# Xen基本設計および運用TIPSについて

株式会社 BeaconNC
國武　功一

# Agenda

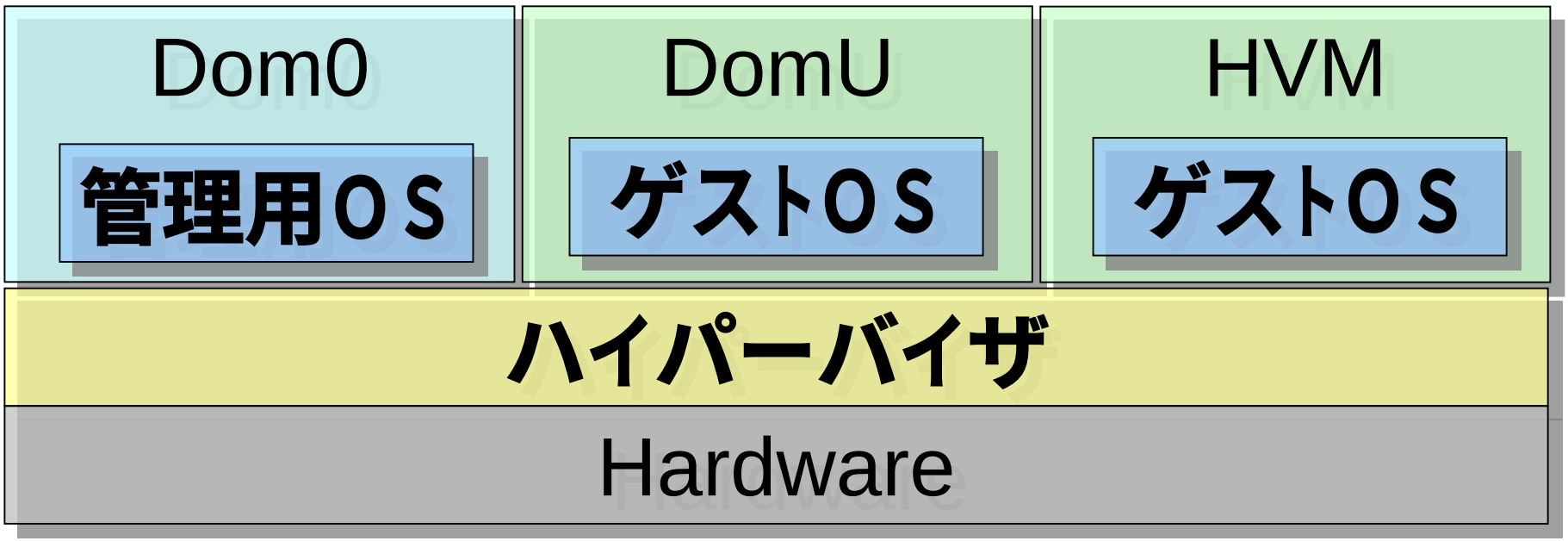
- Xenでできること
- 管理ツールでできること
  - 例：OracleVM
- 基本設計で考慮すべき点
  - なにをサービスとして提供するのか
- 運用TIPS

# Xenで出来ること

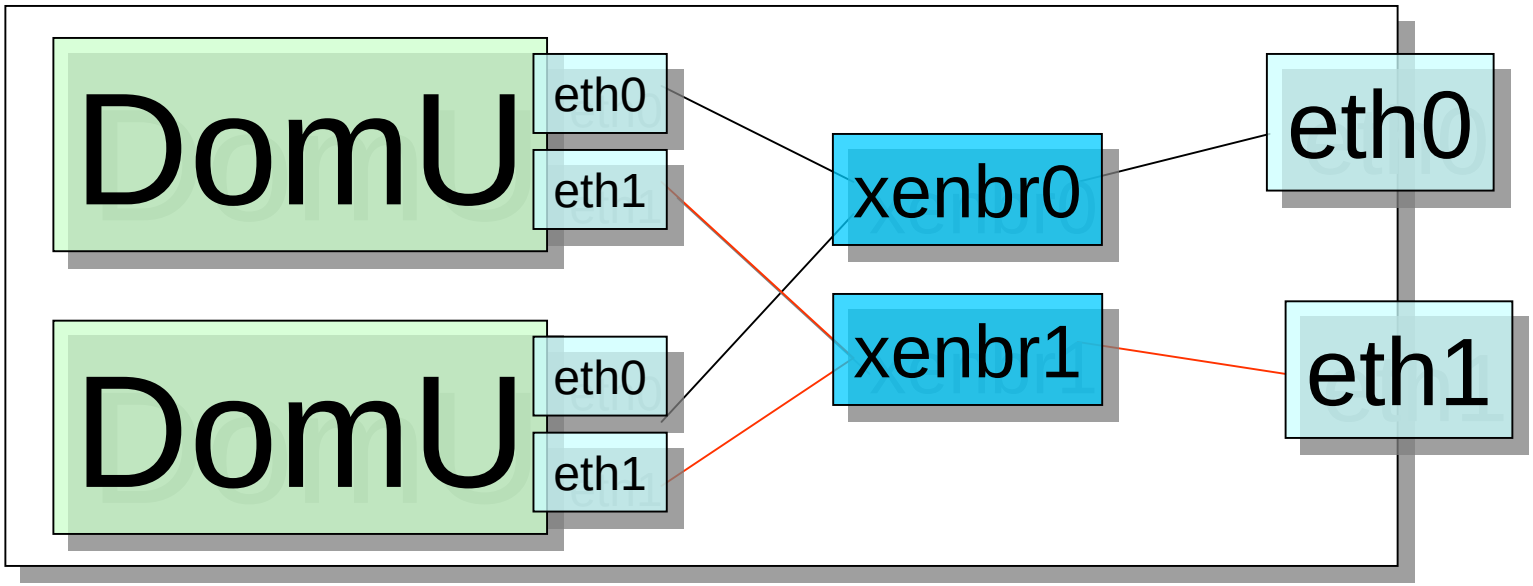
B*NC

- サーバの準仮想化。ＶＴ技術を利用したサーバの完全仮想化
- CPUスケジューラの選択
  - SEDF(Simple Early Deadline First scheduler)
    - default
  - BVT(Borrowed Virtual Timer Scheduler)
- CPU割り当て数の変更
- 物理CPUへのマッピング
- ドメインUの一時停止・リブート・強制停止
- ドメインUの稼動状態の保存、リストア、移動
- ドメインUの別物理サーバへのライブ移行
- メモリ割り当ての動的な割り当て（準仮想化）
- ドメイン０からのコンソール接続
- などなど

# Xenの構成要素(概略)



# Xenのネットワーク例 (概要)

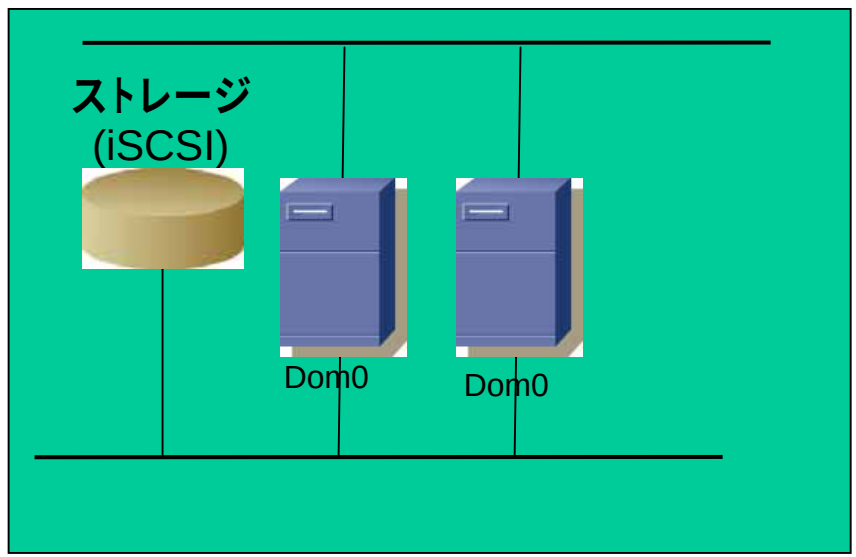



## Bridgeの状態(DomUx1の場合)

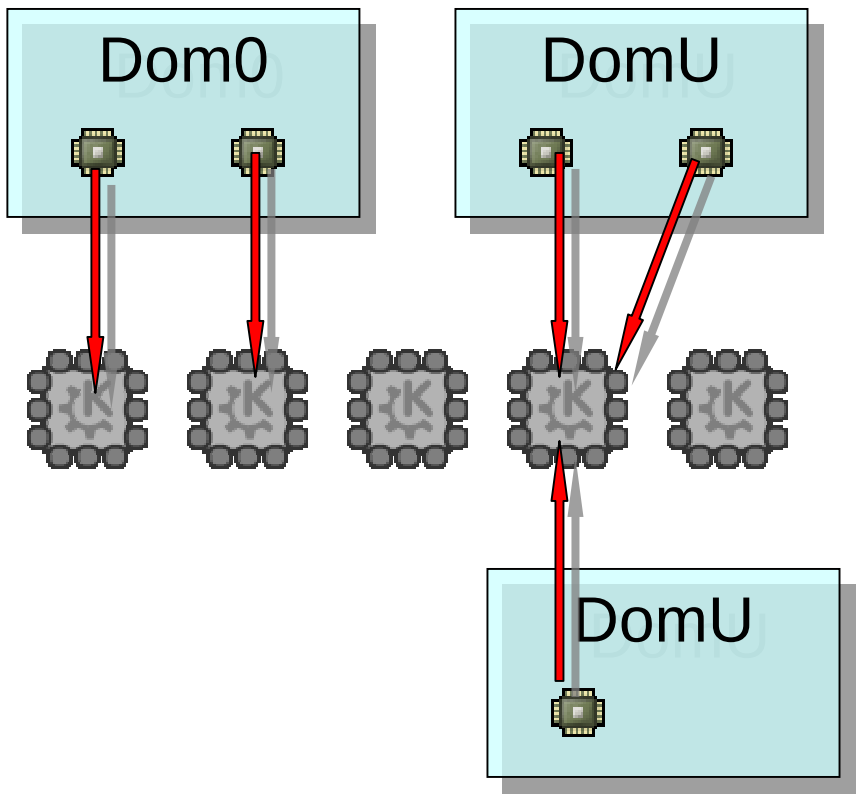
```
[root@dom0 ~]# brctl show
bridge name     bridge id       STP enabled     interfaces
xenbr0          8000.fefffffffff no              vif10.0
                                                 peth0
                                                 vif0.0
xenbr1          8000.fefffffffff no              peth1
                                                 vif0.1
```

(*)IPマスカレード用のvirtbrもあり

- iSCSIなど共有ディスクを使っていると、ライブ移行も可能
- ローカルディスク上に、ファイルを作り、それをディスクとして扱うことも、またパーティション（ＬＶでも可能）などを直接扱うことも可能



- 仮想CPUを物理サーバに固定することも可能



# 管理ツールでできること

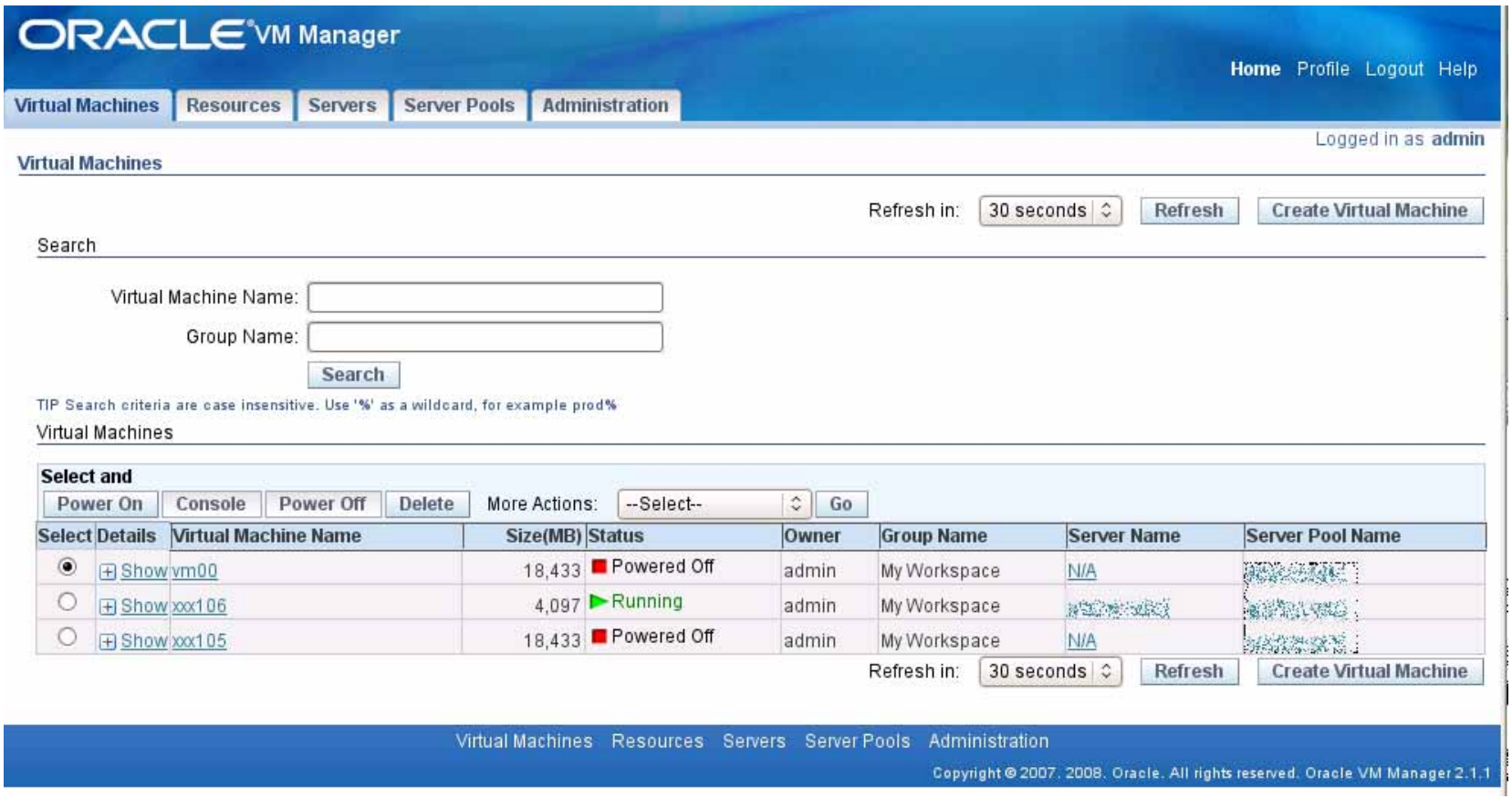
- コマンドラインベースでも運用出来るが、なんらかのGUI/Web UIツールを用いることが多い。
  - RHEL標準のvirt-manager
  - XenServer
  - VirtualIron
  - 3tera
  - OracleVM
  - and so on….

# OracleVMの機能

- Webブラウザベースの管理ツール
- テンプレートを用いたサーバ構築
- P2V , V2Vをサポート
- サーバ間をSSLで保護したライブマイグレーションのサポート
- HAクラスタリングのサポート
- I/Oリソース管理

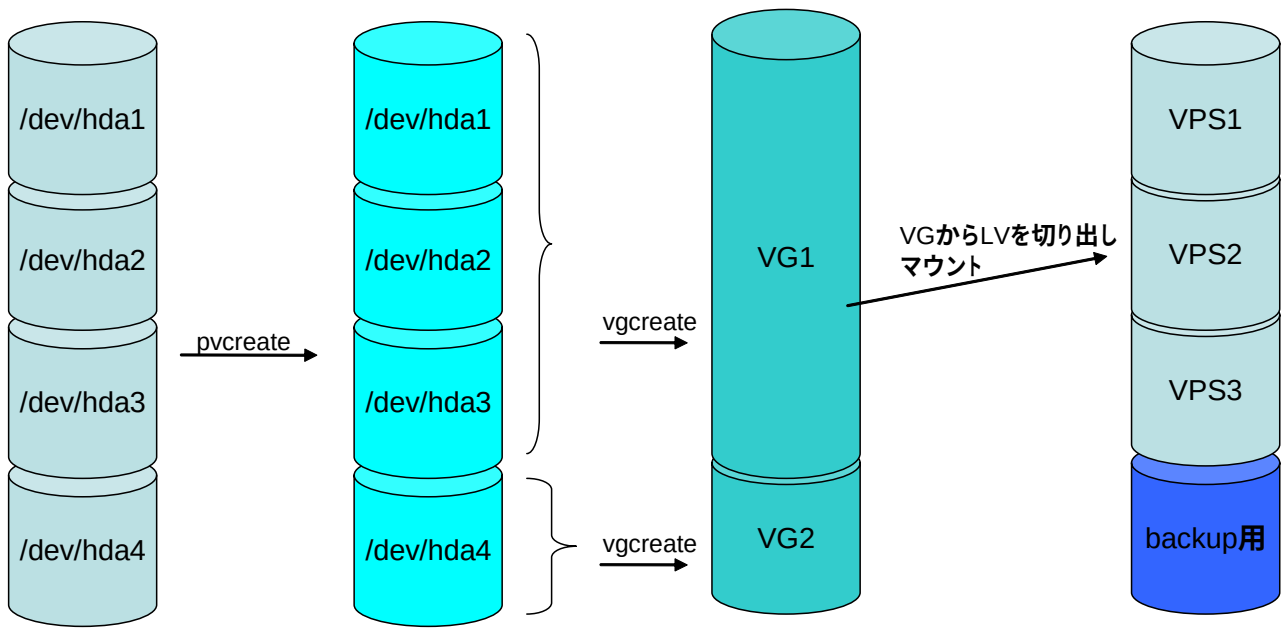
# OracleVM管理画面例

- 権限管理も可能



- なにをサービスとして提供するのかを明確にする
  - ライブ移行は本当に必要なのか？
    - 共有ディスクの持たせ方が重要になる
- リソース配分をどうするのか
  - 固定でよいのか。追加する場合、余剰リソースをどれぐらい確保すべきなのか
- ファイルシステムをどうするのか
  - LVMでスナップショットバックアップ？
- ネットワーク設計
  - すべて同一セグメントでよいのか？

- サーバまるごとバックアップ
- サーバのクローニング
  - LVMのスナップショット機能などを有効活用



# 共有ディスクを持たせる場合

- パフォーマンスに留意
  - iSCSIもよいソリューションだが、本当にパフォーマンスが要求される場合には、不安点がある。
  - 共有ディスクがボトルネックになりがち。そのデメリットをどこまで許容するのか
- ライブ移行
  - ネットワーク設計を事前に行う必要あり
  - 内部ネットワークの処理をどうするのか？（VLAN?)

# 運用TIPS

- 多段LVMにはどう対応？
- いっせいのーで！
- 仮想マシンのDNA

# 多段LVMにはどう対応？

- LVMで切り出した領域をインストール領域に指定すると、CentOSなどは、さらにそこにLVMを作成する。
  - Dom0から、DomUを起動させずにこのファイルシステムにアクセスできるか？

    # mount –o loop /dev/mapper/Xen_guest /mnt はできない
    - kpartxが便利！

# How to use : kpartx

mount**方法**

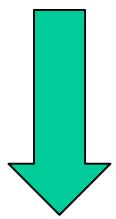
```
# losetup –f  <= 使用されていない最初の loopデバイス名を表示
# losetup /dev/loop0 /dev/mapper/Xen_guest
# kpartx –v –a /dev/loop0
# vgchange –ay
# mount /dev/mapper/XXX-XXXX /mnt
```

umount**方法**

```
# umount /mnt
# vgchange –an
# kpartx  -v –d /dev/loop0
# losetup –d /dev/loop0
```

- 仮想マシンの作成
  - 初期インストール
    - 追加パッケージの導入
    - 不要サービスの停止
    - 管理ユーザの作成
    - ……
  - 実稼動スタート
    - 最初はいいんだけど……

- 特定の時間にいきなり負荷があがる！
  - ログの正確性を担保するため、ntpで正確に時刻同期を実施している。
    - 標準のcronが特定の時刻に一斉に起動。I/Oを中心にボトルネックが発生し、load average が急浮上。

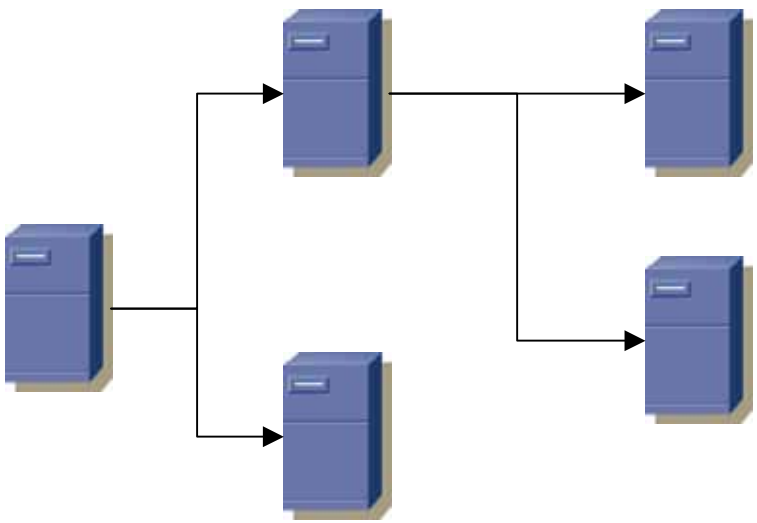


**サーバを新しく追加したら、cronをずらそう！（涙）**

B*NC

- あるとき、メンテナンス時にファイルシステムの障害が多発。
- fsckでファイルシステムにトドメをさしてしまうことも……
- 不安定？では定評があったReiserFSを利用していた。
- ReiserFS出入り出入り禁止で落ち着きそうだったが…..

- それらの仮想マシンは、クローニングで作成されていたマシンだった。
- 実は詳しく追って行くと、ひとつの親サーバに行き着いた……



- Rsyncでベースとなるデータのみをコピーするのとは違い、ファイルシステム毎、コピーすることが多い。
- この場合は、ベースとなるサーバのファイルシステムが壊れていると、大変なことに……
- 問題のトレースは留意する必要あり。

- 基本機能は限られているが、管理ツールでできることの幅が広がる (やるべきことも増える)
  - 自作したくなるけど……
- 管理すべきサーバ数は確実に増えてしまうので、それを考慮して管理ツールを選定すべき (でも意外と泥臭い運用も必要)
- 思わぬ運用TIPS (バッドノウハウ) 習得の必要性は増える……