

一般口演

一般口演12

ネットワーク・システム開発

2017年11月21日(火) 16:15 ～ 17:45 B会場 (12F 会議室1202)

[2-B-3-OP12-4] Hyper-Converged Infrastructure（HCI）による仮想化環境全体の高速化と長寿命化

野口 忠祥¹, 甲斐 聖人¹, 中熊 英貴¹, 東 賢剛¹, 松尾 健司^{1,2}, 佐方 友和²（1.社会福祉法人恩賜財団済生会熊本病院 医療情報システム室, 2.株式会社プレス）

【背景】

当院は2010年に複数台の物理サーバをスイッチ経由で1台のストレージに接続する3階層構成の仮想化基盤を構築し、サーバ仮想化を推進してきた。2011年の電子カルテ導入を機に部門システムサーバ、VDI(Virtual Desktop Infrastructure)など約200台を仮想化、その後もハード追加をしつつ拡大し、2014年には約300台の仮想マシンを稼働させた。3階層の仮想化基盤はストレージに Disk 負荷が集中し、拡大に伴いレスポンスが低下する。また、高速化と安定稼働のために高価なストレージ製品を採用するため、大容量サーバの仮想化は費用対効果が得られにくい。さらに仮想化基盤のハード更新時は、大量の仮想マシンを受け入れ可能な大規模な新環境が必要となり、高額な投資と膨大な移行作業が生じる。途中で増設した機器は使用期間が5年未満となり、活用不十分で役目を終える。更なる仮想化拡大のためには仮想化基盤の刷新が必要であった。

【目的】

仮想マシン増加によりレスポンスが低下することなく、ハード寿命に左右されない仮想環境を構築し、仮想化全体の高速化と長寿命化を実現する。大規模な環境更新を減らし、作業負荷と投資の集中を防止する。大容量サーバを仮想化し、更なる仮想化拡大を図る。

【方法】

Hyper-Converged Infrastructure (HCI)の仮想環境を構築し、各ノード(物理サーバ)に負荷分散し、数台のノード追加で仮想化全体の性能をコントロール可能とする。別途、スケールアウト型 NAS を構築し、容量単価の低い筐体選択を可能とする。

【結果】

9ノード構成の HCI で150台の仮想マシンを稼働し、レスポンス低下は発生していない。新たに70台の仮想マシンを稼働予定であり、6ノード追加し、対応する。HCI、スケールアウト型 NAS により、5年が経過したノードを順次入れ替えながら、長期間使用できる環境となり、投資分散と無駄のない機器利用が可能となった。画像サーバ仮想化において、システム領域を HCI、画像領域を NAS(Network Attached Storage)に割り当て、大容量サーバの仮想化コストを抑えることができた。

Hyper-Converged Infrastructure (HCI) による仮想化環境全体の高速化と長寿命化

野口 忠祥^{*1}、甲斐 聖人^{*1}、中熊 英貴^{*1}、東 賢剛^{*1}
松尾 健司^{*2}、佐方 友和^{*2}

*1 社会福祉法人恩賜財団済生会熊本病院、*2 株式会社ブレス、

Improving speed and extending life of virtualized environment overall using hyper-converged infrastructure (HCI)

Noguchi Tadayoshi^{*1}, Kai Masato^{*1}, Nakaguma Hideki^{*1}, Higashi Kengou^{*1},
Matsuo Kenji^{*2}, Sakata Tomokazu^{*2},

*1 Saiseikai Kumamoto Hospital, *2 BRESS Co., Ltd.

Background

In 2010 this institution constructed a 3-layer hierarchical virtualized infrastructure that connects multiple physical servers to a single storage unit via switches. With the introduction of electronic patient charts in 2011 as a starting point, virtualization was implemented in approximately 200 devices such as departmental system servers and virtual desktop infrastructure (VDI), hardware was subsequently added to expand the network, and in 2014 approximately 300 virtual machines became operational. In a 3-layer hierarchical virtualized infrastructure, however, disk loading aggregates into storage, and expansion slows response. Moreover, cost effectiveness is hard to achieve with virtualization of large capacity servers because higher speed and operational stability require expensive storage equipment. In addition, updating hardware for the virtualized infrastructure requires a large-scale new environment capable of handling a vast number of virtual machines, which means heavy capital investment and a daunting transition task. Equipment added during expansion has a service life of less than 5 years, and is ultimately underutilized. The virtualized infrastructure needed revamping to accommodate further expansion of virtualization.

Goals

One goal was to construct a virtual environment that did not lose response with the addition of virtual machines and did not depend on hardware life span. Other goals were to decrease large-scale revamping of the environment and to control workload and the accumulation of capital investment. We aimed to virtualize large capacity servers and implement greater virtualization.

Method

We constructed a hyper-converged infrastructure (HCI) virtual environment to disperse the load to separate nodes (physical servers) and enable the control of virtualization as a whole through the addition of multiple nodes. Separately, we constructed a scale-out network-attached storage (NAS) architecture to enable low cost-per-gigabyte housing choices.

Results

A total of 150 virtual machines now operate in the 9-node HCI with no loss of response. Operation of 70 new virtual machines is scheduled, and this will be balanced by adding 6 nodes. With HCI and scale-out NAS we realized an environment for long-term use while gradually replacing nodes after 5 years to enable investment diversification and efficient equipment utilization. For the virtualization of image servers, the system domain was allocated to the HCI and the image domain to the NAS to control large capacity server virtualization costs.

Keywords: HCI, Virtual server

1. 背景

病院情報システムは、電子カルテを中心に多数の部門システムで構成され、それらを相互連携することで、一連の病院業務を支援する統合型システムを形成する。各システムは、システム間連携のために、インターフェースサーバや Web サーバをデータベースサーバとは別に設けるため、1つのシステムに対し複数のサーバが必要となる。また、画像処理など専門性の高い領域は次々に先進的なシステムが登場し、IoT

や AI など最新 IT の発展も目覚ましく、医療の質向上や業務効率化のため、迅速なシステム導入も重要である。このように医療機関では膨大な数のサーバ運営と迅速なサーバ構築が求められ、サーバ仮想化はそれを低コストかつ効率的に実現する有効な手段であり、多くの医療機関で導入されている。

済生会熊本病院では、2010 年よりサーバ仮想化を開始し、「使用状況に応じたリソース追加」など運用面の工夫をしながら推進したことで、仮想化の費用対効果を高めることができた

が、一方で「パフォーマンス低下」などの根本解決が必要な課題に直面し、その解決と更なる仮想化推進のため、HCI (Hyper-Converged Infrastructure) ベースの仮想化環境を新たに構築した。本稿では、当院の仮想化運営の過程を踏まえつつ、HCI 導入に至った経緯と導入効果について報告する。

2. 仮想化の価値を高める工夫と推進の経緯

2.1 仮想化推進における運用面の工夫

当院では、仮想化の費用対効果を高めるために「統合率向上」と「使用状況に応じたリソース追加」の2点を運用方針に定め、仮想化を推進してきた。

2.1.1 仮想化の統合率向上

統合率とは、物理サーバ1台あたり何台の仮想マシンが稼働するかを示す値である。統合率が高いほど、物理サーバ数が減り、費用対効果が高まる。マルチベンダーである病院情報システムにおいては、病院主導の仮想化環境構築が統合率向上のポイントとなるため、当院では地場のSI業者に仮想化環境の構築と運用管理を委託し、特定ベンダーの専用環境にならないよう計画した。これにより同一仮想化環境に複数ベンダーの仮想マシンを稼働し、統合率を高めている。

2.1.2 使用状況に応じたリソース追加

仮想サーバは、構築後にリソース(CPU、メモリ、ディスク容量)を随時追加することができ、過剰な設備投資を防止できる。新規に仮想サーバを構築する際、担当ベンダーのヒアリングを元にリソース設計をするが、オーバースペックの要求が多い傾向がある(ピーク時基準のCPU・メモリ数、5年分のディスク容量など)。当院では、ベンダーからの要求を過剰と感じた場合、要求の2割～5割減で設定し、実際の稼働状況に応じて随時リソース追加を行っている。

2.2 仮想化推進の経緯

当院は2010年に、4台の物理サーバをiSCSIスイッチ経由で1台のストレージに接続する3階層の仮想化環境を構築し、サーバ仮想化を開始した。2011年の電子カルテ導入を機に、部門システムサーバ、VDI(Virtual Desktop Infrastructure)など約200台を仮想化、その後も基盤のハード追加をしながら拡大し、2014年には仮想マシン数が約280台となった。前述の通り、統合率向上と使用状況に応じたリソース追加を意識しながら進めたことで、280台の仮想マシンに対し、物理サーバ数11台、統合率25.5となり、仮想化の費用対効果は高まったが、一方で以下の課題に直面した。

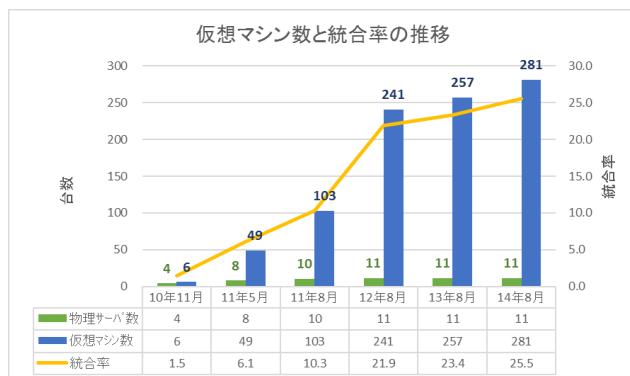


図1 仮想マシン数と統合率の推移

3. 仮想化推進において直面した課題

3.1 ストレージ装置への負荷集中によるパフォーマンス低下

3階層の仮想化環境はストレージにディスクI/Oが集中し、拡大に伴い仮想マシンのパフォーマンスが低下する。当院では、統合率を高めた結果、夜間バックアップで処理が集中する時間帯にVDIの応答遅延やスイッチ・ストレージ間の片側経路切断などの問題が発生した。

3.2 基盤更新時の高額投資と膨大な移行作業

2016年にストレージ装置がサポート終了を迎え、仮想化基盤の更新を余儀なくされた。その際、稼働中の大量の仮想マシンを受け入れる大規模な新環境が必要となり、高額な投資と膨大な移行作業が生じることとなった。さらに仮想化の拡大に伴い、途中で増設した機器は5年未満の使用期間で廃棄することとなり、活用不十分のまま役目を終えた。

3.3 大容量サーバ仮想化へのハードル

3階層の仮想化基盤は、高速化と安定稼働のために高価なストレージ製品を採用するため、PACS(Picture Archiving and Communication System)などの大容量サーバの仮想化は費用対効果が得られにくく、着手できずにいた。

3. 目的

1. 仮想マシン増加によるパフォーマンス低下が少ない仮想化環境を構築する。
2. 仮想化基盤の大規模更新を減らし、基盤更新に伴う作業負荷の軽減と投資の分散を図る。
3. 大容量サーバを仮想化し、更なる仮想化推進を図る。

4. 方法

前述の3階層仮想化で直面した課題は、いずれもストレージ装置がボトルネックであり、「負荷分散」、「製品の長期利用」、「大容量化」が改善のポイントである。「負荷分散」と「長期利用」については、複数のIAサーバを繋ぎ合わせ、大容量を作り出すSDS(Software Defined Storage)の技術を用いることで、改善できると考え、2016年10月のストレージ装置のメーカーサポート終了を機に、SDS(Software Defined Storage)をベースとするHCIの仮想化環境を新たに構築することとした。「大容量化」については、HCIより容量単価が低いNAS(network attached storage)を別途導入し、大容量データの保管コストを低減できる構成とした。NASはスケールアウト型の製品を採用し、HCIによる「製品の長期利用」を阻害しないよう計画した。

4.1 HCIの構成と稼働状況

HCI製品にはNutanixを採用し、負荷分散を目的としたノードを9台導入した。

表1 HCI初期導入ノードのスペック

CPU	Intel Xeon Processor 2.2GHz 12-core
メモリ	256GB
Disk	800GB SSD×2、2TB NLSAS×4

3階層仮想化環境(以下、旧環境)からの移行により、上記HCI環境で148台の仮想マシンを稼働させた(統合率16.4)。また、旧環境のVDIを一時的に延命する必要があったため、HCIを旧環境の物理サーバからNFSマウントし、旧環境のストレージとして使用できるようにした。

表 2 HCI 環境の仮想マシン内訳

仮想マシンの種類	台数
DB サーバ(薬剤部門、検体検査部門など)	26
その他 (IF サーバ、Webサーバ等)	122
VDI(ストレージ利用)	34

4.1 HCI の性能評価

仮想マシン移行前の旧環境と、仮想マシン移行後の HCI 環境において、同一の評価用仮想マシン (CPU:1vCPU、メモリ:2GB、ディスク:40GB) を使い、シーケンシャルアクセス (連続した領域に対する読み書き速度) とランダムアクセス (分散した領域への読み書き速度) を比較したところ、下表の通り旧環境より HCI の方が優れた数値結果となった。

表 3 3 階層環境と HCI 環境の性能比較

	Sequential (MB/s)		Random (MB/s)	
	Read	Write	Read	Write
旧環境	84.946	65.228	0.647	1.916
HCI 環境	457.191	158.987	29.352	6.327

また、旧環境と HCI 環境で同一 VDI の起動速度を比較したところ、旧環境が 2 分 39 秒であったのに対し、HCI 環境では 1 分 50 秒となり、HCI の方が 49 秒速い結果であった。

旧環境のストレージは、2010 年に導入したものであり、SSD も搭載しておらず、一概に HCI 構成の方が 3 階層構成よりも優れていると言う結論には至らないが、当院で新たな仮想環境として使用する上では問題なく、パフォーマンス向上も見込める結果であった。移行前に計測をしておらず体感での報告となるが、実際に HCI に移行した仮想マシンの再起動や画面展開の速度などは、向上している。

5. 結果および考察

5.1 負荷分散による仮想化全体のパフォーマンス向上

3 階層環境では仮想マシン増加によりストレージへディスク I/O が集中したが、SDS をベースとする HCI ではディスク I/O が複数の物理サーバに分散され、仮想マシン増加によるパフォーマンス低下が発生しにくい。3 階層環境での物理サーバ追加は、CPU とメモリの性能強化であったが、HCI では併せてディスク性能も強化し、物理サーバ追加が仮想化環境全体の高速化に繋がる。前述の通り、9 台のノード構成で 148 台の仮想マシンを約 1 年間稼働しているが、パフォーマンスの問題は発生していない。

2017 年 10 月に電子カルテを更新予定であり、100 台以上のサーバ構築が必要となったため、6 ノード追加し、HCI 環境を増強した。初期導入時は、負荷分散重視のノードを選択したが、今回は、DB サイズが大きなサーバが多いため、容量重視のノードを採用した。

表 4 HCI 追加ノードのスペック

CPU	Intel Xeon Processor 2.2GHz 12-core
メモリ	384GB
Disk	1.92TB SSD×2、4TB NLSAS×10

稼働前評価や稼働後にパフォーマンス上の問題が発生した場合は、オールフラッシュモデルのノードを追加し、ピンポイントでのパフォーマンス対策も可能である。このように HCI では、ノードタイプの選択で容量やパフォーマンスを細かくコントロール可能となり、リソース制御の柔軟性が増す。

5.2 仮想化基盤長期利用による投資分散と移行作業の排除

3 階層基盤のストレージは 1 台の制御ユニットに複数の拡張ユニットを下層配置する構成であり、製品サポート終了時に全面入替が必要となる。それに対し、HCI では各ノードが並列の関係で構成され、5 年経過ノードを順次入替ながら、仮想化基盤を長期間使用できる。これまで通り、使用状況に応じたリソース追加を意識することで、ノードの使用期間にずれが生じ、基盤更新の費用を分散できると考えている。また、部分的なノード入替により仮想化環境は無停止のまま継続利用できるため、基盤更新時の仮想マシンの移行が不要となり、作業負荷を大幅に軽減できる。

当院の HCI 環境は、初期導入時に旧環境の移行に必要な容量しか設けず、前述の通り、電子カルテ更新に伴い 6 ノード追加することとなった。追加の 6 ノードについても、メーカーが提示した 5 年間の必要量を満たしておらず、状況に応じ随時追加する計画である。

5.3 大容量サーバ仮想化の実現

100TB の画像ファイルを有する PACS サーバを仮想化するにあたり、システム領域とプログラム領域 (C、D ドライブ) に高速な HCI を割り当て、大容量を必要とする画像領域に比較的容量単価が低い NAS を割り当てた。この PACS サーバは、現在過去データのコンパルト中であり、2017 年 10 月に本稼働予定である。

前述の通り、HCI の長期利用を阻害しないためにスケールアウト型の NAS を採用したが、320 列 CT や 3 テスラ MRI の導入など、PACS サーバは予想外の容量増加も多く、拡張が容易な点でもスケールアウト型のメリットがある。

5.3 仮想化運営における閾値設定

当院では、冗長化対策に必要な余力を残しつつ、仮想化の費用対効果を高めるために CPU・メモリ・ストレージの使用率 80% を閾値と定め、これを超えた場合は、物理サーバやストレージユニットを増設しながら 3 階層仮想化環境を運営してきた。しかしながら、ディスク負荷に対してのモニタリングが不十分で、パフォーマンス低下の発見が遅れたため、今後はディスク負荷に対しても適切な運用管理をしていきたい。

ディスク負荷の閾値として、IOPS などの数値データを定期的に評価することが必要と考えるが、ディスク負荷は曜日や時間帯によって大きく変化し、システムの種類や運用によっても許容範囲が異なるため、現時点で閾値を定めるのは難しい。旧環境では、統合率 25.5 の時点で、ディスク負荷起因のパフォーマンス低下が発生したため、この数値がパフォーマンス低下の限界点の一要素であったと考え、モニタリング項目への追加を検討したい。

SDS という新たな技術をベースとし、リソース追加の柔軟性が増した HCI では、基盤構成の複雑性が増すと予想され、これまで以上にモニタリングが重要となる。3 階層仮想化と同じ閾値では、HCI の価値を引き出せず、不具合を未然に防止できない可能性もあるため、閾値の見直しや管理体制の自動化なども検討し、適切な運用管理方法を確立したい。

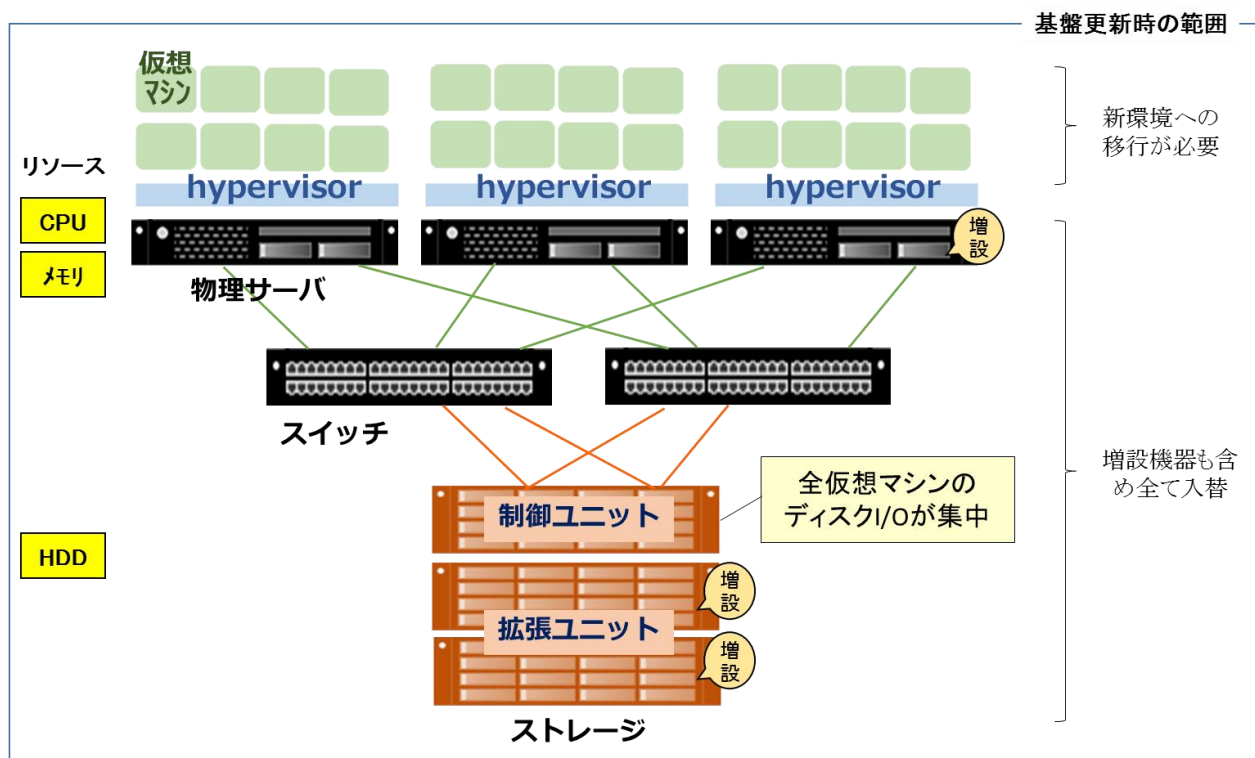


図 2 3 階層仮想化環境の構成と基盤更新時の範囲

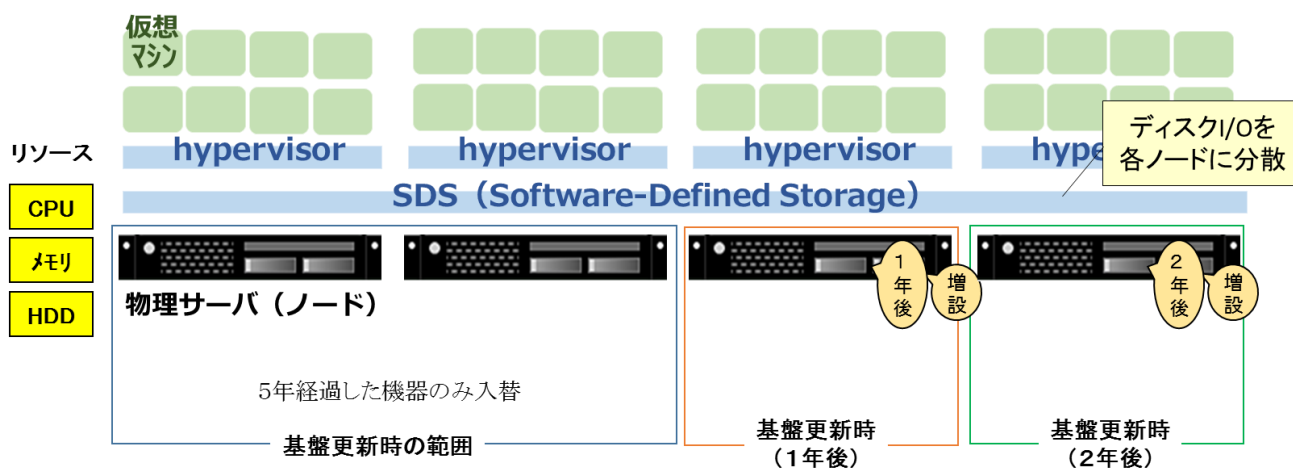


図 3 HCI 仮想化環境の構成と基盤更新時の範囲