

# ヘテロジニアスな分散処理システム における空間ビッグデータ処理の 高速化技術の研究開発

(162205001)

---

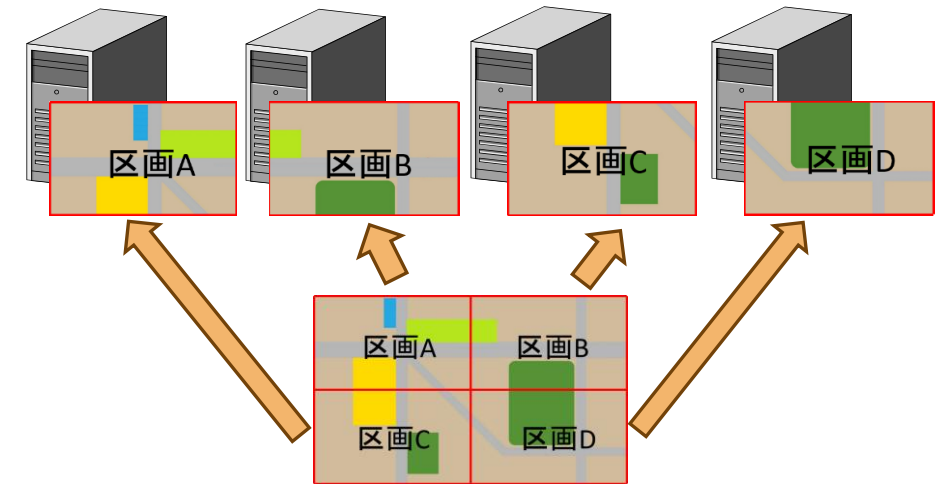
研究代表者: 荒木 光一 (五大開発株式会社 システム事業部 技術研究所)

研究分担者: 新保 泰輝 (石川工業高等専門学校 環境都市工学科)

# 研究開発の背景・目的

## 背景

- 空間ビッグデータはバザードマップ作成や地形解析などで利活用
- 空間ビッグデータの高精度化や増加，解析手法の複雑化により空間ビッグデータの処理時間は長期化しているが，空間ビッグデータ解析の特性を考慮することで分散処理で高速化可能
- ヘテロジニアスな分散処理システム
  - ◆ 社内の遊休PCの活用，クラウドとの併用，マルチクラウド
- 分散処理は専門的な知識が要求されるため，空間ビッグデータ解析者にとってハードルが高い
  - ◆ ヘテロジニアスな分散処理システムにおける負荷分散
  - ◆ 分散処理を行う空間ビッグデータ解析プログラムの開発



## 目的

1. ヘテロジニアスな分散処理システムで空間ビッグデータ解析者に対して迅速に空間ビッグデータ処理の結果を提供
2. 並列プログラミングを気にせずに空間ビッグデータ解析のプログラムを開発

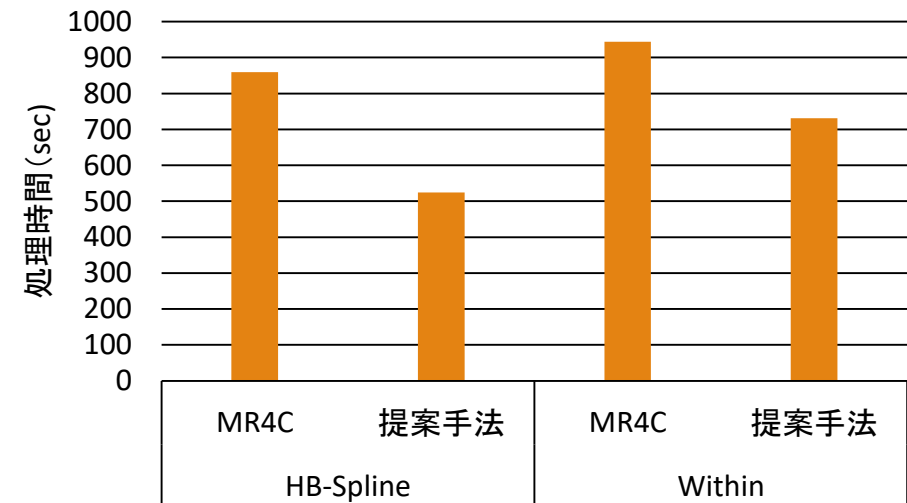
# ヘテロジニアスな分散処理システムにおける 空間ビッグデータの高速化

- 各ノードの処理時間を平準化
- ベクターデータが処理対象
- 分散処理にはMPIを利用
- 各ノード毎に処理時間予測モデルを構築
- 処理時間予測モデルを繰り返し更新することで、処理時間の予測精度を向上させ、平準化を促進

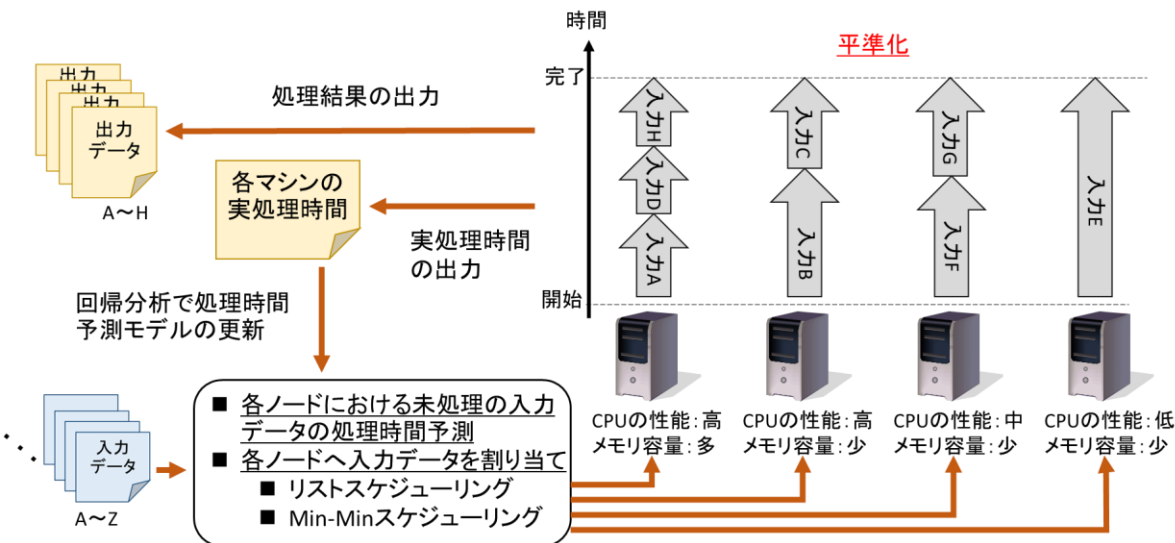
- MR4Cと比較して、提案手法はHB-Splineで1.6倍、Withinで1.3倍高速化
- MR4C: Hadoopをバックエンドとして利用した空間ビッグ

データ処理向けフレームワーク

(google MR4C : <https://github.com/google/mr4c>)



- HB-Spline: 数値標高モデルの曲面推定で標高不連続部分を滑らかに補間
- Within: オブジェクトに他のオブジェクトが包含されているか否かを判断
- 5ノードのヘテロジニアスな分散処理システム(五大開発2台, 石川高専1台, Amazon EC2 2台)



# 並列プログラミングを隠蔽するライブラリ

- MPIプログラミングを隠蔽するC++ライブラリを開発
- MPIプログラミングを隠蔽したメソッドと、属性値の取得などベクターデータを扱う上で利便性を高めるメソッドを定義
- 本ライブラリを利用することでMPIを考慮せずに空間ビッグデータ解析のプログラムが開発可能
- 従来のMPIプログラミングとMR4Cと比較して、コードステップ数を削減

```
1:#include "Mfgv.h"
2:static CMfgv Mfgv;
3:
4:static int UserApp(int argc, char **argv)
5:{
6: for (int key = 0; key < Mfgv.GetNumFile(); ++key) {
7:
8:     std::vector<std::vector<std::string> > Attribute;
9:     std::vector<std::string> Coordinate, Header;
10:    std::stringstream OutputData;
11:    std::string OutFileName
12:
13:    std::string InFileName = Mfgv.GetFileName(key);
14:    Mfgv.GetFileInfo(InFileName, Coordinate, Attribute, Header);
15:
16:    // Write vector data processing.
17:    .
18:    .
19:    .
20:    Mfgv.WriteFile( strOutFileName, OutputData.str(), True );
21: }
22: return 0;
23:}
```

ベンチマーク	実装手法	ステップ数
HB-Spline	MPI	906
	MR4C	892
	提案ライブラリ	831
Within	MPI	151
	MR4C	148
	提案ライブラリ	52

・コメントと空白ラインは含まれていない

# 今後の研究開発成果の展開 及び波及効果創出への取り組み

- 五大開発において社内利用と製品化
- 空間地理情報の専門家が参加するOSGeo.JPのイベントにて本成果を紹介
- 処理時間が長いため躊躇・諦めていた空間ビッグデータ解析を試みる事が可能
  - 多角的な視点で解析・考察
  - 地形解析や空間解析などの研究分野の発展を促進
  - より有用な情報を社会に提供
- 民間企業における業務の短縮化・効率化
  - ex1) 土木業界ではUAV測量で得た数十GBオーダーのデータを扱う業務
  - ex2) 空間ビッグデータを扱うが、守秘義務によりクラウドを利用できない業務
    - 中小企業では社内の遊休PCで構築したヘテロジニアスな分散処理システムを利用