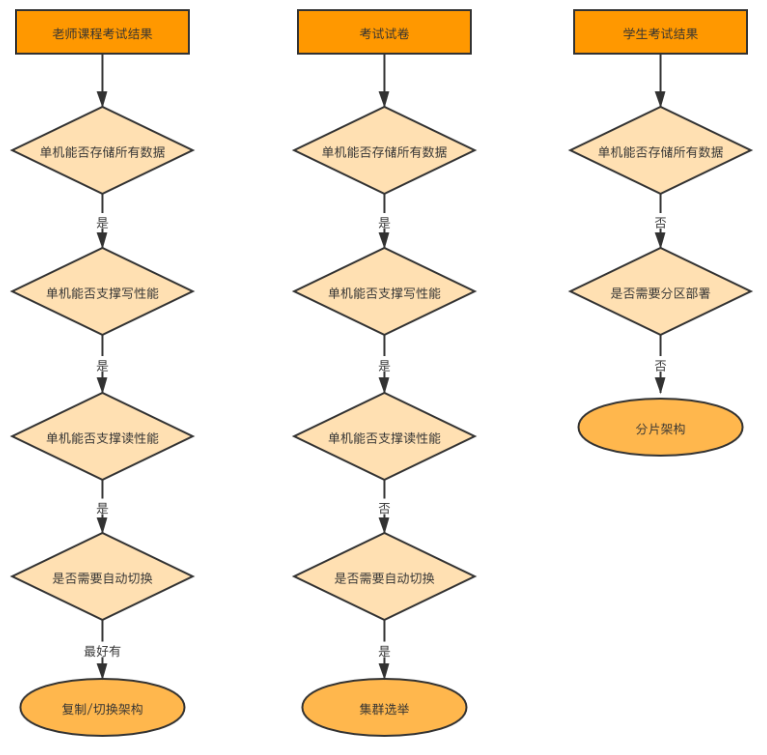
* 在校学生考试结果存储：2.4T
* 离校学生考试结果存储：每年增长 0.6T
* 试卷请求 QPS：5w/s
* 提交试卷 TPS：1700/s
* 试卷存储一年：10M
* 单张试卷：10k

**二、选择存储系统**

存储架构选择逻辑：



考试存储架构分析如下：



考试试卷在开考的时候会有高访问，且都是一样内容和试卷。选择 redis sentinel。

教师的考试成绩需要做相关的成绩分析，选择 mysql 集群方式。

学生考试结果，只需要自己进行查看，选择 hbase。

**三、设计存储方案**

**考试试卷存储**

数据结构设计：

* 存储类型：Redis Sentinel
* 数据类型：string 类型就可以满足需求，且 key-value 都没有超过 redis 使用的建议值。
* Key：学校 ID:课程 ID:考试 ID
* Value：试卷格式为 Json，转为 String 进行存储

读写分析：

* 老师编写完试卷之后，以学校 ID:课程 ID:考试 ID 为 key 写入到 redis 中。
* 学生开始考试后，根据相应的考试科目，系统读取对应 key 的 value。

**学生成绩存储**

数据结构设计：

* 存储类型：Hbase
* key: 学校 ID+学号 ID+考试 ID
* Column Family: test
* Column：result，score，其中 result 是 JSON 格式

读写分析：

* 学生提交考试结果，直接按照 key 保存 result；
* 老师改卷后，直接写入 score
* 学生查看自己的成绩，按照 key 读取 result 和 score，可以看到得分和具体错在哪里

**老师成绩存储**

数据结构设计：

* 存储类型：MySQL 集群
* 相关表结构设计：



读写分析：

* 试卷评分完成后，直接按照表结构保存结果。
* 直接读取成绩表进行相关数据分析。

**四、redis sentinel 集群服务器数量和性能**

试卷读取 5w/s，一般 redis kv 能都达到 5w-10w，带宽也在单机范围内，考虑到高可用，使用 redis sentinel。

根据上面推算的数据，最终机器为 redis 主一台，redis 从一台，sentinel 三台。