

仮想計算機システムの導入効果に関する一考察

—大規模計算センタの場合—

末永正, 景川耕宇

(九州大学大型計算機センター)

VM/370方式の仮想計算機システムは、一台の実計算機のもとで複数の仮想計算機を創り出すことができ、大規模計算センタのように長時間のサービス時間帯の確保と頻繁なシステムレベルアップを両立させなければならない施設では、システムの二重化よりもコストパフォーマンスの点で非常に魅力的なシステムである。反面、その制御方式によってもたらされるシステム効率の低下が懸念される。特に、オープンショップ方式の運用形態をしている計算センタでは、効率低下をエンドユーザが直接感ずるため、その程度を把握しておく必要がある。そこで、九州大学大型計算機センター（FACOM M-200）に仮想計算機システムを導入し、そのもとでの運用サービスによって適用の問題点および得失を考察する。本例では、システム効率を低下させる原因が主としてCPUオーバヘッド増にあるが、システムに対する負荷状況によってはVM導入による計算機資源の分割が効率低下の主要因となることも明らかにしている。

1. はじめに

大規模計算センタの業務内容には、末端の利用者（エンドユーザ）に対する計算サービスはもちろんのこと、そのサービス業務を円滑に遂行するためのシステムメンテナンス、より高度な機能をもつソフトウェアへのバージョンアップ、および、それらに付随するセンタ独自のシステムプログラム開発などが存在する。このようなシステムメンテナンスやシステムプログラム開発などは、同一オペレーティングシステム（Operating System : OS）の下でエンドユーザのジョブと並行処理できない場合が多く、一般に別の時間帯に実施するか、又は別システムで実施せざるを得ない。しかしながら、このためにシステムを二重化することは、平均故障間時間（Mean Time Between Failures : MTBF）が重要であるオンライン系の業務ならばと

もかく、コストパフォーマンスの点からあまり得策ではない。又、別の時間帯を設けて処理することは長時間の計算サービス（たとえば、24時間オーバナイトサービス）に対する妨げになる。そこで、これらの問題の解決策として仮想計算機システムの適用が考えられる。

仮想計算機システムでは、上記の目的にそう実用的なシステムとして、VM/370¹⁾が1972年にIBM社によって世に供されている。近年になって、ようやく、日立製作所、日本電気、富士通といった国内計算機メーカーも次々と商用化を行っている。^{2)~4)}日立製作所と富士通の仮想計算機システムの制御方式はVM/370とほぼ同じであり、日本電気のそれは特別な制御プログラム（Virtual Machine Monitor : VMモニタ）を必要としない形式になっている。VM/370による制御方式で

は、VMモニタによって1つの実計算機システム中に複数個の仮想計算機 (Virtual Machine : VM) が生成され、各VMで独自のOSの動作が可能になっている。

筆者の所属する計算センタ*では、富士通のM-200システムによって計算サービスを実施しているが、利用時間の延長化の前提として、1981年8月17日より仮想計算機システム**での試験的な運用を開始した。本稿では、約1年間の運用経験をもとに、仮想計算機システム導入の得失について報告する。

2. VMモニタの仕様

富士通のVMモニタ (AVMバージョン2) には、以下に示す機能がある。

(1) 動的切り替え機能

VMモニタが動作すれば必然的にCPUオーバヘッドが増加する。そこで、VMモニタの動作の必要性に応じて、システム動作を仮想計算機モード又は、実計算機モードに切り替える。

(2) 自動運転機能

操作性および省力化への配慮から、
i. 自動IPL機能、ii. 自動電源切断機能、iii. 警報機能、iv. 自動ダンプと再IPL機能の4つの自動化を可能にする。

* 大学間全国共同利用の大型計算機センタで、利用登録課題数1,800以上、年間28万件のバッチジョブと約30万件のTSSセッションを処理している。

** 富士通では自社の仮想計算機システムのことをAVM (Advanced Virtual Machine) と呼んでいる。

(3) リソース専用化機能

効率を重視するVM (V=RのVMと呼ぶ) に対して、各種リソース (主記憶装置およびCPU) を専用的に割り当てる。つまり、実記憶装置の1部をV=RのVMに専用化しVMモニタによるオーバヘッドを軽減する。又、MP (Multi processing) の場合にCPUの1台を各VM間で共有させ、他をV=RのVMに専用化させている (専用プロセッサと呼ぶ)。M-200システムでは、この専用プロセッサに対するVMモニタの介入は一切なく、オーバヘッドの軽減を図っている。しかし、その反面、(2)の自動運転機能の一部が制限される。

3. 適用例と効果

3.1 仮想計算機システムの構築

九大計算センタにおける仮想計算機システムの導入目的は、1. 述べたように、システムプログラム開発を含む広義のシステムメンテナンスを運用時間内に並行処理することであったことから、それらに必要な最小の機器構成のシステム (以後、ゲストシステムと呼ぶ) 用のVMと通常のユーザサービスを行うシステム (以後、ホストシステムと呼ぶ) 用のVMの2つだけが動作する単純な仮想計算機システムを構成した。構築のベースとなるM-200システムは、初期導入時 (1981年8月) から現在 (1982年12月) に至るまでに次のように変化した。

- i. 1981年8月～1982年1月
主記憶量16MB, 3CPU,
ディスク15.9GB, MSS,
各種入出力装置
- ii. 1982年1月～1982年10月
主記憶量24MB, 3CPU,
ディスク15.9GB, MSS,
各種入出力装置
- iii. 1982年10月～1982年12月
主記憶量32MB, 3CPU,
ディスク22.8GB, MSS,
各種入出力装置

これらのリソースのうち, ゲストシステム用に3.5MBの主記憶と2ボリュームのディスクを, VMモニタの動作用に0.6MBの主記憶を与えているほかは, 通常ホストシステムに割り当てている.

なお, 仮想計算機システムによる運用は, 1981年12月から翌年4月までの間, 業務繁忙期における混雑回避とAVM改版* 待ちのために, 一時中断した.

3.2 メンテナンス時間の変化

オープンショップ形式の大規模計算センタにおけるシステムメンテナンスには, 概して次のような作業項目がある.

- (1)ハードウェア保守
- (2)OS保守
 - i. 障害修正

- ii. バージョンアップ
- iii. システム編集
- (3)システムプログラム開発
 - i. センタルーチンと称するセンタ運用と密着したプログラムの開発
 - ii. OSの機能強化のためのプログラム開発
- (4)ファイル保守
 - i. バックアップ処理
 - ii. 各種ファイル管理情報(VTCやカタログなど)破壊の復旧
 - iii. ディスク(ファイル)構成の再編成
- (5)データベース保守
 - i. データの追加
 - ii. 階層管理によるデータの移動
- (6)通信回線保守
 - i. 回線属性を定義するシステム編集
 - ii. 通信状況のトレース調査
- (7)ライブラリ管理
 - i. プログラムライブラリのレベルアップとその世代管理

ここで, 1980年11月から1981年7月までの実計算機システムによる運用において, ユーザサービス時間とサービス時間外のメンテナンス内訳を表1に示す.

表1によると, サービス時間外のメンテナンスは平均してサービス時間の30%に達している. また, その内訳を図示すると図1のようになる.

* 16MB以上の主記憶装置を持つシステムに対しては, VMモニタのレベルアップが必要であった.

表 1. サービス時間外のシステムメンテナンス状況
Table 1. System maintenance state at out-of-service time

単位 = 時間

年/月	項目	(1)	(2)	(3)	(4)	(5)	(6)	(7)	合計	サービス 時間
	ハード ウェア	OS	プログラ ム開発	ファイル	データ ベース	通信回線	ライブ ラリ			
'80/11		15.1	15.4	9.4	0.0	4.3	0.0	1.8	46.0	202.8
'80/12		18.5	34.8	6.3	7.9	0.0	0.0	0.0	67.5	211.8
'81/01		18.8	29.2	21.9	1.5	0.0	0.0	0.0	71.4	266.5
'81/02		22.4	0.0	32.1	4.2	0.0	0.0	0.0	58.7	310.3
'81/03		11.0	28.2	107.1	3.6	1.0	0.0	0.0	150.9	188.5
'81/04		14.8	25.6	9.7	5.6	10.1	0.0	0.0	65.8	191.0
'81/05		13.5	50.0	21.3	4.5	5.2	0.3	0.0	94.8	248.7
'81/06		7.8	10.2	7.8	4.9	19.4	17.0	0.0	67.1	240.7
'81/07		6.8	14.8	7.0	4.0	2.9	0.0	0.0	35.5	284.0

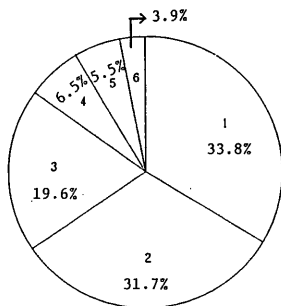


図 1 システムメンテナンスの内容比率
Fig. 1 Ratio of system maintenance contents

1. システムプログラム開発
2. OS保守
3. ハードウェア保守
4. データベース保守
5. ファイル保守
6. その他

図 1 のグラフから解るように、全メンテナンス時間の 65% は OS 保守やシステムプログラム開発に費やされている。

一方、仮想計算機システムにおける運用のうち、インストール時を除いた安定時期（1981年10月～1981年11月）では、サービス時間外のメンテナンス時間は表 2 に示されているように、18%程度にまで減少している。さらに、内訳では OS 保守やプ

表 2. 仮想計算機システム導入後のシステムメンテナンス状況
Table 2. System maintenance state after introduction of a virtual machine operating system

単位 = 時間

年/月	項目	(1)	(2)	(3)	(4)	(5)	(6)	(7)	合計	サービス 時間
	ハード ウェア	OS	プログラ ム開発	ファイル	データ ベース	通信回線	ライブ ラリ			
'81/10		25.3	3.8 (18.1)	0.0 (128.7)	22.3	4.7	2.0	0.0	58.1	297.1
'81/11		16.9	7.3 (15.2)	0.0 (71.3)	7.1	10.9	3.0	0.0	45.1	262.1

() 内はサービス時間内での作業も含めた時間を示す

プログラム開発が全メンテナンス時間の10%までに減少している。しかしながら、実際の作業時間は、実計算機システムの場合よりも逆に多くなっている。又、表2に示されるファイル保守（バックアップ）とデータベース保守は、長時間のユーザーサービスを行った場合には、深夜の閑散な時間帯を利用して一般ユーザのジョブと並行処理することも可能であり、ほとんどサービス時間内に組み込むことができる。このように、仮想計算機システムによる運用では、サービス時間帯の拡大と相まって、実計算機システムにおけるサービス時間外での計算機保守時間の75~80%が縮小可能と思われる。

4. VMモニタの下のOSの処理能力大規模計算サービスを行うOS（以後これをホストOSという）をVMモニタの下で動作させる場合、VMモニタおよびホストOS以外のOS（以後ゲストOSという）の使用するシステム資源の量とVMモニタによるOSに課せられた制限によって、ホストOSを単独で動作する場合より、より少ない使用負荷でシステムのボトルネックが発生する可能性がある。このようなボトルネックの起こりやすいシステム資源として演算時間、主記憶、データ転送能力がある。これらは、いずれもホストOSの総合処理能力を低下させる原因となる。ここでは、ゲストOSのシステム全体に与える負荷は非常に少ないので、VMモニタによる効率低下

を調べるために、ホストOSのみがVMモニタの下で動作している場合の測定を行った。測定時期の運用システムは単独で動作している場合でもそのボトルネックは演算時間にあり、計算志向型のジョブが多い。このため総合処理能力の目安としてTSSセッションおよびバッチジョブによって使用された単位時間当たりのCPU時間を測定した。図2は、

ホストOSを単独で動作させた場合
主記憶28MB --- RV28
主記憶32MB --- R32
VMモニタの下で動作させた場合
主記憶28MB --- VM28
のTSSセッションをふくむ利用者ジョブによって使用されたCPU時間（3CPUのシステムであるので最大180秒）をプロットしたものである。

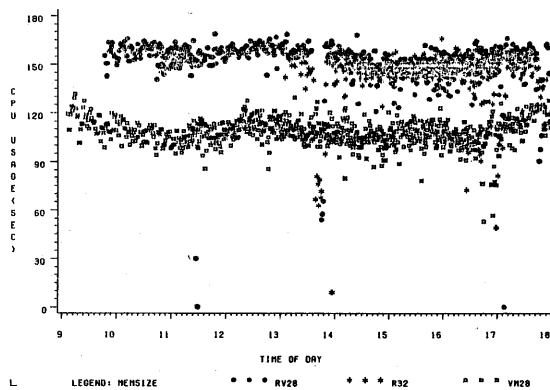


図2 利用者のジョブのために使用された演算時間
Fig.2 CPU usage for user program

なお図3に測定時間におけるTSS端末動作数を示す。バッチ多重度は10から14程度であった。図2からVM

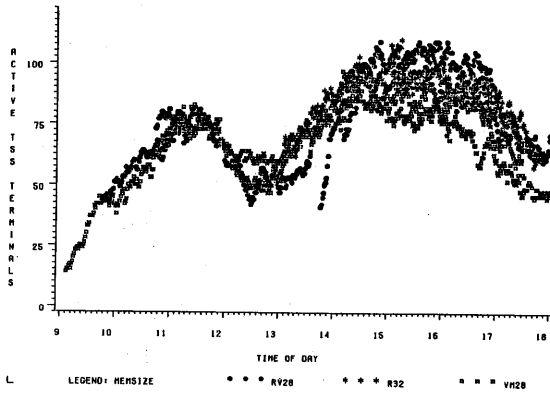


図3 動作端末数
Fig.3 Active terminals

モニタを使用すると利用者に提供できるCPU時間が減少することが認められる。VMモニタの有無による差を予測するために利用者の最も多い14時から16時の2時間のCPU使用時間の平均はそれぞれ110秒、148秒となり差は38秒である。メディアン、Q3(75%点を含む値)求め差をとるとほぼ40秒となることがわかる。比較したシステムは共にTSS端末、バッチ多重度が同程度の利用環境であり、この差はVMモニタの使用したCPU時間と考えてよい。つまりこのような利用環境ではVMモニタを使用することにより40秒程度の計算能力を犠牲にすることになる。言い替えば、測定された運用システムのような演算志向型のジョブの多い使用状態では、利用者用のCPU時間の1/3に近いCPU時間をVMモニタが使用していることになる。CPU全体の62%を利用者に、22%をVMモニタに、残りの16%はOSに使用され、ホストOS単独の場合は利用者に83%、O

Sに17%のCPU時間が使用されている。CPU時間にボトルネックがあり、他の要因はほとんど無視できるようなこの例では、バッチジョブのレスポンスは当然遅れるとしても、TSSのレスポンスについてはTSSの処理優先権をさらに上げることによりホストOS単独の場合と比較して同程度のレスポンスを保証することが、十分に可能である。図4に示すように使用可能な主記憶が少なくなると、急激にCPU使用率が低下する。主記憶の使用がボトルネックの限界値に近い場合の例で、⁵⁾VMモニタを使用することにより、ホストOSのため使用することのできる主記憶が減少して、その結果、ページング回数が増加し、処理能力を低下させている。

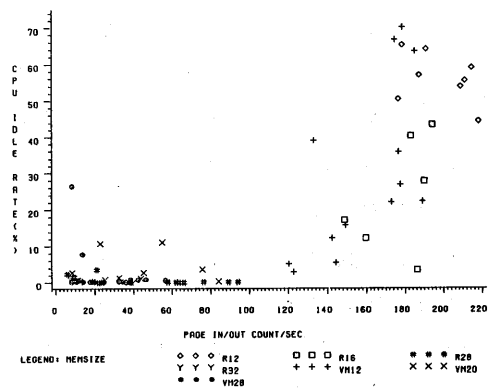


図4 ページングin/out回数とCPUアイドル率の関係
Fig4 Paging in/out vs.CPU idle rate

5. おわりに

以上の結果から、VMモニタの使用によりある程度のCPU時間と主記憶を少々犠牲にしなければならないとしても、VMモニタによって効率低下を

きたす要因の利用率がボトルネックの限界値に達しない場合，VMモニタの使用により，十分な効果を期待できることが分かった．本例でも，CPU時間がボトルネックとなる繁忙期以外は，非常に有効であった．さらに，CPUのVMアシスト機能を強化しVMモニタ使用にかかるCPU時間が少なくなれば，常時VMモニタの下での運用も可能と思われる．

参考文献

- 1) 計算機マニュアル，VM/370
Introduction, GC
20-1800, IBM社
- 2) 計算機マニュアル，VMS概説
8040-3-001, 日立製作
所(株)
- 3) 計算機マニュアル，ACOS-4
システム管理 VM/200説明
書DDG04-1, 日本電気
(株)
- 4) 計算機マニュアル，AVM使用手
引書，64SP-9901-1,
富士通(株)
- 5) 景川，九州大学大型計算機センタ
ーにおけるシステム性能評価につ
いて，富士通サイエンティフィッ
ク・システム研究会資料，No.
15, PP. 6-15 (198
2)