

調査 5 モデルカリキュラムの提言 コースウェア

10. クラスタシステム構築に関するスキル

I. 概要	High Availability Cluster と High Performance Computing Cluster の構築にかかわる技術の習得およびシステム構築の演習を通してクラスタ構築に関する基礎技術の習得を目指す。HPC Cluster の講義では、並列プログラミングに関する講義も行う。
II. 対象専門分野	HA クラスタ、HPC クラスタ
III. 受講対象者、 受講前提	HA クラスタ、HPC クラスタの構築、設計を行う SE。 Linux のインストール、設定が行える。 C、Fortran でのプログラミング経験者。
IV. 学習目標	HA クラスタ、HPC クラスタの構築、コンサルティングが行える技術を習得する。
V. 使用教科書、 教材等	HA クラスタに関しては、適切な市販教材がないため、オリジナル教材を作成するものとする。 HPC クラスタに関して 『Building LINUX CLUSTER』 David HM Spector 著、O'Reilly 刊 『High Performance Cluster Computing Vol. 1』 Rajkumar Buyya 著、Prentice Hall 刊 『PC クラスタ構築法』 トーマス L スターリング他著、産業図書刊 『MPI 並列プログラミング』 Peter S Pacheco 著、秋葉博訳、培風館刊 『OpenMP による並列プログラミングと数値計算法』 牛島省著、丸善刊 『マルチコア CPU のための並列プログラミング—並列処理 & マルチスレッド入門』 安田絹子 他著、秀和システム刊
VI. 習得スキルの 評価方法	HA クラスタおよび HPC クラスタの構築手順書を作成し、評価する。
VII. カリキュラム の構成	レベル 1 第 1 回～第 7 回 レベル 2 第 8 回～第 15 回

講座内容

第 1 回 クラスタシステム概論、HA クラスタ その1 (講義 90 分)

HA クラスタ、HPC クラスタの概要を理解する。HA クラスタの概念およびその基本技術に関する理解を深める。

(1) High Availability (HA) Cluster と High Performance Computing (HPC) Cluster

1. High Availability Cluster とは
2. High Performance Computing Cluster とは

(2) HA クラスタ概論

1. 負荷分散クラスタとフェイルオーバークラスタ
2. 負荷分散クラスタ
 - ・ DNS ラウンドロビン
 - ・ ロードバランサー

◆ 負荷分散クラスタの基本的な構成とその構成要素

➤ リアルサーバを選出するアルゴリズム

- ① round robin
- ② least connections
- ③ hash

➤ リアルサーバの死活監視

- ① ICMP
- ② TCP
- ③ protocol

◆ ロードバランサーの区分

- OSI 参照モデル
- L4 ロードバランサーと L7 ロードバランサー
- L7 ロードバランサーのプロダクト紹介 (2007 年 8 月現在)

① HA Proxy

<http://haproxy.1wt.eu/>

② mod_proxy_balancer

http://httpd.apache.org/docs/2.2/en/mod/mod_proxy_balancer.html

③ POUND

<http://www.apsis.ch/pound/>

④ Perlbal

<http://www.danga.com/perlbal/>

⑤ LVS

<http://www.linuxvirtualserver.org/>

第2回 HA クラスタ その2 (講義+ワークショップ 90分)

HA クラスタ構築を実際に行い、その構築手順を理解する。

- (1) LVS で実現するロードバランサー
 1. ロードバランサーの構成要素
 - ・ IPVS 対応カーネル
 - ・ ipvsadm
 - ・ keepalived
 - ・ iptables
 - ・ iproute
 2. IPVS カーネルの再構築
 3. ipvsadm と keepalived のビルド
 4. テスト環境の構築
 5. LVS-NAT による負荷分散
 6. DSR(Direct Server Return)の構築

第3回 HA クラスタ その3 (講義+ワークショップ 90分)

HA クラスタ構築を実際に行い、その構築手順を理解する。

- (1) ロードバランサーの冗長化
 1. VRRP(Virtual Router Redundancy Protocol)とは
 - ・ VRRP の死活監視
 - ・ マスターの選定方法
 - ・ プリエンプティブモード
 2. VRRP の構築
 - ・ ネットワーク構成
 - ・ keepalived の設定
 - ・ ネットワーク障害時の動作
 - ・ フェールオーバーの影響
- (2) keepalived の運用のテクニック

第4回 コンピュータシミュレーション(講義 90分)

コンピュータシミュレーションの必要性および現在の HPC の動向に関して理解する。

- (1) 科学技術研究基盤としてのコンピュータシミュレーション
 1. 科学技術計算史
 2. コンピュータシミュレーションとは
- (2) スーパーコンピュータ動向
 1. はじめに
 2. スーパーコンピュータハードウェア動向
 - ・ スーパーコンピュータの性能の推移
 - ・ 米国 ASCI(Accelerated Strategic Computing Initiative)プロジェクト
 - ・ 日本 地球シミュレータ計画
 - ・ 日本 京速コンピュータ計画
 - ・ 日米の開発戦略の相違
 - ・ PC クラスターの登場
 - ◆ SCore
 - ◆ 東京工業大学 TSUBAME (Tokyo-tech Supercomputer and UBiquitously Accessible Mass-storage Environment) システム
 - ◆ 筑波大学 PACS-CS プロジェクト
 - ・ グリッドコンピューティング(ネットワーク結合型コンピュータ利用技術の最新動向)
 3. スーパーコンピュータの応用範囲

第5回 並列プログラミング概論(講義 90分)

並列プログラミングの概要を理解する。

- (1) 計算の高速化と並列処理の必要性
- (2) 並列計算機の分類
- (3) 並列計算機の構成方式
 1. 共有メモリモデル
 2. メッセージ交換モデル
- (4) プロセッサ間のネットワーク
- (5) 並列処理のプログラミング
- (6) 並列処理の効率
 1. アムダールの法則
 2. 問題サイズとリアルタイム性
- (7) 数値計算の並列化
 1. 差分法の並列化
 2. 有限要素法における並列化
 3. 粒子法による物質の挙動シミュレーション
 4. モンテカルロシミュレーション
- (8) 並列プログラミング環境とツール

マルチスレッドプログラミングの基礎を習得する。

- (1) マルチスレッドプログラミング入門
 1. シングルスレッドとマルチスレッド
 2. スレッドとプロセスの違い
 3. マルチスレッドの動作原理
- (2) マルチスレッドプログラミングの基礎
 1. スレッドの操作
 - ・ 生成
 - ・ 終了
 - ・ 同期
 - ◆ 終了待ち
 - ◆ 相互排除
 - ◆ 条件変数
 - ・ 強制終了
 2. pthread
 - ・ pthread を利用したプログラムのコンパイル
 - ・ スレッドの生成と消滅
 - ◆ fork-jion モデル
 - ・ mutex によるスレッド間の同期
 - ◆ 共有資源
 - ◆ 複数のスレッドによる共有資源のアクセス
 - ◆ 相互排除
 - ・ 条件変数によるスレッド間の同期
 - ◆ パイプと循環バッファ
 - ◆ 条件変数を使った環境バッファ
 - ・ 再帰的 Mutex
 - ◆ mutex でのデッドロック
 - ◆ 再帰的 mutex
 - ◆ stdio 専用再帰的 lock
 - ・ pthread とメモリ
 - ◆ auto 変数
 - ◆ static 変数
 - ◆ static 変数を使ったライブラリ関数
 - ◆ スレッドセーフ

- ・ pthread でのセマフォ利用

第 8 回 並列プログラミング実践 その2 HPF (High Performance Fortran) と OpenMP (講義 90 分)

HPF および OpenMP の基礎技術を習得する。

- (1) HPF (High Performance Fortran) 入門
 1. 並列プログラムの開発と HPF
 - ・ データの並列化(データマッピング)
 - ・ 処理の並列化(計算マッピング)
 2. 簡単な HPF プログラミング
 3. HPF の仕様と概念
- (2) HPF プログラミング
 1. データマッピング関連の指示文
 2. 計算マッピングと通信関連の指示文
 3. 不規則問題のための指示文
 4. 外來手続き
- (3) OpenMP
 1. OpenMP の概要
 2. API の紹介
 - ・ 並列実行領域 (Parallel Regions) 構文
 - ・ ワークシェアリング (Worksharing) 構文
 - ・ データ環境 (Data Environment) 構文
 - ・ 同期 (Synchronization) 構文
 - ・ 実行時関数/環境変数 (Runtime functions/environment variables)
 3. OpenMP プログラムサンプルの紹介とデモ

第 9,10,11 回 並列プログラミング 実践 その3

MPI (Message Passing Interface) (講義+ワークショップ 90 分 x 3)

HPC クラスタでのプログラミングのトレンドである MPI の基礎技術を習得する。

(1) MPI (Message Passing Interface)

1. MPI の概要
2. MPI のプログラミングモデル
3. MPI の基本機能
 - ・ Hello World での例
 - ・ mpi.h
 - ・ MPI_Init
 - ・ MPI_Finalize
 - ・ rank
 - ・ コミュニケータ
 - ・ MPI_Comm_size
 - ・ MPI_Comm_rank
 - ・ MPI_Send
 - ・ MPI_Recv
4. MPI の先端的特長

(2) MPI によるプログラミング

1. 1 対1通信
2. グループ通信
 - ・ π の計算の並列化
 - ◆ アルゴリズム
 - ◆ MPI での実装例
 - ・ MPI_Bcast
 - ・ MPI_Reduce
 - ・ 演算のバンドル
3. MPI ライブラリの作成
 - ・ MPICH2 のコンパイル
 - ・ MPICH2 の実行環境の整備
4. MPI プログラムのコンパイルおよび実行
 - ・ MPICH2 による MPI プログラムのコンパイル
 - ・ MPICH2 による MPI プログラムの実行

(3) プログラミングの並列化の方法

1. どんなところが並列化できるか
2. 計算部分の並列化パターン
3. 入出力部分の並列化パターン
 - ・ 入力部分の並列化パターン
 - ・ 出力部分の並列化パターン
4. ループの分割方法
 - ・ ループの反復と配列
 - ・ 配列の分割方法
 - ・ ブロック分割でのロードバランスが不均一になる例
 - ・ ブロック分割
 - ・ サイクリック分割
 - ・ ブロック・サイクリック分割
 - ・ 配列の最小
 - ◆ 配列の最小方法
 - ◆ 配列の最小に伴う変更 (I/O 以外の部分)
 - ◆ 配列の最小に伴う変更 (I/O 部分)

(4) 並列化の応用例

1. 差分法の徹底攻略
 - ・ 2次元目でブロック分割した場合
 - ・ 1次元目でブロック分割した場合
 - ・ 1、2次元目でブロック分割した場合
 - ・ 1、2次元目でブロック分割した場合(斜めの要素の参照)
 - ・ 配列を縮小する場合
2. 有限要素法
 - ・ 有限要素法の並列化
3. LU 分解
4. ICCG 法
5. マルチフロントル法
6. SOR 法
7. モンテカルロ法
8. 個別要素法/分指導力学法

第 12 回 Beowulf PC クラスタの構築（ワークショップ 90 分）

一般的な PC クラスタの構築手順を理解する。

- (1) Beowulf PC クラスタのコンポーネント
- (2) システム構築
 1. NIS の設定
 2. NFS の設定
 3. ntp の設定
 4. rsh の設置
 5. MPICH2 のコンパイル・インストール・設定
 - ・ GNU コンパイラ
 - ・ PGI コンパイラ
 - ・ Intel コンパイラ
- (3) mpich2 のサンプルプログラムのコンパイル・実行
- (4) ベンチマーク測定
 1. SkaMPI を使用したベンチマーク測定

第 13 回 SCore クラスタ（講義+ワークショップ 90 分）

国産の HPC クラスタ専用ミドルウェア SCore の構築手順を理解する。

- (1) SCore の紹介
- (2) SCore の優位性
- (3) SCore のソフトウェアアーキテクチャ
- (4) ユーザ環境
 1. シングルユーザモード
 2. マルチユーザモード
- (5) SCore のインストール
- (6) SCore のデモンストレーション

第 14 回 PC クラスタの周辺技術（講義 90 分）

HPC クラスタで使用すると便利なツールを紹介する。

- (1) バッチ処理システムの紹介
 1. バッチ処理システムとは
 2. OpenPBS、Torque
 3. GridEngine
 4. PBS Pro
 5. LSF
- (2) システム監視の紹介
 1. Ganglia
- (3) 並列ライブラリの紹介
 1. ScaLAPACK
 2. NAG Parallel Library
- (4) 並列デバッガ、プロファイラの紹介
 1. インテルトレースアナライザ/コレクタ
 2. TotalView
 3. PARAVAR
- (5) 開発環境の紹介
 1. Eclipse PTP (Parallel Tools Platform)
- (6) ベンチマークソフトウェアの紹介
 1. SKaMPI
 2. Intel MPI Benchmarks
- (7) PC クラスタで使用可能な商用ソフトウェアの紹介

第 15 回 グリッド・コンピューティング (講義 90 分)

グリッド・コンピューティングに関する概念および最新の動向に関して理解する。

- (1) グリッド・コンピューティングの概要
 1. グリッド・コンピューティングとは
 2. グリッド・コンピューティングが登場した背景
- (2) グリッド・コンピューティングの分類
 1. プロセッシング・グリッド
 2. データ・グリッド
 3. ビジネス・グリッド
- (3) グリッド・コンピューティング関連プロジェクトの紹介
 1. 米国 Extensible Teragrid Facility (ETF)
 2. 英国 UK e-Science
 3. 欧州 EU FP6、EGEE、DEISA、CoreGRID、UNIGRID、NEXTGRID…
 4. 日本 NAREGI (National Research Grid Initiative)
- (4) The Globus Alliance と Globus Toolkit

以上