

航技研数値風洞（NWT）における  
ジョブスケジューラについて

2T-9

土屋雅子 末松和代  
航空宇宙技術研究所

1. まえがき

航技研において平成5年2月より運用を開始した数値風洞（NWT：Numerical Wind Tunnel）は、要素計算機（PE）にベクトル計算機を配置する分散主記憶型並列計算機システムであり、中核に140台のPEを並列配置している。また、NWTは、既設の大型電子計算機システム（FACOM VP2600）をフロントエンドシステムとして有機的に結合した複合計算機システム（NSシステムと呼称）を構成する（図1参照）。NSシステムの運用では、先進的な大規模数値シミュレーションを可能にすること、ハードウェアが有する超高速処理性能を十分に引き出し、かつ、最大限に高度有効利用を図ることを最重要課題としている。この課題に基づき、並列計算機システム用ジョブスケジューラが開発された。NSシステムでは、このスケジューラを母体として、航技研独自の各種運用機能を組込んで開発されたジョブスケジューラの運用を平成6年11月より開始した。本報告は、ジョブスケジューラの運用機能と実運用におけるジョブスケジューラの効果について述べる。

2. ジョブスケジューラの運用機能

並列計算機システム用ジョブスケジューラはNSシステムの超高速処理性能を十分に引き出すための本質的な役割を果たす。一方、NSシステムの運用規則を定め、秩序ある運用を実現するためには、各種の運用機能が必要となる。実運用においては規則どおりの運用だけでは対処し得ない事態が多々発生するので、ジョブスケジューラにはそのような事態

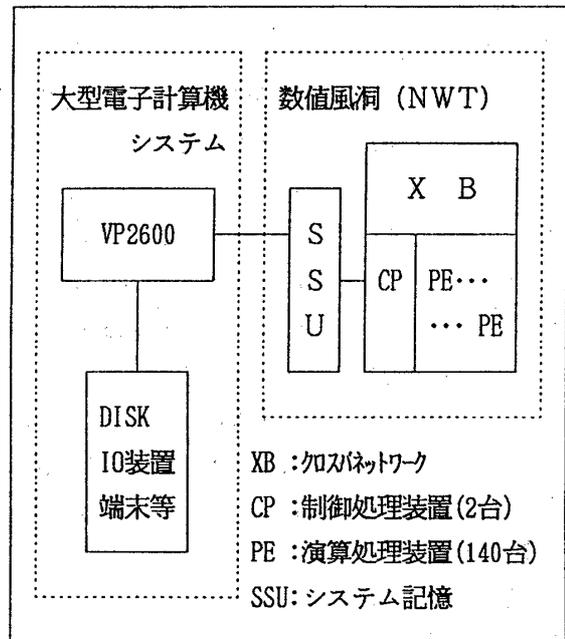


図1. NWT中核部のハードウェア構成概念図

に即応できる柔軟性も要求される。以上のことを考慮して、ジョブスケジューラには下記の運用機能を搭載している。

- (1) 非並列ジョブの起動を制限する機能
- (2) 同一ユーザジョブの起動を制限する機能
- (3) ジョブが要求するPE台数およびCPU時間を制限する機能
- (4) 特定ユーザのジョブを優先実行する機能
- (5) 特定ジョブを優先実行する機能
- (6) ジョブの優先度を調整する機能
- (7) 同一ユーザジョブの実行順序を保証する機能
- (8) 運用時間帯ごとに最大並列度を定義する機能
- (9) ジョブ状態表示に要する情報を出力する機能

3. ジョブスケジューラの効果

新たなジョブスケジューラを搭載したシステム運用を開始して以来、既に9ヵ月を経過し、ジョブスケジューラはバグの除去と小さな改良とによって、

ほぼ所期の目標どおりのものに成長した。実運用の中で確認されている本ジョブスケジューラの有用性を列挙すると以下のとおりである。

(1) 並列計算機システム用ジョブスケジューラは「PEの高效率利用」、「要求台数のPEの割当」および「適切なターンアラウンドタイムの保証」の実現を目標とするスケジューリングアルゴリズムを特徴としている。表1は勤務時間帯において、実行可能ジョブがシステムに十分滞在している運用日におけるPEの平均稼働率を示している。同表に示すとおり、実運用において、ジョブが存在する限り、実行可能なジョブが起動され、PEの稼働率は高く、かつ、コンスタントに維持できることが確認できる。

(2) 優先度を調整する機能により並列度に応じたジョブのスループットを制御することが可能である。これにより、特定の並列度を有するジョブが急激に増加しても柔軟に対処できるので、実運用には極めて有効である。

(3) 実運用では、緊急を要するジョブが予想以上に発生しているが、特定ジョブを優先実行する機能により非常に柔軟に対処できる。

(4) 運用時間帯ごとに最大並列度を定義する機能により、ユーザが多い昼間の勤務時間帯には、小規模並列ジョブのレスポンスを爽快に返せる。また、大規模な高並列度ジョブのデバッグ環境は最善である。図2はNWTのジョブ実行状況をジョブ名、ステップ実行順位、割当PE台数ならびにジョブ実行経過時間で示している。同図に示すとおり、昼間の勤務時間帯には、最大並列度を64に定義し、割当PE台数が64台までの小規模並列ジョブの実行処理を運用している。

(5) 同一ユーザジョブの実行順序を保証する機能により、ジョブ途中結果のチェックポイントファイルを継続し、連続実行するユーザジョブは確実に投入順に実行できる。これにより、結果の収束に長時間を要するジョブは複数ジョブ構成とすることが可能となり、途中結果のスナップショットを可視化処理により検証できる。

(6) ジョブ状態表示に要する情報を出力する機能に

表1. 平均PE稼働率

DATE	稼働率 (%)	NWT execute job			
		JOBNAME	STEP	PE-ALC	TIME
05/16	95.90	A66S364	02/11	6	0:25:46
05/17	90.11	J15S095	03/04	1	3:05:31
05/24	91.14	L07S509	05/07	64	3:05:11
05/25	90.82	L26S957	06/07	16	2:32:59
05/29	94.00	L30S773	02/02	20	1:55:40
05/31	93.92	L30S774	02/02	20	0:26:51
06/02	93.01	N71S194	04/04	5	2:32:16
06/08	92.24	P45S145	02/02	1	0:21:59
06/15	93.90	P92S379	02/02	1	1:05:58
06/16	95.85	P95S397	01/01	1	1:14:05
		P95S398	01/01	1	1:10:44
		P95S399	01/01	1	1:04:50

図2. NWTジョブ実行状況

より、ユーザがジョブのターン・アラウンド・タイムを容易に予測できる。

#### 4. まとめ

NWTは超高速処理性能を有しているので、ジョブの到着がラッシュアワー的に発生した場合にも、また、過負荷なワークロード状況においても、小規模ジョブの処理は瞬間に終了する。システム状態が過負荷なワークロード状況では、ジョブスケジューラは非常に威力を発揮する。しかし、夜間等のジョブ投入が少なく、かつ、ワークロードが軽負荷の状況には、PE稼働率は低下する。このため、システムのワークロードの負荷状況に応じて稼働PE台数を自動的に制御する省電力機能が必要である。ジョブスケジューラの今後の課題として、本機能の開発を検討している。

#### 参考文献

- 1) 末松和代：「並列計算機システム用ジョブスケジューラ」(航技研報告)