



3.1.1 处理机调度的层次

【三级调度模型如图】

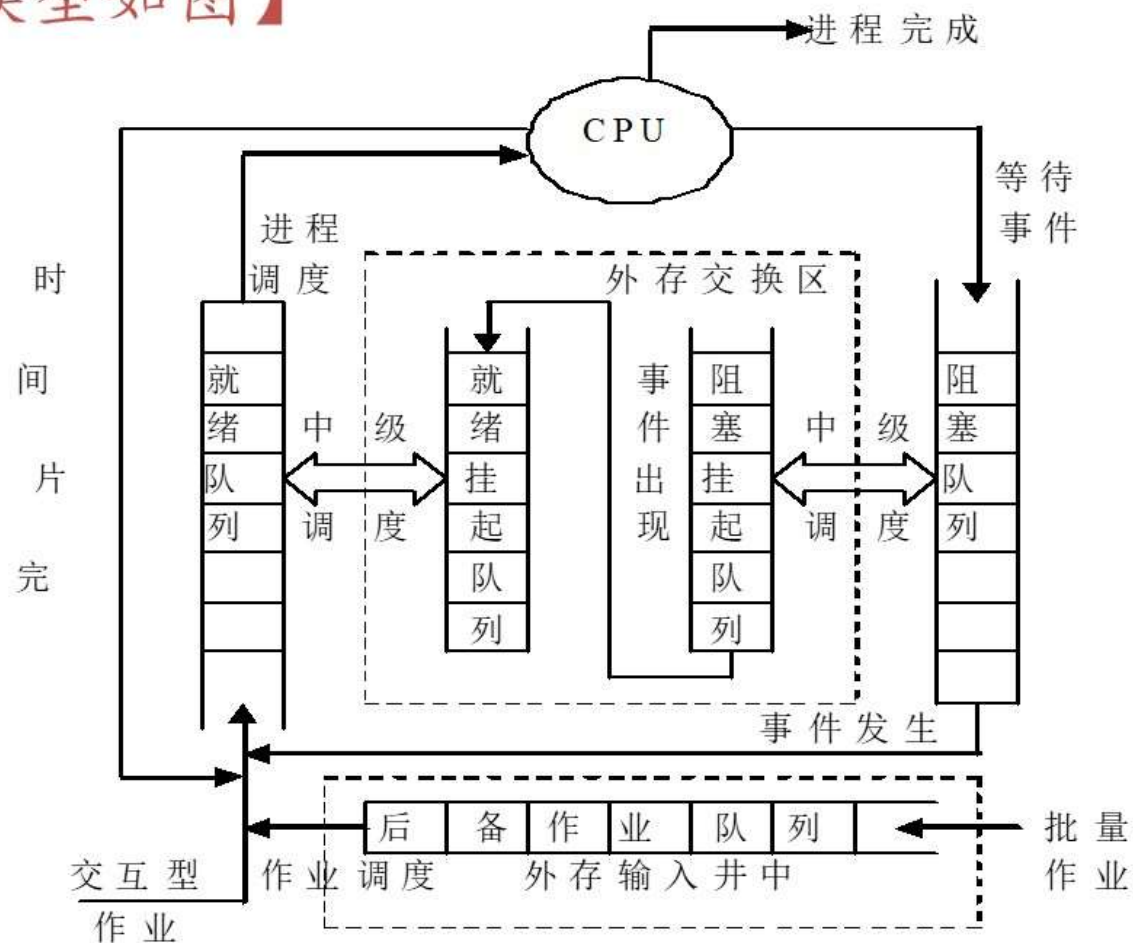


图 三级调度模型



3.1.2 处理机调度算法的目标

在操作系统设计中如何选择调度方式和算法，通常取决于操作系统的类型及其设计目标。

1. 处理机调度算法的共同目标

(1) 资源利用率 系统最重要的资源就是CPU。

$$\text{CPU的利用率} = \frac{\text{CPU的有效工作时间}}{\text{CPU的有效工作时间} + \text{CPU空闲等待时间}}$$

(2) 公平性 公平性是指所有的进程都获得合理的CPU时间，不会发生饥饿现象。但公平是相对的。

(3) 平衡性：保持系统资源使用的平衡性。

(4) 策略强制执行：如安全策略



3.1.2 处理机调度算法的目标

2. 批处理系统的目标

(1) 平均周转时间短

在批处理系统中，**周转时间**就是作业从提交到完成所经历的时间。（公式中 T_{si} 为实际运行时间）

$$\text{平均周转 } T = \frac{1}{n} \sum_{i=1}^n T_i$$

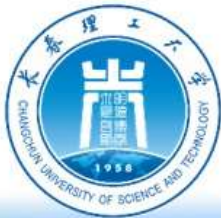
$$\text{平均带权周转 } W = \frac{1}{n} \sum_{i=1}^n \frac{T_i}{T_{si}}$$

(2) 系统吞吐量高

吞吐量：单位时间内所完成的作业数，跟作业本身特性和调度算法都有关系。

(3) 处理机利用率高

对于大、中型计算机尤为重要（CPU昂贵），故调度选择计算量大的作业可提高CPU的利用率。



3.1.2 处理机调度算法的目标

3. 分时系统的目标

(1) 响应时间快

响应时间是从用户从键盘提交一个请求开始，直到屏幕上显示出处理结果为止的一段时间间隔。

(2) 均衡性

均衡性是指系统响应时间的快慢应与用户所请求服务的复杂性相适应。
通常，用户对较复杂的任务的响应时间允许较长。



3.1.2 处理机调度算法的目标

4. 实时系统的目标

(1) 截止时间的保证

截止时间是指某任务必须开始执行的最迟时间，或必须完成的最迟时间。实时任务又分为硬实时任务与软实时任务之分。

(2) 可预测性

实时系统中可预测性非常重要，如视频播放应具有连续性，应具有提前下载的功能，提高并发性。



3.2 作业与作业调度

3.2.1 批处理系统中的作业

1、作业和作业步

1) 作业：用户在一次解题过程中或一个事务处理中要求计算机系统所作工作的总和，它是用户向计算机系统提交一项工作的基本单位。

-用户的观点：在一次业务处理过程中，从输入程序和数据到输出结果的全过程。

-系统的观点（针对作业进行资源分配）：作业由程序及数据（作业体）和作业说明书（作业控制语言）构成。

2) 作业步：是在一个作业的处理过程中，计算机所做的相对独立的工作。

3) 作业流：批量系统中需要将一批作业依次输入到辅助存储器中，形成作业流。



3.2.1 批处理系统中的作业

2、作业控制块(JCB, Job Control Block)

- **预输入程序**从输入设备上接收作业信息（即作业说明书、程序和数据），把作业组织成文件送到磁盘上的特定空间（称为输入井），同时，系统应把所录入的作业进行登记，登记在一张作业控制表中。每一个进入系统的作业均占据作业控制表中的一个表项，该表项称之为作业控制块。此时，作业建立完成。
- **JCB**包含了对作业进行管理控制所必需的信息。作业控制块是作业在系统中存在的标志。作业控制块的内容也是作业调度的依据。
- **从系统管理的角度看**，作业的建立过程就是申请一个空白的作业控制表项，填写作业控制块，作业控制表如下所示：

用户名	作业名	作业优先数	外存中地址	作业长度	资源需求	作业状态	作业进入系统时间	运行时间计数	要求运行时间



3.2.1 批处理系统中的作业

3、作业运行的三个阶段和三种状态

作业从进入系统到运行结束，通常要经历收容、运行、和完成三个阶段。

收容阶段：用户作业从提交到放入作业后备队列中，此时作业为后备状态；

运行阶段：作业被调度进内存，建立进程控制块，进入就绪队列直至运行结束，此期间作业均为运行状态；

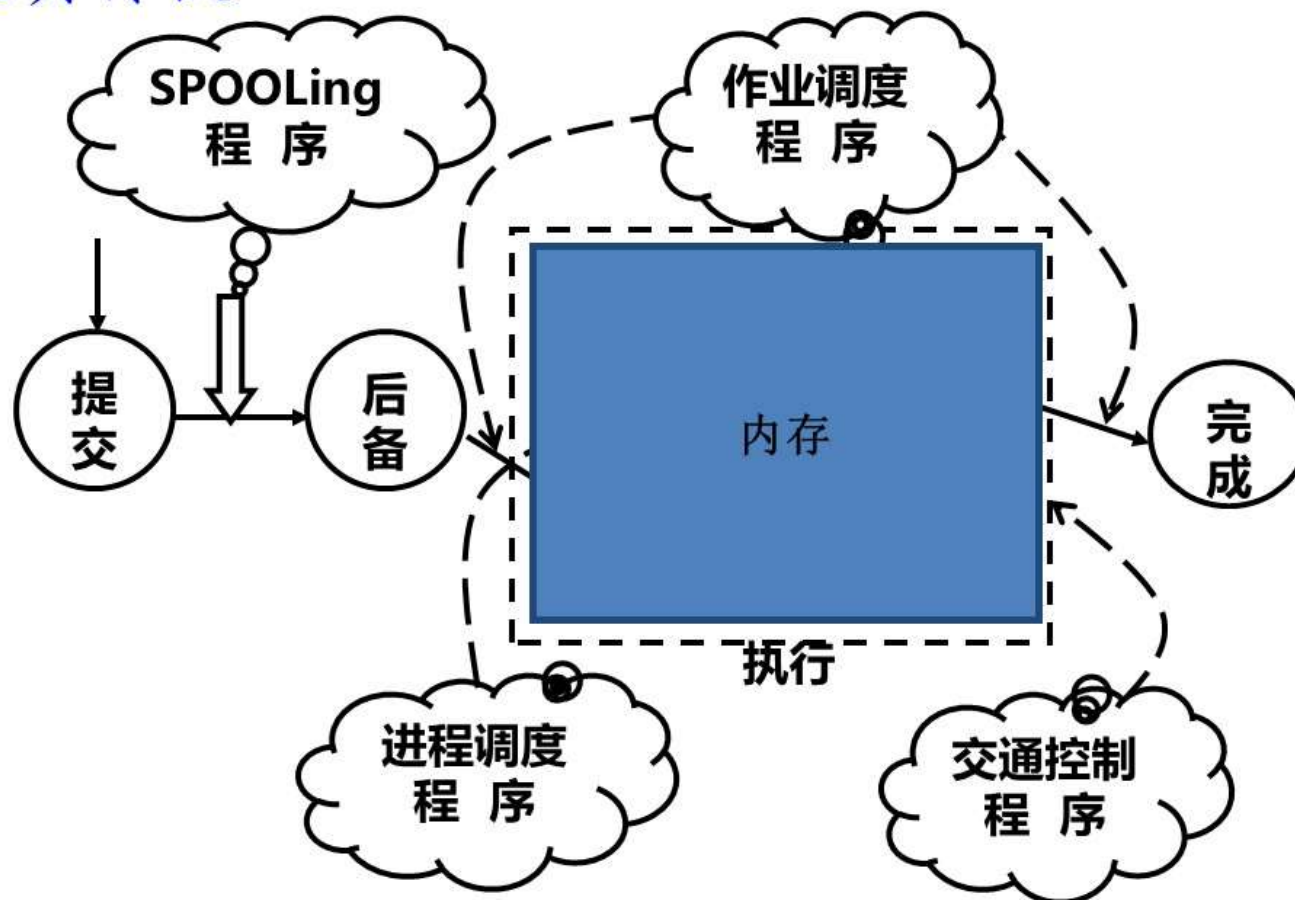
完成阶段：当作业完成或异常结束时，进入完成阶段，此时作业为完成状态，等待回收资源。



3.2.1 批处理系统中的作业

作业状态及其转换

四种状态



主观题 10分



说说慕课中让你困扰或者特别感觉感兴趣的内容有哪些？

11:40

17



3.2.2 作业调度的主要任务

作业调度的主要任务是根据JCB中的信息，检查资源需求，并按一定的调度算法，从外存的后备队列中选取某些作业到内存中来，为其创建进程。故作业调度也称为接纳调度。

作业调度的主要工作：

(1) 接纳多少个作业：从后备队列中选取多少作业进入内存，取决于多道程序度；

(2) 接纳那些作业：选择后备队列中的那些作业进入内存，取决于调度算法。

作业调度的目标是使作业运行最大限度地发挥各种资源的利用率，并保持系统内各种活动的充分并行。



3.2.3 先来先服务和短作业优先调度算法

1、先来先服务调度算法 (first-come first-served, FCFS)

进程名	到达时间	服务时间	开始执行时间	完成时间	周转时间	带权周转时间
A	0	1	0	1	1	1
B	1	100	1	101	100	1
C	2	1	101	102	100	100
D	3	100	102	202	199	1.99