
第4回Linux-HA 勉強会資料

オープンソースHAクラスタソフトウェア Pacemakerの紹介



2011年9月16日
Linux-HA Japan プロジェクト
<http://linux-ha.sourceforge.jp/>
森 啓介

■ 名前: 森 啓介 (Keisuke MORI)

□ twitter: @ksk_ha

■ Linux-HA Japanプロジェクト関連の活動

□ Pacemaker-1.0系(安定版)のパッチメンテナ

■ 現在Pacemaker-1.0.12リリースに向けて鋭意作業中です!

□ 本家Linux-HA開発コミュニティのボードメンバ(でも名前だけ…)

■ 本業

□ 所属会社: NTTデータ先端技術株式会社

■ 所在地: 月島 — ぜひもんじゃを食べにどうぞ!

□ 普段の業務

■ NTTグループ内におけるPacemaker/Heartbeatの導入支援・サポート

■ バグ報告・パッチ作成などによるNTTからコミュニティへのフィードバック・貢献

■ Pacemakerってどんなもの

- Pacemakerの特長
- 適用事例

Pacemakerって何が良いの？
どんな使い方ができるの？

■ Pacemakerのパッケージ構成

- HeartbeatとPacemakerの関係
- Pacemakerのパッケージ構成

Heartbeatなら使ったこと
あるんだけど…
Corosyncとかcluster-glueとか
訳わかんない!

■ 開発コミュニティの動向

- 開発コミュニティの紹介
- 今後の開発動向

どんなヤツらが作ってるん？
今何やってんの？

- (1)オープンソースであること
- (2)数多くのLinuxディストリビューションで利用できること
- (3)複雑な構成にも対応できること
- (4)応援キャラがいること



特長(1) オープンソースであること

■ ライセンス費用が不要

- 気軽に試してみたり、イニシャルコストを抑えることが可能
 - でも「ものすごく安くなる」とは限らないよね…

■ いざというときは全部自分たちで面倒をみることも可能

- 当事者能力を持てることが重要

■ 気に入らないところは自分で直せる

- バグ報告、パッチ作成などでいつでも開発に参加できる
- 使い方の質問も立派なコミュニティ参加です。ぜひ積極的にご参加を!

■ Red Hat Enterprise Linux 5/6

- RHEL5/6用リポジトリパッケージをLinux-HA Japanページから入手可能(1.0系安定版)
- RHEL6 ではTechnology Preview として標準添付(1.1系開発版)

■ CentOS 5/6, Scientific Linux 5/6

- RHELと同様に利用可能

■ Fedora 14/15/16

■ SUSE Linux Enterprise Linux Server 11

- HAE (High Availability Extension)という製品として利用可能

■ openSUSE 11

■ Gentoo 11

■ Debian squeeze / wheezy

■ Ubuntu 11

特長(3) 複雑な構成にも対応できる

■ 一般的なHAクラスタの構成

□ Active/Standby(1+1)構成

□ Active/Active構成

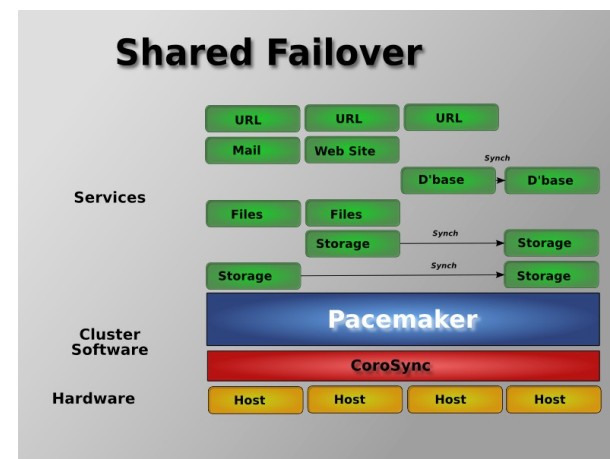
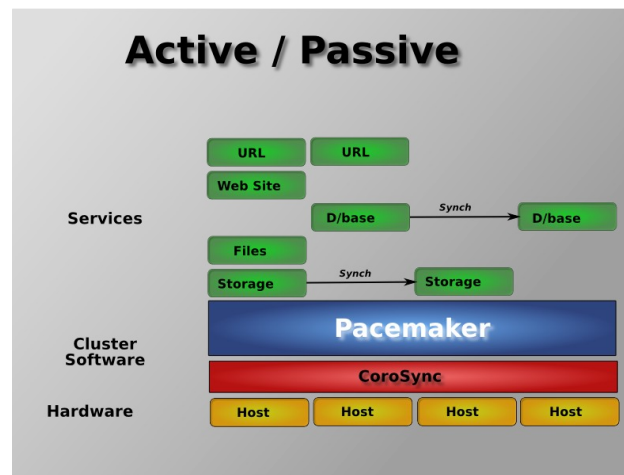
■ 全ノードActive構成

■ 相互待機構成

□ N+1構成

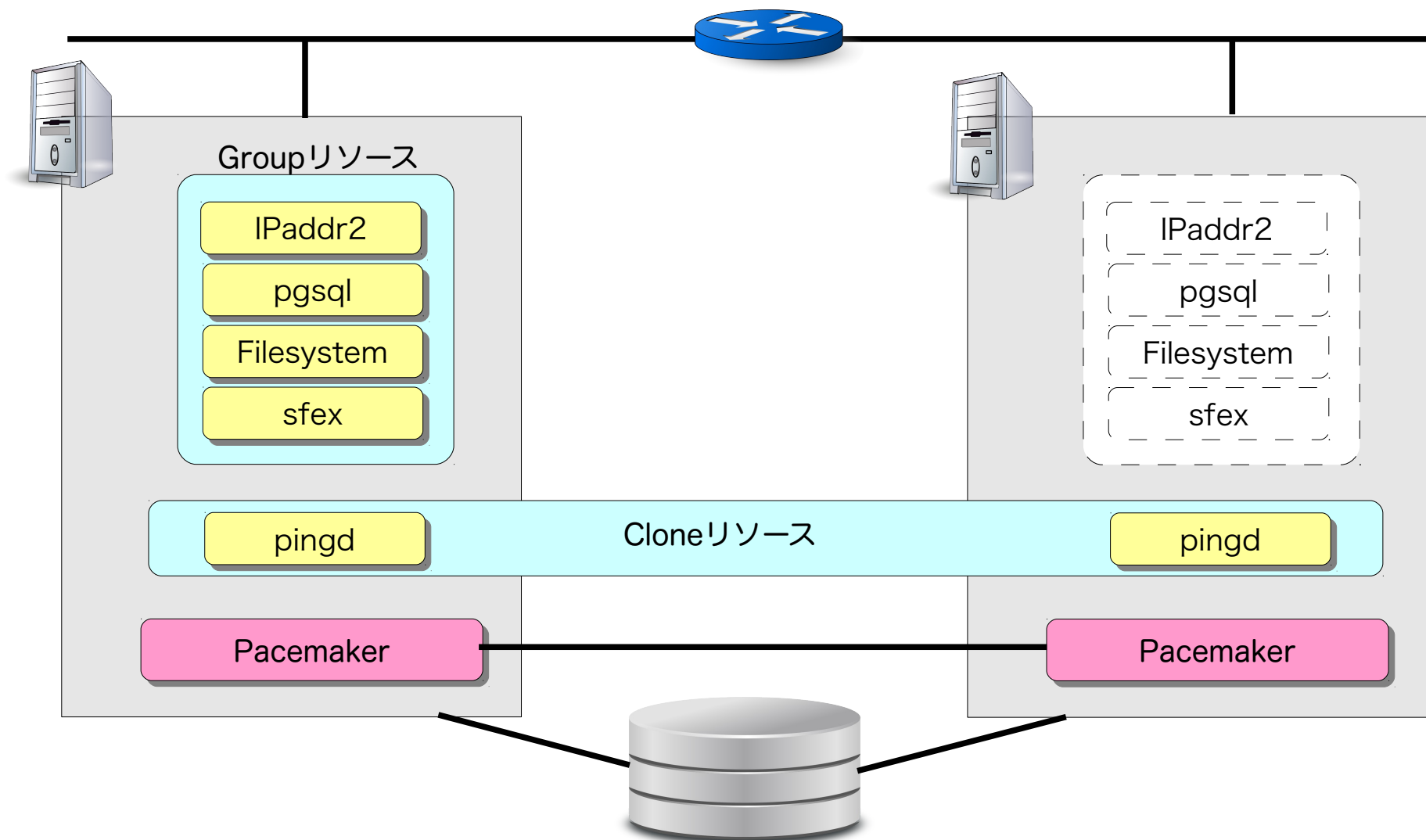
□ N+M構成

□ . . .

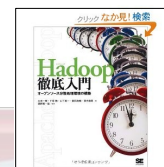
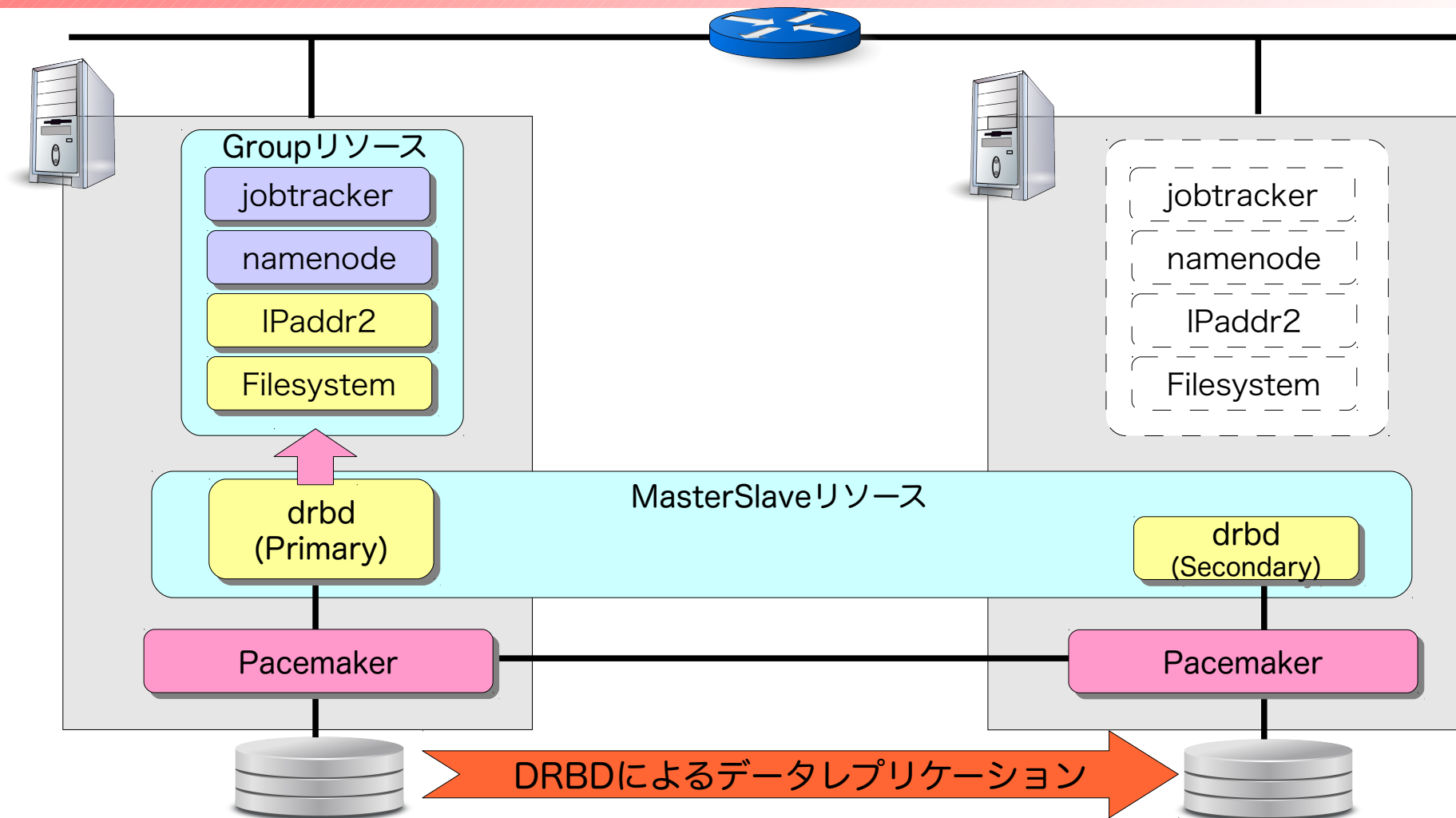


図出典: www.clusterlabs.org (Pacemaker開発サイト)

適用事例(1)データベースサーバの冗長化

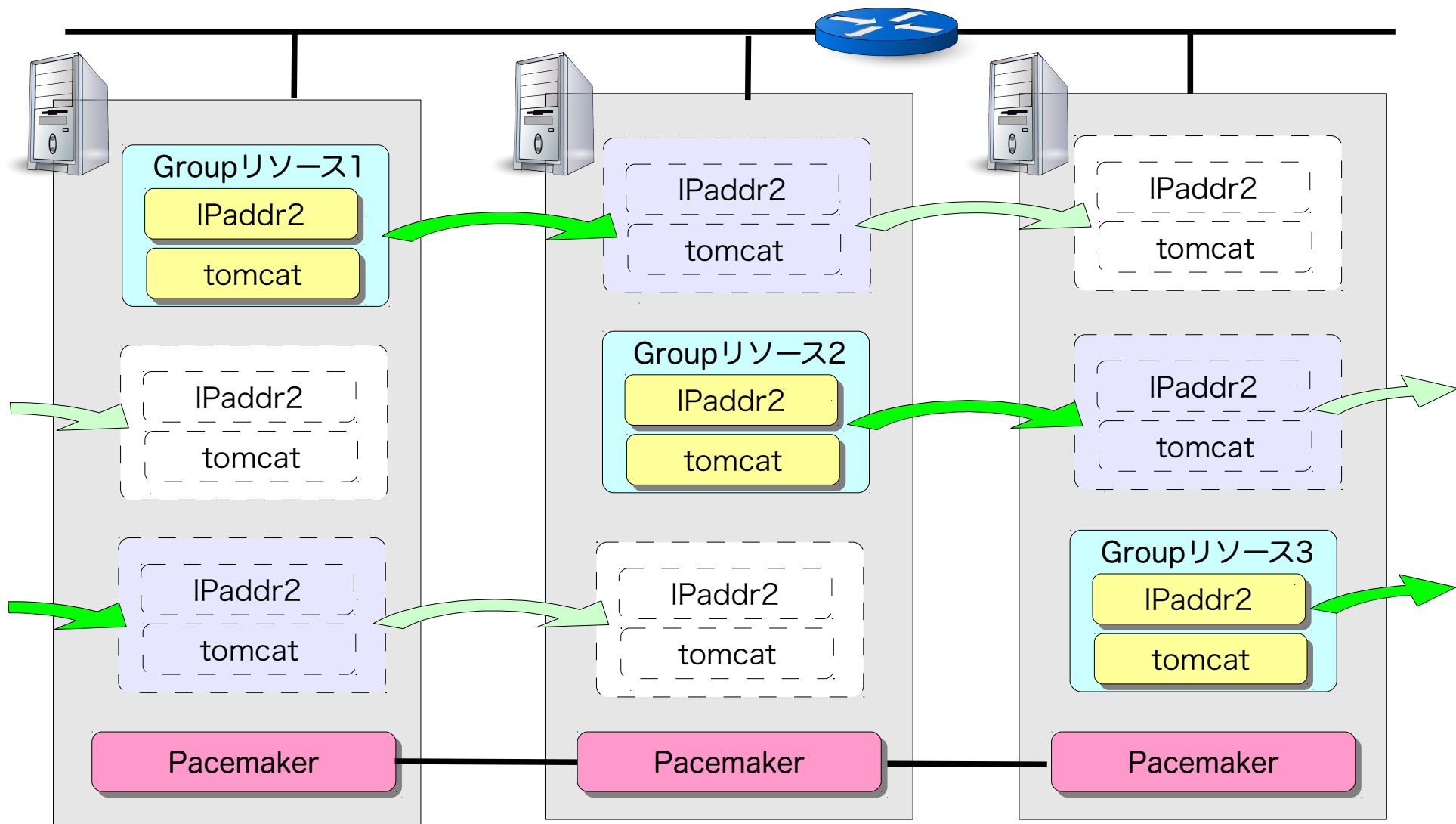


適用事例(2)Hadoopマスターノードの冗長化



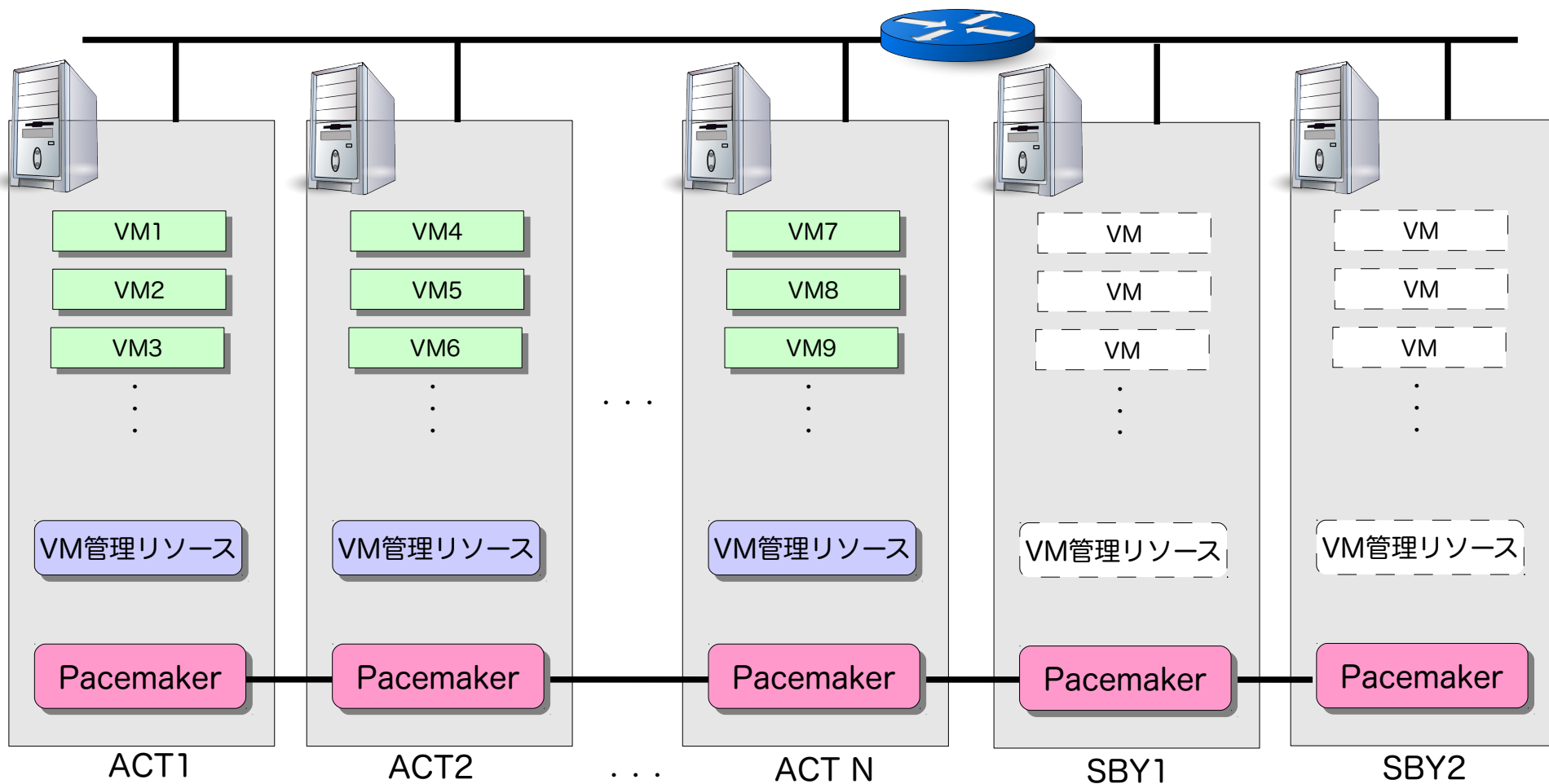
出典: Hadoop徹底入門 翔泳社
※Heartbeat2での事例

適用事例(3) 3ノードActive構成



※Heartbeat2での事例

適用事例(4)クラウド基盤 N+2構成



※Pacemaker+Corosyncでの事例

Pacemakerクラスタ構成の設定要素

- リソース: 監視対象となるアプリケーションの単位
 - primitiveリソース
 - IPaddr2, apache, mysql, pgsql, Filesystem, sfex, VirtualDomain, etc...
 - groupリソース
 - cloneリソース
 - master slave リソース
 - アプリケーションのホットスタンバイ
 - DRBD、MySQLレプリケーション、PostgreSQL 9.1 SR対応(coming soon...)
- 制約条件: リソースの配置先、起動順序を制御
 - location (配置ノード制御)
 - colocation (同居制御)
 - order (起動・終了順序制御)
 - groupリソース=colocation + order を自動的に設定したもの

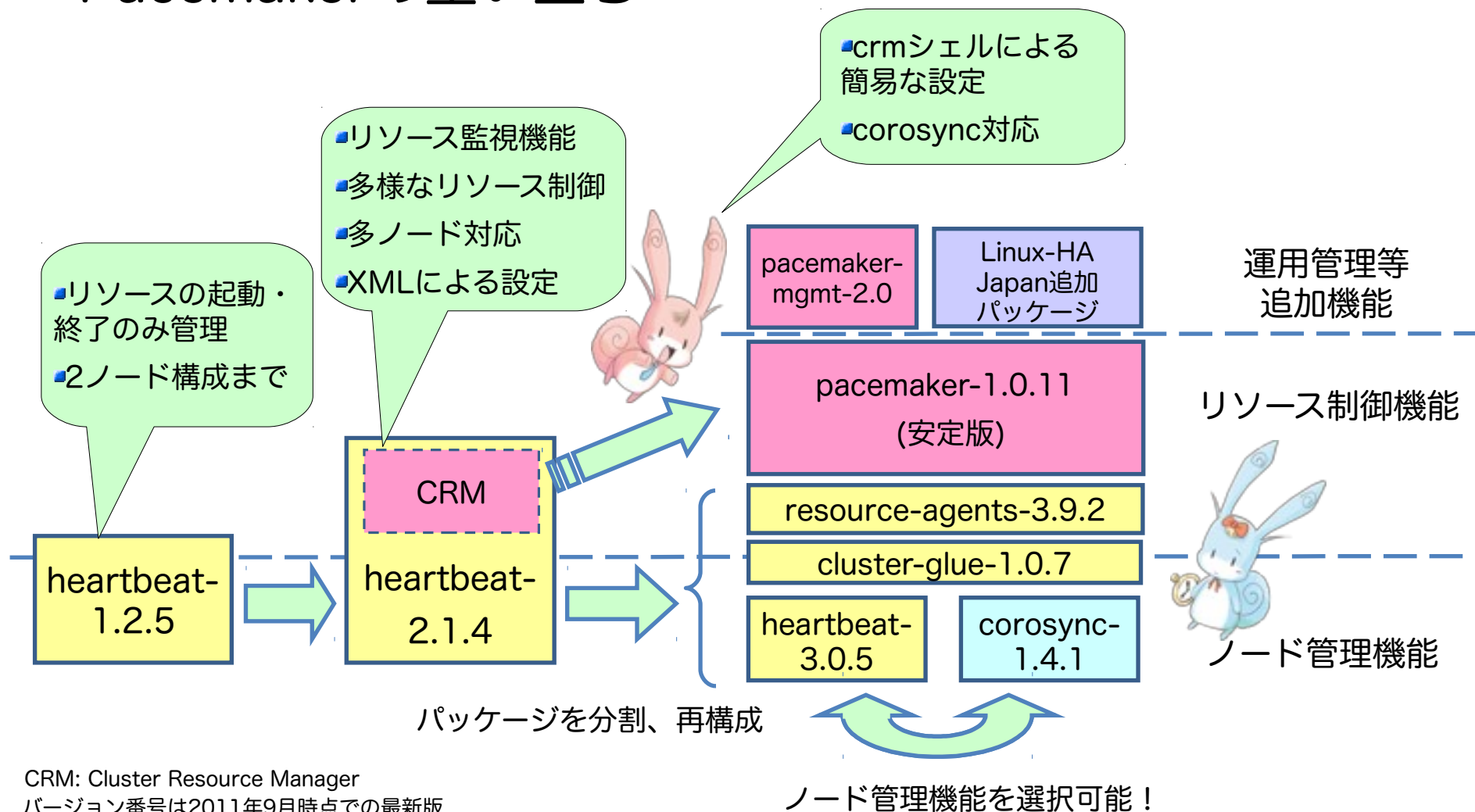
これらの設定を任意に組み合わせることで、
非常に複雑な構成への対応が可能になっています

Heartbeatなら使ったことあるんだけど…
Pacemakerとどういう関係なの？

- Pacemaker は Heartbeat の後継バージョンです。
- Heartbeatバージョン2までは単一パッケージで提供されていましたが、Pacemakerからは複数のパッケージを組み合わせるようになりました。

Pacemakerのパッケージ構成

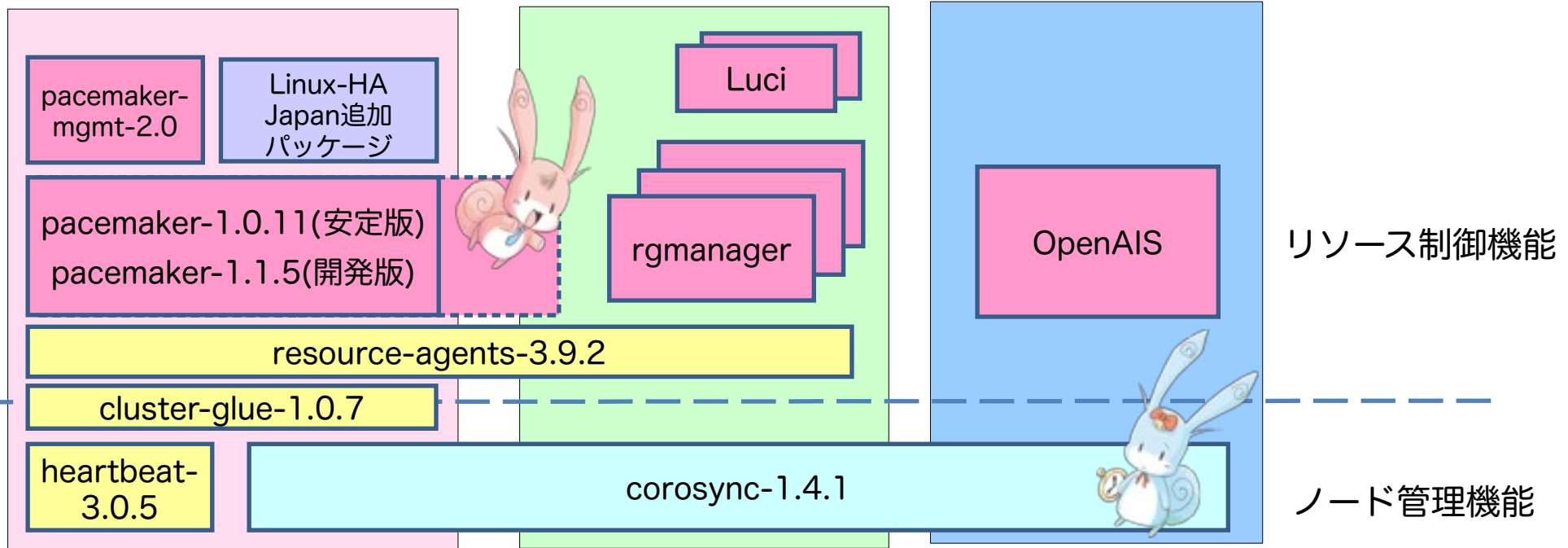
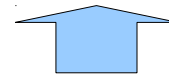
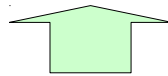
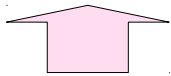
■ Pacemakerの生い立ち



- 必要なパッケージが多すぎて覚えられないんだけど…
 - インストールには“**pacemaker**” だけ覚えてもらえれば充分です。
 - 他のパッケージは Pacemakerの依存パッケージとして自動的にインストールされます。
 - Linux-HA Japan のリポジトリパッケージには必要なパッケージが全て入っています。
- HeartbeatとCorosyncを選べるって、どっちを使えばいいの？
 - (1) 安定版: Pacemaker-1.0.* + Heartbeat-3.0.* ← **今はこちらがオススメ!**
 - より安定動作・実績多し。動作確認済みのバージョン組み合わせをLinux-HA Japanから提供
 - (2) 開発版: Pacemaker-1.1.* + Corosync-1.* ← 将来有望
 - 多ノードの場合はこちらが○。でも機能変更・リリースが頻繁。
- Heartbeatバージョン1と同じ使い方はできないの？
 - もう開発コミュニティでも誰も動作確認してないので、正直お勧めできません。
- そもそも何でコンポーネントをバラバラにしちゃったのさ？
 - それは…

クラスタソフトウェアを超えたコンポーネントの共通化

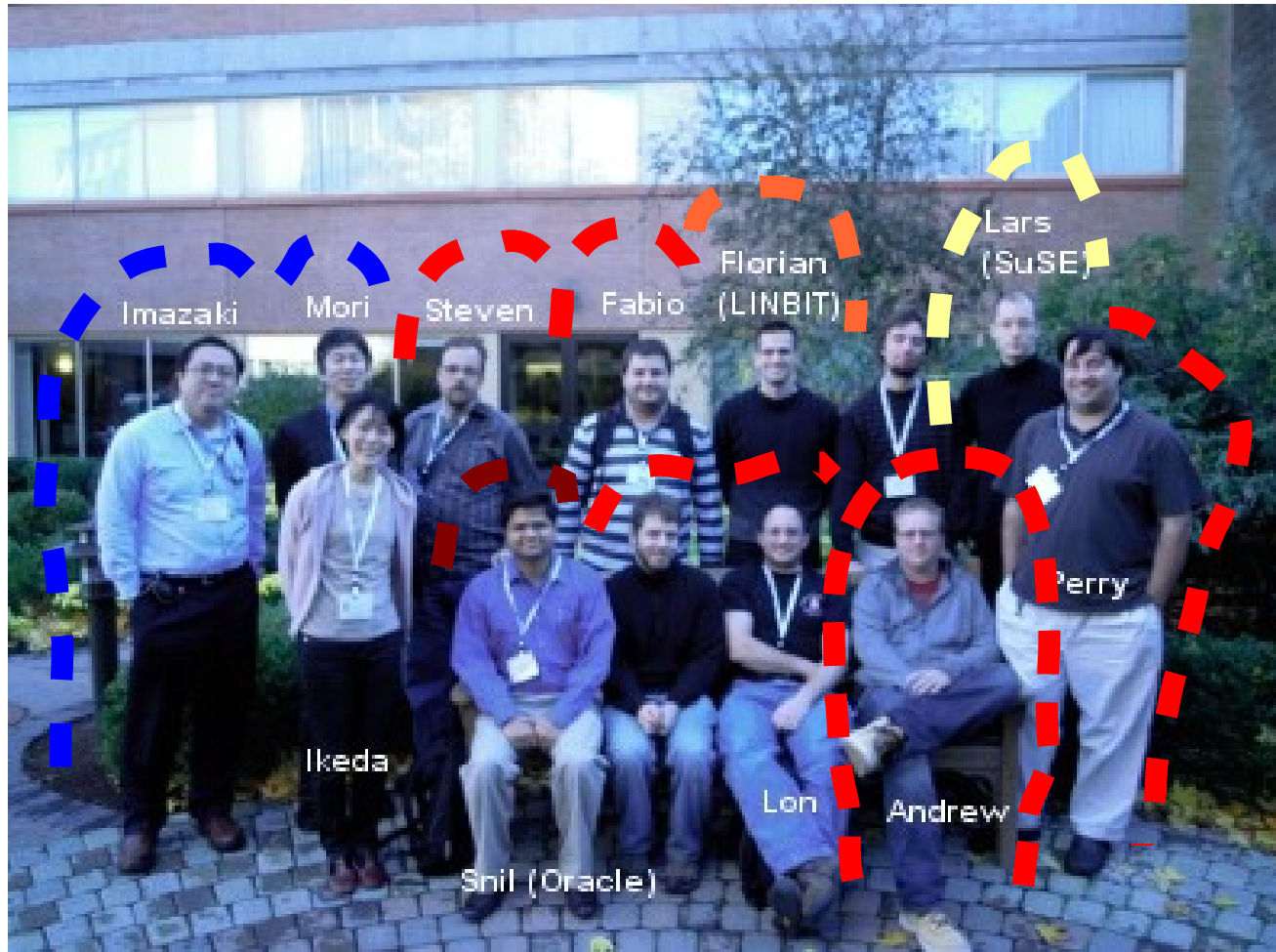
Pacemaker Red Hat OpenAIS
High Availability Add-on
(はあん)



オープンソースクラスタ開発コミュニティの動向

- クラスタの開発コミュニティは今、ゆるやかに統合しつつあります
 - Red Hat や SUSE 他のクラスタ開発者同士がコミュニティを通して連携し、コンポーネントの共通化などが進められています。
- 通常は各コンポーネントごとにML、IRC等で活動
 - Pacemaker : <http://www.clusterlabs.org/>
 - Linux-HA(Heartbeat他) : <http://www.linux-ha.org/>
 - Corosync : <http://www.corosync.org/>
 - linux-cluster(Red Hat) : <https://fedorahosted.org/cluster/wiki/HomePage>
- Linux Foundation HA-WG の設立・全体の活動を統括
 - 正式アナウンス: 2011年3月3日
 - 年に一度程度 F2F ミーティングで集中議論
 - 2008年プラハ、2010年ボストン、2011年・・・スケジュールが合わずあえなく中止
 - 主な参加組織
 - Red Hat
 - SUSE
 - LINBIT
 - NTT (Linux-HA Japan)
 - その他: IBM, Oracle

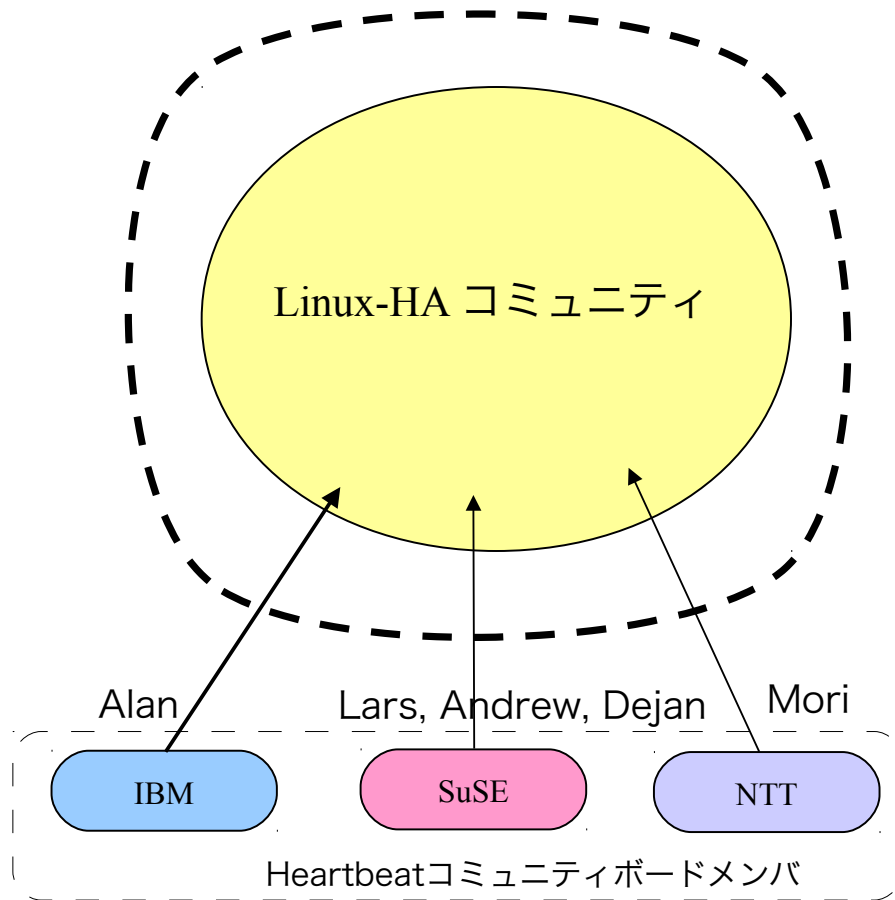
LinuxCon / Cluster Summit 2010 Boston



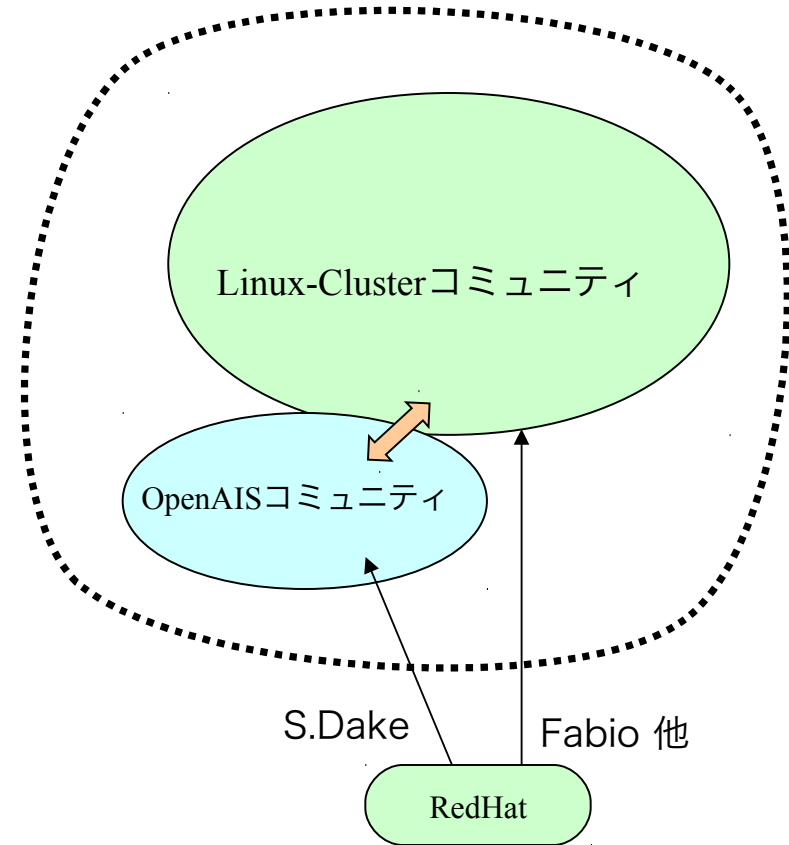
月刊あんどりゅー
くんもよろしく!
Linux-HA Japanサ
イトで絶賛連載中

オープンソースクラスタコミュニティ情勢の変化(2008年前期頃まで)

Heartbeat 陣営



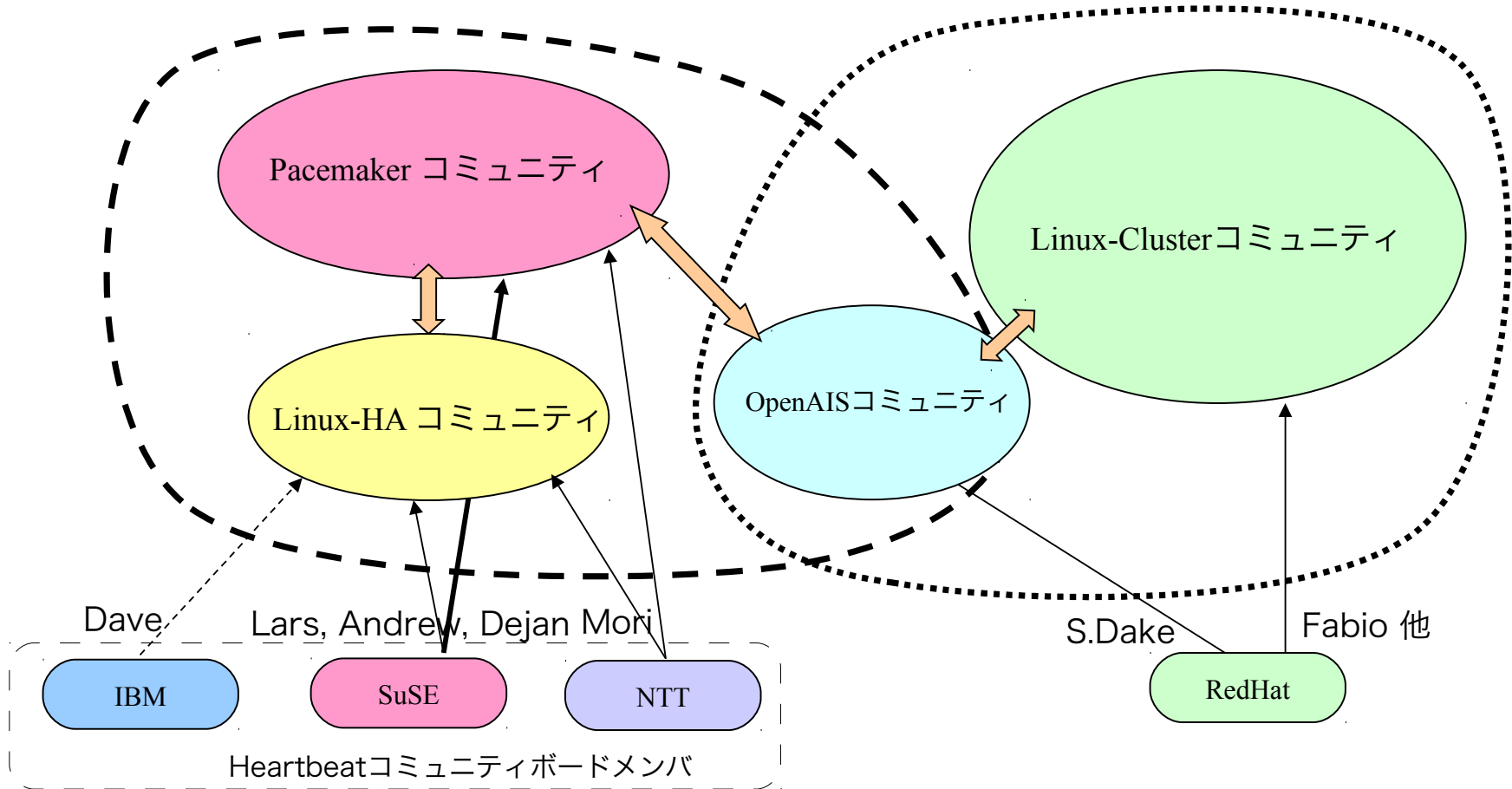
RedHat Cluster Suite 陣営



オープンソースクラスタコミュニティ情勢の変化(2009年前期頃まで)

Pacemaker 陣営

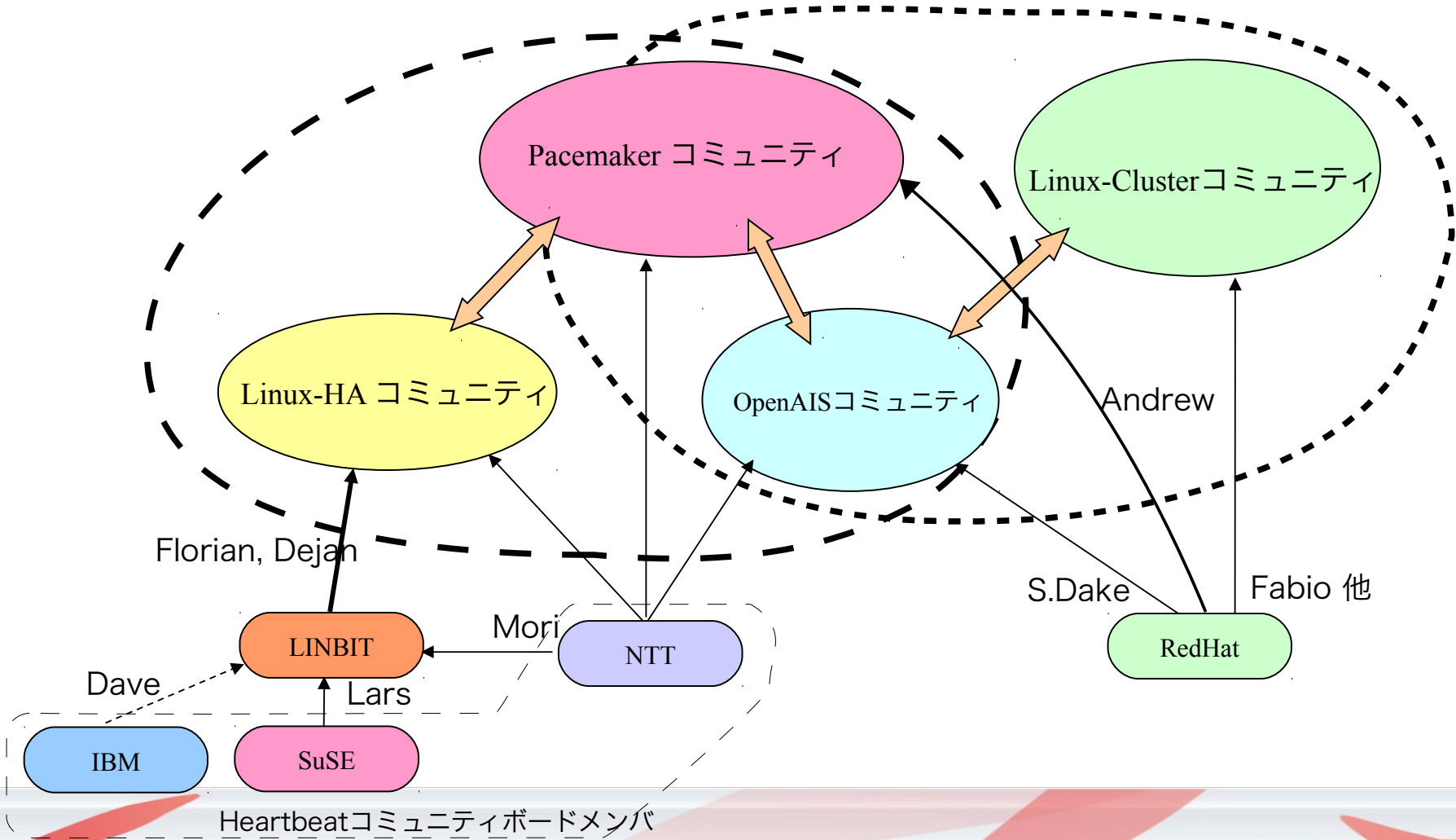
RedHat Cluster Suite 陣営



オープンソースクラスタコミュニティ情勢の変化(2010年以降)

Pacemaker 陣営

RedHat Cluster Suite 陣営



■ Linux-HA Japanでの開発動向

□ PostgreSQL 9.1 SR 対応リソースエージェント開発

- PostgreSQL 9.1正式リリース！(2011/09/12)
- 本家開発コミュニティに提案予定。皆さんからのフィードバックをお待ちしています！

□ KVM対応仮想環境連携機能 本日リリース！(2011/09/16)

- pm_kvm_tools: KVMゲスト内のPacemakerとハイパーバイザ上のPacemakerを連携
- vm-ctl: 仮想マシンリソース管理用のツール

■ 本家開発コミュニティの開発動向

□ ディザスタリカバリ対応：遠隔地を結ぶクラスタリング機能(split-site)

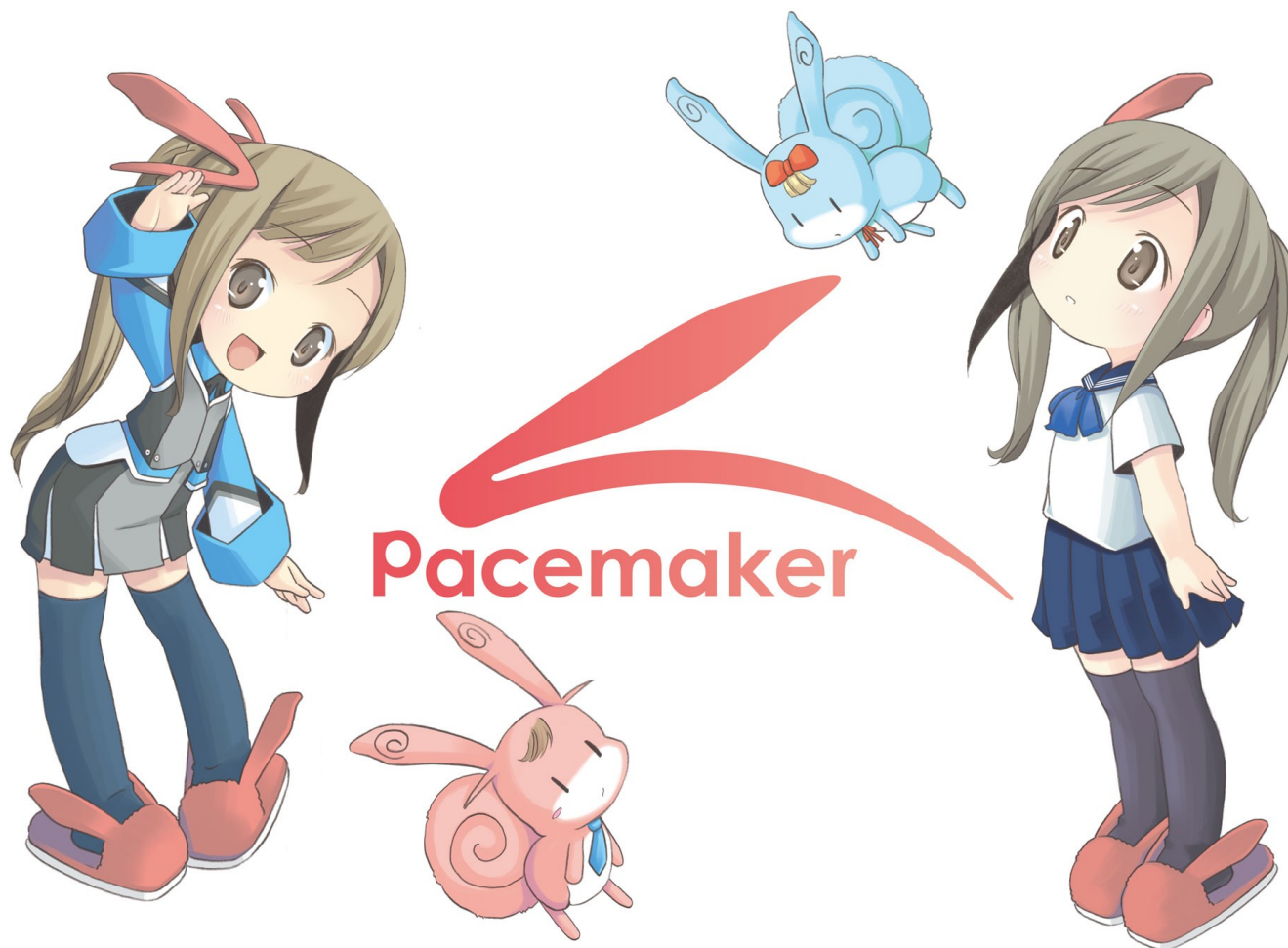
□ クラウド対応：ハイパーバイザ上のPacemakerからのVM内監視機能

- Pacemaker-cloud プロジェクト、Matahariプロジェクト

□ Pacemaker と Red Hat HA add-on の融合

- リソースエージェント(RA)共通化:開発リポジトリの統合は済。今後は共通RAの開発
- rgmanagerからPacemakerへの置き換え

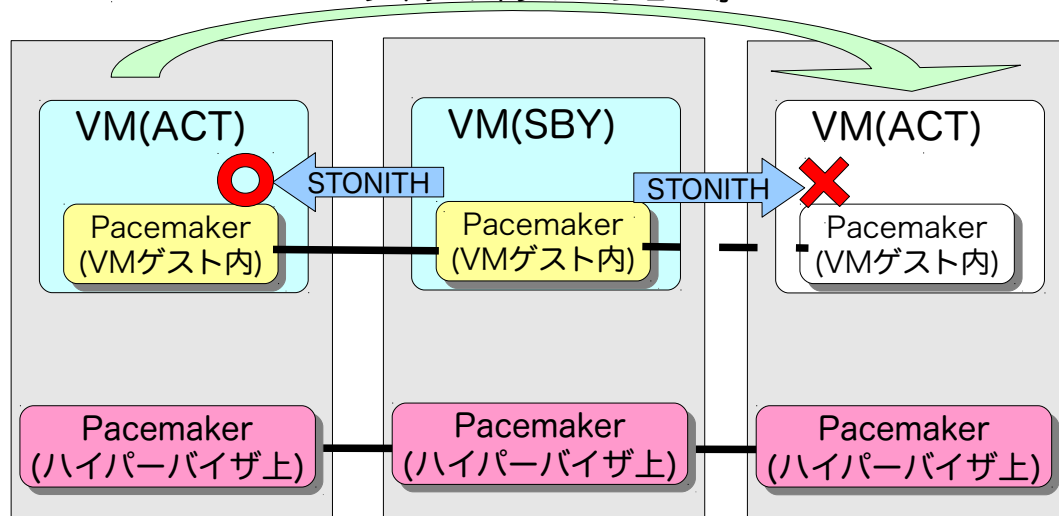
- 以上です。ありがとうございました！



ライブマイグレーション等

■ 仮想環境連携STONITH機能

- 仮想マシン(VM)の移動先に関わらず、PacemakerからのSTONITH実行を可能とする機能
- 従来のSTONITHプラグインではVMが稼働するホスト(ハイパーバイザ)は固定されている必要があった(右図上)。
- pm_kvm_toolsを使うと、ハイパーバイザ上のPacemakerと連携することでSTONITHの実行が可能となる。



ライブマイグレーション等

