

課題番号 12-IS01

分散クラウドシステムにおける遠隔連携技術

棟朝 雅晴（北海道大学）

概要 北海道大学アカデミッククラウドを核として、国内各大学に分散配置されたクラウドシステムとの相互運用を実現する遠隔連携技術を確立する。そのために必要となるインタークラウドマネージャの開発を行うとともに、各拠点間のバーチャルマシン群をバーチャルプライベートネットワークで相互接続した、分散型のバーチャルマシンクラスタを構築し、Hadoop や分散ストレージなどにより評価することで、その運用技術について検討する。

1. 研究の目的と意義

本共同研究課題は、地理的に分散配置されたプライベートクラウドシステムを連携させるために必要となる技術的課題について検討、検証することを目的とするものである。学際大規模情報基盤共同利用・共同研究拠点は、大規模計算機とそれを支える汎用計算機資源が大容量ネットワークで接続されており、さらに近年、いくつかの拠点においてプライベートクラウドシステムの構築を計画している。各拠点におけるプライベートクラウドシステムを超大容量ネットワークで接続することにより、広域分散型のクラウドシステムの構築をめざし、主な技術的課題として以下の項目について研究開発を行う。

(1) ハイパーバイザソフトウェア、クラウドシステム管理ミドルウェアが互いに異なる環境における管理システム間の連携方法、特に大規模学術クラウドシステム連携を実現するインタークラウドマネージャの実現に必要な技術的課題に関する検討を行う。

(2) バーチャルプライベートネットワークとして SINET4 におけるオンデマンド L2/L3VPN サービス等を用いた、分散プライベートクラウドシステムの相互接続に関する検討、検証、接続実験を行うとともに、使用するネットワークに依存した拠点間

の伝送遅延と帯域による影響に関する検証実験を実施する。

(3) 分散配置され、相互接続されたバーチャルマシン群を用いたシステム設計法について検討する。特に、大規模分散クラウドシステム上における Hadoop 等のバーチャルマシンクラスタの構成に関する検討を行い、実験的にクラスタを構成し、その性能について評価を行う。

(4) 大規模分散クラウドストレージの実現に必要な、拠点間でのストレージシステムの連携技術について検討、検証する。

以上の技術的課題について、北海道から九州・沖縄にいたる全国規模の広域分散クラウドシステムのテストベッドを構築することを通して具体的な検討、検証を行うことで、アカデミッククラウドの連携に必要な基盤的技術の開発、および運用モデルの確立を目指すものである。

クラウドコンピューティングについて、海外では大規模なクラウドシステムによるサービスが行われているが、ハイパフォーマンスコンピューティングを実現するものはまだ少なく、さらにその利用条件、利用料金、ネットワーク遅延、セキュリティポリシーなどの問題から、国内の大学や研究機関における要求にすべて合致させることは困難である。このため、各大学、研究機関においてプライベートクラウドシステムの構築が進められ

ているところであるが、予算上の制約等により、それぞれのシステムの規模は比較的小規模なものに留まることが多く、クラウドコンピューティングによるスケールメリットが生かされないという問題が生じる。

そこで、本研究プロジェクトにおいては、各大学においてそれぞれ導入されたプライベートクラウドシステムを連携させ、大規模なアカデミッククラウドシステムを構築することで、各大学のポリシーを生かしつつ、かつスケールメリットを享受できるような分散環境を実現する。そのために必要な技術的課題について、具体的な広域分散クラウドシステムの連携を通して検討、検証するところに、本共同研究の意義が存在する。

さらに、各大学において導入されるクラウドミドルウェアなどのソフトウェア環境が異なる場合に運用方法について、特に複数のクラウドシステムの運用連携を実現するためのインタークラウドマネージャを実現することが求められるが、現状では、そのようなインタークラウドマネージャとして十分な機能を有するものは存在しない。本研究における研究開発を通して、アカデミッククラウド連携に必要な機能や性能の要件について検討するとともに、プロトタイプシステムとしてのインタークラウドマネージャを開発することで、今後の学術クラウドの発展に貢献することが期待される。

2. 当拠点公募型共同研究として実施した意義

(1) 共同研究を実施した大学名と研究体制

北海道大学、東京工業大学、国立情報学研究所、九州大学、東京大学、東京藝術大学、大阪大学、広島大学、北見工業大学、琉球大学

(2) 共同研究分野

超大規模情報システム関連研究分野

(3) 当公募型共同研究ならではの事項など

全国規模での分散クラウド連携プロジェクトとして、ネットワーク型拠点到分散配置されたクラウド資源を活用した研究開発であることが特徴的である。

3. 研究成果の詳細と当初計画の達成状況

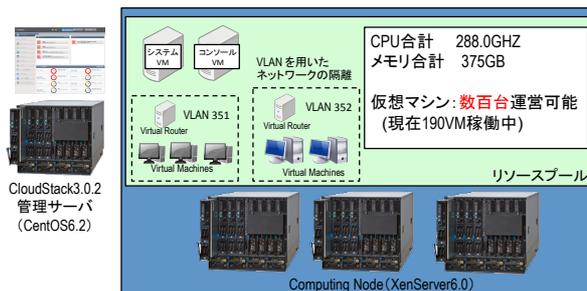
(1) 研究成果の詳細について

(1) 試験システムに関する環境整備

まずプロジェクトの実施に必要なとなる分散配置されたクラウドシステムによるインタークラウド試験システムの構築において、以下の拠点を整備を行った。

インタークラウド試験システム(北大拠点)

- 高性能ノード(40コア, Mem:128GB, 10GbE x2)を4台使用



ネットワーク構成図(北大拠点)

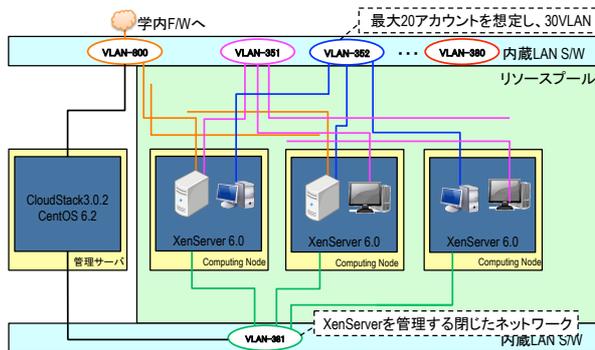


図1 北海道大学拠点におけるシステム構成

- 北海道大学拠点：情報基盤センターにおける運用システムとは別に、図1に示す CloudStack 3.0.2 ベースの試験システムを構築し、最新のク

クラウド管理ミドルウェアを導入した数百台規模の仮想マシンが実行可能なクラウド試験システムを実現した。

さらに、試験システム上において、図 2 に示す 129VM の Hadoop 環境を整備するとともに、ビッグデータ処理のアプリケーションを想定して、機械学習パッケージの Mahout やテキスト処理にひつようなライブラリなどのソフトウェアを導入するとともに、より一般的な Web アプリケーションをスケラブルに実行するためのフレームワークとして、図 3 に示される CloudFoundry による PaaS(Platform as a Service)環境についても 40VM で構築した。

Hadoop環境の整備

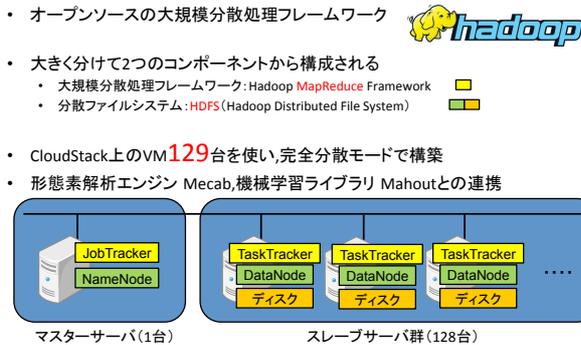


図 2 Hadoop 環境の整備

CloudFoundry環境の整備

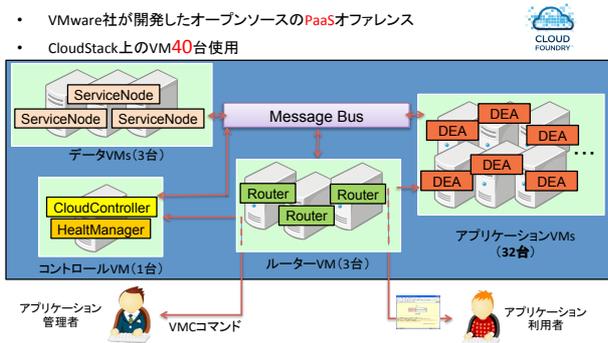


図 3 CloudFoundry 環境の整備

CloudFoundry は近年注目を集めている PaaS 環境を実現するためのミドルウェアであり、利用者がアプリケーションを vmc コマンドによりプシ

ユするだけで、Web アプリケーションが実行可能となる。さらに仮想マシンの数を動的に制御することで、アプリケーションに対するアクセスの増大に対してスケラブルに対応することが可能となる点が特徴的である。

- 北見工業大学拠点：遠隔連携のために、SINET 直結により接続されている試験システム環境を構築している。現在、本拠点では、管理サーバ用と計算ノード用に各 1 台動作している。その環境では、計算ノードを XenServer 6.0.2 により実装している。それを管理するためのサーバは、CentOS 6.2 により構築されており、その上で CloudStack 3.0.2 により構築された管理用コンソールが実装されている。本拠点では、/25 の範囲で、グローバル IP アドレスを VM に割り当てることができるので、今後大規模な遠隔連携試験環境にも対応可能となっている。

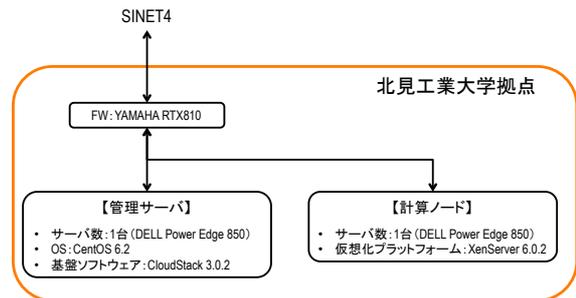


図 4 北見工業大学拠点のシステム構成

図 4 にそのシステム構成図を示す。琉球大学においても北見工業大学と同様の構成で、CloudStack 3.0.2 ベースの環境を整備した。

- 東京工業大学を中心に整備が進められている HPCI-AE (HPCI 先端ソフトウェア運用基盤) についても連携して環境整備を進めており、東京工業大学、東京大学、北海道大学、九州大学におけるシステム構築作業を行った。

- 国立情報学研究所においては、OpenStack ベースの研究クラウドシステムを運用しており、さら

に gunnii+tinii という物理マシンも取り扱え、既存資産を活用できるクラウドシステムの実現に向けた研究開発を行っている。gunnii+tinii の検証実験として、北海道大学⇄国立情報学研究所間で SINET-VPLS 網を経由した遠隔連携試験を実施しているところであり、遠隔サイトへの資源の拡張を実現するための技術的基盤の検証を行った。

(2) インタークラウドマネージャの実現に必要な技術的課題に関する検討

インタークラウドマネージャの実現に必要な技術的課題について、本年度前半においては、その認証連携基盤に関する検討を中心に実施した。異なるクラウド管理ミドルウェアで制御されている複数のクラウドシステムを連携させるため、それらの間の認証連携が必須となる。

単にクラウドポータルへのログイン認証を共通化するだけであれば、Shibboleth の標準的な機能で実現できるが、それぞれ異なるクラウド管理ミドルウェアを API レベルで制御するために必要な認証については、その API キーを含めた連携が必要となる。

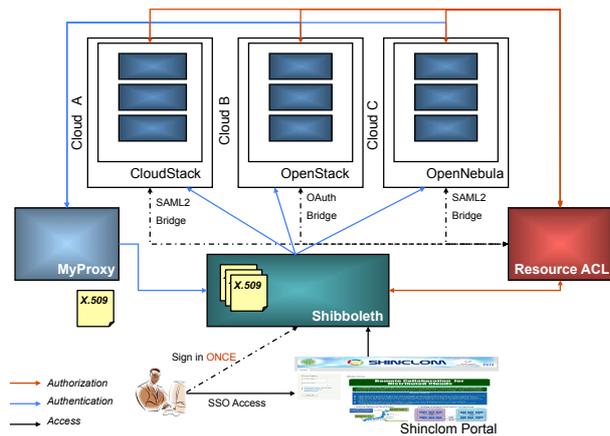


図 5 異なるクラウド管理ミドルウェア間の認証連携方式の検討：概要

図 5 に本研究課題において検討およびプロトタイプの実装を行っているシステムの概要を示す。インタークラウドマネージャ SHINCLOM (Simple Heterogeneous INter-CLOUD Manager) のポータル

に対してシングルサインオンすることで、Shibboleth と連携した認証処理を行う。CloudStack, OpenStack, OpenNebula などそれぞれのクラウド管理ミドルウェアへの API レベルでのアクセス制御に必要な API キーについては MyProxy との連携により認証キーを発行、管理することで、それぞれ異なる管理ミドルウェアの API キーを一括して管理し、利用することを可能とする。Resource ACL (Access Control List) の制御機構によりそれぞれのクラウド管理ミドルウェアの配下にある資源のアクセス管理を実現するものである。

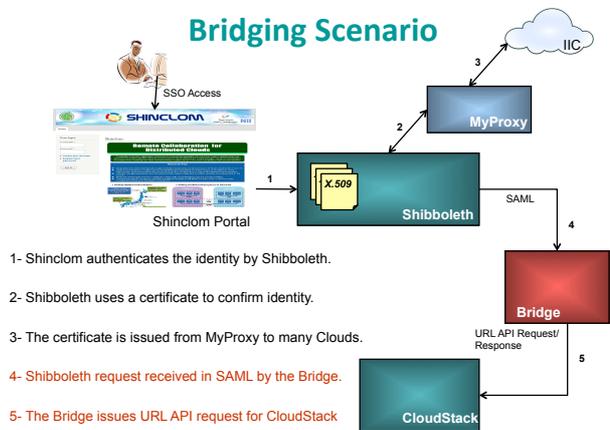


図 6 Shibboleth と CloudStack における認証ブリッジの制御手順

OpenStack においては、Keystone により API キーを管理することが可能となりつつあるが、それ以外のクラウド管理ミドルウェアには現状では対応していない。そこで、北海道大学アカデミッククラウドで採用されている CloudStack における API キー管理のための認証ブリッジ機構を実装し、検証を行っているところである。具体的な制御手順を図 6 に示す。まず、インタークラウドマネージャが Shibboleth により認証を行った後、MyProxy により発行された証明書がクラウド管理ミドルウェアに対して送信され、それを元に認証ブリッジを通してクラウド管理ミドルウェアに対する API アクセスの制御が行われる。

認証ブリッジの構成案を図 7 に示す。

Shibboleth, SAML に関する処理については共通してあり、それらと連携した CloudStack 対応の JAVA のドライバーも含めて開発した。それぞれ異なるクラウド管理ミドルウェアに対応するには、それぞれに対応する JAVA のドライバーを開発することで実現可能である。

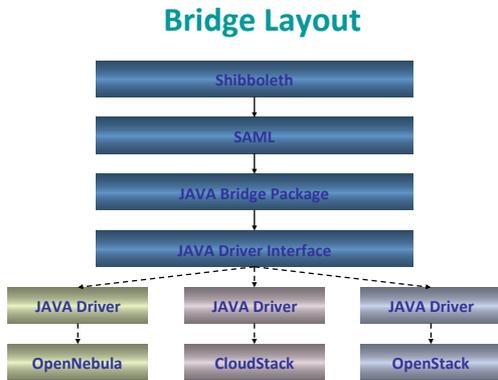


図 7 認証ブリッジシステムの構成案

(3) ネットワークに関する連携については、昨年度から継続して、CloudStack 標準機能として用意されているソフトウェアルータの L2VPN を用いた相互接続試験を実施するとともに、Software Defined Network (SDN) に関する検証として、Virtual Distributed Ethernet (VDE) を用いた相互接続を北海道大学と北見工業大学 (図 8)、さらには Amazon Web Services (AWS) の Tokyo Region と North Virginia Region を加えた構成で検証した (図 9)。

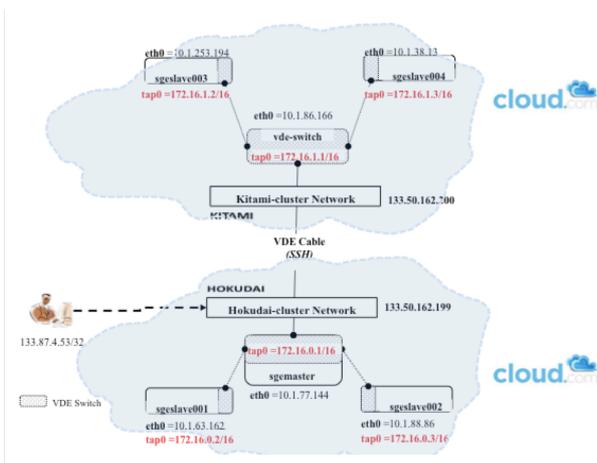


図 8 北海道大学⇄北見工業大学間の仮想ネット

ワーク (VDE) による相互接続環境

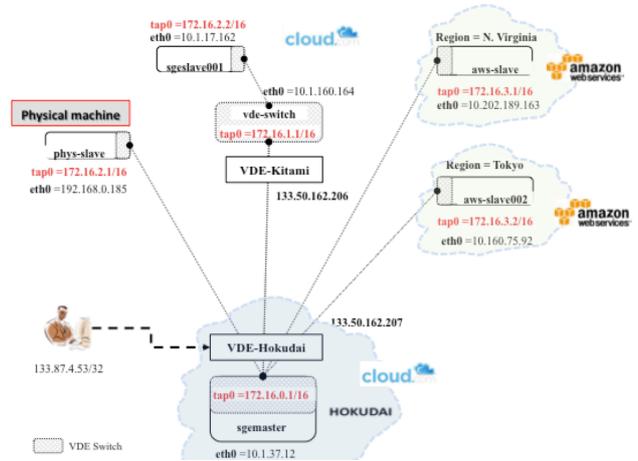


図 9 北海道大学⇄北見工業大学⇄AWS (Tokyo, North Virginia)間での仮想ネットワーク (VDE) による相互接続環境

その結果として、図 10 に示す通り北大と AWS の間ではインターネット経由で遠隔地との通信を行うために、帯域幅が大幅に低下しているが、SINET4 経由を想定した北海道大学⇄北見工業大学については、実用的な帯域幅を確保できていることが分かった。

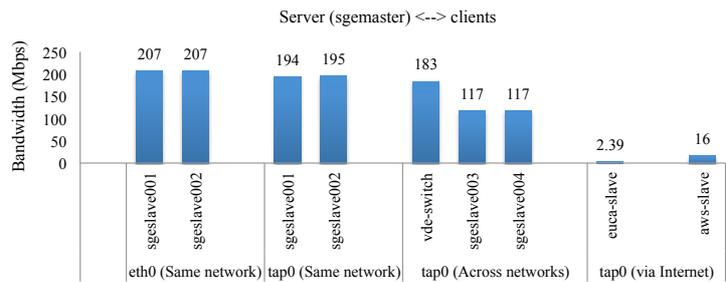


図 10 帯域幅 (Mbps) に関する測定結果 (左から、北大拠点内 (VDE 不使用)、北大拠点内 (VDE 経由)、北大⇄北見工大、北大⇄AWS)

(4) インタークラウド環境下におけるバーチャルマシクラスタの自動構成技術に関する検討

インタークラウドシステム環境下において、多数のバーチャルマシンをクラスタとして自動的に

構成するために必要となる技術的課題について検討、検証を行った。

図 1 1 に、インタークラウドシステム環境下におけるバーチャルマシンクラスタの自動構成を実現するシステムアーキテクチャの全体像を示す。それぞれ異なるクラウド管理ミドルウェアの違いを吸収するため、CPSL (Cloud Provider Specific Layer)を設定し、Cloudinit.d を用いて複数のクラウドの API 制御を担当し、バーチャルマシンや相互接続のために必要となる仮想ネットワークの制御などを行うレイヤーとして設計を行うものである。

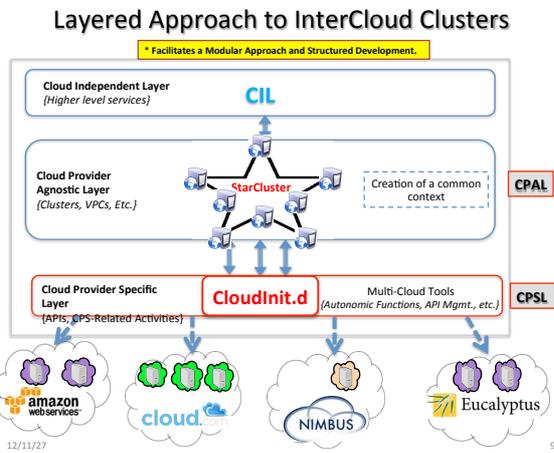


図 1 1 インタークラウド環境下でのバーチャルマシンクラスタの自動構成：概要

次に、CPAL (Cloud Provider Agnostic Layer) については、クラスタや VPC (Virtual Private Cloud) などの利用者が必要とする仮想的なシステムを制御するレイヤーとして定義する。ここでは仮想的なシステムの構築を担当し、クラスタ構築に必要な情報の管理を行うものである。

さらに、最上位のレイヤーとしては CIL (Cloud Independent Layer) を定義し、ここではクラウドは独立に構築された仮想的なシステムの設定や制御などサービスレベルでの情報の管理を担当する。

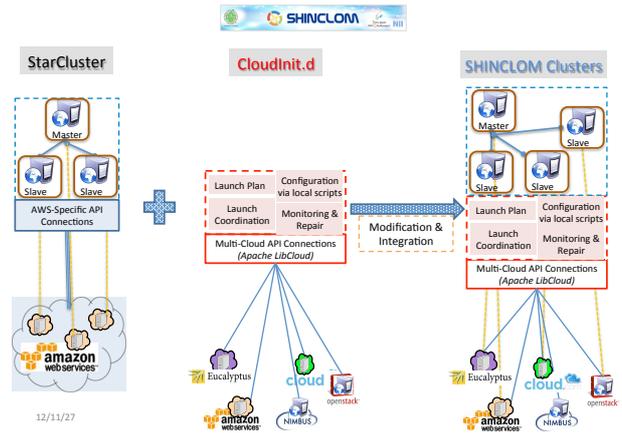


図 1 2 インタークラウドマネージャとバーチャルマシンクラスタ構成システムの融合

クラスタシステムの設定、構成にあたっては既存のクラスタ設定ソフトウェアを利用可能であり、それらと、クラウド基盤を制御する Cloudinit.d と呼ばれる制御ソフトウェアとの連携により互いに異なるクラウド管理ミドルウェアの配下にある複数のクラウドシステムを一括して制御することが可能となる。

(2) 当初計画の達成状況について

インタークラウドマネージャのプロトタイプ開発およびその検証については、認証基盤との連携方式も含めて、当初の目的をほぼ達成しており、基礎的な技術開発に関する検討、検証を完了した。今後の検討としては運用システムとして利用可能なプロダクトレベルへの展開について、民間などとの連携も含めて今後検討を進める予定である。

バーチャルプライベートネットワークなどネットワーク仮想化については、CloudStack の機能向上やさまざまなネットワーク仮想化ソリューションを利用することが可能になりつつあるが、本研究課題では、VDE をベースとした運用上の課題について検討、検証をすすめることで、その典型的なユースケースである遠隔サイト間でのバーチャルマシンクラスタの構築も含めた評価を行うことができ、当初計画を順調に達成している。

バーチャルマシンによるクラスタの構築につい

では、既存の技術を活用しつつ、インタークラウドマネージャとの連携によるクラスタシステムの自動構成技術について基礎的な検証を完了しており、当初の予定通りの成果をあげている。

大規模分散ストレージに関する検証実験については、現在広島大の拠点との連携試験について検討をすすめているところであるが、検討段階にとどまっており、今後の課題となっている。さらに、Mongo DB などの大規模分散データベースについても検討を始めたところであり、CloudFoundry と連携した大規模 Web アプリケーションにおける性能評価が今後の課題となる。

4. 今後の展望

平成 25 年度以降の展望として、北海道大学アカデミッククラウド等の運用システムへの展開を計画している。今後の学術クラウド連携の方向性として、大規模な大学においては共同利用サービスとしてのパブリック（コミュニティ）クラウドシステムが構築、運用され、小規模な大学においては最小構成のプライベートクラウドを構築して、拠点大学の共同利用サービスとしてのクラウドと連携することが想定される。北海道大学アカデミッククラウドと、各大学のプライベートクラウドシステムの連携ポータルを構築する上で、そのユースケースの確認や、相互運用上の課題について検討するためのプロトタイプシステムとして今後、本研究課題の成果を活用する予定である。

さらに、インタークラウドシステムにおける CloudFoundry 等の PaaS 環境の実現や、分散ストレージ・分散データベースの実装等、大規模分散アプリケーションの構築を意識したシステム設計および連携試験の実施についても、今後検討をすすめていく計画である。

5. 研究成果リスト

- (1) 学術論文（投稿中のものは「投稿中」と明記）
- (2) 国際会議プロシーディングス

[1] Omar Abdul-Rahman, Masaharu Munetomo, Kiyoshi Akama: Toward A Genetic Algorithm Based Flexible Approach for the Management of Virtualized Application Environments in Cloud Platforms, Proceedings of the 21th International Conference on Computer Communications and Networks (ICCCN2012), (CD-ROM) IEEE (2012)

[2] Shinichiro Takizawa, Satoshi Matsuoka, Masaharu Munetomo, Taizo Kobayashi, and Hideyuki Jitsumoto: A Virtual Machine Hosting System on e-Science Cyberinfrastructure, Proceedings of the IWCCA2012 (2012)

(3) 国際会議発表

[1] Omar Abdul-Rahman, Masaharu Munetomo, Kiyoshi Akama: Toward A Genetic Algorithm Based Flexible Approach for the Management of Virtualized Application Environments in Cloud Platforms, The 21th International Conference on Computer Communications and Networks (ICCCN2012) (2012)

[2] Shinichiro Takizawa, Satoshi Matsuoka, Masaharu Munetomo, Taizo Kobayashi, and Hideyuki Jitsumoto: A Virtual Machine Hosting System on e-Science Cyberinfrastructure, IWCCA2012 (2012)

[3] Masaharu Munetomo: Hokkaido University Academic Cloud: Largest Academic Cloud System in Japan, Cloud Technical Leadership Forum, (2012)

(4) 国内会議発表

[1] 棟朝 雅晴: 全国規模の学術インタークラウドシステムの構築, アカデミッククラウドシンポジウム 2012@北海道大学 (2012)

[2] 横山 重俊: アカデミックコミュニティクラ

ウドの実現に向けて，アカデミッククラウドシンポジウム 2012@北海道大学（2012）

[3] 滝澤 真一郎：HPCI 先端ソフトウェア運用基盤の設計と構築，アカデミッククラウドシンポジウム 2012@北海道大学（2012）

[4] 伊東 栄典，九大キャンパスクラウドの現状と課題，アカデミッククラウドシンポジウム 2012@北海道大学（2012）

[5] 小林 泰三，UBIM：シンプル且つコンパクトな広域分散環境の管理運用機構，アカデミッククラウドシンポジウム 2012@北海道大学（2012）

[6] 熊谷 悠平，認証フェデレーションに基づく分散ファイル管理システムの開発，アカデミッククラウドシンポジウム 2012@北海道大学（2012）

[7] 棟朝 雅晴：北海道大学アカデミッククラウドにおける大規模データ処理環境について，2012 年度統計関連学会連合大会（2012）

[8] 棟朝 雅晴：アカデミッククラウドとビッグデータ，第 1 回ビッグデータと統計学研究集会（2012）

(5) その他（特許，プレス発表，著書等）