

1 押题

1.1 降低写放大倍数

降低写放大倍数的方法之一是使用写入补偿 (Write Amplification Reduction) 技术, 其中包括以下几个常见的策略:

1. 块重组 (Block Compaction): 通过将已有数据进行整理和重组, 使得块内的数据尽可能填满, 减少擦除的次数和写放大倍数。
2. 块级回收 (Block-level Garbage Collection): 在写入新数据时, 先执行擦除操作, 然后将需要保留的数据和新数据一起写入, 以减少擦除操作和提高写入效率。
3. 数据压缩 (Data Compression): 使用压缩算法对数据进行压缩, 在写入存储器之前对数据进行压缩, 可以减少存储空间的使用, 从而减少写入和擦除操作的次数。
4. 精细映射 (Fine Mapping): 使用更高级的映射算法来管理页面和块之间的映射关系, 可以减少数据的碎片化, 提高数据的写入效率和存储利用率。

1.2 2021设计题

2021 年 5 月 15 日, 天问一号火星探测器着陆火星, 人类火星探测任务成功率目前只有 50% 左右, 而 成功登录火星表面并顺利开展工作的探测任务成功率只有 20%。火星探测任务中技术难度最大, 失败率 最高的关键阶段是着陆器分离抛射到着陆器着陆这一阶段, 时间为 7 分钟, 这被科学家们称为 “黑色7分钟”, 即火星探测器进入火星大气的时速约为 21000 千米/小时, 要在 7.9 分钟内, 让速度在受控状态下 将为 0, 实现火星上安全着陆。考虑到从地球传信号到着陆器要 10 多分钟, 远大于 7 分钟着陆时限, 因 此在地球的工作人员无法操控, 只能依靠着着陆器自身设定系统和程序。这就要求, 在此期间的操作系统在处理输入, 提供输出之间不能有明显的延迟, 为完成着陆, 系统必须 有足够的弹性, 无论发生器件失效、系统错误还是其他错误, 都要能恢复过来, 这一阶段的导航、制 导、控制程序至为关键, 同时还会有系统各类部件/模块持续产生的各种数据流和对应任务。无独有偶, 根据新浪科技报导, 阿波罗 11 号登月时, 阿波罗计算机在登月着陆阶段的最后 4 分钟内连 续重启了 5 次, 所幸计算机在设计之时, 就考虑了优先级调度, 像制导和控制等关键程序优先级最高, 会持续运行, 它们可以被某些数据请求, 比如来自宇航员或地球的控制指令和数据请求所打断, 但该方案中的一个问题是, 程序如果中断太多次可能会堵塞 CPU, 导致系统崩溃。进一步地, 科学家提供了重 启保护 (restart protection) 的解决方案。如果发送给处理器的任务过多, 某些受保护的程序将会把他 们的数据吐入内存库, 然后处理器队列重置, 计算机立即重启, 恢复受保护的任 务并放弃其余任务。因此, 阿波罗计算机在登月着陆阶段的连续重启并未对制导和 控制等关键程序的运行产生致命影响。此外, 在太空探索实践中, 经常出现单粒子翻转 (SEU, Single Event Upsets), 即在太空环境下存在大 量高能带电粒子, 计算机中的电子元器件受到地球磁场、宇宙射线等照射, 引起电位状态的跳变, 0 变 成 1, 或者 1 变成 0。例如 1994 年 2 月 8 日发射的实践四号卫星上搭载的两台用于单粒子事件测量 的监测装置, 在入轨后的 19 天发生了 65 次翻转。这些事件可能造成姿态控制计算机程序混乱, 无法控 制飞行器姿态, 导致在空间翻转, 失控。为抵御单粒子翻转事件, 计算机器件往往采用抗辐照设计, 因 此性能普遍偏低, 现代抗辐照计算机其 CPU 主频往往只有几十兆到上百兆。除此之外, 提升计算性能的 先进制程, 可能受到单粒子影响更大, 而且随着位单元在较小的制程结 点中靠得越来越远, 单粒子翻转 可能会不止一位, 这对校验算法提出了更大的挑战。从以上

例子中，我们可以看到，针对大型科研探索任务，必须有科学合理的计算机系统，有步骤地进行任务调度、计算和存储，针对不同的任务，往往有不同的计算和存出方案

请根据以上材料，结合操作系统所学知识，从以下七个问题中任选其五来回答

1. 试分析阿波罗 11 号计算机采用**优先级调度**的原因，并从你的角度，试给出天问一号计算机响应任务调度设计，并说明其合理性
2. 从**存储的可靠性**出发，从你的角度，试给出在先进制程下，抵御单粒子翻转（可能会不止一位）的存储设计
3. 若天问一号分别采用**单核和多核 CPU**，从你的角度分析，其任务调度设计会有何异同
4. 若天问一号**遭遇异常重启**，请给出重启保护的设计方案
5. 抗辐射器件往往价格昂贵，性能偏低，若天问一号采用民用 CPU（无抗辐照设计），如何在任务执行过程中抵御单粒子翻转（可能会不止一位），保证结果的正确性
6. 抗辐射器件往往价格昂贵，性能偏低，若天问一号采用民用存储（无抗辐照设计），如何抵御单粒子翻转（可能会不止一位），保证**存储内容的一致性和正确性**
7. 假设天问一号采用了 **NAND 存储芯片**，要延长其工作寿命，必然要延长其存储颗粒的寿命，请给出响应的存储颗粒磨损平均化的设计

1.阿波罗11号计算机采用优先级调度的原因是确保关键程序（如制导和控制）的持续运行，优先级最高，不会被其他任务的请求打断。这是因为在登月着陆阶段，这些关键程序的连续运行非常重要，任何中断或延迟可能导致系统崩溃或任务失败。通过优先级调度，这些关键程序能够优先获得CPU的执行时间，保证任务的实时性和可靠性。

对于天问一号计算机，响应任务调度的设计可以采用实时操作系统的概念。关键程序和任务可以设置较高的优先级，确保它们能够及时地响应和执行。同时，使用适当的调度算法（如优先级调度算法）来管理任务的执行顺序，确保关键任务能够及时获得处理器的时间片，并满足其响应时间要求。对于非关键任务，可以采用较低的优先级，以保证关键任务的执行。

这样的设计可以确保天问一号计算机在关键阶段能够高效地执行任务，保证系统的稳定性和可靠性。

2.在先进制程下抵御单粒子翻转的存储设计可以采用冗余校验码（ECC）技术。ECC是一种用于检测和纠正存储器中位翻转错误的技术。它通过在存储数据中添加冗余信息来实现错误检测和纠正能力。

具体实施时，可以在存储数据中添加校验位，计算并存储校验和。在读取数据时，通过校验位和校验和进行数据的校验，如果检测到错误，则可以使用校验和进行错误的纠正。通过使用ECC技术，即使在发生单粒子翻转的情况下，也能保证存储数据的正确性和一致性。

3.天问一号采用单核和多核CPU时，任务调度设计会有一些异同之处。在单核CPU中，任务调度主要是根据任务的优先级和调度算法来决定任务的执行顺序。而在多核CPU中，可以采用并行调度策略，将不同的任务分配给不同的核心进行并行执行。

在单核CPU中，任务的执行是串行的，每个任务依次执行，而在多核CPU中，多个任务可以同时执行，提高系统的处理能力和效率。

然而，在任务调度设计上，无论是单核还是多核CPU，都需要考虑任务的优先级、任务的依赖关系、任务的并发性等因素，以保证任务能够按时执行，并满足系统的实时性和可靠性要求。

4.若天问一号遭遇异常重启，可以采用重启保护的设计方案。当发生异常重启时，可以将当前正在执行的任务的状态和数据保存到非易失性存储器中，以确保重启后能够恢复到异常发生前的状态。

具体实施时，可以在系统中引入一个监控模块，负责检测系统的异常情况。当监测到异常时，监控模块会触发重启保护机制，将当前任务的状态和数据保存到非易失性存储器中。在系统重启后，可以从非易失性存储器中读取保存的状态和数据，恢复到异常发生前的状态，并继续执行任务。

这样的设计方案可以提高系统的容错性，保证任务的连续执行，并减少异常重启对系统的影响。

5.若天问一号采用民用CPU（无抗辐照设计），可以采用软件方法来抵御单粒子翻转，保证结果的正确性。一种方法是通过冗余计算和校验来检测和纠正位翻转错误。

具体实施时，可以在关键数据的计算过程中引入冗余计算，并进行数据校验。例如，通过在计算过程中多次执行相同的计算，并对计算结果进行校验，以检测和纠正可能的错误。

另一种方法是使用重复执行和投票机制。通过多次执行同一任务，并将执行结果进行比较和投票，选择多数投票结果作为最终结果。这样可以在发生错误时纠正错误结果，并保证最终结果的正确性。

这些软件方法可以在抵御单粒子翻转的同时，提高系统的容错性和可靠性。

6.若天问一号采用民用存储（无抗辐照设计），可以采用冗余存储和数据校验的方法来抵御单粒子翻转，保证存储内容的一致性和正确性。

具体实施时，可以将存储数据进行冗余存储，存储多个副本，并使用校验和等校验机制来检测和纠正可能的位翻转错误。通过比较不同副本之间的数据，并使用校验和进行校验，可以发现和纠正存储数据中的错误。

另一种方法是采用磁盘阵列（RAID）技术，将存储数据分布在多个物理存储设备上，并使用冗余数据和校验机制来保护存储数据的完整性。RAID技术可以检测和纠正存储设备中的错误，并保证存储内容的一致性和正确性。

这些方法可以在使用民用存储设备的情况下，提高存储的可靠性和抵御单粒子翻转的能力。

7.若天问一号采用了NAND存储芯片，并希望延长其工作寿命，可以采用存储颗粒磨损平均化的设计。存储颗粒磨损平均化是通过在使用存储颗粒时均匀分布磨损，减少部分颗粒的使用频率，从而延长整体存储器的寿命。

具体实施时，可以采用动态均衡算法来调整存储颗粒的使用频率。通过监测存储颗粒的使用情况，将频繁使用的颗粒转移到不频繁使用的颗粒上，使得存储颗粒的磨损更加均匀。

另一种方法是采用空间换时间的思想，使用更大容量的存储器。通过增加存储器的容量，可以减少每个存储颗粒的使用频率，从而延长整体存储器的寿命。

这些设计方案可以帮助延长NAND存储芯片的工作寿命，提高系统的可靠性和稳定性。

1.3 2022设计题

三、分析设计题（共30分）

在深圳证券交易所2019年推出的第五代交易系统中，其服务器集群需要管理和维护每日交易订单和持仓记录（例如查询、修改、删除、新增）。假设服务器集群为100台服务器，每台服务器有2颗处理器，每颗处理器24核，其上可以运行若干进程（线程），这些进程需要对这些订单 / 记录进行查询、修改、删除、新增操作。该系统的订单委托处理能力约为300000笔 / 秒，紧随NASDAQ和德交所，委托交易时延1.1毫秒。

国家关键基础设施需要有科学合理的计算机系统设计来进行任务调度、计算和存储。而面向不同的任务场景，往往会有不同的计算、存储和I / O设计方案。考虑到系统对于实时性、可靠性和高性能的要求，请对以下两个案例分别进行分析（任选5小题回答，每小题3分，每个案例15分）

Case 1：服务器集群需要管理的每日交易订单和持仓记录为500万条 Case 2：服务器集群需要管理的每日交易订单和持仓记录为50亿条

- (1) 请评估CPU性能要求和内存要求，以及每个进程的任务负载
- (2) 试从你的角度，给出该交易系统的任务调度算法，并简要说明其合理性
- (3) 考虑到订单和持仓记录可以按用户和股票代码进行分类，请根据局部性原理，给出订单 / 持仓记录数据的内存分布设计
- (4) 从数据一致性出发，请给出数据的存储一致性设计
- (5) 从2006年5月，全球证券市场共发生了40余起技术故障导致的交易异常事件，平均每年3起以上，其中大体可以把威胁数据可靠性的因素归为以下几类：硬件故障、软件隐患、运维故障。请从数据存储的可靠性出发，请给出数据存储设计方案
- (6) 从高可用性出发，请给出集群中单点失效的解决方案
- (7) 从低延时出发，请给出订单 / 持仓数据的内存无锁并发访问策略
- (8) 假设该系统采用了NAND存储芯片，若要延长其工作寿命，必然需要延长其存储颗粒的寿命，请给出相应的存储颗粒磨损平均化的设计

1. 评估CPU性能要求和内存要求，以及每个进程的任务负载：

- Case 1：服务器集群需要管理的每日交易订单和持仓记录为500万条。假设每个进程处理能力为1000笔/秒，那么需要大约3000个进程来处理这些订单和记录。对于CPU性能要求和内存要求，需要根据进程的处理能力、并发量以及数据结构的大小进行进一步评估。
- Case 2：服务器集群需要管理的每日交易订单和持仓记录为50亿条。假设每个进程处理能力为1000笔/秒，那么需要大约1500万个进程来处理这些订单和记录。对于CPU性能要求和内存要求，需要更强大的计算和存储资源来满足更大规模的数据处理和管理需求。

2. 任务调度算法的设计：考虑到实时性、可靠性和高性能的要求，可以采用优先级调度算法。根据订单和记录的紧急程度、时间限制和系统负载情况，将优先级高的任务优先调度到可用的处理器上执行。合理性体现在根据任务的特性和重要性进行任务调度，以最大程度满足系统的需求。
3. 订单和持仓记录数据的内存分布设计：基于局部性原理，可以按用户和股票代码进行分类，并将具有相同用户和股票代码的订单和持仓记录放置在相邻的内存区域。这样可以提高数据访问的效率和系统的响应速度。
4. 数据的存储一致性设计：为确保数据的一致性，可以采用事务机制和并发控制技术，如锁机制、MVCC（多版本并发控制）等。通过合理地设计和管理事务的边界和隔离级别，以及采用适当的并发控制策略，可以保证交易系统中的数据操作的一致性和正确性。
5. 数据存储设计方案的可靠性：为提高数据存储的可靠性，可以采用多副本备份和数据冗余技术。通过将数据复制到多个存储节点上，实现数据的冗余存储和容灾备份，以应对硬件故障、软件隐患和运维故障等可能的风险和故障。
6. 集群中单点失效的解决方案：为提高系统的高可用性，可以采用故障转移和负载均衡技术。配置多个服务器节点，当某个节点发生故障时，其他节点可以接替其工作，确保系统的连续性和可用性。同时，通过负载均衡技术，将任务均匀地分配给各个节点，提高系统的整体性能和容错能力。
7. 订单/持仓数据的内存无锁并发访问策略：为实现低延时和高并发访问，可以采用无锁数据结构和乐观并发控制策略。通过使用无锁数据结构和CAS（比较和交换）操作，避免了传统锁机制的开销和竞争，提高系统的并发性能和响应速度。
8. 存储颗粒磨损平均化的设计：对于NAND存储芯片，为延长其工作寿命，可以采用均衡化的磨损分布策略。通过动态均衡算法，根据存储颗粒的使用频率和磨损情况，调整数据的存储位置，使得存储颗粒的磨损更加均匀，延长整体存储器的寿命。