アプリケーションベンチマーク
システム班・ベンチマーク

2012/9/19資料

東京工業大学
松岡聡

アプリfs・システム班としての初年度のマイルストーン

- 1. ベンチマークのデータ・メタデータの(データベース的)スキーマの定義・調査・標準化
- 2. それらの取得のための測定ソフトウェアツールのサーベイ・整備、並びに、並びにそれらによるデータの性能データの標準的な取得手続きの策定・ドキュメント化
- 3. ミニアプリの作成法の策定、およびいくつかのサンプルミニアプリの作成
- 4. 各FSの横断的なベンチマークチームの形成、特にメタデータ等定義の合意形成

HPCI一般利用による京の利用

- 採択(代表・松岡聡): 利用可能資源

	2012下期	2013上期	2013下期
京	53.3万ノード時間 (申請: 100万)	13.3万ノード時間 (申請: 25万)	5.3万ノード時間 (申請: 10万)
TSUBAME2.0	3万ノード時間	3万ノード時間	3万ノード時間
九州大 FX10	48ノード 占有	48ノード 占有	48ノード 占有

ベンチマークのメタデータ及びスキーマの標準化

- ベンチマークの結果の共通化の必要性
  - 単なる「数字」では意味がない⇒気象・天文・地質調査などのサーベイにおけるデータと同様に、意味の統一化が必要
  - 「当社比」では客観性がなく、意味がない
- メタデータの(データベース的)スキーマの標準化
  - メタデータ：結果や計測環境の共通的な意味(セマンティックス)の表記
  - 結果のデータベースへの統一的な格納
  - 結果の統一的な比較や利用を可能に
  - FS後、今後の調達や研究開発にも役立つ、更に国際貢献
- スキーマに応じたメタデータの標準的な取得法の手続きの策定・ドキュメント化
  - Scalasca, TAUなどの標準的な計測ツールを使用
  - ベンダーのカスタムツールも可能、ただしスキーマへ合わせるための作業はアプリFSとしては直接はサポートしない

メタデータのスキーマの例

- 例1: SPECベンチの投稿時データ
  - SPEC MPIから抜粋
  - システム名, 計測日, CPUスペック, ノード内CPU数, メモリ, ディスク, アクセラレータ,
    コンパイラ, コンパイルオプション、MPIライブラリ, ファイルシステム, ネットワーク(ベンダ モデル トポロジ バンド幅),
    チューニングの有無と内容, …, ベンチマーク結果
  - 無論これらでは十分とは言えない。
- 例2: 京の大規模実行時の審査基準
  - 単体性能(ピーク比)、並列化効率（これらはデータ）
  - 一定の並列化効率がないと大規模実行させてもらえなかった⇒統一的な指標

スキーマの標準化に向けて

- 目標
  - 収集すべき情報が何なのかを決定
  - これらの情報を半自動で(=マニュアルの通りにやれば)取れるような手続きの策定・標準ツールを利用
  - 結果を収集して、比較できるリポジトリを作る
- FSで一過性な話ではない
  - FS終了後も我が国のHPC分野が常用できる手法や環境を構築するのが目標
  - 今後の研究開発、論文や調達に利用可能な手法
  - 日米などの国際協力、標準化へ
- Issues
  - 海外や国内のベンダはどのようなデータを集めているのか情報収集
  - すでに標準があれば、それを検討・採用・共同研究等（ただし、あまりない）
  - 難しいデータの取得：電力、Scaling等をどう取得・表現するか等検討
  - どのような性能データ・メタデータをそれぞれの分野で欲しいか、という各方面の需要を容易に反映するにはどうするか

海外におけるベンチマークの結果のメタデータスキーマの事例

- 1. OTF : Open Trace Format that works with a number of tools: http://www.tu-dresden.de/die_tu_dresden/zentrale_einrichtungen/zih/forschung/software_werkzeuge_zur_unterstuetzung_von_programmierung_und_optimierung/otf, including Tau, Scalasca

- 2. Tau: a comprehensive performance tool that includes a performance database(not standard, but commonly used): http://www.paratools.com/TAU

- 3. PAPI: a standard tool that serves as an interface to most hardware counters: http://icl.cs.utk.edu/papi/

計測環境の整備状況

- ScalascaとTAUによる解析をテスト中
  - 特にInstrumentを入れずとも、以下のパラメータは自動的に収集可能

	Scalasca	TAU
実行時間	○ (関数/OpenMPブロック)	○
HWカウンタ (キャッシュミスetc)	○ (PAPIを利用)	○? (PAPIを利用)
通信: ノードごとのIn/Out	○	○
通信トポロジ	×?	○ (Communication Matrix)
I/O	?	?
CUDA	×	○
OpenACC	×	×
計測環境(システム・並列度等)	?	○

  - Instrumentを入れて、Human Readableになるのか
    - 例: C++のTemplate Programming(Thrust), Fortranのループ

そのほかのツール候補と取りうる情報

- Valgrind
  - Working set size, アクセスパターン, …
  - メタデータとしてどう表現するか
- NVIDIA GPU Profiler
  - GPU以外の環境でも同意味のデータはとれるか
- Cray社のプロファイリングツール (XK6/7・東工大に8ノード、年度末40ノード)
  - 何が取れるのか調査中、OpenACC対応？
- Vampire
  - TAUやScalascaで取れない情報(VampirTrace)はあるか調査中
- ASPEN
  - 性能とアーキテクチャのモデリングを可能にするDomain Specific Language (SC12にて発表)
- 別途測り方を考えなければいけないもの?
  - Arithmetic Intensity (= Bytes/Flops)
    - 演算数を手で数えるのは無理
  - I/O: POSIX IO, HDF5, MPI-IO
    - どのデータがどこに書かれたかを把握したい
  - 電力関連

他FSチームとの連携

- システムFSチームと、アプリFSチームの間でベンチマークの実行環境・測定項目・結果の収集方法を共通化
  - ベンチマーク結果には客観性が必要・「当社比」の排除
  - アプリFSチームと同じスキーマおよび取得法
  - 各FSチームからベンチマーク共通化窓口となる人を選出：下記候補(敬称略)
    - 東大FS　片桐
    - 東北大FS 江川・板倉(JAMSTEC)
    - 筑波FS　高橋(大)
  - FS全体として合意できる統一スキーマや手法を制定
  - SC前に一度システム班との横断ミーティングを開催(10月、日程調整必要)