

SCore 7 の概要

PC Cluster Consortium 開発部会
Allinea Software
堀 敦史

SCore 7 で変わるもの

- 新しいネットワークへの対応
 - Myrinet 10G
 - InfiniBand
- マルチコアでの性能向上
- ジョブ管理の信頼性向上
- 細かい機能改良

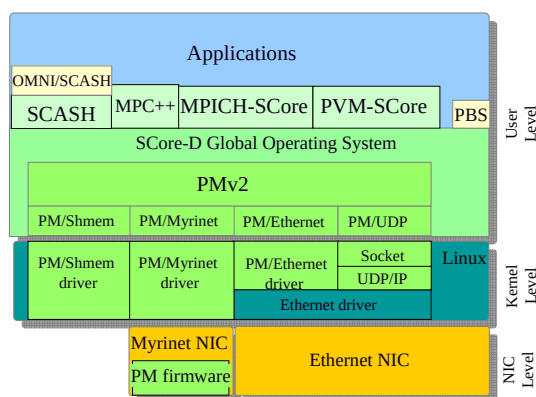
SCore 7 で変わらないもの

- Ethernet, Shmem のサポート
- ギャングスケジューリング
- チェックポイント(パリティ)
- 並列ジョブ
- カーネルパッチは不要

⇒実質的にほとんど変化なし

⇒しかし、内部的にはスクラップ&ビルド

SCore 6 の根本的問題点



- SCoreパッケージ
 - 密接に関係
 - 分割不可
 - 独自API
 - 全て独自開発
- 問題点
 - 外部通信ライブラリの取込困難
 - 新たな通信ハードへの対応はかなりの労力が必要

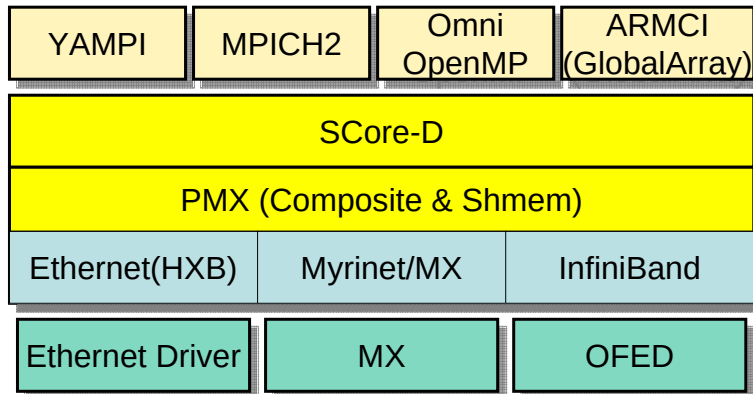
SCore 6 の問題点

- SCore はクラスタオペレーティングシステム
 - SCore の開発がクラスタ用 OS の研究から
 - 特殊(当初はユニーク)な設計思想、API
 - ユーザレベル通信、かつ OS 機能を実現する
- PMv2 の性能上の問題点
 - マルチコアでの性能低下が顕著
 - PMv2 の再設計が必要

SCore 7 技術的チャレンジ

- 新しい SCore 通信ライブラリ *PMX*
 - 第3者開発の通信ライブラリと親和性を高める
 - MX - Myrinet
 - OpenFabrics - InfiniBand
 - マルチコアでの性能向上
 - PMv2 の利点はそのまま
 - マルチネットワークのサポート
 - 複合ネットワークのサポート
 - ギャングスケジューリングやチェックポイントも可能

SCore 7 ソフトウェア階層



ARMCI: Aggregated Remote Memory Copy Interface
OFED: OpenFabrics Enterprise Distribution

SCore 7 : PMX

- 高速化
 - Composite と Shmem の一体化
 - ロック機構の見直し(マルチコアの性能向上)
- 外部通信ライブラリの取込み
 - Myrinet(10G): MX 通信ライブラリ
 - InfiniBand: OFED
 - 他のネットワークへの対応が容易になる
- マルチネットワーク、複合ネットワーク機能は従来通り

SCore 7 : SCore-D

- オペレーティングシステム的な色合いを薄く
 - MPICH-2 のプロセスマネージャに近い
 - PMX の管理は必要最低限
 - プロセスの生成と管理に徹する
 - 標準入出力の管理
 - 並列ジョブ
 - チェックポイントの起動
- ギャングスケジューリング
 - 新たなタイムシェアリング専用スケジューラを別途

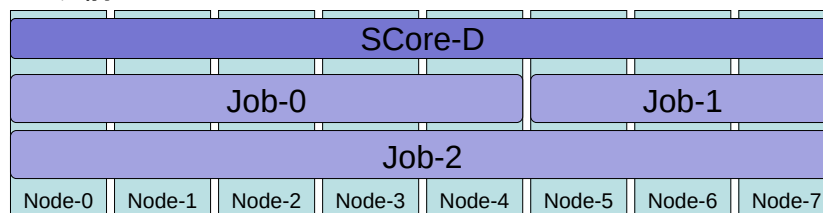
2008年7月25日

PC クラスタワークショップ in 京都

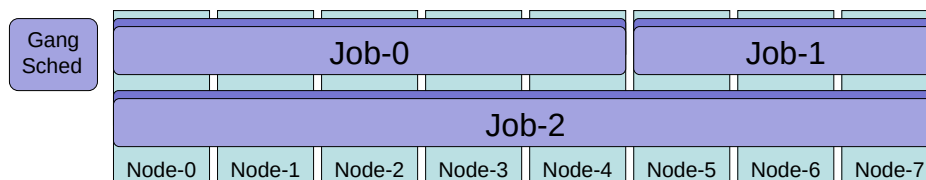
9/14

SCore-D の具体的な違い

SCore 6 以前



SCore 7



2008年7月25日

PC クラスタワークショップ in 京都

10/14

SCore 7 : SCore-D の違い

- ギャングスケジューリング
 - SCore-D はジョブを管理 <-> 以前はクラスタを管理
 - タイムシェアリングスケジューラを別途設ける
- ノードの故障
 - そのノードで走るジョブのみが影響を受ける
- SCore-D の信頼性向上
 - 影響の及ぶ範囲がジョブに限定
 - 機能の単純化による信頼性の向上も期待できる

ギャングスケジューリング再考

- ギャングスケジューリング
 - クラスタでの効率的な時分割スケジューリング手法
- クラスタでのスケジューリングの実態
 - 効率重視
 - ほとんどがバッチスケジューリング
- ギャングスケジューリングが有効とされるケース
 - プログラム開発、実時間シミュレーション
- 効率よりも信頼性が重要 !!

SCore 7 開発状況

- PMX
 - Composite(Shmem) 安定動作
 - Ethernet 安定動作
 - Ethernet-HXB 基本動作確認済み
 - MX (Myrinet) 基本動作確認済み
 - InfiniBand 基本動作確認済み
- SCore-D コーディング中
- MPICH2 基本動作確認済み

リリーススケジュール

- 2007/11 PMX 基本機能 SC@Reno でデモ
- 2008/06 MPICH2 実装完了
- 2008/07 SCore 7 α リリース
- 2008/10 ギャングスケジューリング、
 チェックポイント実装
- 2008/11 SCore 7 β リリース SC@Austin でデ
 モ
- 2009/03 SCore 7 正式リリース