

Japan Lattice Data Grid: 計算素粒子物理分野におけるデータ共有

大野浩史

筑波大学計算科学研究センター

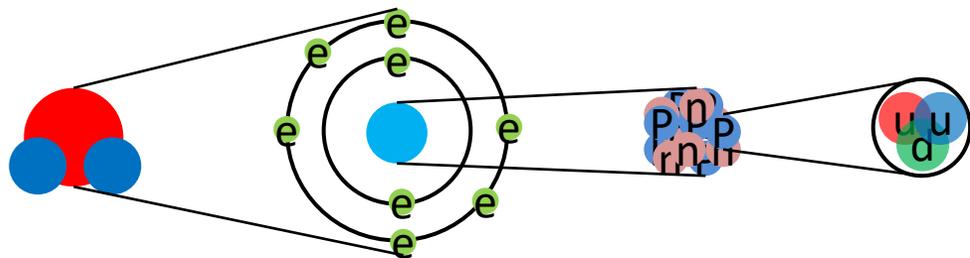
Gfarmシンポジウム2024
富士ソフトアキバプラザ 6F セミナールーム3
2024年9月6日

目次

- 自己紹介
- 素粒子物理学についての簡単な紹介
 - 素粒子標準模型
 - 格子QCD
- Japan Lattice Data Grid (JLDG) と Gfarm の利用
- まとめと今後の課題

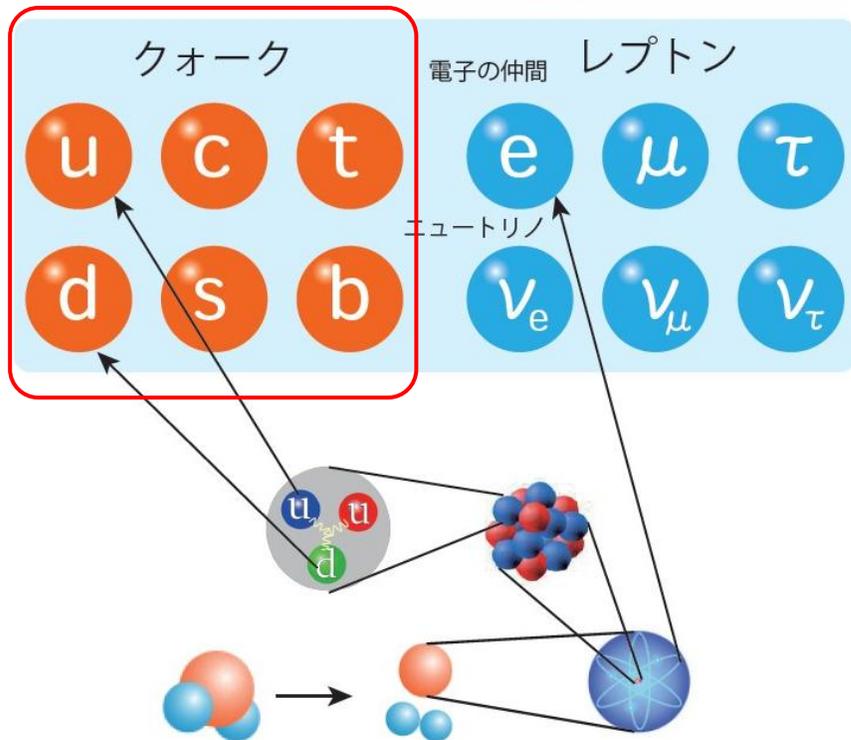
自己紹介

- 氏名：大野 浩史
- 所属：筑波大学計算科学研究センター
- 専門：素粒子理論、特に格子QCDの数値的研究
最近は機械学習の応用にも興味あり
- 活動：Japan Lattice Data Grid (JLDG) 管理者グループ
– 2022年より代表



素粒子物理学の標準模型

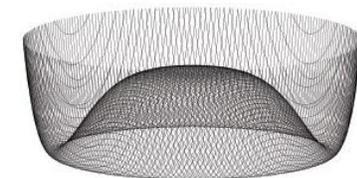
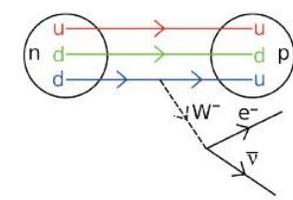
物質を構成する素粒子



力を伝える素粒子



質量を与える素粒子

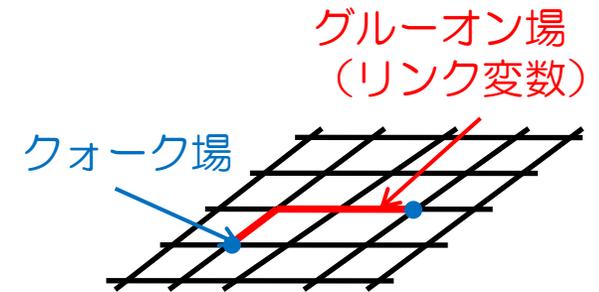


強い力

- 自然界に存在する基本的な力のひとつ
- グルーオンにより媒介され、クォーク・グルーオンの間に働く → 核子 (陽子や中性子) の仲間・中間子を作る
- 核力のもとになっている、核子の質量を作っている

格子QCD

- QCD = Quantum Chromodynamics、量子色力学
 - 強い力 (クォーク・グルーオンの間の相互作用) に関する基礎理論
 - 相互作用が強いので摂動論 (解析計算) が一般に困難 → 数値計算
- 格子QCD
 - 格子状に離散化した時空間上で定義したQCD、数値計算が可能
 - 物理量の計算 → 超高次元の多重積分 → モンテカルロ積分
 - 計算のボトルネック → 大規模疎行列の連立一次方程式を大量に解く
 - **大量の基礎データ (リンク変数の配位) の保存・共有が必要**



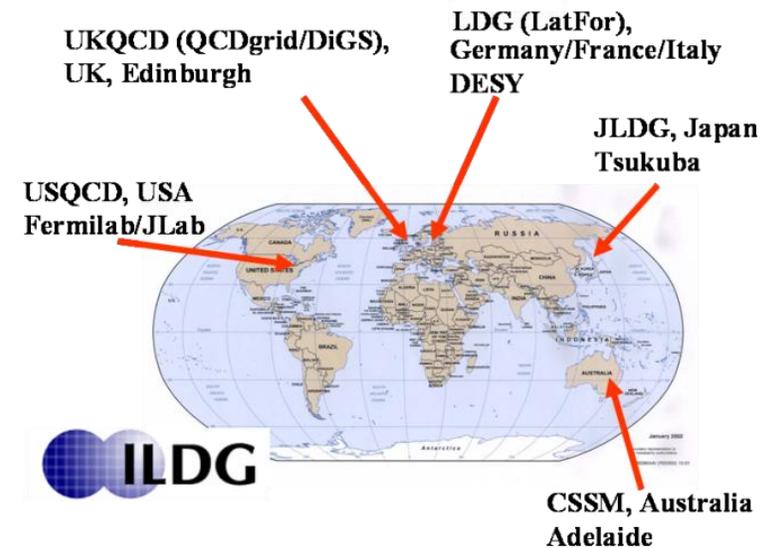
典型的なデータサイズ: $100^4 \times 36 \times 1000 \times 16 \text{ byte} \cong 52 \text{ TiB}$
(格子サイズ) (内部自由度) (配位数) (倍精度複素数)

格子QCDの応用例：

ハドロン質量の計算、クォークの閉じ込め、カイラル対称性の自発的破れ、新粒子の予言、有限温度・密度相転移、核力の計算、標準模型を超えた物理の探索

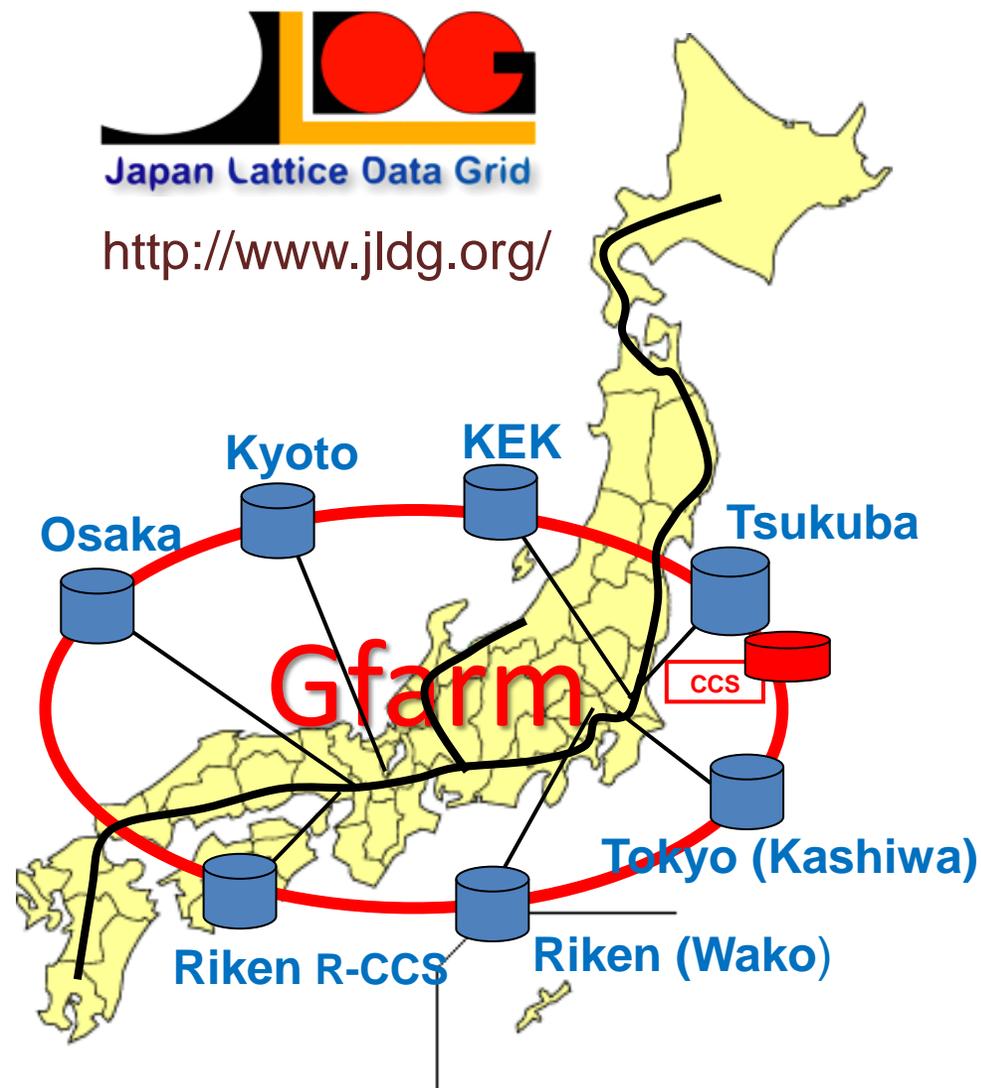
計算素粒子物理分野におけるデータ共有の国際的取り組み

- International Lattice Data Grid (ILDG)
 - 格子QCDの基礎データのアーカイブ・共有のための国際的取り組み
 - 2002年始動
 - 参加国:
 - 日本、アメリカ、イギリス、ドイツ、イタリア、フランス、オーストラリア等
 - ILDG board、Middleware/Metadata working groups
 - 共通のファイル・メタデータフォーマット
 - メタデータカタログ
 - マークアップ
 - データアクセスのためのインターフェース

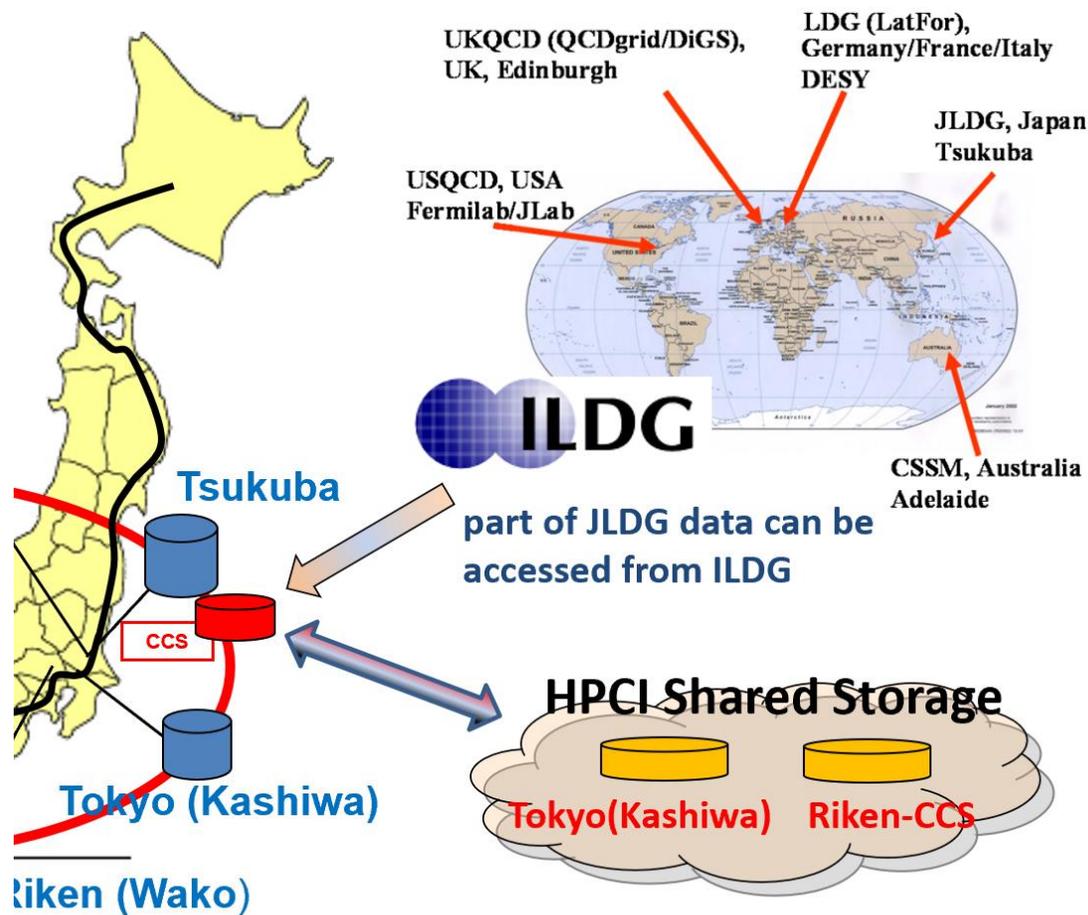


日本での取り組み: Japan Lattice Data Grid (JLDG)

- 2008/5/23 運用開始
- 全国7拠点
 - KEK、筑波大、理研和光、東大、京大、阪大、理研神戸
(かつては名古屋大、金沢大、広大も)
- NII が運用する SINET6 上の VPN (HEPnet-J/sc) 上に構築
- Gfarm グリッドファイルシステム
 - 単一ファイルシステムとして利用可
→ ユーザは保存先を意識しなくてよい
- 利用状況 (2024/8月末 現在)
 - 全ストレージ容量: 21PiB
 - 使用量: 18.3 PiB (87%)
 - 総ファイル数: 2億超
- 素粒子だけでなく宇宙分野でも利用



他システムとの連携



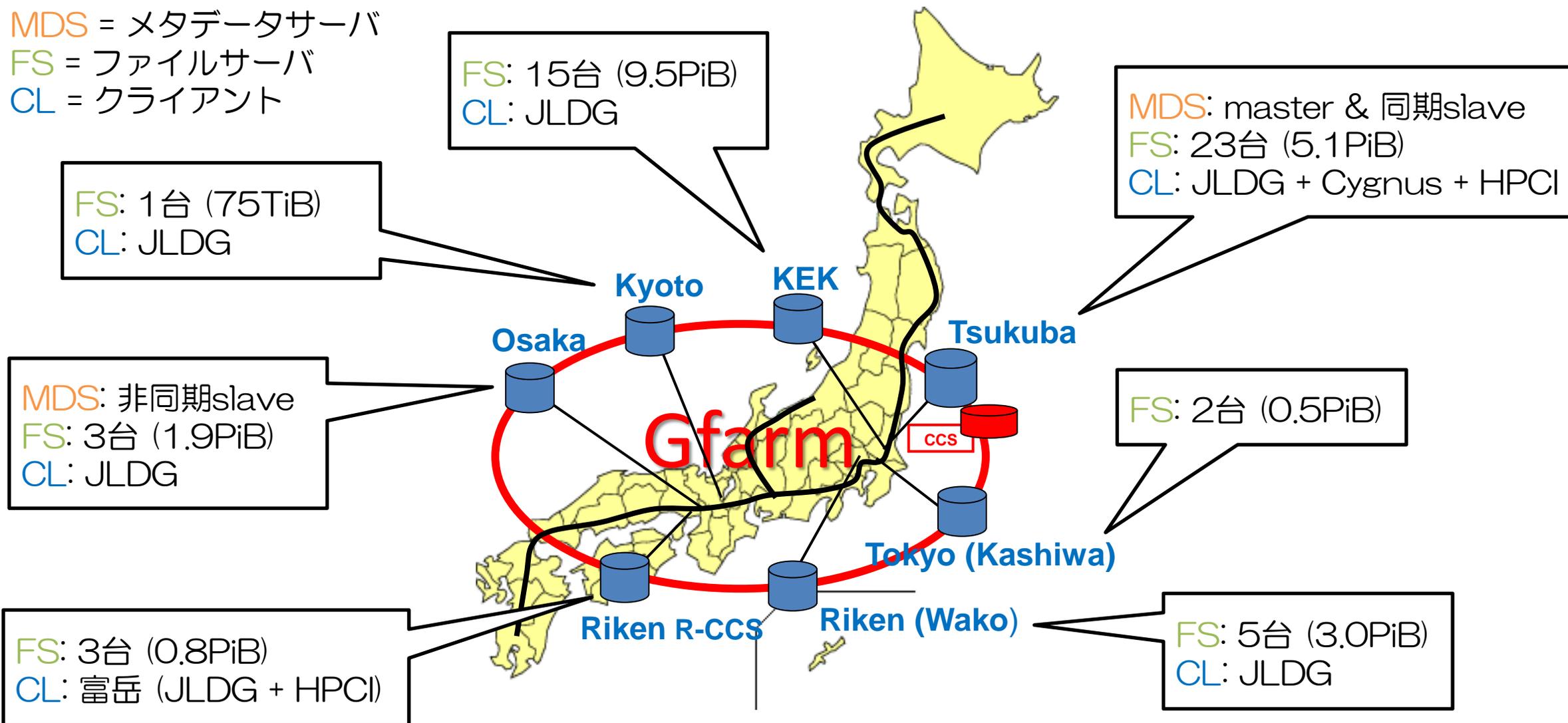
- HPCI 共用ストレージ
 - HPCIの証明書で認証可能
 - 筑波大、富岳に連携システム
- International Lattice Data Grid (ILDG)
 - 筑波大と接続

運用体制

- 管理グループ
 - 各拠点1～4名程度のメンバー
 - 月例ミーティング
 - サーバ保守・管理、ユーザ管理、セキュリティ対応、システム改善、…
- 関連する活動
 - データの公開
 - データの DOI (Digital Object Identifier) 登録
 - 新ILDG (課題: 新認証システム、新データフォーマット、…)

Gfarmの利用 (1): 各拠点の構成

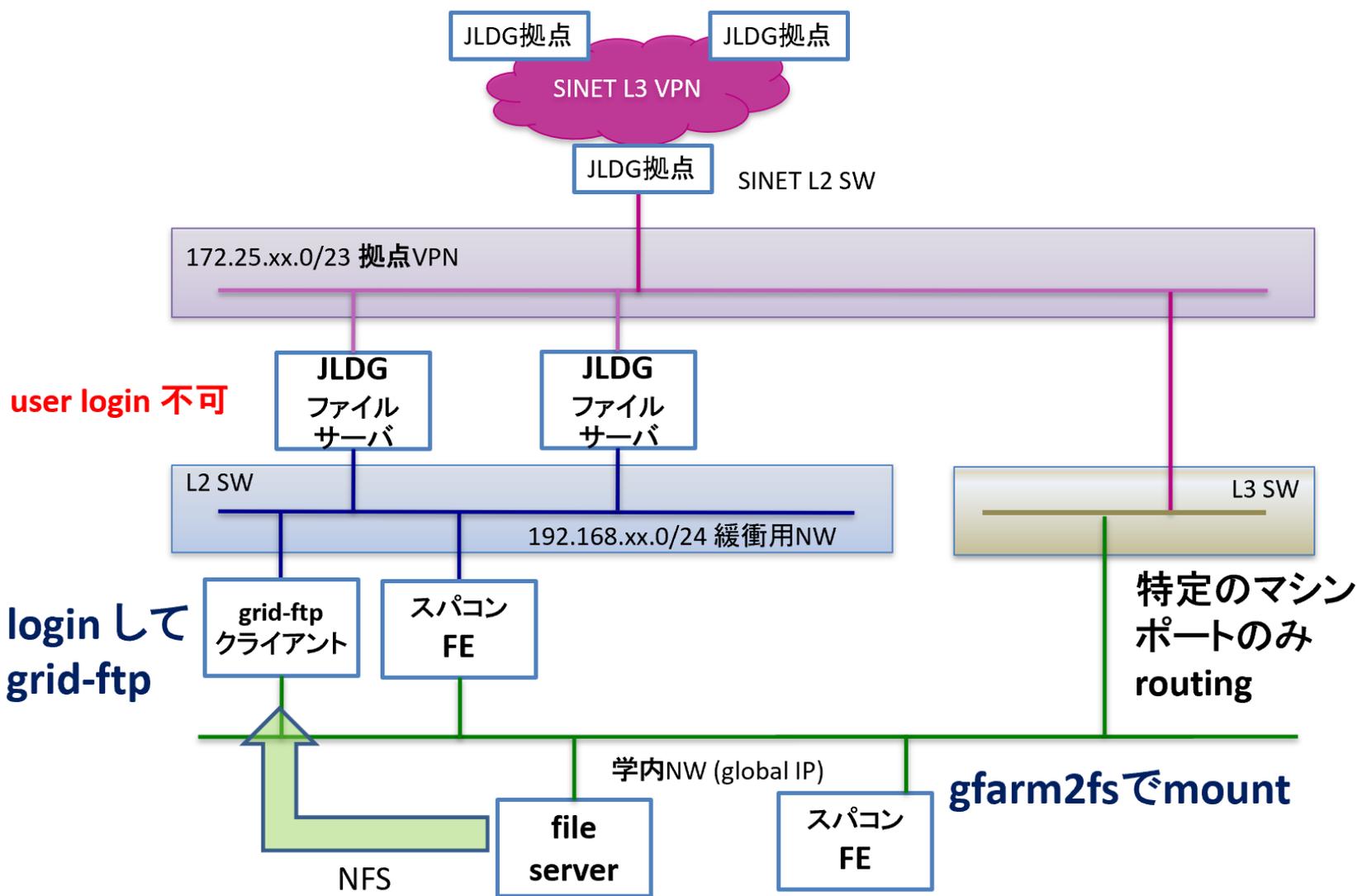
MDS = メタデータサーバ
FS = ファイルサーバ
CL = クライアント



Gfarmの利用 (2): 認証方法

- X.509 公開鍵証明書による認証
 - JLDG独自 + HPCI
- JLDG 認証局 (CA) / 登録局 (RA)
→ 筑波大CCS
- 仮想組織 (VO) を通じてユーザ・グループを管理
- 証明書発行およびJLDG利用開始の流れ
 1. 新規ユーザは所属する研究グループを選択
 2. グループ責任者がユーザの身元を確認、VO管理者にユーザの情報を提供
 3. VO管理者は情報を確認後、ユーザに証明書取得のためのライセンスIDを発行
 4. ユーザはクライアントにて証明書を発行
 5. ユーザは証明書をウェブブラウザにインストール、VO管理ウェブシステムよりVOに登録
 6. VO管理者が承認、登録情報がシステム全体に同期され次第利用開始

Gfarmの利用 (3): ネットワーク構成とデータ



- データアクセス方法
 - GridFTP
 - FUSE mount
 - Gfarm コマンド
- データ管理
 - チェックサムによる整合性チェック
 - デフォルトの複製数: 3
- データの利用
 - グループ内共有
 - ILDGへ公開 (約4万配位が公開済み)

まとめ

- 計算素粒子物理分野では大量の基礎データの保存・共有が必要
 - 世界的な取り組み: International Lattice Data Grid (ILDG)
 - 日本独自の取り組み: Japan Lattice Data Grid (JLDG)
- Japan Lattice Data Grid (JLDG) では Gfarm を利用して日本の各研究拠点間でデータ共有を実現
 - 7拠点、55のファイルサーバ、21PiBのシステム
 - SINET6上のVPN、HEPnet-J/sc で各拠点を接続
 - ユーザからは1つのファイルシステムに見える、保存場所を気にしなくてよい
 - 証明書を用いて認証
 - ILDG、HPCIとの連携

今後の課題

- VO管理システムの再構築
 - システムの大部分は構築完了
 - 移行準備
- 新認証方法の導入
 - 新ILDGとの連携
 - OAuth2認証など (現在はgfarm2.7までを使用)
- 新たなスパコンとの連携
 - Pegasus
 - Miyabi

参考

- JLDG 公式ウェブサイト
 - <https://www.jldg.org/>
- 文献
 - T. Amagasa et al., “Sharing lattice QCD data over a widely distributed file system” , J. Phys. Conf. Ser. 664, 042058 (CHEP2015)
- Gfarm ファイルシステム
 - <http://oss-tsukuba.org/software/gfarm>