

マルチクラスタによる EUCALYPTUSの拡張

2010/5/22



Japan Eucalyptus User Group

株式会社 NTTDATA
技術開発本部
岡本 隆史

日本電信電話株式会社
情報流通プラットフォーム研究所
境 美樹

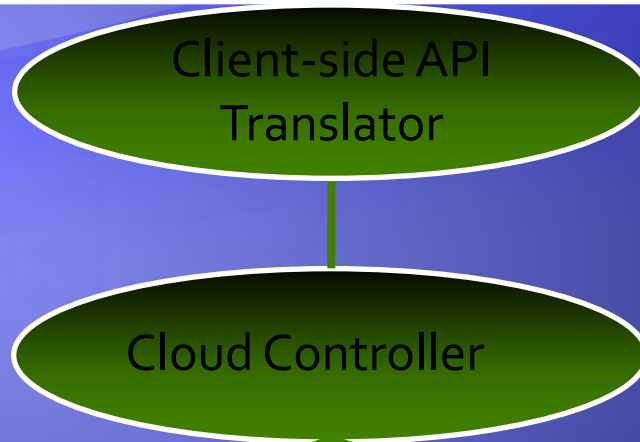
アジェンダ

- Eucalyptusのマルチクラスタ
- マルチクラスタにおけるネットワーク
- 障害発生時の挙動
- まとめ

※本資料はEucalyptus 1.6.2とCentOS 5.4の組み合わせによる検証をもとに作成しています
記載されている会社名、商品名、又はサービス名は、各社の登録商標又は商標です。

Eucalyptus 1.5のアーキテクチャ

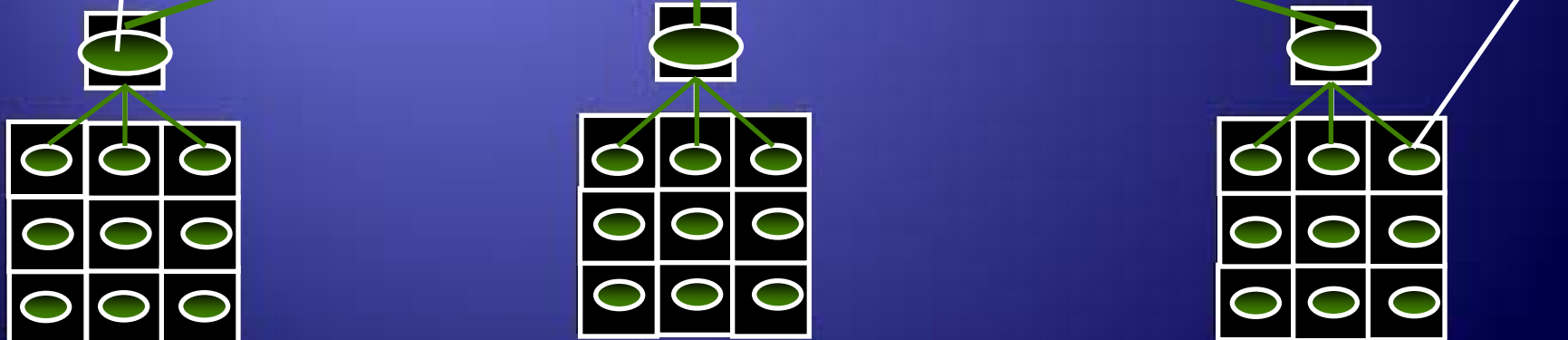
Amazon EC2 Interface



複数のクラスタで拡張できることを売りにしていたが実際にはクラスタは1つしか使えなかった。

Cluster Controller

Node Controller



Eucalyptusのアーキテクチャ(1.6~)

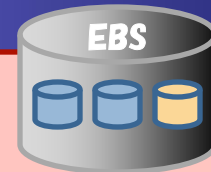
HybridFox/Eclipse AWS
などのクライアントツール



AWS API
Cloud Controller



Warlus(S3)



NAT/
FireWall DHCP
Cluster Controller

Storage Controller

VLANでユーザ毎に
ネットワークを分離

DHCPで仮想マシンにプライベートIPを割り当て、NATで外部アドレスへ変換

PublicIP:10.68.x.x

PrivateIP:192.168.x.x

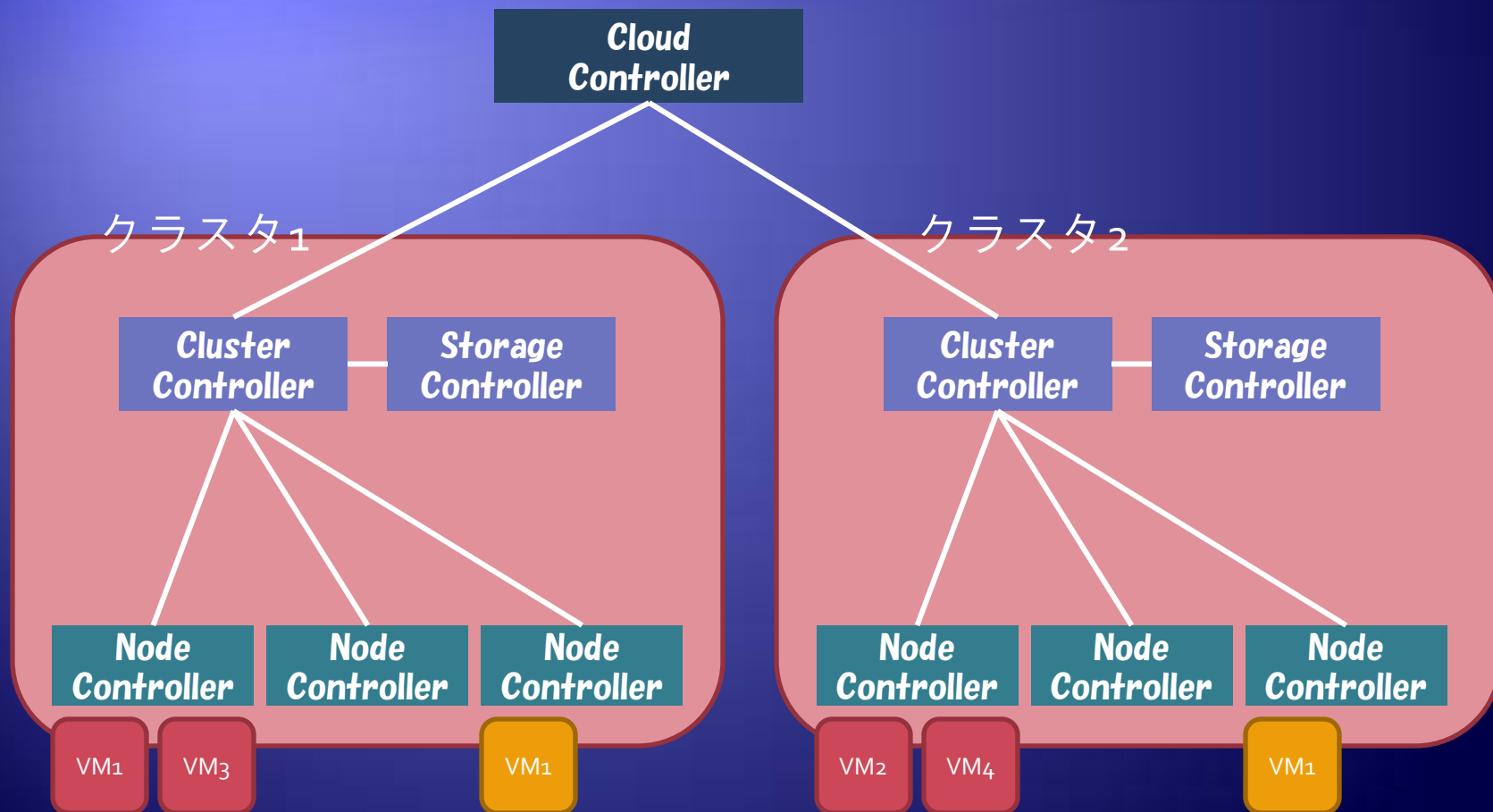
VM VM VM
VMM
Node Controller

VM VM VM
VMM
Node Controller

VM VM VM
VMM
Node Controller

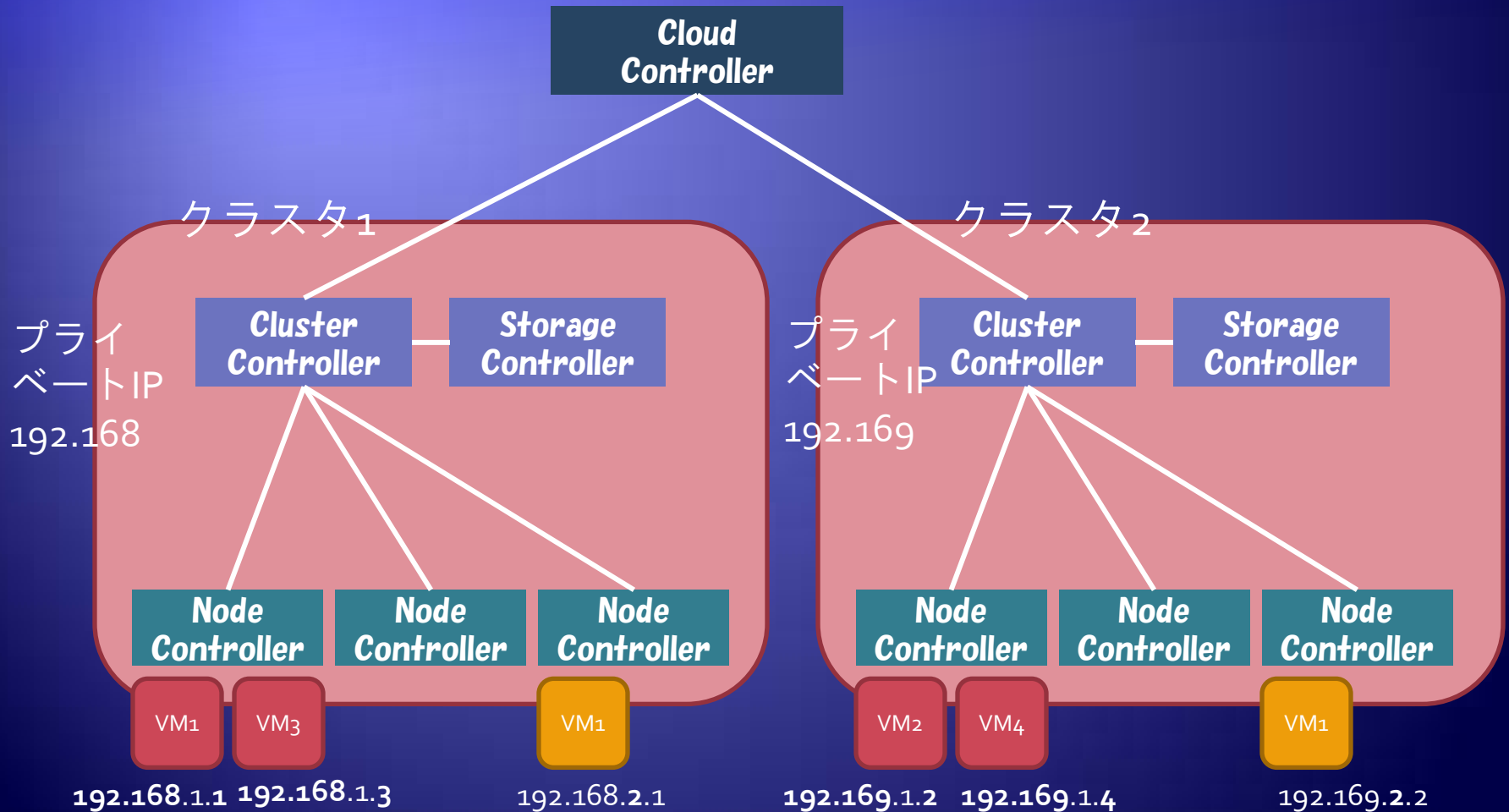
マルチクラスタの概要

- クラウドを複数のグループに分割可能。
- クラスタはAmazon EC2のZone(データセンタ内のグループ)相当
- Cloud Controllerや他のクラスタを停止することなく、クラスタの追加や削除を可能
 - リソースが不足した際にクラスタ単位に動的に追加したり、障害が発生したときに特定のクラスタを停止することが可能



クラスタにおけるVMへのIPの割り当て

- IPの下から2番目の数字でユーザが一意に決まる
 - >クラスタを利用しても利用可能なユーザ数が増える訳ではない
- IPの下から1番目の数字でユーザの仮想マシンが一意に決まる。
 - >クラスタを利用しても1ユーザが利用可能な仮想マシン数が増える訳ではない。



【参考】ユーザ毎のIP割り当ての仕組み

ユーザ	割り当てられるPrivateIP
ユーザ ₁	192.168.1.1-192.168.1.126
ユーザ ₂	192.168.1.129-192.168.1.254
ユーザ ₃	192.168.2.1.-192.168.2.126
...	
ユーザ ₄₉₉	192.168.255.1-192.168.255.126
ユーザ ₅₀₀	192.168.255.129-192.168.1.254

ネットマスクの下7ビットで各ユーザの仮想マシンIDを設定
下16~8ビット目でユーザのIPを設定

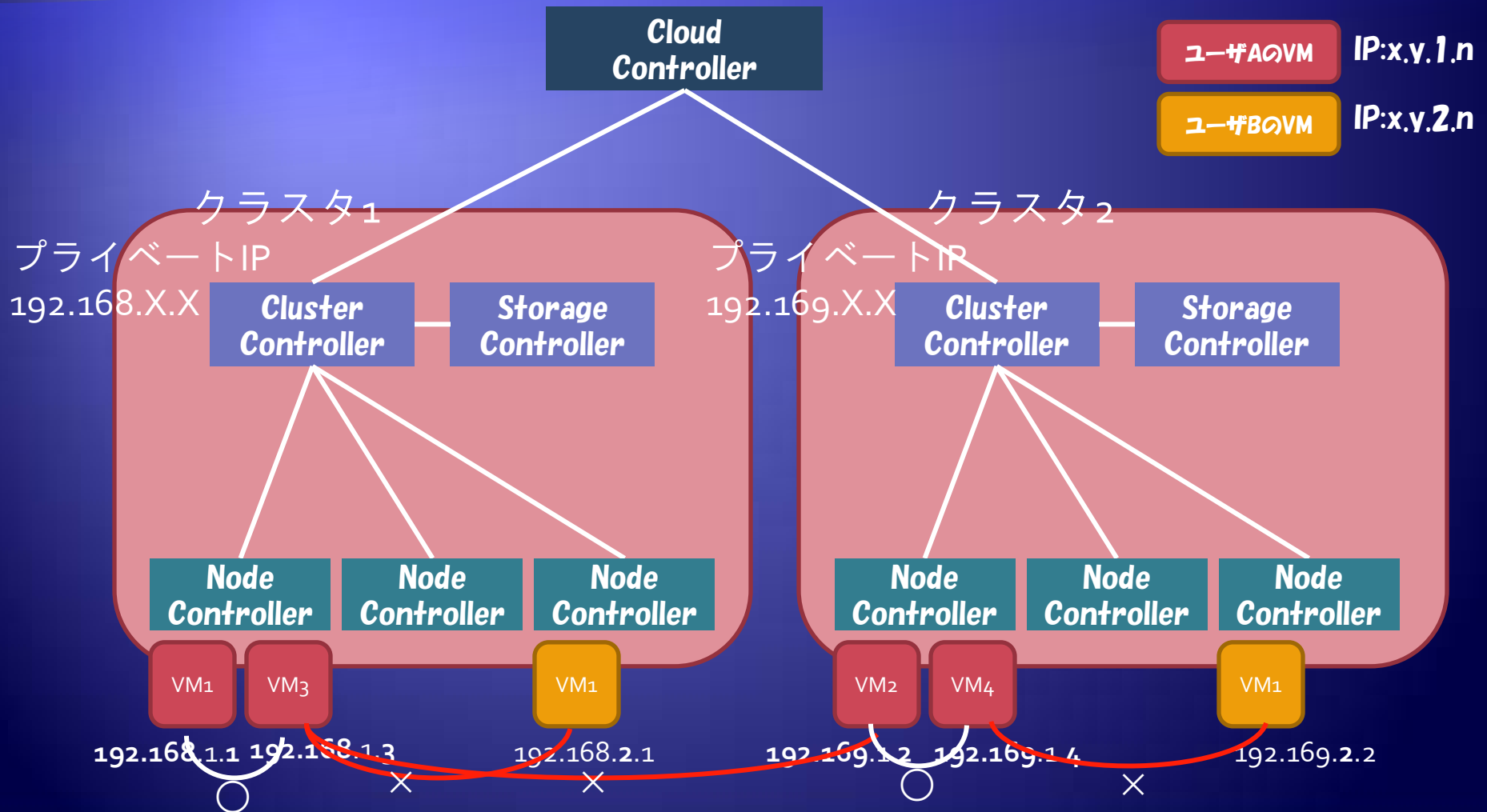
- ・最大ユーザ数510
- ・1ユーザ辺り126台まで

(設定ファイルで割り当て範囲の変更は可能だが、現バージョンでは反映されない)

マルチクラスタ対応してもユーザ数・ユーザ当たりの仮想マシンの最大数は増えない

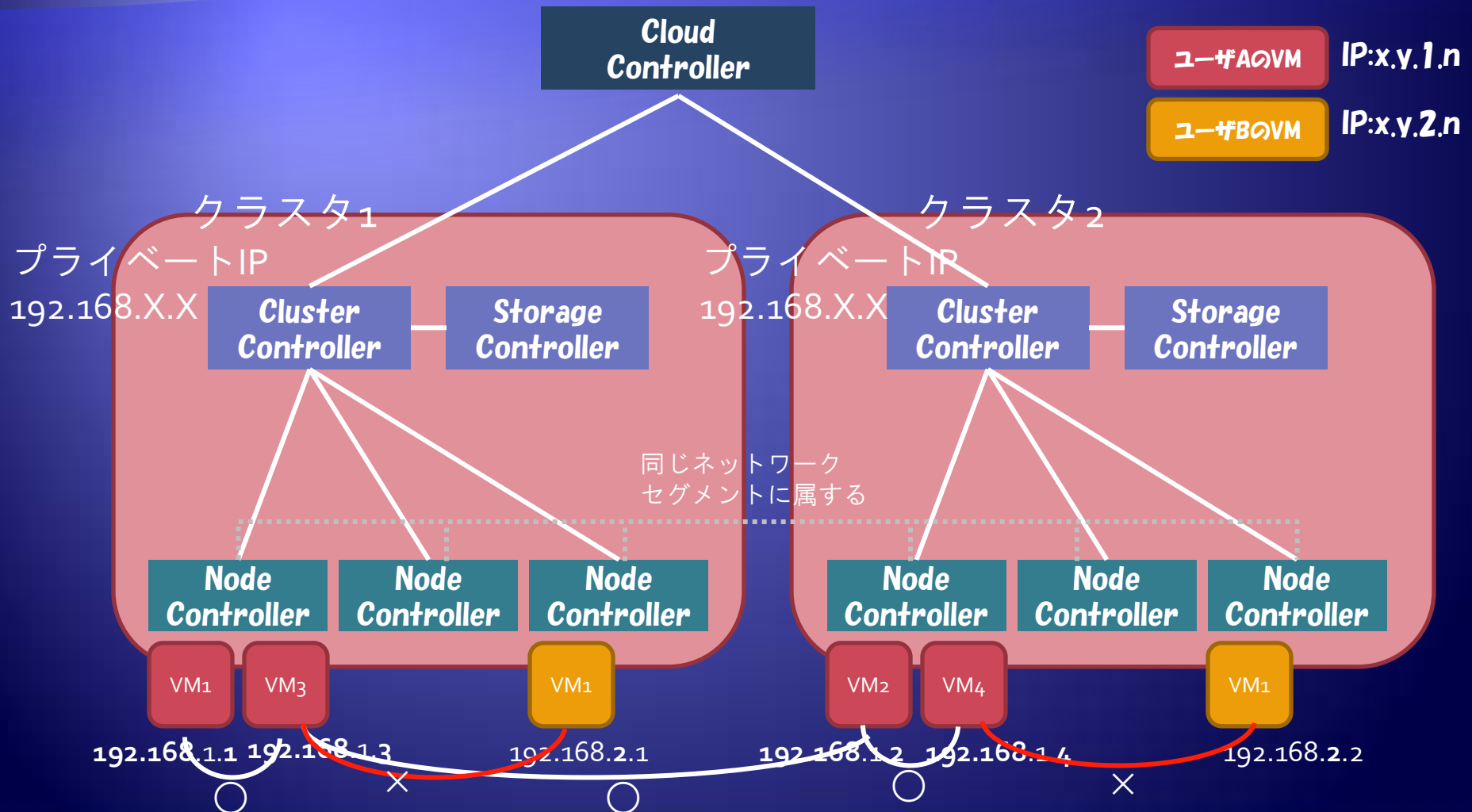
クラスタ間のVMの通信(1)

- クラスタのプライベートIPを異なるセグメントにするとクラスタ間のVMの通信はNG (異なるグループ間のVMの通信もちろん不可)



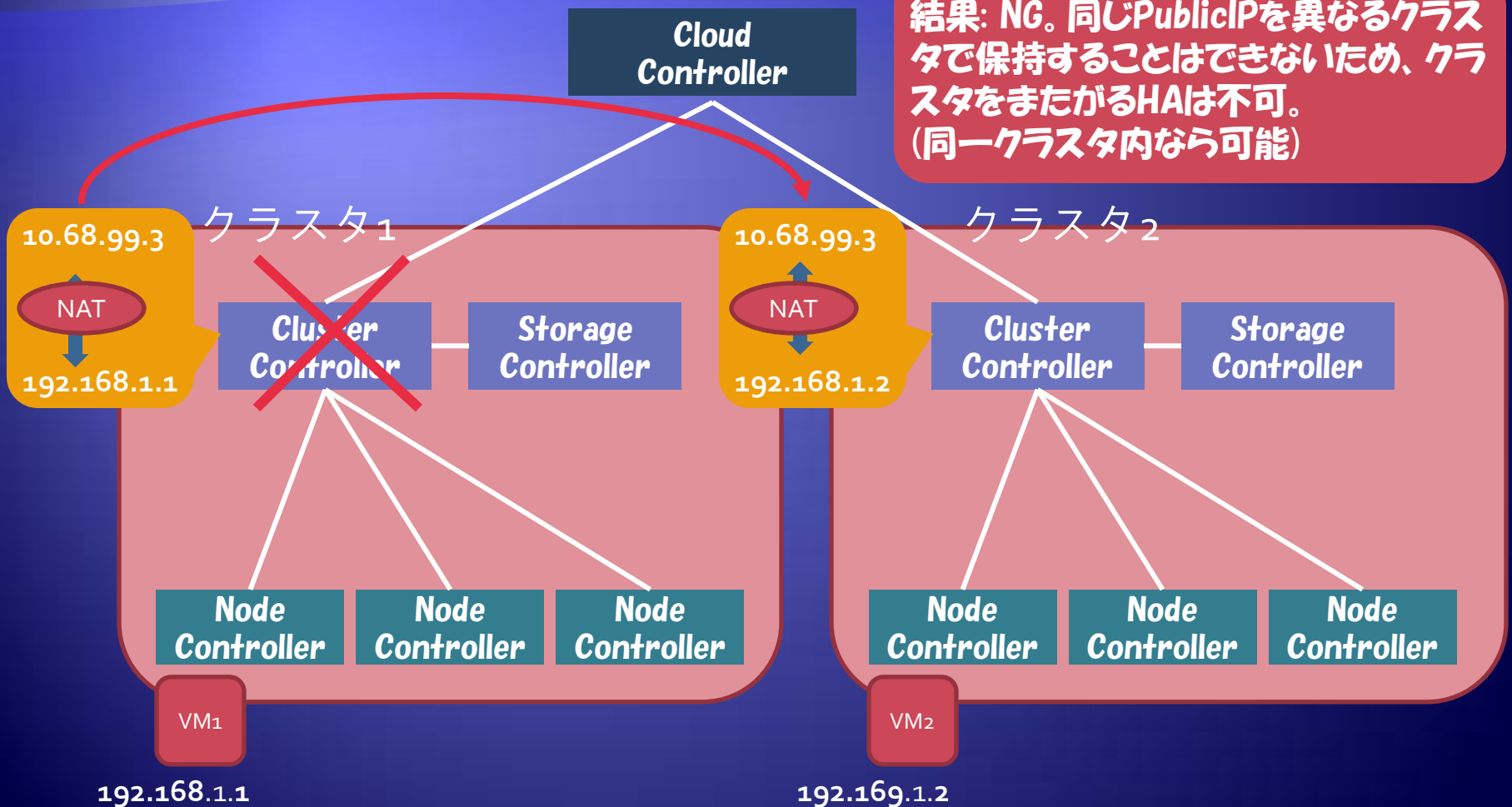
クラスタ間のVMの通信(2)

- クラスタ1とクラスタ2で同じPrivateIPを付与するように設定し、同じネットワークセグメント上にノードコントローラを配置するとき、クラスタを跨ってアクセス可能。



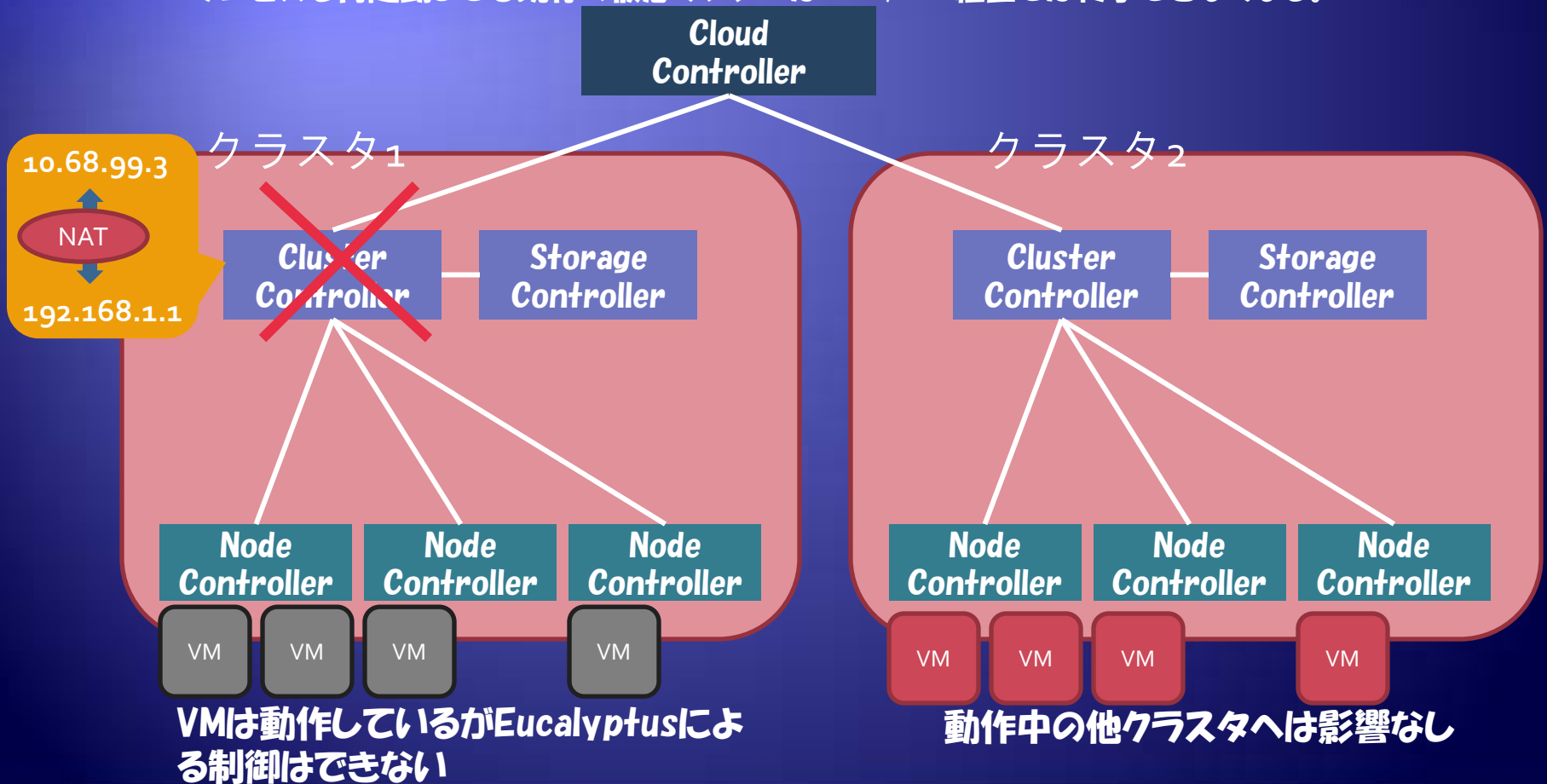
クラスタをまたがるHAは可能か？

- クラスタ障害発生時に他のクラスタのVMへフェールオーバーできるかを想定し、他のクラスタで動作する待機系のVMにIPを付け変えられるか検証。



クラスタ障害発生時の挙動

- Cluster Controllerが動作するハードが停止すると、クラスタへのアクセスは全て不可能になる。ユーザからはCluster Controller経由でアクセスするため。
- Cluster Controllerのプロセスが停止した場合、NAT / FW、DHCPDは動作しているので、仮想マシンはそのまま仮想マシンへのアクセスは可能。
 - プロセスを再起動しても既存の仮想マシンへはAWS / API経由では終了できなくなる。



まとめ

- マルチクラスタ機能により、ネットワーク帯域やストレージが不足したときに、クラウドを拡張することが可能。
 - ネットワーク帯域やストレージが不足していない場合は、NCの増設で十分
- 異なるクラスタのVMのネットワークセグメントは別なので、通信はPublicIP経由となる
- 障害発生時に、特定のクラスタを切り離すことはできるので、クラスタ単位のメンテナンスができる。
 - ただし、Cluster Controllerのプロセスに障害が発生した場合、Clusterの状態の整合性が取れなくなってしまうため、再起動が必要となる。