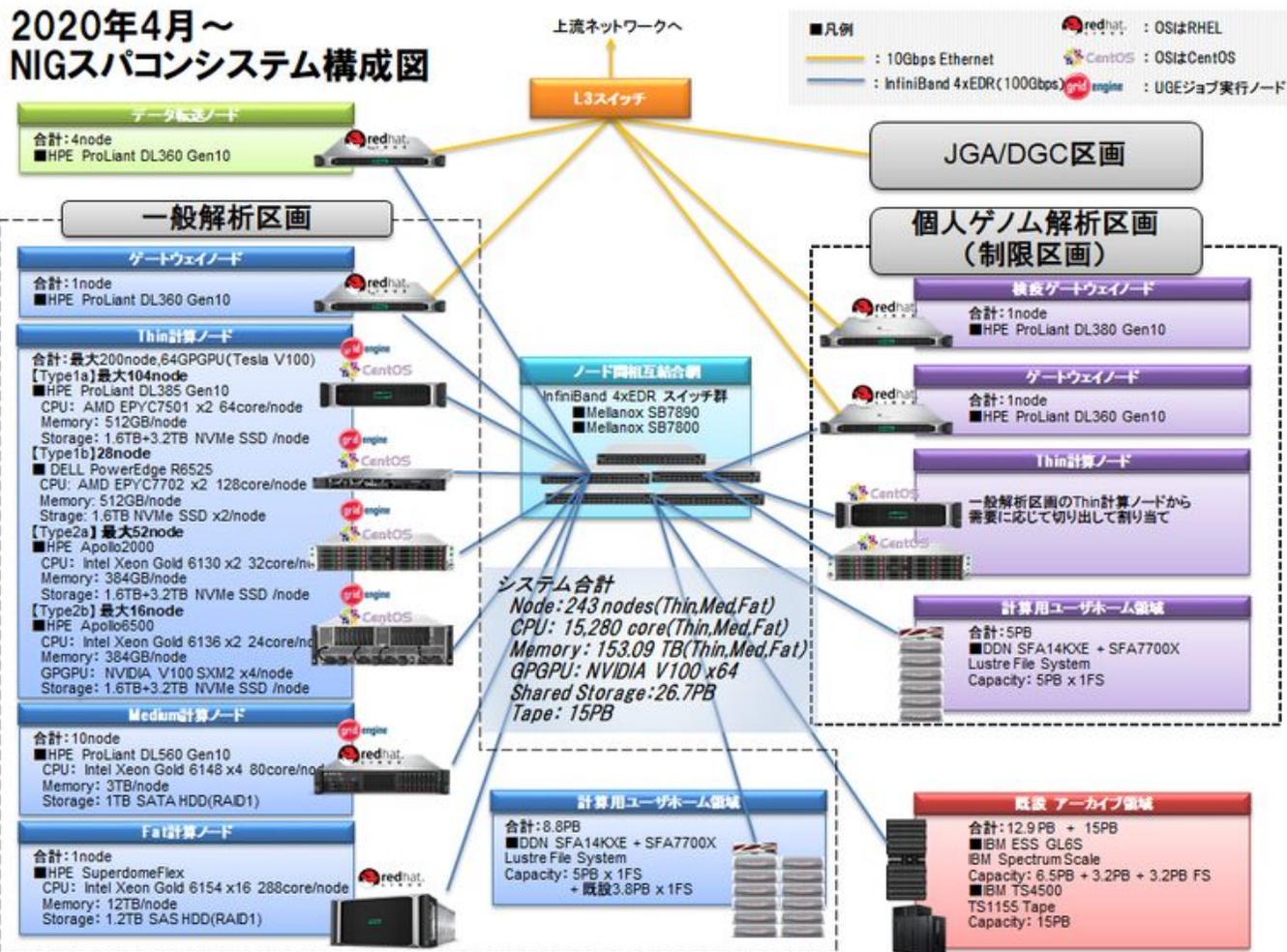


○大田達郎¹, 末竹裕貴², 児玉 悠一³, 藤澤 貴智³, 小笠原 理³, 清水 厚志³, 有田 正規³

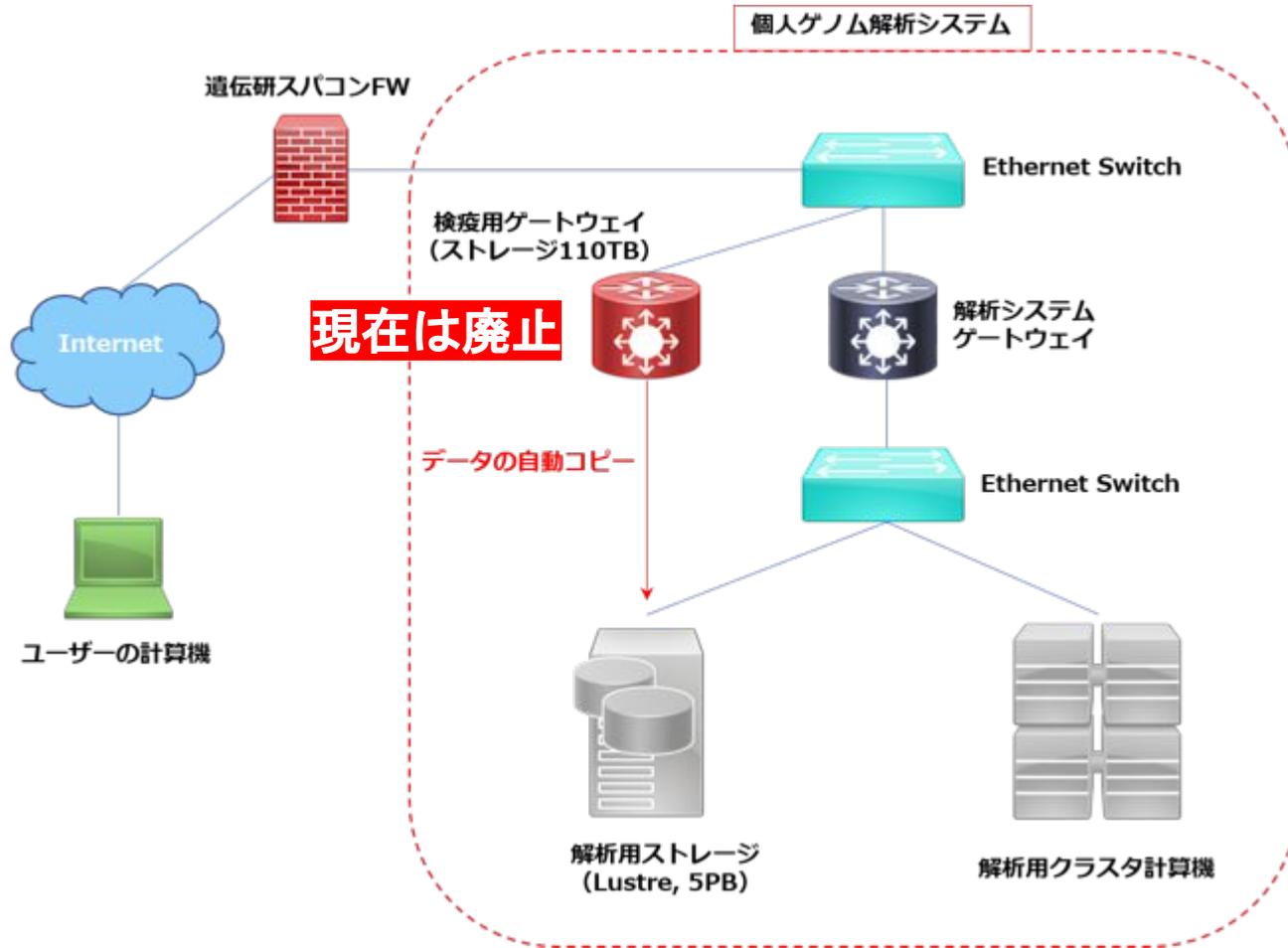
1. ライフサイエンス統合データベースセンター, 2. 東京大学 情報理工学系研究科, 3. 生命情報・DDBJ センター

概要

- ヒトゲノム研究におけるデータ増に伴う大型計算機環境の需要
 - **セキュアかつパワフル**な計算機の調達は容易ではない
- 遺伝研スーパーコンピューター 個人ゲノム解析環境の提供
 - NBDC のガイドラインに準じて構築・運用
 - NBDCヒトデータ取扱いセキュリティガイドライン
 - NBDCヒトデータグループ共有データ取扱いセキュリティガイドライン
 - しかしネットワーク接続に制限があるため、**解析者の作業は通常の計算機環境と比較すると困難**
- DDBJ Workflow Execution Service (WES) の開発
 - コマンドライン操作をすることなく、ユーザはデータとワークフローを指定するだけで容易に解析を行うことができる

2020年4月～
NIGスパコンシステム構成図

- [遺伝研スパコンシステム利用案内](#) より
- 2020年4月に新CPUの導入などを行い現在の構成に
 - 一般ユーザ: 798名
 - 一般ユーザ(大規模): 165名
- 個人ゲノム解析区画
 - 遺伝研スパコンシステムの課金サービス
 - 一般解析と区画が分けられている
 - 利用者の申請に応じて計算ノードが切り出して割り当てられる
- NBDCのガイドラインに準拠
 - NBDCヒトデータベースの [機関外サーバ](#) として認定されている



個人ゲノム解析システムの初期構成

- [遺伝研スーパーコンピュータ個人ゲノム解析環境の利用方法](#)より
- 不正なプログラムがサーバ内で稼働することを防ぐため、検疫用ゲートウェイが持ち込まれるデータやプログラムをスキャン

個人ゲノム解析システムの現在の構成

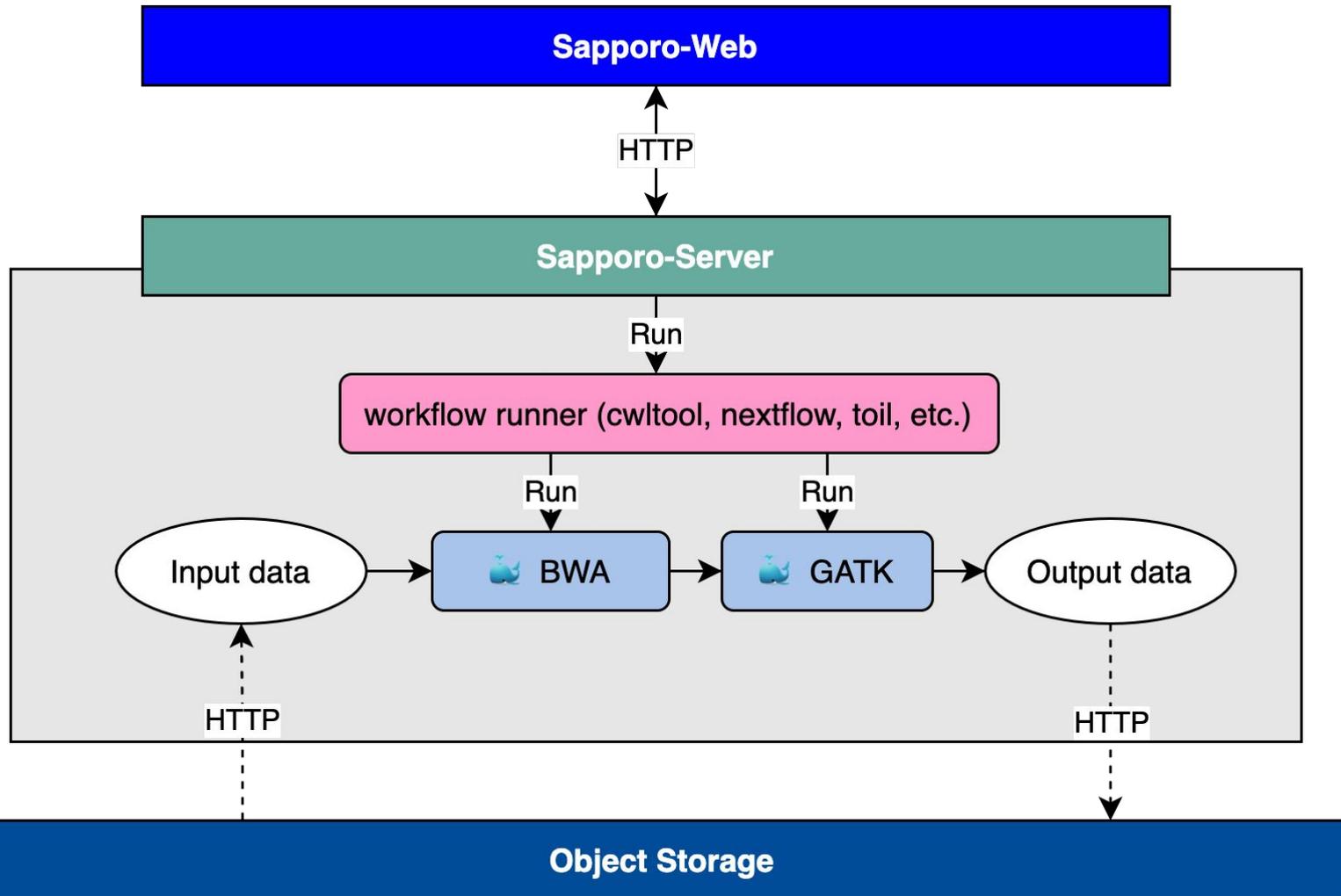
- 検疫用ゲートウェイは廃止
- ユーザはVPNを利用して個人ゲノム解析システムにログインして作業
- ファイルの持ち込み・持ち出しは解析ゲートウェイから
 - 解析ゲートウェイ上でファイルのスキャンが走る

個人ゲノム解析における問題点

- 閉鎖環境では外部とのネットワーク接続が遮断されている
 - データの取得やツールのインストールなど**環境構築に手間がかかる**
- データの取り扱いに専門的な知識が必要
 - **ヒューマンエラー**でデータが流出する可能性がある

問題の解決:Workflow Execution Service (WES) の提供

- 環境構築が不要
 - 新規のワークフロー構築だけでなく、既存のワークフローを再利用できる
 - 解析者はデータとワークフローの構築/選択に集中できる
- データ取り扱いにおけるヒューマンエラーのリスクを下げられる
 - ワークフローは指定したデータを入力に自動的に実行されるため、**解析者はデータを操作しない**
 - 管理者が指定したワークフローのみ実行を許可することも可能



DDBJ-WES Code: Sapporo

- Web/Server の2つのコンポーネント
 - Server を分散することで多拠点稼働が可能
- Web からのリクエストを受けた Server が workflow を起動、Docker コンテナで実行される
- 入力と出力は計算ノードのファイルシステムではなくオブジェクトストレージを介してやり取りする
 - デプロイの簡易化と再現性の向上
- Server API は [GA4GH Cloud WS](#) の [WES Standard](#) に準拠

今後

- 一般区画で DDBJ WES を試用するための
デモサーバーを構築中
- ツールとワークフローを共有するための
仕組みの整備を進める
 - WF言語ごとに既存の資産を活用
- 遺伝研スパコンだけでなくオンプレミスと
クラウドを組み合わせた複数の拠点に
跨ったワークフロー実行

