社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

OpenStackを用いたクラウド間のデータ転送手法の検討

西出　彩花†　　小口　正人†

†

あらまし　近年の情報技術の発達に伴い，様々なデータ量が日々増え続けている．これに伴い企業のシステムは，「変動する可能性の高い，一般的なデータ」をコストや導入スピードに優れるパブリッククラウドに保管し，「会社内部の機密情報を含むデータ」をプライベートクラウドに保管するような，プライベートクラウドとパブリッククラウドで提供されるサービスを組み合わせた「ハイブリッドクラウド」の形へと進んでいる．このハイブリッドクラウド環境では，従来プライベートクラウドで構築，運用されていた様々な業務システムが，複数のクラウドに分散配置されることになる．その際にクラウド間でのデータベースの冗長的な同期が必要不可欠である．また，日本は地震などの自然災害の影響を受けることが多い．災害時に重要なデータを失わないためには，冗長的に遠隔地にバックアップを行っておく必要がある．そこで本研究では，通常時の冗長的なデータベースのバックアップに加え，緊急時には外部情報をトリガに，データベース処理を実行するインスタンスのマイグレーションを行うことにより，データベースアクセスの継続を実現するための手法を検討する．

キーワード　オープンスタック，クラウド，Pangea

Data contorol with OpenStack between clouds

Sayaka NISHIDE† and Masato OGUCHI†

†

Abstract　These days, we use a lot of amount of data on computer systems. It makes many company to use "hybrid cloud" which is focusing on only the good points or strong points of "public cloud" and "private cloud". In addition, we have many earthquakes. At the time of earthquake, we might lose impotant information if we put them only on one data center. To avoid this, I try to construct the system which can migrate impotant data before we lose them.

Key words　OpenStack，CloudComputing，Pangea

1. まえがき

近年，コンピュータシステムにおける情報量が爆発的に増加している．その処理プラットフォームとして，クラウドの利用が注目を集めている [1]．中でも，変動する可能性の高い一般的なデータをパブリッククラウドに保管し，会社内部の機密情報を含むデータをプライベートクラウドに保管する，といった，複数のクラウドを効率的に使い分ける環境である，「ハイブリッドクラウド」に特に着目する人が多い．

ハイブリッドクラウドを利用する際には，パブリッククラウドとプライベートクラウド間で，データベースの冗長的な同期が必要である．一般にパブリッククラウドとプライベートクラウドは別の場所に構築され，多くの場合，両者の距離は離れている．従って両者間でデータベースの遠隔同期や遠隔バックアップが行われることになる．

一方で，日本は火山やプレートに囲まれており，地震などの自然災害の影響を受けることが多い．2011 年の東日本大震災で多くのデータが失われた [2] ことなどから，災害発生時には迅速にデータを災害地から別の場所へ移す必要があることがわかる．このとき，災害地付近でバースト的に大量のデータが転送されることが考えられる．

また，SNS や高機能デバイスの普及に伴い，災害地にはバースト的なアクセスの増加が起こることが考えられる．これは，ネットワークに大きな負荷をかけ，サーバなどの機能の低下を引き起こす．この際に安全にデータを転送するためには，帯域が不足する経路を迂回してパケットを別の経路に振り分けたり，マイグレーションの性能を上げたりする必要がある．そこで本研究では，データベースの冗長的な遠隔バックアップと，災害時の外部情報をトリガとするマイグレーションによって継続的なデータアクセスを実現するシステムを提案する．

OpenStack [3] [4] を用いて仮想環境上にクラウド基盤を構築し，データベースの遠隔バックアップを動作させて，インスタ

ンスマイグレーションを行うことにより，提案システムを実現する．さらに OpenFlow のスイッチ制御で，マイグレーション性能を向上させることにより，緊急時に確実なデータ転送を行う．

OpenStack とは，クラウドを構成する仮想マシンや物理サーバの運用管理を実行し，それを効率的に行うためのオープンソースのクラウド構築ソフトウェアである．OpenStack の利用者は，KVM などで構成されるハイパーバイザ上で動作する仮想マシンに外部ネットワークからアクセスし，CPU，メモリ，HDD，IP アドレス等の計算資源を利用することができる．OpenStack は複数のコンポーネントから構成され，これらのコンポーネントが連携することで IaaS のサービスを提供するアーキテクチャである．OpenStack の構成を図 1 に示す．

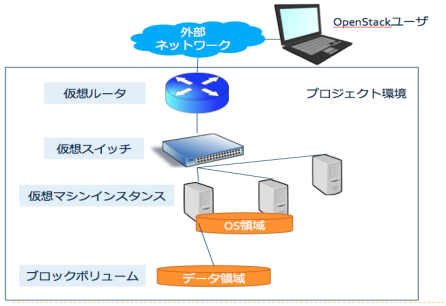

図 1　OpenStack の構成

2. 関 連 研 究

2.1 Pangea

ここで本研究で用いる Pangea [5] について説明する．Pangea は，NTT 研究所で開発されている，LAN 環境を前提としたデータベース同期ミドルウェアである．コンシステンシを保ちながら，複数台のサーバでデータベースのクエリ処理を並列実行する事により，性能向上を実現した．これら複数台のサーバは，同一のデータベースイメージを保持しており，同一 LAN 上に接続されている．サーバの 1 台を Leader，その他は Follower としており，クライアントからこのミドルウェアを介してサーバにアクセスをして同期をとる．全てのクエリは照会処理と更新処理に分類され，照会処理はどれか 1 台のサーバで，更新処理は全てのサーバで実行される．更新処理の場合は Leader に対して更新をした後に，Follower に対しても同様に処理を行う．

2.2 Pangea**

次に Pangea** [6] について説明する．Pangea**とは，前述の Pangea をベースに開発された，リモートバックアップ機能を有したクラスタデータベースシステムのことである．Pangea は LAN 環境では非常に良い性能であるが，遠隔地にあるサーバにバックアップをとろうとすると，遅延の影響を多く受けてしまい大幅な性能低下をもたらすことが予想される．これを受けて，非同期バックアップを取り入れた Pangea**が開発されている．Pangea**では，照会処理と更新処理に区別された後にこの処理をマスタとスレーブに分離して，遠隔地に配置したスレーブへは非同期データ転送とし，さらにスレーブから遠隔ストレージへのアクセスを並列化することにより，遠隔バックアップを行ってもデータベース処理性能が落ちないようになっている．この様子を表したものが図 2 である．遠隔地でバックアップをとることで，災害時のデータ損失を防ぐことができる．Pangea**は 2 台のサーバに展開する．本研究ではまず Pangea をミドルウェアとして利用する．

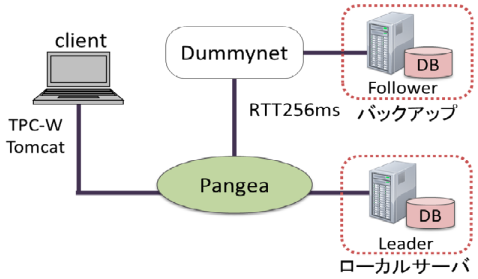

図 2　Pangea**

2.3 外部情報を用いたトラフィックの制御

本研究では，緊急時のトラフィック制御にも着目している．災害時のバースト的な負荷変動はネットワークに大きな影響を与える．そこで重要となるのが，外部情報をトリガとして緊急時に自動でデータを転送するシステムを用意しておくことである．

本研究では Twitter などのソーシャル情報から関連する情報を引き出し，それを元に分析を行うことで，負荷のかかるノードや障害発生場所の情報を入手し [7]，これらを元に Open vSwitch [8] を用いて通信経路の制御を行う [9]．

本研究では，コントローラの代わりに RabbitMQ に notify キューを作り，ここで Open vSwitch をコントロールする．

2.4 負荷分散による効率的なデータマイグレーション

データマイグレーションを行うことは，緊急時のデータ保護と災害地付近のバースト的なアクセス負荷の分散に有用であると考えられる．しかし同時に，マイグレーションに伴うディスクアクセスやネットワーク転送により，システム性能を一時的に低下させ得る．

この性能低下を抑えることを目的とし，データマイグレーション時のみ複製データへアクセス要求を一部回送する手法を，2004 年に東京工業大学の小林らは発表している [10]．これにより負荷集中時の円滑なデータ移動が可能となる．

また彼らは 2007 年までに，ストレージノード中に格納されるバックアップデータを用いたデータの移動経路の変更とアクセスの回送により，円滑にマイグレーションを行う資源を確保する手法について [11] [12] も発表している．

これを大規模災害時のマイグレーション時にも適応させることが可能であると考えられる．

3. 提 案 手 法

3.1 クラウドにおけるインスタンスマイグレーション

OpenStack ではコンポーネントの 1 つである Nova のマイ

グレーション機能を用いることが可能である．Nova のマイグレーション機能は大きく分けて以下の 3 つである．

（1） コールドマイグレーション

停止している仮想マシンを他のノードに移動させる．

移動先は nova-scheduler により自動的に指定される．

- チャンススケジューラ

nova-compute が稼働しているノードから任意の 1 台をランダムに選択する．

- フィルタスケジューラ

「フィルタ」と「重み」の 2 種類の条件で，最適なノードを選択する．

このコールドマイグレーションが Horizon による GUI からマイグレーションを行った時の標準の仕様となっている．

（2） ライブマイグレーション

稼働している仮想マシンを他のノードに移動させる．

共有ディスクの利用が必須である．

（3） ブロックマイグレーション

稼働している仮想マシンを他のノードに移動させる．

ディスクとメモリのみをコピーするので共有ディスク環境は必要がない．

本研究では，コールドマイグレーションを中心に実験を行ったが，将来的にはブロックマイグレーションへの応用が有用であると考えられる．

3.2 本研究でのクラウドの利用

本研究では複数のクラウドを用いて，異なるクラウド間でのマイグレーションに着目する．外部情報をトリガに，プライベートクラウドからパブリッククラウドへ仮想マシンをマイグレートする場合について考える．

これを図 3 に表す．共有ストレージとの間に Pangea**を介することで，マイグレーションの際に遠隔地にあるクラウド内のデータベースへのアクセスを行う必要がなくなるため，負荷の軽減が予想される．

まず，元々仮想マシンが置かれていた compute1 上の DB1 と仮想マシンを切り離す．DB1 の上のデータでパブリッククラウドからもアクセスが可能であるべきだと判断されたデータに関わるクエリは Pangea**をベースとするミドルウェアを介し，共有ストレージにバックアップされる．DB2 がここにアクセスすることで，必要なデータを取得することができる．同時に OpenStack クラウド上のマイグレーション機能を用いて仮想マシンをプライベートクラウドからパブリッククラウドへマイグレートする．

これにより仮想マシンとデータベースのクラウドを跨いだマイグレーションが可能となる．

4. OpenStack を用いた実験

4.1 実験環境

本研究で想定するクラウド環境を，IaaS のクラウド環境構築ソフトウェアの OpenStack（Icehouse）を用いて構築した．12 台の物理サーバを用意し，6 台ずつを用いてクラウド環境を構築し，それぞれをパブリッククラウドとプライベートクラウドとする．本研究ではパブリッククラウドには 2014 年 4 月リリースの Icehouse を用いて，コントローラノード 1 台，ネットワークノード 1 台，コンピュートノード 4 台の計 6 台のサーバからなるクラウド基盤を構築した．コントローラノードには RabbitMQ,NTP，Keystone，Glance，Nova，Neutron，Cinder，Ceilometer，Swift，Heat，セキュリティグループ，フロントエンドを，ネットワークノードには Neutron を，コンピュートノードには Nova，Neutron，Ceilometer，Swift を配置した．

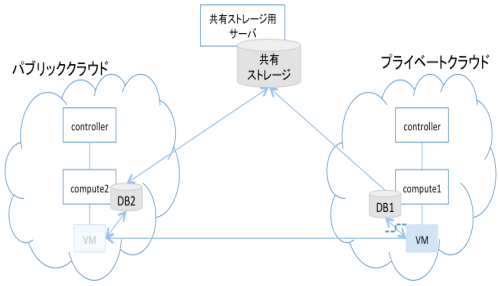

図 3 クラウドを挟んだデータベースマイグレーション

一方で，プライベートクラウドには 2015 年 10 月リリースの Liberty を用いてコントローラノード 1 台とコンピュートノード 4 台の計 5 台のサーバからなるクラウド基盤を構築した．コントローラノードには RabbitMQ, NTP, MariaDB, Keystone, Glance，Nova，Neutron，Linux Bridge Agent，L3 Agent，DHCP Agent，Metadata Agent，Cinder を，コンピュートノードには Linux KVM，Nova Compute，Linux Bridge Agent を配置した．

そのほかにマイグレーション時に共有ストレージとするためのサーバを 1 台用意した．この 2 つのクラウド間でリソースの転送を行うことで異なるクラウド構築環境をまたいだハイブリッドクラウドの環境が実現される．構築に用いた物理サーバのスペックを表 1 に，構築した環境を図 4 に示す．

表 1 使用サーバのスペック

OS	Linux3.13.0-43-generic Ubuntu14.04.3 64bit
CPU	Intel(R)Xeon CPU E3-1270V2 @3.50GHz 4C/8T
Memory	16GB
Disk	500GB

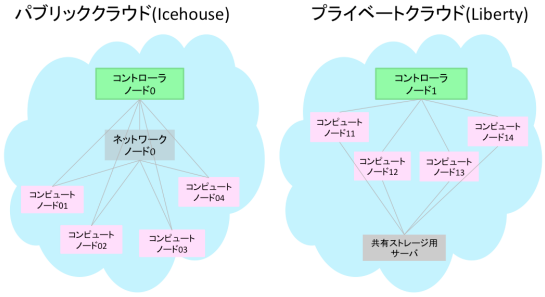

図 4 実験環境 (物理サーバ)

4.2 基本性能検証

まずパブリッククラウド内で，インスタンスマイグレーションに関する性能の検証を行った．3 章で述べた提案手法において，インスタンスのマイグレーションは重要な役割を持っており，その性能が提案手法のパフォーマンスに大きな影響を与える．

複数のコンピュートノード下に複数の仮想マシンを作り，一方からもう一方へ Iperf [13] コマンドを用いてパケットを流しながらマイグレーションを行った．ライブマイグレーションの際にはマイグレーション先の指定が可能であるが，そうでないときには OpenStack 内の Nova のスケジューラ機能がマイグレーション先のノードを指定する．本研究ではコールドマイグレーションを行うため，マイグレーション先のノードはスケジューラ機能に委ねることとなる．この実験のイメージ図を図 5 に表す．

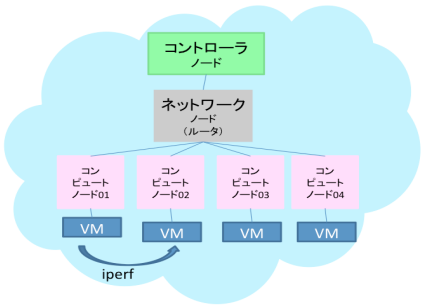

図 5　実験イメージ図

結果を図 6 に示す．

横軸の数字のコンピュートノード上にある仮想マシンをライブマイグレーショする際に，マイグレートされた先のコンピュートノードを縦軸に，またそのマイグレーションにかかった時間（秒）を縦軸に示した．この結果から，パケットを流さずにマイグレーションを行った際には平均 26.3 秒程度だったものと比較することから，対象のマイグレーションに関係のないパケットについて，他の経路を経由する様なパケット制御をすることによるマイグレーション効率の向上が期待できることが分かった．

また，本研究では compute11 上に共有ディスクを置き，ここに NFS マウントすることでマイグレーションを行うように環境を構築した．これが構築時に最も安定する構成であった．compute11 から，または compute11 へのマイグレーション時間が極端に短いため，compute11 に共有ディスクが置かれていることが実験結果に与える影響が大きいことがわかる．

このことから，共有ディスクを介するマイグレーションを必要最小限にすることで，通信性能を安定化させることが可能であると考えられる．

4.3 OpenStack への Pangea**の導入

クラウド上で Pangea を用いる性能考察を行うために, 構築したクラウド環境内に仮想マシンを立ち上げる．OpenStack においては，コントローラノードから指示を出し，この指示に従ってコンピュートノード上でインスタンスが起動される．本実験

VM@12->VM@11 パケット送信

	11	12	13	14
13	17.57	65.87		47.65
14	16.87	57.54	60.56	

VM@12->VM@13 パケット送信

	11	12	13	14
11			55.81	
13	22.11	58.25		105.58

VM@11->VM@13 パケット送信

	11	12	13	14
11		55.65	57.06	56.37
13	17.76	52.99		52.65
14			57.15	

図 6　VM 間で Iperf でパケットを送信しながらマイグレーションを行った結果

では OpenStack のコンピュートノード 4 台のうち，1 台をクライアントサーバ，1 台を Pangea 配置サーバ，2 台をデータベース配置サーバとする．この際，仮想マシンのインスタンスとしては，ダウンロードした Ubuntu14.04 のインストールイメージを元に，80G のディスク領域を使用した Linux OS が立ち上がるように設定する．

データベース配置用のサーバに PostgreSQL サーバをインストールし，クライアントサーバでは TPC-W 用の Tomcat を動かし，Pangea 配置用のサーバ上で Pangea を立ち上げる．今回はこの環境下でデータベース間での転送を行い，スループット値やレスポンス時間を計測することで，Pangea の利用による転送の性能の低下が起こらず Pangea の導入が有用であることを確認する．構築した環境の物理イメージを図 7 に，作成した仮想マシンのネットワークイメージを図 8 に示す．

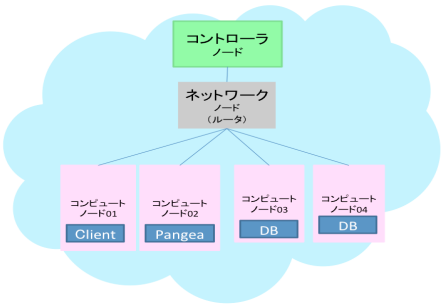

図 7　物理イメージ図

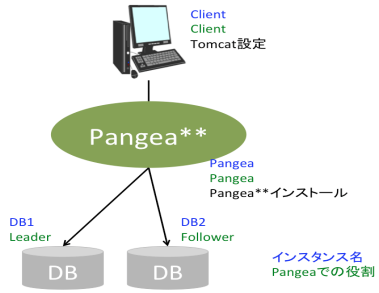

図 8　仮想マシンイメージ図

4.4 実 験 結 果

4.3 節で紹介した環境下で性能検証を行う．TPC-W におけ

る仮想ブラウザ数 EB（クライアントからの負荷量）を変化させた時のスループット値を図 9，レスポンス時間を図 10 に示す．

インスタンスを介さず物理サーバ上で直接同じ実験を行った結果である図 11，12 と比較すると，最大スループット値が 39%低下するなどの大きな性能の低下がみられる．

この要因として，今回の測定環境ではメモリサイズの大きさが仮想マシンと物理サーバで大きく異なることが考えられる．これは今後の測定の際に仮想マシンと物理サーバのメモリサイズを調整することで，正確な比較ができるようになる事が期待される．

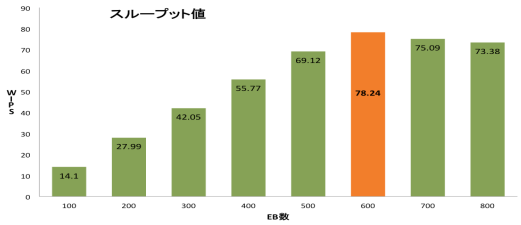

図 9　VM 上で Pangea を動かす際のスループット

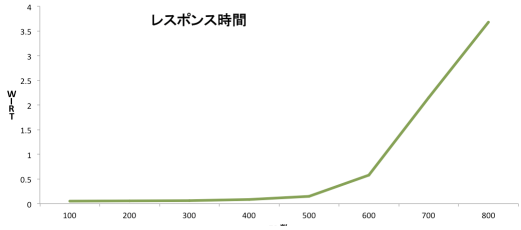

図 10　VM 上で Pangea を動かす際の応答時間

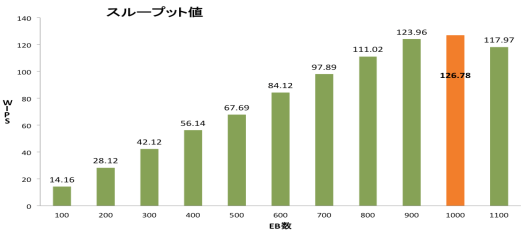

図 11　物理サーバ上で Pangea を動かす際のスループット

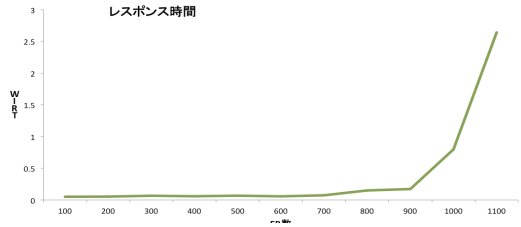

図 12　物理サーバ上で Pangea を動かす際の応答時間

また，仮想化によるオーバヘッドも性能低下の要因の一つであると考えられる．そこで，DB までクライアントサーバと Pangea サーバと共にインスタンス上に置いていたため，ストレージアクセスのコストが高くなっていたと考えられることから，これを物理サーバに直接置き，同様に実験を行った．この実験結果を図 13，図 14 にあらわす．

この結果から，仮想マシンが用いるデータベースを外部の物理サーバ上に置くことで，インスタンスをクラウド上に置く際におこる仮想化によるオーバヘッドを軽くすることができると考えることができる．なお，各データベースをクラウド上に配置するのか物理サーバ上に直接配置するかのは，そのデータベースが用いるデータの特性に依存する．

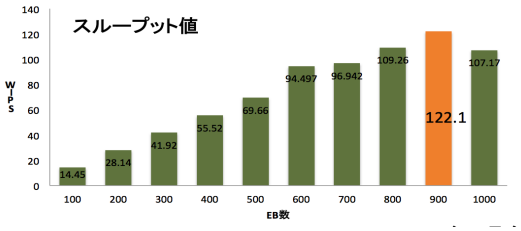

図 13　システムのみをクラウド上に置き Pangea を動かす際のスループット

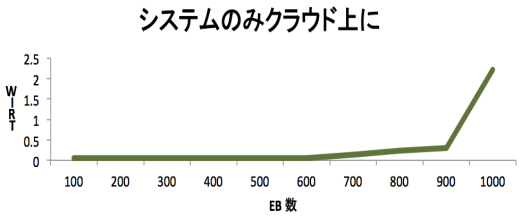

図 14　システムのみをクラウド上に置き Pangea を動かす際の応答時間

5.　まとめと今後の課題

OpenStack(Icehouse) を用いて構築した環境内での仮想マシンマイグレーションの性能検証を行った．

この結果から，汎用的な利用には相応しくないと判断し，ブロックマイグレーションを行う環境を整えるために OpenStack の最新バージョンである Liberty を用いて共有ストレージを用いないデータベースマイグレーションについて検討している．

今後は，まず第一に，Pangea の利用により容量の大きなデータベースの転送の効率が上がることを確認する必要があるためこれを行う予定である．また，DB1 と DB2 の間にダミーネットを挿入することで，遠隔地でのマイグレーションの性能についても再考する必要があると考える．本稿では，データベースサーバ内の全てのデータを同期することを前提としている．しかし，実際にハイブリッドクラウドとしてクラウドを利用する際にはセキュリティの問題などから，パブリッククラウドとプライベートクラウドに保存されるデータは一部異なることが予想される．そのため，要求されたクエリがパブリッククラウドからのアクセスが認められたデータであるかどうかも Pangea** 内で区別することにより，テーブルごとのデータ管理を可能としていきたい．これは，Pangea**がパケットを元にクエリの種類の判断をしていることから，可能であると判断される．

謝辞　本研究を進めるにあたり，ＮＴＴ研究所 細谷柚子様，三島健様に数多くの助言を賜りました．深く感謝いたします．

文　　献

[1] Naoto Matsumoto："クラウドコンピューティングへの誤解と注意点"，IEICE Communications Society Magazine，Vol．7 No．3，March 2014，pp．175–177.
[2] Masugi Inoue："頼れる情報通信インフラストラクチャの実現を目指して"，IEICE Communications Society Magazine，Vol．5 No．3，March 2012，pp．203–208.
[3] OpenStack：http://www.oprnstack.org/
[4] 中井悦司，中島倫明：「オープンソース・クラウド基盤 OpenStack 入門」2014 年 7 月 29 日第一版第三刷
[5] T.Mishima and H.Nakamura : "Pangea:An Eager Database Replication Middleware guaranteeing Snapshot Isolation without Modication of Database Servers", Proc.VLDB2009,pp.1066-1077, August 2009. PVLDB2009.
[6] 細谷 柚子，三島健，小口 正人：「リモートバックアップ機能を有したクラスタデータベースシステムにおける性能向上手法の提案」 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016)，D5-2，2016 年 3 月．
[7] Chihiro Maru, Miki Enoki, Akihiro Nakao, Shu Yamamoto, Saneyasu Yamaguchi, and Masato Oguchi, "Network Failure Detection System for Traffic Control using Social Information in Large-Scale Disasters," ITU Kaleidoscope Conference 2015: Trust in the Information Society, S5.3, December 2015.
[8] Open vSwitch：http://openvswitch.org/
[9] 原瑠理子，小口正人：「緊急災害情報に基づく OpenFlow を用いたバックアップシステムの実装と評価」 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016)，E7-5，2016 年 3 月．
[10] 小林大，渡邊明嗣，山口宗慶，田口亮，上原年博，横田治夫：「複製データを併用した効率的なデータマイグレーションの検討」，DBSJ LettersVeL3，No .2，2004
[11] 小林大，田口亮，横田治夫：「並列ストレージにおけるサービス性能を保った複製利用負荷均衡化に対する更新リクエストの影響」，DEWS2007 論文集，pp.L2-1，March 2007
[12] 小林大，渡邊明嗣，山口宗慶，田口亮，上原年博，横田治夫：「負荷分散のためのデータ移動による性能低下を抑制するアクセス回送制御」，TECHNICAL REPORT OF IEICE，DE2004-112，DC2004-27
[13] Iperf - The TCP/UDP Bandwidth Measurement Tool :https://iperf.fr/