Vol.2013-CLE-10 No.6
2013/5/25

スモールスタートで始める情報教育のための仮想化基盤の構築と展望

柏崎 礼生[1,a)]

概要：情報システムのための計算機資源を仮想化し集約することで初期導入・維持運用コストの低減を図る取り組みが 2000 年代後半から様々な大学で行われている．しかし一般的に導入されたシステムは 5 年間の稼働を前提としているため，キャパシティプランニングと設計に困難が付きまとう．本稿では小規模大学における仮想化基盤の導入を通して，計測・設計・導入の高速サイクルによるコスト削減の方策を検討する．

Hiroki Kashiwazaki[1,a)]

1. はじめに

2000 年代中盤に CPU の動作周波数の競争が一段落するとマルチコア化が進んだ．マルチコア化した計算機資源を最も活用し得るアプリケーション市場は仮想化ハイパーバイザであると目した CPU メーカーは仮想化ベンダーと協調し，仮想マシン (VM) を動作させる際の様々なボトルネックを解消した．これにより VM は物理マシンと遜色ないパフォーマンスで動作することが可能となった．このような仮想化技術の成熟とともに情報システムを稼働させる物理マシンを仮想化環境へと移行し，さらにはパブリッククラウド事業者が提供する IaaS[*1] へと移行する試みが行われている [1]．クラウドコンピューティングという言葉は，Gartner，UC Berkeley，そして NIST による定義が引用されることが多いが [2–4] ，本稿では「仮想化技術を用いて実現されるスケールアウト可能な基盤の上に構築された，規模を収縮可能なサービス」の意味で用いることとする．

組織外部のパブリッククラウドサービスを使うだけでなく国内の教育・研究機関の情報センターや研究科でのパブリック・プライベートクラウドの構築が行われている．静岡大学はクラウドコンピューティングを全面採用した情報基盤システムを構築した [5]，北陸先端科学技術大学院大学 (JAIST) では仮想デスクトップサービスを提供するためにプライベートクラウドを構築している [6,7]．佐賀大学は専用線で接続された外注先にプライベートクラウドを構築し，メールサービスの提供を行っている [8]．一方で，東京工業大学の TSUBAME2 に代表されるクラウド型 (スケールアウト型)HPCI や北海道大学アカデミッククラウド [9] など計算能力の大きさに重点をおいたパブリックサービスも提供されている．

本稿では小規模な大学におけるスモールスタートによる仮想化基盤の構築事例を解説する．この大学では，幸いなことに仮想化基盤の導入が遅れていた．予め確保された巨大予算に従った構築でもなかったため，様々な実験的手法を試行する余地があった．既存のパブリッククラウド，アカデミッククラウドとの連携や，広域分散クラウドコンピュータを利用する構築・運用モデルについての提案，そして大規模大学の情報教育システムへの適用可能性について考察を行う．

2. 仮想化基盤の構築 (2010 年度)

本稿で事例を紹介する大学は 2 学部 3 研究科からなる国立大学である．学生数は学部・大学院あわせて約 3,200 人，正規教職員数は約 300 人，非正規教職員数は約 1,000 人で，総計約 4,500 のアカウントを持つ．組織運営のための情報システムとしては，財務会計システム，人事給与システム，教務情報システムなど 11 システムが稼働している．2010 年 7 月までこれらの情報システムはすべて物理マシンで動作しており，VM で運用される情報システムは少数の部局の実験的な情報システムに限定されていた．

[1] 大阪大学
Osaka University
[a)] reo@cmc.osaka-u.ac.jp

[*1] Infrastructure As A Service

2.1 コストメリットの算出

学内の情報システムを物理マシンから VM に移行する際に，仮想化ハイパーバイザーを提供するメーカーによる P2V(Physical to Virtual) ツールを用いて VM へ移行する方法がある．しかしこの方法を用いて，物理マシン上で稼働している情報システムを VM 上での稼働に移行する場合，式 (2.1) により得られる損失 L_{P2V} を評価することが望ましい．

$$L_{P2V} = C_{P2V} + T_{P2V} \times S_c(t) + P_{P2V}(t) \times Ux(t) \quad (1)$$

上式において
(1) C_{P2V}: P2V に要する費用
(2) T_{P2V}: P2V により発生するサービス断時間
(3) $S_c(t)$: サービスが提供する価値関数 (費用/時間)
(4) P_{P2V}: P2V 後のサービスのパフォーマンス変化
(5) $Ux(t)$: ユーザエクスペリエンスの価値関数
である．P2V においては，この損失 L_{P2V} よりも VM へ移行することによるメリットが上回ることを，該当情報システムの構築業者および運用担当者に納得してもらうことが必須である．情報システムの更改のタイミングで，初期構築から VM で行う場合，C_{P2V} および T_{P2V} が最小化されるため VM への移行を比較的潤滑に行う事が可能となる．しかし大抵の情報システムはハードウェア・ソフトウェアのサポートの観点から，構築されてから 5 年以上利用される．全ての情報システムの更改が同時期に揃って行われる大学や研究組織もあるが，本稿の題材となる大学においては各情報システムごとに導入時期やサポート期限が異なっていた．

大規模な大学が大規模な予算を獲得した場合，大量の計算機，広帯域のネットワーク，そして高性能なストレージをふんだんに駆使したクラウドコンピューティング環境を実現できるが，小規模な大学における単年度のスタートアップ予算で開始する仮想化基盤の構築では全く異なる戦略が求められる．不必要に強大な計算機リソースや広帯域のネットワーク帯域，ストレージの高い I/O 性能を実現する必要性はどこにもなく，その組織の情報システムの総規模に相応な仮想化基盤を構築することが求められる．物理環境の計測結果は設計根拠として有効である．

実際に稼働しているシステムを計測するのではなく，システムが稼働するサーバのスペックから仮想化基盤に必要なホスト数の推定を行う方法がある．この手法は推定の正確さには欠けるが，費用もかからず即座に見積もる事が可能である点が特長である．この手法で概略を把握した上で，部分的にでも実システムの計測を行う事で，まずまず正確な見積もりを低予算・低労力で実現することができる．

2.2 物理環境のリソース計測

2010 年度に更改が予定されているシステムとしてグループウェアがあり，このシステムの運用は広報係が行っていたが，ハードウェアを保有していたのは情報センターであるといういびつな構造になっていた事も幸いし，グループウェアのバージョンアップのタイミングで物理マシンから VM に移行することに快諾して頂くことができた．システムの OS やソフトウェアのインストールは情報センター側で行っていたため，物理マシン上の OS に計測ツールをインストールすることについても，運用担当の反発もなく了承して頂けた．当初，このグループウェアの利用は事務職員のみに限られていたが，有期雇用の事務職員も含めると常時 200 ユーザが利用しており，そのほぼ全てのユーザが常駐アプリケーションであるリマインダーを使って定期的な Web サーバへのアクセスを発生させていた．学内の Web システムとしては有数の負荷を誇るであろうと推測されていたため，計測を行うテストケースとして意義があった．計測に用いたツールは Munin[*2] で，Web サーバへのアクセス頻度，トラフィック要求量，CPU 利用率，メモリ利用率を 1 ヶ月計測した (図 1)．Web サーバへのアクセス頻度は秒間平均 10 アクセス未満，トラフィック要求量は朝のピーク時で 3 Mbps 程度，CPU は 95%以上がアイドル状態で，4 GB のメモリは 100%使われる事もあるが，1 ヶ月の総量ではメモリの 82%が cache/buffers 状態にあった．またこのグループウェアの仕様は 2 コア以上の CPU と 4 GB のメモリを要求しており，計測結果からもこの規模の VM を作成すれば良いことが根拠付けられた．

2.3 仮想化環境とのリソース比較

VM を動作させる仮想化ハイパーバイザとして VMware ESXi (vSphere) を選択し，これを動作させる物理サーバとして hp 社の Proliant DL360 G7(Xeon E5640, 18 GB, 500 GB SATA HDD × 3 (RAID1 + Hot spare)) を調達した．仮想化ハイパーバイザと管理ソフトウェアの組み合わせとして KVM，Xen と OpenStack, CloudStack の組み合わせではなく VMware vSphere と vCenter を選択した理由は学内情報システムの動作する OS が多種多様であり，構築する仮想化基盤に合わせて各情報システムの構築をしてもらうのは本末転倒であると判断したことと，KVM，Xen と OpenStack，CloudStack の組み合わせでは当時は正式な日本拠点のサポートがなかったため，自分が構築できたとしても後任が運用を持続できる保証がなかったためである．この調達は仮想化基盤の本番環境ではなく，あくまで仮想化基盤上の VM に移行した場合における計測が主目的であるため高いスペックである必要はなかったが，管理サーバである VMware vCenter を導入することを想定していたおり，この物理サーバを vCenter 用に転用することを前提としたため，前述のようなスペックとなった．

[*2] http://munin-monitoring.org

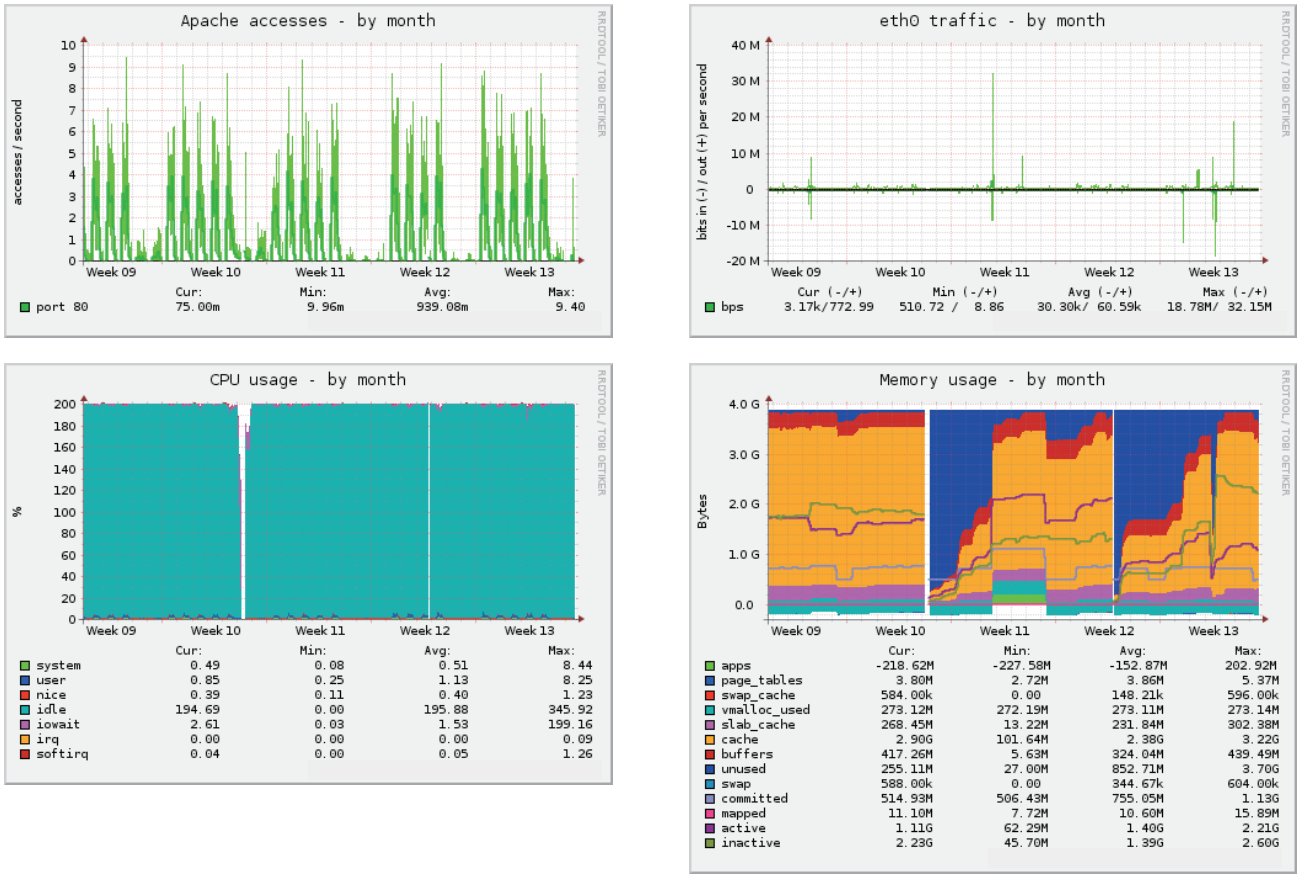

図 1　グループウェアサーバ (物理) のアクセス数，トラフィック量，CPU 使用率およびメモリ使用量

Fig. 1　Apache Accesses, amount of traffic, CPU usage and Memory usage status of a Physical Server of Groupware

この物理サーバに VMware vSphere を導入し，2 core, 4 GB のメモリからなる VM を作成して Debian (GNU) Linux 5.0 をインストールしてグループウェア環境を構築した．構築後，1 ヶ月間計測を行った結果，CPU 使用率とメモリ使用率は図 2 に示す結果となった．

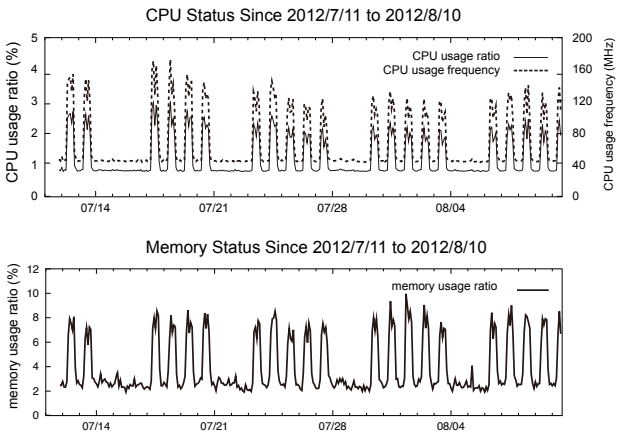

図 2　グループウェアサーバの CPU とメモリ状態

Fig. 2　CPU and Memory status on Groupware Server

CPU の使用率は数%，動作周波数にして数十〜200MHz 程度であった．またメモリも割り当てられた 4 GB に対して使用率は数%と非常に少ないものであり，これは物理サーバでの稼働時に計測した結果に十分近い結果である．学内にある他の 10 の情報システムにおける平均的なメモリ搭載量は 8 GB 程度であったので，今年度に全てのシステムが仮想環境に移行されたと仮定しても合計 80GB 程度のメモリを確保した仮想化サーバを用意すれば良いと見積もりを立てた．CPU の性能については評価が困難であったが，プロセッサーを後から増設することの困難さから 2 プロセッサ構成とすることを前提とした．同様に動作周波数についても後から不足が判明した場合に買い換えと換装をする事が困難である事を加味して当時における最高モデルの CPU を調達することとした．またサーバの物理障害に対応するため HA(High Availability) 構成とすることを前提としていたので調達する仮想化サーバ数は 2 台とした．サーバ数が今後も増えると仮想化サーバのパッチ当てやバックアップの運用コストが問題となるので，それらを統合管理できる VMware vCenter を含めた構成とし，vMotion によるライブマイグレーションが可能な VMware

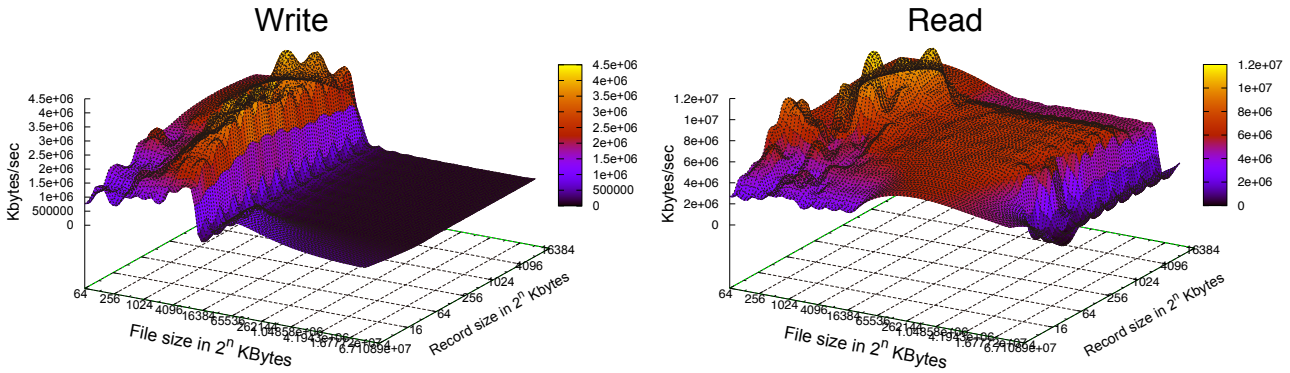

図 3　Netgear 社 ReadyNAS Pro 2TB × 6 の iozone による評価結果
Fig. 3　Results of Evaluation on Netgear ReadyNAS Pro

vSphere 4.1 の Standard Edition を仮想化ホストの OS として選んだ．購入した仮想化サーバのハードウェアスペックは下記の通りである．

- CPU: Xeon X5670 2.93GHz (2CPU)
- メモリ: 36GB
- HDD: 2TB (RAID5 + HotSpare)

2.4　ストレージの選定

技術的な情報収集を十分に行わなかった事に起因し，この構成には無駄な箇所がある．vSphere 4.1 では NFS もしくは iSCSI により接続されたデータストア上に配置された VM を vMotion によりライブマイグレーションすることができるが，仮想化サーバのローカルストレージに配置された VM はライブマイグレーションできない [*3]．HA 構成を実現し，ライブマイグレーションを行うためには NFS もしくは iSCSI を利用可能なストレージ製品を調達する必要がある．残予算の問題から高価なストレージ製品を調達することが困難であった．そこで Netgear 社の安価な NAS である ReadyNAS Pro を購入し，2TB の HDD を 6 台搭載して RAID-X を構築し，NFS でマウントした際における iozone によるパフォーマンステストを行った (図 3)．

64MB 以上の sequential write を行うとパフォーマンスは格段に低下する点が特徴的である．sequential read や random write, random read においてはファイルサイズ，ブロックサイズに関わらずそこそこの性能を発揮する事が分かった．本来ではあれば iozone だけでなく bonnie++[*4] や fio[*5] を使って単位時間あたりの I/O 性能についても比較するべきであった．この結果から初期導入においては Netgear 社の ReadyNAS Pro でも数十 MB のファイルの sequential write が頻繁に行われない限りは運用可能であろうと考え，ReadyNAS Pro を 2 台購入した．数十 MB のファイルを sequential write する場面として想定されるものは初期構築時において旧環境のデータをアーカイブしたファイルを転送する場面や，そのファイルを展開する場面などが挙げられる．ファイルを scp で転送した場合は TCP/IP 用通信バッファサイズに起因すると考えられる転送速度の抑制が，この問題においては良い方向に作用し深刻な問題となることはなかった．しかしアーカイブファイルの展開の際にはリモートコネクションの切断など様々な問題が発生した．ReadyNAS PRO ではなく安価な 1U サーバを調達し，Linux などでファイルサーバを構築することも考慮したが，バックアップも含めた構成を実現した上で管理ツールも含めて構築する場合のコストを勘案し，ReadyNAS PRO の購入を決断した．

ReadyNAS Pro は NIC が 2 ポートついており，片方のポートに障害が発生した場合においても対応することが可能である．これらを GbE で接続するために Cisco 2960S-48TD-L も購入した．この製品は 10GbE が 2 ポートついており，将来的に 10GbE ポートを持ったより高性能なストレージを拡張することを見越したものである．これらの機器によって図 4 のような構成で仮想化基盤を構築した．機器の調達においては総額を 500 万円未満にすることで事務手続きを簡略化することが可能となり [*6]，より早い調達と，調達された機器による計測と，年度内のもう一段階の増強を実現することが可能となる．

2.5　構築ポストプロセス

構築当初から仮想化ハイパーバイザが動作停止する不具合が発生したが，これらはメーカーから提供された診断ツールと仮想化ハイパーバイザのログを解析することにより原因が判明し，アレイコントローラの交換によって対応が行われた．調達や構築に時間を要してシステムの仮想化基盤への移行を急ピッチに行ってしまうとこういった初期

[*3] VMware vSphere Storage Appliance 5.1 を用いると，ESXi ホストの内部 (ローカル) ハードディスクリソースを抽象化して VSA クラスタを構成することができ，これを用いて vMotion を行うことができる．
[*4] http://www.coker.com.au/bonnie++
[*5] http://freecode.com/projects/fio
[*6] 東京大学 [企業のみなさまへ] 調達・契約について
http://www.u-tokyo.ac.jp/fin03/g04_j.html

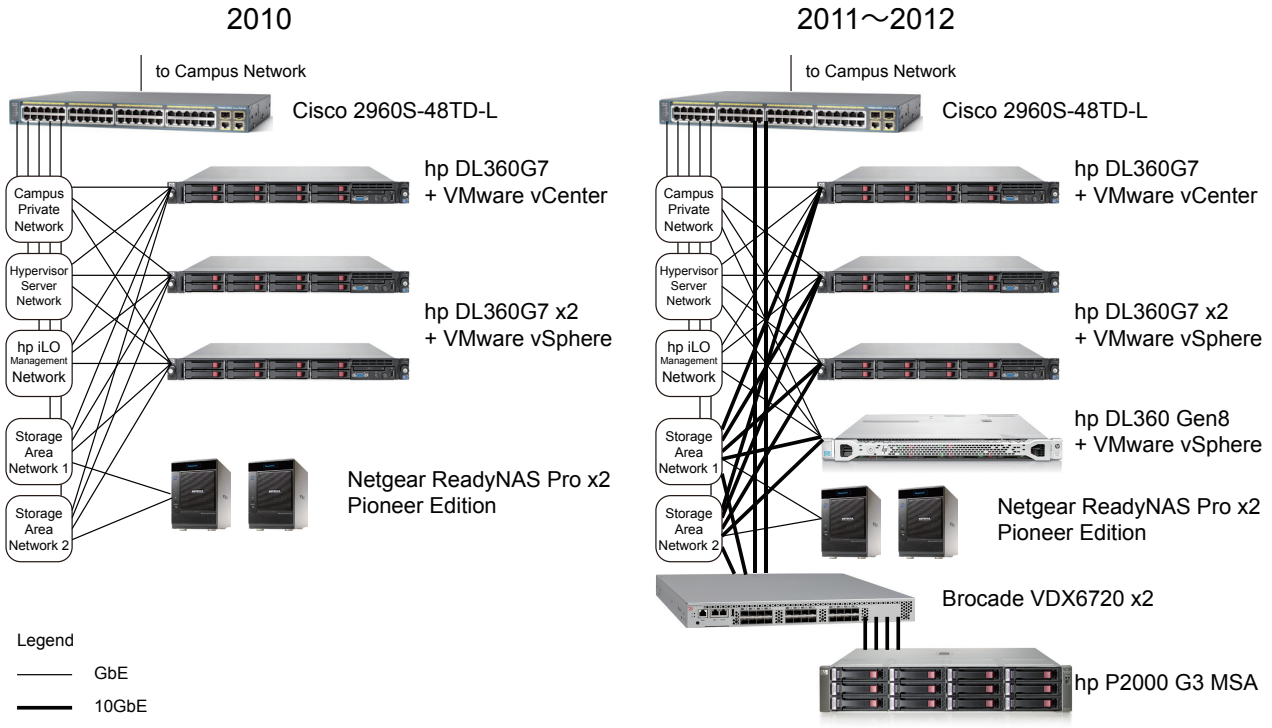

図 4　仮想化基盤の機器構成変遷
Fig. 4　A Change of Configuration of Virtulization Infrastructure.

動作不良に対して検証を時間をかけて行う事が困難となる．スモールスタートでの構築の利点はハードウェア障害の特定という点においても有利である．仮想化基盤において当初はグループウェアサービスを動作させる VM のみが稼働していたが，仮想化基盤でのサービスの開始とともに学内での宣伝活動を並行して行った結果，2010 年度中に 4 つの情報システムが仮想化基盤への移行，あるいは新規構築を希望してきた．この 4 システムは以下の通りである．

- 教員総覧システム
- 中期目標・中期計画進捗管理システム
- 図書館貴重資料データベース
- 美術館収蔵品データベース (バックアップ)

一方でまだ仮想化基盤に移行していない財務会計システムおよび教務情報システムはデータベースへのアクセスが頻繁に行われるシステムであることを勘案し，より高性能なストレージの導入とストレージエリアネットワークの 10GbE 化，および大学外部でのバックアップの試行などが来年度の課題となった．

3. 2011 年～2012 年度の増強

2011 年度の組織内予算案の暫定確定とほぼ同時に東北地方太平洋沖地震が発生し，ほぼ全ての新規予算案は凍結された．仮想化基盤は増強されないが，仮想化基盤への移行を希望するシステムは増えた．2011 年度に仮想化基盤への移行を希望した情報システムは以下の通りである．

- 図書館情報システム
- 美術館収蔵品データベース (本番環境)
- 財務会計システム

図書館情報システムは図書館蔵書の購入・寄付受け入れに伴う情報の入力の他，学生・教職員による書籍の貸し出し管理も行うため，重要度，アクセス頻度ともに高いシステムである．財務会計システムは入力担当者の数こそ多くはないが，大学運営において極めて重要なデータを扱うため，高いデータの保全性と短いサービス停止時間を求められる．最終的にこれらのシステムの移行を完了した時に問題となったのは VM 用のメモリ容量が逼迫したことにあった．また，各システムは 1～2 個の VM を要求し，2011 年度末時点で VM 数は 14 となっていた．これに対してストレージは ReadyNAS Pro 1 台のみであったために，ある VM がストレージに対して高負荷をかけた時の他 VM への影響が顕在化していた．こういった状況になることは 2011 年度第 2 四半期には推測されていたためストレージの増強を予定していたが，2011 年 7 月から始まったタイ洪水の影響を受けて 10 月にはストレージの高騰および製品調達が困難な状況が発生した．この問題に対処するため，ストレージ製品のシャーシのみを調達し，ハードディスクを翌年度に調達することとした．増強用のストレージは 10GbE 対応の hp P2000 G3 MSA とし，それに対応するため 10GbE 対応のスイッチ，Brocade VDX6720 を 2 台調達した．

機器の調達後，2012 年度に入ってから構築を行った．図

4 に示したように，hp DL360G7 の 10GbE 化も併せて行い，全情報システムの仮想化基盤への移行に備えて CPU，メモリ資源を増強するため仮想化サーバを 1 台増設，耐障害性に不安のあった Netgear ReadyNAS Pro はバックアップストレージとする構成としてデザインを行った．実際にはこれらの増強に必要な機材が揃う前に任期が完了したために，これらの構成が実稼働している様を筆者は観測することができない．このデザインは小規模な予算で増強をすることが可能な設計となっているのが特徴である．2011 年度分の増強機器と，2012 年度に申請をした機器費用を合計しても 500 万円未満で調達することが可能である．

4. 広域分散クラウドコンピューティング環境実現への取り組み

東北地方太平洋沖地震の教訓から，災害回復を実現するための取り組みとして，日本学術振興会産学協力研究委員会インターネット技術第 163 委員会 (ITRC) 地域間インタークラウド分科会 (RICC) を中心に，NII，金沢大学，大阪大学，広島大学は広域分散クラウドコンピューティング環境を実現するための実証実験を 2012 年度から開始した (図 5)．

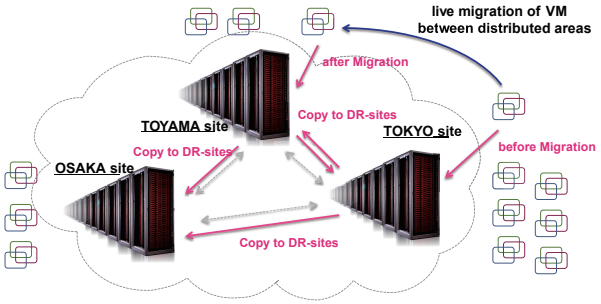

図 5　広域分散ストレージを利用したライブマイグレーション実験
Fig. 5　A Diagram of an Experiment of Live Migration by Using Wide Area Distributed Storage

我々は広域分散クラウドコンピューティング環境を実現するための層を下記のような 4 層で定義しており，2012 年度に行った実験でこの 1 層を実現した．

(1) 広域分散ストレージ層
(2) 障害回復 Traffic Engineering 層
(3) ID，ロケータ分離・結合層
(4) 分散プラットフォーム層

実証実験参加組織の仮想化サーバや広域分散ストレージソフトウェア，マイグレーションに必要な要件を満たすため，各拠点は SINET を介して接続され，SINET L3VPN と L2VPN を利用して環境を構築している．現在 JGN-X を併用したマルチホーム環境での実験も計画中である．既に 1000km 超の距離での広域ライブマイグレーション実験を行い，秒オーダーのネットワーク不通時間が発生するものの VM の移動を実現している．この距離でのライブマイグレーションが現実的になれば，各拠点は自拠点の仮想化基盤から任意の割合の計算機資源を提供することで，VM がどの拠点で稼働しているかを意識せずに広域分散クラウドコンピューティング環境利用を実現することができる．

5. まとめと考察

限られた時間と資源の中での小規模大学における仮想化基盤構築の取り組みと，広域分散クラウドコンピュータ研究を利用した広域ライブマイグレーションとの連携について紹介した．大学における災害回復手法として遠隔地のデータセンターを利用する例があるが，そういった取り組みだけでなく，研究開発段階の取り組みを本番環境に乗り入れさせる構築・運用モデルを示した．

謝辞　本研究は平成 25 年度北海道大学情報基盤センター共同研究「インタークラウド環境での広域分散ストレージ実験と検証」による支援を受けました．

参考文献

[1] 柏崎礼生: スモールスタートで始める大学の仮想化基盤の構築と運用の実情, インターネットと運用技術シンポジウム 2012 論文集, pp.94-101 (2012).
[2] Daryl C. Plummer, Thomas J. Bittman, Tom Austin, David W. Cearley and David Mitchell Smith: Cloud Computing: Defining and Describing an Emerging Phenomenon, Gartner Research, G00156220 (2008).
[3] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica and Matei Zaharia: Above the Clouds: A Berkeley View of Cloud Computing, UCB/EECS-2009-28 (2009).
[4] Lee Badger, Tim Grance, Robert Patt-Corner, Jeff Voas: DRAFT Cloud Computing Synopsis and Recommendation, NIST Special Publication 800-146 (2012).
[5] 坂田智之, 長谷川孝博, 水野信也, 永田正樹, 井上春樹: 情報セキュリティの観点からみた静岡大学の全面クラウド化, 情報処理学会研究報告, 2011-IOT-14, Vol.7, pp.1 (2011).
[6] 松原義継, 大谷誠, 江藤博文, 渡辺健次, 只木進一: プライベートクラウドによる電子メール管理コストの低減とサービスレベルの改善 —佐賀大学の事例—, 情報処理学会研究報告, 2011-IOT-14, Vol.8, pp.1-6 (2011).
[7] Shikida Mikifumi, Miyashita Kanae ,Ueno Mototsugu, Uda Satoshi: An evaluation of private cloud system for desktop environments, Proceedings of the ACM SIGUCCS 40th annual conference on Special interest group on university and college computing services (SIGUCCS '12), pp.131-134 (2012).
[8] 宮下夏苗, 上埜元嗣, 宇多仁, 敷田幹文: 大学におけるプライベートクラウド環境の構築と利用, 第 3 回インターネットと運用技術シンポジウム, pp.17-24 (2010).
[9] 棟朝雅晴, 高井昌彰: 北海道大学アカデミッククラウドにおけるコンテンツマネジメントシステムの展開, 第 10 回情報科学技術フォーラム 情報科学技術レターズ pp.15-18 (2011).