

戦略的高性能計算システム開発に関するワークショップ
2010.08.02 (Mon) @キャッスルイン金沢

データセントリックスーパーコンピュータ ～データ・ユーザとの親和性の向上を目指して～

NEC システムプラットフォーム研究所

大野 善之

自己紹介

大野 善之（おおの よしゆき）

- NEC システムプラットフォーム研究所

- 1983年生まれ（若手）

- 研究分野

- 学生時代 ハイブリッドシステム(離散系・連続系)のモデリング言語
 - システム検証, ハイブリッドシステム, 制約プログラミング, モデル検査
- 入社後 スパコンのノード間ネットワーク
 - スパコン, 集団通信 (MPI)
- 現在(入社3年目) クラウド基盤向けの大量データの処理・分析システム
 - データセントリックコンピューティング, クラウド
 - 処理性能・省電力性の相反する2軸を評価軸におき研究
 - 明日からのSWoPP2010 で発表予定
大野, 小林, 菅: “データインテンシブコンピューティングの省電力化に向けたGPUノードの活用”, SWoPP2010金沢 (CPSY)

現在の研究テーマ: データセントリックコンピューティング

コンセプト:

“計算”中心のアーキテクチャ思考から“データ”中心のアーキテクチャ思考へ

近年の情報量増大の流れから、コンピュータシステムの性能ボトルネックの要因は、何らかのデータI/Oボトルネックが多くなるのではないか？

- ディスクI/O
- ノード間インターコネクトによるデータ転送
- 記憶媒体間のレイテンシ差

これらのボトルネックを回避するために、ボトルネックを生み出さないための様々な技術革新が行われてきた。

- インターコネクト帯域増加, トポロジの工夫, 分散ストレージ...etc.

データ中心で考えてみると、データの配置戦略・データの構造制御を従来よりインテリジェンスに実現できれば、ボトルネックの幾つかを回避できるのではないか？

時間の関係上、
詳しいお話は後ほど自由時間に

菅, 小林, 鳥居, 小川, 板橋, 宮田, 山川, 長谷部:

“スケーラビリティと高効率性を備えたクラウド基盤を実現するデータセントリック分散制御”,
DEIMForum 2010:第2回データ工学と情報マネジメントに関するフォーラム, pp. C2-2 (2010).

現在のスパコンと将来のスパコン

現在のスパコン

- とても高速な計算機で、とても高速に計算ができる
- 膨大な計算処理に対する高速な処理要求
- ベストパフォーマンスを要求するアプリケーション
 - 気象予測, 分子動力学, …

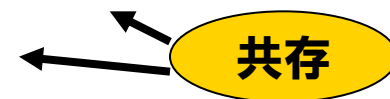
将来の“計算機”では、
計算だけではなくデータ・ユーザとの親和性も必要となるのでは

溢れ出てくる大量のデータに対する高速な処理要求

ベストパフォーマンスを要求するアプリケーション（既存ユーザ／アプリ）

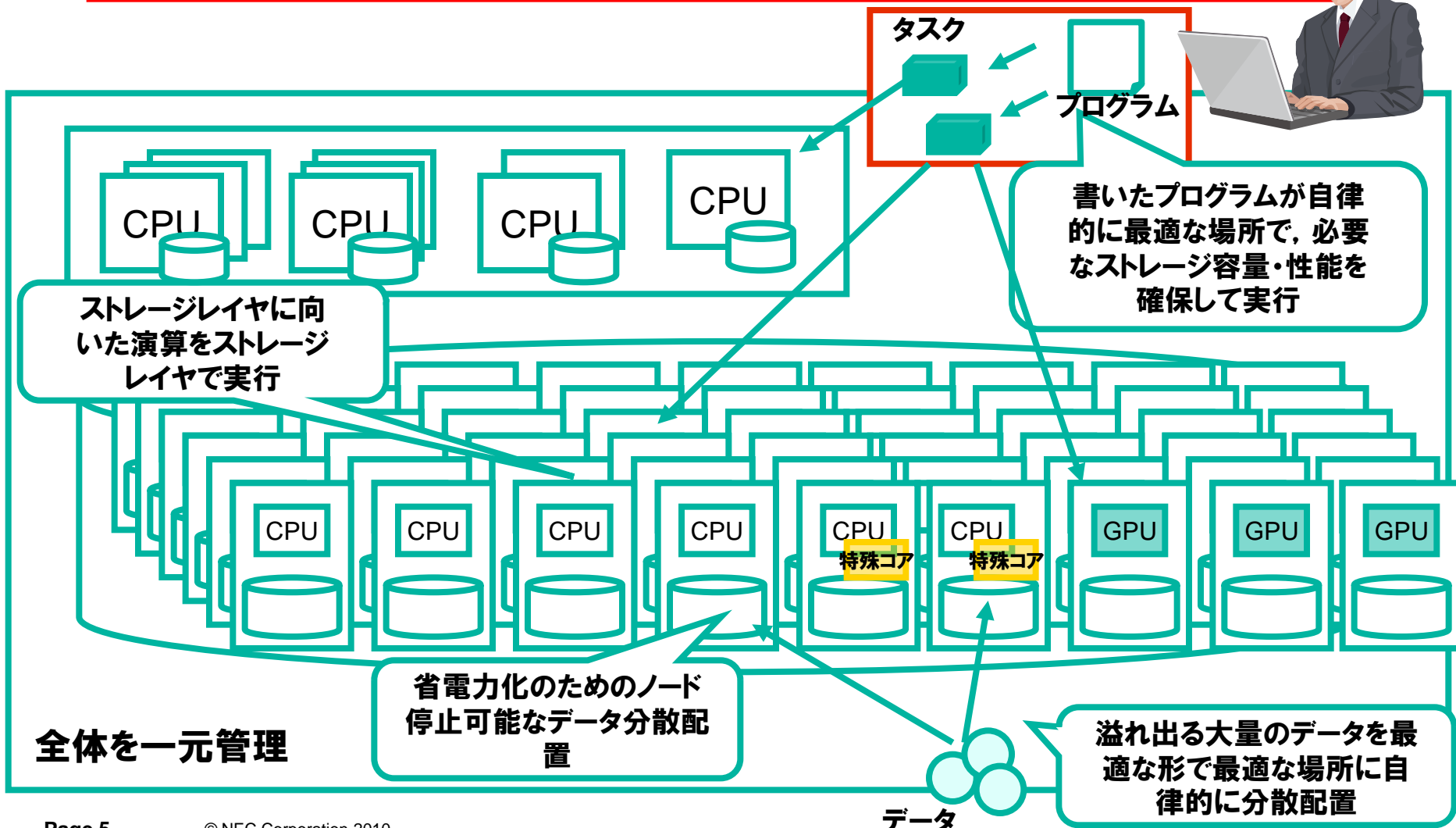
新たな計算ニーズの出現（未知のユーザ／アプリ）

- 簡単に記述したプログラムで十分高速に処理
- 大量のデータを十分高速に処理



データセントリックスーパーコンピュータ

計算だけではなくデータ・ユーザとの親和性も提供できる
データセントリックスーパーコンピュータ



書いたプログラムが自律的に最適な場所で、必要なストレージ容量・性能を確保して実行

ストレージレイヤに向けた演算をストレージレイヤで実行

省電力化のためのノード停止可能なデータ分散配置

溢れ出る大量のデータを最適な形で最適な場所に自動的に分散配置

全体を一元管理

データ

データセントリックスーパーコンピュータの実現のためには

プログラムを自律的に最適な環境で実行させるプラットフォーム

- ヘテロジニアスなリソースの有効利用
- データの配置/格納方式の自律制御

柔軟にデータ格納／データ利用できるプラットフォーム

他に必要となる要素技術

- ユーザに対する利便性を向上させるようなプログラミング言語
- 大量のデータを高速に分散配置／配置転換できるノード間ネットワーク
- 個々のノード(プロセッサ)の計算能力
 - ・ プロセッサ性能向上
 - ・ アプリケーション／ライブラリの高高速化

さまざまな分野の研究成果をうまく結びつけることで
データセントリックスーパーコンピュータを実現