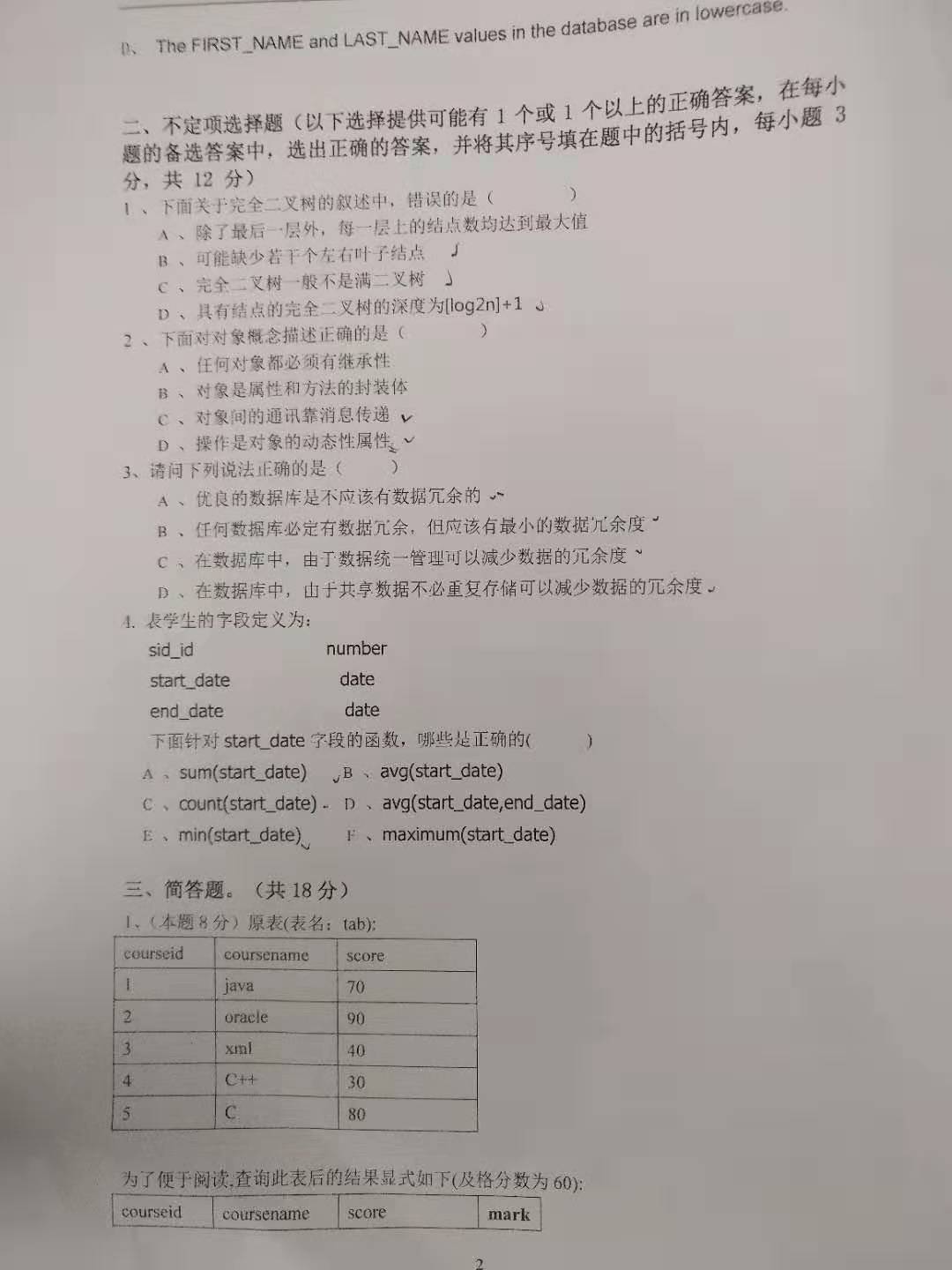
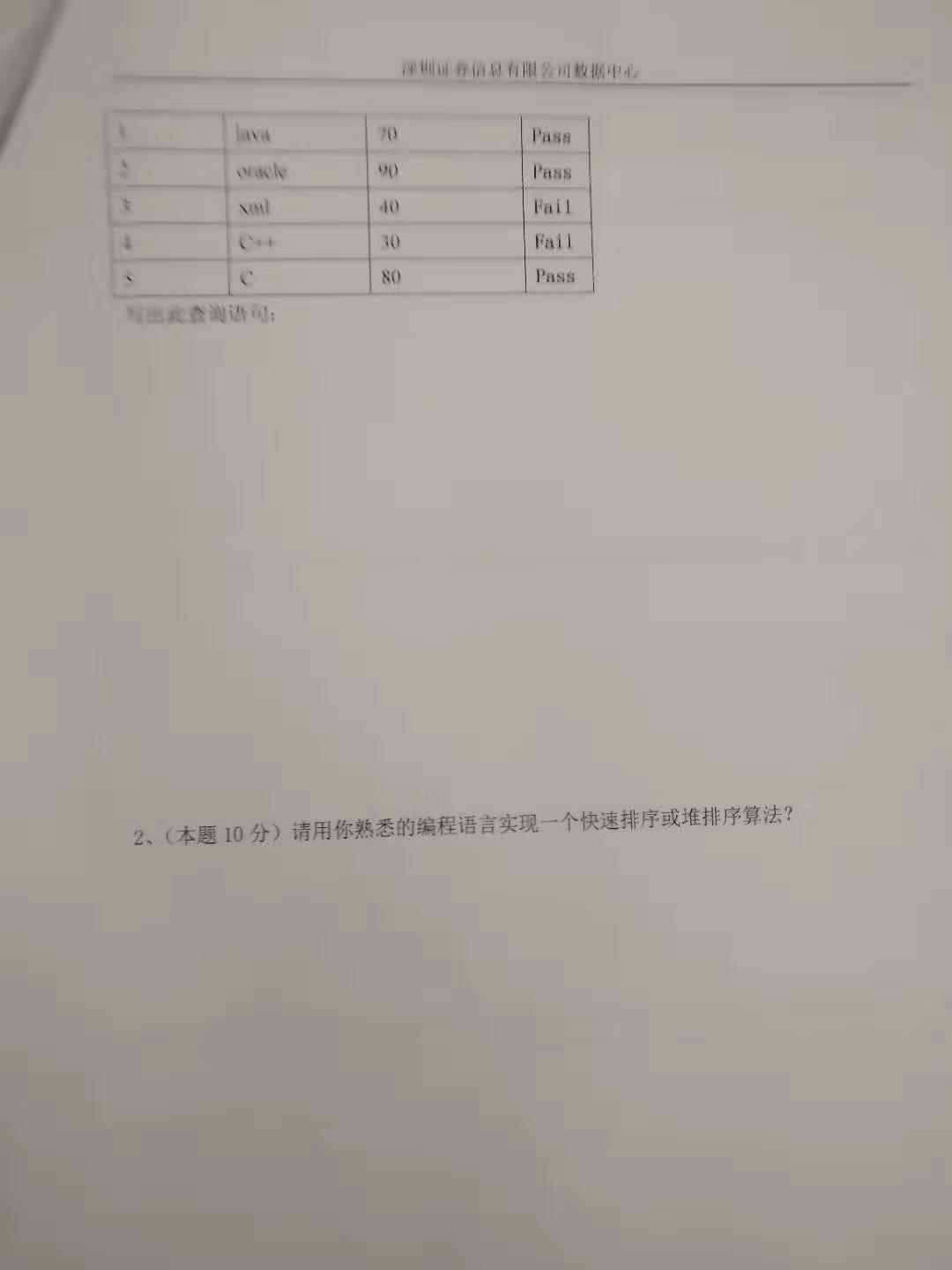


答案：1.A 2. D 3. D 4.C 5.D 6.D 7.B 8.C 9.C 10.D



答案：1.A 2.BCD 3.BD 4.CE



1.

SELECT T1.\*

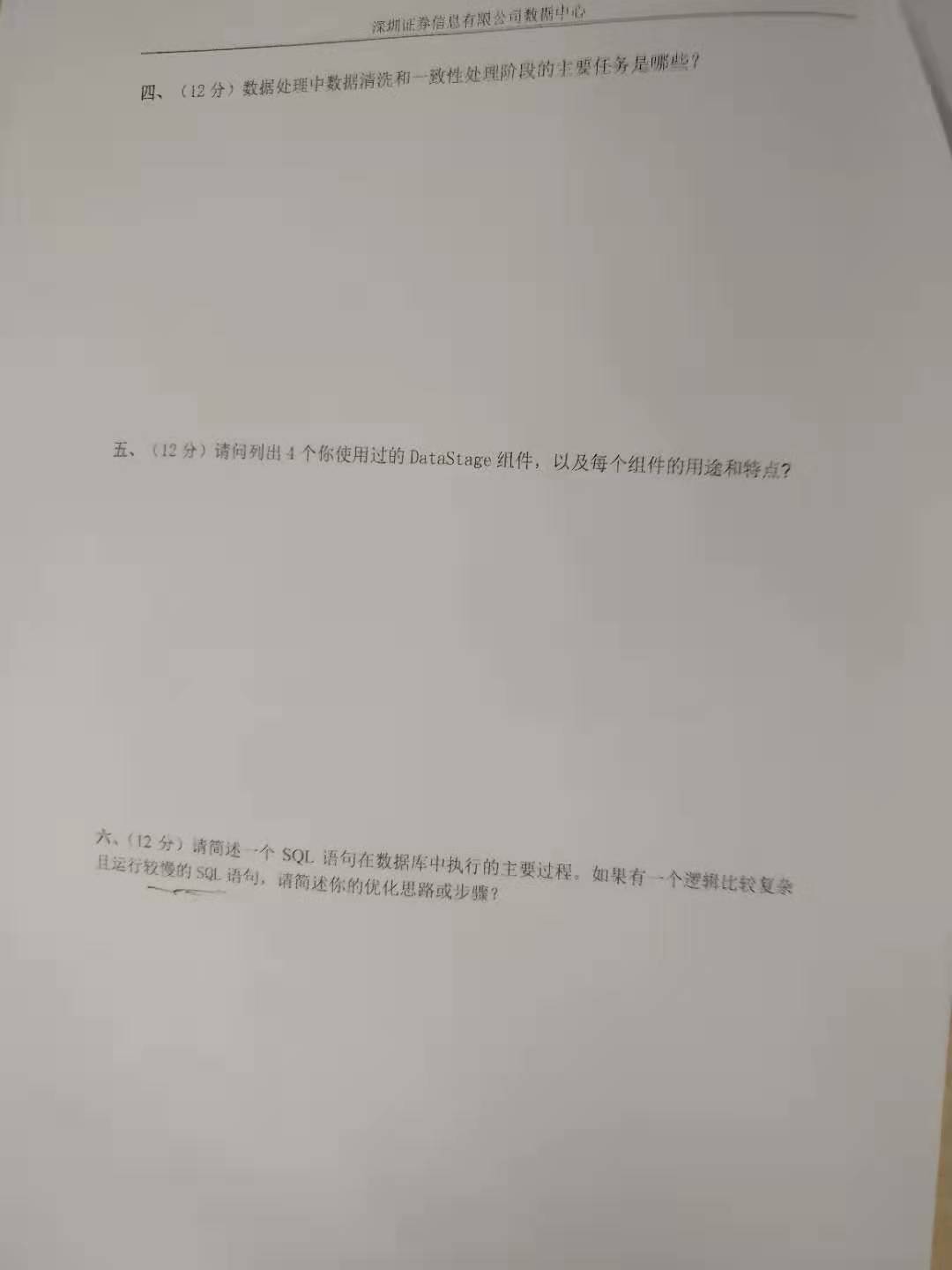
,CASE WHEN SCORE>=60 THEN 'PASS'

ELSE 'FAIL' END MARK

FROM TAB

2.

SELECT \* FROM 表 ORDER BY 排序字段 DESC



4.数据清洗和一致性处理阶段的主要任务：数据补全（补充数据的完整性）；去除重复数据；去除或修改格式和内容错误的数据

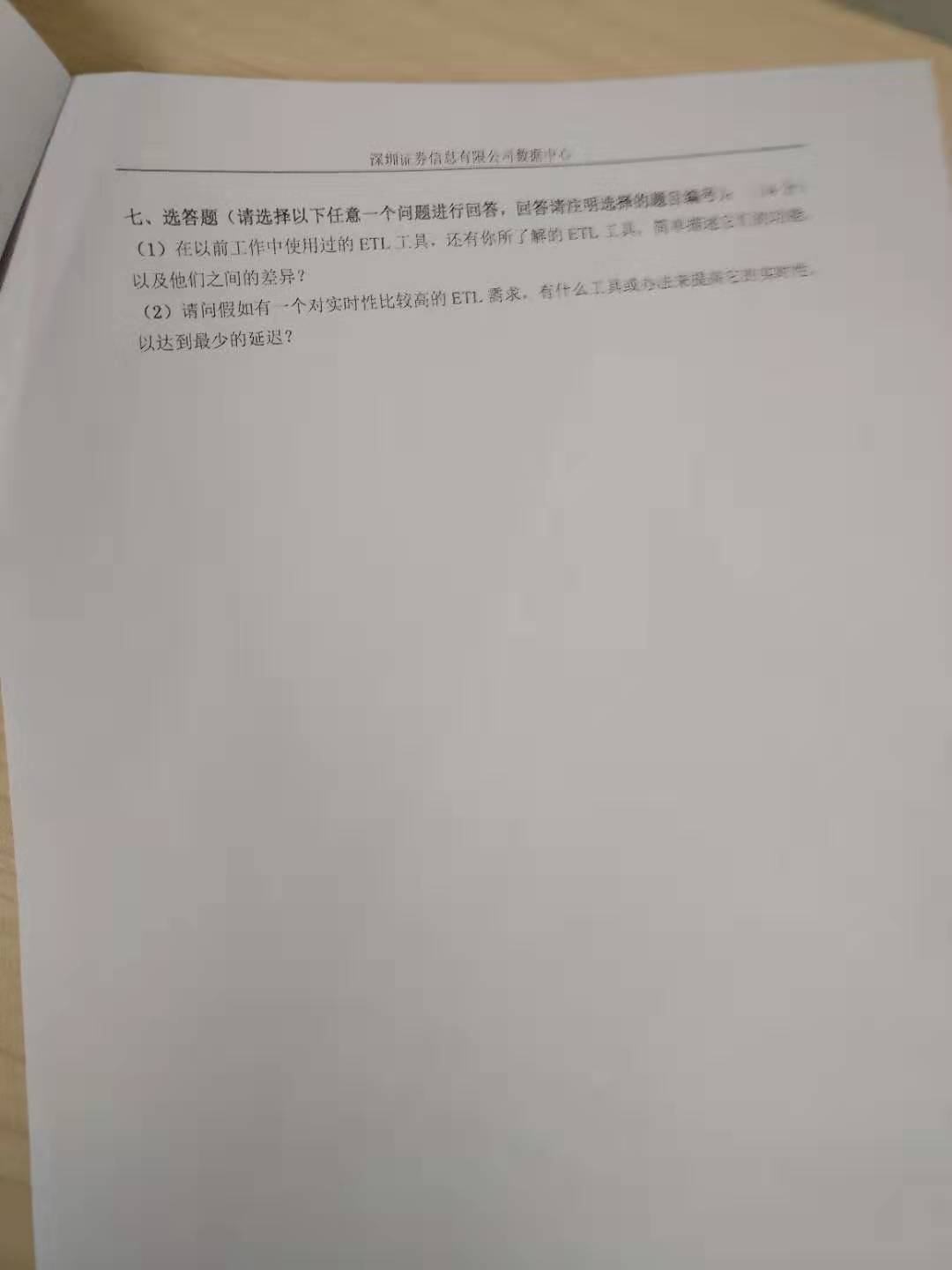
5.lookup stage针对小表进行数据匹配，装载小表进内存，效率高，但是费内存；JOIN STAGE通常用于多张同级别的表进行关联，轻内存；MERGE STAGE要求主数据集与更新数据集无重复，其余同JOIN STAGE

Copy Stage可以有一个输入，多个输出。他可以在输出时改变字段的顺序，但是不能改变字段类型。

6. 执行顺序FROM🡪WHERE🡪GROUP BY 分组🡪HAVING过滤🡪SELECT查询字段

优化思路：少用子查询，多用表关联，尽量减少查询次数；

看扫描表方式，是全表扫描还是索引扫描，走索引扫描更优，看是否索引失效，找出索引失效原因；



ETL工具：Kettle、DataStage、Informatica

Kettle 开源免费，扩展性差，需自行实施、维护

DataStage 不开源，扩展性好，主要为第三方的实施和售后服务

Informatica 不开源，扩展性好，主要为第三方的实施和售后服务