

西安电子科技大学

考试时间 120 分钟

试 题

题号	一	二	三	四	五	六	七	八	总分
分数									

1. 考试形式：闭卷 ☒ 开卷 ☐

2. 考试日期： 年 月 日 (答题内容请写在装订线外)

一、关于分布式系统的定义和特性，回答下列问题。(共 16 分)

1. 简述分布式系统的定义。(4 分)
2. 同一主板上包含多个 CPU 的计算机系统是不是分布式系统？为什么？(4 分)
3. 同一个物理主机上是否可以包含多个分布式计算节点？为什么？(4 分)
4. 用多个计算节点构成分布式系统可以带来哪些好处？(4 分)

二、关于分布式系统的构架模式，回答下列问题。（共 12 分）

1. 简述主从模式（Master-Slave）和对等模式（Peer-to-Peer）这两种分布式系统构架模式的定义和各自的优缺点？（8 分）
2. Google 中的 GFS 和 MapReduce 系统采用了哪种构架模式？（2 分）
3. 比特币系统采用了哪种构架模式？（2 分）

三、关于远程过程调用（RPC）技术，回答下列问题。（共 12 分）

1. 从应用程序员角度看，RPC 技术的主要作用是什么？（4 分）
2. 简述 RPC 的实现原理和 RPC 中间件的主要作用。（6 分）
3. 支持跨编程语言调用的 RPC 中间件系统中，如 CORBA 和 gRPC，都会定义专用的接口定义语言（IDL: Interface Definition Language），请问 IDL 语言的作用是什么？（2 分）

四、关于基于消息中间件的分布式通信技术，回答下列问题。（共 11 分）

1. 简述发布/订阅模式和消息队列模式的区别。（4 分）
2. 消费者接收消息的三种方式是什么？（3 分）
3. 分布式节点之间基于消息中间件进行通信有哪些优点？（4 分）

五、关于 HDFS 分布式文件系统，回答下列问题。（共 11 分）

- 1. HDFS 的 NameNode 节点和 DataNode 节点的主要功能分别是什么？（4 分）**
- 2. HDFS 采用什么技术提高数据存储的可靠性？（3 分）**
- 3. 简述 HDFS 客户端读取一个 HDFS 文件时与 NameNode 和 DataNode 的交互过程。（4 分）**

六、根据下面某软件工程在数据分布式存储方面的需求描述，回答下列问题。（共 14 分）

需求描述：为加强疫情防控，某热门景区计划在关系型数据库中保存每日到访的游客实名信息。游客信息数据表的联合主键包含两个字段：身份证号，访问日期。为分担存储和查询负载，计划将游客信息分布存储在 7 台硬件配置接近的服务器中，每台服务器上运行一个关系型数据库系统软件（如 MySQL）。另外部署一台 Web 服务器运行前端软件模块以处理客户端插入和查询游客信息的交互请求。系统支持两类查询。查询类型 1：给定身份证号查询游客信息；查询类型 2：给定日期查询当日到访的游客信息列表。

1. 假定采用基于哈希的数据分区法，即根据身份证号的哈希值进行分区。请简述 Web 服务器中前端软件实现游客信息插入的基本思想。（4 分）。
2. 在上述分区法中，简述前端软件实现类型 1 查询的基本思想。（2 分）
3. 在上述分区法中，简述前端软件实现类型 2 查询的基本思想。（4 分）
4. 另一种数据分区方法是根据游客所属的地理区域进行分区，即 7 台服务器分别对应全国 7 大地理分区（华北、东北、华中、华东、华南、西南、西北）。游客所在省份（通过身份证号可以判断）归属那个地理分区就由哪个服务器进行存储。与基于哈希的数据分区法相比，该分区方法有什么缺点？（4 分）

七、根据需求，描述 MapReduce 程序的实现方法。（12 分）

假定作为输入的 HDFS 大数据文件的每一行都存放了一个形如<ID1, ID2>的二元组，表示微博用户 ID1 关注了微博用户 ID2，或者说 ID1 是 ID2 的粉丝。要求编写 MapReduce 分布式计算程序，其功能是根据输入文件输出具有成为大 V 博主资格（粉丝数超过 50 万）的所有博主 ID 及其对应的粉丝数量。请给出该程序的 Map 方法和 Reduce 方法的伪代码。

八、根据需求，描述 Spark 程序的实现方法。（12 分）

假定作为输入的 HDFS 大数据文件中保存了学生成绩信息，格式如下：

班级 1, 姓名 1, 科目 1, 必修, 成绩 1
 （注：
 为换行符）

班级 2, 姓名 2, 科目 1, 必修, 成绩 2

班级 1, 姓名 1, 科目 2, 选修, 成绩 3

.....,,,,

现要求编写一个 Spark 程序，统计学生必修课平均成绩在：90~100, 80~89, 70~79, 60~69 和 60 分以下这 5 个分数段的人数。请给出该程序的伪代码。