MVS 系统下计算机资源的管理及性能调整

贺小鹏　（成都飞机工业公司计算中心）

摘要:本文介绍了在 *IBM* 大中型计算机 *MVS* 操作系统资源管理设施*(SRM)*的作用机制及其使用,以及如何利用系统活动检测工具*(MF/1)*对系统进行分析和调整,最后对笔者从事以上工作的经验进行了总结。

一、引言

贵刊 1993 年第 5 期登载了杨连波同志的文章<<关于 MVS/SP 系统下计算机资源管理的探讨>>。该文对 SRM 的概念和系统资源管理的概念进行了介绍,其中包括存储器管理、地址空间管理、CPU 负荷平衡管理和 I/O 通道平衡管理。重点介绍了对几种常见系统问题的分析和调整。但对 SRM 的作用机制、参数设置以及如何利用系统活动检测工具对系统进行分析调整没有论述。签于资源管理和性能分析调整是系统维护的一个重要课题,本文拟从上述几个方面对这一课题做进一步的探讨。

二、SRM 的作用机制

SRM 是 MVS 系统下的资源管理设施,用于监控系统资源的使用及对系统性能进行调整。但是 SRM 必须有一定的基础才能起作用。这就好比是调整一台复杂的机器要使这台机器正常运转,首先在安装机器时就要确定好基本框架,安装好主要部件,然后再进行局部调整。SRM 就相当于这样一种局部调整工具。一个系统的调整首先应从选择适当的系统生成参数及各子系统的参数开始,并以此建立一个性能基础,然后 SRM 在此基础上管理系统的资源,优化资源的分配与使用。下面将对 SRM 的主要方面进行介绍。

1.输入

SRM 通过一些参数来对系统资源进行分配和监控,其输入参数来自系统数据集 SYS1。PARMLIB 中,分为 ICS、IPS 和 OPT 三类,其成员名分别为 IEAICSXX,IEAIPSXX,IEAOPTXX。

(1)IPS 参数:用于指定性能参数。它根据各地址空间对系统资源使用方式的不同将其划分为不同的 DOMAIN(域),并为每个域指定一限制:CNSTR=(MINMPL, MAXMPL,DOMAIN WEIGHT),该参数同时也表明了这个域与其它域相比较在 SRM 进行调度时的重要程度。每个域都属于某一 PGN(性能组),IPS 通过性能组为每个域指定其性能目标,SRM 就根据这些目标对地址空间进行资源的分配和管理。

(2)OPT 参数:用于控制系统吞吐量与响应时间之间的关系,包括系统负荷平衡参数和资源相关系数(RFC:RESOURCE FACTOR COEFFICIENT),SRM 根据这些参数确定系统是否处于不平衡状态并对资源的使用进行相应的调整。这些调整包括多道程序级调整、CPU 负荷调整和 I/O 负荷调整。

(3)ICS 参数:用于告诉 SRM 各子系统处于哪些 PGN 中,这些 PGN 叫控制性能组,在 IEAIPSXX 中定义。此外,IEAICSXX 中还定义报告性能组,它们只用于报告资源使用情况,不进行控制,因而不是 IPS 参数。

2.目标

SRM 只与系统可用资源打交道,其目标有两个:

(1)根据响应时间、作业轮转和优先级的要求将系统资源分配到各地址空间;

(2)从系统吞吐量方面考虑获得资源的最优使用。

总的来说,SRM 的作用在于监视系统的运行并平衡资源的使用。

3.控制

SRM 实施控制的种类有四种:

(1)DOMAIN(域)控制,这是控制的基本单位;

(2)WORKLOAD(工作负荷)控制,在同一个域中控

制地址空间对资源的竞争;

(3)DISPATCHING PRIORITY (调度优先级)控制,在整个系统范围内调度各地址空间;

(4)THROUGHPUT(吞吐)控制,对资源瓶颈问题进行处理。

这四种控制同时起作用,但何种控制起主要作用要视系统活动情况而定。

4.方法

SRM 实施控制的主要手段是对地址空间在实存与辅存之间进行 SWAP(交换)。当系统处于超负荷运行时,SRM 将选择一个地址空间从内存中换出;当系统处于欠负荷时,选择一个地址空间从辅存中换入。系统的负荷包括工作负荷、CPU 负荷、I/O 负荷三种,SRM 依据公式:

SWAP recommendation value= workload level recommendation+ CPU * CPUload ajusting recommendation+ IOC * I/Oload ajusting recommendation

计算各地址空间的 SWAP 推荐值,确定系统不平衡时地址空间换入换出的次序。该值大者将被换入,小者被换出。

5.事务

SRM 通过所谓事务(TRANSACTION)来量度一个地址空间所消耗的系统服务。在 MVS 下,对于批作业,一个事务相当于一个作业步或一个作业(只有一个作业步);对于交互方式,通常一个事务相当于一条 TSO 命令。一个地址空间在某一时刻只能有一个事务活动。地址空间所消耗的服务用如下公式计算:

SERVICE=CPU * CPU SERVICE UNITS+IOC * I/O SERVICE UNITS+ MSO+ STORAGE SERVICE UNITS

其中 CPU、IOC、MSO 叫服务定义系数,它们之间的比例关系反映了 SRM 对 CPU、I/O 及内存三项活动进行统计的重点是什么。如果要使这三项活动在统计中所占比重平均,可根据公式:CPU * CPU SERVICE UNITS=IOC * I/O SERVICE UNITS= MSO * STORAGE SERVICE UNITS 确定 CPU、IOC 和 MSO 的值。CPU/I/O 及 STORAGE SERVICE UNITS 的值随机型的不同而不同,它们都是常数。

6.参数指定

以上介绍了 SRM 的一些主要方面,下面指出在设置 SRM 参数时应注意的一些问题:

(1)DOMAIN 的划分原则:

· 不可 SWAP 的地址空间,比如各子系统牵力(JES2、TSO、STC 等),应单独分配一个 DOMAIN。

· 短 TSO 命令(要求快速响应)应有自己的 DOMAIN,并令其 MINMPL=最大就绪用户数。

· 大量占用系统资源的工作应放在一个 DOMAIN 中,并令 MAXMPL=1,如为批作业,可令 MAXMPL=2

· DOMAIN WEIGHT 之间的差别应足够大,以反映各个 DOMAIN 的重要程度。域的重要性用冲突索引表示:CONTENTION INDEX=就绪用户数 * WEIGHT / TARGET MPL,大者重要。可首先确定批作业的 WEIGHT,再决定其它的域。对于批作业,如果作业队列足够多,则其 TARGET MPL=1。

(2)IPS 参数中的 ISV(分段服务值)表示一个 SWAPIN 的地址空间最少应得到的服务,它用于控制 SWAP 的次数,因而不能太小。在服务定义数 CPU=10.0,IOC=5.0,MSO=3.0 的情况下,对于短作业,其 ISV≈600,中 TSO 命令,600<ISV<2K。

(3)IEAIPSXX 中指定的 CPU、IOC 与 IEAOPTXX 中指定的 CPU、IOC 含义是不同的,前者叫服务定义系数,后者叫资源相关系数。资源相关系数 CPU 与 IOC 的值要小心选择,否则因为它们对 SWAP 推荐值的影响很大,将导致 SWAP 很频繁,从而抵消 ISV 的作用。一般该值都较小。

(4)PERFORMANCE OBJECT 应少一些好,以便于管理。一般对于批作业,只需定义两个 OBJECT,对于 TSO,一个就够了。

(5)RTB(响应一吞吐偏置)=0,表示 CPU 负荷和 I/O 负荷调整不起作用;RTB=1 则表示要起作用,此时 SRM 考虑的是系统吞吐量而不是响应时间。通常情况下不要令 RTB=1,因为这将增加系统负担而效果并不是很好。

(6)此外,PGN 中的 PERIOD 可以适当多设一两个,这样可以减少 SWAP 次数,并能有效地降低 ISV、DUR

和 OBJ 参数指定不恰当时产生的影响。

三、系统活动检测工具 MF／1

FM／1 是对系统活动及资源使用进行记录的一个软件工具，用于对系统性能进行观察和分析。利用 MF／1 所产生的报告，我们可以了解整个系统的活动情况，系统活动包括 CPU 活动、通道活动及其与 CPU 的重叠、I／O 设备活动、页活动以及系统的工作负荷。我们也可以利用 MF／1 报告在对系统进行调整后观察系统的反应，比较调整前后的差别。因此 MF／1 报告为我们分析和调整系统提供了一个定量和科学的依据。但是 MF／1 报告可记录的东西很多，究竟哪些数据对我们是有用的，应该如何利用它们对系统活动进行分析和调整？下面就这些问题进行探讨。

1.MF／1 参数的选择

MF／1 的参数在 SYS1.PARMLIB 的成员 ERBMF1XX 中指定，包括 CPU、通道、I／O 设备（又包括字符输入设备、通讯设备、图形设备、磁盘、磁带和打印机）、页活动、工作负荷等等，MF／1 对这些参数指定的活动进行记录并产生报告。我们也可以根据我需要指定对系统某一部分的活动进行记录。此外还有 CYCLE 参数。指定 MF／1 进行数据收集的时间区间；INTERVAL 参数，指定产生报告的时间间隔；STOP 参数，指定 MF／1 的活动时间。这里要注意的是，CYCLE 参数不能太大（＞1 秒）也不能太小（＜50 微秒），否则结果都不很准确，一般取值为 250 微秒；对于 INTERVAL 参数，则不妨取大一些，例如 20 分钟就比较合适；对于 WKLD 参数，最好指定为 WKLD(PERIOD)，以便于观察 SWAP 的情况。

2.MF／1 报告的分析及对系统的调整

对于 MF／1 报告，具体的分析和调整步骤如下：

(1)首先检查 CPU ACTIVITY 报告，若发现"WAIT TIME PERCENTAGE"下显示的数字大于 20，则说明 CPU 没有得到充分的使用，这可能是由于系统任务不饱满，用户不多；也可能是由于性能参数的设置有问题，需要进行下一步的检查。如果该值小于 15，则说明 CPU 过载，转入第四步的检查。

(2)检查 PAGING ACTIVITY 报告，若页活动的时间超过了 CPU 时间的 5%（对于 IBM3031 机器而言，相当于换页率大于 15 页／秒。换页率为"PAGE IN"和"PAGEOUT"栏中"NONSWAP"栏下数值的总和），则说明存在过度的页活动，因此应该增大 ISV、ERV 以及 RFC 的 CPU 值，减少 MINMPL。如果不存在过度页活动，则检查"SWAP OUT COUNT"一栏，该栏包括如下一些引起 SWAP 的事件类型：

• INPUT TERMINAL WAIT：该值一般都较大，在 TSO 环境下大部分 SWAP 属这种类型，这是正常情况。

• OUTPUT TERMINAL WAIT：若该值大于零，则是由于 TSO 命令或命令过程产生输出太快，可以动态分配给输出的虚存不够。应增大 TSO KEY 中的 HIBFEXT 值。

• LONG WAIT：若该值较大，一般是由于用户程序的问题；也可能是由于所需的资源被一个 SWAP OUT 的地址空间占用了，可减小 IEAOPTXX 中 ERV 的值。

• DETECTED WAIT：若该值很大，则可能有两个原因，一是启动的作业队列太多；二是系统操作员对作业请求响应得太慢。

• UNILATERAL：在批方式下，该值应等于结束的作业数，若该值超过了结束的作业数，则是由于 MINMPL 太小，使有些地址空间在工作还没完成时就被强迫换出了。检查 WORKLOAD ACTIVITY 报告，找出最后一个 PERIOD 中"NUMBER OF SWAPS"值较大的 PGN 及相应的 DOMAIN，将其 MINMPL 增大。

• EXCHANGE ON RECOMMENDATION：若该值较大，则是由于 ISV 值太小或 OBT 指定不当。再次检查 WORKLOAD ACTIVITY 报告，可以找出这个 DOMAIN，将其 ISV 增大，或者将其 OBJ 指定为与上一个 PERIOD 的 OBJ 一样。

• ENQ EXCHANGE：该值一般应为零，否则说明对需进行排队访问的资源（串行资源）竞争很激烈，如果同时发现"UNILATERAL"值也大大超过了结束的作业数，则说明存在资源瓶颈。解决办法是减小 ERV 值，并找出引起严重竞争的资源，增加该资源的 RFC 值。此时可能还应该修改 IEAIPSXX 中的 ISV、DUR 和 OBJ 值，并令 RTB＝1。

(3)如果"SWAP OUT COUNT"中没发现什么异常情况，则检查 CHANNEL ACTIVITY 报告，该报告显示

通道的活动情况。如果发现通道也不忙,则说明系统中活动的地址空间很少,任务不饱满,应增大 MPL 的值,并多启动一个作业队列;也有可能是上机的用户本来就不多,系统没有处于高峰期。反之,如果通道很忙,则说明系统过份注重 I/O 处理,应在 IEAOPTXX 中减小 IOC、增大 CPU 的值。

(4)检查 WORKLOAD ACTIVITY 报告,观察"AVERAGE TRANS SERV RATE"一栏,如发现服务率很低,则表明系统过份注重 CPU 的使用率,从而降低了系统的吞吐量。应减小 RFC 的 CPU 值。

(5)检查 CHANNEL ACTIVITY 报告,如发现"PERCENT CHANNEL BUSY"一栏中某一通道比其它通道大得多,则显示系统中存在 I/O 不平衡的情况,应增大 RFC 的 IOC 值。如果修改后情况并没有改善,说明 SWAP 策略无效,应进行下一步的检查。此外,该报告中显示的"PERCENT CHAN BUSY& CPU WAIT"值表示 CPU 活动与 I/O 活动的重叠情况,该值越小说明重叠程度越大,系统的吞量就越大。若该值较大,则应检查 IEAIPSXX 中调度优先级的设置,将可有批作的 DP 参数都设置到 MEAN TIME TO WAIT 组。

(6)检查 DIRECT ACCESS DEVICE ACTIVITY 报告,如果"%DEV BUSY"栏和"AVG Q LENTH"栏中的数字显示某一盘特别大,则说明其上有一些频繁使用的数据集,应移动一些到其它盘上。如果不是这种情况,则说明数据集的分布是较合理的,那么就应该重新配置系统,将一些盘挂到另外一个通道上,以改善通道的不平衡状况。

除了以上所介绍的步骤外,还可以利用 PAGING 报告观察调整 SRM 参数对 SWAP 的影响,辅存中可用页槽的变化,也可以观察换页率与工作负荷之间的关系。利用 WORKLOAD 报告可观察事务和服务在各个域中的分布以及 WORKLOAD 与"SERVICE IN INTERVAL"之间的关系。

总之,MF/1 是一个非常有用的工具,对于它所产生的报告,我们应该加以仔细的分析,并根据其记录的系统资源的不正常使用情况,对系统进行相应的调整。这里要注意的是,每做一次修改,就应产生一个报告,以便于进行比较,验证调整结果。

四、其它工具

在对系统活动及资源使用进行记录方面,MVS 还提供了两个工具,也有助于我们了解系统的运行状况。这两具工具是:

1.RMFMON

这是一条 TSO 命令,实际上是一个动态监测系统活动的实用程序。利用该程序,我们可以观察整个系统的动态变化,资源的分布和使用以及各地址空间的活动情况;还能对单个的作业进行观察。

RMFMON 具有如下一些窗并在其中动态显示有关数据:

- ARD:显示地址空间资源数据
- ASD:显示地址空间状态数据
- ASRM:显示地址空间 SRM 数据
- DDMN:显示系统中各 DOMAIN 的情况
- SPAG:显示页活动
- SRCS:显示实存、CPU 及 SRM 活动

对于前三个功能,还有相应的 ARDJ、ASDJ、ASRMF 窗口用于对单个作业进行动态显示。

2.SMF(SYSTEM MANAGEMENT FACILITY)

这是一个系统管理软件,是 MVS 系统的一个标准设施,用于收集和记录面向系统和面向作业的各种信息。面向系统的信息包括系统配置、页活动和工作负荷情况;面向作业的信息包括 CPU 时间、SYSOUT 活动以及对数据集的存取。我们可利用 SMF 来生成用户账单,分析系统的负荷或资源的使用情况,所以,SMF 对于分析系统也是一个很有用的工具。SMF 共有约 80 种记录类型,用 SMFN 表示,例如类型 71 就表示为 SMF71,它记录的是页活动的情况。

相比较而言,MF/1 可用于显示某一时间段内某些系统资源的不正常使用情况,而 SMF 用于显示系统的负荷情况。对照两报告也许可发现在资源的不正常使用与某个问题程序(批作业或交互会话)间存在一定的关系。

SMF 与 MF/1 的联合使用至少可用于以下目的:

(1)比较用户程序和系统任务的换页率(SMF3.34 和 SMF70,71),从中也许可以发现在问题程序中存在不正常的页活动。

(2)比较用户程序和系统的 I/O 活动,可以找出

SMF 记录的 EXCP 次数与 MF／1 记录的 I／O 次数的对应关系。

(3)比较用户程序所消耗的服务与整个系统所消耗的服务(SMF5.35 与 SMF72),可以确定服务在各地址空间的分配是否合理。

(4)决定是否对系统配置进行修改(SMF0、8、9、10、11、22、70、73、74),这些 SMF 信息可以解释 MF／1 报告中的重要变化,因而对于分析 MF／1 数据是很有用的。

五、一些经验

虽然笔者维护的是 IBM 3031 大型机,但这些经验对于运行 MVS 操作系统的其它型号的 IBM 机器也是适用的。现在把它们介绍给大家,希望能对有关的系统人员有所帮助。

1.在对系统进行调整前首先要对系统性能有总体的考虑,其次要有明确的目标,并一定要预计调整后的结果。

2.要选准参数,只有系统比较敏感的参数对系统性能才有真正的影响。不要对每个参数都试图进行修改。

3.对参数的修改要有准确的记录,建立完善的档案。因而 SRM 参数很多,很容易出现混乱。

4.对于选好的参数,尤其是 OPT 参数,在进行调整时,不要同时修改许多个,否则将判断不清究竟是哪一个参数在起作用,而应该一次只选择一个进行修改,同时要使其修改前后的差别足够大,这样才便于观察系统的反应,并进行有效的比较。

5.参数调整后,观察系统应尽量在系统负荷大致相同的情况下进行,比如在同一时间段内,或用户数相同的时间段内,以免得出不正确的结论。

6.可设置几组 OPT 参数和 IPS 参数,并根据系统资源的使用情况进行合理的组合,以达到某些特殊的目的,增加灵活性。

7.再次指出 SRM 参数的调整只是辅助性的,诸如系统生成参数,系统数据集、页空间等在盘上的分配情况等对系统的性能都有很大的影响。

六、结束语

系统性能的调整涉及的因素很多,并且互相牵连、关系复杂。要调整一个系统,首先要善于利用系统提供的各种工具,分析查找出系统的问题所在,才不致于盲目行动;其次要认识到系统的调整是个反复不断的过程,不能指望一次就将系统调整得很好,而应该进行耐心和细致的观察比较,并不断进行修正。只有这样,才能使系统处于一种良好的工作状态。

参考资料:

[1]OS／VS2 MVS Initialization and Tuning Guide,GC28-0681-4,IBM

[2]OS／MV2 MVS System Management Facility,GC28-1