マルチベンダー環境の運用事例

北陸先端科学技術大学院大学
情報科学センター　松澤　照男

2005年10月26日

アウトライン

- 情報環境の概要
- 計算サーバの変遷
- 計算サーバの導入の基本的な考え方と経緯
- 代表的な計算サーバ
  - Cray XT3
  - SGI Altix3700　他
- 運用の基本的な考え方と稼動統計
- 姫野ベンチマーク
- おわりに

JAIST情報環境の目的

- 本学構成員（教職員や学生など）が世界最高水準の研究を組織的に推進するために必要となる、高度かつ先端的な情報環境を提供する。

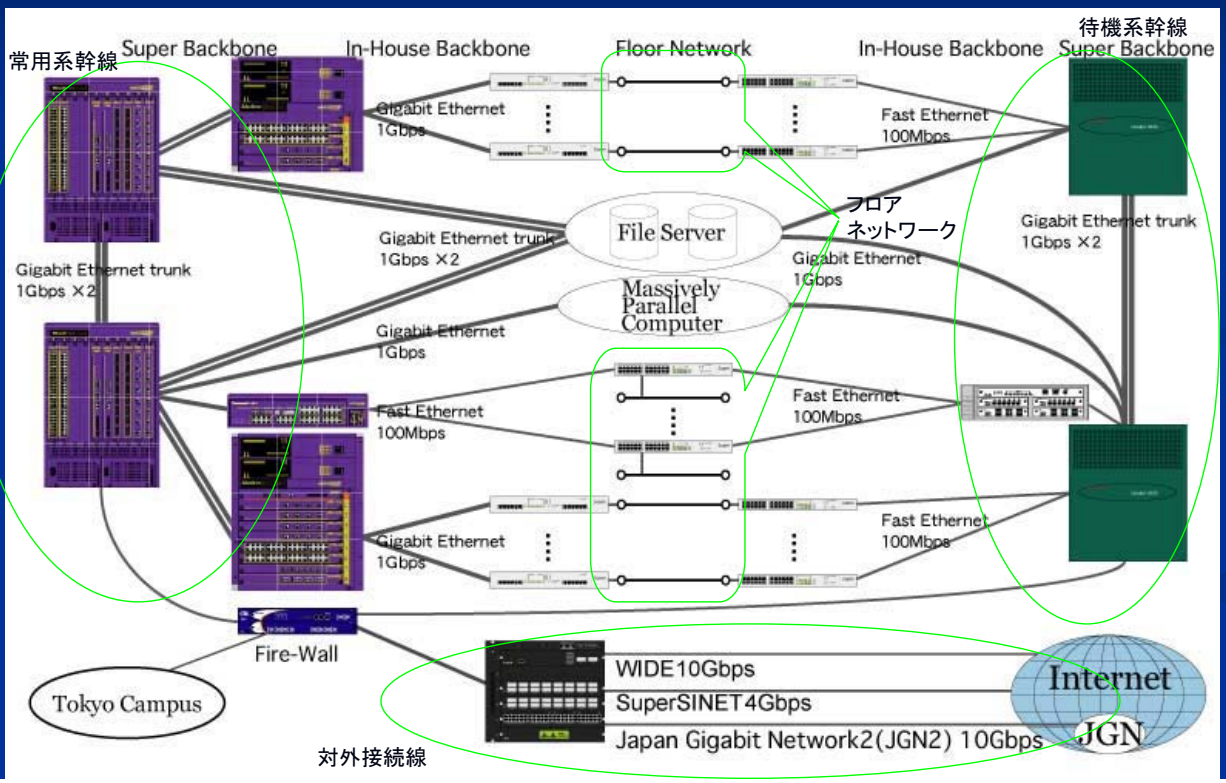
JAIST情報環境の特徴

- 日本におけるインターネットの重要な研究拠点
- 教職員、学生 各自一台（材料の学生はおよそ3人に一台）のワークステーション 24時間利用可能
- 1500台以上のワークステーションの超高速ローカルエリアネットワークによる結合
- 超並列計算機群へのアクセス
- 各種サーバによる、データや情報の集中管理
- コラボレーションルームにおける遠隔会議や遠隔教育の実施
- インターネット、ギガビットネットワーク への接続
- 無線ネットワーク等様々な先端的ネットワーク

本学のネットワーク

- Frontnet
  - 幹線、対外接続線、無線ネットワーク、フロアネットワーク（各研究室）から構成
  - 高速性、高可用性、利便性を満たすための機構
  - 情報科学センターによる集中管理

Frontnet（全体図）

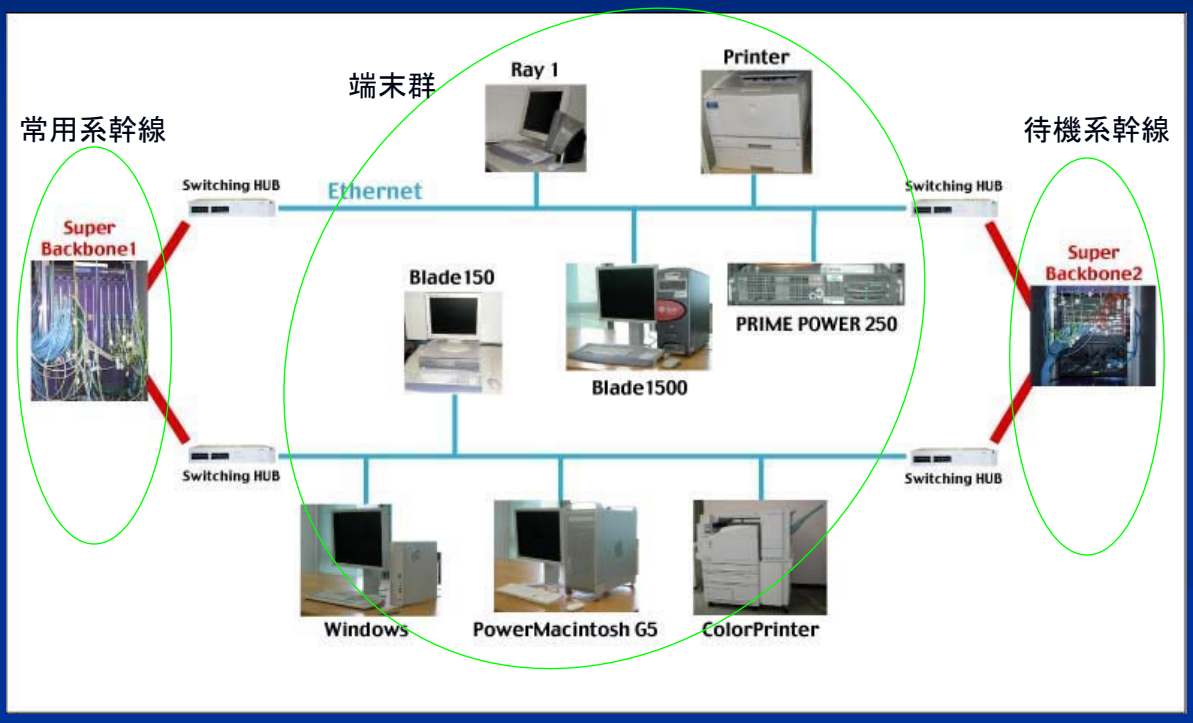


高速なネットワーク

- 幹線
  - 2Gbpsイーサネットによる高速光ファイバ接続
  - 常用系・待機系の2系統
  - ファイルサーバへ高速アクセス

  本年度中に10GbE化の予定
- 対外接続線
  - WIDE
  - SuperSINET
  - JGN2
- 無線ネットワーク
  - ノートパソコン等の機動力を活用

Floor Network



フロアネットワーク

- 100Mbpsイーサネットによってユーザの端末を接続
- 各フロアネットワークは常用系・待機系両方に接続され、トラブル時には自動で待機系の利用を開始

本年度中に1GbE化の予定

JAIST情報環境の特徴
超並列計算機群

- 情報科学での高効率な超並列処理研究
- 材料設計やゲノム情報処理、並列処理などに現在、活用されている
- シミュレーションオリエンテッド・エンジニアリングに対応するコンピューティング環境の整備が計画されつつある
  - ナノテクノロジー
  - ゲノムデータベースを対象とした検索、知識発見
  - 第1原理分子動力学シミュレーション
  - 金属の薄膜成長および相変態のシミュレーション

アウトライン

- 情報環境の概要
- 計算サーバの変遷
- 計算サーバの導入の基本的な考え方と経緯
- 代表的な計算サーバ
  - Cray XT3
  - SGI Altix3700　他
- 運用の基本的考え方と稼動統計
- 姫野ベンチマーク
- おわりに

超並列計算機

Cray
XT3

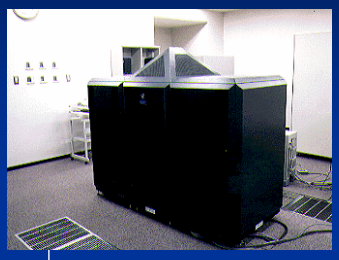
Cray
T3E-1200

Cray
T3E

Thinking Machines
CM-5

1992　1996　2000　2004　2008　2012

高度データベース処理計算機



SGI

Altix3700

IBM

RS/6000 SP

nCUBE

nCUBE-3

nCUBE

nCUBE2

1992 1996 2000 2004 2008 2012

超並列ソフトウェア研究用システム / PCクラスタシステム

APPRO

HyperBlade

Best Systems

HPC2000

Sun

SunFireV890

Sun

SunFire 15K

Fujitsu

GP7000S

Parsytec

GC/MPC-128

1992 1996 2000 2004 2008 2012

小規模計算サーバシステム

アウトライン

- 情報環境の概要
- 計算サーバの変遷
- 計算サーバの導入の基本的な考え方と経緯
- 代表的な計算サーバ
  - Cray XT3
  - SGI Altix3700　他
- 運用の基本的な考え方と稼動統計
- 姫野ベンチマーク
- おわりに

導入の基本的な考え方

- 導入のタイミングで最新の機器を選定（多少のリスクが伴う）。
- 教育的効果からなるべくアーキテクチャーの異なる計算機を導入する。

  ⇒　ユーザが最適な計算機が選択する。
- 情報環境の一部として導入　⇒　最終責任の所在を明確にする。
- 更新に際には、前計算機の資産にとらわれない。

主たる計算サーバの導入の経緯

- JAISTの超並列システム
  - 超並列処理研究システム
    - 超並列アルゴリズムの研究用
  - 高度データベース処理研究システム
    - 幅広い研究のプラットホームとして利用

超並列処理研究システム

- 超並列処理研究システム
  - 超並列アルゴリズムの開発・検証用
  - 並列化を陽に用いて，高い性能を引き出すことが目的
  - ユーザーは，基本的にMPI, SHMEMなどが使えることを仮定
  - 流体力学，分子動力学など，数値シミュレーション系の研究が主
- 要求要件
  - 分散メモリマシンであること
    - アルゴリズムの挙動が分かりやすいこと
      - 共有メモリシステムだと，分析が難しい
  - MPIを前提としたシステムであること.
  - 浮動小数点演算性能に優れること

→ **Cray XT3**

高度データベース処理研究システム

- 幅広い研究プラットホームとして導入
- 必ずしも並列処理が目的ではない.
  - 処理の高速化するための手段として，並列処理を用いる
- ユーザーは，並列処理の専門家とは限らない
- データベース処理を始め，画像処理や暗号処理などに利用

- 要求要件
  - 複雑なプログラムをせずに，大きなメモリ空間が利用できること.
    - 現代では，計算能力はCPU性能ではなく，メモリ容量で規定される
    - SMPまたはcc-NUMAだと有難い
  - MPI以外でも，簡易な並列化が可能なこと
    - 性能は出ないかもしれないが…
  - 整数演算性能が優れていること

→ **SGI Altix3700**

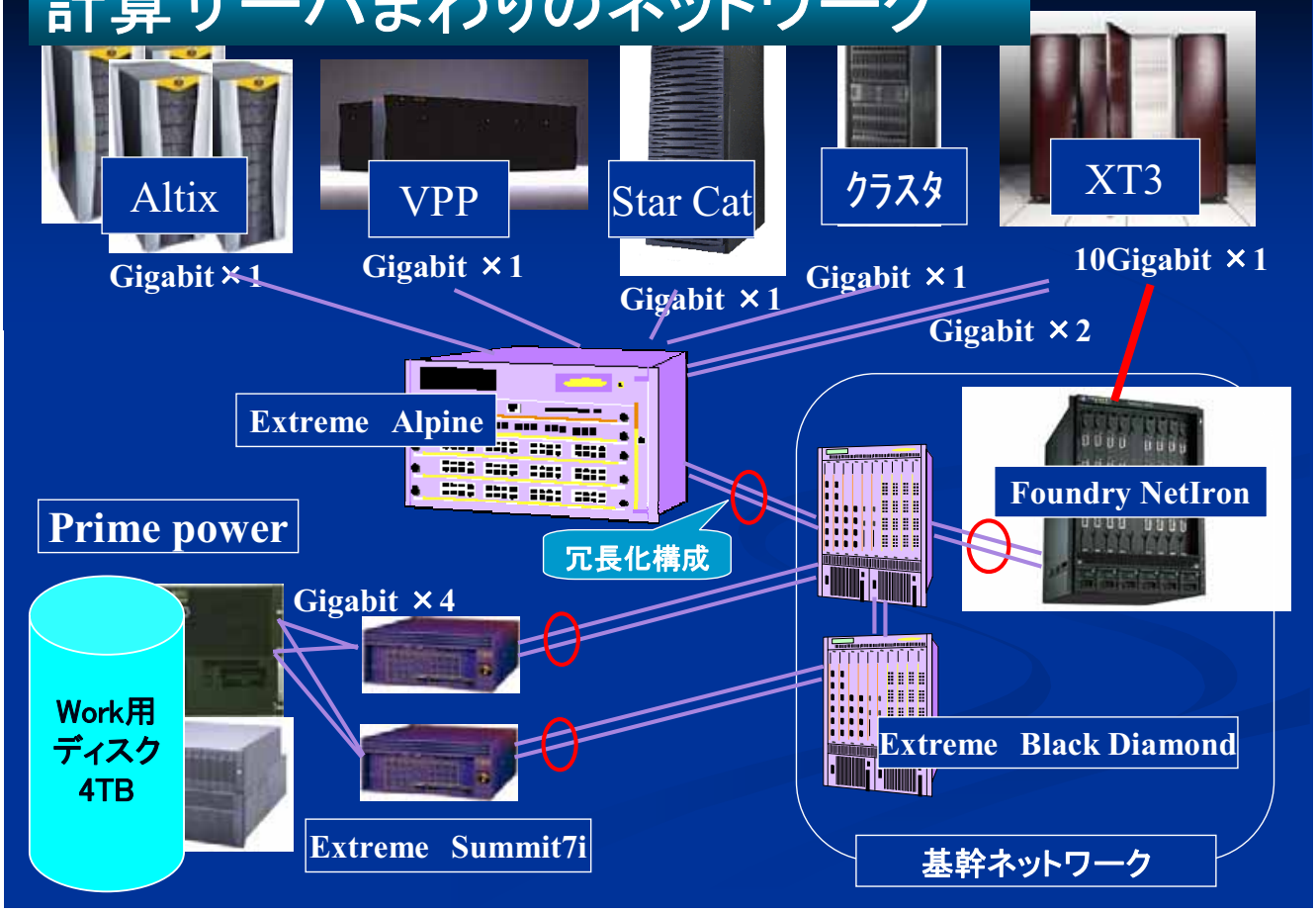
アウトライン

- 情報環境の概要
- 計算サーバの変遷
- 計算サーバの導入の基本的な考え方と経緯
- 代表的な計算サーバ
  - Cray XT3
  - SGI Altix3700　他
- 運用の基本的な考え方と稼動統計
- 姫野ベンチマーク
- おわりに

現在稼動中の主な計算サーバ

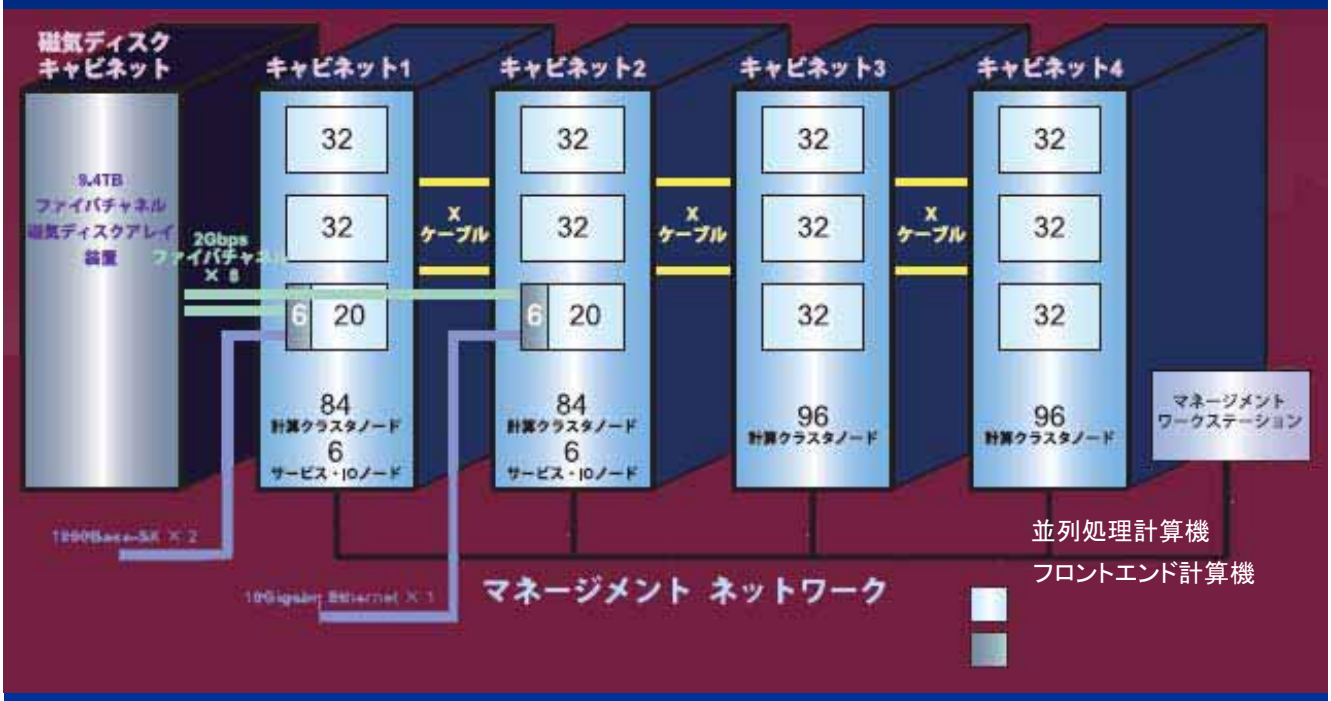
- Massively Parallel Computers
  - **SGI Altix3700**（ 128 cpus）
  - **CRAY　XT3**　（ 360 cpus）
  - **SUN Fire15K**（ 32 cpus）
- Computing Servers
  - **Fujitsu VPP5000**（ 2 cpus）
- PC cluster
  - **BestSystems HPC2000**（ 32cpus）

計算サーバまわりのネットワーク



Cray XT3

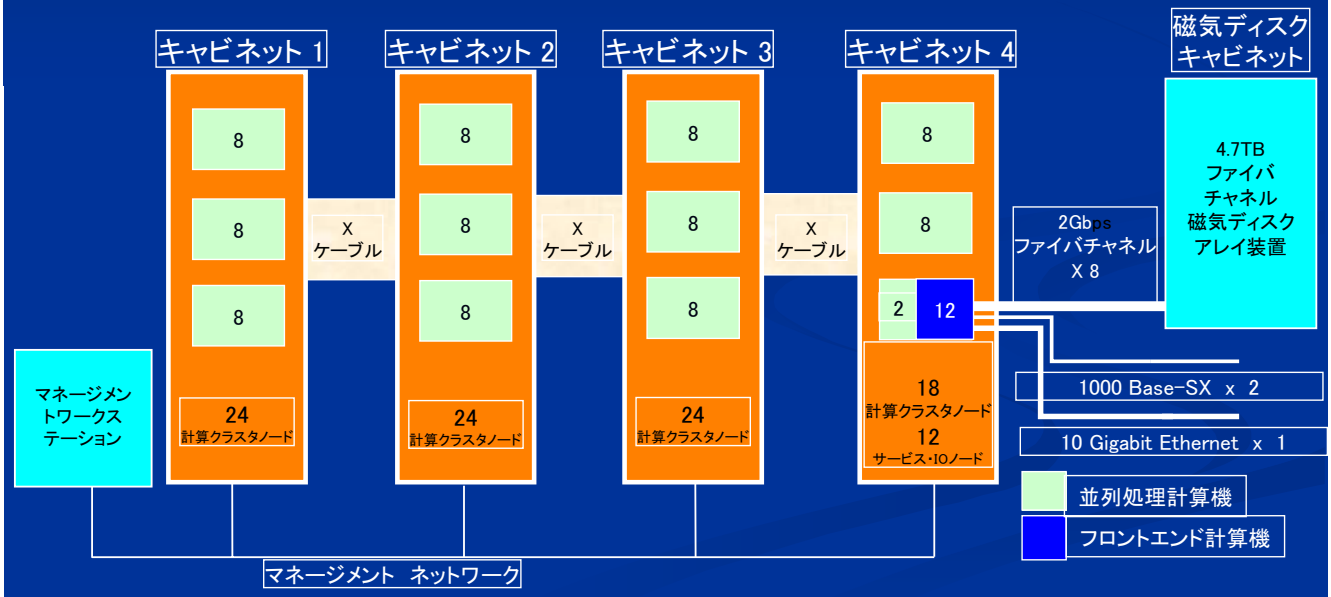
Cray XT3 システム構成



Cray XT3 システム構成

- 並列処理計算機
  - 90 計算クラスタノード（360CPU）
  - 4 CPU/計算クラスタノード
  - 2.88TB 主記憶容量（32GB/計算クラスタノード、8GB/CPU）
- フロントエンド計算機
  - 12 サービス・IOノード
  - 96GB　主記憶容量（8GB/ノード）
- 二次記憶装置
  - 4.7TB RAID 5 ファイバチャネルディスクアレイ装置
- システム相互結合ネットワーク
  - 4 x 12 x 8 3Dトーラスネットワークトポロジー
  - 491.52GB/sバイセクションバンド幅

Cray XT3 システム構成



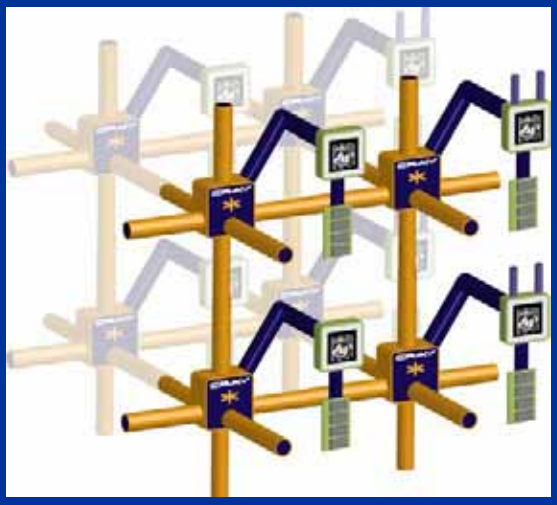
Cray XT3 プロセッサノード

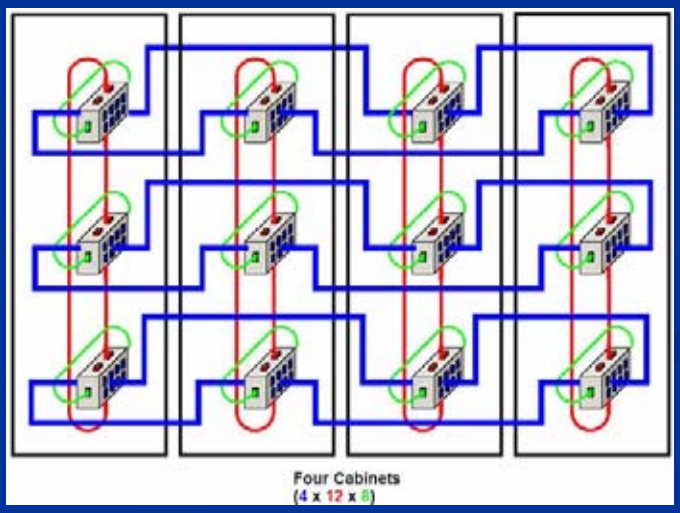
- 計算プロセッサ
  - 64ビット高性能プロセッサ（AMD社Opteron、4.8 Gflops）

    プライマリキャッシュ（L1）＝ 64KB （命令用）
    プライマリキャッシュ（L1）＝ 64KB （データ用）
    セカンダリキャッシュ（L2）＝ 1MB
- 内部ネットワーク・通信ハードウエア
  - 専用の高速ルータ（Cray SeaStar）
  - 高帯域（6方向計45.6 GB/秒）

Cray XT3 内部ネットワーク

- X,Y,Z全ての方向に環状接続を成す（3次元トーラス）
- どのプロセッサノード間でも最も効率の良い経路で
  データ通信を実現





SGI　Altix 3700

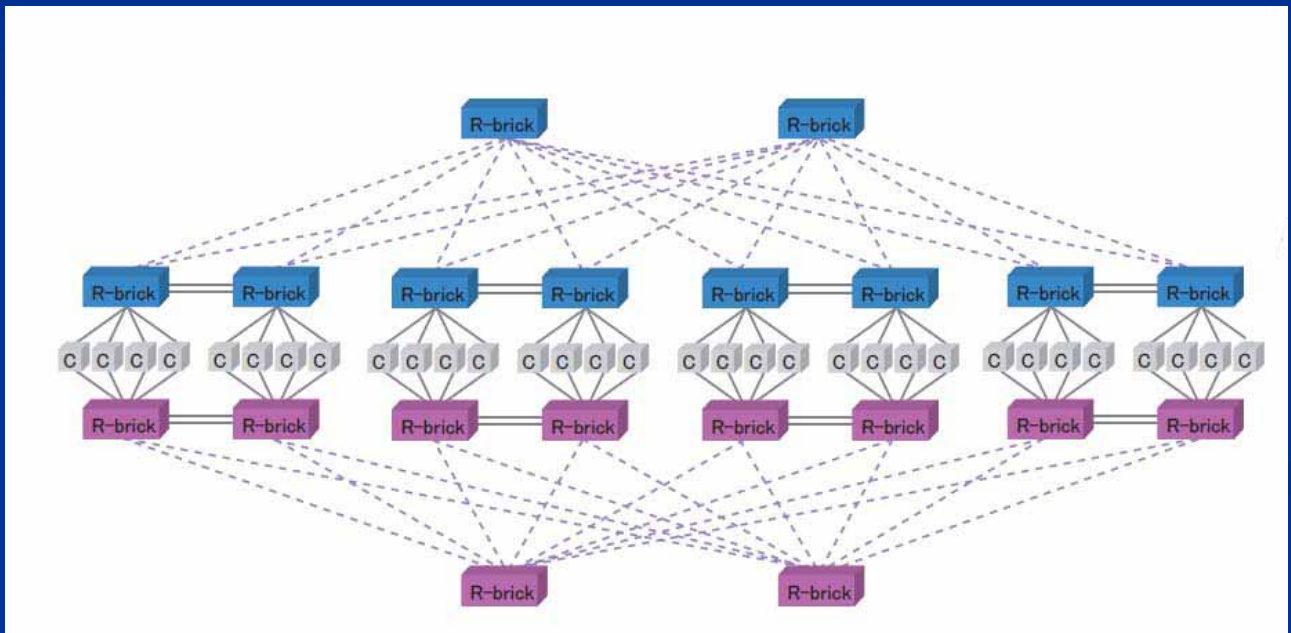
SGI Altix3700 の特徴

- Cーブリック32台をNUMAlink3(3.2GB/秒)で結合させた共有メモリ型の並列計算機
- Cーブリック（4個の64ビット Intel(R) Itanium(R) 2プロセッサ 1.3GHz、24GBのメモリ

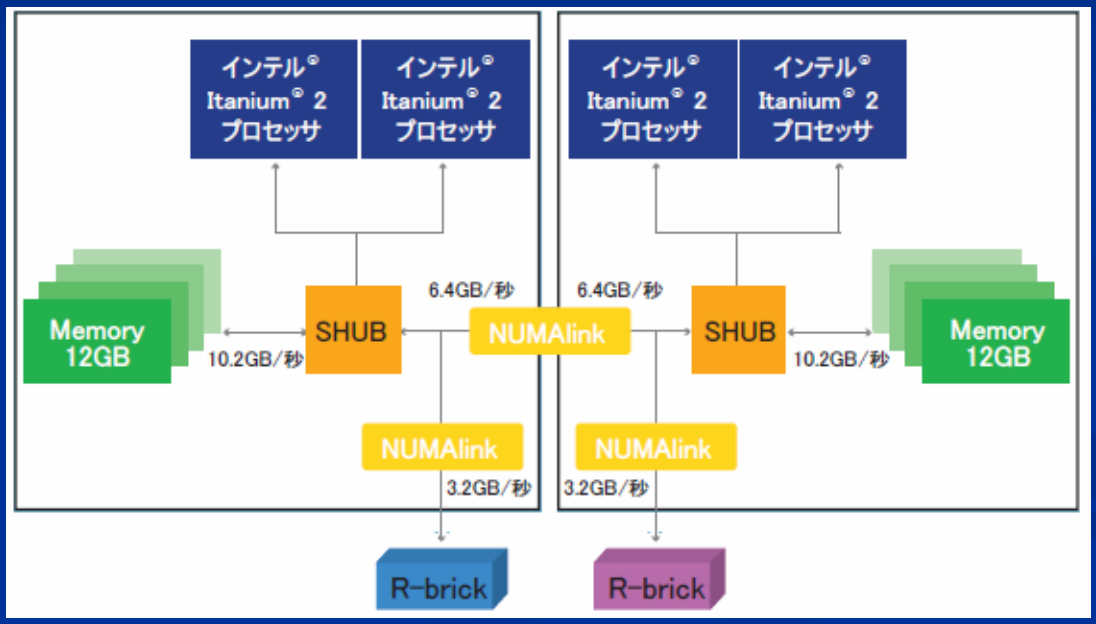
  プライマリキャッシュ（L1）＝　32KB　（レイテンシ1クロック）
  セカンダリキャッシュ（L2）＝ 256KB　（レイテンシ5クロック）
- システム合計128個のCPUと768GBのメモリ
- 単一のLinux オペレーティング・システム

  64Bit Linux（SGI Linux Environment with SGI ProPack)
- **36GB x 4 = 144 GB の内臓ディスク　⇒　OS用**

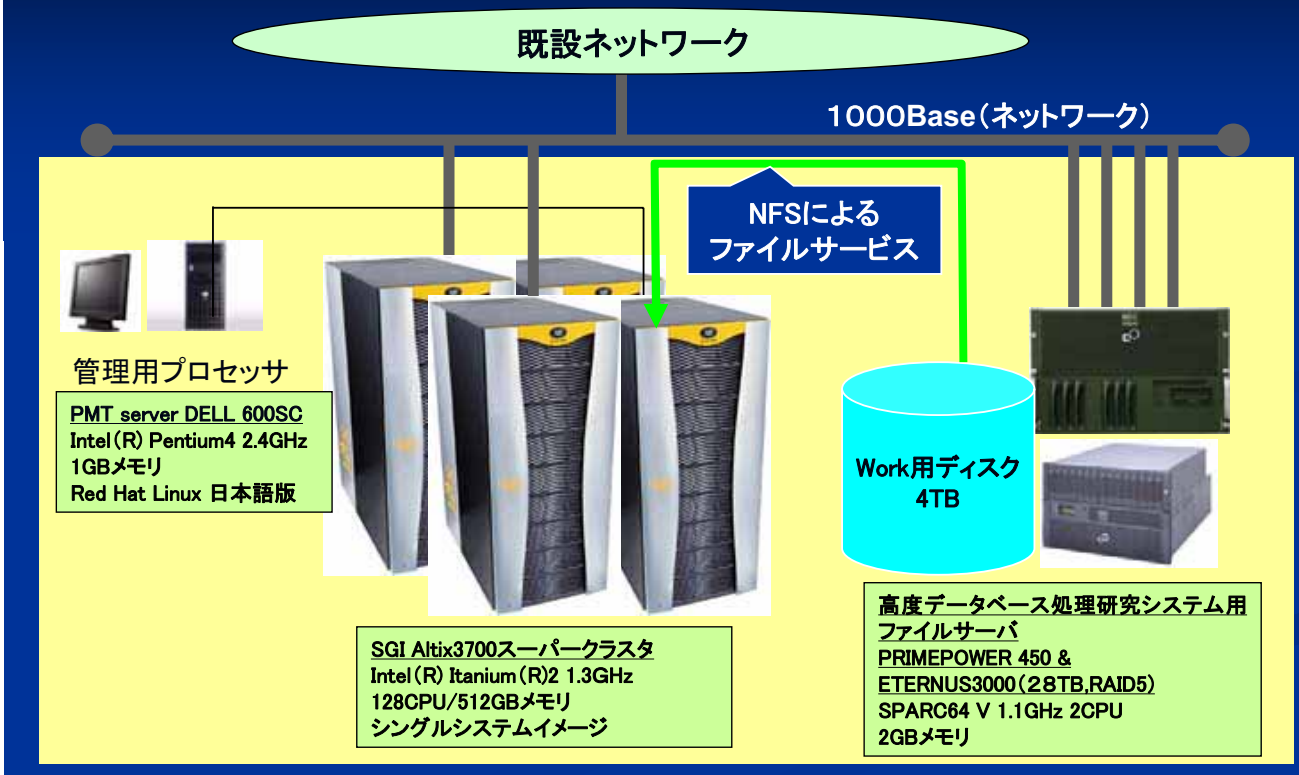
SGI Altix3700トポロジー



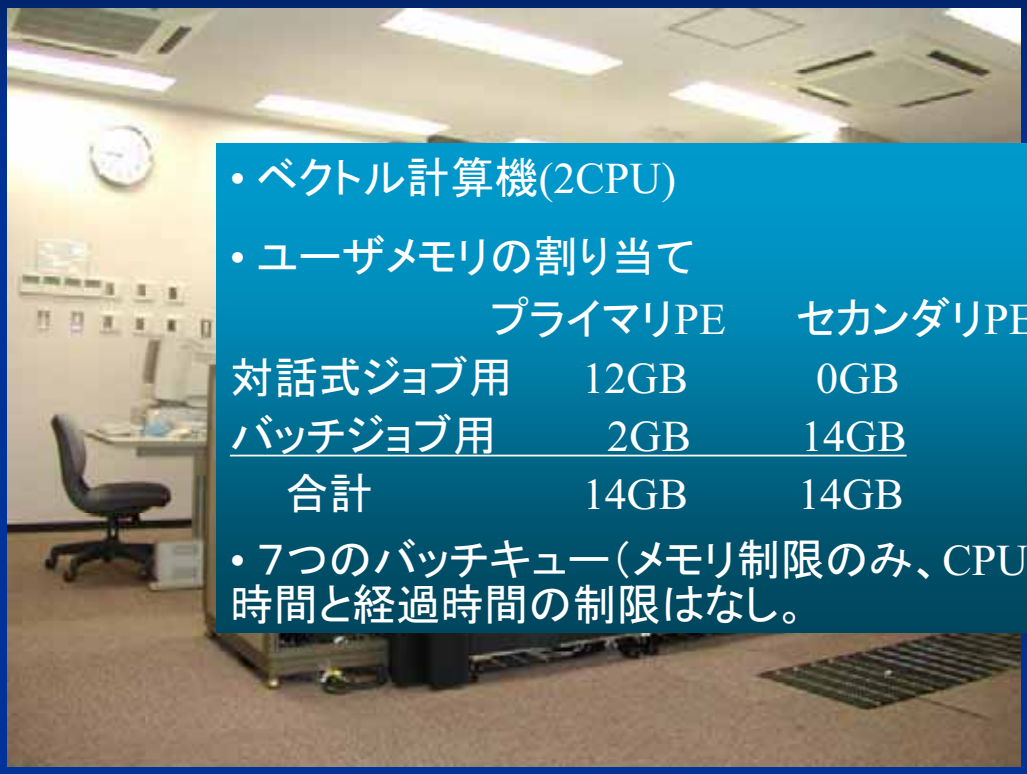
SGI Altix3700 Cブリック構成



SGI Altix3700 構成



計算サーバ(Fujitsu VPP5000)



- ベクトル計算機(2CPU)
- ユーザメモリの割り当て

	プライマリPE	セカンダリPE
対話式ジョブ用	12GB	0GB
バッチジョブ用	2GB	14GB
合計	14GB	14GB

- 7つのバッチキュー(メモリ制限のみ、CPU時間と経過時間の制限はなし。

計算サーバ(SunFire 15K)

- メモリー共有型計算機
- Ultra5バイナリ互換な超並列システム
- UltraSparc III Cu(900MHz)×32CPU, 32GB memory

PCクラスター(Best Systems HPC2000)

- 32CPU
- 各CPU
  Pentium3 1GHz
  512MB memory
  40GB disk,
- Myrinet2000 と Gigabit Ethernet

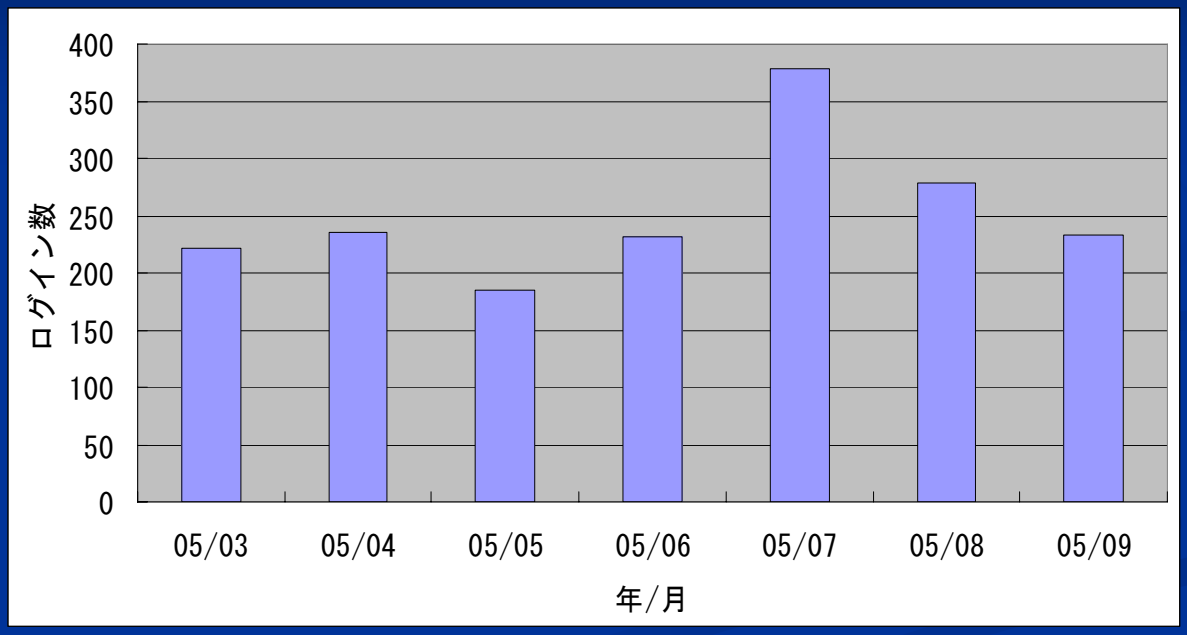
アウトライン

- 情報環境の概要
- 計算サーバの変遷
- 計算サーバの導入の基本的な考え方と経緯
- 代表的な計算サーバ
  - Cray XT3
  - SGI Altix3700
- 運用の基本的考え方および稼動統計
- 姫野ベンチマーク
- おわりに

運用の基本的な考え方

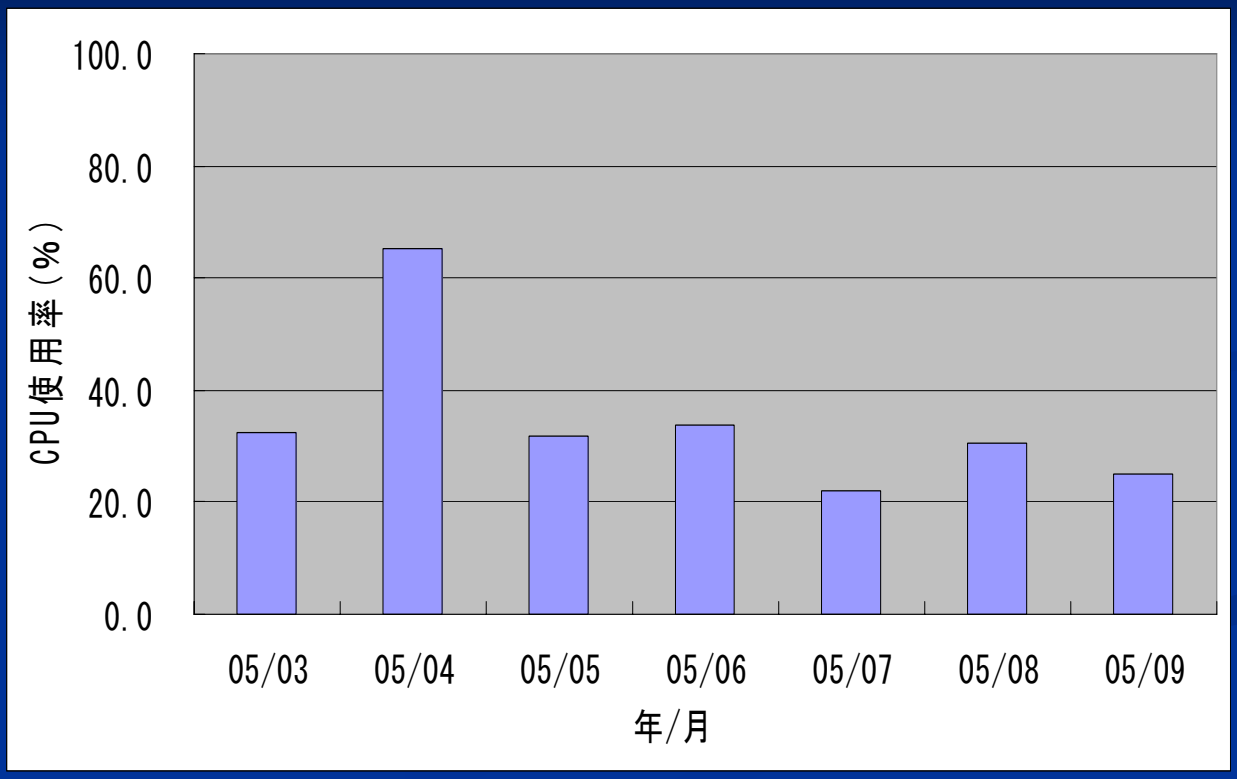
- センターはユーザ利用に関してなるべく干渉しない
- ユーザグループの育成
  - メーリングリストで情報交換
  - 利用法についての質問 ⇒ ユーザで解決
  - 処理時間などの計測で占有するときは他ユーザの了承を得る
  - ディーラーへの質問はセンター経由で行う
- 障害はセンターで対応

SGI Altix3700 の月別ログイン数

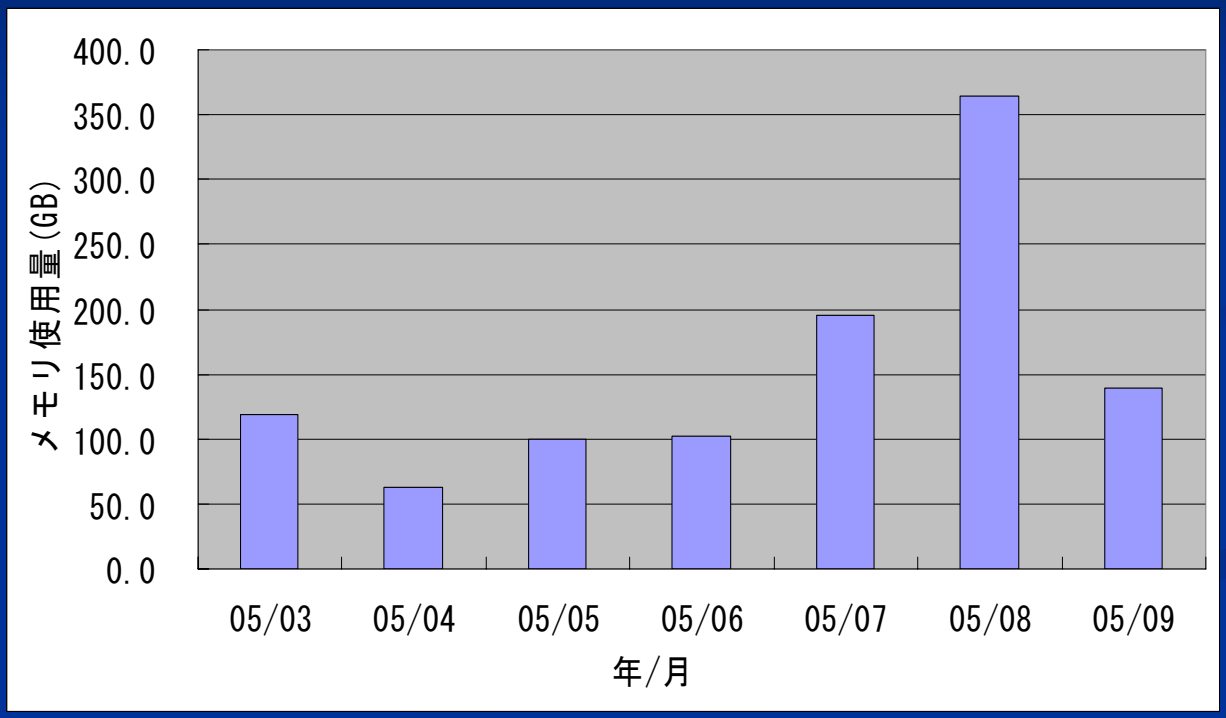


（1利用者のログインを1日1回のみカウント）

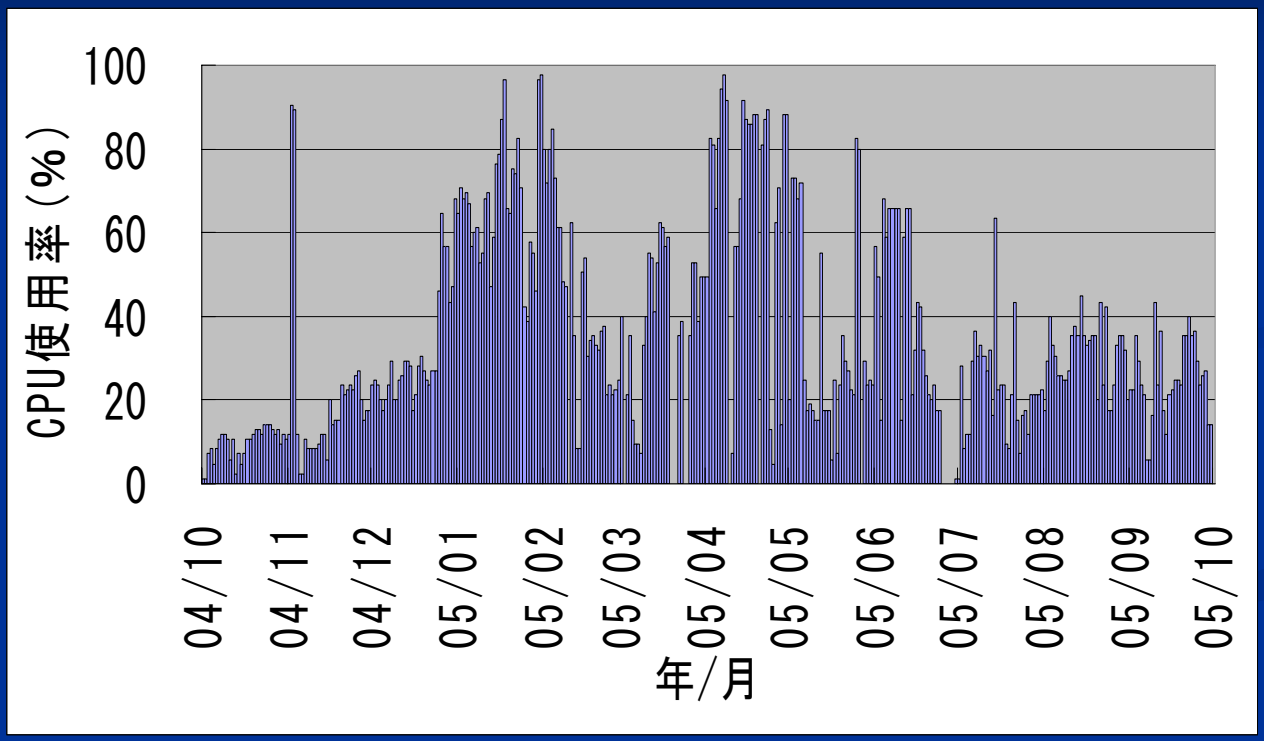
SGI Altix3700 の月別平均CPU使用率



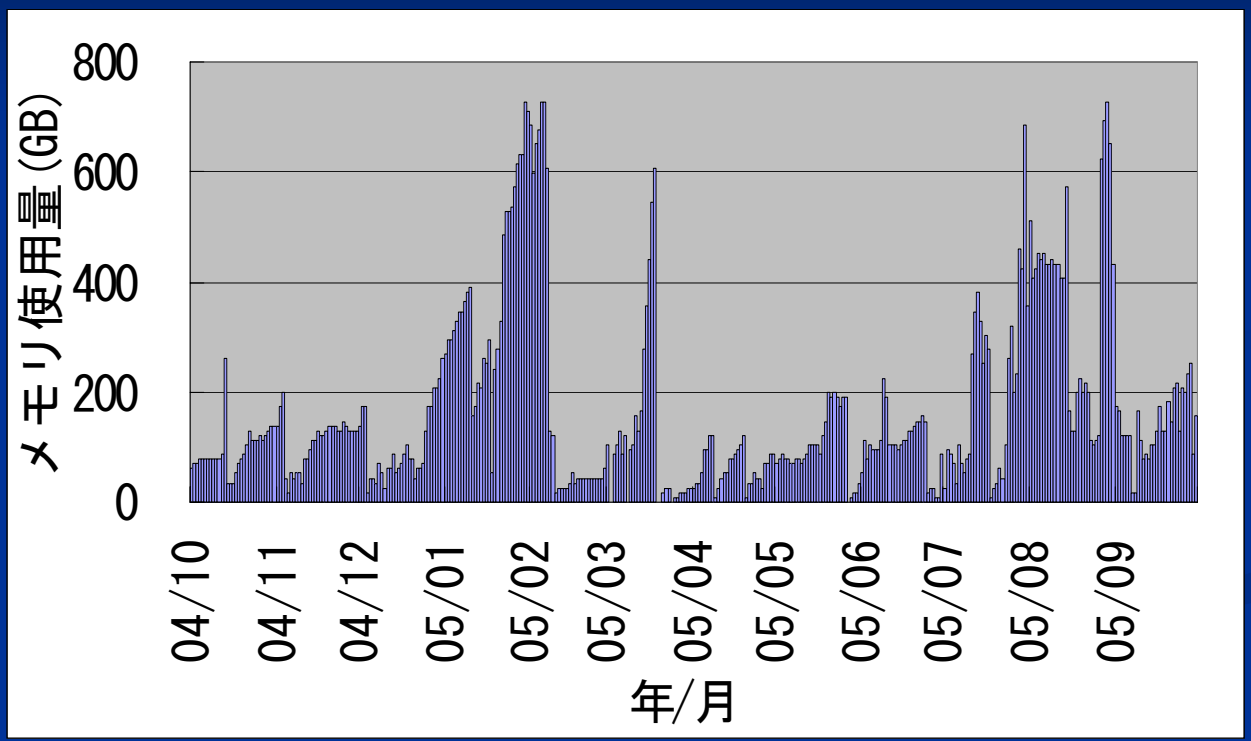
SGI Altix3700 の月別平均メモリ使用量



SGI Altix3700 の日別平均CPU使用率



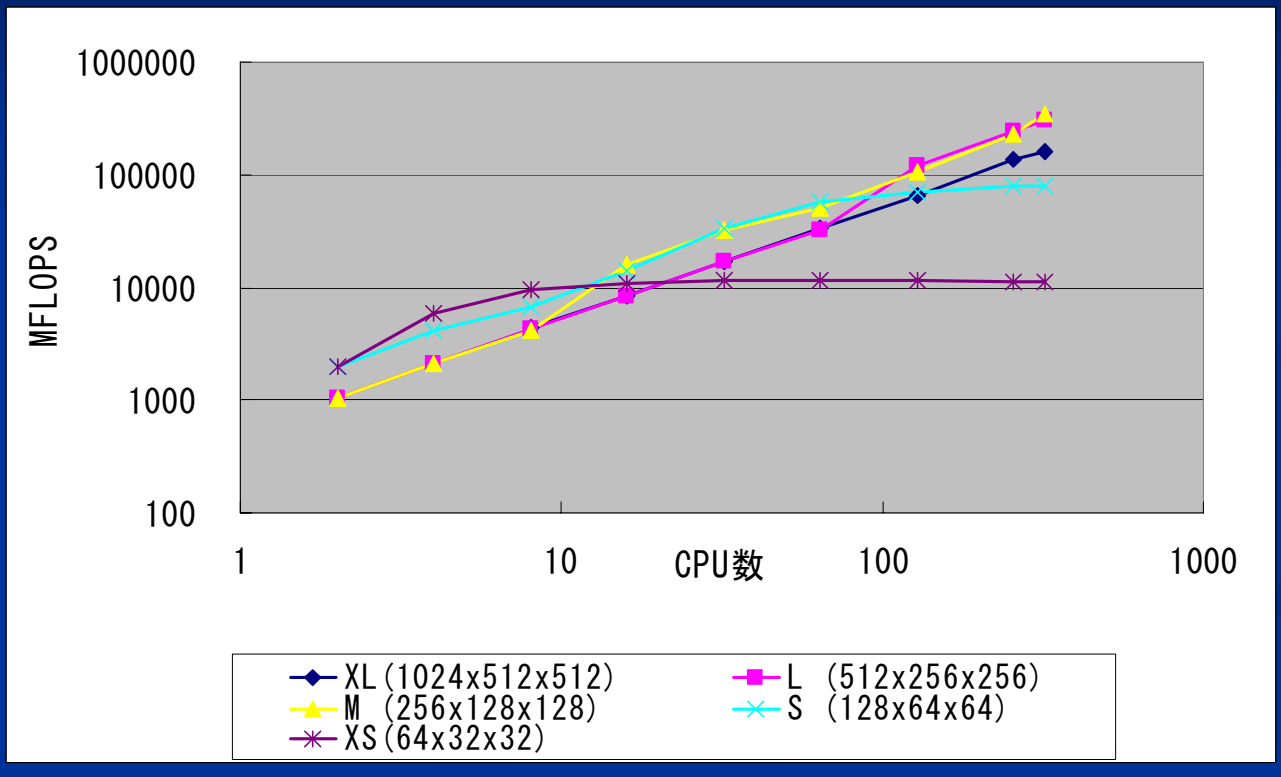
SGI Altix3700 の日別平均メモリ使用量



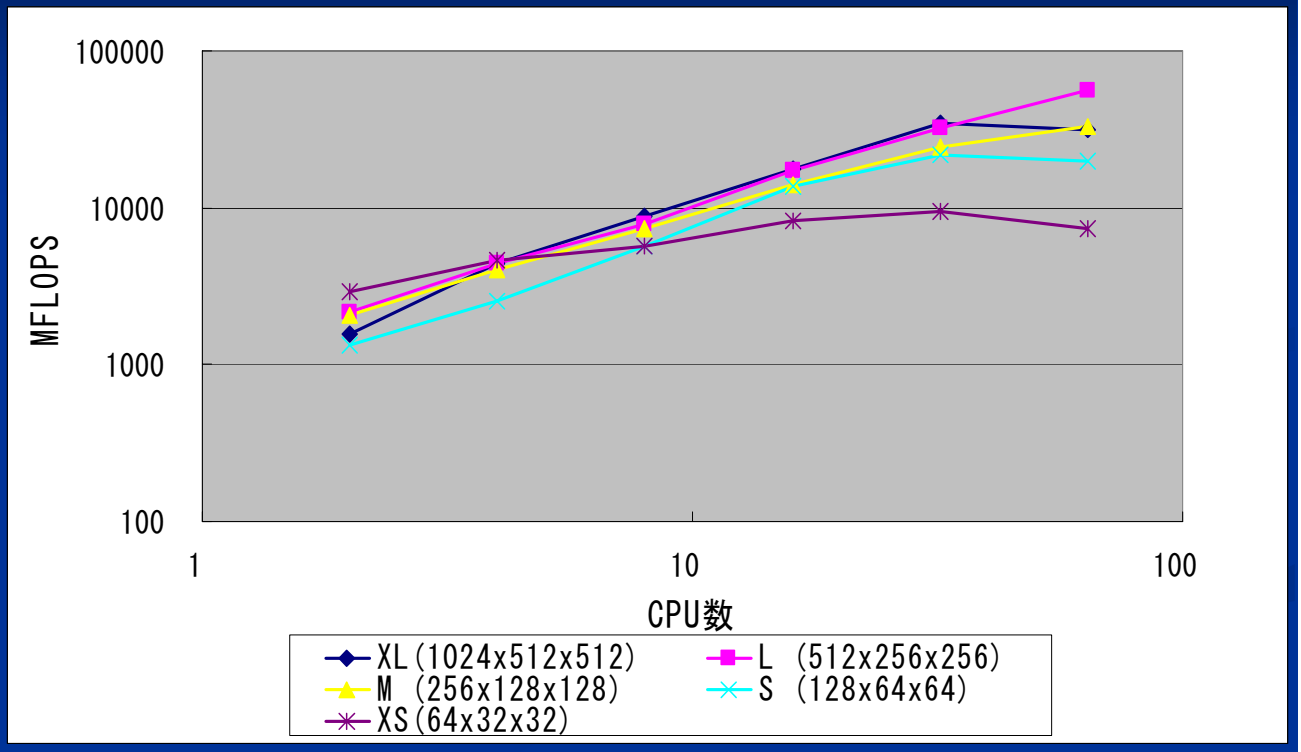
アウトライン

- 情報環境の概要
- 計算サーバの変遷
- 計算サーバの導入の基本的な考え方と経緯
- 代表的な計算サーバ
  - Cray XT3
  - SGI Altix3700　他
- 運用の基本的考え方および稼動統計
- 姫野ベンチマーク
- おわりに

姫野ベンチ (Cray XT3　MPI)



姫野ベンチ (SGI Altix3700, MPI)



おわりに

- JAIST の情報環境および計算サーバの紹介
  - Cray XT3
  - SGI Altix3700
  - ベクトル計算機、SMP計算機、PCクラスターなど
- 計算サーバの導入の基本的な考え方および導入の経緯

  ⇒ 情報環境の1部(マルチベンダー)、最新の計算機
- 計算サーバ運用の基本的な考え方

  ⇒ センターの関与をなるべく少なく、ユーザが主体
- 稼動統計と姫野ベンチマークテストの紹介