Vol.2014-CSEC-65 No.26
Vol.2014-IOT-25 No.26
2014/5/23

大阪大学の仮想化基盤の増強と設計

柏崎 礼生[1,a)]　宮永 勢次[1,b)]　森原 一郎[1,c)]

概要：大阪大学では 2010 年にサーバ集約を目的として，VMware 社の VMware ESX 4.0 をハイパーバイザとして利用する仮想化基盤を構築した．この基盤上で 2014 年 3 月末時点で 32 システムによる 105VMs が稼働し，227 仮想 CPU を利用している．また本基盤の上で稼働するメールシステムは学内 38 部局，8,450 アカウントを提供する．この基盤は当初 2014 年度末にリプレイスされる予定であったが，2014 年 9 月に竣工予定の学内データセンタ「IT コア棟」への移設に伴い，半年繰り上げて「拡張」が行われる事となった．本稿ではこの仮想化基盤の増強設計と構築・運用の方針について解説する．

キーワード：設計, 仮想化基盤

Extension and design of a virtualization infrastracture in Osaka University

Hiroki Kashiwazaki[1,a)]　Seiji Miyanaga[1,b)]　Ichirou Morihara[1,c)]

Abstract: A virtualized infrastracture on Osaka University started its service to aggreagte server machines inside the university in March 2010. A hypervisor software of the infrastracture have been VMware ESX 4.0 even now and 105 VMs of 32 systems that using 227 virtual CPU is working on the infrastracture. An university mail systems is one of the systems and this mail system provide 8,450 mail accounts with 38 departments. Although the infrastracture was supposed to be replaced until March 2015, replacement was moved up in September 2014 because of datacenter of the university a.k.a. IT core building will be completed of construction in Septmeber 2014 and the current infrastracuture have to be moved to the building. In this paper, authors show the design and specification of new virtualized infrastracture and its plan of moving.

Keywords: design, virtualization infrastracture,

1. はじめに

組織に於ける業務の電算化が進んで幾星霜，社会と情報通信技術 (Information and Communication Technology: ICT) は密接な関係を築き上げている．必然的に業務における ICT 投資は肥大し，それに対して組織は ICT 投資を効率化することを目指すこととなった．この潮流と，計算機科学において古くから研究開発されていた仮想化技術の親和性は高く，2000 年代からエミュレーションによるレガシー OS 動作環境の仮想化や，コスト削減のためのサーバマシンの集約を目的として用いられるようになった．米 VMware 社 *1 は代表的な仮想化ソフトウェア企業であり，VMware 社が VMware Infrastructure 3 をリリースした 2006 年からサーバ集約のための仮想化技術の利活用が様々な組織で進められるようになった．奇しくも当時 Google の CEO であった Eric Emerson Schmidt 氏が Search Engine Strategies Conference において Danny Sullivan 氏との対談で「クラウドコンピューティング」という言葉を紹介したのも 2006 年のことである *2.

日本国内の大学におけるサーバ資源集約のための仮想化

1 大阪大学
Osaka University
a) reo@cmc.osaka-u.ac.jp
b) miyanaga-s@office.osaka-u.ac.jp
c) morihara@cmc.osaka-u.ac.jp

*1 http://www.vmware.com
*2 http://www.google.com/press/podium/ses2006.html

© 2014 Information Processing Society of Japan

技術の利用例としては，2007 年から「クラウド化」を推進している静岡大学の事例が有名である [1]．また北陸先端科学技術大学院大学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おり [2,3]，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うなど [4]，仮想化技術を用いた様々なアプローチで ICT 投資の効率化を図っている．一方，情報技術に関する研究とコンサルティングをもっぱら行う米ガートナー社の research director である Michael Warrilow 氏は，2014 年はプライベートクラウドからハイブリッドクラウドやパブリッククラウドへと移行するようになるだろうという趣旨の発言をしている [*3]．2000 年代後半に仮想化基盤を構築した組織の中には，数年間の運用とリソースの計測を通して，その基盤構築への投資が期待した効果を提供しないことに気付いた組織もある．2013 年 8 月に米ガートナー社が公表した “Hype Cycle for Emerging Technologies, 2013” によると，クラウドコンピューティングは過剰な期待のピーク (Peak of inflated expectations) を越えて幻滅・目覚めの谷 (Trough of disillusionment) に入っている (図 1)．2014 年からの「プライベートクラウド」あるいは仮想化基盤の構築には，これまでの 8 年の過程を考慮に入れた設計が求められる．

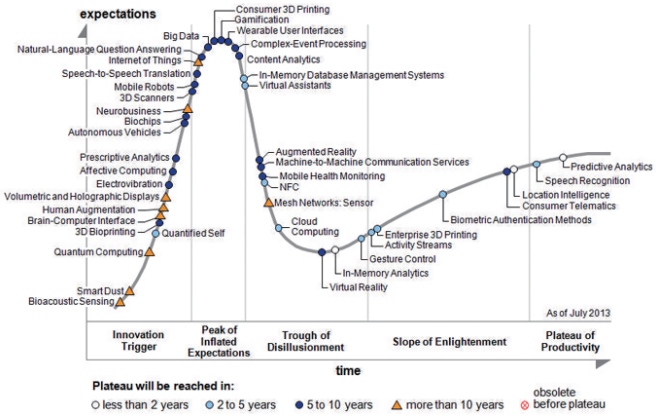

図 1 米ガートナー社による先進テクノロジのハイプサイクル 2013 年度版 [*5]

Fig. 1 Hype Cycle for Emerging Technologies, 2013

2. 大阪大学の仮想化基盤

大阪大学は学部学生約 1 万 5 千人，大学院生約 7 千 5 百人，教職員数約 6 千人 (非常勤職員等を含めると約 9 千 3 百人) からなる国立大学である．大阪大学もまた ICT 投資を効率化するという目的のもと，大阪大学全体として業務フロー全体の最適化を行い，業務の効率化を目指すという目標が掲げられた．大阪大学程度の規模の総合大学では，部局ごとに独自の ICT 投資が行われ，事務業務フローも部局独自で構築されるケースがある．大学全体を俯瞰すると ICT 投資が分散し，業務改革も局所的な最適化に留まり，非効率な状態にあることもある．業務の全体最適化と ICT 投資の集約を実現する手段として，大阪大学は情報推進機構を設置し，仮想化技術を中心に据えたクラウド技術の活用に取り組んだ．本機構は，将来的には ICT リソースを外部にアウトソースする可能性も選択肢の 1 つとして考えながら，プライベートクラウド方式のプラットフォームシステムの構築を目指していた [6,7]．

大阪大学では上述の背景のもと，事務業務の効率化・改革の第一歩として事務系基幹システムを 2010 年に刷新した．事務系基幹システムの刷新の際には，今後の大学の様々なサービスを集約し実行可能なシステムの構築を目指し，仮想化技術を採用した共通基盤プラットフォームシステム「大阪大学キャンパスクラウド」(以下，キャンパスクラウド) を設計，構築した．このキャンパスクラウドは事務系基幹システムの計算機資源だけでなく，財務会計システムからも計算機資源の提供を受けて 2010 年からサービスを開始した．このキャンパスクラウドは hp のブレードサーバを 12 台利用し，物理コア数は合計 96，主記憶容量は合計 432GB で動作するものである．ここにさらに 2010 年度に教員基礎データベースシステムと呼ばれる計算機資源が追加され，2 つの業者が導入した 3 つのシステムによる 3 ブレードシャーシからなるキャンパスクラウドが出来上がった．この基盤上で 2014 年 3 月末時点で 32 システムによる 105VMs が稼働し，227 仮想 CPU を利用している．また本基盤の上で稼働するメールシステムは学内 38 部局，8,450 アカウントを提供している．

一方，これらの機器が設置されている大阪大学吹田キャンパスのサイバーメディアセンター本館では耐震・改修工事が 2014 年度に行われる事が決定し，スーパーコンピューター等を収納する地上 2 階建て，建物延べ面積 2,040m^2 の IT コア棟が新設されることとなった．IT コア棟の竣工は 2014 年 9 月末を予定しており，その直後にサイバーメディアセンター本館の耐震・改修工事が着工することとなっている．サイバーメディアセンターの改修にあたって，全ての計算機資源は IT コア棟に移設される必要があるが，移設作業中，キャンパスクラウドのサービスが全て停止することが懸念された．移転作業は週末の休日と祝日を含めた 3 日間で完了することが可能と見積もられたが，メールシステムを 3 日間ダウンさせることにはユーザからの反発が想像することが難くない．それ以外にもソフトウェアのサポート上の問題点も指摘された．

2010 年のサービス構築時において仕様として要求されたハイパーバイザソフトウェアは VMware 社の VMware Infrastructure 3 である．このソフトウェアはバージョンアップを繰り返し，現在では VMware ESXi^{TM}5.5 となっ

*3 http://www.cio.com.au/article/534196/where_cloud_computing_heading_2014_/

ているが，キャンパスクラウドは VMware ESX 4.0 で稼働し続けている．VMware 社ではハイパーバイザソフトウェアのメンテナンスアップデート，アップグレード，不具合とセキュリティの修正，および技術的な支援が提供される “General Support” の期限を製品ごとに設定しており，VMware ESX 4.0 の General Support は 2014 年 5 月 21 日で終了する．その後は 2016 年 5 月 21 日まで技術的な支援は提供されるが，その他のサポートは提供されなくなるため，ハイパーバイザソフトウェアに深刻な脆弱性が発見された場合は対処を行うことが極めて困難となる．

これらの問題を解決するために，基幹系プラットフォームの機器を拡張することとなった．この拡張について説明するために，キャンパスクラウドの現在の構成を解説する．

2.1 キャンパスクラウド現構成

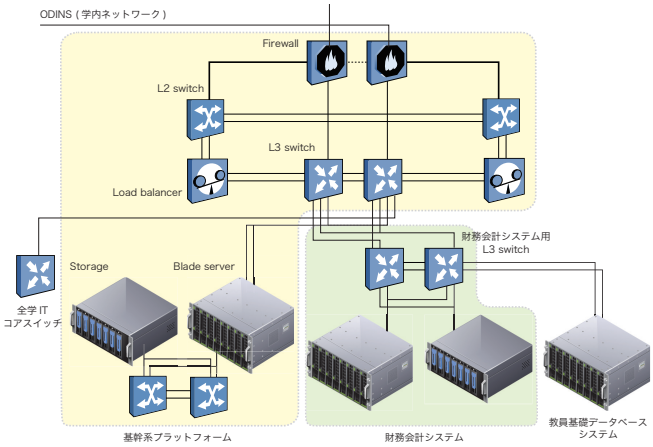

図 2 現在の大阪大学キャンパスクラウド概念図
Fig. 2 A present diagram of the campus cloud

キャンパスクラウドの構成図を図 2 に示す．本システムでは，仮想マシン (Virtual Machine: VM) を動作させるハードウェアとしてブレードサーバシステム複数台導入し，ファイアウォールやコアスイッチ，ロードバランサなど全てのネットワーク機器を冗長構成にすることでシステムの高可用性を実現している．前述のようにキャンパスクラウドは基幹システム，財務会計システム，教員基礎データベースシステムから構成されており，それぞれがファイアウォール，スイッチ，ロードバランサーを調達して，さらに冗長度の高い構成となっている．そしてこの冗長度の高さはこれらの機器の年間保守費用の増大に繋がる．

ブレードサーバシステムは VM が利用する CPU およびメモリ資源を高密度に配備できることから，キャンパスクラウドのように様々なサーバを集約するようなシステムにとっては非常に効率がよい，とされた．また，システム全体の要求リソースに応じてブレードを追加導入するといった拡張が可能であり，効率のよい ICT 投資を計画することが可能である，と考えられた．そういった背景もあり 2009 年頃の日本の ICT 業界においては，ブレードシャーシを無料同然の価格で売り，ブレードサーバを購入させ，段階的に増強させるというビジネスアプローチが存在した．

しかしブレードシャーシにもファームウェアが存在し，このアップデートを行う際にはそのブレードシャーシに格納された全てのブレードサーバは停止状態となる．そのためブレードサーバシステムで可用性を高めるためには 2 台以上のブレードシャーシが必要となる．また，ブレードシャーシのファームウェアをアップデートする際には，そのシャーシに格納されたブレードサーバ上にある全ての VM を，他のブレードシャーシに格納されたブレードサーバ上のハイパーバイザソフトウェア上にマイグレーションさせる必要がある．ブレードシャーシ n 台で動作させる場合，1 台のブレードシャーシのアップデートに備えて，他 $n-1$ 台のブレードシャーシに格納されるブレードサーバは常に $\frac{1}{n}$ の余裕をもって稼働させなければならないため，ブレードサーバが持つ計算機リソースの $\frac{n-1}{n}$ しか VM に提供できないこととなる．例えば $n=2$ の場合は各ブレードサーバは 1/2 の計算機リソースしか VM に提供できない．サーバ集約を目的とした仮想化基盤の設計において，ブレードシャーシ数と費用対効果は比例する．

2.2 拡張の設計

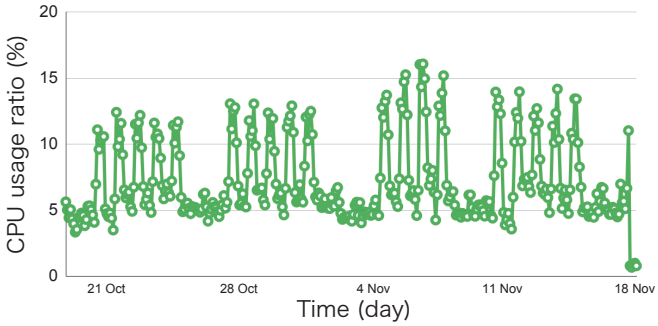

図 3 仮想化ホストの CPU 利用率の平均値
Fig. 3 Average of CPU usage ratio of 12 virtualization hosts

拡張したキャンパスクラウドの性能的な要求要件を策定するため，まず現キャンパスクラウドの計測を行った．キャンパスクラウドを構成する 12 台の仮想化ホストの中の CPU 利用率を 4 週間計測し，その平均値をグラフ化した (図 3)．仮想化ホスト 1 台は 2 つの CPU から構成されており，Intel Xeon 5500 番台，2.53GHz の CPU が仕様で要求されている．この数値は 2 時間ごとの平均値から算出されているため，より短い時間でのピークは図よりも大きくなるが，参考値として利用するものとする．より短い時間での CPU 利用率を計測し，偏差を算出することが望ましい．この図では平日の 9 時から 17 時にピークがあり，ピーク時の平均 CPU 利用率は 20%未満，夜間および休日

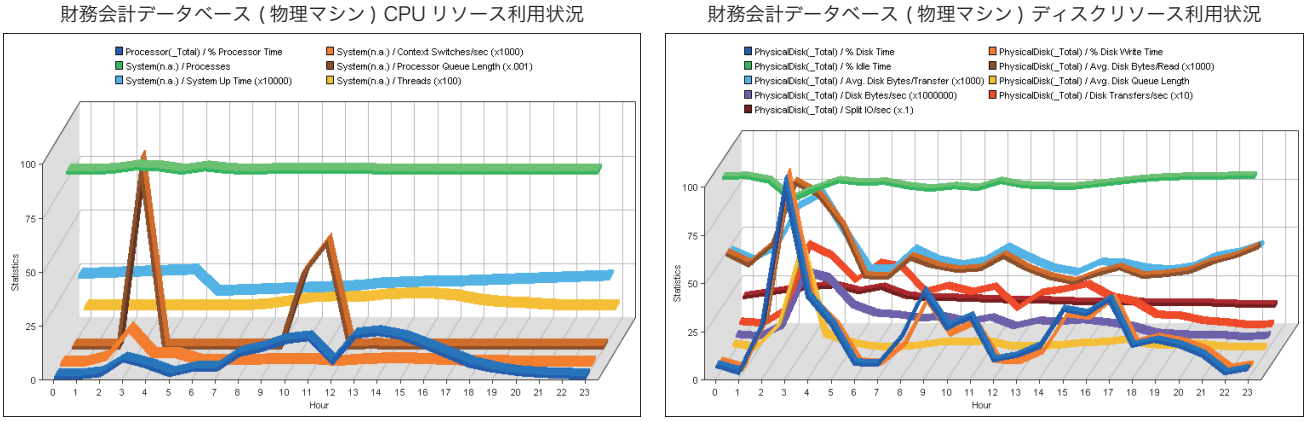

図 4 物理ホストの CPU・ディスクリソース利用状況
Fig. 4 Status of CPU and storage resource usages on a physical database server

の平均 CPU 利用率は 10%未満であることが分かる [5].

次に今後キャンパスクラウドへの移転を検討する物理マシンの計測を行った．業務系システムの中でも負荷が高いとされる財務会計システムのデータベースサーバを対象とし，CPU 負荷だけでなくストレージの IOPS も評価対象とした．ここではネットワンシステムズ株式会社の協力のもと，VMware 社が提供する Capacity Planner を利用して計測を行った．財務会計システムのデータベースサーバは 2 つの CPU から構成されており，動作周波数は 2.93GHz，メモリ容量は 24GB である．2 週間の計測を行い，毎日 0 時から 23 時までの平均値を計算した (図 4)．図の左は CPU リソース利用状況であり，右はディスクリソース利用状況である．夜間バッチが実行される 2 時から 5 時が最も高いプロセス待ち行列長 (茶色) を示している他，業務開始後の午前中もまた比較的高い値を示している．一方 CPU 利用率 (青色) は業務時間内である日中は比較的高い値を示すが，24 時間の平均 CPU 使用率は 8.3%，日中に限定しても 14.9%，ピークは 20.5%である．ディスクリソースも，夜間バッチが実行される 2 時から 5 時の値が高くなっているが，日中は比較的安定しており，IOPS の平均値は 216，ピーク値は 490 であった．また Read と Write の割合は平均時で 77:23，書き込みのピーク時においても 65:35 であった．

この結果を前提とし，今後 3 年間のシステム数の増大を推測した上で，拡張したキャンパスクラウドは物理コア数は現行と同程度の 120 コアとし，この上で動作させる VM 数を約 4 倍の 400 にすることとした．また潤沢な主記憶容量の要求に応じることができるように，拡張当初は合計 640GB とし，400VMs が平均 200IOPS を要求するものとして 80kIOPS の要求に応じることのできるストレージを具備するものとした．その上で費用の増大を抑制することを考える必要がある．

2.3 ライセンス費の抑制

現構成において導入されたハイパーバイザソフトウェア群である VMware vSphere 4.0 のエディションは Enterprise Plus である．このソフトウェア群は，動作する物理 CPU の数に比例したライセンス費用体系となっている．Enterprise Plus は vSphere における最高級のエディションであるが，その理由は Storage vMotion や Distributed Resource Scheduler を利用するためではなく，Cisco Nexus 1000V を利用することが目的であった．

Cisco Nexus 1000V は Cisco 社の分散仮想スイッチソフトウェアである．複数の仮想化ホストをクラスタ構成する場合はネットワークの管理・設定作業が非常に煩雑になるためモダンなハイパーバイザソフトウェアにおいてはその内部で仮想スイッチを作成し，VM をその仮想スイッチに接続する機能を標準で提供する．各仮想化ホストに分散した仮想スイッチの設定を一元管理するのが Nexus 1000V である．キャンパスクラウドにおいては主に VM のライブマイグレーション後の VLAN の設定を主な目的として Cisco Nexus 1000V が導入されたが，全ての VM が 1 つの VLAN の配下にあるキャンパスクラウドにおいて利活用される場面は極めて限定的であった．そして Cisco Nexus 1000V もまた物理 CPU の数に比例したライセンス費用体系である．キャンパスクラウドは 24 物理 CPU からなる計算機資源により構築されているため，これらのライセンス費用が高額であることもまた問題点として挙げられていた．

このため物理コア数の大きな CPU を調達するメリットを算出する必要がある．現環境では 96 物理コアのために 24 個の CPU を利用している．2013 年に発売された Ivy Bridge の Xeon では 1CPU あたり 12 コアの製品がリリースされている．これを用いることにより 120 コアは 10 個の CPU で実現することが可能である．比較的安価な Xeon は 8 コアのものがリリースされているが，こちらを使うと 15 個の CPU が必要である．5 個の CPU の差により年間 5

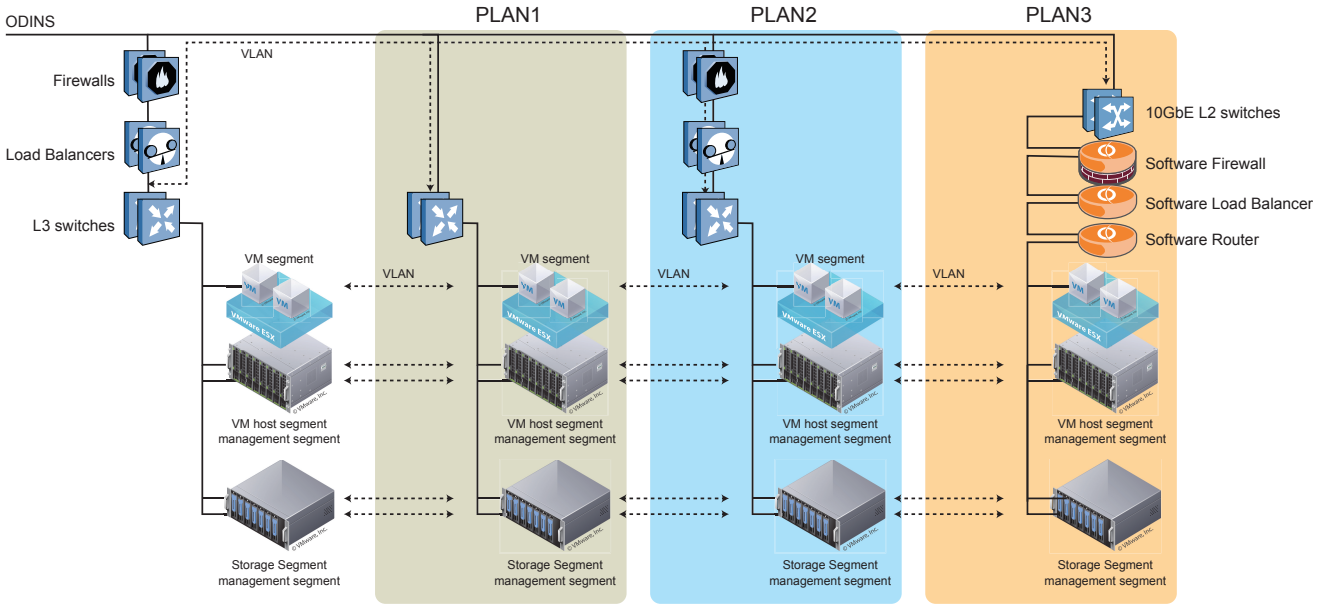

図 5 移行計画の議論の過程模式図
Fig. 5 A diagram of discussion process of the migration

ライセンス分の運用費用差が発生するので，5 年間運用することを前提とするならば，25 ライセンス・年分の VMware vSphere Enterprise Plus Edition の価格が 8 コア Xeon 構成にメリットとして存在するならば 8 コア Xeon 構成を選択すべきである．価格を検討した結果，12 コア Xeon で構成することとなった．

3. 移行プラン

現行のキャンパスクラウドにおける問題点の 1 つとして，構成する機器が使用するリソース情報のほとんどを長期間に渡って収集，蓄積する環境になっていないことであった．VMware vSphere および vCenter の利用により仮想ホストのリソース利用率は計測できたが，その他のネットワーク機器，ファイアウォール，ロードバランサについてはログも極めて限定的にしか保存されていなかった．そのため拡張したキャンパスクラウドの要件として，全ての機器の情報を収集可能であることを要求している．

ファイアウォールやロードバランサにどの程度の負荷がかかっているかを把握することが出来ないため，根拠となるのは現行機種と同程度以上の機能を有することとしか要求することができない．移行に関する試行実験としては，

(1) 現行のファイアウォール，ロードバランサを継続して利用する方法
(2) 新たにファイアウォール，ロードバランサを購入し，現環境とは VLAN で仮想化ホスト，ストレージが通信可能な状態とする方法
(3) ファイアウォール，ロードバランサはソフトウェアで実現する方法

が提案された (図 5)．最終的にプラン 2. が採択され，その後の移行のステージについても同様に図を用いて設計部門内で共有され議論が進められた．言葉による議事録ではなく，このような図を用いて共有をすることで，設計や仕様策定に携わる担当者が異動したり，あるいは新たに加わった時でも迅速に理解を共有することができ，また異動にまつわるドキュメントの執筆で作業が遅延することも未然に防ぐことが可能となる．

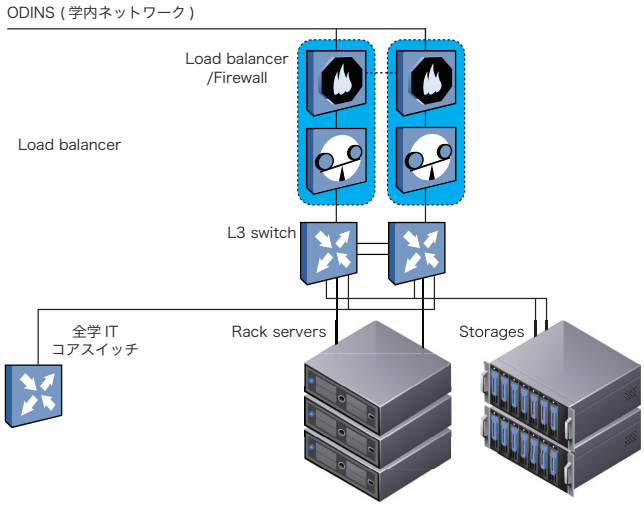

図 6 キャンパスクラウド拡張版の模式図
Fig. 6 A diagram of enforcement of the campus cloud

また，当初はファイアウォールとロードバランサーが個別の機器として調達することを前提としていたが，昨今のファイアウォール機器には十分なロードバランサー機能を具備するものが少なくないため，最終的に図 6 に示すシンプルな構成で仕様書が執筆されることとなった．

© 2014 Information Processing Society of Japan

4. 拡張設計と仕様書

情報基盤の担当者が任期付き教員である場合が多い国公立大学においては，仕様書は単に調達する物品を要求するだけのものではなく，次期システムを構築する担当者が現行のシステムが「なぜ」作られたのか，その背景を知るための文書ともなり得る [8]．そのためこの拡張仕様書においては第一章において，現状 3 つのシステムにより構成されているキャンパスクラウドの歴史的背景を述べつつ運用インターフェイスの統一を目的とすることを明記してある．

拡張設計に基づき学内での予算要求を行い 2013 年 11 月に暫定的な予算決定が行われたが，どういう経緯によるものか 2013 年 12 月末にその予算のうち，およそ 25% が 2013 年度中に提供されることとなった．そのため，事前に調達したてとしても 2014 年度の調達物品と組み合わせて利用することが可能な機器で，かつ 2013 年度に購入してから 2014 年度に構築が始まるまでの 6 ヶ月間で有効活用することが可能な機器を選別し，見積もりを行うこととなった．その結果，仮想化ホスト 5 台を事前に調達することとなり，この環境を用いて，VMware が提供する Software defined storage である VSAN，およびネットワーク仮想化プラットフォームである VMware NSX を検証することとした (図 7)．

図 7 導入された VSAN，NSX 検証環境
Fig. 7 A verification environment for VMware VSAN and NSX

2014 年 4 月 1 日にキャンパスクラウドの機器拡張仕様 (基幹系プラットフォーム機器拡張) が官報に掲載されたが，仕様書に不備があることが分かった．その不備とは RAID(Redundant Array of Inexpensive Disks) の記載である．RAID は 1988 年の David A. Patterson による ACM SIGMOD での発表 [9] が初出だが，その後国際標準規格として策定されてはいない．業界団体である SNIA が Common RAID Disk Data Format(DDF)[*6] という仕様文書を公開しているが国際標準ではない．そのため仕様書において RAID という文言を使う場合は，要求したい RAID レベルについて仕様書内で定義するか，あるいは上述の DDF に従うといった記載が求められるのではないだろうか．

*6 http://www.snia.org/tech_activities/standards/curr_standards/ddf

5. 動機付けとまとめ

大阪大学における ICT 投資効率化のための仮想化基盤の増強計画について解説し，2010 年代中盤における仮想化基盤の設計について述べた．今や情報基盤の設計において組織に依存する特殊性が備わることは少ないため，学術組織内に人材を確保して設計を行う積極的必然性はない．だが，もし下記のインセンティブが担当者に与えられるならば，より良い設計が実現できる可能性は高まるかもしれない．

$$I = k \times (E_{max} - P_{bid}) \times S^n \quad (1)$$

ここで E_{max} は整理された要望をもとにした要件定義を複数の業者に見せて見積もりをとった時の最大見積金額，P_{bid} は最終的な落札金額，S は導入後の顧客満足度 (%)，k は雇用者が指定することが可能なパラメータであり，I が担当者に支払われるインセンティブである．

参考文献

[1] 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).
[2] 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).
[3] 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).
[4] 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).
[5] 柏崎礼生, 森原一郎: 大阪大学における仮想化基盤の設計とその増強計画, インターネットと運用技術シンポジウム 2013 論文集, Vol. 2013, pp. 47-50 (2014).
[6] 市川昊平, 江原康生, 長岡亨, 森原一郎: 大阪大学のキャンパスクラウドへの取り組み, 大学 ICT 推進協議会 2011 年度年次大会論文集, pp. 312-325 (2011).
[7] 宮永勢次, 市川昊平, 小林兼: 大阪大学のキャンパスクラウドシステムについて, 全国共同利用情報基盤センター研究開発論文集, No. 34, pp. 77-82 (2012).
[8] 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, Vol. 2012, pp. 94-101 (2012).
[9] David A. Patterson, Garth Gibson, Randy H. Katz: A Case for Redundant Arrays of Inexpensive Disks (RAID), Proceedings of the 1988 ACM SIGMOD International Conference on Management of Data, pp. 109-116 (1988).

© 2014 Information Processing Society of Japan