マルチノードマルチコア向け分散共有メモリにおけるデータ分散配置機能稼働実験

白澤 卓磨　緑川 博子（成蹊大）

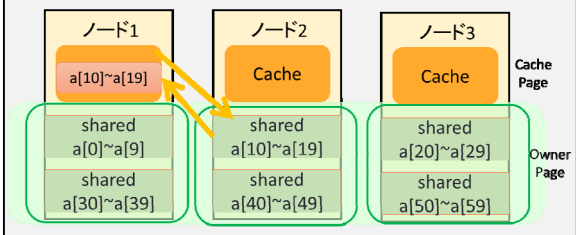
マルチスレッド・マルチノードプログラミングのための分散共有メモリシステム（M-SMS）

ノード内並列（OpenMP/pthreadマルチスレッドプログラム）＋ノード間並列（クラスタにおけるPGAS・共有メモリプログラミング）を提供する新しいページベースの分散共有メモリシステム[1]

- 従来のMPI＋OpenMPハイブリッドプログラミング：MPIではデータをローカルビューで扱う → 並列プログラムの生産性が低くなる
- 共有するデータを全ノードが認識するグローバルビューの提供
- 共有データの各ノードへの分散配置機能（分散マッピング）[2] → 応用プログラムのデータアクセス局所性を利用した性能向上

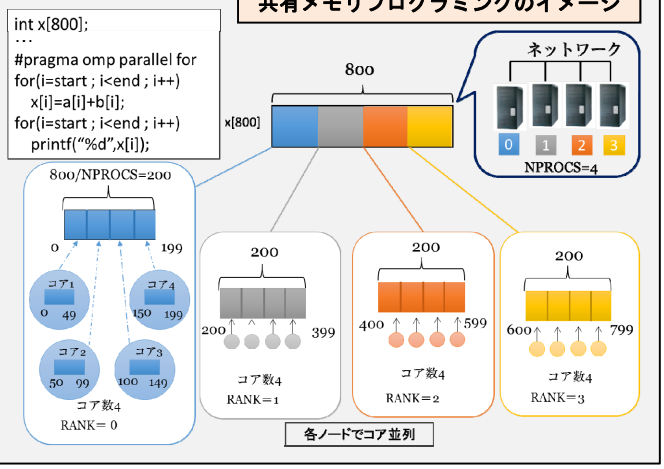
ページベース分散共有メモリ

共有データの仮想共有メモリシステムへの分散配置



ノード間で共有するデータは各ノードに分散して管理
Owner: 各ノードが管理する共有データ
Cache: 他ノードのOwnerページのコピー領域

マルチノード・マルチスレッド 共有メモリプログラミングのイメージ



sms_mapalloc関数

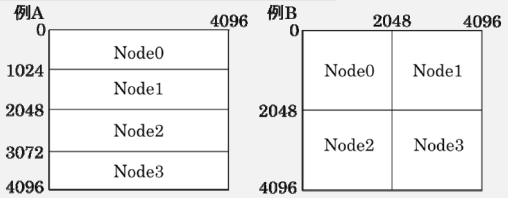
```
void* sms_mapalloc(  int dim[],  int div[],
size_t data_size,  int start_node,  int num_node  );
```

dim[]・・・各次元の要素数，div[]・・・各次元の分割数
data_size・・・型の大きさ，start_node・・・始点ノードランク
num_node・・・利用ノード数

```
int num=4096,dim[3]={num,num,-1},div[3]={4,1,-1};
int(*array)[num];
・・・
array=(int (*)[num])sms_mapalloc(dim,div,sizeof(int),0,4);
・・・
#pragma omp parallel for private(j)
 for( i=sms_rank*(num/4); i<(sms_rank+1)*(num/4); i++ ){
  for(j=0;j<num;j++){
   array[i][j]=i+j;
  }
 }
```

変換後

分散マッピング例（二次元配列）

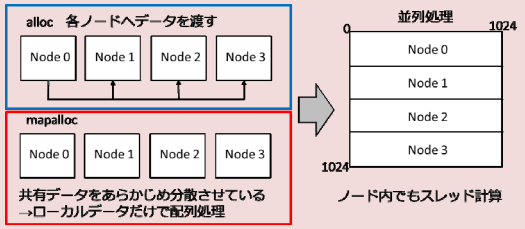


分散データマッピングAPI（MpC）[2]による記述（組み込み予定）

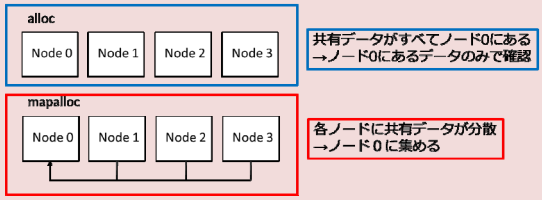
```
shared int arrayA[4096][4096]::[4][1]; //例A
shared int arrayB[4096][4096]::[2][2]; //例B
```

稼働実験

配列処理：各ノードで共有データ配列への代入を行う

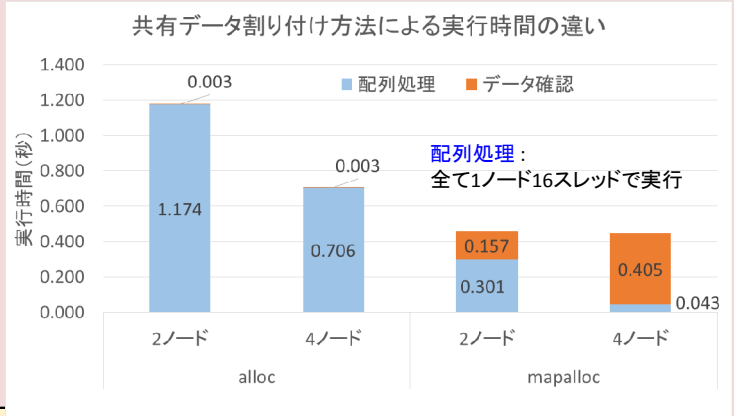


データ確認：共有データを1ノードに集めて実行



ベンチマーク
共有データ配列に対する代入処理
→ノード1つに共有データを集めデータが正しいか確認

共有データ（int 1024x1204）割付方法
sms_alloc関数 → ノード0のみ
sms_mapalloc関数 → 例Aのように割付



CPU	Intel Xeon E5-2687W 3.10GHz, 2CPU x 8Core / node
Memory	128GB/node (2node) or 256GB/node (4node)
Network	Infiniband singleFDR(56Gbps)
OS	CentOS 7.1.1503
Compiler	gcc version 4.8.3

今後の展開

- システムの安定稼働（正しい共有データの反映、システムが止まってしまう原因の調査）
- プログラムインターフェース、API：　既存開発のSMSシステム、MpC[2]、SMS[3]のAPIを踏襲
- 共有メモリー貫性とは別経路による遠隔メモリ・ローカルメモリ間直接コピー操作Get/Putの実装
- メモリー貫性実装方式の検討（現在はall-invalidate方式）
- アクセス局所性を考慮した応用アルゴリズムによる性能評価

[1] 緑川,岩井田："マルチスレッド対応型分散ン共有メモリシステムの設計と実装", HPCS2015, HPCS2015論文集,（2015,5-19）
[2] MpCインターフェース: 緑川他"メタプロセスモデルに基づくポータブルな並列プログラミングインターフェースMpC",情処論文誌：Vol.46 No.SIG4(ACS9), pp.69-85,（2005,3）
[3] 緑川，飯塚："ユーザレベル・ソフトウエア分散共有メモリSMSの設計と実装", 情処論文誌,Vol.42, No.SIG9(HPS 3), pp.170-190 （2001,8）