

三、分析设计题 (共 30 分)

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仅有 50%左右,而成功登陆火星表面并顺利开展工作的探测任务成功率则只有 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降为 0,实现在火星上安全着陆。考虑到从地球传信号到着陆器要 10 多分钟,远大于 7 分钟的着陆时间,因此在地球的工作人员无法操控,只能依靠着陆器自身设定的系统和程序。

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。所幸其计算机在设计之初,就考虑了优先级调度,像制导和控

制等关键程序优先级较高,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它们的数据吐入内存,而后处理器队列重置,计算机立即重启,恢复受保护的程序并放弃其余任务。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),即在太空环境中存在着大量高能带电粒子,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 1995 年 2 月 8 日发射升空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 天内共发生了 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。为了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节点中靠得越来越近,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,针对大型科研探索实践,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储方案。

请根据以上资料,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相应的任务调度设计,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) 抗辐照器件往往价格昂贵,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) 抗辐照器件往往价格昂贵,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,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

三、分析设计题 (共30分)

FAST (Five-hundred-meter Aperture Spherical radio Telescope),即500米口径球面射电望远镜,又被称为“中国天眼”,是由国家科教领导小组审议确定的国家九大科技基础设施之一,坐落于我国贵州省平塘县克度镇大窝凼洼地,是目前世界上口径最大、最灵敏的单口径射电天文望远镜。

近日中国科学家团队利用FAST观测数据,发现了纳赫兹引力波存在的关键证据,使纳赫兹引力波研究进入了新时代。中国天眼FAST在试运行期间,一年产生的数据大约在15PB左右,每秒产生3GB数据。在采用L波段的19波束馈源接收机之后,其每秒最高传输基带数据从此前3GB左右,提升到每秒38G,平均每小时接收到的有效科学数据近3.6TB。随着FAST步入常规运行,观测数据量和科学目标的持续增加,FAST多科学目标巡天每天将会产生约500TB的零级未压缩数据。

从以上材料可以看出,国家关键大科学装置必须有科学合理的计算机系统设计来支撑,从而有步骤地执行调度、计算、存储和I/O任务。请从以下七个问题中任选其五,从存储要求、延迟要求、I/O要求、调度要求、同步要求、任务特性,任务的并发性等方面考虑

- (1) 试分析中国天眼的数据存储需求,并给出相应的兼顾高性能访问和可靠性的存储设计
- (2) 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
- (3) 考虑到系统处理数据时的时间开销大多来自于数据拷贝和上下文切换,请给出针对数据访问和存储的低延迟设计
- (4) 从高可用性出发,请给出存储集群中单点失效的存储解决方案
- (5) 从高吞吐量出发,请给出能够支持高速并发读写的数据I/O设计
- (6) 越近发布的数据越常被访问 vs. 所有数据的访问频率均等。试说明不同条件会导致数据存储分布和数据访问的具体设计有何不同
- (7) 数据的访问模式为:研究者们会并发访问某些数据文件(如研究者们合作分析某些数据,多对多) vs. 单个研究者一次只会访问一个数据文件(如独立研究者,一对一)。请根据所学知识,给出适用于读请求场景的多进程并发竞争设计,以及适用于写请求(修改数据)场景的多进程并发竞争设计

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方案来进行任务调度、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)

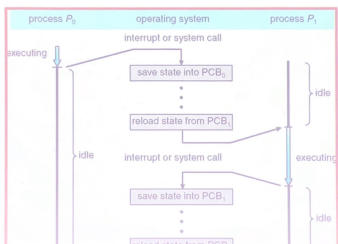
Case 1: 服务器集群需要管理的每日交易订单和持仓记录为50万条

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

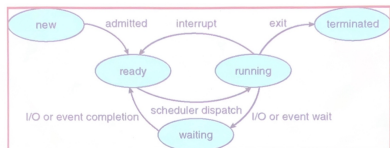
- (1) 请评估CPU性能要求和内存要求,以及每个进程的任务负载
- (2) 试从你的角度,给出该交易系统的任务调度算法,并简要说明其合理性
- (1) 请评估CPU性能要求和内存要求,以及每个进程的任务负载
- (2) 试从你的角度,给出该交易系统的任务调度算法,并简要说明其合理性
- (3) 考虑到订单和持仓记录可以按用户和股票代码进行分类,请根据局部性原理,给出订单/持仓记录数据的内存分布设计
- (4) 从数据一致性出发,请给出数据的存储一致性设计
- (5) 从2006年5月,全球证券市场共发生了40余起技术故障导致的交易异常事件,平均每年3起以上,其中大体可以把威胁数据可靠性的因素归为以下几类: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

一、简答题 (共35分)

- (6分) 请简要说明单道批处理系统、分时系统和你日常使用的操作系统(多道程序系统, 例如Linux、Windows或MAC) 的特点, 并简要分析其优缺点
- (2分) 请从操作系统结构设计角度, 简要分析层次结构和微内核结构的优缺点
- (3分) 请以系统调用read(file)为例, 结合下图来简要说明中断处理及其上下文切换过程, 并简要说明操作系统在设计上引入中断的原因



- (4分) 小明的系统中有一些计算密集型进程, 也有一些交互型进程, 各个进程用时长短不一, 小明希望长工作之间可以公平的分享CPU, 又能给短工作或者交互型工作很好的响应时间. 请结合下图给小明介绍进程状态模型和调度时机; 并根据你的理解, 从以下5种



算法中为小明推荐一个综合性能较好的进程调度算法, 并简要给出你的推荐依据

2022--2023 学年第二学期 共5页

(5) 多级队列反馈调度算法 (MLFQ)

- (4分) 请就操作系统中的同步、互斥和死锁概念, 从日常生活中给出类比例, 并简要分析操作系统在设计上引入临界区/临界资源的原因
- (2分) 小明说, 网页浏览器程序(如Chrome/Firefox浏览器程序)应该采用多线程的方式来运行, 请根据进程和线程的区别和联系, 分析小明上述话语的对错
- (4分) 请简要说明连续存储空间管理和分页式存储管理各自所面临的主要性能瓶颈/挑战, 并说明现代操作系统设计中是如何解决相关性能瓶颈/挑战的
- (2分) 请简要说明局部性原理, 并结合下图来简要分析现代操作系统通常采用层次化存储结构来进行存储管理的原因
- (2分) 某系统采用分页式存储管理, 有一作业大小是8KB, 页大小为2KB, 相关页面依次

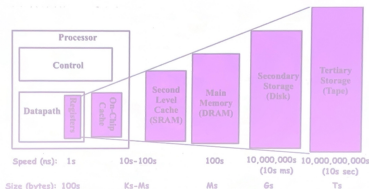
多线程的线程

图片类进程

文本类进程

;

每类进程有多个线程



装入内存的第7、9、B、5块/帧, 试将逻辑地址0ABCH, 1BDDH转换成物理地址。

- (2分) 小明计划配置一个Raid存储阵列, 购买了16个大小相同的同款硬盘, 要求
 - 相比单硬盘存储整个文件的方式, 大文件的访问速度需快10倍以上
 - 存储可靠性比单个硬盘好
 - 单块硬盘损坏不应影响数据安全
 小明希望你帮助完成设计, 请给出相关设计并简要说明其基本特性。

- (4分) 6月份中区附近打印业务繁忙, 小明计划开设一家打印店承接打印业务, 添置了一批打印机, 希望你对其店内的操作系统及其打印资源进行分析和升级
 - 如果打印店可以被抽象为x个进程共享的具有y个同类资源的系统, 其满足:
 - 对 $i=1, 2, 3, \dots, x$, 进程 P_i 至少需要1个资源, 最多需要 y 个资源;
 - 在任意时刻, 所有进程对资源的需求量之和小于 $x+y$
 试向小明证明, 该系统抽象是死锁无关的。
 - 如果打印机一次只能服务于一个请求,
 - 如何使一台打印机变成共享设备?
 - 请给出采用询问(程序直接控制)、中断、DMA、通道方式的打印工作流程

二、问答题 (共35分)

1. (5分) 科大中区深夜食堂有多个档口, 每个档口有一位服务员。当档口没有师生来打饭菜时, 服务员会空闲; 若有师生来打饭菜, 服务员则开始为师生服务; 每个服务员同一时刻最多只能服务一位师生。当服务员服务时, 若有更多师生到来, 后来的师生需按到达时间来依次排队。

1. 请找出以上服务过程中的同步与互斥关系 (1分)
2. 请以信号量的方式来描述以上服务过程, 并给出相应的信号量定义 (伪代码即可) (2分)
3. 如果科大中区深夜食堂每个档口的服务员空闲时, 可以帮相邻档口的排队师生打饭菜 (即该档口最多可以有3位服务员同时服务3位师生), 这时候繁忙档口的师生排队情况可以大大缓解, 请再次给出以上服务过程的伪代码描述 (2分)

2. (5分) 根据作业到达时间和作业长度, 请分别用先进先出 (FIFO) 调度算法、时间片轮转 (RR) 调度算法 (假定时间片大小为5)、最短剩余时间优先 (SRTF) 调度算法、优先级调度算法 (Non-Preemptive) 对以下作业进行调度, 并计算

(1) 各调度算法的作业完成时间

(2) 各调度算法的**平均周转时间** (周转时间: 从作业被提交给系统开始, 到作业完成为止的这段时间间隔称为作业周转时间。平均周转时间 = 作业周转总时间 / 作业个数)

参数				调度算法完成时间			
作业名	作业长度	作业到达时间	作业优先级	FIFO	RR	SRTF	PR (Non-Preemptive)
P0	35	0	2				
P1	10	5	4				
P2	35	10	5 (最低)				
P3	10	50	1 (最高)				
P4	5	80	3				
平均周转时间							

4. (5分) 在一个32位请求分页式系统中, 页表均存储在内存中, 磁盘访问时间为5ms, 内存访问时间为200ns, TLB访问时间为 x ns。如果TLB访问命中, 则访问时间为 $x + 200$ ns, 即TLB访问 (获取数据所在帧号) 和一次数据访问 (假定TLB命中时对应数据位于内存)。

给定TLB Hit命中率为 a ($a \leq 1$), 则TLB miss的概率是 $(1-a)$, 其中又有10%的概率会出现缺页 (数据不在内存中, 缺页中断)。如果系统采用单级页表, 因此通过页表进行数据读写会导致至少两次访存, 即一次页表访问 (用于获取数据所在的帧号) 和一次数据访问, 如果数据位于内存, 则时间为 $x + 200 + 200 = x + 400$ ns。

(1) 假设TLB访问时间可以忽略不计, 请问当 x 分别为99.99%和99.9999%时, 该系统在采用单级页表和两级页表时的有效内存访问时间。

(2) 假设TLB访问时间为10ns, 请问当 x 分别为99.99%和99.9999%时, 该系统在采用单级页表和两级页表时的有效内存访问时间。

5. (5分) 给定数据请求地址序列为 (0073, 0335, 0210, 0632, 0614, 0608, 0222, 0068, 0130), 例如0103表示请求访问第103字节位置的数据。假设页大小为100字节, 则0103数据请求地址将要访问的页面为第1页, 0068则对应第0页 (注: 页面序号从第0页开始)

(1) 请给出对应的页面访问序列

(2) 当内存的物理帧数为2时, 采用LRU页替换算法产生的缺页中断次数为?

(3) 当内存的物理帧数为4时, 采用LRU页替换算法产生的缺页中断次数为?

6. (5分) 有一个矩阵为100行, 200列, 即:

```
var A: array[1..100, 1..200] of integer;
```

在一个采用LRU页替换算法的虚拟存储管理系统, 系统分给该进程十个页面来存储数据 (不包含程序), 设每页的大小可以存放200个整数, 该程序要对整个数组初始化, 数组存放时是按行存放的。试计算下列两个程序各自的缺页次数。(假定所有页都以请求方式调入)

```
程序1:          程序2:  
for i=1 to 100 do      for j:=1 to 200 do  
  for j:=1 to 200 do    for i:=1 to 100 do  
    A[i, j]:=i*j;      A[i, j]:=i*j;
```

7. (5分) 磁盘的I/O请求队列的柱面号依次为9, 73, 46, 132, 114, 72, 68, 87, 磁头初始位置为50, 假设磁盘有200个柱面号 (0-199), 磁头沿移动臂每移动一个柱面 (磁道) 需时1ms, 请根据以下调度算法, 给出以上请求序列的平均寻道时间

(1) 先来先服务 (FCFS)

(2) 电梯调度算法 (SCAN扫描算法, 磁头当前向磁道0移动)

重点: CPU管理, 内存管理 (PPT)

优先做最后一题