

一、填空题

1. QoS
2. 块存储
3. Shuffle
4. 虚拟化
5. 中间键
6. NAS
7. 结构化数据
8. 静态能耗
9. 作业跟踪器
10. Next Fit Decreasing (NFD)
11. 半群
12. 分布式一致性模型

二、判断题

1. 不正确。虽然用户上传的照片可能具有一定的结构，比如人物、场景等，但照片本身是一种非结构化数据，因为它们通常是以像素的形式存储，而不是按照特定的结构或格式。结构化数据通常指的是以表格、数据库等形式存储的数据，具有明确定义的字段和值，而照片通常没有。
2. 正确。如果虚拟机 V1 和 V2 被部署在同一台物理服务器上，并且操作系统支持共享内存的机制，那么 V1 上的进程 P1 和 V2 上的进程 P2 可以通过共享内存的方式进行通信。共享内存是一种机制，允许多个进程访问同一块内存区域，从而实现数据的共享。
3. 不正确。MapReduce 是一种并行计算框架，旨在处理大规模数据集，通过将数据处理任务分解为 Map 和 Reduce 阶段，并在多台计算节点上并行执行来提高处理速度。虽然 MapReduce 可以在多线程环境下运行，但其设计初衷是为了分布式环境下的并行计算，而不是单线程。在单线程环境下，传统的实现方式可能更加高效，因为单线程的开销比较小，而 MapReduce 框架需要额外的管理和调度开销。
4. 正确。在 MapReduce 框架中，InputFormat 模块负责读取输入数据，并将其分割成适当大小的数据块供 Map 函数处理。InputFormat 模块通常定义了数据的输入方式，比如从文件系统中读取、从数据库中读取等，以及数据的切分方式。

三、分析题

- (1) 这个不符合顺序一致性模型，因为写操作 $W(x)a$ 和 $W(x)b$ 都各只有一次，这说明 a 和 b 这两个值在每一个进程中能读取的顺序是一样的，要么都是 $R(x)a$ 、 $R(x)b$ ，要么都是 $R(x)b$ 、 $R(x)a$ 。但 P3 进程和 P4 进程的读取顺序是不一样的。
- (2) 这个符合顺序一致性模型，指令的交织顺序如下表（表 1）

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

表 1

四、计算题

(1)

会使用 5 个箱子，分别装入

0.2 0.5

0.4

0.7 0.1

0.3

0.8

(2)

会使用 4 个箱子，分别装入

0.2 0.5 0.1

0.4 0.3

0.7

0.8

五、程序设计题

function reduce(int n, Iterator partialCounts):

 for i from 2 to ceil(sqrt(n)): //sqrt(n)表示 n 的平方根 ceil()表示对数字向上取整

 if n%i == 0

 return

 end if

 end for

 sum = partialCounts(n) //等号后表示 n 在列表中的计数

 emit(n, sum)

end function