サイエンテイフィック・システム研究会
科学技術計算分科会 第2回会合資料

新スカラサーバGP7000F
ハードウエアについて(報告)

1999年11月5日

富士通(株)コンピュータ事業本部
第三コンピュータ事業部
安藤 努
tando@cs.fujitsu.co.jp

All Rights Reserved, Copyright © FUJITSU LIMITED 1999

昨年度の科学技術計算分科会第２回会合で 富士通のUNIXサーバへの取り組み」としてスーパサーバを中心にプラットフォームの展開について説明を行った。

このスーパサーバを今年7月にGP7000Fモデル2000として発表したので、本報告ではモデル2000の特長・実現技術を中心に説明し、またIO関連についても説明する。

本報告では、まずHPC環境での要件、開発方針及びGP7000Fファミリでのモデル2000の位置づけを説明した後、モデル2000の特長・実現技術を説明する。

更に、I/O関連ではGP7000FでのIOへの取り組み及び10月に発表したRAID装置GR720の概要を説明する。最後に今後のUNIXサーバの展開計画を紹介する。

All Rights Reserved,Copyright ©）富士通株式会社 1999

急速に大規模、高速化して来ているスカラーサーバーがHPC環境で注目されて来ている。

このHPC環境でのスカラーサーバへの要件としては、

高性能 ･スケーラビリテーが最も重要な要件であり、プロセッサ、システムアーキ、並列処理での実行性能が重要な要素となる。

また、プロセッサ性能への追随、業務変化、解析量の増大への対応への関心も強く即応性も重要となっている。

更に、全世界からの利用に代表される24H365日のサービス提供から信頼性の要件、

標準性も重要な要件である。

これらの要件を満足出来るスカラーサーバは、大規模UNIXサーバである。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

この様なHPC環境へ適合させるUNIXサーバの開発方針を説明する。

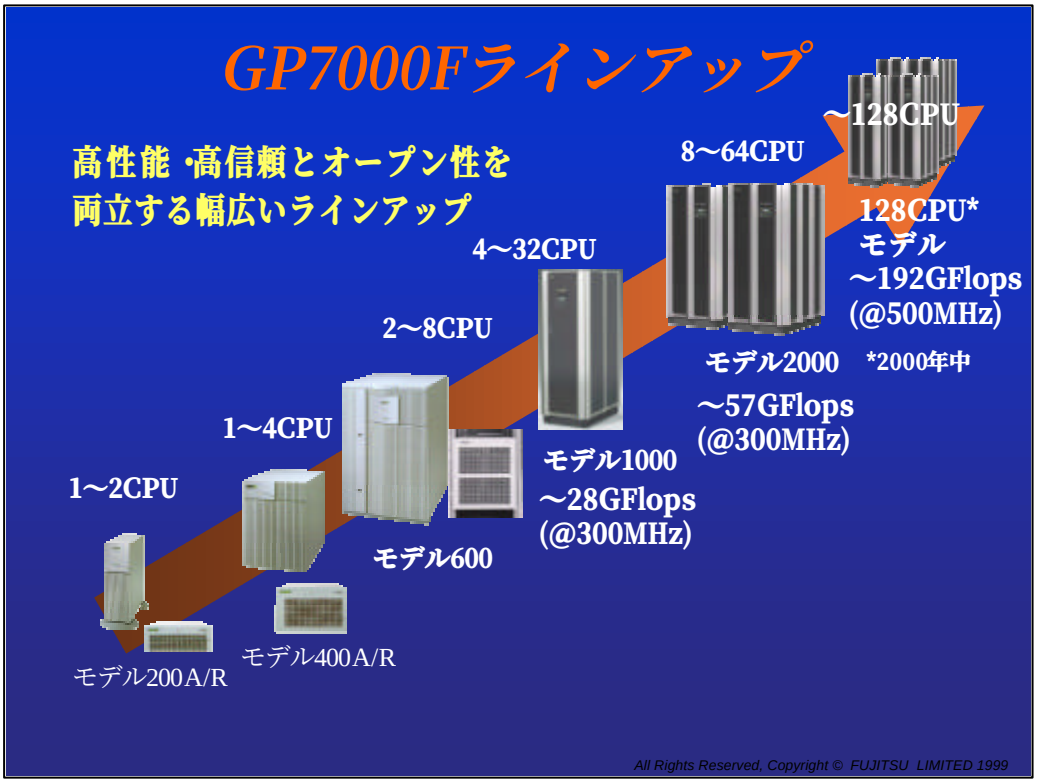
標準性の観点から、Solaris を基本とし、ISV及びIHVを活用して行く。

ハードウエアプラットフォームとしては、GP7000Fを提供する。

性能/スケーラビリテー,信頼性の観点からは、プロセッサ、システムアーキテクチャ等重要となるキーテクノロジは自社技術で開発し提供する事である。

これにより、グローバル性との両立を図り、世界に先駆けたスーパスケーラブルサーバを開発・提供する。

All Rights Reserved,Copyright (C) 富士通株式会社 1999



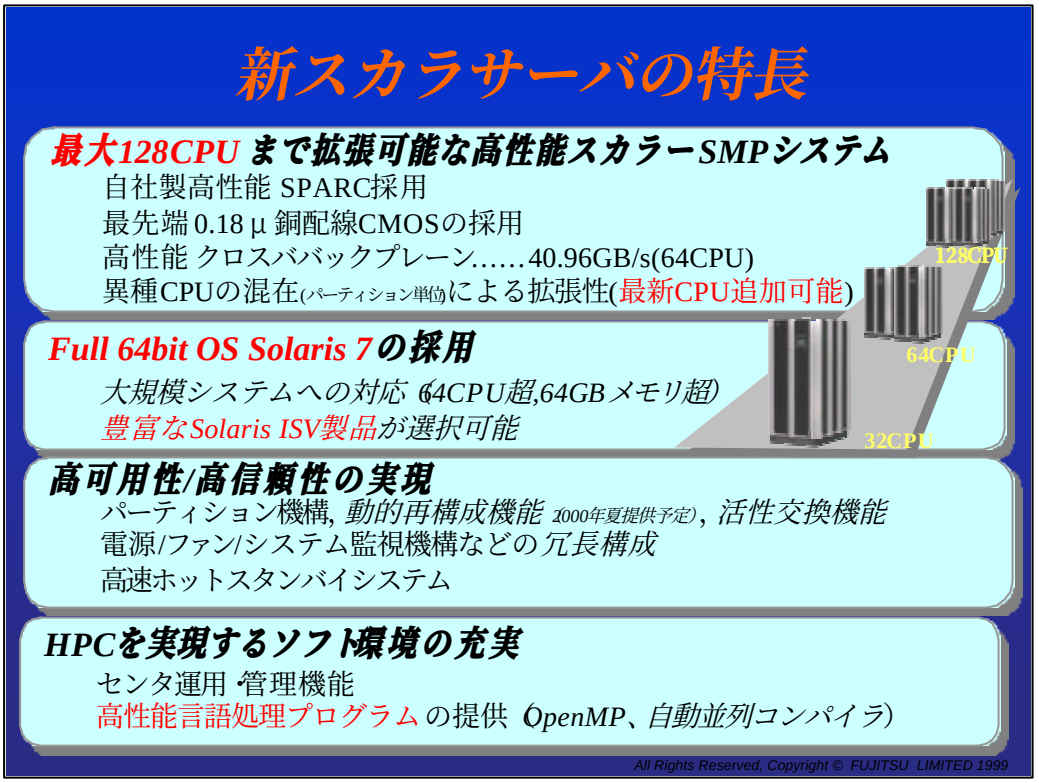
GP7000Fの製品ラインナップとモデル2000の位置づけを説明する。

GP7000Fは、2CPUから8CPUのミッドレンジサーバ、モデル2000及び10月に発表したモデル1000を追加し同一のアーキテクチャを採用したハイエンドサーバと幅広いラインナップで構成されている。

更に、2000年中には次期プロセッサである500MHzのSPARC64 GPを搭載する業界最大の128SMPサーバを加えて更に充実する予定である。

本説明では、ハイエンドサーバであるモデル2000を中心に説明する。

All Rights Reserved,Copyright ©）富士通株式会社 1999



新スカラサーバモデル2000の特長を説明する。

最大の特長は、最大128CPUまで拡張可能としたアーキテクチャで実現していることである。

キーテクノロジであるプロセッサには、自社開発のSPARCを採用し、自社の最先端0.18μ銅配線CMOS技術を採用している。また、異種CPUの混在を可能とし、システムの初期投資を保護すると共に柔軟な拡張性を提供している。

第二の特長は、64bit版Solaris7を採用し、大規模システムへ対応可能とすると共に豊富なSolaris ISV製品が利用可能なことである。

第三に、高可用性・高信頼性の実現である。

第四の特長は、HPCを実現するソフト環境が充実している点であり、これに関しては「新スカラサーバGP7000Fソフトウェアと利用環境について」の報告で説明する。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

ハードウェアの特長

- **クロスバスイッチ技術の採用**
  - すべてのデータパスを並列動作 Processor / IO / Memory
  - 1対1接続により他ポートの動作待ち時間を最小化
- **高周波数動作**
  - 先進テクノロジの採用・クロスバ構成により200MHz動作を実現
- **SMPスケーラビリティ**
  - 2-levelクロスバ構成により最大128プロセッサまでのスケーラビリティを実現
  - 均一なメモリアクセスレイテンシを実現
  - パーティション単位に異なる周波数のCPUが混在可能
  - 高性能メモリサブシステム(SDRAM採用、64wayインターリーブ)
- **高性能IOサブシステム**
  - 66MHz / 64bit PCIをサポート (PCI Rev 2.1 compatible)
- **システム制御サブシステム**
  - 各種コンポーネントの監視・制御

All Rights Reserved, Copyright © FUJITSU LIMITED 1999

ハードウエアの特長を説明する。

新規に開発したクロスバスイッチ技術を採用している事である。

このクロスバスイッチには、先進テクノロジを採用し、200MHzの高周波数で動作させている。

第二の特長は、SMPスケーラビリテイであり、2レベルのクロスバ構成で128プロセッサまでのスケーラビリテイを実現している。また、均一なメモリレイテンシ、64Wayインターリーブの高性能メモリサブシステムを実現し、高性能を実現した事である。

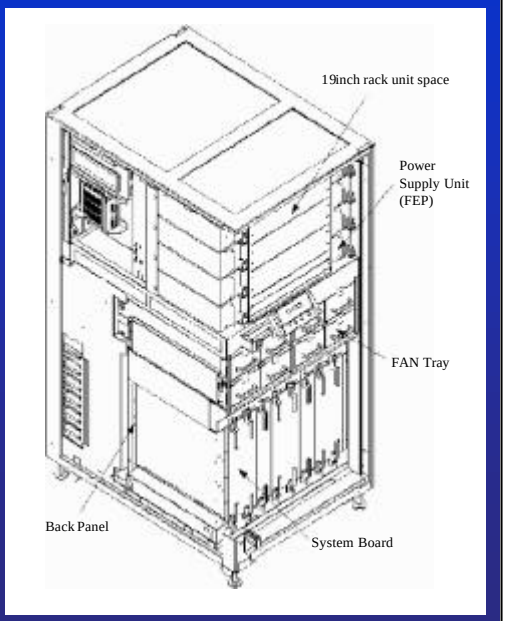
第三の特長は、64bitPCIバスをサポートし、高性能Oサブシステムを実現している。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

実装構造

- 筐体仕様（～32cpu）
  - システムボード× 8枚
  - ファントレイ
  - 電源ユニット
  - システム監視機構×2枚
  - バックパネル
  - クロスバボード
- 主要コンポーネントは活性交換可能
- 33～64cpu搭載時は筐体を密結合

2筐体接続時



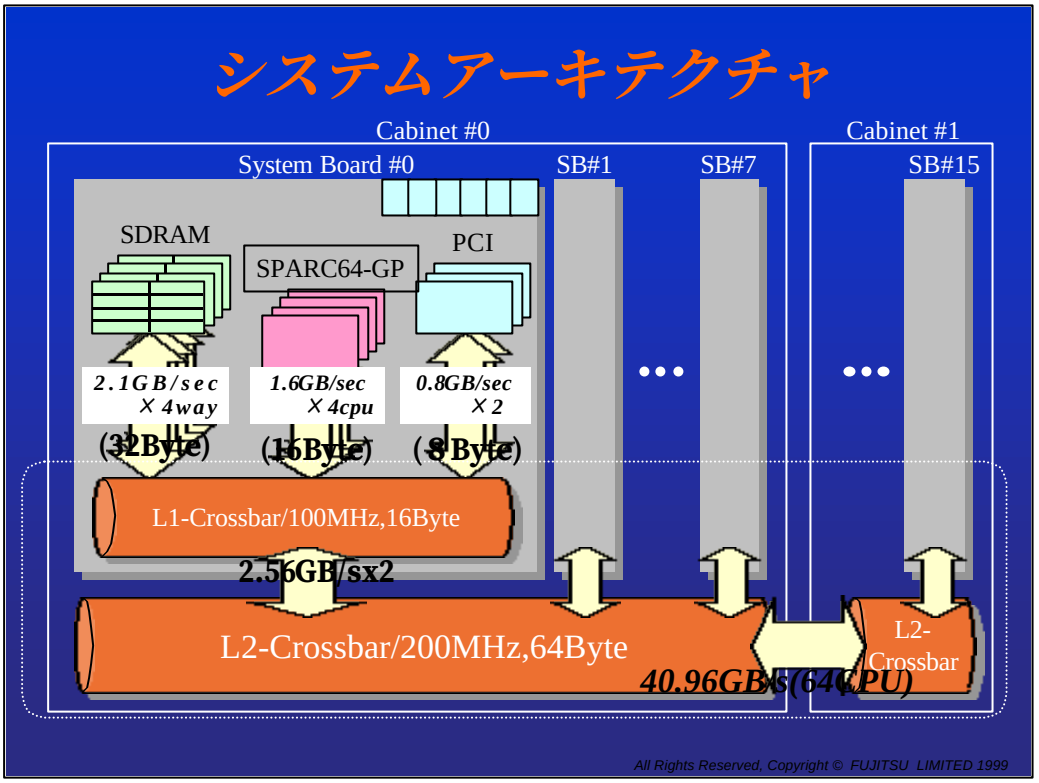
All Rights Reserved, Copyright © FUJITSU LIMITED 1999

モデル2000での実装構造を説明する。

1筐体は32CPUまで搭載され、64CPU構成では、この筐体を2個蜜結合して構成される。

筐体内は、システムボード、電源ユニット、クロスバボード等で構成され、主要なコンポーネントは、活性交換を可能となっている。

All Rights Reserved,Copyright（C）富士通株式会社 1999



システムアーキテクチャを説明する。ここでは、モデル2000の64CPU構成で説明している。

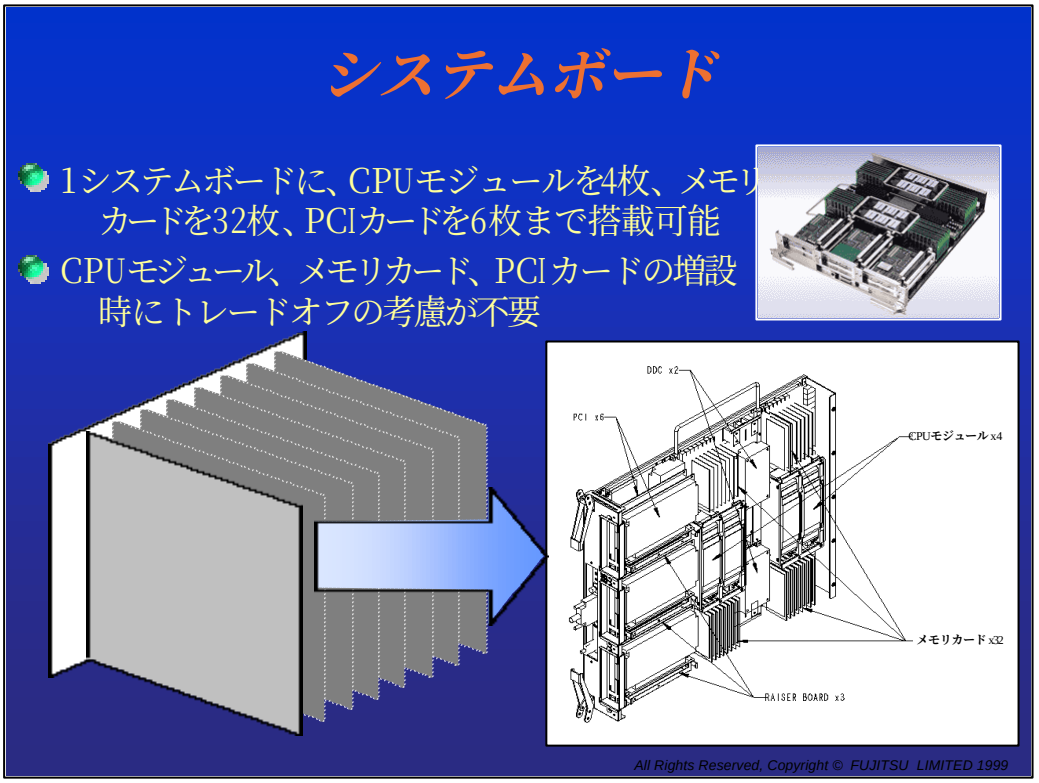
モデル2000では、最大64CPUのSMPのスケーラビリティを実現すること、及び32CPUを単位とした増設を可能とする為に2レベルのクロスバ構成を採用している。

第一のレベルのクロスバ(L1-crossbar)はシステムボード内部のクロスバで、CPU、メモリモジュール、PCIブリッジを相互接続している。このL1クロスバは、16Byte幅、100MHzで動作している。第二のレベルのクロスバ(L2-crossbar)は、最大16枚のシステムボードのL1クロスバを相互接続するクロスバである。

このL2クロスバは、64バイト幅、200Mhzで動作し、最大約41GB/秒のバスバンド幅を持っている。

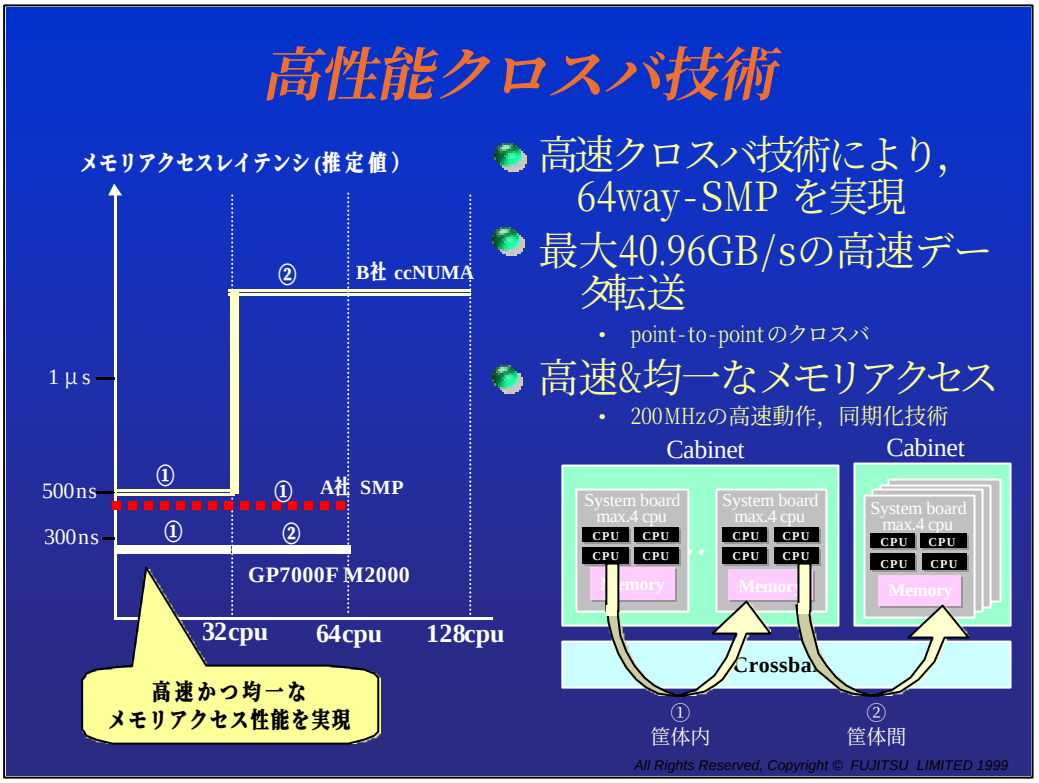
今回採用したクロスバシステムアーキにより、高性能化を実現している。

All Rights Reserved,Copyright (C) 富士通株式会社 1999



システムボードには、最大4CPUモジュール、32個のメモリスロット、6個のPCIスロットが搭載可能となっている。

All Rights Reserved,Copyright (C) 富士通株式会社 1999



今回開発した、クロスバでの性能を説明する。

SMPシステムでは如何に高速で且つ均一なメモリアクセスを実現するかが重要な要件であり、モデル2000ではこれら要件を実現している。

左のグラフで、他社のccNUMA及びSMPでのメモリレイテンシの比較を示している。

A社は、約500nsecで64SMPを実現している。

B社は、ccNUMAを採用しており、32CPU以上で極端にメモリテイテンシが悪くなっており、レスポンスの均一性が保証出来なくなっている。

一方、モデル2000では、300nsでの均一性を保証し、高速処理を可能としている。更に、このレイテンシで128CPUまで実現可能としている。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

SPARC64™ GP CPUモジュール

- SPARC V9完全準拠
- 高性能実現のための先進機能のサポート
  - 4命令同時発行
  - 3浮動小数点演算同時実行
  - register renaming
  - Out-of-order
  - 大容量一次キャッシュ
  - direct-mapped 8MB E-cache
  - Branch prediction精度の向上

- 高信頼性
  - **1次キャッシュData，2次キャッシュTag/DataをECCにより保護**
  - **1bitエラー時でも動作を継続**

All Rights Reserved, Copyright © FUJITSU LIMITED 1999

富士通が開発した「SPARC64 GP CPUモジュール」について説明する。

SPARC64 GPは、SPARC V9に完全準拠し,

4命令同時発行、浮動少数点演算同時実行が3個まで可能とし、Out-of Order実行等各種の先進機能を提供し高性能化を実現したSPARCプロセッサである。

また、大容量のキャッシュメモリを搭載可能とし高性能を実現すると共に、各レベルのキャッシュメモリには、ECCをサポートし、高信頼性を実現している。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

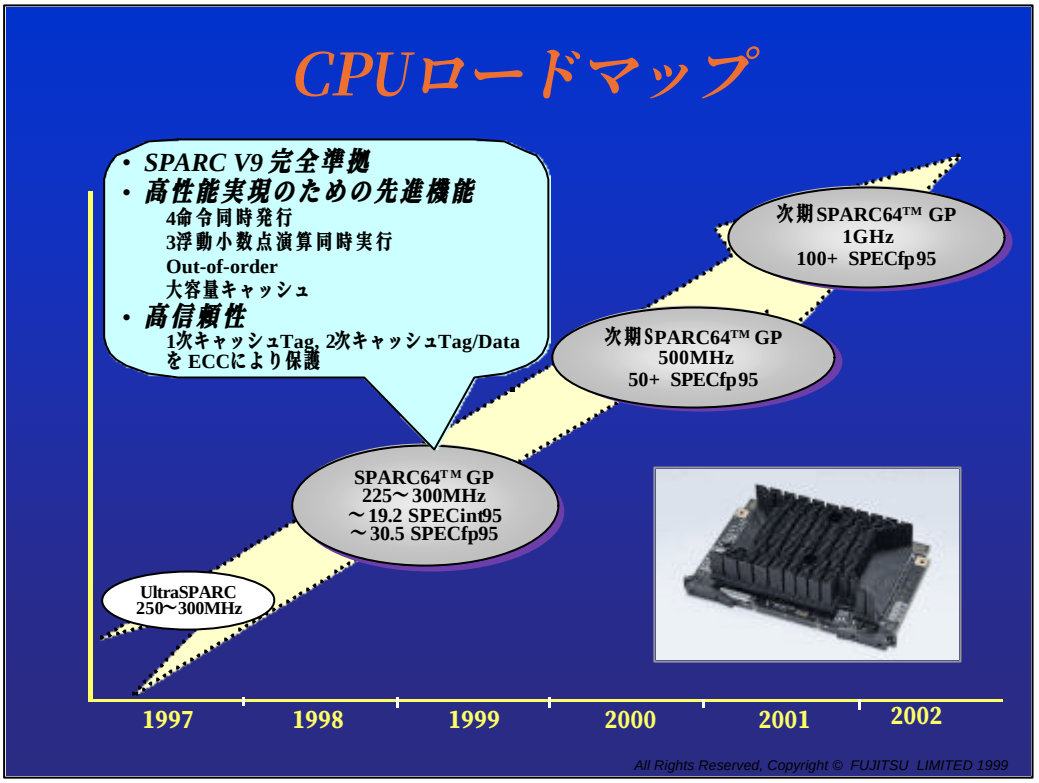
CPUモジュール比較

CPU	SPARC64 GP 300MHz	UltraSPARCII 400MHz	PA8500 440MHz	R12000 300MHz	21264A 700MHz	Power3 200MHz
搭載機種	FUJITSU GP7000F	SUN E10000	HP V2500	SGI Origin2000	COMPAQ GS140	IBM RS6000SP
アーキテクチャ	SMP	SMP	ccNUMA	ccNUMA	SMP	クラスタ
最大CPU数 (システム当り)	～128	～64	～128(32x4)	～512(4x256)	～14	～512
オンチップキャッシュ	I:64KB D:64KB	I:16KB D:16KB	I:0.5MB D: 1MB	I:32KB D:32KB	I:64KB D:64KB	I:32KB D:64KB
外部キャッシュ (命令+データ)	8MB/cpu	8MB/cpu	無し	8MB/cpu	8MB/cpu	4MB/cpu
同時実行 浮動少数点演算数	3	2	4	2	2	2
ロードストア ユニット数	2	1	2	1	2	2
Out-of-order実行	あり	無し	あり	あり	あり	あり
SPECint95	19.2	18.3	34.0	18.4	39.1	13.1
SPECfp95 (1CPU)	30.5	30.1	51.4	未公表	68.1	30.1

All Rights Reserved, Copyright © FUJITSU LIMITED 1999

各社の現時点での、スカラサーバに採用されているCPUモジュール仕様を紹介する。

All Rights Reserved,Copyright ©）富士通株式会社 1999

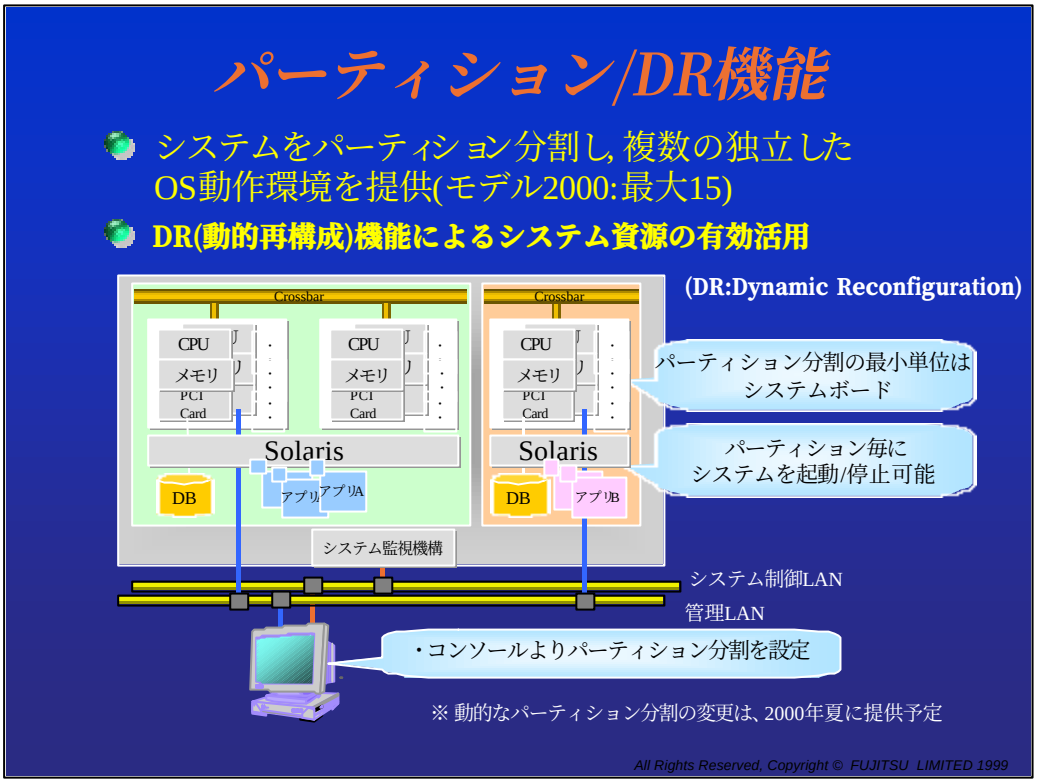


SPARC64 GPの今後の展開を説明する。

現時点は、300MHzの動作周波数で提供しており、動作周波数の向上へ向けて開発を進めており、2000年中には500Mhzの次期SPARC64を、2001年には各社とも1GHzクラスがターゲットとなっており、1GHz以上を実現する予定である。

性能的には、各々約2倍が目標で、さらなる高性能化へ向けて継続的な開発を実施している。

All Rights Reserved,Copyright ©）富士通株式会社 1999



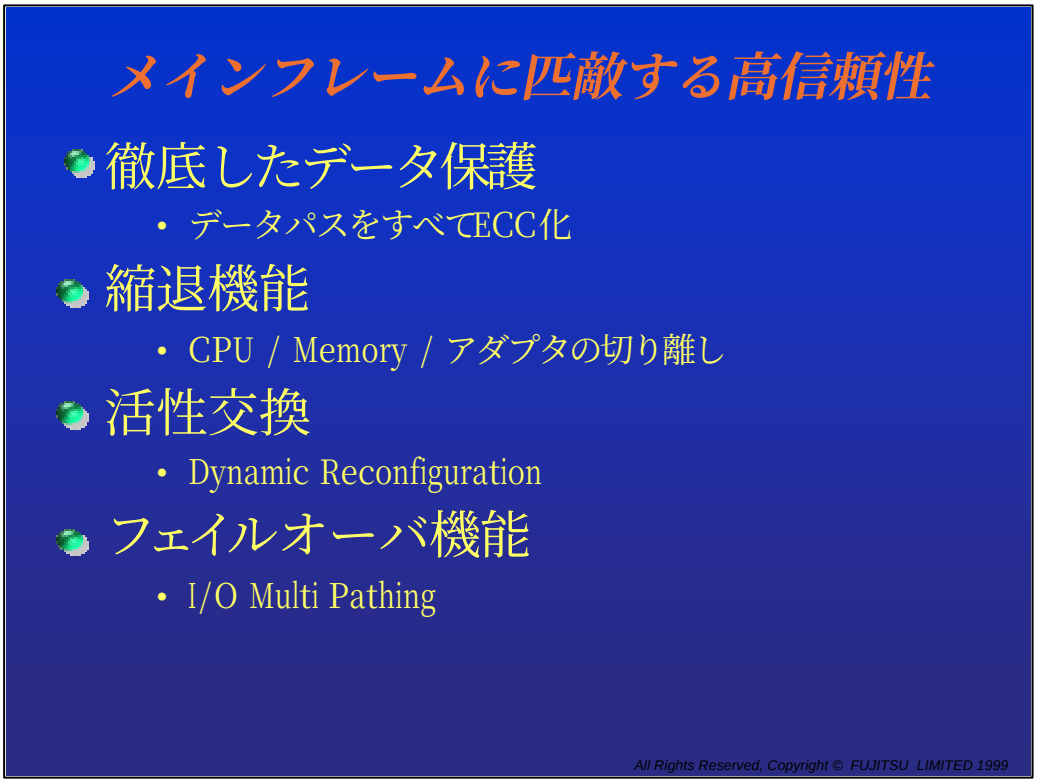
今回のモデル1000以上のモデルでは、パーティション機能及びDRと呼ぶ動的再構成機能

を提供する。

複数のシステムボードの組み合わせをパーティションと呼ぶ1つの論理的なハードウエアシステムとすることが出来、各パーティションは独立したSolaris環境で動作可能となっている。モデル2000では、最大15パーティションまで分割可能となっている。このパーティション機能を使用する利点は、サーバ統合による運用管理の一元化、障害発生での影響の局所化、運用環境に応じたシステム資源の有効活用等が上げられ、新しい利用環境が可能となる。

DR機能は、運用業務を停止する事無く、パーティションを変更したり、またハード故障により縮退したボードの動的交換等を実現している。

All Rights Reserved,Copyright (C) 富士通株式会社 1999



メインフレームに匹敵する高信頼性を実現している。
高信頼化を実現するに当り、徹底したデータ保護、主要構成コンポーネントの縮退機能、これら縮退したコンポーネントの活性交換をDyanmic Reconfiguration機能で、及びI/O Multi Pathingによるフェールオーバ機能の提供により実現している。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

徹底したデータ保護

Single point failureを完全救済

- データパス全てをECC化
- **CPU障害率を1/30に低減**

	CPU		メモリ	システムバス
	1次キャッシュメモリ	2次キャッシュメモリ		
富士通 GP7000F Model 2000	**ECC (128KB)**	**ECC (8MB)**	**ECC**	**ECC**
A社 最上位サーバ	なし (32KB)	パリティ (8MB)	ECC	ECC
B社 最上位サーバ	ECC (1.5MB)	―	ECC	パリティ

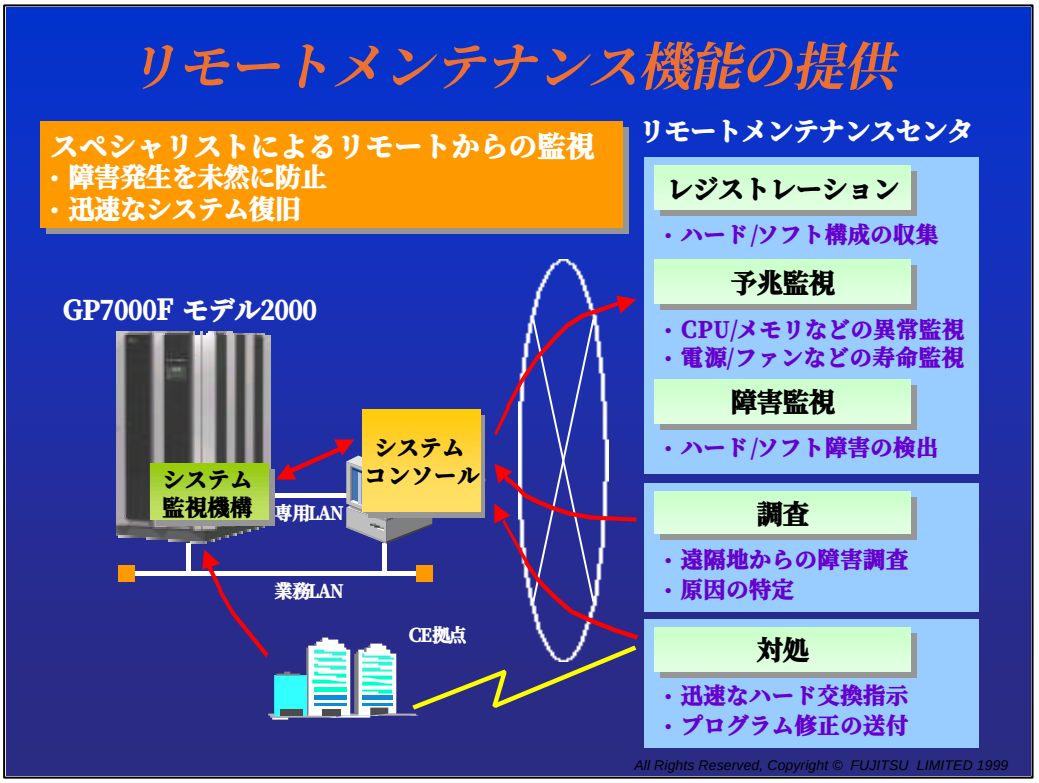
All Rights Reserved, Copyright © FUJITSU LIMITED 1999

高信頼性の実現については、徹底したデータ保護を例として説明する。

この表では、他社2社との比較を上げている。GP7000Fではメモリ、システムバスに加えCPU内の一次、二次キャッシュメモリにもECCを採用し、データパス全てをECCで保護している。

特に、キャッシュメモリの増大により、メモリのソフトエラー率が高くなりシステム障害となることがある。ECCのサポートにより、CPU障害率は1/30に低減可能となり、大幅に信頼性を向上させている。

All Rights Reserved,Copyright (C) 富士通株式会社 1999



リモートメンテナンス機能を提供する。

GP7000Fモデル2000に内蔵するシステム監視機構の情報をシステムコンソールを介して、リモートメンテナンスセンタからスペシャリストにより監視を行う。遠隔地からのハード異常・寿命の監視により障害発生を未然に防止する。また、障害監視を行い、障害の調査・原因の特定により迅速なシステムの復旧を実現する。

本機能の提供により、保守性の向上を図っている。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

新スカラサーバの仕様

		モデル1000	モデル2000	128CPUモデル
CPU	プロセッサ	SPARC64™ GP		
	周波数	300MHz/500MHz *1		500MHz *1
	プロセッサ数	～ 32	～ 64	～ 128
キャッシュメモリ容量 (プロセッサあたり)		1次:命令/データ各64KB 2次:命令+データ:8MB		1次:命令/データ各128KB 2次:命令+データ:8MB
メモリ容量		～32GB	～64GB	～128GB
システムバス	帯域幅	最大20.48GB/sec	最大40.96GB/sec	最大51.2GB/sec
	遅延時間	300ns以下		
ディスク容量		内蔵基本 :18.2GB, 内蔵最大 :1TB, システム最大700TB		
PCIスロット		6～48	12～96	～192
信頼性	冗長構成	システムボード, システム監視機構, 電源ユニット, ファントレイ, ディスク装置（ミラー構成), 電源系統（二系統受電機構）		
	活性交換	システムボード*2, システム監視機構, 電源ユニット, ファントレイ, ディスク装置		
外形寸法 (W×D×H)		766x1161x1800(16CPU) 1470x1161x1800(32CPU)	1044x1300x1800(32CPU) 2172x1300x1800(64CPU)	2172x3338x1800 (128CPU)
消費電力		5.7KW(16CPU)	9KW(32CPU)	約36KW(128CPU)

*1 :2000年中予定, *2 :システムボードの活性交換は2000年夏に対応予定

All Rights Reserved, Copyright © FUJITSU LIMITED 1999

これまでに説明した、モデル2000及びモデル1000、来年中を予定している128CPUモデルの仕様を示す。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

GP7000FでのI/O関連装置を説明する。

ファイル関連のI/O装置では、GR720等のデイスクアレイ装置及びLT300等のテープライブラリ装置をサポートしている。

ネットワーク関連では、GigabitEthernet,ATM等高速ネットワークへの接続が可能となっている。

また、AP-Netのインターコネクトを介してクラスタ構成を可能としている。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

I/O装置

	富士通製品	IHV製品
	品質保証,サポート	動作確認
デイスクアレイ装置	GR720/700 GP-D16/35/40	
テープライブラリ装置	LT300 L3500 L1800 GP-DLL351	PetaSite(Sony) PowderHorn* (STORAGETECK) TimberLine* (STORAGETECK)
その他I/O装置	プリンタ装置 (VSPシリーズ) UPS装置 ・ ・	CD-ROMライブラリ (日本ビクター:MC2100/2200等) DTFテープ装置 (Sony:GY2120/DMS-B9) プリンタ装置 (富士ゼロックス, セイコーインスツルメント等)

(IHV:Independent Hardware Vender)　*)予定製品

All Rights Reserved, Copyright © FUJITSU LIMITED 1999

GP7000Fで接続するIO装置は、富士通が品質保証し、サポートも行う富士通製品に加え各種IHV製品を接続し利用可能とする環境を提供している。

IHV製品は、IHVによる動作確認を行い接続を可能としており、現状Sony製のPetasite等が確認済となっている。

今後サポートI/Oの拡大を図って行く予定である。

All Rights Reserved,Copyright (C) 富士通株式会社 1999

10月に発表した、新デイスクアレイ装置GR720を説明する。

GR720は、Storplexの概念に基づき開発したハードRAID装置である。本装置は、SAN環境に対応した最初の製品で、ファイバーチャネルをサポートし高速に接続する事を可能としいる。大容量、高速で、高可用性にも配慮された点が特長である。

また、業務停止を不要とした瞬時バックアップが可能となっている。具体的には 新スカラサーバGP7000Fソフトウェアと利用環境について」の報告で説明する SystemWalker/StorageMGRにより可能となる。

All Rights Reserved,Copyright ©）富士通株式会社 1999

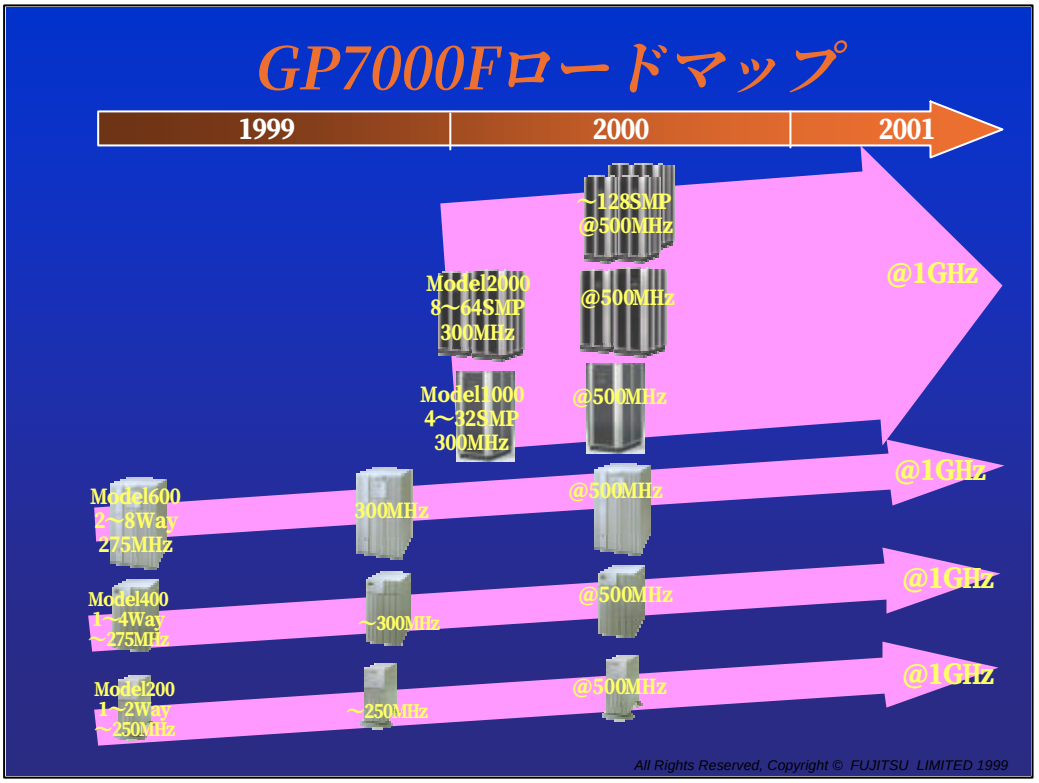
GR720（仕様）

		GR720 (ﾗｯｸﾓﾃﾞﾙ)	GR720 (ﾀﾜｰﾓﾃﾞﾙ)	GR700 (ﾗｯｸﾓﾃﾞﾙ)
記憶容量(RAID5構成時)		最大2.3TB	最大0.57TB	最大1.44TB
ﾃﾞｨｽｸﾄﾞﾗｲﾌﾞ	容量/回転数	9GB/10000rpm 18GB/10000rpm 36GB/10000rpm		9GB/10000rpm 18GB/10000rpm
	ﾄﾞﾗｲﾌﾞｲﾝﾀﾌｪｰｽ	FC		UltraSCSI
ｺﾝﾄﾛｰﾗ構成	キャッシュ容量	最大2GB	最大512MB	最大512MB
ﾎｽﾄｲﾝﾀﾌｪｰｽ		UltraSCSI Ultra2 SCSI FC-AL FC-Fabric		UltraSCSI FC-AL FC-Fabric
ﾎｽﾄﾎﾟｰﾄ数	SCSI	最大4ﾎﾟｰﾄ		最大4ﾎﾟｰﾄ
	FC	最大4ﾎﾟｰﾄ		最大4ﾎﾟｰﾄ
ｻﾎﾟｰﾄｻｰﾊﾞ		GP7000F/S GP5000 HP,RS/6000,NCR 各社NTｻｰﾊﾞ		GP7000F/S GP5000

All Rights Reserved, Copyright © FUJITSU LIMITED 1999

この表は、GR720の仕様を説明している。

All Rights Reserved,Copyright（C）富士通株式会社 1999



最後に、今後のGP7000Fの展開を説明する。
先程説明した、CPUのロードマップをベースにプラットフォームも展開する。
2000年には500MHzの次期CPUを搭載して展開する。
2001年には1GHz次次期CPUを搭載して展開する。
以上で、新スカラサーバのハードウエアの説明を終わります。

All Rights Reserved,Copyright ©）富士通株式会社 1999