

仮想化環境の設計手法 ～プロのテクニック教えます～

日本仮想化技術株式会社
代表取締役社長兼CEO 宮原 徹
miyahara@VirtualTech.jp

VirtualTech Japan

日本仮想化技術株式会社 概要

- 社名: 日本仮想化技術株式会社
 - 英語名: VirtualTech Japan Inc.
 - 略称: 日本仮想化技術/VTJ
- 設立: 2006年12月
- 資本金: 14,250,000円
- 本社: 東京都渋谷区渋谷1-1-10
- 取締役: 宮原 徹 (代表取締役社長兼CEO)
- 伊藤 宏通 (取締役CTO)
- スタッフ: 8名 (うち、5.5名が仮想化技術専門エンジニアです)
- URL: <http://VirtualTech.jp/>
- 仮想化技術に関する研究および開発
 - 仮想化技術に関する各種調査
 - 仮想化技術に関連したソフトウェアの開発
 - 仮想化技術を導入したシステムの構築

日本初の独立系
仮想化技術専門会社
(当社調べ)

2

仮想化環境構築をトータルサポート

設計

↓

導入

↓

運用保守

- 設計
 - サーバ、ストレージからネットワークまでアプリケーションまで考慮した設計
 - キャパシティプランニング (ベンチマーク)
- 導入
 - 仮想化ソリューションパッケージの提供
 - 仮想化統合 (P2Vレガシーマイグレーション)
- 運用保守
 - エンジニア教育
 - 技術サポートの提供
 - Xenソースコードレベルサポート

ベンダーニュートラルなワンストップ・サポートをご提供

4

本日のアジェンダ

- 仮想化でコスト削減するには
 - 現行システムインフラの課題と仮想化
 - サーバー仮想化のコスト削減効果
- 仮想化環境の設計手法
 - キャパシティプランニング
 - 性能を考慮したハードウェア選択

4

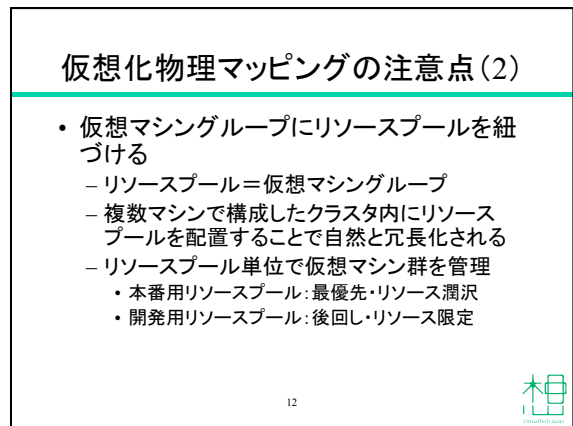
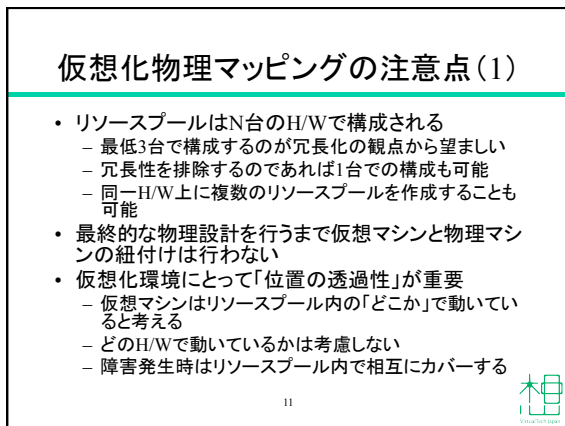
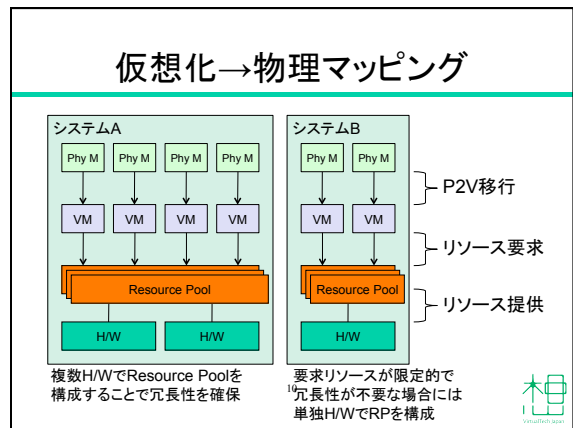
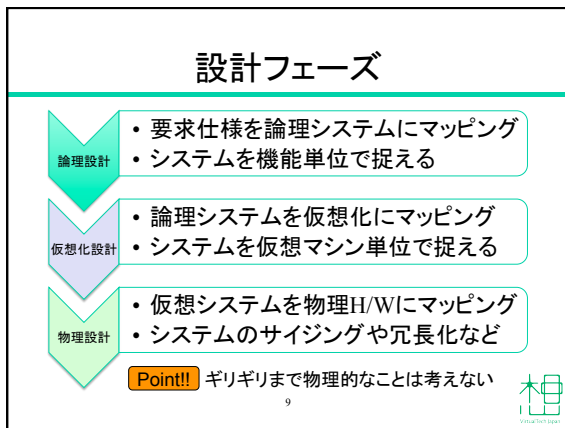
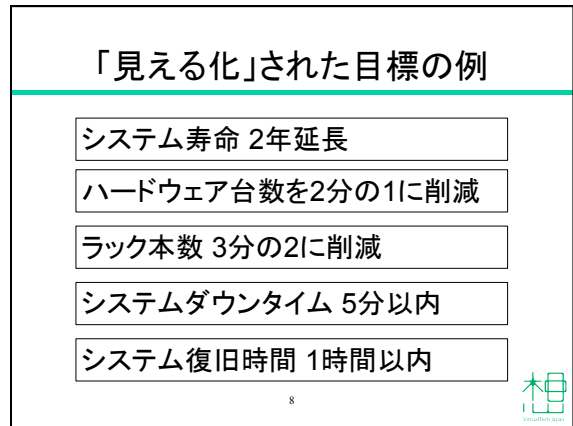
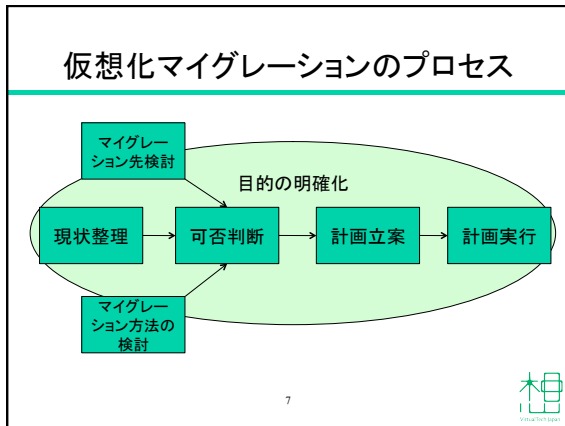
仮想化環境の設計手法

VirtualTech Japan

仮想化環境設計のポイント

- 目的の明確化
 - 既存環境移行 ← 現状こちらの方が多い
 - 新規システム構築
- 何を優先するのか
 - コスト: 導入コスト、ランニングコスト
 - 機能: 運用管理、可用性
 - 性能: 実行速度、処理容量

6



設計手順詳細(1)

1. 移行対象サーバーのリストアップ
 - マシンスペックおよび使用率
 - 重要度および負荷率をABCランク分け
 - 移行不要サーバーは除外
2. 特記事項の確認
 - 冗長構成
 - スタンバイ構成 など
3. マシングループ毎に仕分け
 - システムを構成するWeb+DB+その他をグループ化

13



サーバーリストの例

サーバ名	OS	CPU	メモリ	ディスク
サーバ1	Windows Server 2012 R2	12%	20%	10%
サーバ2	Windows Server 2012 R2	15%	25%	12%
サーバ3	Windows Server 2012 R2	10%	18%	8%
サーバ4	Windows Server 2012 R2	18%	28%	15%
サーバ5	Windows Server 2012 R2	12%	20%	10%
サーバ6	Windows Server 2012 R2	15%	25%	12%
サーバ7	Windows Server 2012 R2	10%	18%	8%
サーバ8	Windows Server 2012 R2	18%	28%	15%
サーバ9	Windows Server 2012 R2	12%	20%	10%
サーバ10	Windows Server 2012 R2	15%	25%	12%

14



設計手順詳細(2)

4. グループ毎の要求リソース量を算出
 - CPUクロック数、メモリ量、ディスク容量、ディスクI/O、ネットワークI/Oなどの積算値
5. ターゲットH/Wによるリソースプールと比較
 - リソースプールは最低3台の仮想マシンホストで構成(冗長構成を考慮した場合)
 - CPUクロック数、メモリ量の積算値×60%程度
 - 不足の場合はリソース追加を検討
 - ストレージはパスの帯域幅やディスク本数などで性能値が左右されるので別途検討が必要

15



ハードウェアの選定

VirtualTech Japan

VirtualTech Japan

VirtualTech Japan

仮想化環境の性能要因

- ハードウェア
 - CPU: コア数、クロック速度、仮想化支援技術
 - メモリ: 容量、速度、VMへの割当量
 - I/O: 帯域、デバイス速度、仮想化支援技術
- ソフトウェア
 - VMM種別: ホストOS型、ハイパーバイザー型
 - 内部構造: ドライバモデル、スケジューラー
 - その他: OS、アプリケーション、データ量

17

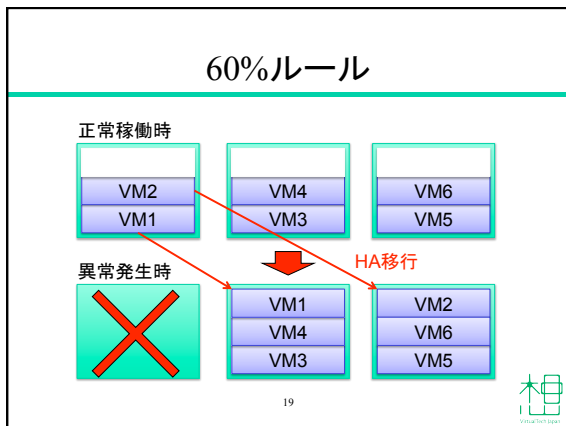


ハードウェアの選定ポイント

- CPUは仮想化しても性能が下がりにくい
 - マルチコアは4コアが標準になりつつある
- メモリ量が仮想マシン収容力を大幅に左右
 - 32GB~64GBが当たり前に
 - RVI、TLBなど仮想化支援技術も標準に
- 高速なI/Oの装備が必須
 - FCまたは10G Ethernet
 - Intel VT-d/AMD IOMMUのサポート

18





CPU使用率の計算方法

- CPUクロック数の世代間性能差に留意
 - CPUの性能指標であるクロック数は製品の世代によって性能が異なる
 - 実質的なクロック対性能は大きく変わっていないので、クロック数比で計算してもよい?

CPU	コア数	MHz	実質MHz	SPECint2000	MHz/SPECint2000
Xeon 3.0GHz(推定)	1	3000	3000	1,429	2.10
Xeon 3.4GHz	1	3400	3400	1,617	2.10
Xeon 3.0GHz	1	3600	3600	1,718	2.10
Xeon 5110(1.6GHz)	2	1600	3200	1712	1.87
Xeon5160(3GHz)	2	3000	6000	3,025	1.98

20

CPU使用率の計算例

- CPU使用率30%の物理マシンを、ほぼ同性能・同クロックの仮想マシンホストに移行
 - CPU使用率60%まで可能なので、2VMまで収容可能

収容可能VM目安 = CPUコア数 × 2

むしろメモリ容量の制約の方が大きい

21

ブレード or ラックマウント?

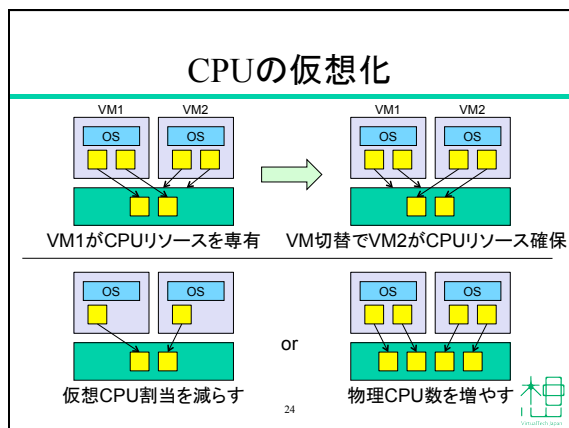
- 仮想化環境を組むなら最低3台構成
 - RAID 5の考え方と同じ(最低HDD3台)
 - 60%ルールを基準にCPU、メモリなどを選定
- 4台以上ならブレード?
 - 以後の増設計画にも左右される
- ブレードの弱点も克服されつつある
 - HP BL495cではメモリスロット16本、最大128GBメモリ搭載可能、10Gb Ethernet標準

22

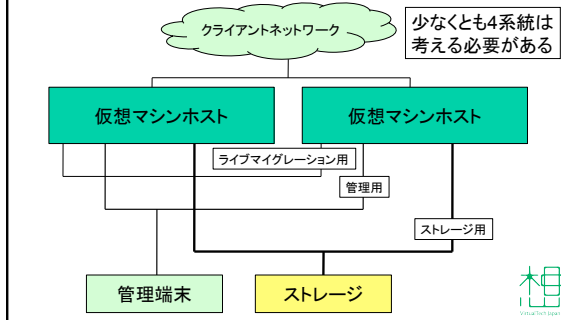
その他サーバ選定のポイント

- 高クロックよりもコア数
 - CPUロックが減少し、仮想マシンの実行並列度が高まる → 全体の性能が向上
 - 低消費電力型を選択できる → ランニングコストの削減
- スタンバイサーバは不要
 - 物理的な区切りで考えず、全体のリソース容量で考えること
 - 60%ルールの徹底

23



ネットワーク構成例

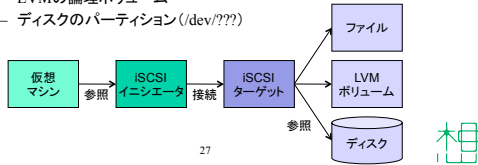


ストレージの選定

- 高速なストレージの選定
 - どの程度の速度が必要か
 - I/Oの性質は？(I/Oブロックサイズの大小)
 - HDDの台数
- 接続方法は？
 - FC接続: 高速だが気軽ではない
 - iSCSI接続: 小規模向けに人気急上昇
 - NFS接続: 扱いやすい、意外と高速

iSCSIとは

- SCSIコマンドをTCP/IPでやり取りする仕組み
- iSCSIイニシエーターからiSCSIターゲットに接続
 - SCSIディスク(/dev/sd?)として参照可能
- iSCSIターゲットが参照可能なもの
 - ファイル
 - LVMの論理ボリューム
 - ディスクのパーティション(/dev/???)



ストレージ比較表

	扱いやすさ	性能	コスト	規模
DAS	簡単	普通	安い	小規模
NAS	普通	やや速い	安い	小~大規模
FC SAN	難しい	速い	高い	中~大規模
iSCSI	やや難しい	やや速い	やや高い	中~大規模

大まかな比較(構成によって例外あり)

- 性能: FC SAN > NAS ≥ iSCSI ≥ DAS
- 扱いやすさ: DAS > NAS > iSCSI > FC SAN

ストレージの追加機能

- 無停止での容量追加
 - ゲストOS側での対応も必要
 - ダイレクト接続+動的ボリューム管理で対応できるが、仮想マシンの可搬性は低下
- スナップショット
 - 仮想マシンのバックアップに有効だが、費用対効果の見極めが肝心
- レプリケーション
 - DRサイト構築に有用

ベンチマークによる性能検証



ベンチマークとは

- 処理速度の測定作業
 - あるハードウェア、ソフトウェアの処理速度を知りたい、比較したい
 - ある処理に必要なハードウェア、ソフトウェアを知りたい
- 測定対象
 - CPU、ストレージ、ネットワーク、メモリ
 - ソフトウェアの処理速度

31



仮想化で使えるベンチマーク

- VMmark
 - VMwareの開発したベンチマーク
 - DBやWebなど、複数のVMを1セット(タイル)にして、ハードウェアのキャパシティを測定
- SPEC*
 - 業界標準のベンチマークを多数提供
- Iometer
 - ディスクI/Oを測定するベンチマーク
- ab (Apache Bench)
 - Apacheの性能測定用
- 他にも多数

32



仮想化ベンチマークの注意点

- ベンチマーク負荷作業は外部から
 - 仮想マシン内では時間測定が不正確
 - 内部で実行する場合でも、時間は外部から実行時に取得する
- キャッシュバッファの影響を考慮
 - レイヤーが1段増えることで構造が変わり、キャッシュが増える場合もある=見かけ上の性能が物理サーバーを超える場合がある
 - 扱うデータ量や、VM数が増えればキャッシュが効かなくなり、性能低下を起こす場合もある

33



参考: Xen+NFSを使用した
DBベンチマーク結果

VirtualTech Japan

VirtualTech Japan

検証環境

- ML350 G5 x 2台
 - Xeon 5150(2.66GHz/Dual Core) x 1プロセッサ
 - 16GBメモリ
 - RAID 0(SAS HDDx8/128MBキャッシュ)
 - ギガビットイーサネット
- 仮想マシン
 - 1仮想CPU
 - 2GBメモリ
- PostgreSQL 8.3.4
 - pgbench(TPC-B相当のDBベンチマーク)

35



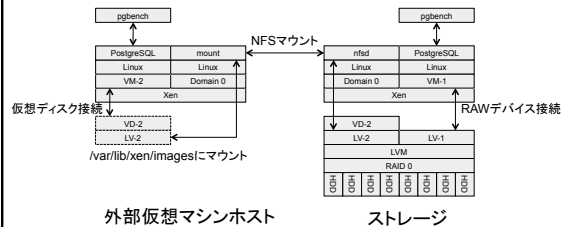
ストレージ環境

- NFSサーバーにてRAID 0ディスクを構築
- LinuxのLVMにて、LVを2つ作成
 - LVは10GBサイズ
 - LV-1はVM-1にRAWデバイス接続
 - LV-2はext3で初期化後、NFSでエクスポート
 - オプション: rw, sync, no_root_squash, no_wdelay
 - VM-2でLV-2領域をマウント
 - オプション: 無し、または rsize=8192, wsize=8192
 - 仮想ディスクファイル(8GB)を作成

36



ベンチマーク環境模式図



37



ベンチマーク作業

- pgbenchを使用
 - TPC-B相当のトランザクション
 - SELECT, UPDATE, INSERT
 - 検索(SELECT)のみも実行可能
- 今回の実行条件
 - スケール: 20 (2,000,000行/約300MB)
 - クライアント数: 20 (スケールと合わせる)
 - トランザクション数: 1000
 - 最初の5回の結果は破棄し、6回~10回のうちの上下1件ずつを取り除いた3回の平均値を取得

38



pgbenchのトランザクションSQL

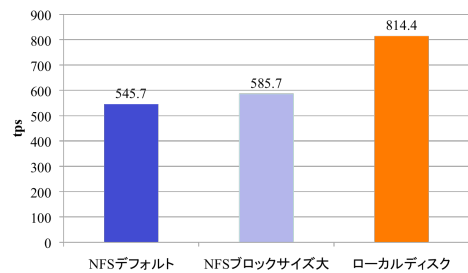
以下のSQLを1つのトランザクションとして実行

1. BEGIN;
2. UPDATE accounts SET abalance = abalance + :delta WHERE aid = :aid;
3. SELECT abalance FROM accounts WHERE aid = :aid;
4. UPDATE tellers SET tbalance = tbalance + :delta WHERE tid = :tid;
5. UPDATE branches SET bbalance = bbalance + :delta WHERE bid = :bid;
6. INSERT INTO history (tid, bid, aid, delta, mtime) VALUES (:tid, :bid, :aid, :delta, CURRENT_TIMESTAMP);
7. END;

39



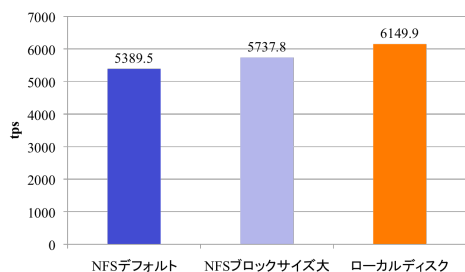
検索更新処理結果



40



検索のみ処理結果



41



結果の考察

- NFSを使用した場合、検索更新処理ではローカルディスクの65%程度の性能
- NFSブロックサイズを8KBにすると、ローカルの70%まで性能向上(5%アップ)
- 検索のみ処理の場合、顕著な性能差が認められないため、UPDATEおよびINSERT処理のオーバーヘッドによる性能劣化と考えられる

42



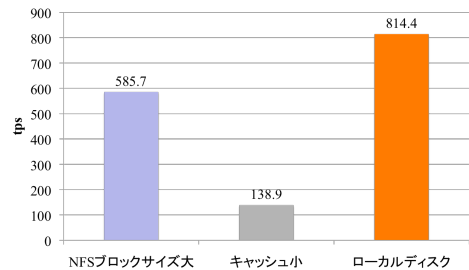
NFSサーバーキャッシュの影響

- NFSサーバーの利用するキャッシュの影響を測定
- NFSサーバーの搭載するメモリを16GBから512MBに変更し、同一内容を測定
 - システム起動時でキャッシュバッファが130MB程度
 - VM-2の仮想ディスクサイズは8GB、DBサイズが300MBのため、キャッシュにすべてが乗り切らなくなる
- NFSサーバーキャッシュが不足している場合、検索更新処理ではローカルディスクの17%まで性能劣化
- 検索処理はキャッシュ容量が少なくてもあまり影響を受けないが、性能ピークへの到達に時間がかかる
 - PostgreSQLのローカルバッファやインデックスが影響？

43



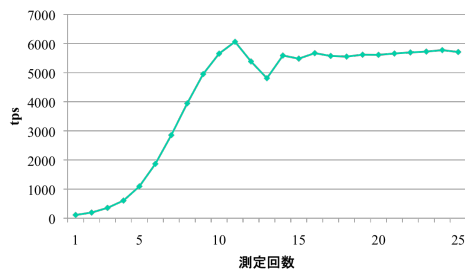
検索更新でのキャッシュの影響



44



検索のみ処理の結果推移



45



考察

- 検索主体のDBでは、NFSをストレージとして使用しても性能劣化は少ない
 - NFSストレージのキャッシュが少ない場合でも、ローカルバッファやインデックスの効果が認められる
- 検索更新のDBでは、NFSをストレージとして使用すると性能劣化が認められる
 - 今回の条件では30%程度の劣化
 - NFSサーバーのキャッシュ容量が不足すると性能劣化の度合いが激しくなる
 - 仮想マシン数が増えたと、NFSサーバーのキャッシュ容量が不足する可能性があるので注意

46



仮想化について相談したい

無料コンサルティング実施中

- ブレードサーバと組み合わせたサーバ仮想化ソリューションのデモも可能

お気軽にお問い合わせください

47



お問い合わせ先

「仮想化環境を構築したいが、どこに相談すればいいの？」

まずは我々にご相談ください

日本仮想化技術株式会社

<http://VirtualTech.jp/>
sales@VirtualTech.jp
050-7571-0584

48

