

## IBM zSeries WorkLoad Manager を活用した CPU 資源抑止手法の検討と評価

加藤 礼基 †

古くからメインフレームの世界では「保有する資源を使い切る」事が最良のチューニングであった。資源を効率良く使う為の技術は非常に重要である。zLinux 等を活用した zServer へのサーバー統合は「資源を使い切る技術」の上に存在している。ただ、処理装置の能力が非常に高くなった今日では「供給する資源を適切な量で抑止する」というアプローチが必要になるケースもある。資源を使い切るのか、或いは抑止して適量のみ使うのかをユーザーが選択出来る事が望ましい。当報告書では資源抑止の手法に焦点をあて、WorkLoad Manager を活用して資源抑止を実現した事例を紹介する。

### Evaluation and assessment of CPU resources optimization using IBM zSeries WorkLoad Manager

Reiki Katoh †

“Resources have been exhausted for as efficiently as possible” was considered to be the best tuning technique in the mainframe-world for a long time. Also in today the approach for “using up” is important. The server consolidation to zServer ( zLinux etc. ) exists on “the technology of exhausting resources.” However, today “The method of giving only a proper quantity, having the technique which exhausts resources” may be required. End user wants to choose “using up” or “under control”. In this report, I introduces about the method of control CPU resources using WorkLoad Manager.

#### 1. はじめに

処理装置の性能向上は驚くべきスピードで進んでいる。1990 年代前半にメインフレームにおける処理装置の主役が Bipolar から CMOS へと推移したが当時の CPU と比較すると現在の CPU 処理能力は 100 倍近くにまで向上している。単体 CPU の能力が飛躍的に向上した今日では CPU 資源の供給を適切な量で制御する事が重要な課題である。当報告書では最初に資源抑止の必要性に述べ、その後 WorkLoad Manager の機能を活用した実際の資源抑止事例を紹介する。

#### 2. 資源抑止の必要性について

本章では処理装置の能力向上による影響を整理する事で資源抑止の必要性を示す。

##### 2-1. 処理能力向上による応答時間の劣化

一般的に処理装置の能力が向上すればオンライン業務の応答時間は短くなる。ただし現実にはシステム構成によって応答時間が長くなる可能性もある。ここでは具体的にどのようなケースにおいて応答時間の劣化が起こり得るか紹介する。

##### 2-1-1. 磁気ディスク制御装置の進化

処理装置同様に磁気ディスク制御装置も進化を続けている。ただし磁気ディスク制御装置の進化は処理装置のそれとは全く種類の異なるものである。磁気ディスク制御装置においては I/O 処理能力の向

上と筐体収容可能容量の拡張が並行して進められている。処理装置が処理速度向上を追求しているのに対して、磁気ディスク制御装置は複数の観点での進化を余儀なくされている。筐体当たりの I/O 処理能力向上と筐体当たりの収容容量拡張が並行して行われるということは単位容量当たりの I/O 処理能力が必ずしも向上しない事を意味する。

一方で機能拡張も進められているが、それによる性能劣化の可能性もある。特にコピー関連の機能は RAID RANK 当たりの大容量化も手伝って RANK 内高負荷状態を引き起こす可能性を持っている。メインフレームに接続する磁気ディスク制御装置にはスナップショット的なコピー機能や遠隔転送の為のコピー機能が標準的に搭載されている。それぞれ単独で利用する場合には性能への影響が少なくとも組み合わせる場合には無視できない影響が出るケースもある。以上を考慮すると磁気ディスク制御装置の進化は性能向上への貢献という観点では処理装置のそれとは明らかに異なっている。

##### 2-1-2. オンライン業務応答時間への影響

処理装置の性能向上がもたらす結果の一つとしてバッチ業務の ELAPS 時間短縮がある。バッチ業務の ELAPS 時間短縮がシステムに与える影響は単位時間あたりの I/O 要求数の増加である。非常に多くの I/O を要求する業務において単位時間あたりの I/O 要求数が 2 倍になる事による磁気ディスク制御装置への影響は非常に大きい。加えて書き込み対象の業務ボリュームが前述のコピーサービス(スナ

† 日本アイ・ピー・エム株式会社  
IBM Japan, Ltd.

ップショット、遠隔コピー等)の対象ボリュームであれば磁気ディスク制御装置の負荷及び該当 RANK 内の負荷を増大させる原因になる。このような高負荷状態においてはオンライン・トランザクションからの I/O 要求に対する応答も当然遅くなる。I/O 要求の優先度を加味するデザインであったとしても、現実には RANK 高負荷状態の影響は避けられない。つまりオンライン・トランザクションからの I/O 要求に対して通常よりも応答時間が長くなってしまふ危険がある。オンライン業務の応答時間に占める I/O 応答時間の割合は CPU 時間に比べて遙かに大きい為、このようなケースでは結果として応答時間が悪化する結果になる。

## 2-2. ソフトウェア費用への影響

メインフレームの世界ではオン・デマンドの考え方が徐々に浸透してきている。ソフトウェア料金制度においては使用量に応じて課金される「ワークロード使用料金方式」を採用するユーザーが増えている。IBM の WLC(Workload Licence Charge)では 4 時間平均の使用量に基づいてソフトウェア料金が請求される。従量課金制度に移行したユーザーにとっては前述のバッチ業務の Elaps 時間短縮はソフトウェア費用を高騰させる危険を含んでいる。例えばオンライン業務の少ない休日等に 12 時間かけて実施している週次処理が処理装置の能力向上により 8 時間に短縮されるケースを考えてみる。単純に計算すると CPU 資源消費量の単位時間(4 時間)あたりの値が 1.5 倍になる。その結果処理装置の能力向上がソフトウェア料金の増額を引き起こすことになる。

## 2-3. Disaster Recovery 環境への影響

処理装置の能力が増強された場合に磁気ディスク制御装置に与える影響に関しては前述したが、遠隔コピー等の災害対策を実現している環境においては単位時間あたりの I/O 要求数が増える事は遠隔地への転送にも影響が及ぶ。殆どの遠隔コピー手法で回線帯域を見積る根拠は単位時間あたりの Write 要求数である。単純に考えると処理装置の能力向上により単位時間あたりの Write 要求が倍になるのであれば回線の帯域が不足したり、リモート側の機器の能力が不足したりする可能性がある。IBM の遠隔コピーソリューションを例にすると例えば XRC(eXtended Remote Copy; zSeries Global Mirror)の構成においてはリモートサイトからの読み込み処理即ち SDM(System Data Mover)からの RRS(Read Record Set)による変更トラックの READ 処理が間に合わず Volume Pair が Suspend になる可能性がある。MetroMirror のような同期転送方式であればそ

のまま業務 I/O の遅延を引き起こす可能性もある。処理能力向上が磁気ディスク制御装置への I/O 要求数増加につながるケースにおいては災害対策システムへの影響も考慮する必要がある。

## 3. WLM(WorkLoad Manager)による資源抑止

WLM(WorkLoad Manager)は一連のジョブ群を資源グループとして定義し、該当グループに対する供給資源量を制御する仕組みを持っている。特定の資源グループに対して「最小資源量」や「最大資源量」を定義する事が出来て、「最大資源量」を定義する事により特定のジョブグループに対する資源供給量を抑止する事が可能である。処理装置の能力向上によりバッチジョブ群に対して CPU 資源を与えすぎる事が問題につながるのであれば、供給量を抑止するのが最も適切な対応策と言える。この章では WLM の機能を簡単に紹介し、次章以降ではどのような方法で設定値を決めればよいかという具体的な方法について論じる。

### 3-1. 稼働ジョブと資源グループの関連

Figure3-1 は WLM Goal Mode の全体像を簡単に示したものである。図の通りそれぞれのバッチジョブはクラシフィケーション・ルールによって関連するサービスクラスが決定される。一つ以上のサービスクラスが特定のリソースグループ(資源グループ)と関連付けられており資源グループに対する CPU 資源の供給量を制御する事でバッチジョブ群に対する抑止が可能となる。

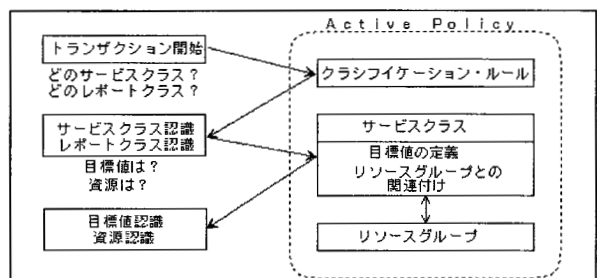


Figure3-1. WLM 環境における資源グループの決定

### 3-2. 資源グループに対する設定値

資源グループには 3 つのタイプがありどのタイプを使用するかはユーザーが選択出来る。使用するタイプにより設定する値の持つ意味も異なる。

① 資源グループ・タイプ 1: 値として Sysplex 全体における 1 秒当たりの CPU Service UNIT 数を指定して制御する。WLM Goal Mode 登場時から提供されている資源グループのタイプである。後で紹介する事例はこのタイプの資源グループを使用している。

- ② 資源グループ・タイプ 2: 値としてそれぞれの区画に配分された資源の割合を指定する。百分率による指定になるので 0 から 99 までの数字を指定する。
- ③ 資源グループ・タイプ 3: 標準 CPU の数を指定する。100 という指定値が 1CPU の能力を意味する。例えば MAX150 という指定は標準 CPU1.5 個の能力を上限とする事を意味する。

### 3-3. 資源グループに対する設定値の変更方法

資源グループに対する設定値の変更は複数の WLM Policy を準備しておいてそれを切り換える事で実現する。Policy の変更はシステム・コンソールからのコマンド投入で可能である。またユーザープログラムからも制御が可能である。「IWMPACT」という MACRO が提供されており、プログラムから適切な Policy を活性化できる。また「IWMPQRY」といった照会用の MACRO も準備されており、現在活動化されている Policy に関する情報を入手する事が出来る。

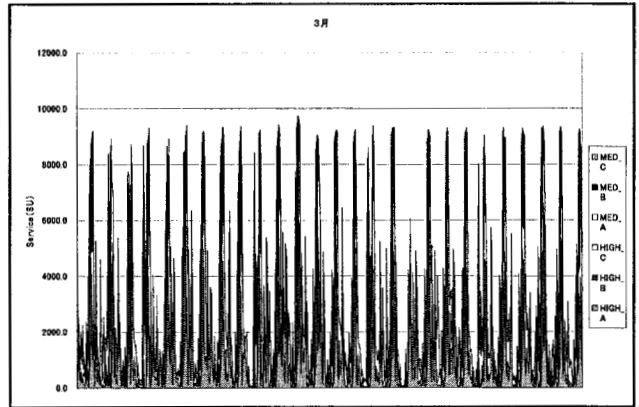
### 4. WLM によるバッチジョブ抑止手法の検討

WLM の機能を活用してバッチジョブ群への資源供給を抑止出来る事は前述の通りである。この機能を使用する為の次の課題はジョブ群のグルーピングと「最大資源量」の決定である。既に「最大資源量」を設定している環境に置いては、基本的にその値をそのまま利用すれば良い場合もある。現在設定していない環境において新規に値を設定する場合は現行の資源使用量を参考に値を決定する。この章では現行の使用状況の調査手法及び設定値の評価方法に関して論じる。

#### 4-1. 現行の使用量調査

使用量の調査は RMF のレポート(SMF Report Type72)を利用して実施する。調査対象期間を長くする事により設定値の適性は高まる。後で紹介する事例では年間で資源消費量の偏りがある可能性を考慮し 1 年間という期間を調査対象とした。Table4-1 は分析結果の一部であるが該当サービスクラスが消費した資源推移を調査し値を決定する事になる。Graph4-1 はある月の消費資源をグラフにし

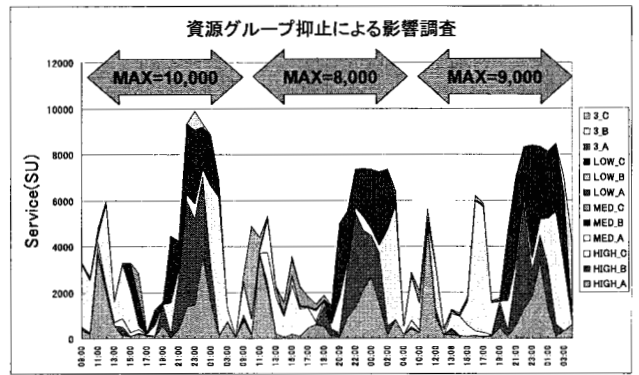
たものである。1年を通してグラフを作成すると該当グループに対してどの程度の値を設定すればよいか把握出来る。



Graph4-1. Service UNIT 消費状況

#### 4-2. 設定値変更の影響確認について

供給資源を制御する事で業務終了時間が予想以上に遅延し、後続業務に影響を与えるような設定をしてしまう可能性も否定出来ない。影響を少なくする為には大きな値を設定して監視しながら値を変更する方法を取るべきである。Graph4-2 は 11 月 15 日に 10,000 と大きめに設定した値を 8,000、9,000 と変化させて影響を確認したものである。このように監視しながら値を調整する事でより適切な値を見つける事が出来る。



Graph4-2. 資源グループ抑止の影響確認

Table4-1. 現行使用量の分析例

DATE	TIME	SID	TRANSACTIONS			SERVICE				
			AVG	MPL	%CPU	IOC	CPU	MSO	SRB	SU/SEC
2006/11/15	00:00	SYSB	3.36	3.36	24.8	14123K	117M	0	2453K	3323
2006/11/15	01:00	SYSB	1.40	1.40	8.1	8908K	37801K	0	1243K	1084
2006/11/15	02:00	SYSB	0.00	0.00	0.0	17510	103K	0	4580	3
2006/11/15	03:00	SYSB	0.00	0.00	0.0	17485	96817	0	3827	2

## 5. WLMによる資源抑止の適用事例

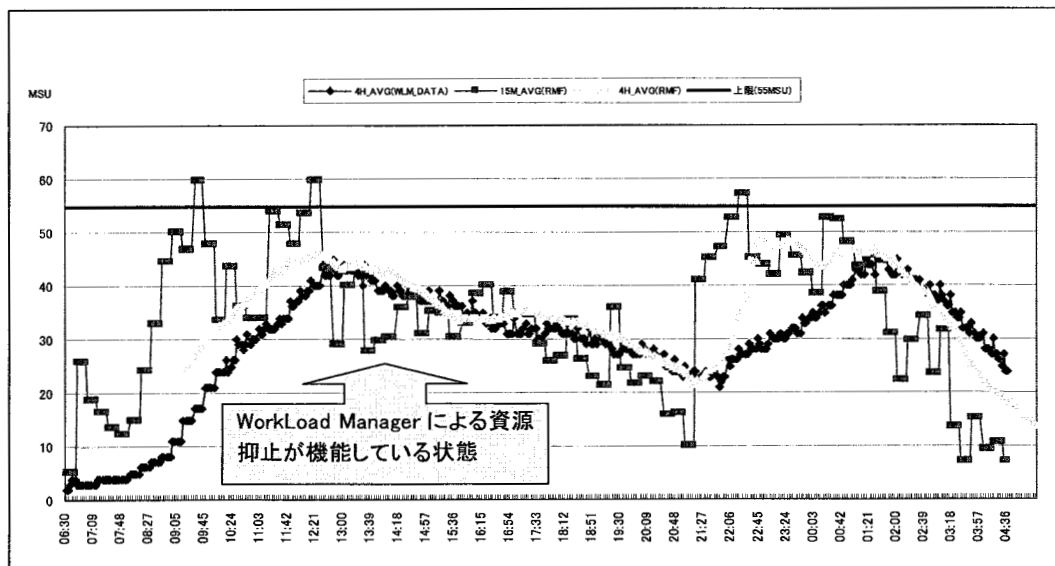
最後に実際にWLM(WorkLoad Manager)を活用して資源抑止を実現している例を紹介する。Graph5は該当システムの1日のCPU資源状況を示したものである。このシステムではソフトウェア料金の極端な高騰を防ぐためにWLMの機能を利用してCPU資源を制御している。具体的には複数システムの資源使用状況をモニターし、一定の閾値を超えたタイミングで資源グループに対する抑止を活動化させている。この例では13:00~14:00の時間帯において資源グループへの制御が機能して全体の使用量を一定量で抑えている状況が確認出来る。

## 6. おわりに

処理装置の能力向上による影響が論じられる事は少ないが、サーバー統合やストレージ統合が進み複数の業務がシステム資源を共有するようなシステム構成においては非常に重要な検討項目である。当報告書では処理装置の性能向上による影響に対してz/OSのWorkLoad Managerを活用する事で適切な資源制御を実現している例を紹介した。WorkLoad Manager自身もz/OS v1.8において機能拡張され複数システムグループの使用量制限等が可能になっている。従来のLegacy Systemと呼ばれているAreaにおいては現在のCPU能力は大きすぎる場合がある。今後CPU処理能力が大きくなるに従って資源抑止へのアプローチもより重要視されていくと思われる。

## 参考文献

- [1] MVS Planning: Workload Management SA22-7602-12
- [2] MVS Programming Workload Management Services SA22-7619-05
- [3] MVS System Management Facilities SA22-7630-13



Graph5. WLMを活用した資源抑止の例