

作业提交注意事项

1. 作业以邮件附件的形式提交至 SDU_Distributed_Computing@outlook.com, 邮件标题为“**分布式计算第 X 次作业_姓名_学号**”, 附件的标题相同。请将 X 替换为作业次数(阿拉伯数字, 如“分布式计算第 1 次作业”), 不要漏掉标题中的下划线, 不要弄错姓名和学号的顺序。提交格式错误的作业将扣除不超过 5 分的卷面分(每次作业按百分制计分)。
2. 附件可以以.doc、.docx、.pdf 格式提交。使用 Word、WPS 等富文本编辑器或使用 LaTeX 完成均可。不允许直接在邮件正文中作答。不允许直接编辑试卷 PDF 进行作答。除非题目中明确要求截图, 否则所有的答案应该以键盘录入, 禁止提交截图或自己手写答案的拍照。
3. 作业需要在布置之后两周内提交, 以两周后的 19:00 为界(例如, 4 月 1 日布置的作业应在 4 月 15 日 19:00 前提交)。迟交 24 小时的作业将扣除一半分数。不接受超过 24 小时的迟交。
4. 如果在 Word 等富文本编辑器中完成作业, 中文字体建议选择宋体或微软雅黑(五号), 英文字体建议选择 Times New Roman。数学公式请使用 Word 的公式环境进行输入, 或在 LaTeX 编辑器中输入。禁止以截图形式提交数学公式。如果使用 Word 完成作业, 建议将数学公式内的字体统一为 Cambria Math(请注意, 从他处复制的数学符号的字体可能并不统一)。对于每道题目的回答应写清题目题号, 伪代码以及证明的书写应注意适当的缩进。对于排版方式过于糟糕的作业, 将扣除不超过 5 分的卷面分。
5. 作业应独立完成, 禁止小组内部合作完成一份作业, 禁止通过搜索引擎直接复制黏贴答案, 禁止通过大语言模型获取答案。对于明显的违规者, 将完全扣除本次作业之成绩。

山东大学 2023-2024 学年 二 学期 《分布式计算》第 2 次作业

| | | | | | | | | | | | |
|----|---|---|---|---|---|---|---|---|---|---|----|
| 题号 | 一 | 二 | 三 | 四 | 五 | 六 | 七 | 八 | 九 | 十 | 总分 |
| 得分 | | | | | | | | | | | |

| | | | | | | | | | | | | |
|----|---|---|---|---|---|---|---|---|---|----|----|----|
| 题号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 得分 | | | | | | | | | | | | |

一、填空题（本题共 36 分，每空 3 分）

- 在云服务的技术指标中，_____（英文缩写）最早是网络通信中的一种机制。
- _____技术将所有数据按照固定的大小分块，每一块赋予一个用于寻址的编号。
- 在 MapReduce 框架中，能够降低网络通信开销的操作为分布式_____操作。
- IBM 认为，_____技术是云计算的基础。
- MapReduce 框架中的 sort 模块根据_____（中文名称）对 mapper 产生的中间键值对进行排序。
- 在网络存储技术体系中，_____（英文简称）通过瘦服务器提供文件存取服务。
- 具有标准化格式、能够以表格形式高效管理的数据被称之为_____。
- 对于云计算服务器的动态能耗与静态能耗而言，_____在服务器利用率较低时占主导。
- MapReduce 框架的任务调度策略是由任务追踪器主动向_____（中文全称）请求任务。
- 课程中所介绍的装箱问题的三种启发式算法中，_____（英文全称）算法的近似比被证明不超过 1.7。
- 如果将 MapReduce 框架的 Reduce 操作视作定义在中间键值对的值列表集合上的二元运算符，那么该集合与运算符组成的代数结构是一个_____。
- 分布式对象存储系统 Ceph 采用的一致性模型是_____。

| | | | | |
|----|---|---|---|---|
| 题号 | 1 | 2 | 3 | 4 |
| 得分 | | | | |

二、判断题（本题共 16 分，每小题 4 分）对于下面的每种说法，判断其是否正确，并简要写出理由。

- 用户在社交网站上上传的照片往往具有清晰的构图结构，因此属于结构化数据。
- 对于虚拟机 V_1 上的进程 P_1 和虚拟机 V_2 上的进程 P_2 ，如果 V_1 和 V_2 被部署在同一台物理服务器上，那么 P_1 和 P_2 可以以共享内存的方式进行通信。
- MapReduce 是一种优秀的框架，其单线程实现比传统实现更加高效。
- 在 MapReduce 框架中，InputFormat 模块负责读取输入数据。

| |
|----|
| 得分 |
| |

三、分析题（本题 16 分）

下图给出了 4 个进程（P1, P2, P3, P4）对于对象 x 执行的一系列读写操作，其中 W(x)表示写操作，后面的值代表写入的值，例如 W(x)a 表示向写入值 a；R(x)代表读操作，后面的值代表从对象 x 中读到的值，例如 R(x)b 代表从 x 中读取到的值为 b。越是靠近左侧的指令，在全局时钟中发生的顺序越靠前。

| | | | | |
|-----|-------|-------|-------|-------|
| P1: | W(x)a | | | |
| P2: | | W(x)b | | |
| P3: | | | R(x)b | R(x)a |
| P4: | | | R(x)b | R(x)a |

已知上图的读写操作符合顺序一致性模型，因为可以将它们按照以下顺序交织后按顺序执行，得到的结果（即读操作读到的结果）保持不变。

| | |
|-----|-------|
| P2: | W(x)b |
| P3: | R(x)b |
| P4: | R(x)b |
| P1: | W(x)a |
| P3: | R(x)a |
| P4: | R(x)a |

下面，针对以下两张图中的指令执行结果，试分别分析其是否符合顺序一致性模型。如果符合，请模仿上图给出一种指令的交织顺序；如果不符合，请说明理由。

(1)

| | | | | |
|-----|-------|-------|-------|-------|
| P1: | W(x)a | | | |
| P2: | | W(x)b | | |
| P3: | | | R(x)a | R(x)b |
| P4: | | | R(x)b | R(x)a |

(2)

| | | | | |
|-----|-------|-------|-------|-------|
| P1: | W(x)a | | | |
| P2: | | W(x)b | | |
| P3: | | | R(x)a | R(x)b |
| P4: | | | R(x)a | R(x)b |

| |
|----|
| 得分 |
| |

四、计算题（本题 16 分）

在装箱问题中，箱子的容量为 1，现有 7 个物品，其尺寸分别为 0.2, 0.5, 0.4, 0.7, 0.1, 0.3, 0.8。

- (1) 如果使用 Next Fit 算法，请计算会使用几个箱子，并写出每个箱子中装载了哪些物品。
- (2) 如果使用 First Fit 算法，请计算会使用几个箱子，并写出每个箱子中装载了哪些物品。

姓名

学号

级

专业

学院

密

封

山东大学 2023-2024 学年 二 学期 《分布式计算》第 2 次作业

得分

五、程序设计题（本题 16 分）

在本题中，我们将基于 MapReduce 框架，以伪代码的形式实现素数统计。输入的文件中包含了一系列正整数，其中包括素数和合数（一个正整数仅能被 1 和其自身整除，那么这个数字是一个素数，否则是一个合数），要求我们统计其中每个素数出现的次数（合数不做统计）。请你补完下面的 Reduce 函数。

```
function map(string name, string document):
```

```
    for each number n in document:
```

```
        emit (n, 1)
```

（注意，在上面的 map 函数中，我们忽略了将 document 中的字符串转化成数字的过程。）

```
function reduce(int n, Iterator partialCounts):
```

```
    // n: a number
```

```
    // partialCounts: a list of aggregated partial counts
```

在你完成的 Reduce 函数中，请在函数内完成对于数字 n 是否为素数的判断（可以通过简单枚举 2~n-1 能否整除 n 来实现，也可以使用更复杂或更优越的素数判别算法，但请给出说明）。如果 n 是素数，那么请使用 emit 函数发射 (n, sum)，其中 sum 为 n 出现的次数；如果 n 是合数，请不要产生输出。你完成的代码行数不必和划线行行数相同。

学院 专业 级 学号 姓名

.....封.....密.....

山东大学 2023-2024 学年 二 学期 《分布式计算》第 2 次作业

| | |
|--|--|
| | |
|--|--|