

# HPC分野への取り組みと、 APPRO社Xtreme-Xのご紹介

～PCクラスタワークショップin大阪～

平成21年3月13日  
株式会社HPCソリューションズ  
根本 雅樹



## アジェンダ

- 株式会社HPCソリューションズのご紹介
- 米国APPRO社のご紹介
  - APPRO製品のご紹介
- スーパーコンピュータXtreme-Xのご紹介



## 会社概要

事業所	本社 東京都中央区日本橋大伝馬町3-2 タウンチ小伝馬町 1階 TEL:03-5640-7858 FAX:03-5640-2578
	大阪事業所 大阪府大阪市北区西天満3-5-10 オフィスポート大阪602号 TEL:06-6364-3025 FAX:06-6364-3026
	つくば事業所 茨城県つくば市二宮1-2-2 酒井ビル TEL:029-861-0250
役員構成	代表取締役会長 藤澤 義麿 代表取締役社長 河野 証 専務取締役 東 孝司 取締役(非常勤) 中田 義弘 取締役(非常勤) 伊藤 俊彦 監査役(非常勤) 酒巻 千丈
設立	2006年7月1日
事業内容	PCクラスタシステムの販売およびサポート、 クラスタシステム構築等のコンサルティング、 保守



3

## 業務内容

- 各ベンダーのハードウェア、ソフトウェアの販売
  - ◆ ⇒米国APPRO社製品も取り扱っています。
- マルチベンダーサポート
  - ◆ お客様のHPC用途コンピュータシステムの高速化、安定稼働を目指し、あらゆるベンダーのあらゆるコンピュータシステムをワークステーションから大規模クラスタまで一貫してサポート。
    - システム設計、クラスタ設計
    - システム構築、クラスタ構築
    - システム開発
    - システムサポート
    - サイトプランニング
    - HPCアプリケーションベンチマーク
    - コンサルテーション



4

# 国内初 Windows HPC Server 2008クラスタシステム の構築・導入

## ナカシマプロペラ様

- OS :
  - ◆ Windows HPC Server 2008
- ノード :
  - ◆ PowerEdge1950 32台
    - CPU : Intel Xeon5406
- インタコネクト :
  - ◆ InfiniBand(4X DDR)(Voltaire)
- アプリケーション :
  - ◆ SCRYU/Tetra(ソフトウェアアクレイドル)



日刊工業新聞, 2008年12月29日付 5面

## Appro社のご紹介

APPRO INTERNATIONAL INC

---

# 会社概要

:: HPC専門の会社

- 1991年設立
- 本社:Milpitas - カリフォルニア(シリコンバレー)  
事業所:Houston - テキサス
- ハイパフォーマンスコンピューティングに特化したソリューション提供会社
  - 大学官公庁/ 防衛
  - 製造
  - バイオ
  - EDA
  - 金融
  - 石油・ガス



# APPRO社のHPC製品

- スーパーコンピュータ
  - Appro Xtreme-X (Nehalem)
- クラスタ
  - HyperGreen Clusters
- サーバ
  - 1U/2U DP RM Servers (Harpertown)
  - 1U/2U DP RM Servers (Nehalem)
  - 1U/2U DP/QP RM Servers (Shanghai)
- ワークステーション
  - Entry WH300 Intel & AMD WS



## Appro 1124H (1UTwinサーバ)



### Appro 1124H

ツイン (2x) 2ウェイ クアッドコア,  
4x ホットスワップ SATA HDD

- 1U 2P Shanghai 高性能サーバ
- 2つのマザーボードを1つの1Uシャーシに実装  
=> 1Uに2台のサーバを実装

#### マザーボードあたりの仕様:

- 2個のAMD Shanghaiプロセッサをサポート
- NVidia MCP55V Proチップセット
- 最大64GB DDR2 667/800MHzのメモリ (16 DIMMスロット)
- 最大2台の3.5" SATA ホットスワップドライブをサポート
- 1x LP PCIe Gen1 x16
- IPMI 2.0 リモート管理 (オプション)



## Appro 1224X (1UTwinサーバ)



### Appro 1224X

ツイン (2x) 2ウェイ クアッドコア/デュアルコア  
4x ホットスワップ SATA HDD

- 1U 2P Harpertown 高性能サーバ
- 2つのマザーボードを1つの1Uシャーシに実装  
=> 1Uに2台のサーバを実装

#### マザーボードあたりの仕様:

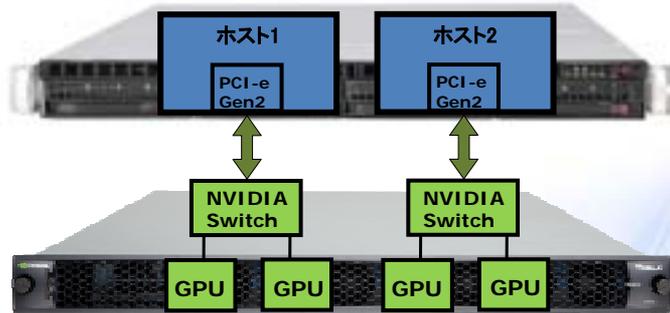
- 2個のIntel Xeon Wolfdale/Harpertownプロセッサをサポート
- Intel 5400 "Seaburg" チップセット - 1600MHz FSB
- 最大32GB FBDIMM 667/800MHz のメモリ (8 DIMMスロット)
- 最大2台の3.5" SATA ホットスワップドライブをサポート
- 1x LP PCIe Gen2 x16
- IPMI 2.0 リモート管理 (オプション)



## GPGPUコンピューティング ～APPRO-1224X～

Nvidia Tesla S1070のホストには  
1Uサイズのシャーシにホストを2つ実装したAPPRO-1224Xが最適！  
S1070の性能を1Uサーバで最大限活用できます。

### APPRO-1224X



### Nvidia Tesla S1070



## 大規模共有メモリコンピューティング ～APPRO-1143H～

16コア・128GB共有メモリを1Uサイズのシャーシで実現！  
→4ソケット x 4コア  
低価格・省スペースな大規模共有メモリ計算環境を提供。  
→1U



### Appro 1143H

4ウェイ クワッドコア,  
3x ホットスワップ SATA HDD

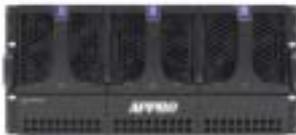
- 1U 4P Shanghai サーバ
- 4個のAMD Shanghai CPUサポート
- NVidia MCP55 Pro チップセット
- 最大128GB DDR2 667/800MHz メモリ(32 DIMM スロット)
- 最大3個のSATA ホットスワップドライブ
- 1x LP PCIe Gen1x16 拡張スロット
- 1000W PS
- IPMI 2.0 リモート管理(オプション)



# GreenBlade™

:: High-level Summary

Front View



Rear View



- Greenソリューション
  - 共有冷却ファンで、ノードあたりの消費電力を低減
  - 共有パワーサプライで、ノードあたりの消費電力を低減
- 信頼性向上
  - 冗長/ホットスワップ可能な共有冷却ファン(3+3 構成)
  - 冗長/ホットスワップ可能な共有パワーサプライ。(2+1または3+1 構成)
  - より大きい(120mm)ファン採用で、低消費電力・信頼性向上
- 保守容易
  - ホットスワップ可能な計算ノード
- より高密度
  - 標準19インチラックに、80台の計算ノードを搭載可能
- CPU
  - Shanghai, Nehalem
- 諸元
  - 5U、67.05kg(最大)、4.875kW(最大)



## 会社概要

:: HPCでのプレゼンス



	Vendors	Count	Share %	Rmax Sum (GF)	Rpeak Sum (GF)	Processor Sum	Avg. Efficiency
1	Hewlett-Packard	209	41.80%	4193870	7250162	707056	58%
2	IBM	186	37.20%	6511376	10059679	1498276	65%
3	Cray Inc.	22	4.40%	2486731	3226223	376309	77%
4	Dell	19	3.80%	569640	774686	82250	74%
5	SGI	17	3.40%	1218446	1479534	141560	82%
6	Sun Microsystems	7	1.40%	542762	724744	82936	75%
7	<b>Appro International</b>	<b>4</b>	<b>0.80%</b>	<b>158700</b>	<b>191536</b>	<b>30736</b>	<b>83%</b>
8	Fujitsu	3	0.60%	84140	100310	10240	84%
9	Bull SA	3	0.60%	115090	137523	19696	84%
10	Self-made	3	0.60%	52740	82556	8880	64%

Vendors by Performance – November 2008

**APPRO社のクラスシステムはTOP500にも浸透中**



# HPC 事例

## :: APPRO導入事例-1

 <p><b>2006. 9</b> <b>NOAA Cluster</b></p> <ul style="list-style-type: none"><li>• 1,424 Cores</li><li>• 2.8TB System Memory</li><li>• 15 TFlops</li></ul>	 <p><b>2006. 11</b> <b>LLNL Atlas Cluster</b></p> <ul style="list-style-type: none"><li>• 9,216 Cores</li><li>• 18.4TB System Memory</li><li>• 44 TFlops</li></ul>
 <p><b>2007. 6</b> <b>LLNL Minos Cluster</b></p> <ul style="list-style-type: none"><li>• 6,912 Cores</li><li>• 13.8TB System Memory</li><li>• 33 TFlops</li></ul>	 <p><b>2008. 2</b> <b>DE Shaw Research Cluster</b></p> <ul style="list-style-type: none"><li>• 4,608 Cores</li><li>• 9.2 TB System Memory</li><li>• 49 TFlops</li></ul>



# HPC 事例

## :: APPRO導入事例-2

 <p><b>2008. 4</b> <b>TLCC Cluster</b></p> <ul style="list-style-type: none"><li>• 72,462 Cores</li><li>• LLNL, LANL, SNL</li><li>• 638 TFlops</li></ul>	 <p><b>2008. 6</b> <b>Tsukuba University Cluster</b></p> <ul style="list-style-type: none"><li>• 10,784 Cores</li><li>• Quad-rail IB</li><li>• 95 TFlops</li></ul>
 <p><b>2008. 7</b> <b>Renault F1 CFD Cluster</b></p> <ul style="list-style-type: none"><li>• 4,000 Cores</li><li>• Dual-rail IB</li><li>• 38 TFlops</li></ul>	 <p><b>2008. 8</b> <b>LLNL Hera Cluster</b></p> <ul style="list-style-type: none"><li>• 13,824 Cores</li><li>• 120 TFlops</li></ul>



## SC08会場のAPPRO社のブース



新型Xtreme-Xを展示



APPRO HPC PRESENTATION

17

# *Appro Xtreme-X* スーパーコンピュータご紹介

APPRO INTERNATIONAL INC

---

## クラスタの課題

:: Xtreme-X によって克服

### ◆ クラスタはまだ利用や管理が難しい

- 電源、冷却、設置面積への対応が重要。
- RAS機能の充実が急務。
- ストレージやデータ管理がボトルネックになりつつある。
- インタコネクトの性能が重要。



**APPRO社 Xtreme-X で解決!**

Source IDC, 2008

APPRO HPC PRESENTATION

19



## Xtreme-X スーパーコンピュータ

:: 概要



- 完全に統合した、プリパッケージソリューション  
→それぞれのラックをスケラブルユニット(SU)として構成
  - TCOの削減
  - クラスタを容易に構成、増設可能
  - クラスタの信頼性、利用可能性、保守性の改善
- 1000TF以上へのスケラビリティ
- 業界標準製品の活用により、アプリケーションソフトウェアの移植が容易
  - Xeon、Opteron
  - インターコネクト
  - ネットワーク

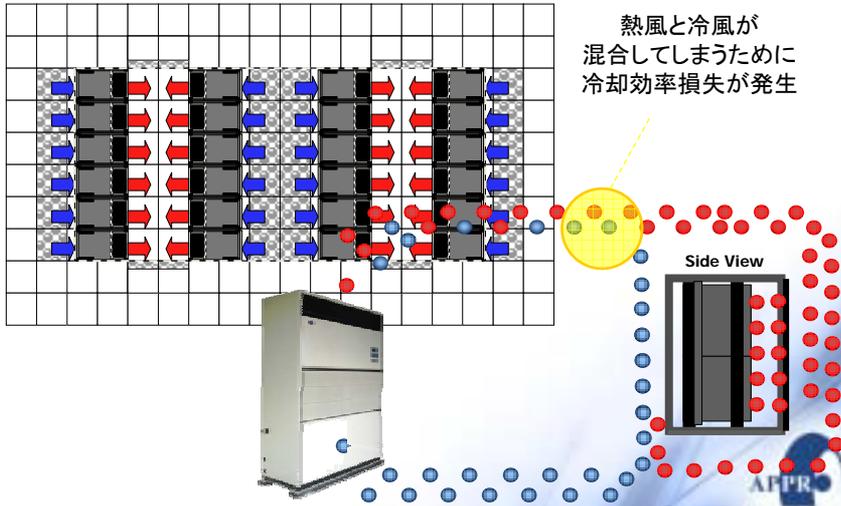
APPRO HPC PRESENTATION

20



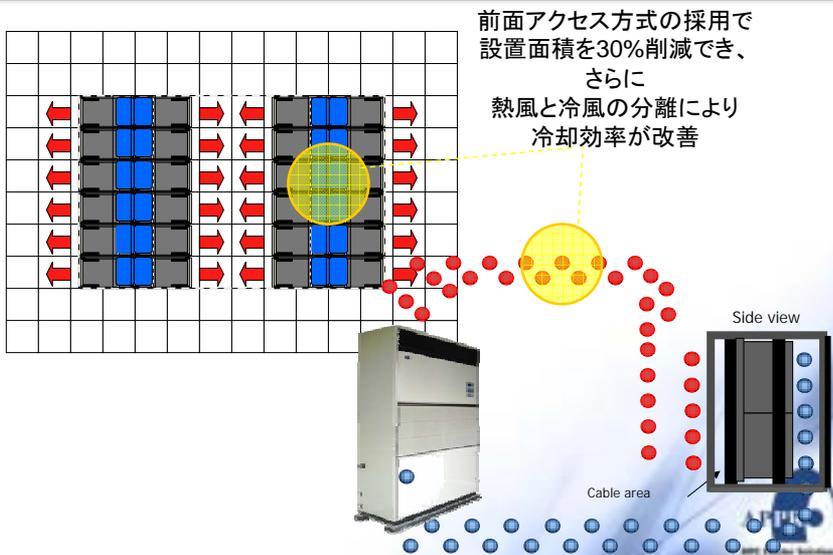
# 電力と冷却

:: 従来の冷却方法



# 電力と冷却

:: Xtreme-Xの新冷却方法 Directed Airflow™ Cooling



# クラスタの信頼性への対策

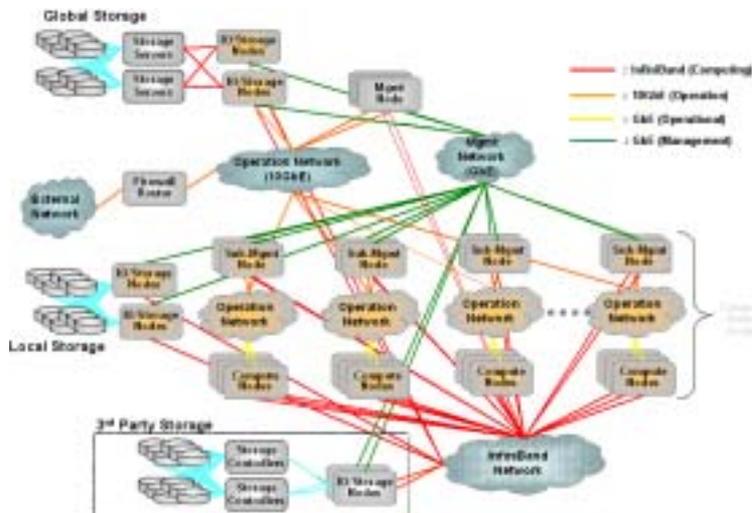
## :: Xtreme-Xの信頼性向上対策

- 不具合発生個所に対する信頼性向上の対策
  - ディスクドライブ
    - ディスクレスアーキテクチャ(計算ノード)
  - ファン
    - N+1 冗長構成
  - パワーサプライ
    - サブラックにはホットスワップ可能なN+1 パワーサプライを配備
  - ネットワーク
    - InfiniBandとイーサネットについて冗長構成採用のXtreme-X アーキテクチャ
  - サーバ
    - RAIDディスクアレイを管理ノードに採用のXtreme-Xアーキテクチャ
    - 冗長構成(アクティブ/スタンバイ)の管理ノード
    - 冗長構成(アクティブ/アクティブ)のサブ管理ノード
    - 計算ノードはホットスワップ可能、および、他のノードへの影響なしに簡単に取り外し可能



# Xtreme-X バランスのとれたソリューション

## :: 高性能かつ安定性を追求した設計



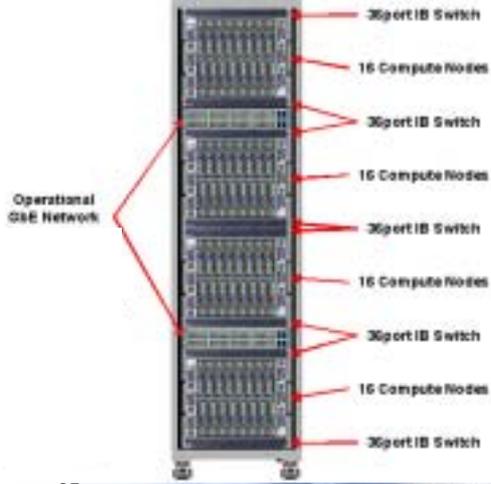
# Xtreme-Xスーパーコンピュータ

:: スケーラブルな設計

## ストレージユニット

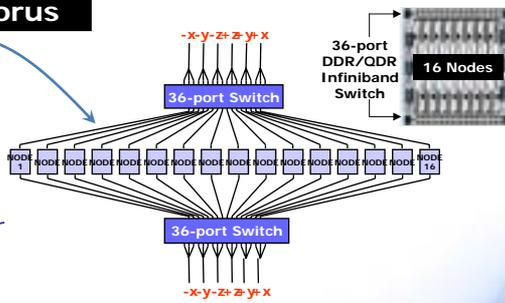
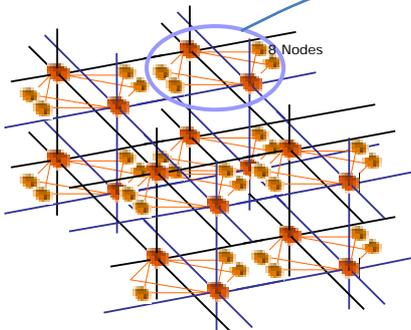


## 計算ユニット



# インタコネク トポロジ

## Dual-Rail DDR(QDR) 3D Torus



### IB インタコネクとサブラックの構成

- 2つの独立したInfiniBandサブネット
- MVAPICH2についてポートを統合可能
  - フェールオーバー及びロードバランス
- フォールトトレランスの提供
- サブラック(16ノード)に対して2倍のバンド幅を提供

### 3D Infiniband 接続イメージ図

- Dual-Rail DDR, 4X links
- スイッチ各軸間バンド幅:  $2 \times 12\text{GB/sec} = 24\text{GB/sec}$   
(QDRでも構成可能。その場合は、48GB/secになります。)



# ペタフロップスへの展開

:: スケーラブル クラスタ管理ソフトウェア ACE

3D トーラス ネットワークトポロジ



“Appro Cluster Engine™ software turns a cluster of Servers into a,  
functional, usable, reliable and available computing system”  
Jim Ballew, CTO Appro



# ACE (Appro Cluster Engine) 概要

:: スケーラブル クラスタ管理ソフトウェア

1. ネットワーク管理	<ul style="list-style-type: none"> <li>a. インターコネクットのハードウェア構成自動検出</li> <li>b. ネットワーク冗長構成</li> <li>c. ロードバランスとフェイルオーバ</li> <li>d. ネットワークステータスの管理システム統合</li> </ul>
2. サーバ管理	<ul style="list-style-type: none"> <li>a. ノードのハードウェア構成自動検出</li> <li>b. 遠隔サーバ管理 (パワーオン、パワーオフ、パワーサイクル)</li> <li>c. 遠隔サーバ初期化 (リセット、リブート、シャットダウン)</li> <li>d. 大規模クラスタでのスケーラブルな高速ディスクサポート</li> <li>e. サーバ冗長化、フェイルオーバー</li> <li>f. サーバステータスの管理システムへの統合</li> </ul>
3. クラスタ管理	<ul style="list-style-type: none"> <li>a. クラスタをパーティショニングし、複数の論理コンピュータとしての運用が可能</li> <li>b. 複数OSのサポート</li> <li>c. 論理コンピュータの管理、監視</li> <li>d. クラスタステータスの収集</li> </ul>
4. ノード管理	<ul style="list-style-type: none"> <li>a. ジョブスケジューリング及び管理</li> <li>b. ノードの監視、管理</li> <li>c. ノードステータスの管理システムへの統合</li> </ul>
5. ストレージ管理	<ul style="list-style-type: none"> <li>a. ディスクレスノードへのスケーラブルなrootファイルシステムのサポート</li> <li>b. 複数のグローバルストレージ構成サポート</li> <li>c. サーバステータスの管理システムへの統合</li> </ul>
6. システム管理	<ul style="list-style-type: none"> <li>a. システム構成の管理</li> <li>b. 管理サーバの冗長化及び自動フェイルオーバー</li> </ul>



## Appro Cluster Engine による管理

The screenshot displays the Appro Cluster Engine management console. At the top, there are tabs for 'Switches', 'Servers', and 'Clusters'. Below this is a table listing server details:

Name	Type	L1	L2	L3	N	Op	Rack	SubIP	Host	Net1	Net2	State
server-0001	management	1	1	1	1	1	1	1		●	●	● active
server-0002	management	1	1	1	2	1	1	2		●	●	● unknown
server-0003	group	1	1	1	3	1	1	3	tyngp-0003	●	●	● active
server-0004	group	1	1	1	4	1	1	4	tyngp-0002	●	●	● active
server-0005	compute	1	1	1	5	1	1	5	prod-9001	●	●	● active
server-0006	compute	1	1	1	6	1	1	6	prod-9002	●	●	● active
server-0007	compute	1	1	1	7	1	1	7	prod-9003	●	●	● fail on
server-0008	group	1	1	1	8	2	1	8	tyngp-0003	●	●	● active
server-0009	group	1	1	1	9	2	1	9	tyngp-0004	●	●	● active
server-0010	compute	1	1	1	10	3	1	10	prod-9004	●	●	● active
server-0011	compute	1	1	1	11	3	1	11	prod-9005	●	●	● active
server-0012	compute	1	1	1	12	3	1	12	prod-9006	●	●	● fail on

Below the table is a 'Console' window showing log messages:

```

ACI Log Console
2008-05-08 15:54:36.546 [INF] Host prod-0805 started on server-0011
2008-05-08 15:54:36.546 [INF] Host prod-0806 started on server-0012
    
```

## NEC様との欧州でのパートナーシップ

- EMEA(欧州、中東、アフリカ地域)において、APPRO社とNEC様は、NEC様のHPCソリューションのひとつに、APPRO社のXtreme-X及びACEを加えるという戦略的パートナーシップを発表しました。

The screenshot shows a press release on the APPRO website. The header includes the APPRO logo and the text 'High-Performance and Enterprise Computing'. The main content of the press release is as follows:

**Press Release**

**SUBJECT: 3/4/08 - APPRO and NEC Partner to Strengthen Supercomputer Sales**

Strategic Partnership to Capitalize on Growing Market for Large HPC Cluster Installations

Hayes, CA – 3/4/2008 – APPRO (http://www.appro.com), a leading provider of high-performance enterprise computing systems, and NEC Corporation, one of the world's leading providers of Internet, broadband network and enterprise business solutions, announce a strategic partnership today that will see APPRO Xtreme-X™ Supercomputer and Appro Cluster Engine™ Software Management added and branded as part of NEC's HPC solution offering.

The partnership will enable APPRO and NEC to work together towards a common goal, focus on reducing complexity of technology integration when deploying and managing integrated solutions, while lowering customer total cost of ownership (TCO). Commencing September 2008, APPRO Supercomputer products will be added to NEC's HPC offering as a first step in this partnership.

<http://www.appro.com/press/view.asp?Page=1&Num=160>



