名古屋大学大型計算機センター新システムの運用管理について

長谷川 明生
名古屋大学大型計算機センター

梗概

名古屋大学大型計算機センターの新システムは、汎用機、４２台のベクトルプロセッサを持つスーパーコンピュータ、２台のサーバー型ワークステーションおよび画像処理用ワークステーション７台を中心に構成されている。汎用機用オペレーティング・システムおよび複数種類のＵＮＩＸオペレーティング・システムが共存した大規模分散処理環境である。

本論文では、このような分散処理環境下での利用者管理や課金管理の方式について議論する。

On the Management Facilities of the Computer System of the Nagoya University Computation Center

Hasegawa Akiumi
Nagoya University Computation Center

Abstract

The system of the Nagoya University Computation Center consists of one mainframe, one super computer with 42 vector processors, two server type workstations, and seven graphic workstations. The operating system MSP and the UXP/M, a kind of UNIX, are operated on the mainframe system simultaneously. The super computer is operating under the UXP/VPP, a kind of UNIX operating system. The Operating systems of the other systems are UNIX of different kinds. These systems are connected with each other by LAN and constitute a large distributed system.

This paper describes the management facilities of our new system.

はじめに

昨年度，名古屋大学大型計算機センターは，大規模な計算機システムの更新を行った。現システムは，４２個のベクトル・プロセッサを持つベクトル型並列計算機システム，３ＣＰＵの汎用計算機，２台のサーバー型ワークステーション，および７台の画像処理用ワークステーションをＬＡＮで相互接続した大規模分散処理システムである。オペレーティング・システムは，汎用機用のＭＳＰおよびＵＸＰ／Ｍ，スーパーコンピュータ用のＵＸＰ／ＭおよびＵＸＰ／ＶＰＰ，ワークステーション用にＳｏｌａｒｉｓ２およびＩＲＩＸと変化にとんでいる。

本論文では，大規模分散処理システムでの利用者管理および課金管理方式について，実際の運用経験に基づいて報告する。

システムの概要

図１に，現システムの構成図を示す。

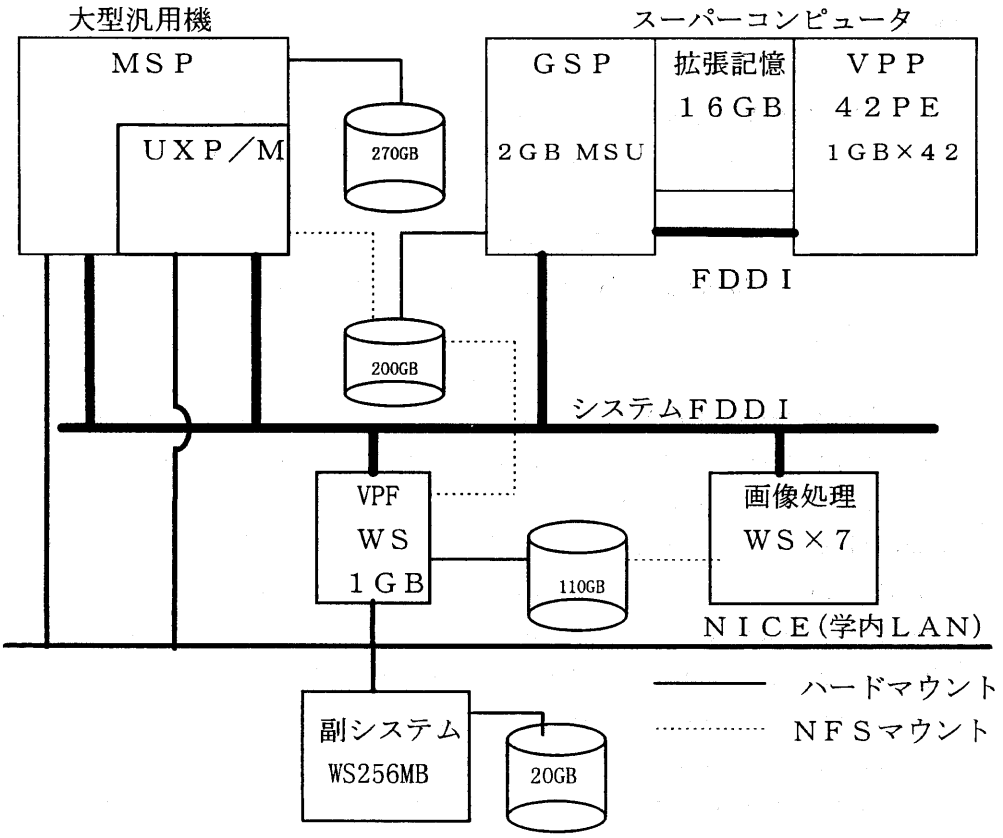

図１ システム構成図

スーパーコンピュータ（ＶＰＰ５００／４２）は、バッチ計算専用のバックエ

ンド・プロセッサとして動作する。スーパーコンピュータへのジョブ投入は、

●汎用機からM－ＶＰＰ連携機能による汎用機のバッチ・ジョブ投入

●汎用機のＵＮＩＸからのＮＱＳによるジョブ投入

●ＶＰＦと呼ぶサーバー型ワークステーションからのＮＱＳによるジョブ投入

が可能となっている。

主な機器の構成を表１に示す。

表1 センターの機器構成

機種	ＯＳ	ネットワーク	備考
M1800/30(3CPU, 1GB)	MSP, UXP/M	FDDI, Ether	汎用機
GSP(2CPU+1VU, 2GB, 16GBSSU)	UXP/M	FDDI	VPPサービスプロセッサ
VPP500/42(42PE, 1GB/PE)	UXP/M VPP	FDDI	ベクトル並列型計算機
S4/1000E(8CPU, 1GB)	Solaris2.4	FDDI, Ether	VPPフロントエンド
S4/1000E(2CPU, 256MB)	Solaris2.4	Ether	UNIX副システム
S4/20H(256MB)×5	Solaris2.4	FDDI	画像処理システム
S4/10(256MB)	Solaris2.4	FDDI	画像処理システム
S4/10(196MB)	Solaris2.4	FDDI	画像処理システム
SGI Reality Engine(256MB)×2	IRIX5.3	FDDI	画像処理システム

ＶＰＰは単独ではファイル処理機能等を持たないので、ハウスキーピングのためのＧＳＰシステムと一体で運用する必要がある。これらのシステムはすべてＦＤＤＩおよびＥｔｈｅｒｎｅｔにより結ばれており、大規模な分散処理システムを構成している。

利用者管理システム

これらのシステム間で利用者管理は自動化されている。ＵＮＩＸシステムでは、ＮＩＳによる利用者管理を行っている。ＮＩＳのマスター・サーバーは、ＵＮＩＸ副システムで、ＶＰＰフロントエンド・システムがスレーブ・サーバーとして動作し、システムＦＤＤＩ上のＵＮＩＸホストにＮＩＳマップをサービスする。ただし、ＶＰＰおよびＧＳＰシステムはバッチ専用のシステムなのでＮＩＳマップではなくパスワード・ファイルによる管理としている。

利用者登録の流れを図２に示す。利用者登録は、課金処理とあわせて毎朝実行される。ＵＮＩＸ上では、利用者登録プログラムは/etc/passwd ファイルやＮＩＳマップの更新だけでなく、ホーム・ディレクトリの作成、.login ファイル等の準備、quota の設定等をおこなう。

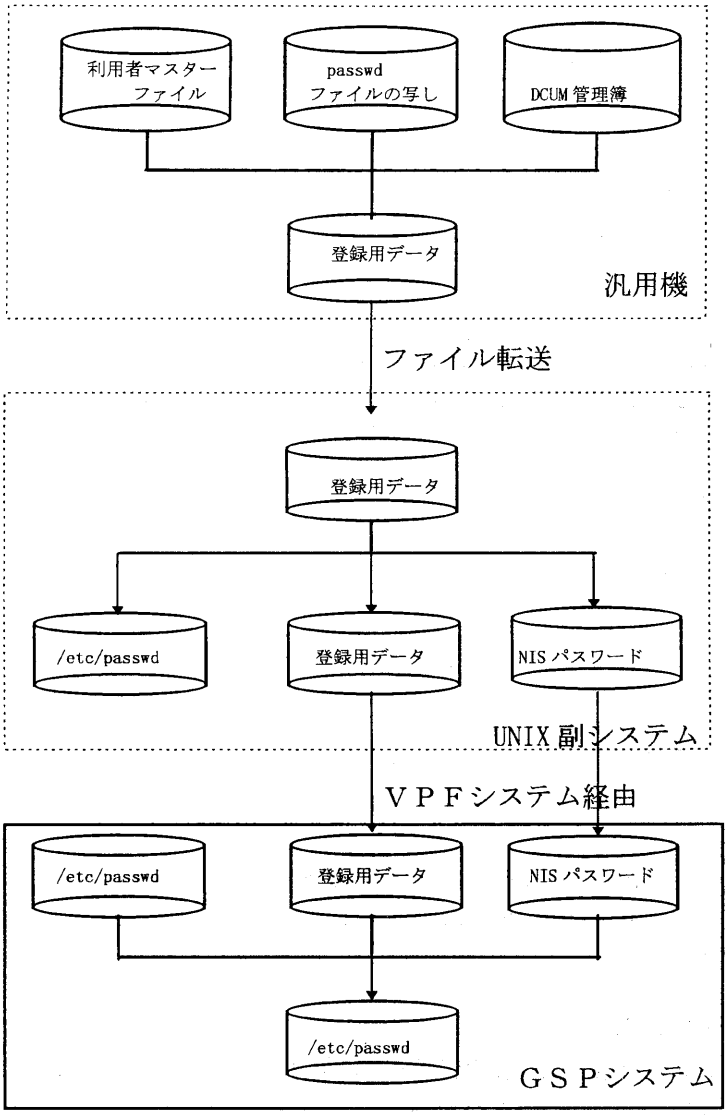

図２ 利用者登録の流れ

課金処理

本センターの課金項目は、ＣＰＵ使用料金、出力負担経費（プリンター出力、XY プロッタ）およびディスク使用料金である。汎用機およびスーパーコンピュータのＣＰＵは、ジョブ単位の課金処理をおこなっている。他のＵＮＩＸホストについては、プロセスのＣＰＵ時間を１日分について合算処理する。

ディスク使用量は、毎日定時にパトロール・プログラムを走らせて調べる。

ＵＮＩＸ系の印刷出力は、センター内のネットワークプリンタのどれかに出力する。これらのネットワークプリンタは、学内ＬＡＮとは独立したネットワーク

に接続している。そして、このＬＡＮと学内ＬＡＮとの間にプリンタ・サーバを兼ねたワークステーションを用意した。このような構成により、利用者チェックおよび課金管理が容易になっている。

これらの独立にシステムごとにあがってくる課金データは、すべて汎用機に転送し汎用機上の課金プログラムで課金処理している。

管理システムの設計方針

利用者登録システムおよび課金管理システムの開発にあたって、極力システムの標準機能を手を入れずに利用することを原則とした。また、新システムへの移行およびテスト期間を最小とするために、既存の汎用機上で動作している課金や登録管理システムの機能を前提にＵＮＩＸホスト等の利用者登録システムや課金データの形式を設計した。

このような方針で開発したため、システム全体を通してのテスト期間は10日程度しか確保できなかったが、新システムへの移行を短時間で行うことができた。

ジョブクラス設計と運用状況

新システムの中心は、ベクトル並列型スーパーコンピュータである。このようなシステムの大規模計算機センターでの本格運用は、本センターが初めてである。複数プロセッサを並列に動作させるジョブに対する利用者の要求は、システム設計時には、ほとんど予想できなかった。表２に現在のＮＱＳジョブ・クラスとその制限値を示す。

表２ ＮＱＳキューと制限値

キュー名	利用目的	CPU時間	メモリー	PE数
c	Fortranの翻訳	30分	100MB	-
x	VPPでの実行	600分	300MB	1
y	VPPでの実行	600分	900MB	1
z	VPPでの実行	600分	900MB/PE	2～16

このキュー設計は、本格的な並列プログラムの開発に時間がかかると予想し、シングルＰＥジョブ主体を想定して行われた。

実際のジョブ処理状況を表３に示す。

表３ジョブ処理状況表

システム	年月	計算サービス時間(時:分)	処理件数	CPU時間 時:分	CPU時間/件 (秒)
VPP500	H7.12	599:12	3759	5454:47	5224.05
	H8.1	605:41	4549	7293:53	5772.25
	H8.2	620:12	4677	10012:03	7706.52
	H8.3	445:27	3430	6661:31	6991.68
	計	2270:32	16415	29422:14	6454.64
VP2600	通年	5037:13	25537	4857:25	684.76

また、表４にＰＥの使用状況を示す。

表４　ＰＥ使用状況表

年月	利用種別	ベクトル化率	平均ＰＥ数	メモリサイズ
１月	Ｍ－ＶＰＰ連携	77.0%	1.29	131.6MB
	ＮＱＳ	37.1%	2.63	409.1MB
２月	Ｍ－ＶＰＰ連携	72.5%	1.08	108.5MB
	ＮＱＳ	29.5%	2.94	376.9MB

ジョブ統計の平均ＰＥ数の値から、当初の予想と異なり複数ＰＥを使用するジョブがスーパーコンピュータ利用のかなりの部分を占めていることがわかる。複数ＰＥを使用した場合、最大ＣＰＵ時間を消費したＰＥのみＣＰＵ課金されるという誘導的な課金体系の効果が予想以上に表れたとも言える。また、複数ＰＥジョブでは、４ＰＥ以上が同時に使用されることが多い。導入当初からの並列ジョブの利用者に確認したところ、同一プログラム中で複数データを同時並列処理しているとのことであった。このような形態のジョブでは、プロセスの起動処理とファイル入出力の競合だけが問題となるだけで、ほぼＰＥ数に比例して高速化がはかれる。このようなジョブとは、別に複数ＰＥで並列処理することによって高速化をねらったジョブも投入されている。

スーパーコンピュータ利用者は、現在のところ、シングルＰＥジョブ主体の旧システムからの移行者とフロント・エンド・システムや汎用機のＵＮＩＸシステムからＮＱＳジョブを投入する並列ジョブ利用者とに2分される。現行のジョブ・プロフィールから、ＮＱＳキューの資源配分の再検討をしなければならないが、４２プロセッサというのは、並列ジョブの増加を予想すると、けっして十分な数ではないと考えている。

おわりに

本センターの新システムについて運用管理方式を中心に報告した。現在は、新年度の新規利用者登録も済み安定に稼動している。

短期的課題として、スーパー・コンピュータに関連した資源配分の再検討およびジョブ管理パラメータとしての課金体系の見直しがある。

長期的には、汎用大型計算機の役割の検討を含め、将来を見通した計算機環境の検討しなければならない。