



SCore の紹介 および今後の開発

開発部会委員長
石川 裕
(東京大学)



発表概要

- ◆ クラスタとは？
- ◆ SCoreの概要
- ◆ 新リリース
- ◆ 今後の開発





クラスタとは？

- ◆ ネットワークに繋がったコンピュータを使って
並列コンピュータ環境を実現したシステム
- ◆ HA (High Availability)
- ◆ HPC (High Performance Computing)



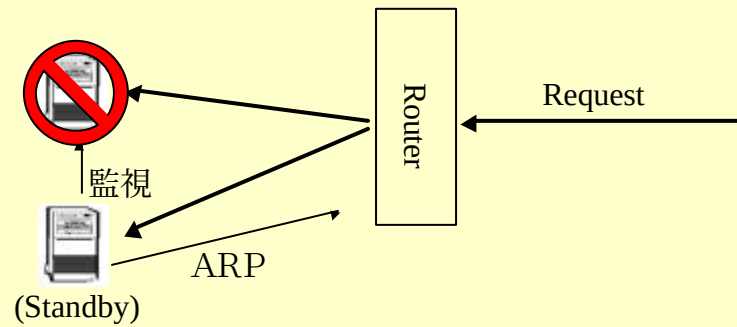
並列処理市場

- ◆ HA (High Availability)
 - ◆ アプリケーション
 - ◆ Mail/Web/News/File Server
 - ◆ プラットフォーム
 - ◆ サーバコンピュータ
 - ◆ クラスタ
- ◆ HPC (High Performance Computing)
 - ◆ アプリケーション
 - ◆ 科学技術計算
 - ◆ プラットフォーム
 - ◆ ベクタ型スーパーコンピュータ
 - ◆ サーバコンピュータ
 - ◆ クラスタ
 - ◆ Job Dispatching(分散コンピューティング)
 - ◆ Parallel Computation
 - ◆ Beowulf(ベオウルフ)
 - ◆ SCore



High Availability

◆ Fail Over Services



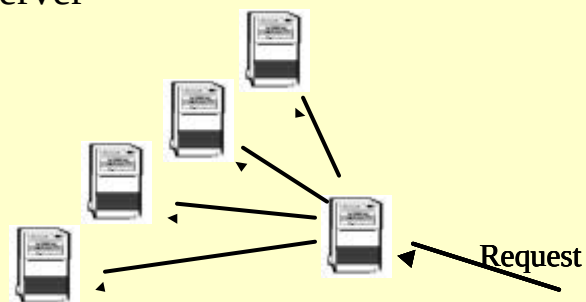
02/03/20

5



High Availability

◆ Virtual Server



負荷分散および対故障性向上

02/03/20

6

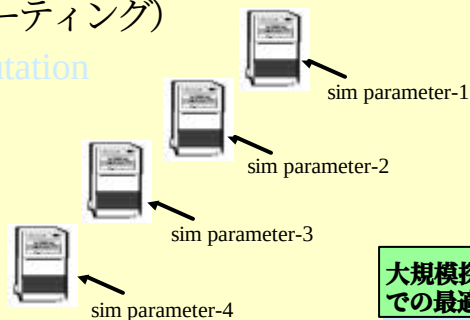




High Performance Computing

◆ Parallel Job Dispatching (分散コンピューティング)

◆ Parallel Computation



大規模探索空間上
での最適解探し

02/03/20

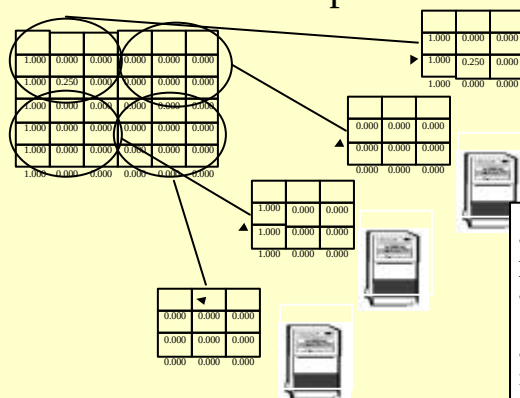
7



High Performance Computing

◆ Parallel Computation

SCoreが目指してきたもの

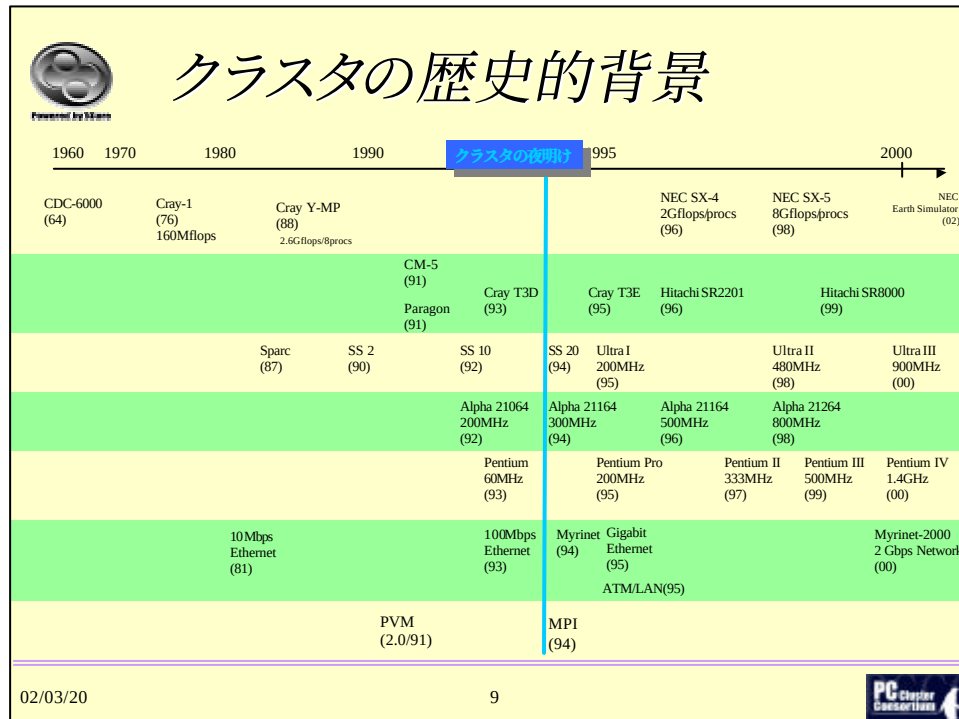


- コンピュータ間でのデータ交換にはMPI通信ライブラリが一般に使われる。
- MPI通信ライブラリを使って並列処理するプログラムが必要。
- Omni OpenMP on SCoreを使うと共有メモリ型コンピュータ上のOpenMPプログラムは、そのままタスクで動く。

02/03/20

8





クラスタの夜明け

- ◆ 1994年以前
 - ◆ Ethernet上でTCP/IPを使ったクラスタの試みがあった
- ◆ 1994年
 - ◆ カリフォルニア大学バークレイ校
 - ◆ David Culler率いるNOW (Network Of Workstation)プロジェクト
 - ◆ ハードウェアとしてSun, Myrinetを使用
 - ◆ イリノイ大学
 - ◆ Andrew Chien率いるプロジェクト
 - ◆ ハードウェアとしてSun, Myrinetを使用。後にPC, Myrinetを使い、OSは Windows NT
 - ◆ NASA
 - ◆ Thomas Sterling 率いるBeowulf プロジェクト
 - ◆ ハードウェアとしてPC, Ethernetを使用。OSはLinux
- ◆ 1995年
 - ◆ 新情報処理開発機構
 - ◆ SCoreクラスタ
 - ◆ ハードウェアとしてSun, Myrinetを使用。1996年からPC, Myrinetを使用。OSはLinux。
 - ◆ 1995年夏Myrinetを導入する。国内で最初。

02/03/20 10



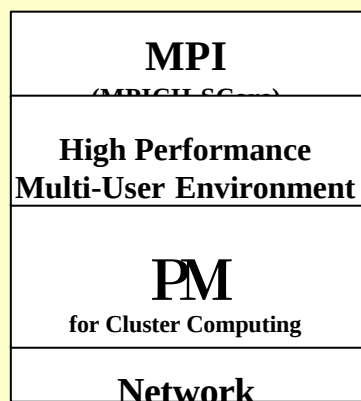
SCoRE vs Beowulf

- ◆ クラスタコンピューティングのためのシステムソフトウェア環境を提供
 - ◆ PM通信機構
 - ◆ MPI
 - ◆ SCoRE-Dグローバルオペレーティングシステム
 - ◆ ギャングスケジューリングによるTSS環境
 - ◆ バッチスケジューリング
- ◆ 既存のフリーソフトウェアの組み合わせ
 - ◆ TCP/IP
 - ◆ MPI
 - ◆ グローバルオペレーティングシステムなし

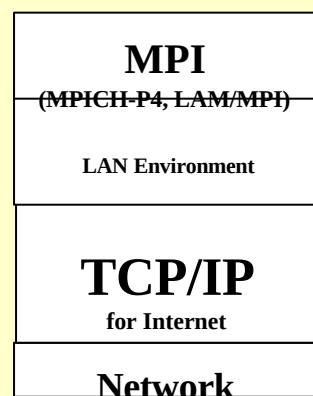


SCoRE vs. Beowulf

SCoRE



Beowulf





SCore System Software

高性能通信ライブラリ

PM/Myrinet

15.0 usec Round Trip time
233 MB/s Bandwidth

MPICH-SCore MPI Library

24.4 usec Round Trip time
228 MB/s Bandwidth

PM/Ethernet Network Trunking

複数のNICを有効利用

グローバルOS

SCore-D

シングル・マルチユーザ環境
ギャングスケジューリング
チェックポイント・リスタート

並列プログラミング言語

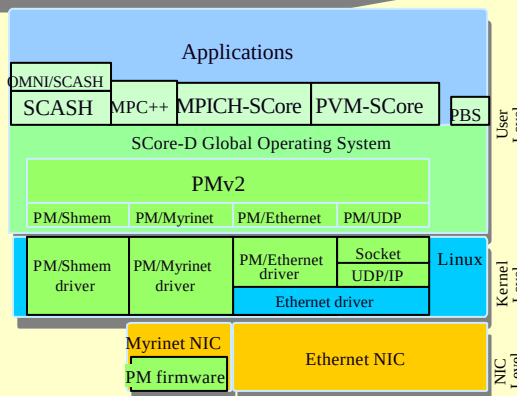
MPC++ Multi-Thread Template Library

共有メモリ・マルチスレッドプログラミング言語

Omni OpenMP on SCASH

10 times faster than Fast Ethernet + TCP/IP

Three times as fast as Gigabit Ethernet + TCP/IP



02/03/20

13



SCoreの特長

- ◆ 高性能通信機能
- ◆ 単一システムイメージ
- ◆ シームレスクラスタ群環境
- ◆ ハイアベイラビリティ
- ◆ シームレス並列プログラミング
- ◆ インストレーションツール

02/03/20

14





Omni OpenMP on SCore

- ◆ OpenMP

- ◆ 共有メモリ型並列コンピュータ上での並列処理用言語

```
#pragma omp for
for (int y = 1; y <= N; y++)
  for (int x = 1; x <= M; x++)
    u[y][x] = (uu[y][x - 1] + uu[y][x + 1]
              + uu[y - 1][x] + uu[y + 1][x]) / 4.0;
```

- ◆ 商用 OpenMP on Clusters:

- ◆ OpenMP + MPI

Omni OpenMP on SCore
クラスタ上でOpenMPプログラム
を修正せずに動く



SCoreのアプリケーション例

- ◆ バイオインフォマティクス

- ◆ MolTreC, PAPIA: 産業技術総合研究所



- ◆ 先端医療

- ◆ MRI画像解析: 先端医療振興財団、産業技術総合研究所



- ◆ 先端画像処理

- ◆ ボリュームレンダリング: 産業総合技術研究所、三菱プレシジョン



- ◆ シミュレータ

- ◆ リアルタイム電力系統シミュレータ 三菱電機





新リリース SCore 5.0

◆ 著作権

◆ PC クラスタコンソーシアム

◆ 新情報処理開発機構から移譲

- ◆ 新情報処理開発機構は1992年から10ヵ年計画で始まった通商産業省 現在の経済産業省)主導プロジェクトである「リアルワールドコンピューティングプログラム」を実施した組織

◆ 利用条件

◆ LGPL

- ◆ 再配布、ソースプログラムの改変可

02/03/20

17



新リリース SCore 5.0

◆ シングルユーザモードでの利用勝手向上

◆ scout コマンドを使わずにアプリケーション実行

- ◆ `scrun -group=pcc1@host2:4+pcc2:4 ./a.out`

◆ チェックポイント機能サポート

- ◆ 従来はマルチユーザモードでしか使用できなかった

◆ ドキュメント構成の変更

◆ 新EITインストレーションツール

- ◆ DHCP用にIPアドレスを必要としなくなった
- ◆ 分かりやすいGUI

02/03/20

18





今後の開発部会の活動ポリシー

- ◆ コンソーシアムにおける活動
 - ◆ 新情報処理開発機構から引き継いだソフトウェアの開発を継続・保守し、コモディティハードウェアに対応
 - ◆ 国際標準化のための作業
 - ◆ 開発したソフトウェアはコンソーシアムの下、LGPL
- ◆ 会員における活動
 - ◆ SCore上でのツール、ミドルウェアの開発
 - ◆ 開発したソフトウェアは会員の著作物
 - ◆ 会員の開発したソフトウェアをSCore標準リリースに含めるかどうかは会員とコンソーシアムとの間で決める



今後の開発

- ◆ コンソーシアム活動
 - ◆ IA64対応
 - ◆ OSDL(Open Source Development Lab)の機材を使用させていただき、SCoreの移植ならびに性能評価を行う
 - ◆ <http://www.osdl.jp> のホームページから抜粋
- ◆ 新My
 - ◆ OSCARのマージ
 - ◆ ISI (インストレーションツール)のマージから始める

OSDLは、Linux とLinux 上で動作するアプリケーションとを、データセンタやキャリアクラスへ展開させることを目的としています。Linux を通信事業やエンタープライズレベルへさらに浸透させるため必要なリソースを提供し、テストや開発において重要なハードウェアを世界中のオープンソース開発者へ開放します



Powered by Xilinx

今後の開発

- ◆ 会員における活動の一部紹介
 - ◆ MPI-2およびGlobus 対応
 - ◆ MPICHからの脱却
 - ◆ Linda/SCore
 - ◆ Gaussianを動かすために
 - ◆ センター運用ツールの整備
 - ◆ アカウンティング、運用管理ツール、診断ツール
 - ◆ Omni OpenMP on IA64
 - ◆ クラスタにおける本格的マルチスレッドプログラミング環境を目指して
 - ◆ PM/Infiniband
 - ◆ 次世代ネットワーク
 - ◆ クラスタファイルシステム

02/03/20

21



皆様のご参加をお待ちしております

<http://www.pccluster.org>