作业三 处理机调度与死锁

1. **什么是处理机高级调度？什么是低级调度？什么是中级调度？它们分别具有哪些主要功能？**

高级调度又称长程调度或作业调度，它的调度对象是作业。其主要功能是根据某种算法，决定将外存上处于后备队列中的哪几个作业调入内存，为它们创建进程、分配必要的资源，并将它们放入就绪队列。

低级调度又称为进程调度或短程调度，它所调度的对象是进程（或内核级线程）。其主要功能是，根据某种算法，决定就绪队列中的哪个进程应获得处理机，并有分派程序将处理机分配给被选中的进程。

中级调度又称为内存调度。目的是提高内存利用率和系统吞吐量。

1. **处理机调度算法共同目标中的平衡性是什么含义？**

由于在系统中可能具有多种类型的进程，有的属于计算型作业，有的属于I/O型。为使系统中的CPU和各种外部设备都能经常处于忙碌状态，调度算法应尽可能保持系统资源使用的平衡性。

1. **什么是周转时间？平均周转时间？带权周转时间？平均带权周转时间？**

周转时间是指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。

平均周转时间：

带权周转时间，即作业的周转时间T与系统为它提供服务的时间Ts之比，即W = T/Ts。

平均带权周转时间则可表示为：

1. **对于批处理系统、分时系统、实时系统，处理机调度算法的目标分别是什么？**

批处理系统的目标(1)平均周转时间短(2)系统吞吐量高(3)处理机利用率高

分时系统的目标(1)响应时间快(2)均衡性

实时系统的目标(1)截止时间的保证(2)可预测性

1. **作业运行有哪三个阶段，分别是什么含义？这三个阶段对应哪三种状态？作业在运行阶段，系统为其分配必要的资源和建立进程，这些进程放入什么队列？**

收容阶段：操作员把用户提交的作业通过某种输入方式或SPOOLing系统输入到硬盘上，再为该作业建立JCB，并把它放入作业后备队列中。此时状态为“后备状态”。

运行阶段：当作业被作业调度选中后，便为他分配必要的资源和建立进程，并将它放入就绪队列，此时为“运行状态”

完成阶段：当作业运行完成、或发生异常情况而提前结束时，作业便进入完成阶段，相应的作业状态为“完成状态”。

放入就绪队列

1. **FCFS、SJF分别是什么作业调度算法？高响应比优先调度算法中的优先级是如何计算的？既考虑作业的运行时间，又考虑作业的等待时间，这是指哪种算法？**

先来先服务(FCFS)和短作业优先(SJF)调度算法



高响应比优先调度算法(Highest Response Ratio Next，HRRN)

1. **作业和进程有何不同？它们之间有什么关系？**
2. 不同： 作业：是用户在一次上机活动中，要求计算机系统所做的一系列工作的集合。也称作任务（task）。 进程：是一个具有一定独立功能的程序关于某个数据集合的一次可以并发执行的运行活动。 作业是一个宏观的执行单位，它主要是从用户的角度来看待的。作业的运行状态是指把一个作业调入内存，然后产生若干个进程可以去竞争CPU。 进程是微观的执行单位，它主要从系统的角度来看待的，它是抢占CPU和其他资源的基本单位。进程的执行状态是指一个进程真正占用了CPU。

（2）关系：一个作业调入内存以后，处于执行状态，则此作业对应在系统建立若干个进程。进程的所有状态对应作业的执行状态，通过这若干个进程的执行，来完成该作业。

1. **进程有哪两种调度方式？分别是何含义？**

1) 非抢占方式：在采用这种调度方式时，一旦把处理机分配给某进程后，就一直让它运行下去，决不会因为时钟中断或任何其它原因去抢占当前正在运行进程的处理机，直至该进程完成，或发生某事件而被阻塞时，才把处理机分配给其它进程。

2) 抢占方式：这种调度方式允许调度程序根据某种原则，去暂停某个正在执行的进程，将已分配给该进程的处理机重新分配给另一进程

1. **进程时间片轮转调度算法中，时间片的大小对系统性能有何影响？**

若选择很小的时间片，将有利于短作业。但时间片小，意味着会频繁的执行进程调度和进程上下文的切换，会增加系统的开销。

若时间片选择的太长，且为使每个进程都能在一个时间片内完成，RR算法将退化为FCFS算法，无法满足短作业和交互式用户的需求。

1. **进程优先级调度算法中，什么是静态优先级？什么是动态优先级？**

静态优先级是在创建进程时确定的，在进程的整个运行期间保持不变。

动态优先级是指在创建进程之初，先赋予其一个优先级，然后其值随进程的推进或等待时间的增加而改变，以便获得更好的调度性能。

1. **什么是死锁？产生死锁的必要条件有哪些？分别是什么含义？**

如果一组进程中的每一个进程都在等待仅由该组进程中的其他进程才能引发的事件，那么该组进程是死锁的。

(1) 互斥条件：进程对所分配到的资源进行排它性使用，即在一段时间内，某程序只能被一个进程占用。  
(2) 请求和保持条件：进程已经保持了至少一个资源，但又提出了新的资源请求，而该资源已被其他进程占有，此时请求进程被阻塞，但对自己已获得的资源保持不放。  
(3) 不可抢占条件：进程已获得的资源在未使用完之前不能被抢占，只能在进程使用完时由自己释放。  
(4) 循环等待条件：在发生死锁时，必然存在一个进程——资源的循环链。

1. **处理死锁的方法有哪些？分别是什么含义？**

1）、预防死锁——设置某些限制条件，去破坏产生死锁的四个必要条件之一或几个来预防死锁。

2）、避免死锁——资源动态分配过程中，用某方法防止系统进入不安全状态。

3）、检测死锁——允许发生死锁，但通过系统设置的检测机构，检测死锁的发生，采取措施，把进程从死锁中解脱出来。

4）、解除死锁——当检测到死锁时，采取措施，将进程从死锁状态下解脱。

1. **讨论避免死锁时，什么是安全状态？什么是不安全状态？**

安全状态：在该方法中，允许进程动态地申请资源，但系统在进行资源分配之前，应先计算此次资源分配的安全性。

不安全状态：如果不按照安全序列分配资源，则系统可能会由安全状态进入不安全状态。

1. **什么是死锁定理？**

用资源分配图加以简化的方法来检测系统是否处于死锁状态。S为死锁状态的充分条件是，当且仅当s状态的资源分配图是不可完全简化的。该充分条件称为死锁定理。

1. **资源分配图如图所示，系统是否处于死锁状态？**

• •

•

• •

•

•

**r1 r2 r3 r4**

答：对该图进行化简，得到如下图所示的结果。由于该图是不可完全简化的，所以根据死锁定理，系统处于死锁状态。

• •

r1 r2 r3 r4

•

•

• •

•

**16、设系统中仅有一类数量为M的独占型资源，系统中有N个进程竞争该类资源，其中各进程对该类资源的最大需求量为W。当M、N、W分别取下列值时，试判断哪些情形可能会发生死锁，为什么？**

**(1)M=2，N=2，W=1； (2)M=3，N=2，W=2；**

**(3)M=3，N=2，W=3； (4)M=5，N=3，W=2；**

1.不会

2.不会

3.会，因为有两个进程，当一个进程拿到2个资源，另一个拿到1个资源时，会互相等待，发生死锁。

4.不会

**17、系统中有3个进程，4个相同类型的资源，每个进程最多需要2个资源，该系统是否会发生死锁？为什么？**

答：该系统不会发生死锁。因为4个资源分配给3个进程，无论如何分配，总会有1个进程能够分配到2个资源，该进程获得其最大资源数后，完成并释放其资源，剩余2个进程就可获得最大资源数，顺利完成，系统始终存在安全序列，故系统不会死锁。

**18、设系统中有五个进程、3种资源，总数分别为A 17，B 5，C 20，T0时刻系统状态如下。**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **最大资源需求** | | | **已分配资源** | | | **剩余资源数** | | |
| **A** | **B** | **C** | **A** | **B** | **C** | **A** | **B** | **C** |
| **P1** | **5** | **5** | **9** | **2** | **1** | **2** | **2** | **3** | **3** |
| **P2** | **5** | **3** | **6** | **4** | **0** | **2** |  | | |
| **P3** | **4** | **0** | **11** | **4** | **0** | **5** |
| **P4** | **4** | **2** | **5** | **2** | **0** | **4** |
| **P5** | **4** | **2** | **4** | **3** | **1** | **4** |
|  | | | | **15** | **2** | **17** |  | | |

* + 1. **完成剩余资源数的计算：**
    2. **T0时刻是否安全？**
    3. **若P2请求资源（0，3，4），系统如何处理？**

解答：T0时刻的向量见图中粗体数字。

need[i,j]=max[i,j]-allocation[i,j]

利用银行家算法对此资源分配情况进行分析，可得此时刻的安全性分析情况：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Work | | | Need | | | Allocation | | | Work+allocation | | | Finish |
| P4 | 2 | 3 | 3 | **2** | **2** | **1** | 2 | 0 | 4 | 4 | 3 | 7 | True |
| P5 | 4 | 3 | 7 | **1** | **1** | **0** | 3 | 1 | 4 | 7 | 4 | 11 | True |
| P1 | 7 | 4 | 11 | **3** | **4** | **7** | 2 | 1 | 2 | 9 | 5 | 13 | True |
| P2 | 9 | 5 | 13 | **1** | **3** | **4** | 4 | 0 | 2 | 13 | 5 | 15 | True |
| P3 | 13 | 5 | 15 | **0** | **0** | **6** | 4 | 0 | 5 | 17 | 5 | 20 | True |

因为T0时刻存在安全序列p4,p5,p1,p2,p3，故T0时刻安全。

按照银行家算法，在T0时刻P2请求资源（0，3，4），

因请求资源数（0，3，4）<最大请求资源数（1，3，4），继续。

请求资源数（0，3，4）>剩余资源数（2，3，3），所以系统没有足够的资源，不能分配。

**19、在一个单道批处理系统中，一组作业的到达时间和运行时间如下表所示。试计算使用先来先服务、短作业优先、高响应比优先算法时的平均周转时间和平均带权周转时间。**

|  |  |  |
| --- | --- | --- |
| **作业** | **到达时间** | **运行时间** |
| **1** | **8.0** | **1.0** |
| **2** | **8.5** | **0.5** |
| **3** | **9.0** | **0.2** |
| **4** | **9.1** | **0.1** |

采用先来先服务作业调度算法时，作业的运行情况如下表所示：

作业执行次序 提交时间 运行时间 开始时刻 完成时刻 周转时间 带权周转时间   
1 8:00 1.0 8:00 9:00 1.0 1.0   
2 8:50 0.5 9:00 9:30 0.67 1.34   
3 9:00 0.20 9:30 9:42 0.7 3.5   
4 9:10 0.10 9:42 9:48 0.63 6.3   
所以，平均周转时间为：  
T=(1.0+0.67+0.7+0.63)/4=0.75  
平均带权周转时间为：  
W=(1.0+1.34+3.5+6.3)/4=3.04  
(2)采用短作业优先调度算法时，作业的运行情况如下表所示：  
作业执行次序 提交时间 运行时间 开始时刻 完成时刻 周转时间 带权周转时间   
1 8:00 1.0 8:00 9:00 1.0 1.0   
2 9:00 0.2 9:00 9:12 0.2 1.0   
3 9:10 0.10 9:12 9:18 0.133 1.33   
4 8:50 0.50 9:18 9:48 0.97 1.94   
(在9:00时，作业2和作业3就绪，作业3因为时间短而先执行。同理，9:12时作业4开始执行。)  
所以，平均周转时间为：  
T=(1.0+0.2+0.1+0.97)/4=0.57  
平均带权周转时间为：  
W=(1.0+1.0+1.33+1.94)/4=1.32  
(3)采用响应比高者优先作业调度算法时，作业的运行情况如下表所示：  
作业执行次序 提交时间 运行时间 开始时刻 完成时刻 周转时间 带权周转时间   
1 8:00 1.0 8:00 9:00 1.0 1.0   
2 9:00 0.20 9:00 9:12 0.2 1.0   
3 8:50 0.50 9:12 9:42 0.87 1.74   
4 9:10 0.10 9:42 9:48 0.63 6.3   
所以，平均周转时间为：  
T=(1.0+0.2+0.87+0.63)/4=0.675  
平均带权周转时间为：  
W=(1.0+1.0+1.74+6.3)/4=2.51

**20、P1，P2，P3，P4四个进程同时依次进入就绪队列，它们所需要的处理器时间和优先数如下，若不计调度等所消耗的时间，请回答：**

**进程 处理器时间（秒） 优先数**

**P1 20 2**

**P2 15 3**

**P3 10 5**

**P4 12 3**

**①分别写出采用先来先服务和非抢占式的优先数调度算法时进程执行的次序。**

**②分别计算每个进程在就绪队列中的等待时间和平均等待时间。**

解答：（a）进程执行次序为：

|  |  |
| --- | --- |
| 先来先服务法 | 非抢占式的优先数法 |
| P1、P2、P3、P4 | P3、P2、P4、P1 |

（b） 先来先服务法：

每个进程在就绪队列的等待时间分别为

P1：0（秒）

P2：0+20=20（秒）

P3：20+15=35（秒）

P4：35+10=45（秒）

平均等待时间为（0+20+35+45）/4=25（秒）

非抢占式的优先数法：

每个进程在就绪队列的等待时间分别为

P1：25+12=37（秒）

P2：0+10=10（秒）

P3：0（秒）

P4：10+15=25（秒）

平均等待时间为（37+10+0+25）/4=18（秒）