

# 広域分散ファイルシステムに基づく「ビッグテーブル」型の 超大規模データ処理系の構築と機能および性能評価



## JHPCN-HPCIシステム (先端ソフトウェア運用基盤) における Hadoop-Gfarmプラグインを用いたHadoop分散ファイルシステムの構築性の検証

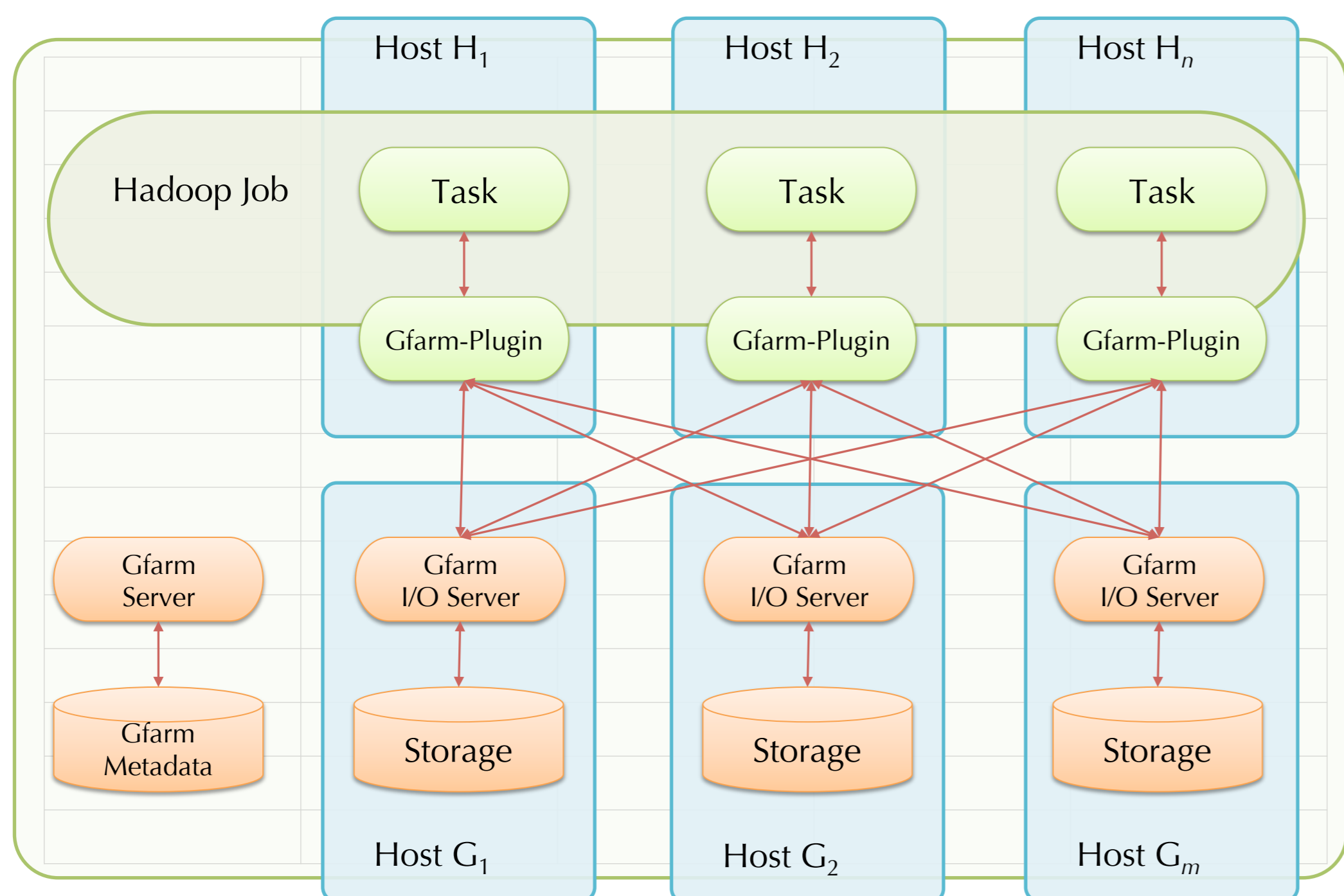
狙い

- ① HPCI先端ソフトウェア運用基盤上にHadoop/HBase環境を構築
- ② Hadoop-Gfarmプラグインを導入し、HPCI共用ストレージ (Gfarm) へデータを蓄積
- ③ HPCI認証基盤・ネットワーク基盤との連携性・相互運用性の検証

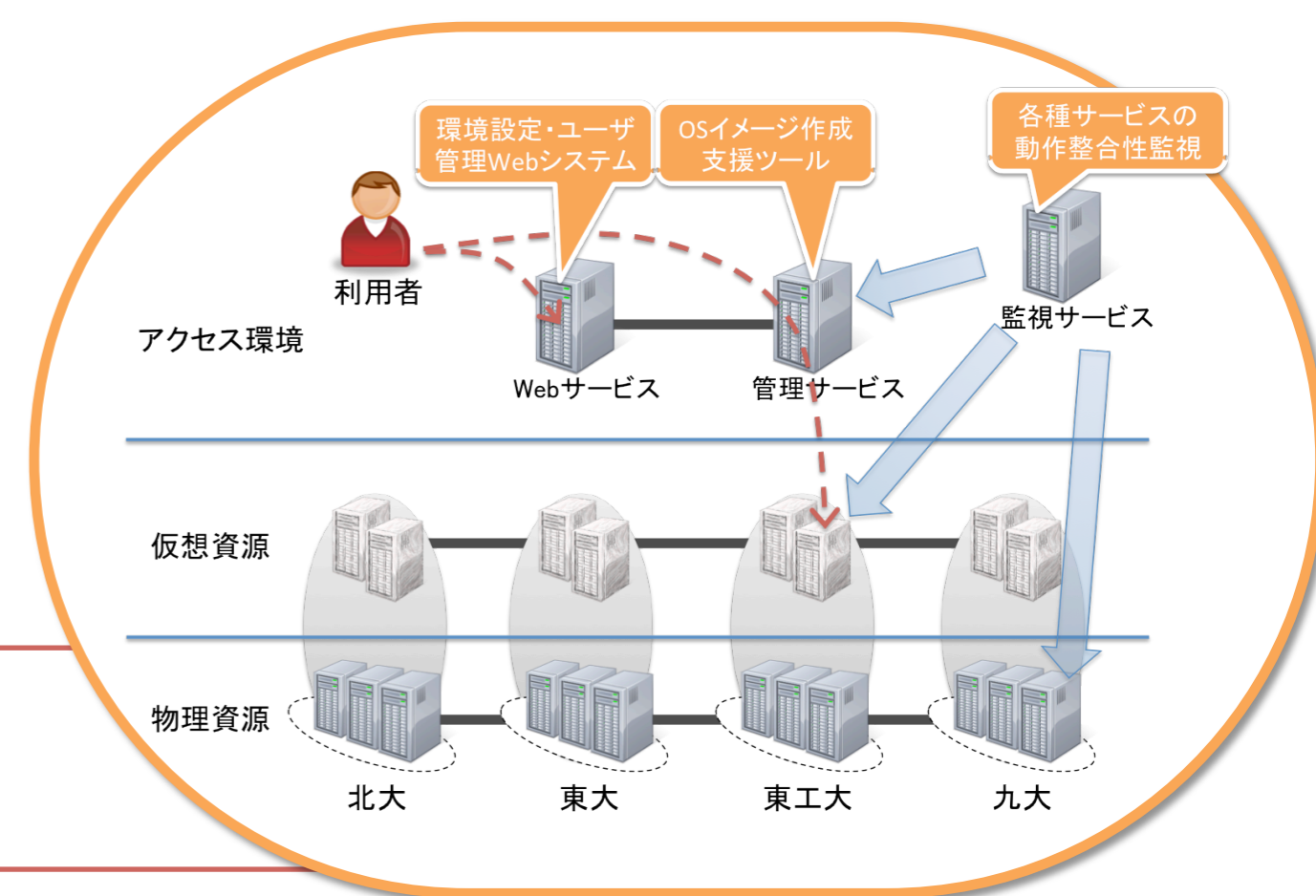
メンバー：滝澤 真一郎 (理研AICS)、棟朝 雅晴 (北大)、實本 英之 (東大)、小林 泰三 (九大)、合田 憲人 (NII)

### Hadoop-GfarmプラグインによるHPCIシステム上のビッグデータ・ビッグテーブル処理系

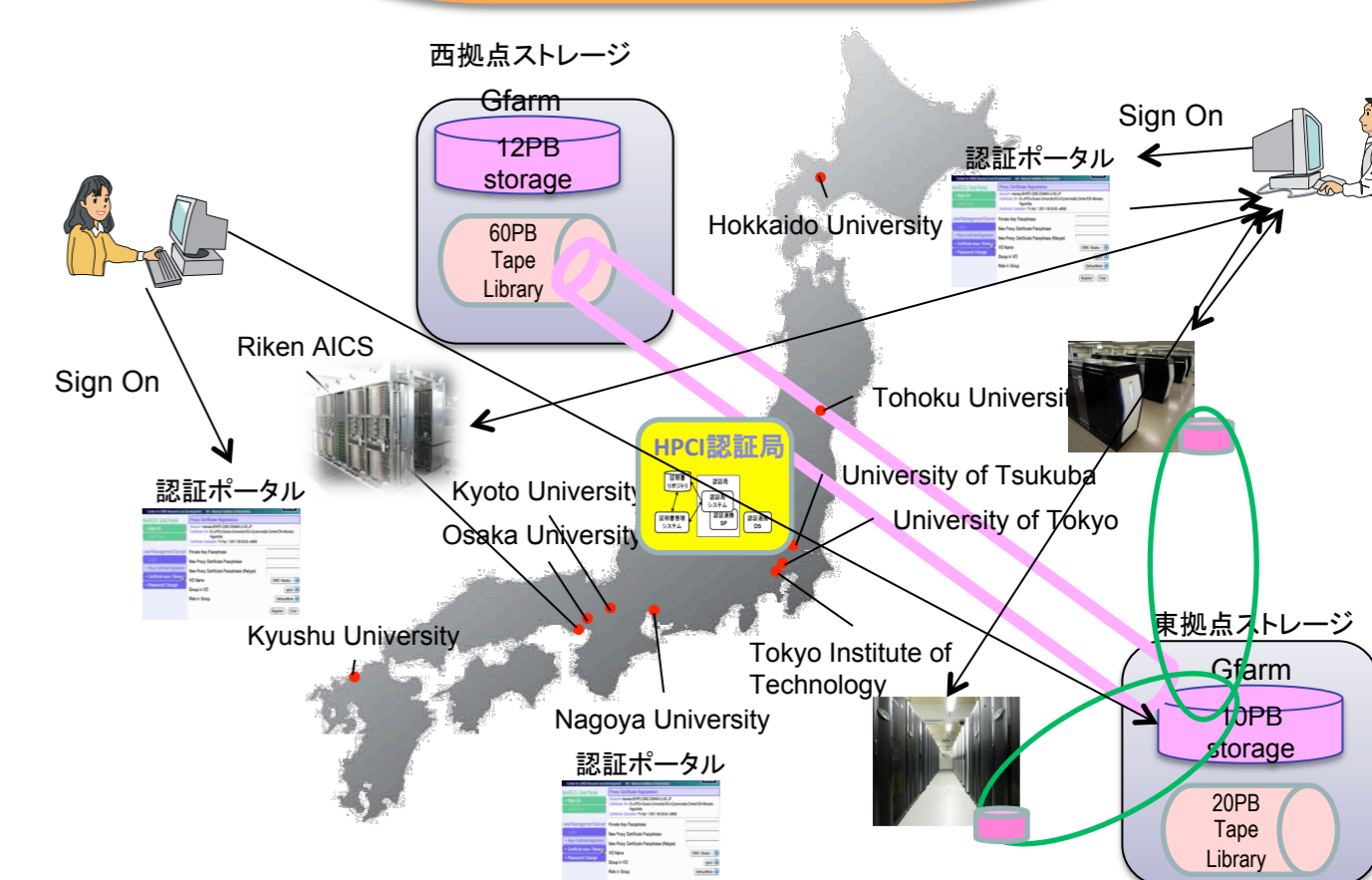
- ☑ 筑波大によるHadoop-Gfarmプラグインを拡張したHadoop/HBase-Gfarmプラグインの開発
- ☐ HPCI先端ソフトウェア運用基盤上にHadoop/Hbase環境を構築



Big Table by Hadoop/HBase on Gfarm

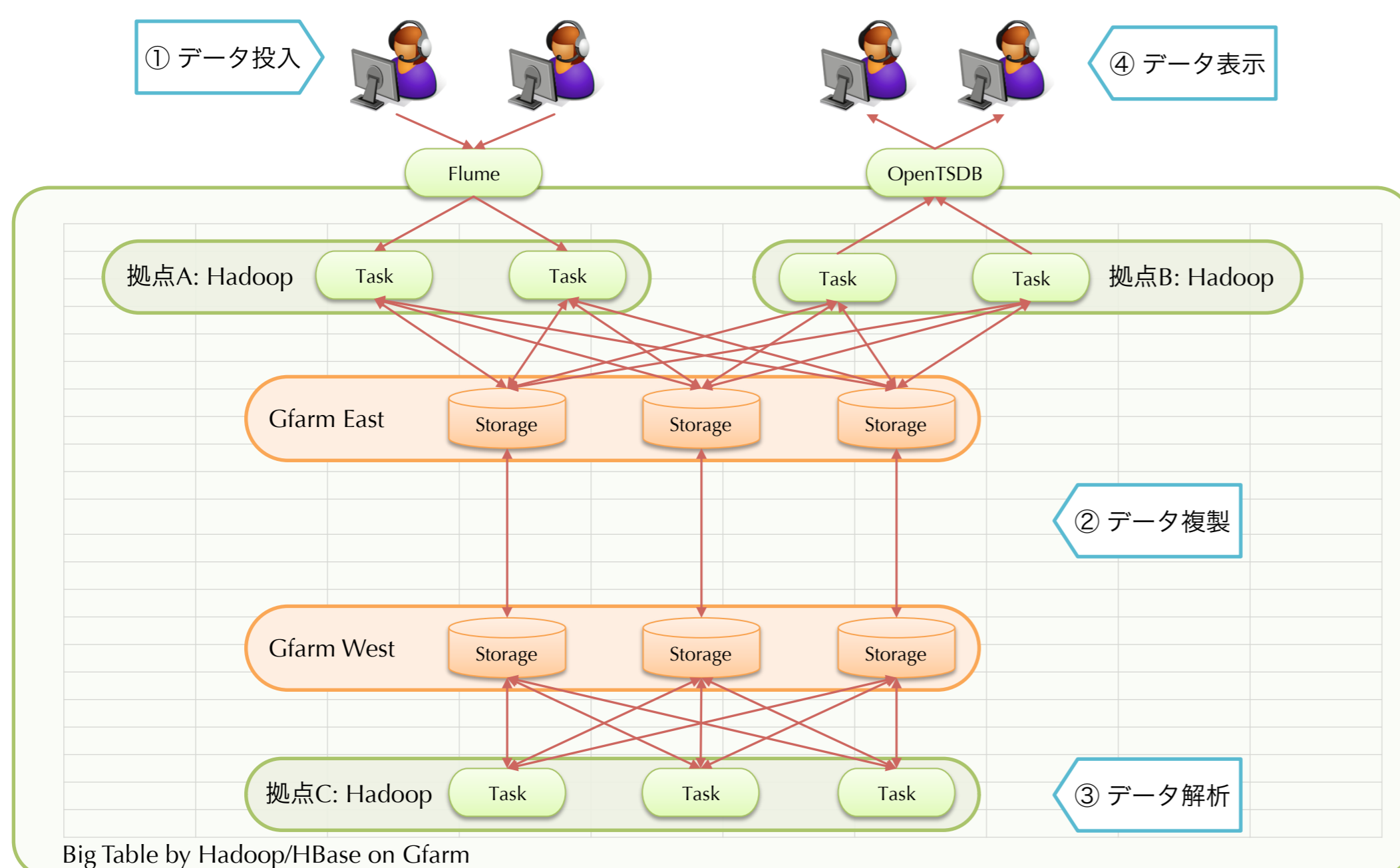


- 先端ソフトウェア運用基盤
- 認証基盤
- ネットワーク
- 共用ストレージ



HPCI共通運用システム

データ投入・複製・解析・表示という実運用に即したシナリオ遂行を通じて、  
地理的に離れた多拠点間での運用性を検証



Big Table by Hadoop/HBase on Gfarm

### 超大規模数値計算システムと連携可能な 大規模データ処理システムとしての有用性を検証

- (1) 北海道大学、東京大学、東京工業大学、九州大学で整備されているRENKEI-VPEによるVMホスティング環境に広域分散ファイルシステムであるGfarmをHadoop分散ファイルシステムとして構成したHadoop/HBase環境を導入し、その際の導入・構築性の検証とそれぞれの拠点内での性能検証を行う。
- (2) さらに、Gfarmのファイル複製機能を活用し、拠点間でのデータ連携性や移行性を検証する。これらの検証結果を基に機能・性能改善を行った上で、HPCI共通運用システム整備において開発を進めているHPCIアカウント集計システムを導入し、データの投入・解析・表示機能を多拠点に分散させた際の運用性の検証を行う。