

PRIMEPOWER600/400/200のハードウェア

Hardware of PRIMEPOWER600/400/200

あらまし

富士通で開発した高性能マイクロプロセッサ“SPARC64 GP”を搭載し、エン트리ユー
スから基幹業務に至る幅広い性能レンジをカバーするUNIXサーバPRIMEPOWER600/
400/200について、ユニットの高性能化、高速接続技術、コストパフォーマンスおよび
信頼性の観点から紹介する。

プロセッサ、メモリ、IOチャンネルなどの各ユニットレベルでの高性能化とこれらのユ
ニットの高速接続技術により、各モデルにおいてクラス最高水準のシステム性能を達成
している。また、モデルごとに最適化されたハードウェアは、コストパフォーマンスお
よび設置性の面においても優れ、柔軟なシステム構築を可能としている。さらに、信頼
性の面においては、プロセッサの1次・2次キャッシュにECC(Error Checking and
Correcting)を採用するほか、メモリ、IOチャンネルとのデータ授受にもECCが採用され
高い信頼性が保証されている。

Abstract

The UNIX server “PRIMEPOWER600/400/200” uses Fujitsu’s SPARC64 GP high-performance microprocessor. The new models can meet a wide range of performance requirements and are suitable for applications ranging from entry systems to backbone systems for business. This paper describes the high-performance, high-speed connection feature, cost effectiveness, and reliability of the hardware of these models. Each model of the PRIMEPOWER600/400/200 provides the highest system performance among the servers of the same class thanks to the enhanced performance of the processor, memory, I-O channels, and other units and the high-speed technology incorporated in each unit. The hardware optimized for each model provides a high cost-effectiveness, convenience at installation, and flexible system construction. To guarantee a high reliability, the models use an error checking and correcting (ECC) feature not only for the first and second cache memory of the processor but also for data transfer between memory and I-O channels.



須藤一義（すどう かずよし）

第三コンピュータ事業部
第一開発部 所属
現在、UNIXサーバの開発に従事。



福田一郎（ふくだ いちろう）

(株)PFU
第一事業部
第一技術部 所属
現在、UNIXサーバの開発に従事。



池上雅雄（いけがみ まさお）

(株)PFU
第一事業部
第四技術部 所属
現在、UNIXサーバの開発に従事。

ま え が き

UNIXサーバは、性能、信頼性、拡張性が評価され、インターネットやデータベース市場の急速な拡大に伴い、そのシェアを拡大させている。しかし、近年、IA(Intel Architecture)サーバの高性能化・低価格化により市場が激化、CPUの高性能化はもちろんのこと、同時に高いシステム性能と高信頼性が要求されている。

このような状況の中で、富士通が開発したPRIMEPOWER600/400/200は最先端の技術を駆使し、性能および信頼性を高い次元で提供する。また、近年、設置性に対する要求も非常に強く、小型・高性能なラック搭載型モデルについても提供する⁽¹⁾。

システムの概要

● システム諸元

PRIMEPOWER600/400/200のシステム諸元を表-1に示す。ラインナップの大きな特徴は、すべてのモデルで自立型とラック搭載型モデルを提供していることである。

● システム構成

CPUにはSPARC V9 64ビットアーキテクチャに準拠した富士通製の高性能マイクロプロセッサを採用した。CPUとメモリおよびIOチャンネルとは、UPA(Ultra Port Architecture)バスインタフェースで接続されている。こ

のUPAバスインタフェースには、システムコントローラとクロスバスイッチとIOチャンネルが接続される。システムコントローラは、CPUおよびIOチャンネルからのアドレスの制御とメモリの制御を行っている。

クロスバスイッチはデータのルートを制御している。メモリはウェイ^(注1)という構成でこのクロスバスイッチに接続されており、モデルにより2ウェイから最大8ウェイのメモリ構成をとることができる。IOチャンネルには66 MHz・64ビットで動作可能なPCI(Peripheral Component

表-1 システム諸元

		PRIMEPOWER200	PRIMEPOWER400	PRIMEPOWER600
CPU	プロセッサ	SPARC64 GP		
	プロセッサ数	基本	1	1
		最大	2	2
メモリ	容量	256 Mバイト		
		基本	8 Gバイト	16 Gバイト
PCI	スロット数	基本	6	10
		最大	6	12
内蔵装置	基本	3.5インチFPD(保守用)/CD-ROM		
	オプション	DAT/8 mm/DLT*		
基本インタフェース		LAN(100BaseTX または10BaseT)		
		シリアルポート × 2		

★：DLTは400/600のみ

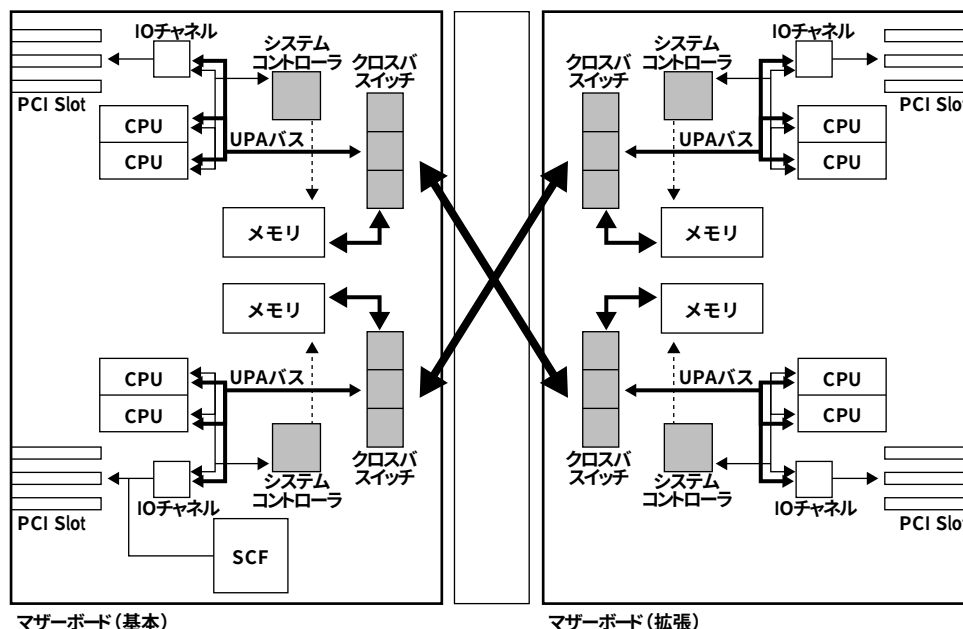


図-1 PRIMEPOWER600のシステム構成図
Fig.1-PRIMEPOWER600 system configuration.

(注1) 64バイトアドレス単位で最大8個のメモリサブシステムに分散配置する。この個々のメモリサブシステムをウェイと呼ぶ。

Interconnect)バスを採用し、IOの高性能化を図っている。また、CPUとは独立したSCF(システム監視機構：System Control Facility)を実装し、システム監視機能を強化するとともに、システムの可用性・耐故障性を向上させ、システム全体の信頼性をより強固なものとしている。PRIMEPOWER600のシステム構成を図-1に示す。

【PRIMEPOWER600システム構成の特徴】

PRIMEPOWER600は最大2枚のマザーボードで構成される(マザーボードについては後述)。この1枚のマザーボードは4CPU分のスロットと6個のPCIスロット、16個のDIMMスロットを持つ。基本構成はマザーボード1枚で構成されるが、業務の負荷やIOの増設に柔軟に対応できるようにマザーボードの増設が可能になっている。この増設により、最大8CPU、最大16 Gバイトのメモリおよび12スロットのPCIスロットを使用することが可能となる。

【PRIMEPOWER400/200システム構成の特徴】

PRIMEPOWER400/200システムでは1枚のマザーボードとIOチャンネル部で構成される。最大4CPU(PRIMEPOWERは2CPU)までの線形的な性能が得られるよう、ひとつのシステムコントローラによって集中管理するシステム構成をとっている。

ハードウェアの概要

● CPUモジュール

本モジュールは、富士通製高性能マイクロプロセッサ“SPARC64 GP”と大容量2次キャッシュメモリである高速SRAM(Static RAM)で構成される。本モジュールの外観を図-2に示す。以下にCPUモジュールの特徴を示す。

(1) 最大4命令実行と大容量キャッシュによる高性能化

本CPUは、最大4命令を同時実行可能なスーパースケラ方式、アウトオブオーダー命令実行、高度な分岐予測機構など先進のマイクロアーキテクチャの採用により

高性能化を実現している。また、大容量1次キャッシュ(命令・データそれぞれ128 Kバイト)・2次キャッシュ(最大8 Mバイト)の採用により、キャッシュミス率を低減し高いメモリ性能を実現している。

(2) ECCの採用による高信頼化

CPUの1次および2次キャッシュメモリにECCを採用し高信頼化を図っている。また、UPAバスにもECCを採用することによりCPU内部からメモリおよびIOチャンネル間のすべての転送がECCにより保護される。

● システムコントローラ

システムコントローラはモデルごとに最適化されたパッケージとテクノロジーにより、高速な周波数での動作を可能としている。ここでは、PRIMEPOWER600に使用されているシステムコントローラの高性能化と高信頼化の手法について述べる。

【高性能化】

(1) 高密度実装技術の採用

PRIMEPOWER600用のシステムコントローラには、CMOS LSIを使用し、このシステムコントローラチップを高密度731ピンBGAパッケージに実装している。これを4個使用することにより8CPU・メモリ8ウェイ分の大規模な回路と多数の信号線を分散管理し、しかも高速な動作を可能としている。

(2) UPAバスの並列動作によるバス性能の向上

バスの高いスループットを得るため、UPAバス上のコマンドは並列に効率良く動作させることが要求される。PRIMEPOWER600のシステムコントローラは合計4本のUPAバスを制御し、最大104個のコマンドを並列に処理することができる。システムコントローラは、メモリアクセス制御機構・キャッシュデータ転送機構・DTAGアクセス制御機構などに分割して、それぞれを多重化するとともに、各資源を効率良く動作させることで並列動作性を高めている。

【高信頼化】

・DTAGのECC化

DTAG^(注2)はメモリセルより構成されるため、ソフトウェア^(注3)はシステム動作に大きなダメージとなる。このため、DTAGにECCを採用することにより信頼性を大幅に向上させている。

● クロスバスイッチ

クロスバスイッチもシステムコントローラと同様にモデルごとに最適化された構成をとっている。クロスバス

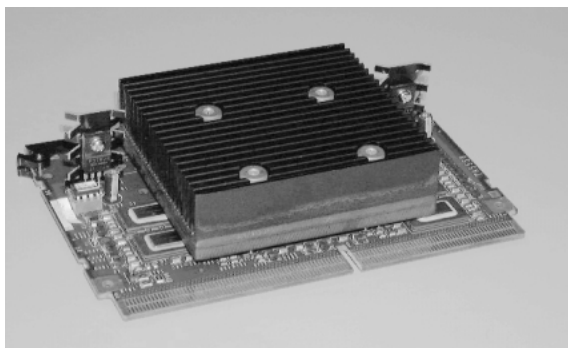


図-2 SPARC64 GP(400 MHz)CPUモジュール
Fig.2-SPARC64 GP(400 MHz) CPU module.

(注2) CPUすべてのキャッシュの状態情報を保持する部分。

(注3) α線などにより散発的に起こる誤り。

イッチにはCPUとIOチャンネルおよびメモリが接続され、それぞれの間のデータ転送を制御する。高性能化の特徴としては以下が挙げられる。

すべてのメモリウェイは並列動作が可能である。したがって、CPUまたはIOチャンネルからのメモリアクセスにおいて、メモリ競合が極めて少なく、非常に高いメモリ性能を実現している。PRIMEPOWER600は、最大8ウェイのメモリ構成をとり、最大9.6 Gバイト/秒というメモリリード性能を実現している。

● マザーボード

PRIMEPOWER600のマザーボードを図-3に示す。

【PRIMEPOWER600マザーボードの特徴】

(1) 高密度実装

4個のCPUモジュール用スロット、16個のDIMM(Dual In line Memory Module)スロット、システムコントローラ、クロスバスイッチ、IOチャンネル、6個のPCIスロットを1枚の基板上に効率良く実装する。

(2) マザーボード間的高速伝送

最大構成時は2枚のマザーボードがバス接続される。それぞれのシステムコントローラ間はアドレスやキャッシュの状態情報を、クロスバスイッチ間はデータを、UPAバスと同じ高速な周波数で伝送する。本バスは以下を特徴とし、2枚のマザーボード構成時でも高いシステム性能を得ることが可能である。

- ・144ビットのデータバスを2系統実装
- ・システムコントローラ間バスを2系統実装
- ・100 MHzの高速伝送が可能

【PRIMEPOWER400マザーボードの特徴】

- ・大型18層基板の採用

サイズが422 mm × 420 mmの18層基板に、4個のCPU

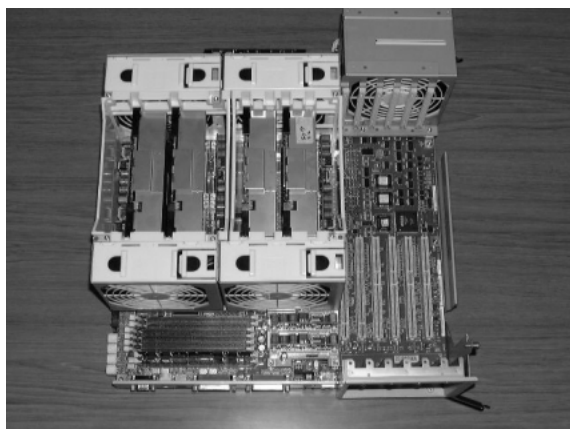


図-3 PRIMEPOWER600のマザーボード
Fig.3-PRIMEPOWER600 motherboard.

モジュール用スロット、16個のDIMMスロット、システムコントローラおよびクロスバスイッチが高密度に実装されている。この高密度実装技術により、各ユニットは最短ルートで配線されるため100 MHzという高速な周波数での動作が可能となった。

● 主記憶

【高性能・大容量メモリシステム】

(1) メモリアクセス時間の短縮

PRIMEPOWER600では、メモリアクセス時間の短縮を図るため、コマンド(リードかライトかを示す情報とアドレス)の入力からメモリ制御回路の起動を最短1クロックで動作するように、コマンドキューイング制御部のパイパス回路を設けた。

また、データがクロスバスイッチを通過する時間を最短1クロックとしている。これは、通常では入り口と出口で2クロックかかるところを、データの入り口のバッファをバイパスさせることで実現している。

(2) 大容量化

メモリ素子はDIMM上に高密度実装され、大容量化を実現している。1枚のDIMMに最大で18個のメモリ素子が実装され、256 MビットSDRAM(Synchronous DRAM)素子を使用したDIMMでは512 Mバイトの大容量化を実現している。これにより、1枚のマザーボードに8 Gバイト、最大構成時は16 Gバイトもの主記憶を持つことが可能となった。

(3) データ幅の拡大

PRIMEPOWER600では1ウェイあたり、上記のDIMMが2枚で構成される。1枚あたりのデータ幅は72ビットであり、2枚で144ビットのデータ幅を持つ。データ幅の拡大により、メモリアクセス時のスループットを向上させている。

(4) バンク制御

すべてのモデルでは、4バンクのSDRAM素子を使用している。システムコントローラは効率よく4バンク制御を行うことでメモリのスループットを向上させている。

【メモリシステムの高信頼化】

全モデルでメモリはECCにより保護される。ECCコードは、8バイト単位に8ビットで定義され、1ビットエラーの検出と訂正、2ビットエラーの検出機能を備える。

● IOチャンネル

IOチャンネルはUPAバスに接続されて、UPAからPCIバスへのブリッジ機能を持つ。IOチャンネルは、PRIMEPOWER200で2個、PRIMEPOWER400/600で最大4個の実装が可能となっている。以下に特徴を述べる。

(1) 66 MHz PCIバスの高性能化

すべてのモデルに66 MHzで動作可能な64ビットPCIスロットを持ち、このPCIバス1本あたりのスロット数を1～3に抑えることにより安定した性能を実現する。PRIMEPOWER400では6本、PRIMEPOWER600では、最大8本のPCIバスを持ち、IOの負荷を効率良く分散させることが可能となっている。

(2) 高信頼化

アドレス・データのパリティチェック機構を有しデータの正当性を保証する。また、PCIバスを監視しバスのハングアップを防止する。

● SCF (System Control Facility)

SCFはCPUとは独立した専用のプロセッサで構成され、ファン、電源、システム状態の監視、周辺装置の電源投入・切断、異常監視を一括して制御している。

(1) 電源、ファンの冗長化・活性交換

冗長化により高い可用性を実現するとともに、活性交換により保守作業時間の短縮、作業性の向上が図られている。活性交換時はマシン管理ソフトウェアにより保守作業を支援する。

(2) 監視機能の強化

ファンの状態、CPUの動作状態、本体内部温度、環境温度などを監視し異常時の通知を行う。

(3) ハード故障情報の採取

マザーボードのシステムコントローラで、アクセスの異常などを検出した場合、自動的にアクセスを停止させる。SCFはこのハード故障情報を速やかに吸い上げSCF内部に保存する。そして、すぐにシステムを再起動する。このハード故障情報により故障か所の特定を容易にすることが可能になった。

● システム実装技術

CPUは周波数とともに性能も向上している。その一方で高速な周波数になるにつれて消費電力も増大し、その冷却方法・実装構造が大きな課題の一つとなっている。そのため、以下のような技術・設計手法を採用したので、次世代CPUにも対応する設置性に優れた、小型のラックマウントタイプモデルの開発も可能となり、開発期間も大幅に短縮することができた。

【冷却技術】

(1) 実装構造設計に精度の高い冷却シミュレーションを実施しCPUの高速化に対応した。

(2) CPU用に高性能な冷却用フィンを開発し冷却効率を向上させた。

(3) 高性能なファンを採用することにより、冷却性能を

より向上させるとともに、信頼性も向上させている。

【3Dコンカレント設計】

構造設計および構造図面は、すべてPRO-E(3次元CADシステム)ツールを使用することによりデータ化が図られ、ほかの様々なツールとのデータ連携が可能となっている。PRIMEPOWER600/400/200の構造検証では、PRO-EデータとVPS(バーチャルプロダクションシミュレータ)を連携させることにより開発効率の大幅な向上が図られた。VPSの特徴を以下に記す。

- ・組立て、分解による製造性の検証が可能
- ・稼働部の干渉、工具の操作性の検証が可能
- ・保守作業性の検証が可能(人間作業も表示可能)
- ・部品ごとに強調表示、角度を変えた表示が可能
- ・シミュレーション過程の記録・再現が可能

上記により、製造からユーザフィールドに至るあらゆる場面での問題点を設計時点で検証することが可能となった。これらのツールと製造・試験・保守部門とのコ

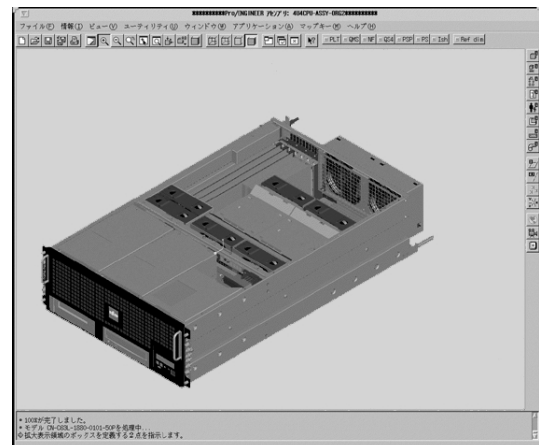


図4 PRIMEPOWER400 VPS画像
Fig.4-PRIMEPOWER400 VPS Image.

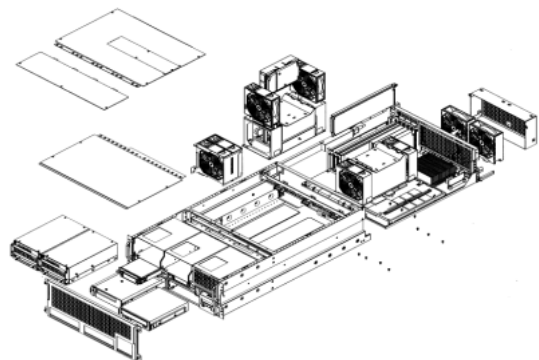


図5 PRIMEPOWER400の分解斜視図
Fig.5-PRIMEPOWER400 decomposition squint chart.

ンカレント設計を強化することにより、大幅な開発期間の短縮と品質の安定を同時に実現することが可能となった。PRIMEPOWER400のVPS画像を図-4に、分解斜視図を図-5に示す。

む す び

PRIMEPOWER600/400/200のシステム構成、ハードウェアの概要について述べた。今回採用した高性能・高信頼化の技術を実現することで、安心して使用できるシ

ステムを提供し、顧客のニーズに十分応えられるものと確信している。

今後も性能・信頼性に対する要求は強まるばかりであり、この要求に応え、顧客のニーズを先取りしたシステムの設計をしていきたいと考えている。

参考文献

- (1) 牟田：GRANPOWER7000シリーズの高性能化技術。
FUJITSU, 49, 1, pp.26-30(1998).

