



北海道大学

日本機械学会 設計情報学研究会
DI Lecture Series 6
2014/10/31
東北大学流体研究所

スパコンとインターネットの連携 による大規模分散設計探査フ レームワークの構築に向けて

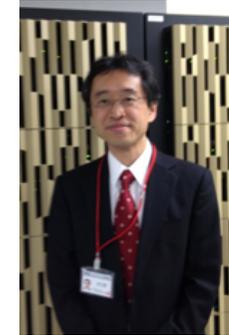
棟朝 雅晴

北海道大学 情報基盤センター
(情報科学研究科)

自己紹介

棟朝 雅晴(むねとも まさはる)

北海道大学情報基盤センター 副センター長
デジタルコンテンツ研究部門 教授



専門:クラウドコンピューティング・分散処理

進化計算・メタヒューリスティクス・最適化

「北海道大学アカデミッククラウド」の設計・構築を担当

情報処理学会数理モデル化と問題解決(MPS)研究会 主査

情報基盤センター クラウドコンピューティング研究会 主査

大学ICT推進協議会クラウド部会 主査

クラウド利用促進機構 総合アドバイザー

日本MSP (Managed Service Provider) 協会 顧問

Open Compute Project Japan 発起人・運営委員



北海道大学

本日のお話

- 「北海道大学アカデミッククラウド」のご紹介
- 「インタークラウド」への展開について
- JHPCN「スパコンとインタークラウドの連携による大規模分散設計探査フレームワークの構築」について
- 今後の展望



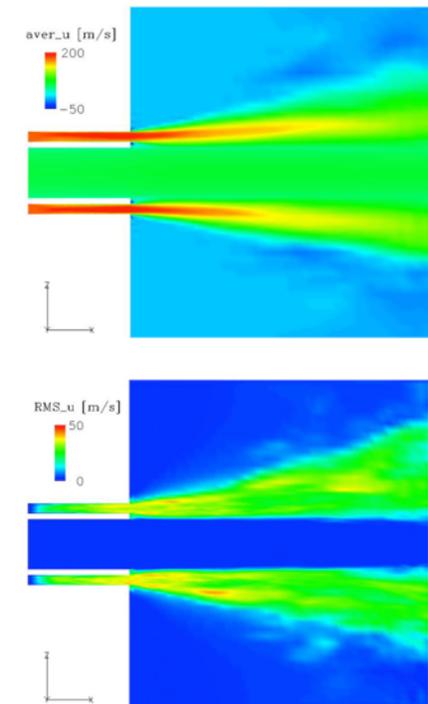
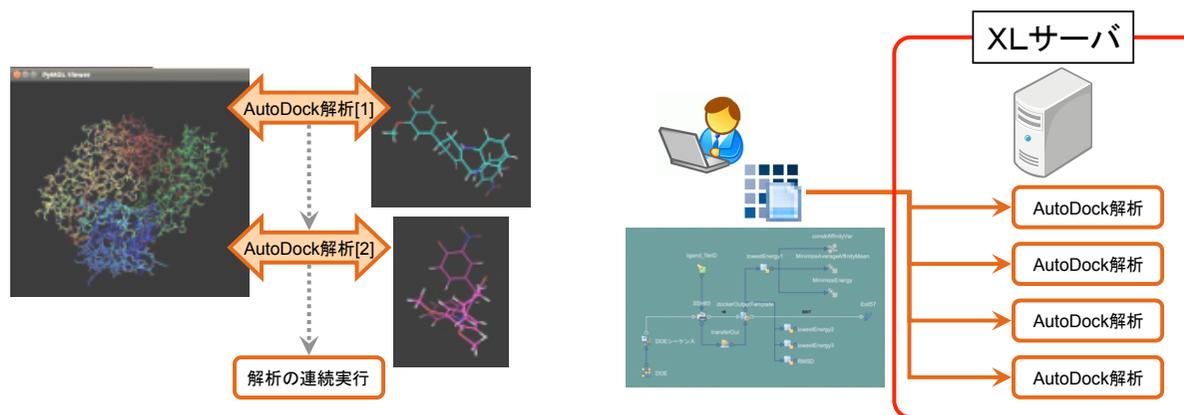
北海道大学アカデミッククラウド

- 全国の学術研究者が利用できる, スパコン並みの性能 (43.8TFlops)を有する国内最大規模の学術クラウドシステム
- 単なる計算資源の仮想化にとどまらず, クラウドミドルウェア (Apache CloudStack)を導入することで本格的なIaaS, PaaSのオンデマンドサービスインフラとして提供
- ビッグデータ処理システム (Hadoopクラスタ+機械学習パッケージ)を自動的に設定し利用者が占有して利用できる基盤を整備



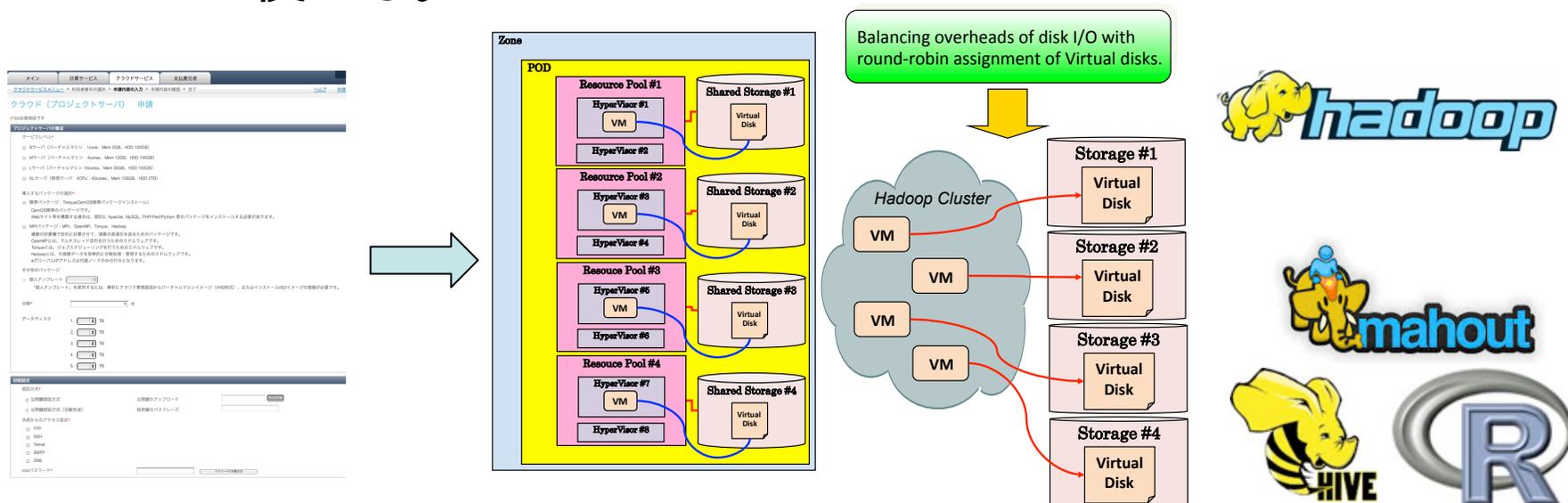
アカデミッククラウドの活用事例(1)

- 計算サーバやシミュレーション専用システムをクラウド上に構築
- 研究室のサーバをクラウドシステムの高性能サーバに移行することで、コスト・電力削減および研究開発プロセスの大幅なスピードアップ(システム構築に数ヶ月→数時間)
- 創薬化学の”in silico screening”など、セキュリティ要件が厳しく、共用スパコン・クラスタ等の利用が難しい用途に適する



アカデミッククラウドの活用事例(2)

- ビッグデータ処理パッケージの提供: Hadoop + Mahout (機械学習パッケージ) + R (加えてMPI, Torqueなど並列計算も可能)
- 自動的に構築して利用できるクラスタパッケージを提供 → 数百台規模のクラスタでも、申請後約1時間で設定終了、遅くとも翌日には使える。



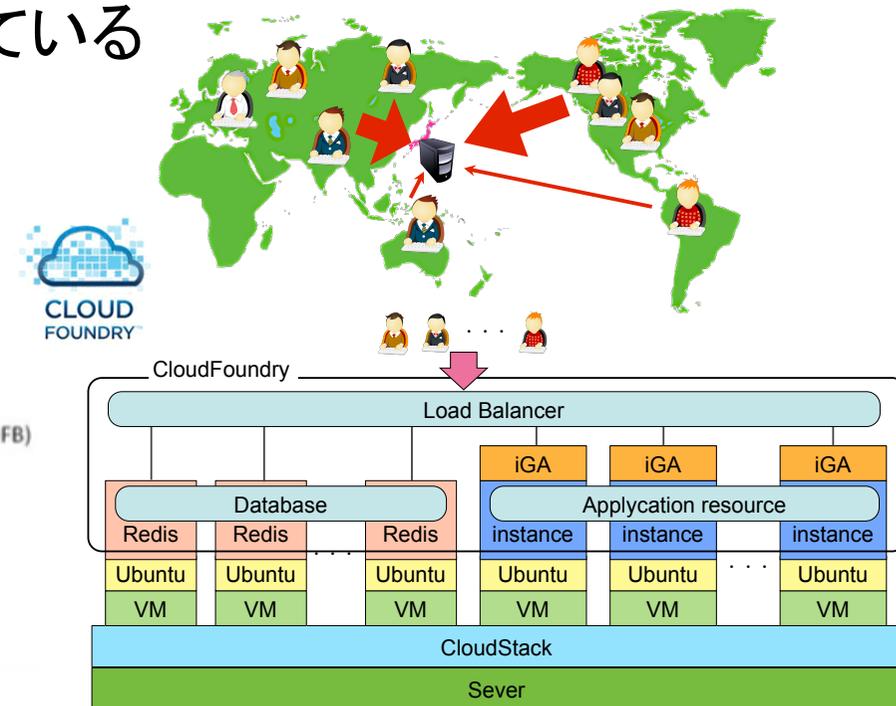
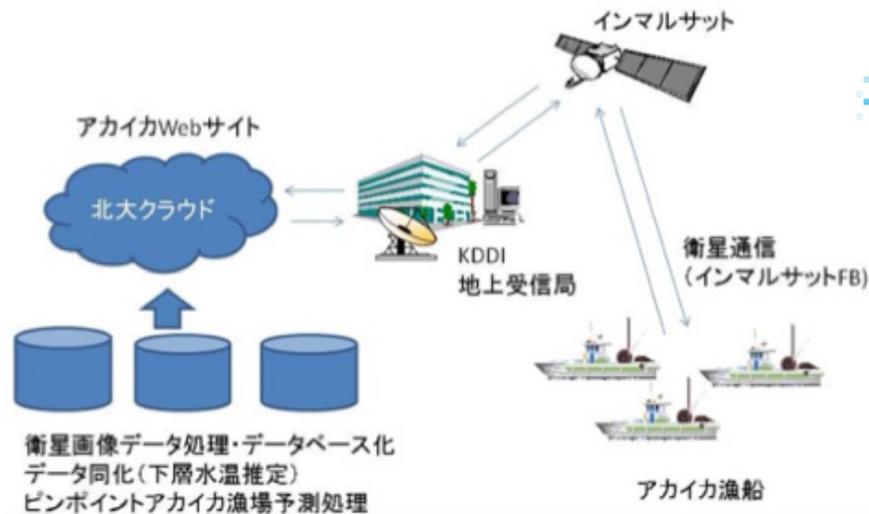
ポータルから申請・設定

システムの自動構築・設定



アカデミッククラウドの活用事例(3)

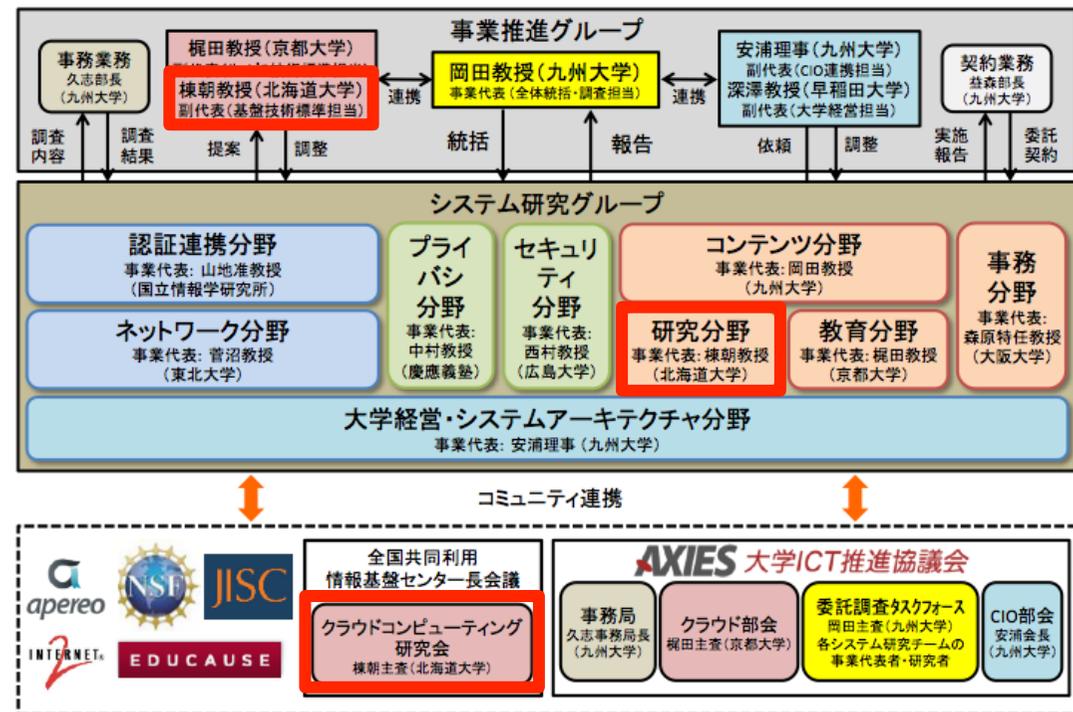
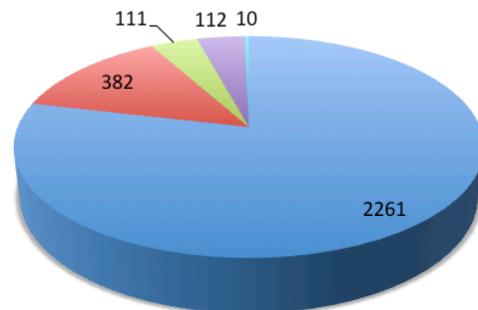
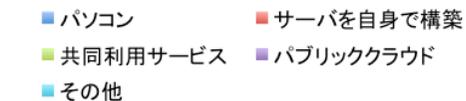
- 先端的な汎用 PaaS (Platform-as-a-Service) を活用した、機械学習と連携による大規模な情報システムの構築など、先端的な情報システムに関する研究開発を推進
- 水産科学(漁場予測)におけるデータの集約など当初予想していなかった活用事例が出てきている



アカデミッククラウド委託調査(文科省)

- 研究支援に係るアカデミッククラウドの調査検討を担当
 - 研究支援組織・科研代表者全員にアンケート調査
 - 海外等の事例調査(大学, クラウド関連企業等)
 - 要求仕様の提案

データ処理に用いている計算資源について
(研究課題数・重複有り)



研究者からのコメント(例)

- データがどんどん増えているが、適切な管理方法が分からず、エクセルファイルでいくつものファイルを保存している。
- データはそのものは基本的に無料で公開されているものが多いが、必要データのダウンロードに、数ヶ月要し、ボトルネックとなる場合が多い、との指摘あり。
- Dropboxなど商用のオンラインストレージは便利だが、不安も有り、容量制限や料金、転送速度の問題もあるため、高速、大容量、安価で、セキュアなオンラインストレージサービスを求める声が多い。
- データのバックアップについて課題を抱えている例が多い、大容量のデータを簡単かつ安価にバックアップできるようにしてほしいとの要望が多い。特に災害対応まで手が回らず不安を抱えている。
- 証拠能力のある「研究ノートの電子化」(研究データの保全・を担えるシステム)や、情報漏洩を防止する、高度に暗号化されたファイルシステムの要望あり。



ビッグデータの活用がイノベーションの中心

- facebook:写真・動画で300PB以上を蓄積 → 毎日500TB以上のデータを処理 → オンデマンドでビッグデータ処理インフラを構築し, 新サービスの開発



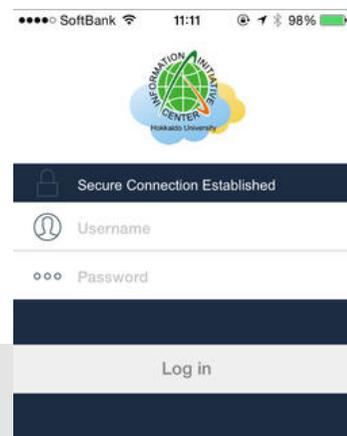
- Imperial College London: Research Data Service
として, 100PB超の共有研究データベースを構築中
- CERN: LHC(加速器)のデータが1ZB/year
→ スイスとハンガリーのデータセンターに分散配置



ペタバイト級データサイエンス統合クラウドストレージシステム@北海道大学

10

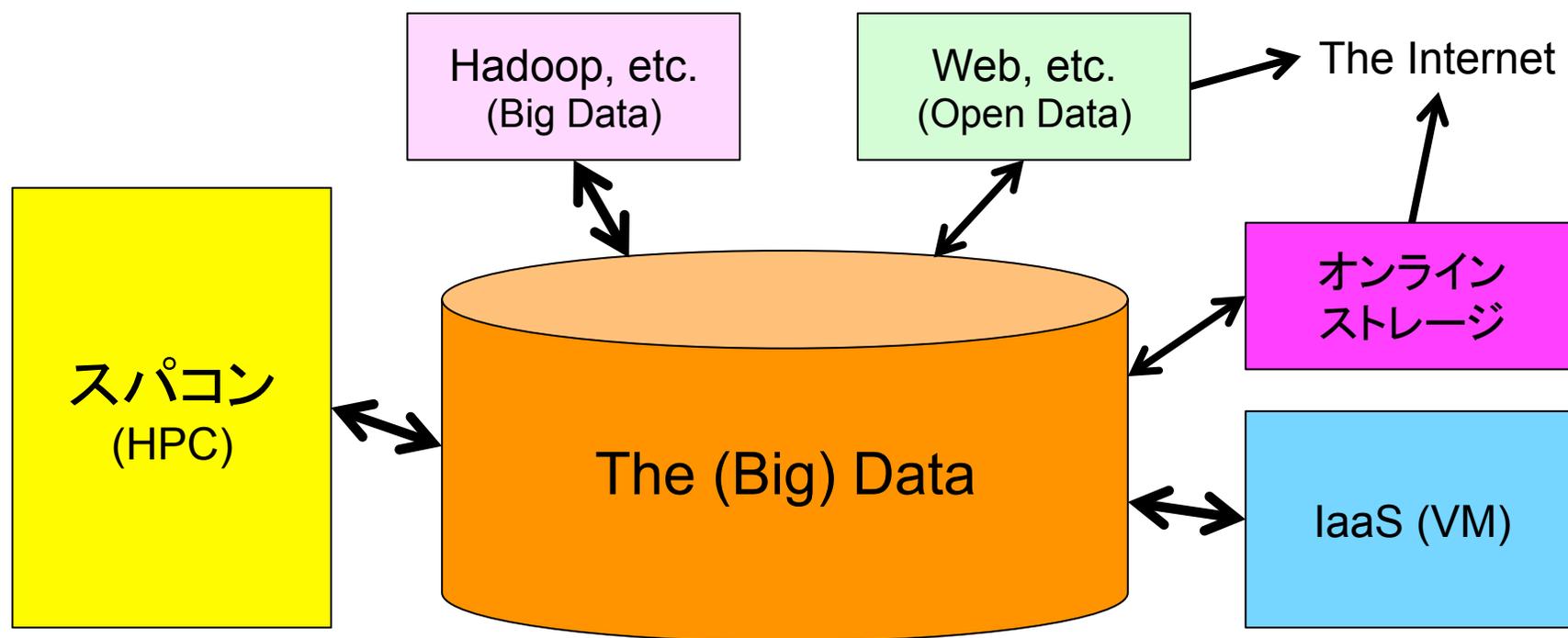
- 2014年4月からサービス開始:ビッグデータ・データサイエンスに関連する研究(HPCI/JPHCN等)を支援
- 物理容量2PBの高速クラウドストレージ+CloudStack最新版によるクラウドシステム(500コア)
- Amazon S3互換, Gfarm, WebDAVに対応し, さらにDropboxライクのオンラインストレージとして利用可能
→ 同期クライアント(iOS, Android, Windows, Mac対応)



北海道大学

次期北大クラウドの向けて:データ中心アーキテクチャ

- “(Big) Data” を中心に全てを考える
 - 大学の「知的財産」としてデータを生成, 蓄積, 活用するシステム
 - ストレージと計算・処理・配信インフラを密結合する



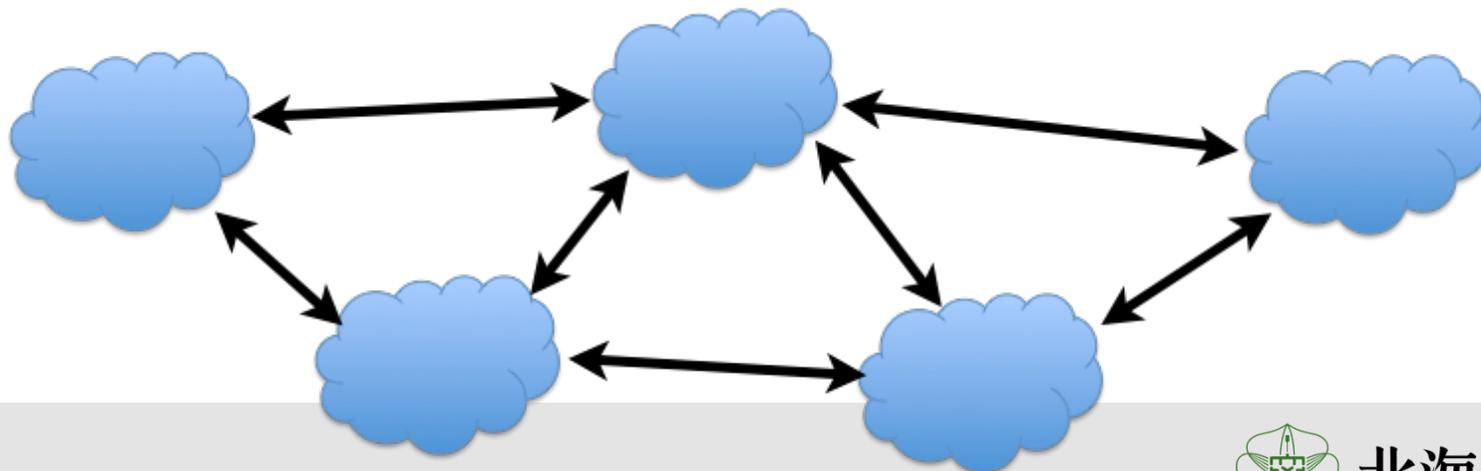
本日のお話

- 「北海道大学アカデミッククラウド」のご紹介
- 「インタークラウド」への展開について
- JHPCN「スパコンとインタークラウドの連携による大規模分散設計探査フレームワークの構築」について
- 今後の展望



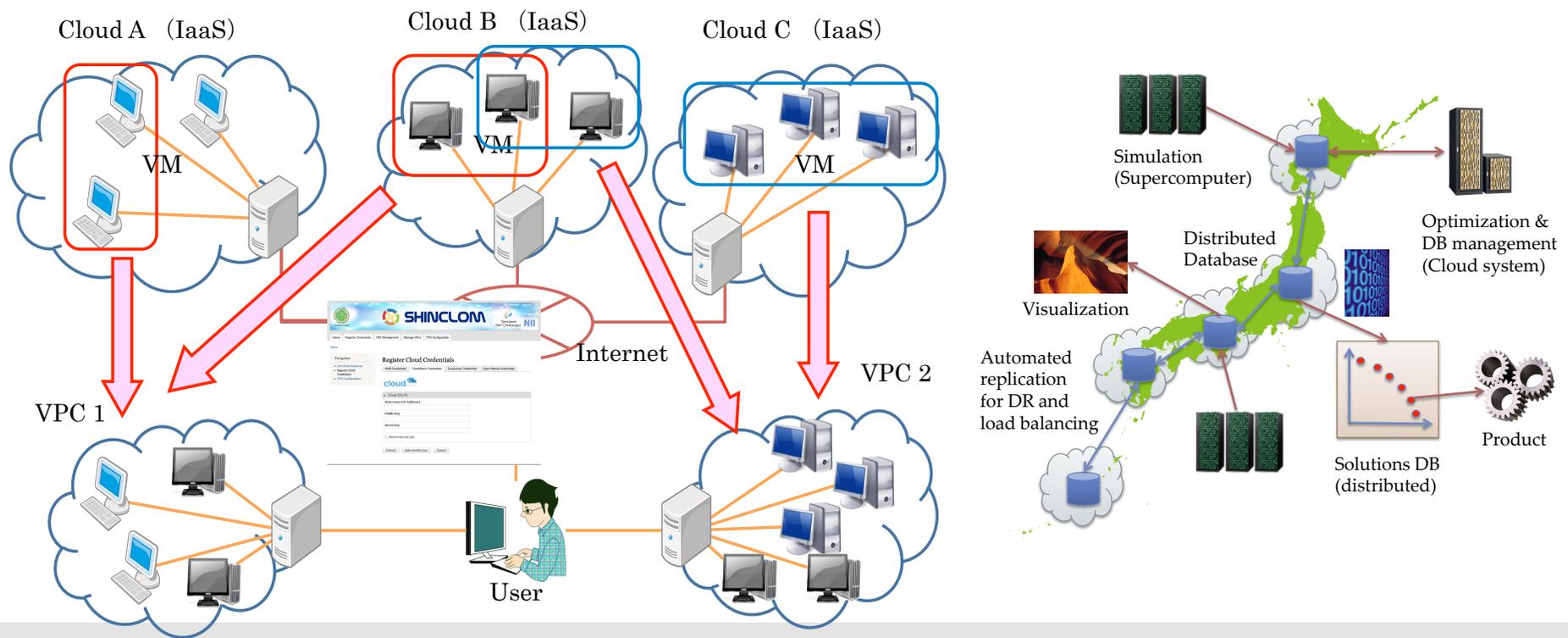
インタークラウド (Inter-Cloud)

- インターネット (Inter-net) がネットワークを相互接続することで世界規模のインフラを構築したように, クラウドシステムが相互接続, 連携することで世界規模のインフラとしての **”Cloud of Clouds”** を実現 → インタークラウド
- 相互運用性 (Inter-Operability) が重要な条件となる: サービス API の共通化, 認証連携などの要素技術に加えて, 資源割当ポリシー, 運用モデルの確立など, 運用システムとしての実現にあたっては課題も多い



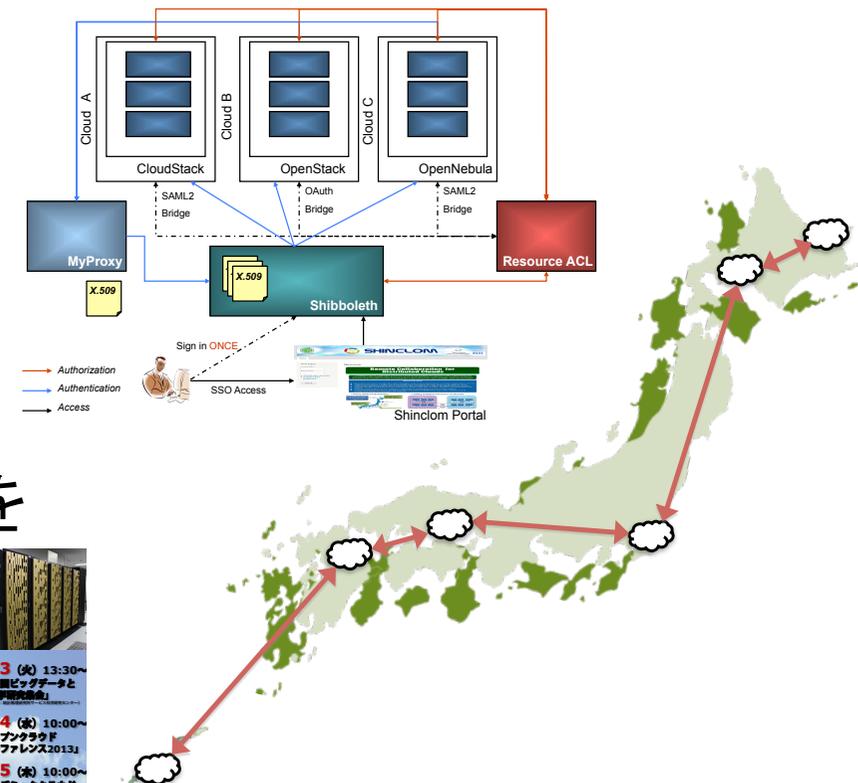
インタークラウドの実現に向けた取り組み(1)

- 各大学のプライベートクラウドを連携させた「インタークラウド」の実現 → 研究者がプロジェクト専用のシステムとして VPC (Virtual Private Cloud) をインタークラウド上で自由に構築する技術の確立, 大規模設計探査などアプリケーションでの実現



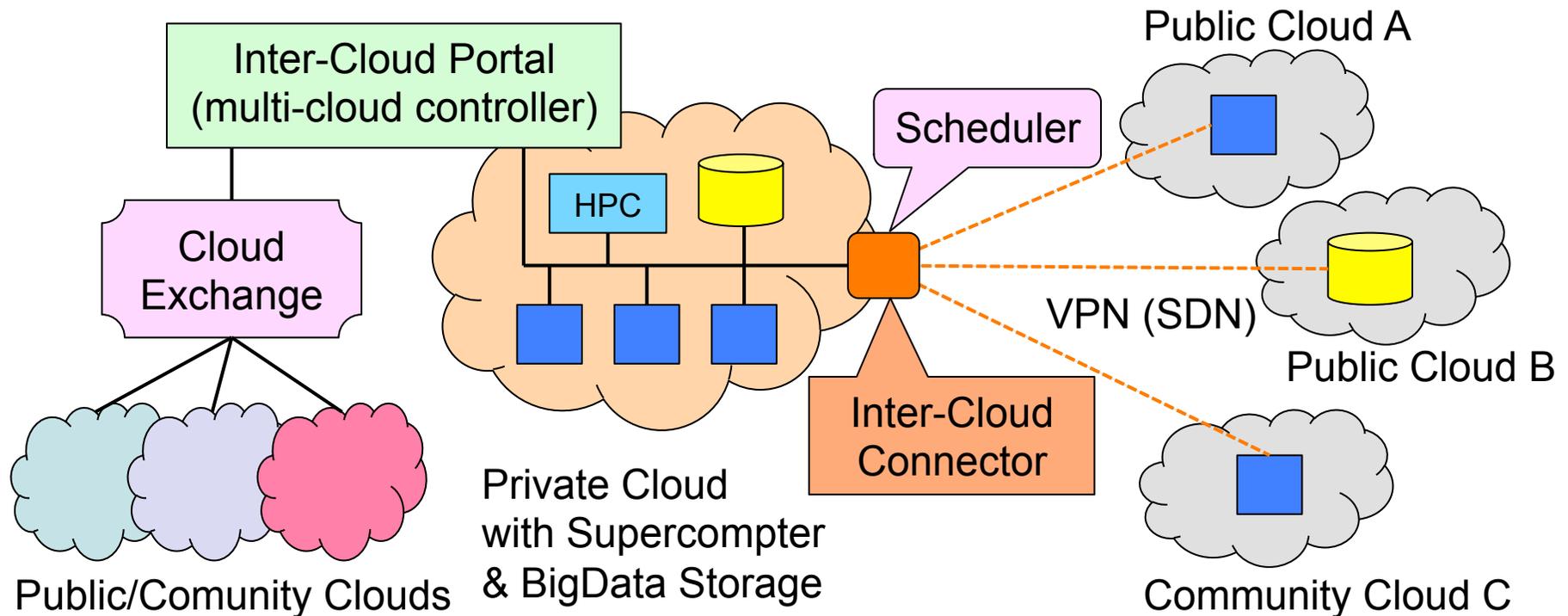
インタークラウドの実現に向けた取り組み(2)

- SINET4の北端(北見工大)～南端(琉球大学)に至る全国規模の大学クラウドの連携に向けたインタークラウド試験システムの構築 → 認証連携, 仮想ネットワーク連携, 仮想マシンクラスタ, 分散データベースの検証
- コミュニティ活動: 大学ICT推進協議会クラウド部会, アカデミッククラウド委託調査, 民間との連携(クラウド利用促進機構など)
- ビッグデータとクラウドの連携: “ビッグデータと統計学研究集会”を CloudWeek@北海道大学にて主催



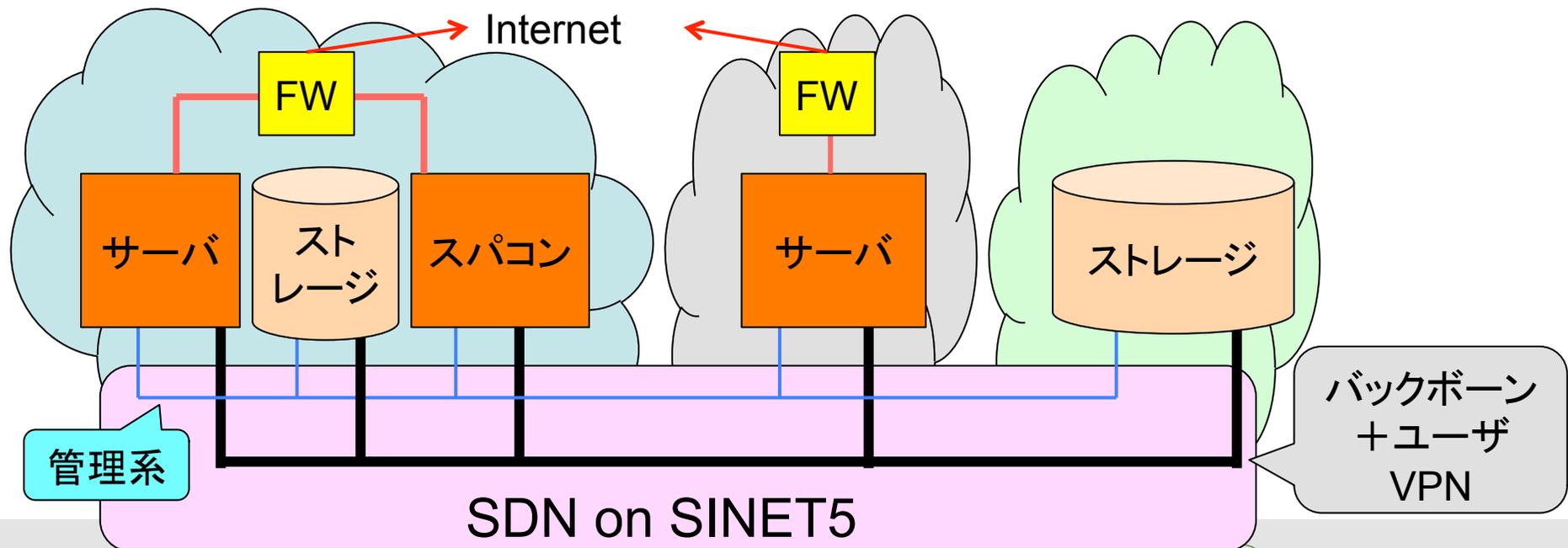
北大次期クラウドのシステムイメージ(案)

- 「北大ハイパフォーマンスインタークラウドシステム」として、サービスレベル、インフラレベルの両方でインタークラウド対応



クラウドインフラの広域分散化

- クラウドインフラネットワークの基本構成を広域分散環境へ拡張: バックボーン+管理系をSDN運用
 - バックボーン, 管理系ネットワークはセンター管理下で直結
 - ユーザ・プロジェクト毎のVPNをバックボーン上に動的構築



本日のお話

- 「北海道大学アカデミッククラウド」のご紹介
- 「インタークラウド」への展開について
- **JHPCN「スパコンとインタークラウドの連携による大規模分散設計探査フレームワークの構築」について**
- 今後の展望



- JHPCN 平成26年度採択課題
- スパコンとクラウドシステムを連携させた、最適設計のフレームワーク、システムの構築を目的とする
 - シミュレーションを各拠点のスパコンで実行
 - 全国規模で実現されたスケーラブルな分散データベース上で設計パラメータに関する情報を管理
 - スケーラブルなオブジェクトストレージ上でシミュレーションにおいて生成される付随的なデータ, 情報を管理
 - 多目的設計探査を実現する多点探索による最適化エンジンをクラウド上で並列分散実行
 - 可視化システムとも連携し、設計者との密なインタラクションを実現



研究組織(敬称略)

- 棟朝 雅晴(北海道大学)研究統括及びシステムの全体設計
- 千葉 一永(北海道科学大学)流体シミュレーション及び多目的最適化アルゴリズムに関する検討
- 大林 茂, 竹島 由里子(東北大学)可視化に関する検討
- 伊藤 貴之, 桑名 杏奈(お茶の水女子大学)可視化に関する検討
- 金崎 雅博, Atthaphon Ariyarit, 大司 早織(首都大学東京)流体シミュレーションに関する検討
- 渡邊 真也, 長舟 和馬, 佐藤 哲也, 中平 顕士, 奥寺 将至(室蘭工業大学)多目的最適化あるボリズムに関する検討
- 三浦 克宜(北見工業大学)クラウドコンピューティング及び分散データベースに関する検討
- 阿部 友哉(北海道大学)クラウドコンピューティング及び分散データベースに関する検討



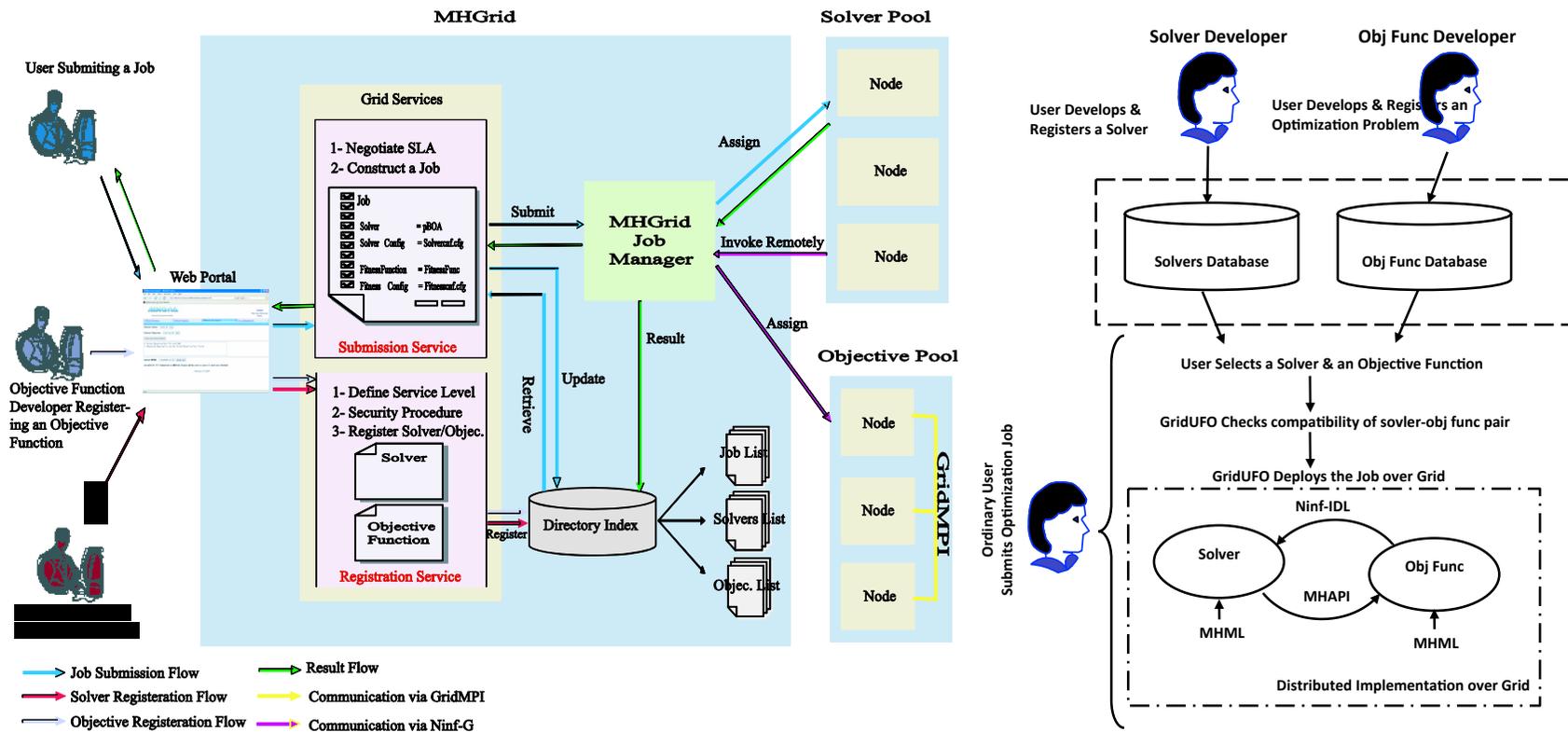
JHPCN <https://jhpcn-kyoten.itc.u-tokyo.ac.jp/>

- 「学際大規模情報基盤共同利用・共同研究拠点」
- 北海道大学、東北大学、東京大学、東京工業大学、名古屋大学、京都大学、大阪大学、九州大学にそれぞれ附置するスーパーコンピュータを持つ8つの施設を構成拠点とし、東京大学情報基盤センターがその中核拠点として機能する「ネットワーク型」共同利用・共同研究拠点
- 毎年、スパコン、クラウドを活用した共同研究の公募を実施（原則として、計算機利用料の補助のみ）
- H27年度は11月7日から公募案内開始予定



Grid Unified Framework for Optimization (Grid UFO)

- グリッドコンピューティング環境において、大規模なシミュレーションと連携したパラメータサーベイ、最適化の連携フレームワークを提案 (Asim, Wahib, Munetomo, 2008-2010)



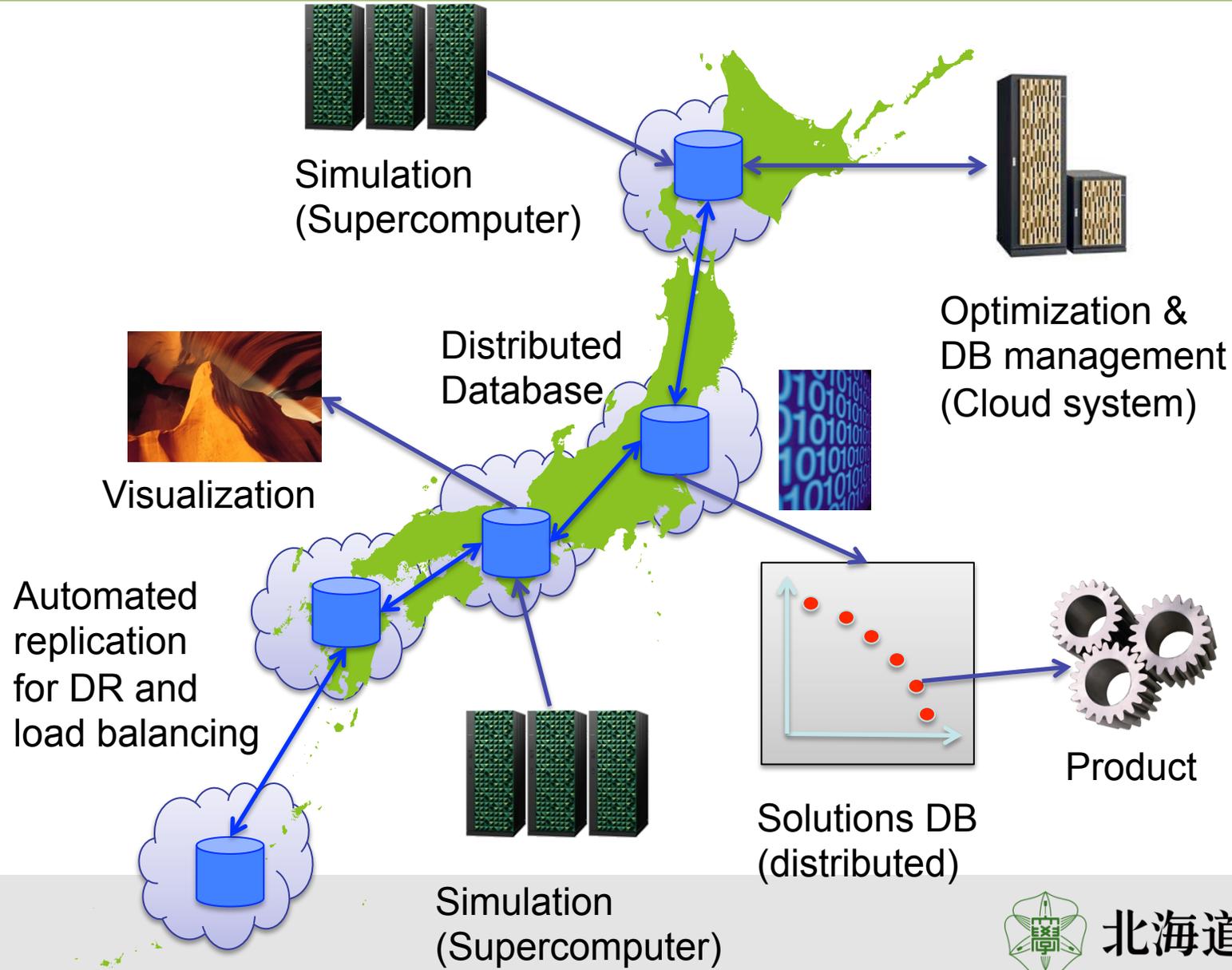
「設計探査の情報・データ共有」

を中心に

「クラウドの作法」で再設計する

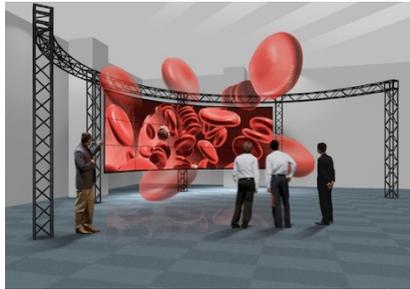


全体イメージ

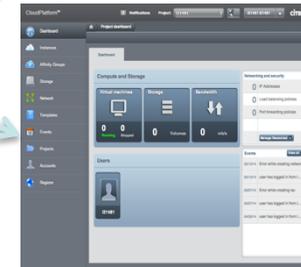


計算資源の概要 (JHPCN)

大阪大学サイバーメディアセンター
可視化システム+PCクラスタ



九州大学情報基盤
研究開発センター
PCクラスタ



北海道大学情報基盤センター
ペタバイト級データサイエンス統合
クラウドストレージ
・CloudStack (IaaS, 500コア)
・Amazon S3互換ストレージ(2PB)

東北大学サイバーサイエンスセンター
スーパーコンピュータ



※ 本研究の一部は、学際大規模情報基盤共同利用・共同研究拠点の支援による。



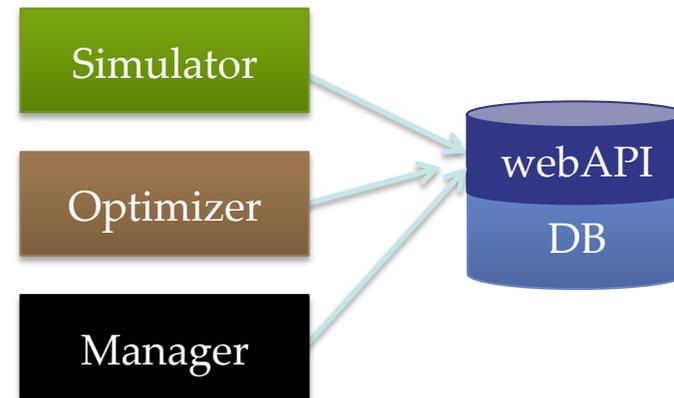
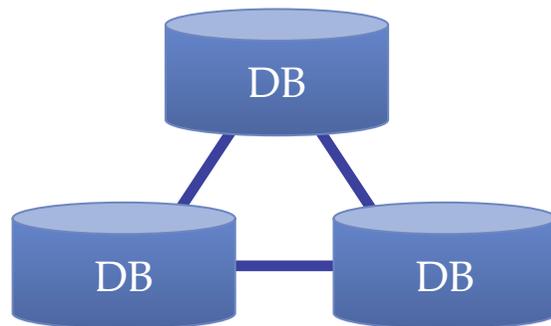
主な課題

- 解(適応度)の評価にかかる時間が数時間から、場合によっては数週間になる
- シミュレーションにより得られる情報が膨大であり、その管理(バックアップ等も含む)が困難
- スパコンへのアクセスがSSHなどに限定され、Webサービス等での連携が難しい
- 1台のスパコンに閉じていれば問題ないが、複数のスパコンやクラウドでの情報共有が難しい
- シミュレーション, 最適化, 可視化の専門家などが、簡単に遠隔連携して共同研究できるシステムが必要
- 派生問題など, 類似した問題を多数解きたい



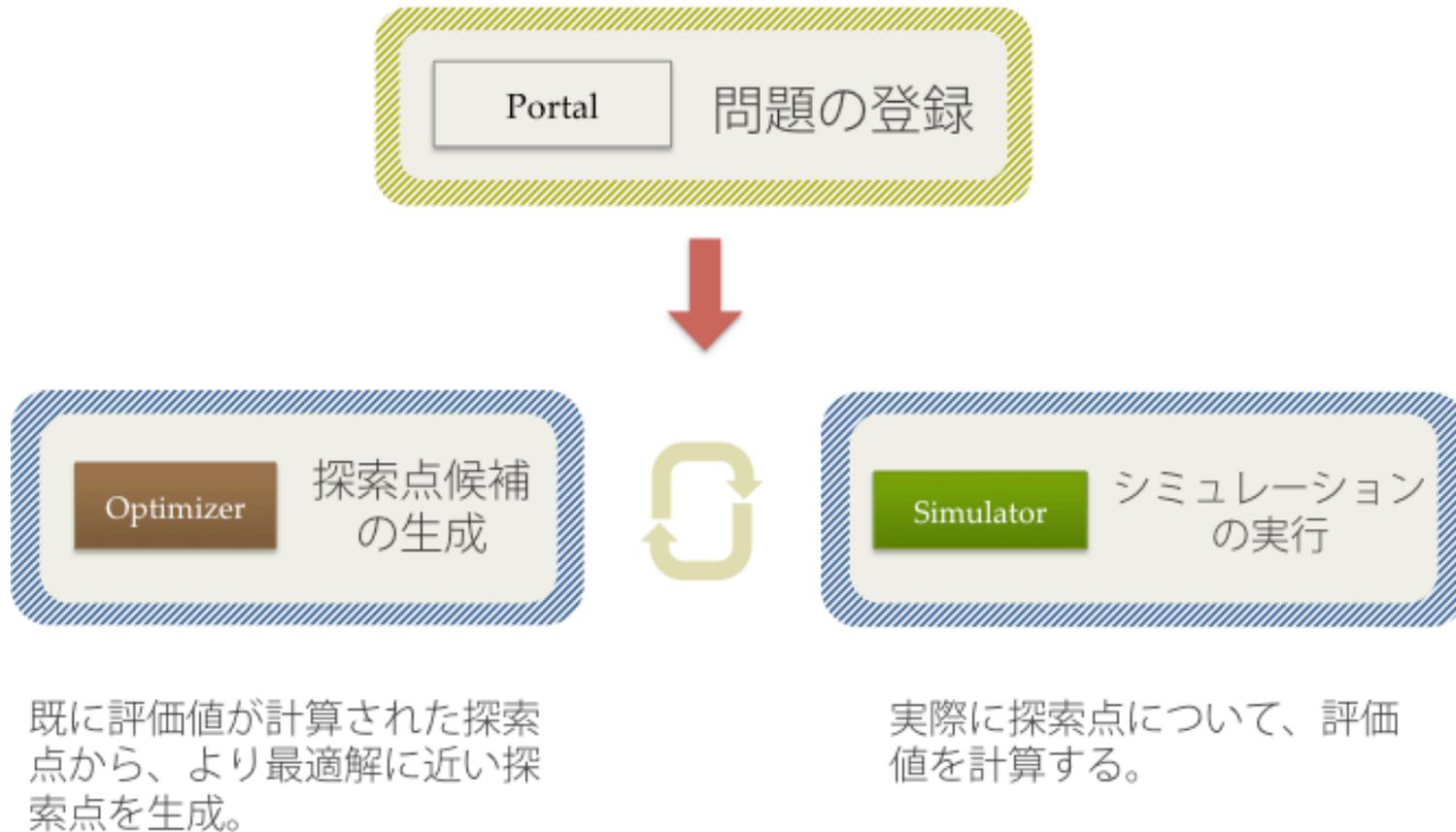
設計上のポイント

- 完全な非同期処理による制御を前提とする
 - スパコン上での長時間処理, バッチ処理待ち
 - ジョブの終了時間の予測ができない
- NoSQLデータベースによりスケーラビリティを確保
- WebサービスAPIでの連携による汎用性の確保



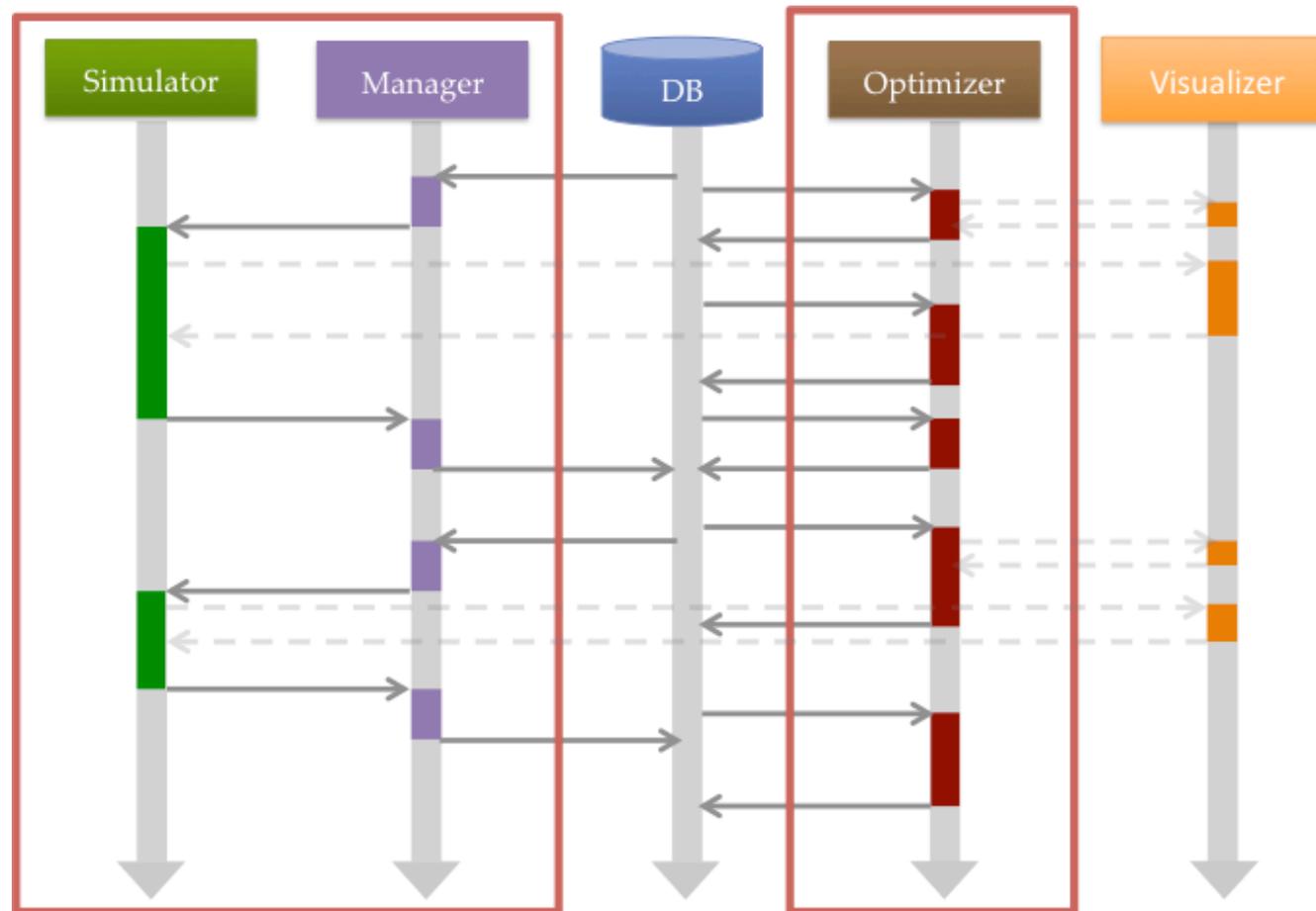
ユーザから見たシステムの動作

- 問題の登録, シミュレーション, 最適化の実行を簡単に



非同期処理の制御

- 分散データベースを中心に複雑な非同期処理を制御



NoSQL型データベース

- NoSQL (Not only SQL) : 通常のリレーショナルデータベース (RDBMS) 以外のデータベース管理システム
- スケーラビリティ: 通常 of RDBMS を大規模化すると, コストが急激に上昇, 超大規模化が困難
- CAP定理 (Consistency, Availability, Partition-tolerance) → 3つの要素のうち, 一つをある程度諦めることで, スケーラビリティを確保する
- クラウドシステムの大規模化に伴い, 従来のRDBMSに加えて, NoSQL型のデータベース管理システムが用いられることが多くなってきている → 完全分散化



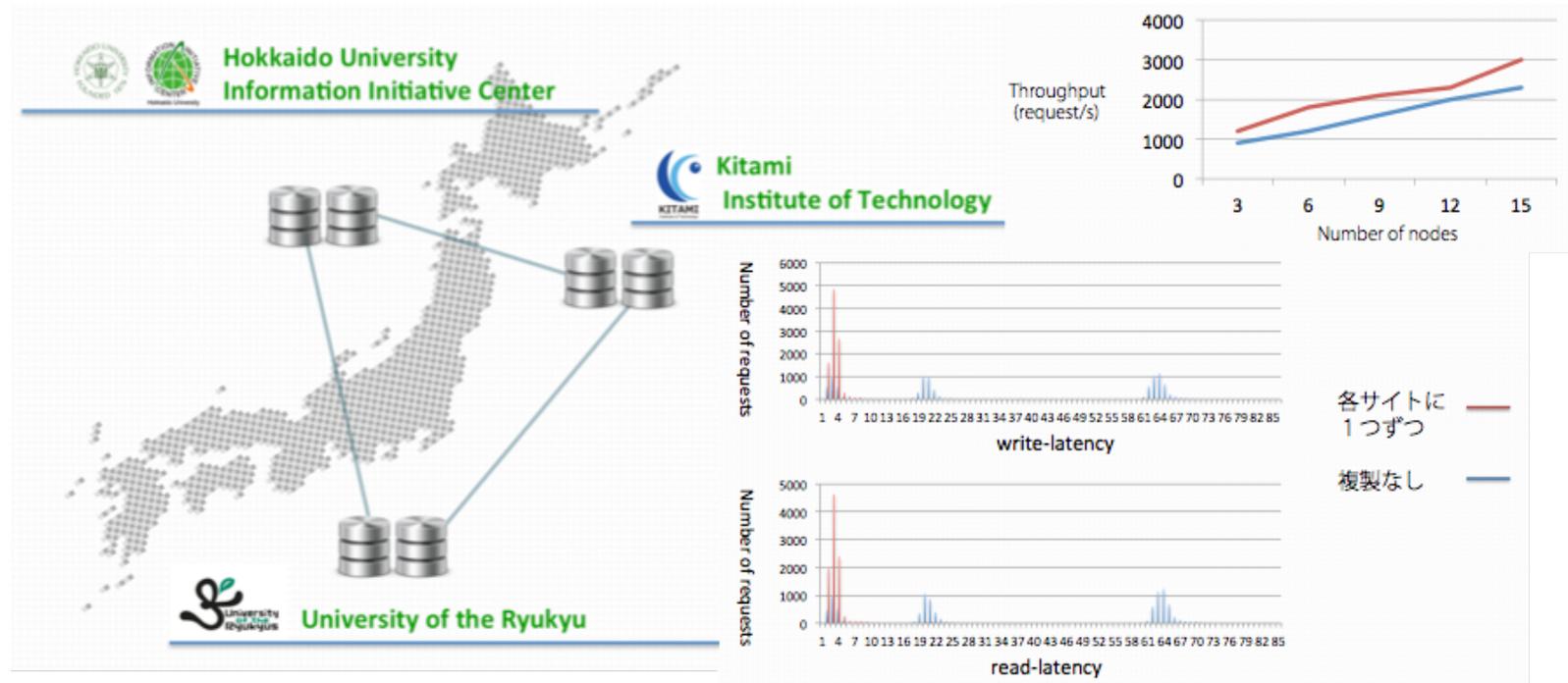
NoSQL型データベースの例

- Apache Cassandra : 分散データベース, カラム指向型
- Apache HBase : 分散データベース, カラム指向型
- Amazon Dynamo : 分散データベース, 商用サービス
- Basho Riak : 分散データベース, キーバリュー型
- Mongo DB : 分散データベース, ドキュメント指向型
- Redis : インメモリデータベース, キーバリュー型
- Neo4j : グラフ型



全国規模での分散データベースの実現

- SINET4の北端～南端(北見工大～北大～琉球大)に渡って, 分散データベース (Cassandra)基盤を構築
→ 応答性能, 対故障性, 災害対応などについて検証済み



(阿部友哉「インタークラウド環境下における広域分散データベースの構築に関する研究」, 2014)

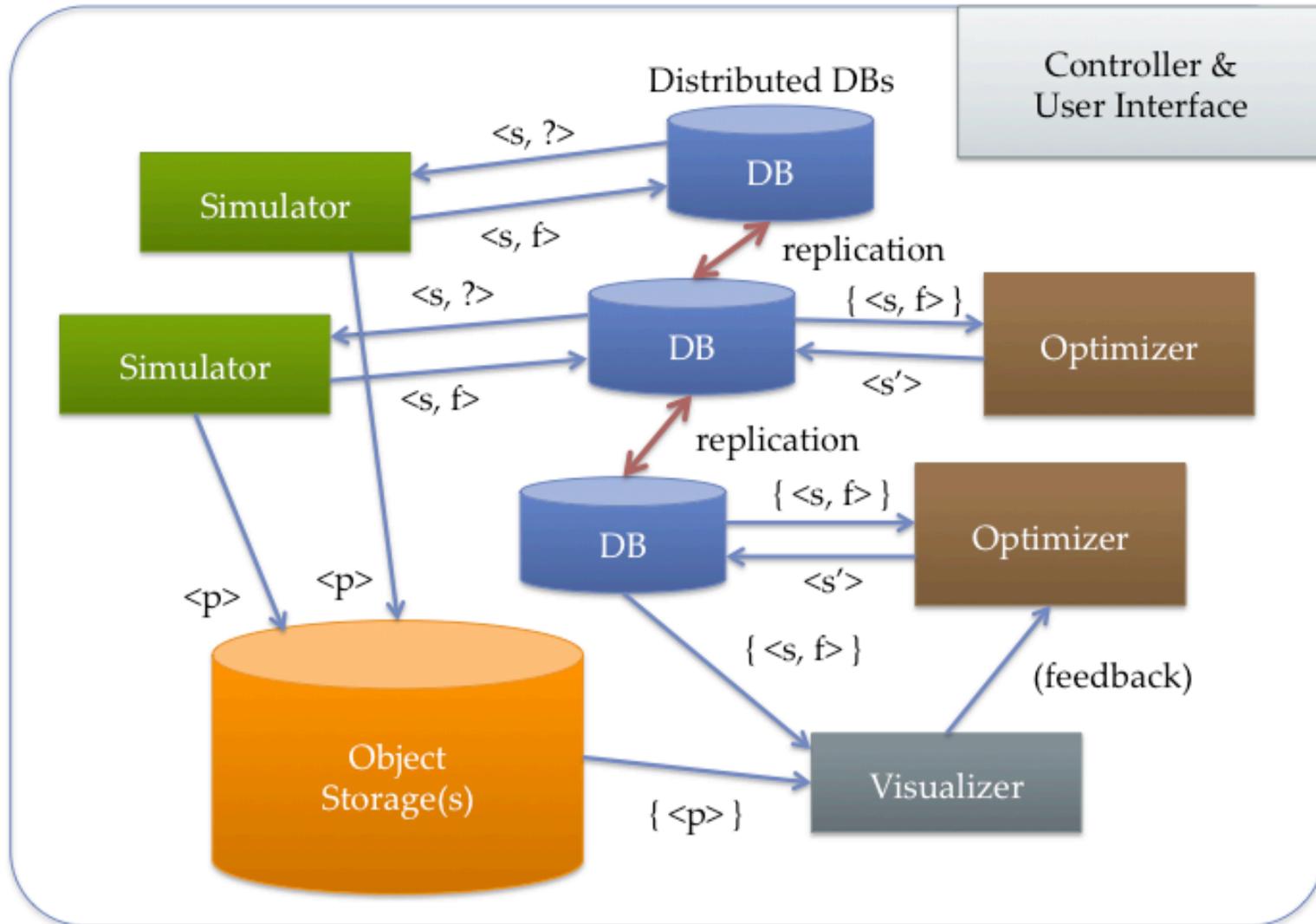


Webサービスによる連携

- 全ての通信および制御について, REST (Representation state transfer) 形式のWebサービスAPIにより記述
例) `http://sample.web-service.com/api?parameters=values`
- スパコンについては, SSH (GSI-SSH) のみのアクセスとなっているため, SSH->HTTPの変換を行うエージェントを配置する
- 分散データベースのCassandraについても, RPCでのアクセスとなるため, WebサービスAPIへの変換を行う
- オブジェクトストレージについてはREST APIが利用可能



ソフトウェアコンポーネント間の連携



ソフトウェア(コンポーネント)



解情報データベース

設計最適化問題を解く上で発生する設計情報を格納する。
マルチテナントデータベースとして、多数の問題に関する設計情報を一括管理。
完全分散アーキテクチャによりノードを理論上無限に拡張可能。

Simulator

シミュレータ

ポータルサイトから送信されたシミュレーションプログラムを格納し、制御システムからの ssh 接続によってシミュレーションを実行する。

Optimizer

最適化エンジン

解情報データベースの情報をもとに次の探索点の候補集合を生成する。

Visualizer

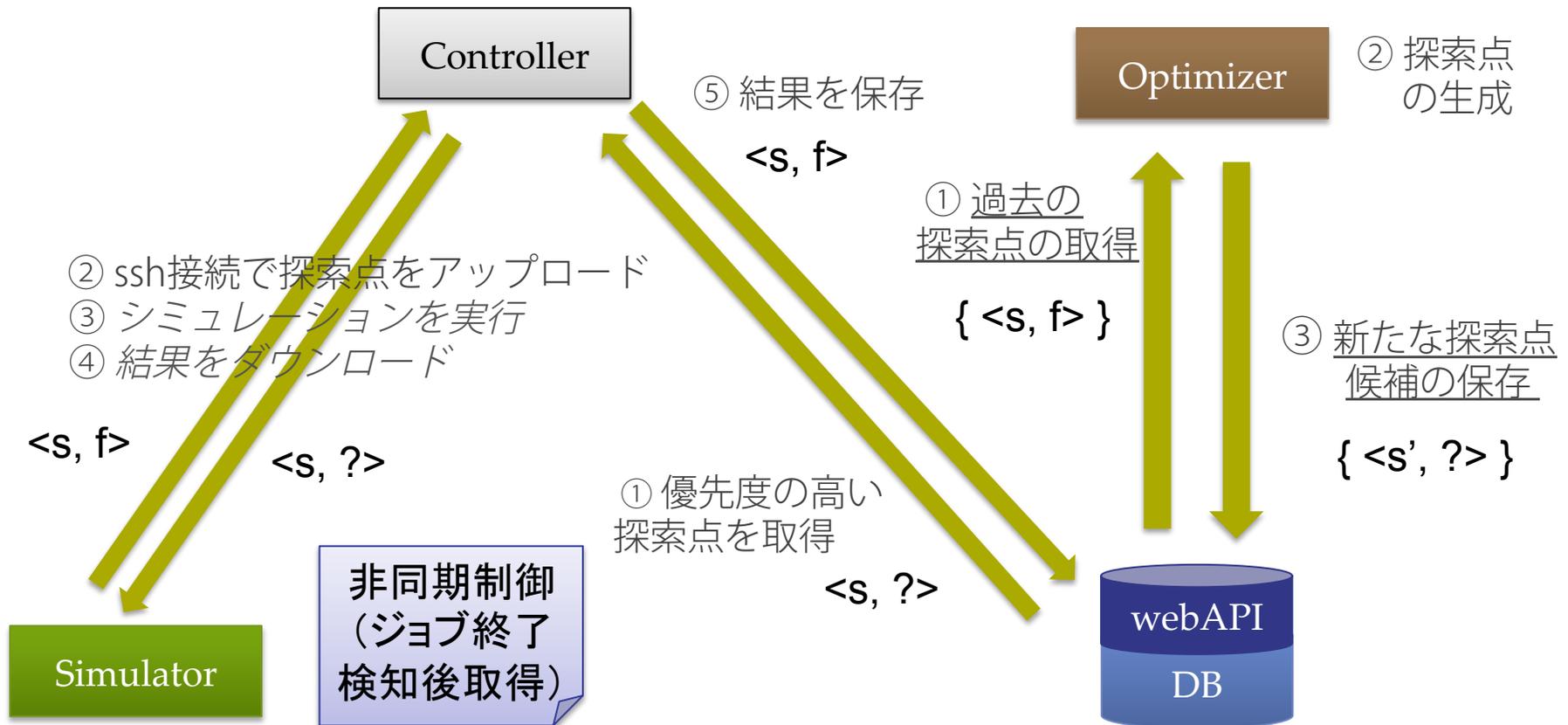
可視化装置 (ユーザインターフェイスを含む)



データベースとシミュレーション, 最適化との連携

シミュレーション実行・連携制御

最適化エンジンによる探索点候補生成



設計したwebサービスAPIの例

- 過去の解候補の取得：getIndividual
入力：設計パラメータ、取得する探索点の個数、取得方法（探索点の選択方法）
出力：評価値が計算済みの設計パラメータの情報
- 新たな解候補の保存：postIndividual
入力：設計パラメータ、優先度情報
出力：解情報データベースへの登録処理の実行ステータス
- 優先度の高い解候補を取得：getTopPriority
入力：N/A
出力：計算優先度が最も高く、評価値が未計算の設計パラメータの情報
- シミュレーションの計算結果を保存：postFitnessValue
入力：設計パラメータ、評価値の計算結果
出力：解情報データベースへの登録処理の実行ステータス



データの保存形式



ユーザーごとに1つ



問題ごとに1つ

POSTパラメータから「ユーザー」と「問題」を判断し、これらのデータを読み込む、もしくは変更を行う

column_family	super_column	row	column_name	value
user1	problem1	10111101100111...	fitness_value	311
			priority	21
		11101101110111...	fitness_value	320
			priority	20
		10101001010110...	fitness_value	222
			priority	5
	10111101100111...	fitness_value	null	
		priority	7	
	10001010010100...	fitness_value	112	
		priority	11	
	10111101100111...	fitness_value	null	
		priority	24	
user2	problem2			
	problem1			
	problem2			
	problem3			

最適化エンジンの実装例 (steady-state GA)

P <- generate a population of individuals randomly
while stopping criterion has not been met:

```
parent1 <- selection(P)  
parent2 <- selection(P)
```

(選択) 解情報データベースの計算済みのパラメータから、評価値をもとにした確率で2つの個体を選択。

```
child1, child2 <- with probability  
cross_rate crossover parent1, parent2
```

(交叉) 2つの個体のパラメータ値を入れ替える。

```
child1 <- mutate child1  
child2 <- mutate child2
```

(突然変異) 一定確率でパラメータ値を変更する。

```
P <- add child1 child2
```

新しい探索点候補として解情報データベースに加える。

```
P <- garbage individual
```

評価値の低いパラメータを削除する。



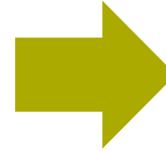
最適化エンジンと分散データベースの連携



11101101110111	fitness_value	320
	priority	20
11101101110111	fitness_value	320
	priority	20

Optimizer

problem1	10111101100111	fitness_value	311
		priority	21
	11101101110111	fitness_value	320
		priority	20
	10101001010110	fitness_value	222
		priority	5
	10111101100111	fitness_value	null
		priority	7
	10001010010100	fitness_value	112
		priority	11
	10111101100111	fitness_value	null
		priority	24

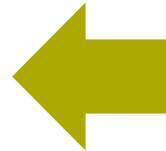


①選択

過去の解候補の取得
getIndividual

②交叉

③突然変異



新たな解候補の保存
postIndividual

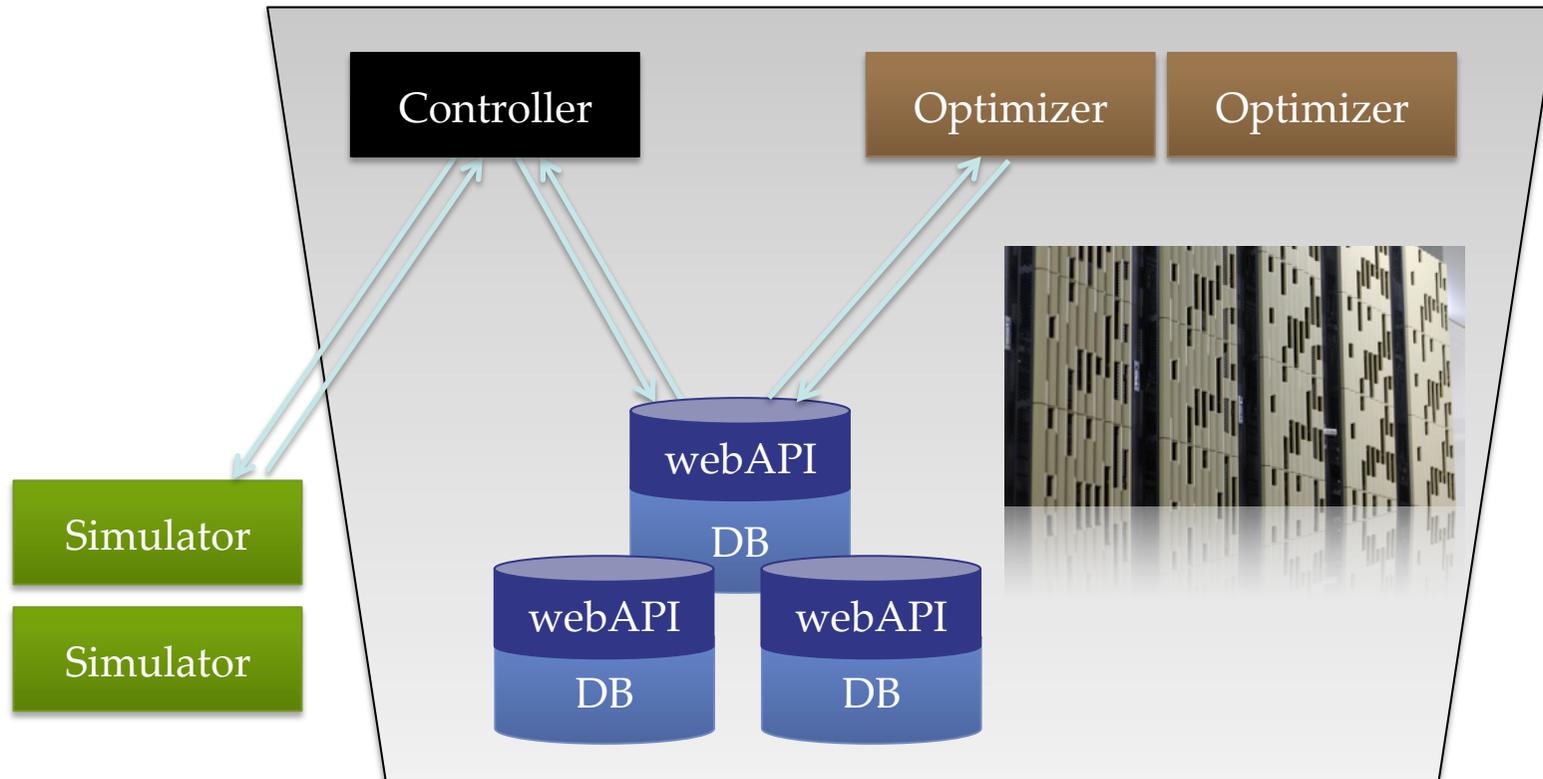
評価値の低いパラメータを削除

11101101110111	fitness_value	320
	priority	20
11101101110111	fitness_value	320
	priority	20



検証用システム（構築中）

北大アカデミッククラウド



2Lサーバー（ペタバイト級データ
サイエンス統合クラウドストレージ）



北大・九大スパコン等を利用予定



北海道大学

現状の課題と今後の方向性

- クラウド等のシステム基盤や要素技術は整備されつつあるが、それらの間の「連携」が課題（運用，研究への活用も含む）
- 特に，スパコン等の従来型システムとの連携において課題が多い。一台のスパコンに閉じていれば簡単なことが，クラウド等と連携する場合にできない。
- 設計探査の文脈において，全国規模のスパコンを連携させたシミュレーションと，インタークラウドにおけるデータ，最適化，可視化連携を組み合わせることで，特徴的な「ユースケース」を提案したい
- さらに，インタークラウド上で実現されたシステムと，研究者コミュニティとのインタラクションによる「共進化」システムとしての実現を目指したい



共同研究の公募について(JHPCN, 北大)

- 北大共同研究課題の公募

http://www.iic.hokudai.ac.jp/kyodo_kenkyu/

- 北大のスパコン, クラウドシステムを活用した研究
- 計算機使用料(～50万円)＋旅費(～50万円)の補助
- 例年, 年末～年明けに公募開始, 2月初め締切

- JHPCN共同研究課題の公募

<http://jhpcn-kyoten.itc.u-tokyo.ac.jp/ja/>

- 北大, 東北大, 東大, 東工大, 名大, 京大, 阪大, 九大の計算資源(スパコン, クラウド等)を活用した研究
- 計算機利用料の補助(成果発表旅費は採択後別途応募)
- 12月中旬より公募開始, 1月初め締切

