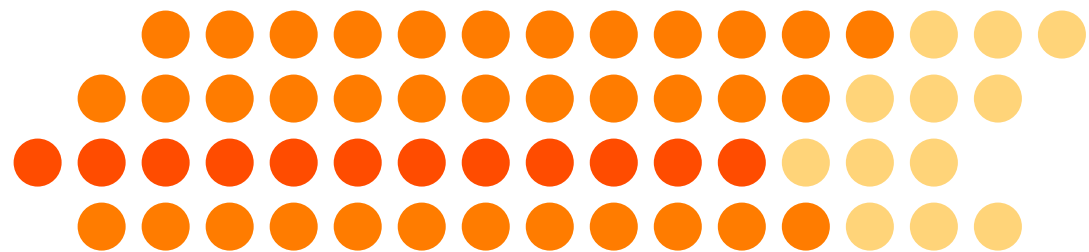


クラウドコンピューティングの 「今」と「これから」

東京大学 情報基盤センター
WIDE Project

関谷 勇司 <sekiya@nc.u-tokyo.ac.jp>



自己紹介 (1)



- 本職
 - 東京大学 情報基盤センター 准教授
 - ネットワークに関する研究を行う
 - 学内ネットワークの設計と運用
- 主な研究テーマ
 - 次世代ネットワークプロトコルの設計と開発
 - 自律分散システムの耐障害性向上に関する研究
 - クラウドコンピューティングに必要とされる要素技術の研究



自己紹介 (2)



- 主な活動

- WIDE Project ボードメンバー

- <http://www.wide.ad.jp/>

- 産学連携の研究プロジェクト

- (Founder : 村井純, 代表 : 江崎浩)



- Interop Tokyo ShowNet NOC メンバー

- <http://www.interop.jp/>

- 2000年から NOC メンバーとして活動

- 2010年から NOCジェネラリスト



本日のイントロ



- クラウドコンピューティングって何でしょう
 - 便利に使っていますよね
 - コストダウンの魔法？
 - 便利な言い訳？
 - 責任のたらい回し？
- クラウドは「クラウド」として運用され、「クラウド」として使われているのでしょうか
 - 研究側の立場から見た意見です
 - でも実運用もやっています
 - 「何のため」のクラウド化なのか



本日の話題



- クラウドの工学的アーキテクチャ
- クラウドを支える技術
- 学術・研究におけるクラウド要素技術動向
- 適切にクラウドを利用する

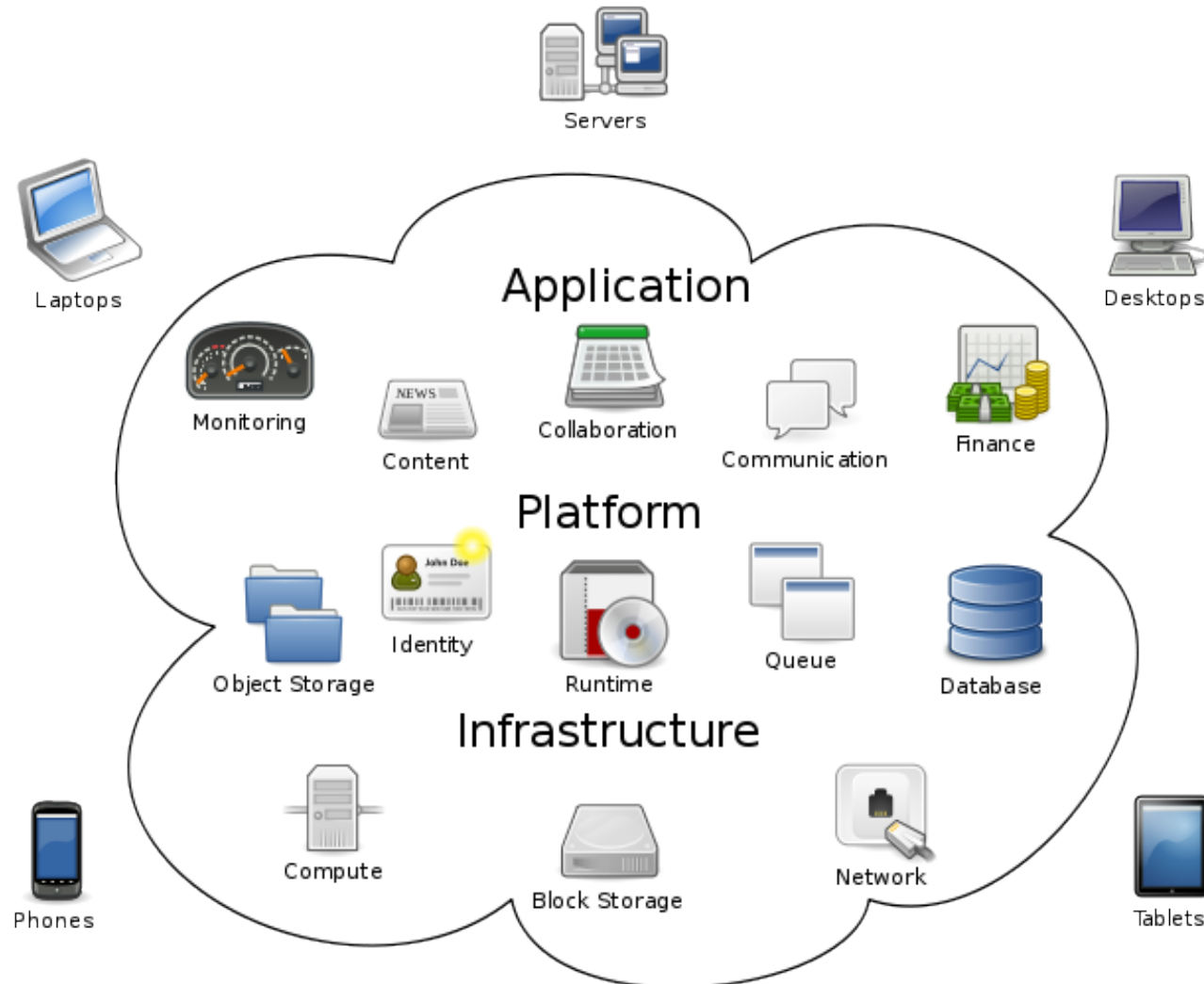




工学的アーキテクチャから見た クラウド



クラウドって何なんだろう？



Cloud Computing

Yuji Sekiya (The Univ. of Tokyo / WIDE Project)

出展 : http://en.wikipedia.org/wiki/Cloud_computing



クラウドの定義とは？



- Above the Clouds: A Berkeley View of Cloud Computing
 - Electrical Engineering and Computer Sciences University of California at Berkeley, Technical Report No. UCB/EECS-2009-28
- 規模性
 - あたかも無限であるかのようにリソースを利用することができる
- 即時性
 - 必要なとき必要なだけ時間単位で利用することができる
- 可用性
 - 単一障害に影響されずサービスを継続することができる



クラウドの特色



- Availability of Service
 - Use Multiple Cloud Providers; Use Elasticity to Prevent DDOS
- Data Lock-In
 - Standardize APIs; Compatible SW to enable Surge Computing
- Data Confidentiality and Auditability
 - Deploy Encryption, VLANs, Firewalls; Geographical Data Storage
- Data Transfer Bottlenecks
 - FedExing Disks; Data Backup/Archival; Higher BW Switches
- Performance Unpredictability
 - Improved VM Support; Flash Memory; Gang Schedule VMs



クラウドの特色 (cont.)



- Scalable Storage
 - Invent Scalable Store
- Bugs in Large Distributed Systems
 - Invent Debugger that relies on Distributed VMs
- Scaling Quickly
 - Invent Auto-Scaler that relies on ML; Snapshots for Conservation
- Reputation Fate Sharing
 - Offer reputation-guarding services like those for email
- Software Licensing
 - Pay-for-use licenses; Bulk use sales



ソフトウェア工学的視点から見たクラウド



SLA(要求)
Service Level Agreement



要求仕様
Requirement Specification



技術仕様
Technical Specifications

種別	項目例	設定例
可用性	サービス時間 計画停止予定通知	24時間365日 30日前
信頼性	平均復旧時間(RTO) 障害発生件数	1時間以内(基幹) 1回以内(基幹)
性能	応答時間	3秒以内(内部)
拡張性	外部接続性 提供リソースの上限	APIを公開 1TB, 40,000view/page
サポート・管理	backupデータ保存 データ漏洩の補償 ...	5年以上 有(保険加入) ...

項目	現状	短期(2,3年)	中長期
プロビジョニング技術	半自動	自動	インテリジェント
プロセスマイグレーション	VM指定	LAN内自動	WAN間自動
データ匿名化技術	標準なし	仮名・曖昧化	標準確立
...	...		

クラウドを取り巻く技術要件



データセンタに求められるネットワーク技術

プロビジョニング技術(自動化・統合管理)
データ高速読み込み
(ストレージ階層化・CDN等)
プロセスマイグレーション
(LAN内・WAN間の自動マイグレーション)

運用・保守技術の高度化に関する技術

データセンタ運用自動化
(監視ツール・仮想ハイパーバイザとの連携・
遠隔管理技術・障害自己修復技術)

Cloud Computing

高信頼化技術に関する技術

マルチテナントにおける脆弱性対策
(VM機能分割・ミドルウェア層のマルチ
制御・仮想化の統合管理)
データ匿名化技術
(構文解釈での切り落とし・暗号化・
データ分割保管)
高信頼性・データ保護技術
(モニタリング・分散ファイルシステム・
秘匿分散計算・物理的データ保護)

データ間相互連携・分析に関する技術

大量データ処理・分析技術
(イベント収集・フィルタリング技術・
分散データ解析・データ可視化)
APIも含むクラウド技術の標準化

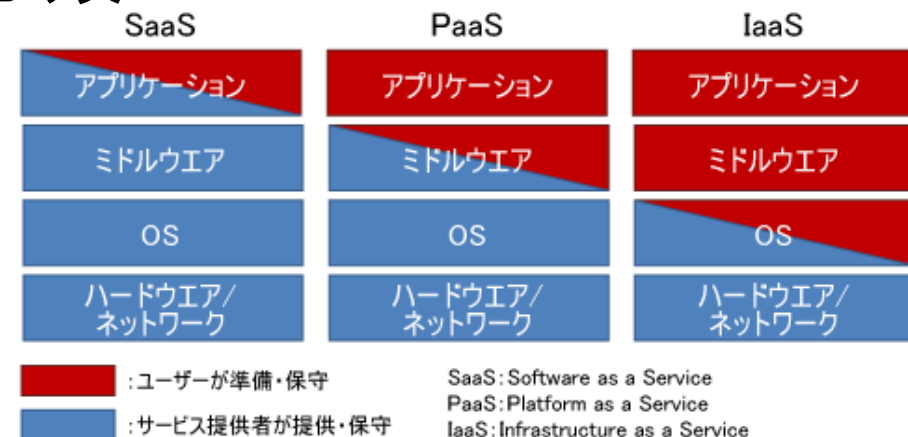
環境負荷低減に関する技術

センサーによる電力管理の自動化
サーバ冷却技術
代替電力の活用(廃熱利用等)
リソース最適化



クラウドの分類

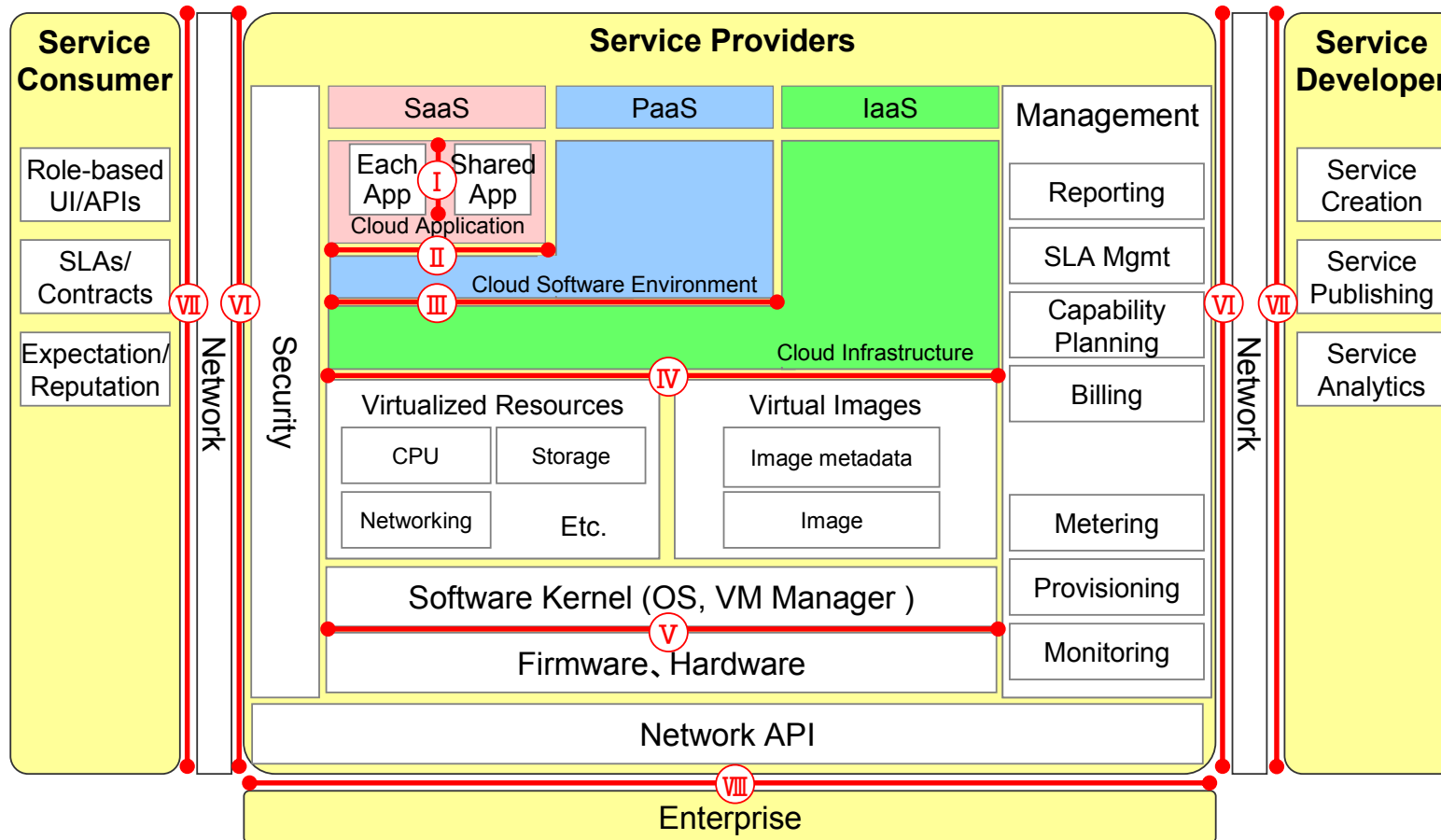
- サービス形態による分類
 - SaaS (Software as a Service)
 - PaaS (Platform as a Service)
 - IaaS (Infrastructure as a Service)
 - HaaS (Hardware as a Service)
- テナント方式による分類
 - Private Cloud
 - Off / On Premise
 - Hybrid Cloud
 - Public Cloud



出展 : <http://itpro.nikkeibp.co.jp/article/Keyword/20110216/357282/>

クラウドのレイヤリングモデル

Cloud Function Reference Model* におけるレイヤー



凡例

- I. 共通アプリと個別アプリ間の標準化
- II. アプリケーションとプラットフォーム間の標準化
- III. プラットフォームとクラウドインフラ間の標準化
- IV. クラウドインフラと仮想化環境間の標準化
- V. 仮想化環境とハードウェア間の標準化
- VI. ネットワークとプラットフォーム・アプリケーション間の標準化
- VII. ネットの外側と内側間の標準化
- VIII. 企業アプリとクラウドサービス間の標準化

*Open Cloud Manifesto"Cloud Computing Use Cases White Paper"を参考に作成




クラウドを支える技術



クラウドの SLA (Service Level Agreement)



- SLA は存在するのか？
 - 稼働率
 - リソースの保証 (CPU, Network)
 - 信頼性
 - サーバ障害、ストレージ障害、ネットワーク障害
- 
- 集約化されているゆえの被害拡大
- 当然事故も起こっている
 - むしろ大規模な障害となっている

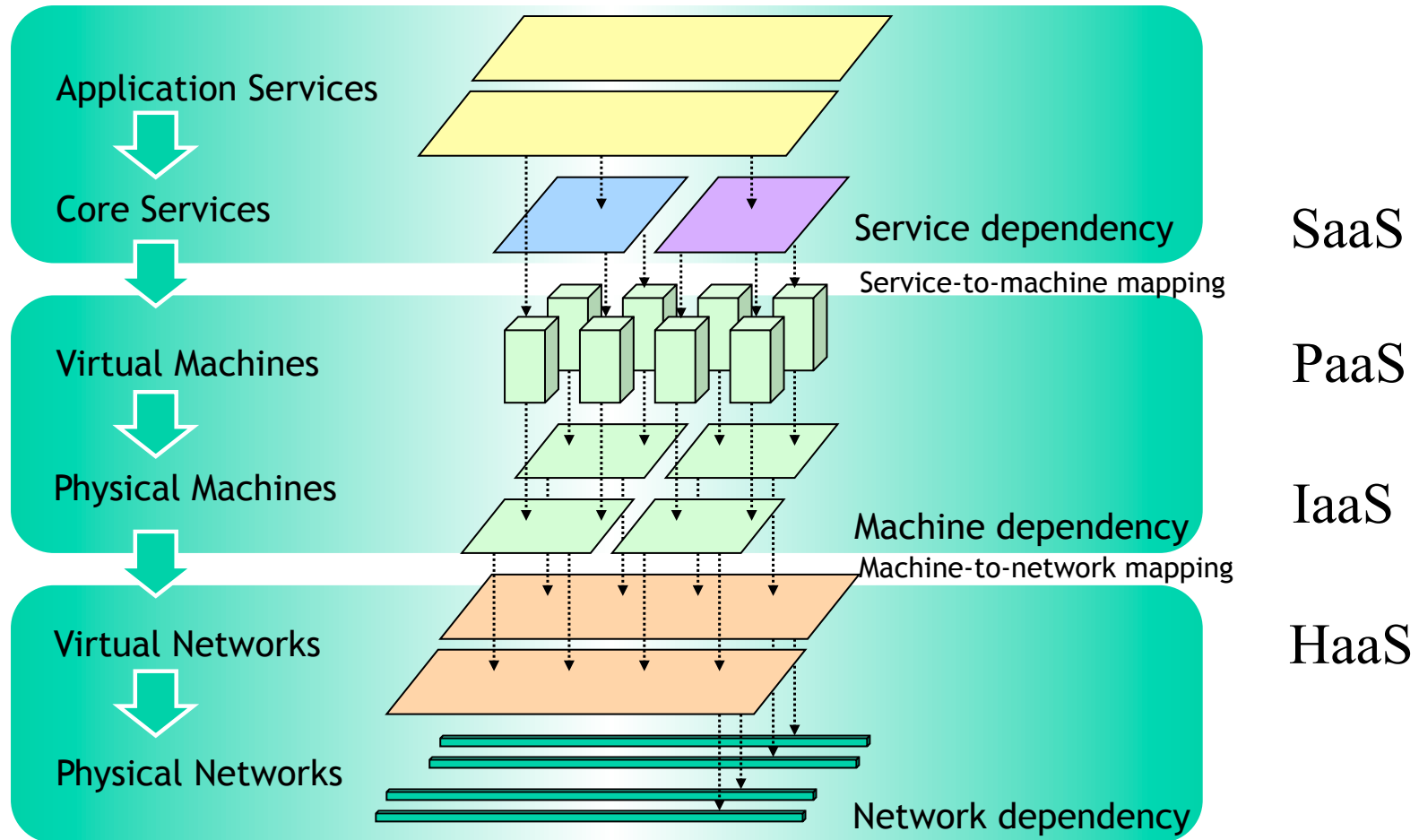
ユーザの VM / データ消滅



対処療法的な最善復帰



階層構造ゆへの依存性



Amazon EC2 / S3 の障害例

2008年2月15日 :

Amazon S3停止
復旧に数時間を要する

2008年7月20日 :

Amazon S3停止
復旧に数時間を要する

2009年12月9日 :

バージニア州 EC2 サイト停電
インスタンス(VM) 復旧に数時間を要する

2009年12月9日 :

ワームに感染したインスタンスを利用して Botnet 制御

2010年4月1日 :

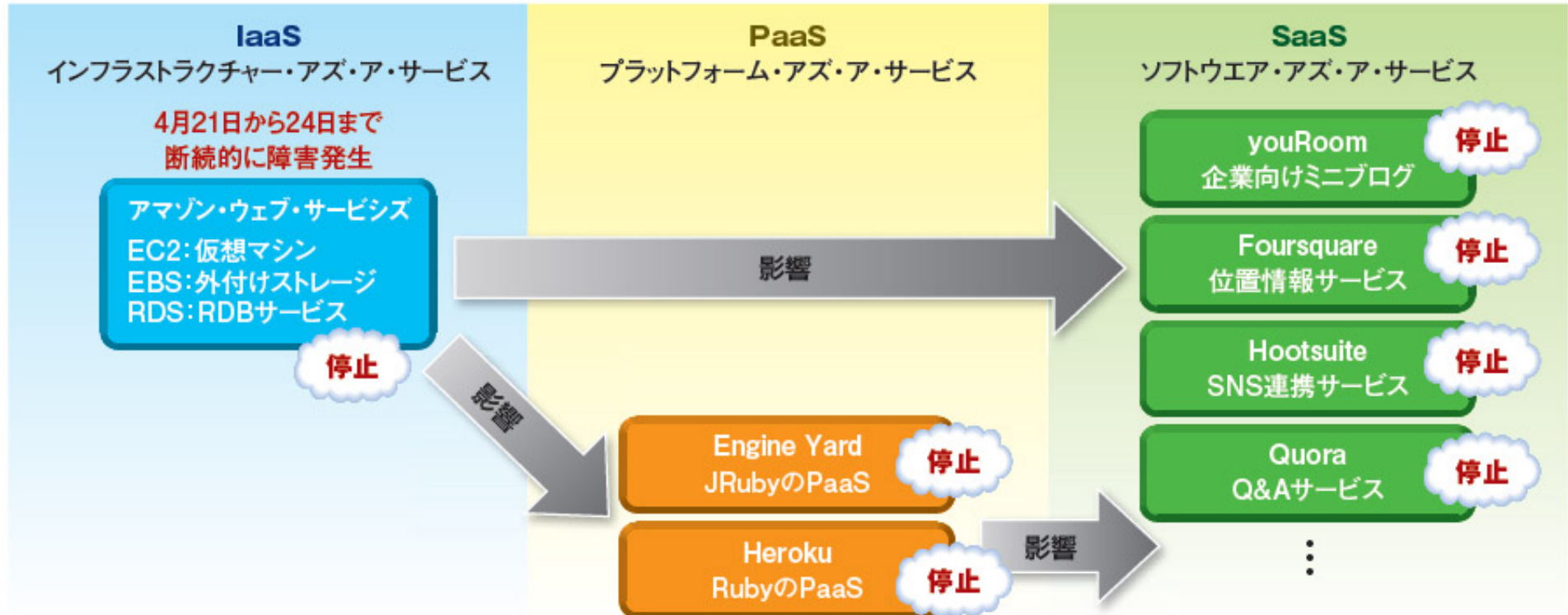
US東海岸 EC2 サイトにてインスタンス制御不能
数時間インスタンスのコントロール/モニタリング不能に

2011年4月21日 :

US東海岸 EC2 サイトにて再ミラーリング頻発による高負荷障害
4日ほど高負荷状態が続きユーザサービスに障害

小規模な障害 (インスタンス消失) は数多く発生

その障害が波及



出展 : <http://itpro.nikkeibp.co.jp/article/COLUMN/20111122/374827/?k2>

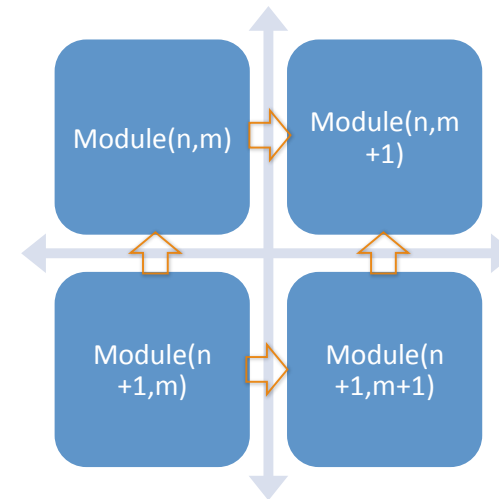
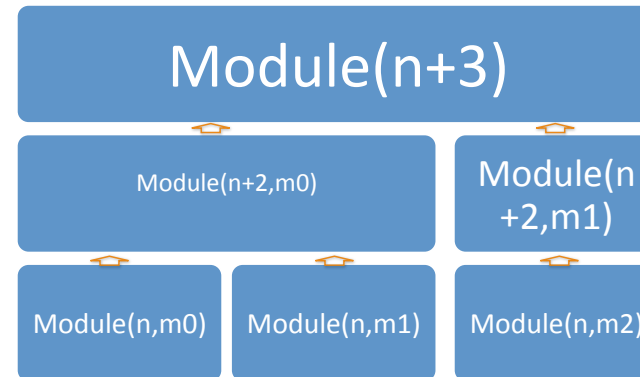


その他のクラウドサービスも

- WebARENA Cloud 9 (NTT PC コミュニケーションズ)
 - 2011年5月にストレージ障害発生
=> サービス停止
- NIFTY Cloud (@nifty)
 - 2011年11月にDNS 設定による障害発生
- さくらのクラウド (さくらインターネット)
 - 2011年12月にストレージに起因する障害
=> 現在もサービス無償化
- Windows Azure (microsoft)
 - 2012年2月にうるう年に起因する障害発生
- Gmail (google)
 - 2009年5月にほぼ全サービスに影響する大規模障害
 - 2011年2月にソフトウェア更新に起因する障害

クラウドにして大丈夫なのか？

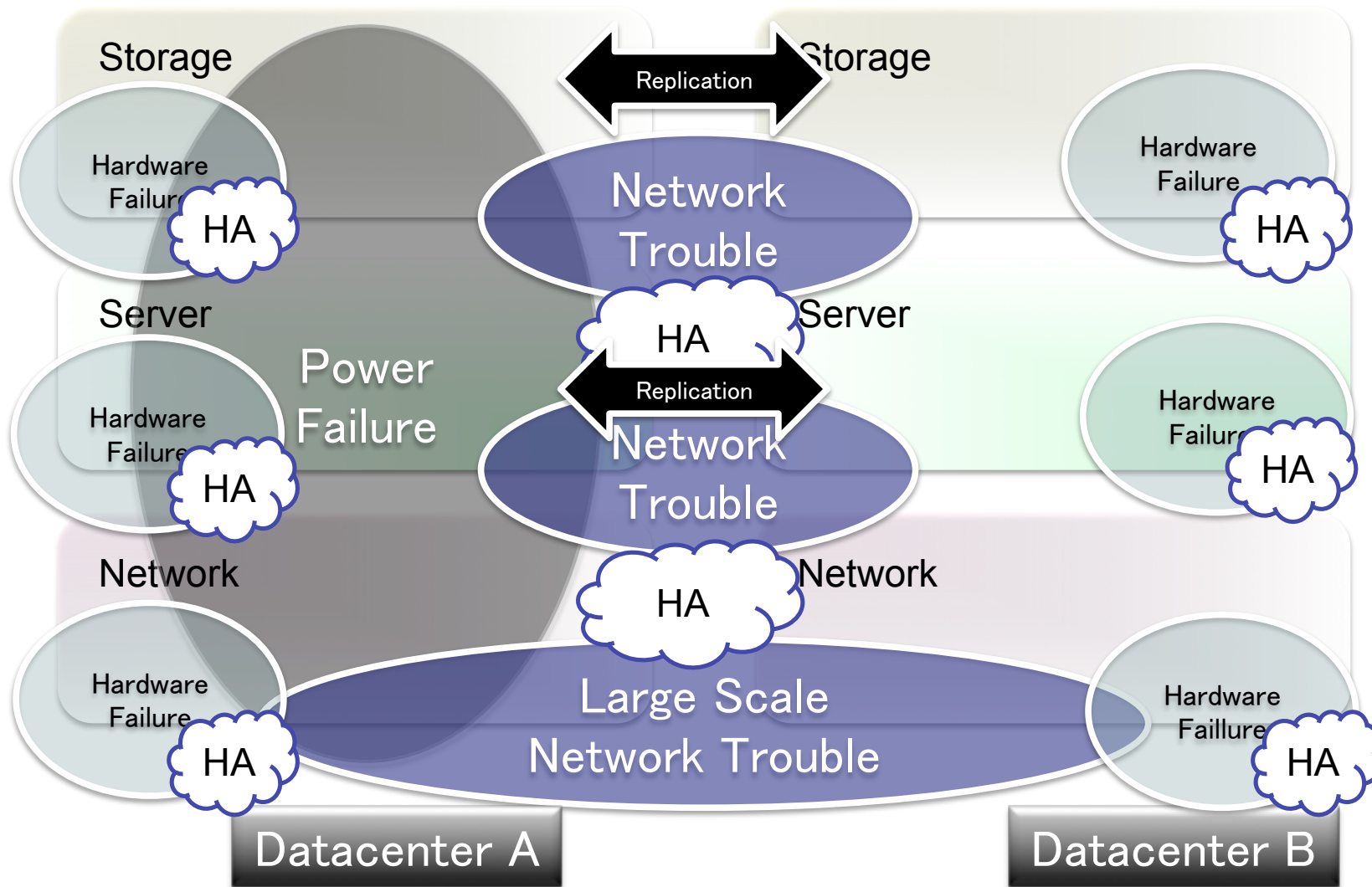
- 信頼できるものなのか
 - 障害は数多く発生している
- 従来のホスティングやVPS に比べて
 - 冗長性
 - 可用性
 - 秘匿性
- 障害発生時
 - アーキテクチャの依存関係が大きいため障害が広範囲になりがち



クラウドとしての運用

- クラウド？
 - 本当にクラウドとして利点を生かした運用がなされているのか
- 障害の影響
 - 機器故障
 - 拠点レベルの障害
- パブリッククラウドサービスとしての IaaS
 - 広域分散 IaaS クラウドへの挑戦
 - 大学間にて構築されるインタークラウド

広域分散 IaaS の冗長性



IaaS の信頼性・冗長性確保への課題

- ハードウェア
 - 壊れたら交換という割り切りも多い
- VMの冗長化
- ストレージの冗長化
 - key-value ストアによる多重化、分散化
 - Block Device は多重化、分散が困難
 - **ストレージの Migration は大きな課題**
- 現状は「ゾーン」という概念が多用されている
 - ユーザがサービスを起動する「ゾーン」を選択できる
 - 異なったゾーンに分割してサービスを複製することによる冗長性確保
 - **拠点単位の冗長性も確保できるはず**

IaaS の可用性を上げるためには



- オープンなクラウド環境
 - 特定クラウドに閉じた技術、環境提供はユーザの利便性を下げる結果となる
 - Open Cloud Framework
 - 信頼性の確保と検証
 - SLA
 - コストとのバランス
 - 「VPS」ではなく
- 「IaaS クラウド」としてのサービス提供、運用**



必要とされる技術



- ネットワーク移動透過性
 - VM が利用するネットワークを拠点間にて透過的に提供
- ストレージ可用性
 - 単一ハードウェアや単一拠点に頼らないストレージの提供
- 計算資源冗長性
 - 単一ハードウェアや単一拠点に頼らない計算資源
 - COLD standby vs. HOT standby






学術・研究における クラウド要素技術動向



大学におけるクラウド利用



- 事務系端末
 - シンククライアント (IaaS)
- 事務系アプリケーション
 - 財務会計・学事 (PaaS, SaaS)
- 研究者
 - 計算資源・実験・サービス提供 (IaaS, PaaS)
- 授業
 - 授業端末 (IaaS, SaaS)



大学間クラウドの利点



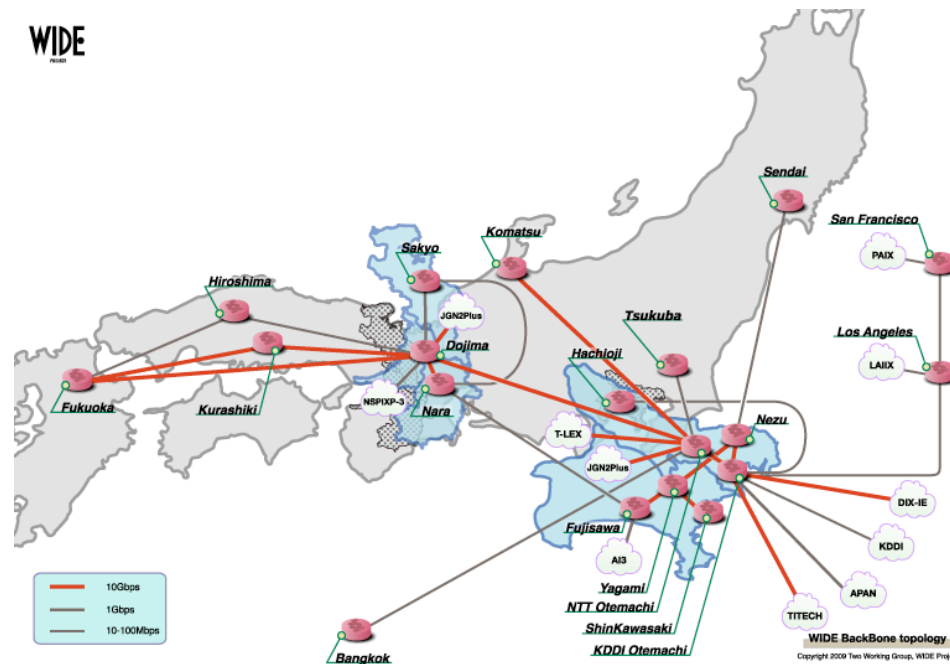
- リソースの融通
 - 大学が持つリソースを融通し合いサービスを提供
- 耐障害性
 - ハードウェア障害によるデータ喪失やサービス断への対応
 - ネットワーク障害によるサービス継続不能への対応
- 運用コストの削減
 - 法定点検による停電対応
 - 重複したリソースの共通化



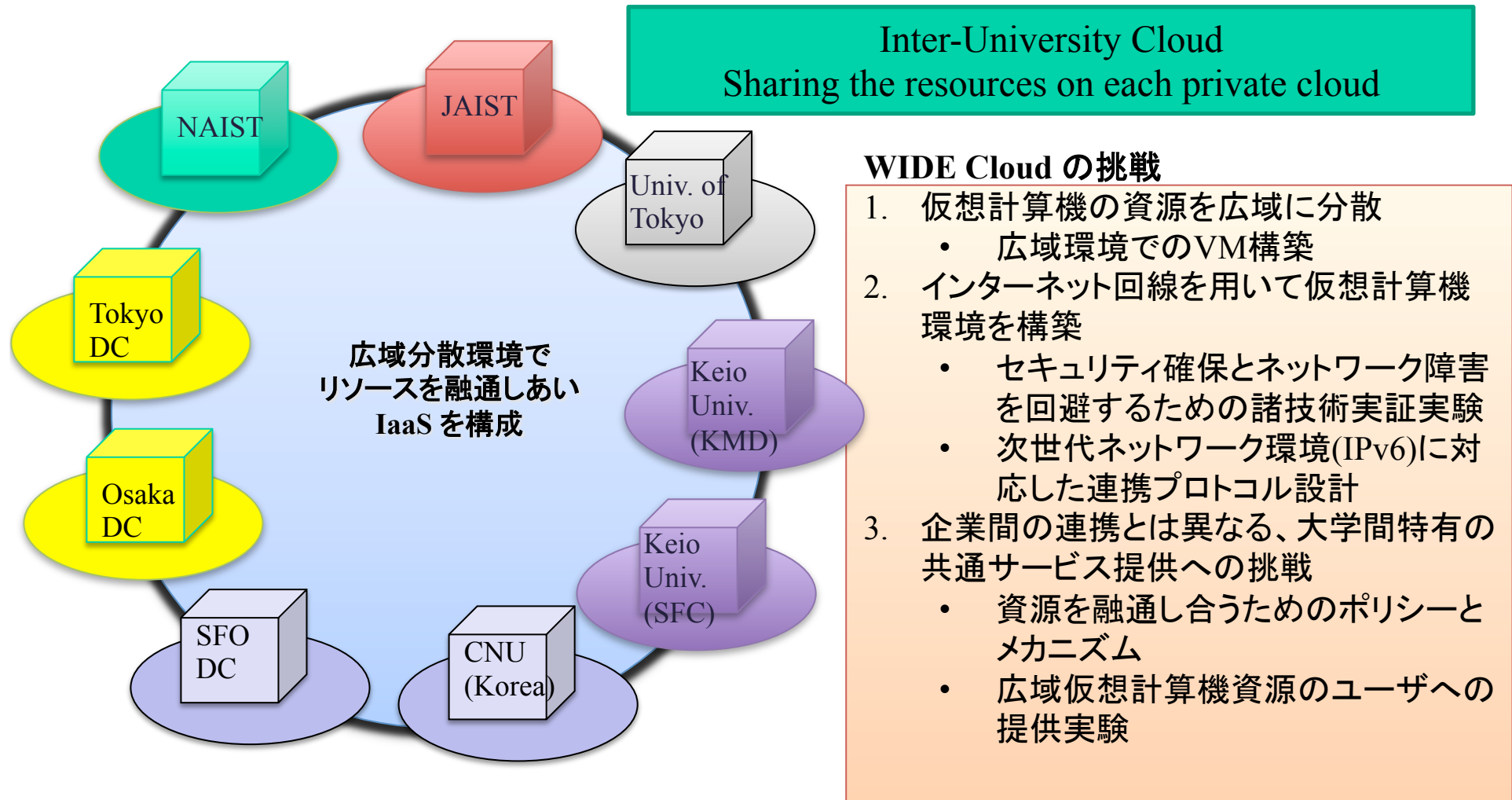
大学間クラウドのテストベッド



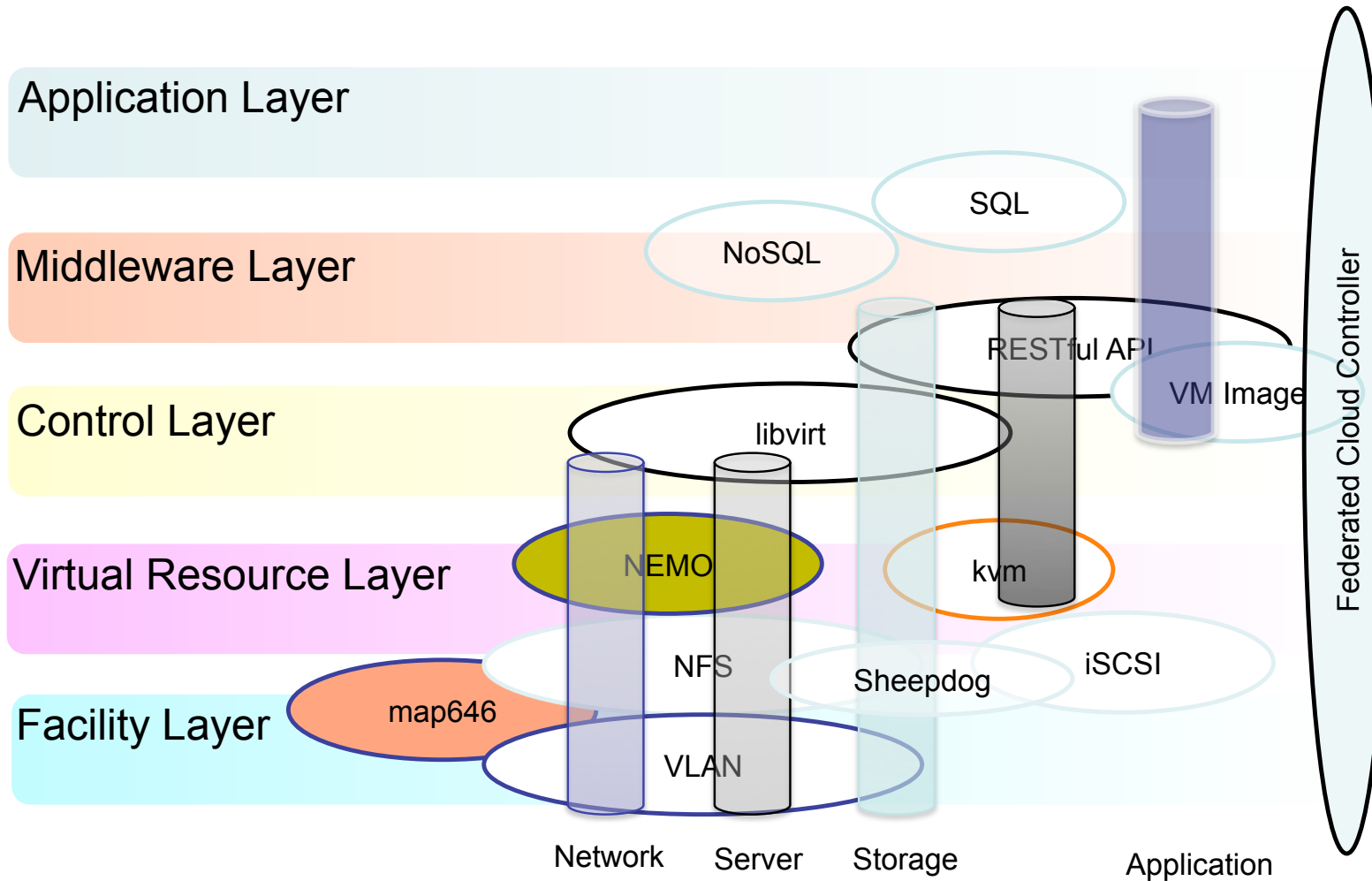
- WIDE / JGN バックボーンを利用した広域 IaaS クラウド
 - 慶応大学
 - 東京大学
 - NAIST
 - JAIST
- 接続形態
 - 専用線
 - JGN
 - Tunneling



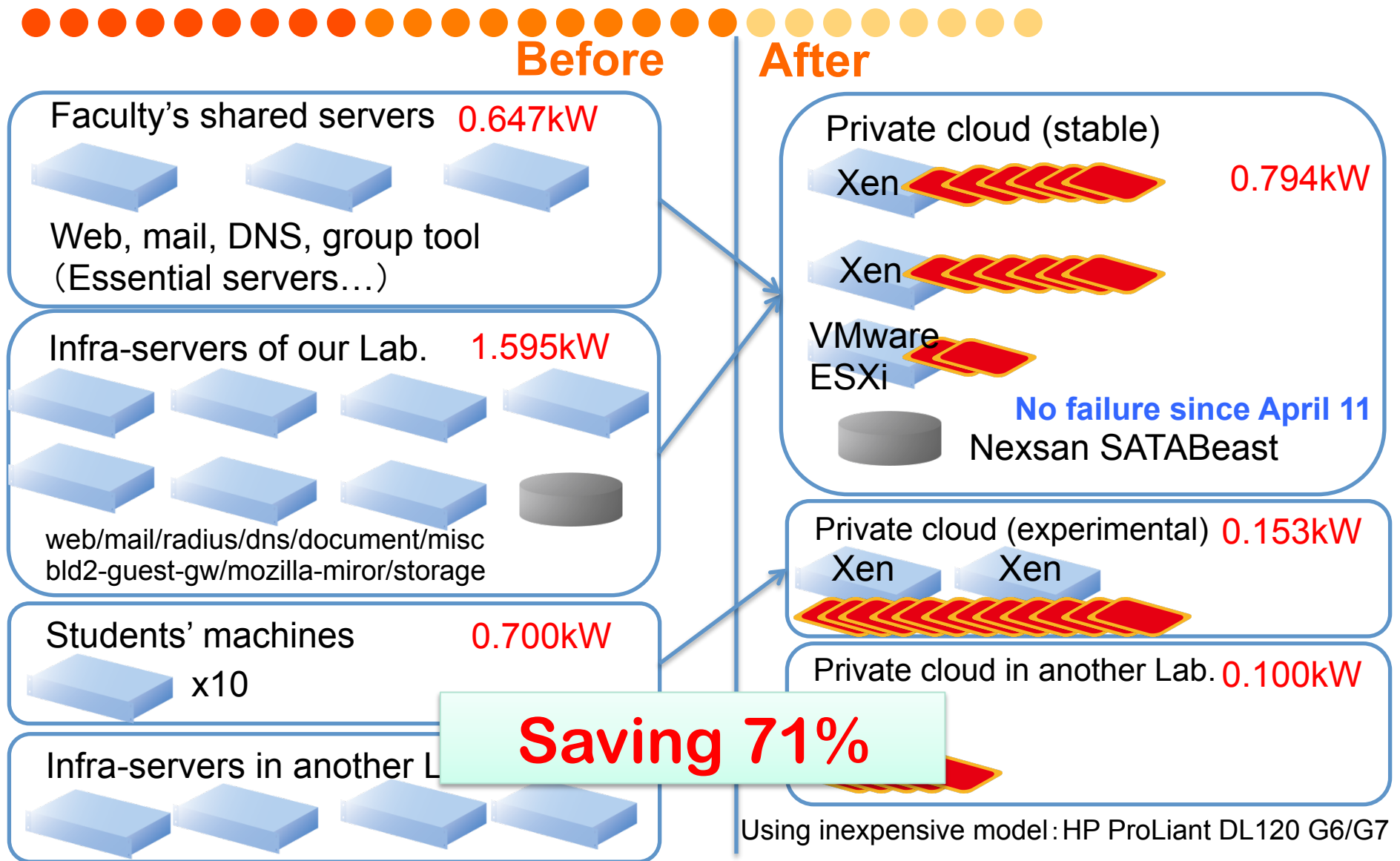
WIDE クラウド



WIDE クラウドの要素技術



東京大学におけるクラウド化



4 Outputs in the summer 2011

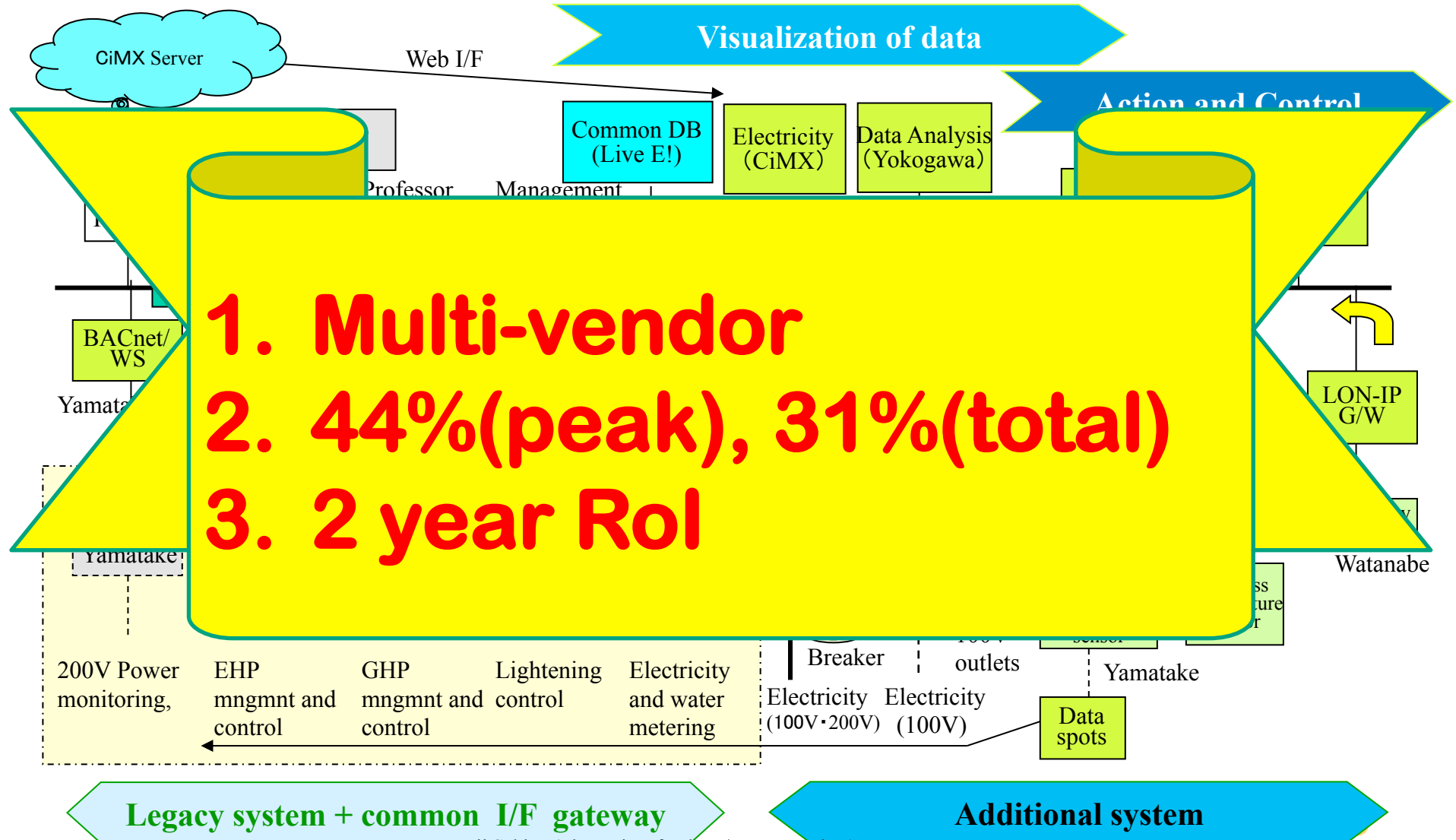


1. Preserve (and Improve) Performance
 - As **"Risk Management"**, i.e., against energy shortage/outage
2. Performance
 - **Δ44%** for Peak, and **Δ31%** for total (at Eng.No.2 Bldg)
 - **2 years RoI** (at Eng.No.2 Bldg) ;
0.6 MUSD investment, 1.0 MUSD Electricity Bill
 - **0.5 year RoI** (for virtualization of servers at Esaki Lab.)
3. Sustainability
 - **Open** architecture
 - **Multi-vendor** operation (**more than 10 vendors**)
4. New Business Development
 - Improvement of business activities



東京大学 工学部2号館の節電システム

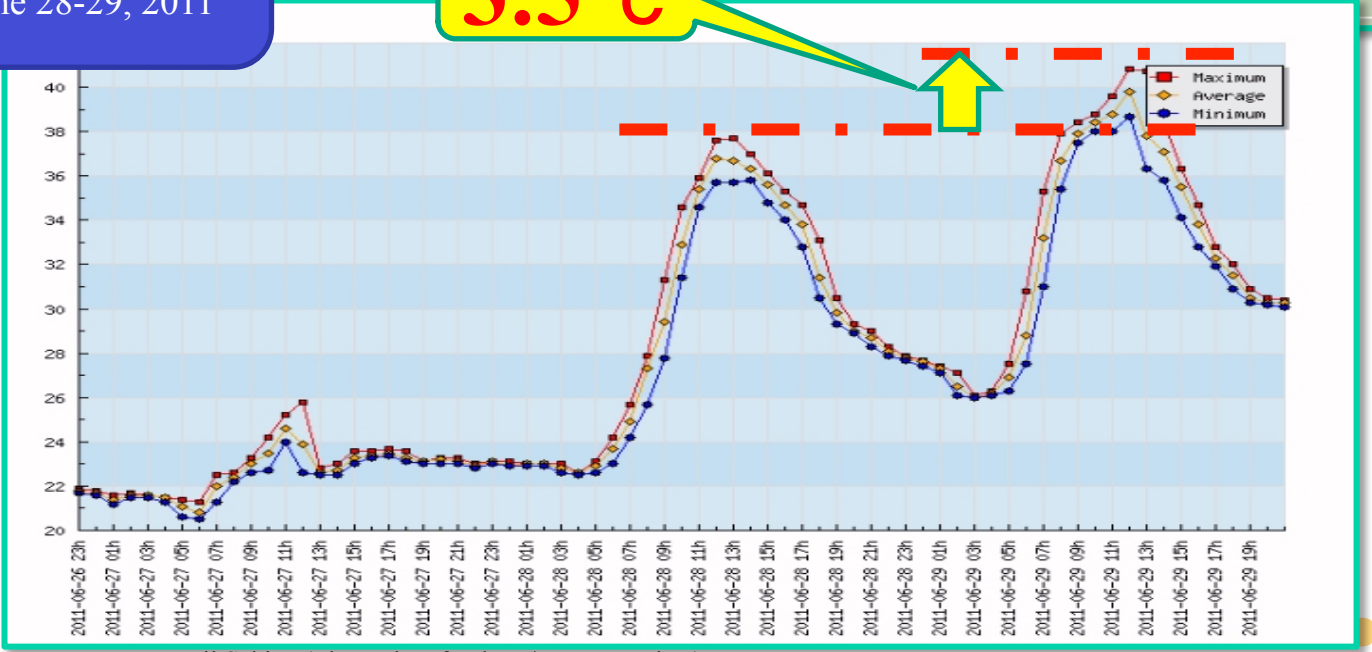
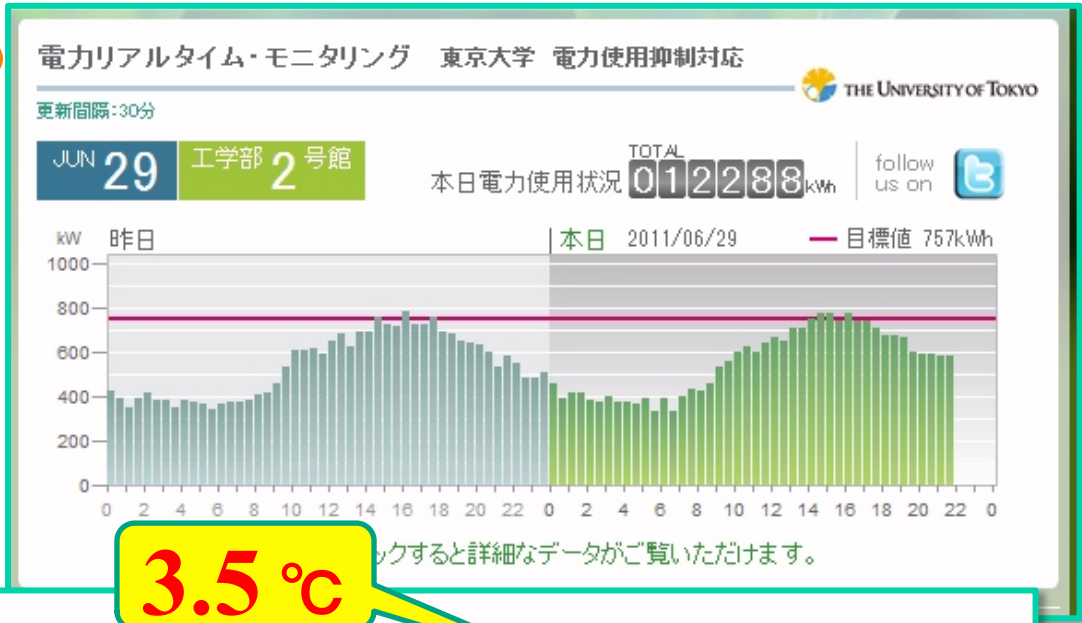
Data Integration among legacy sub-systems



監視システム



Effect of visualization?
 ↓
 Even with different
 temperature, power
 usage is the same !
 June 28-29, 2011

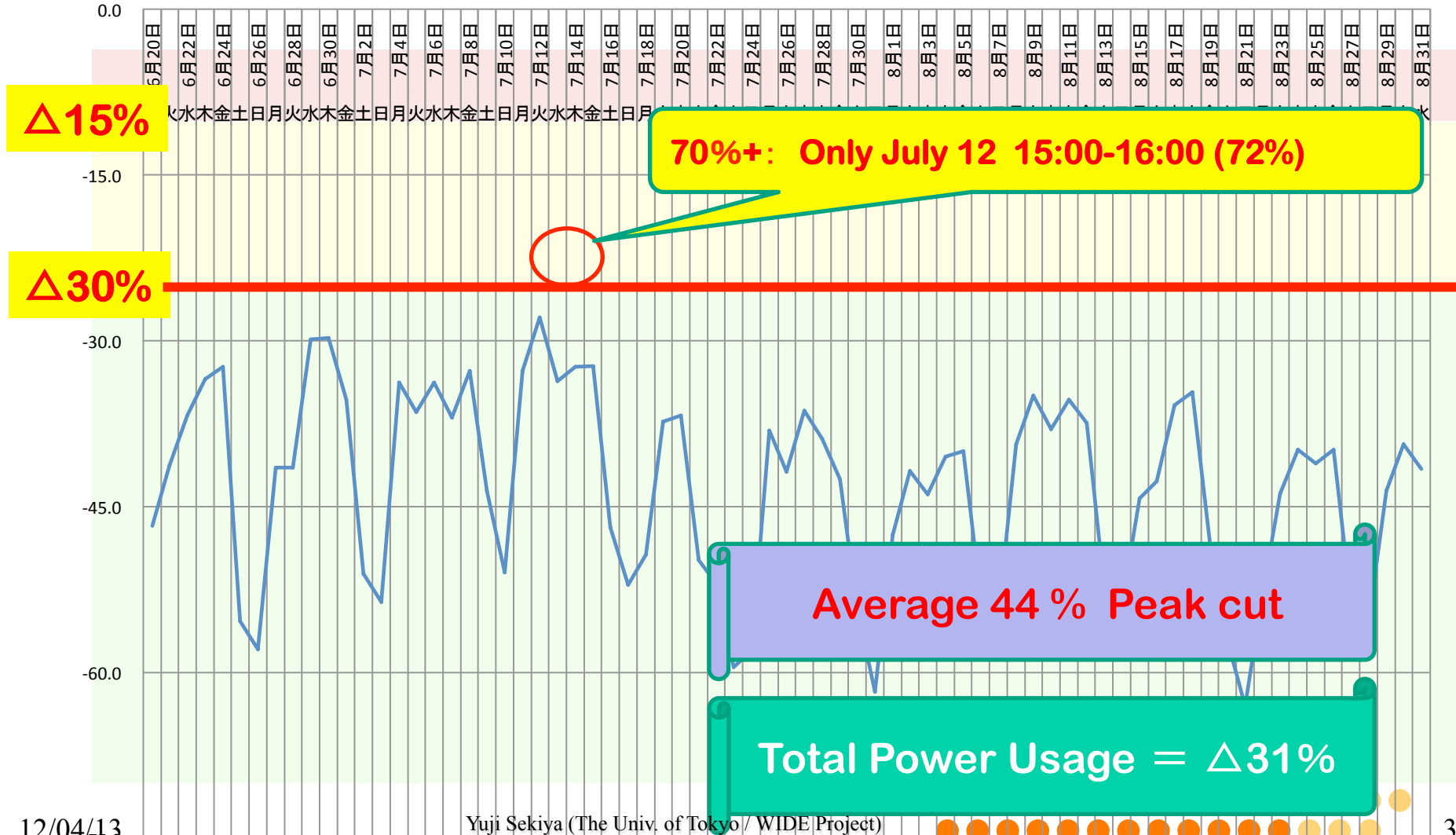


昨年度夏の節電状況



平均削減率(%)

2011/6/20-8/31 Eng.No.2 Bld.



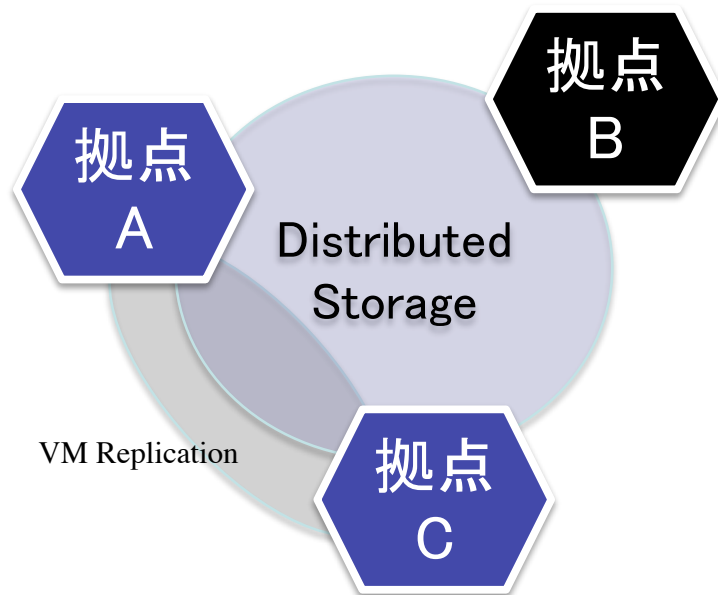
挑戦と課題

- 広域分散 IaaS への挑戦
 - 「クラウド」としての利点を生かした運用
 - リソースの融通
 - 拠点単位の障害に対応できるクラウドとしての技術
- 課題
 - 計算資源
 - 拠点をまたがったマイグレーション
 - 冗長化のためのスタンバイ手法
 - ネットワーク
 - I/O としてのネットワーク
 - サービスとしてのネットワーク
 - 光の速度は超えられない
 - ストレージ
 - 閉域 (SAN) vs. 広域分散
 - 運用コストと耐障害性のトレードオフ

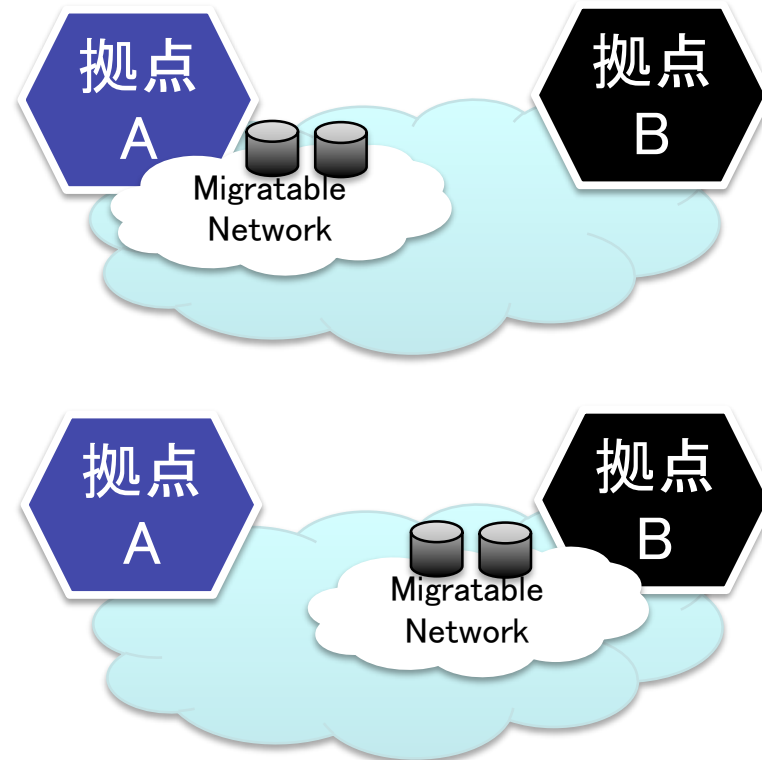
クラウド要素技術の研究



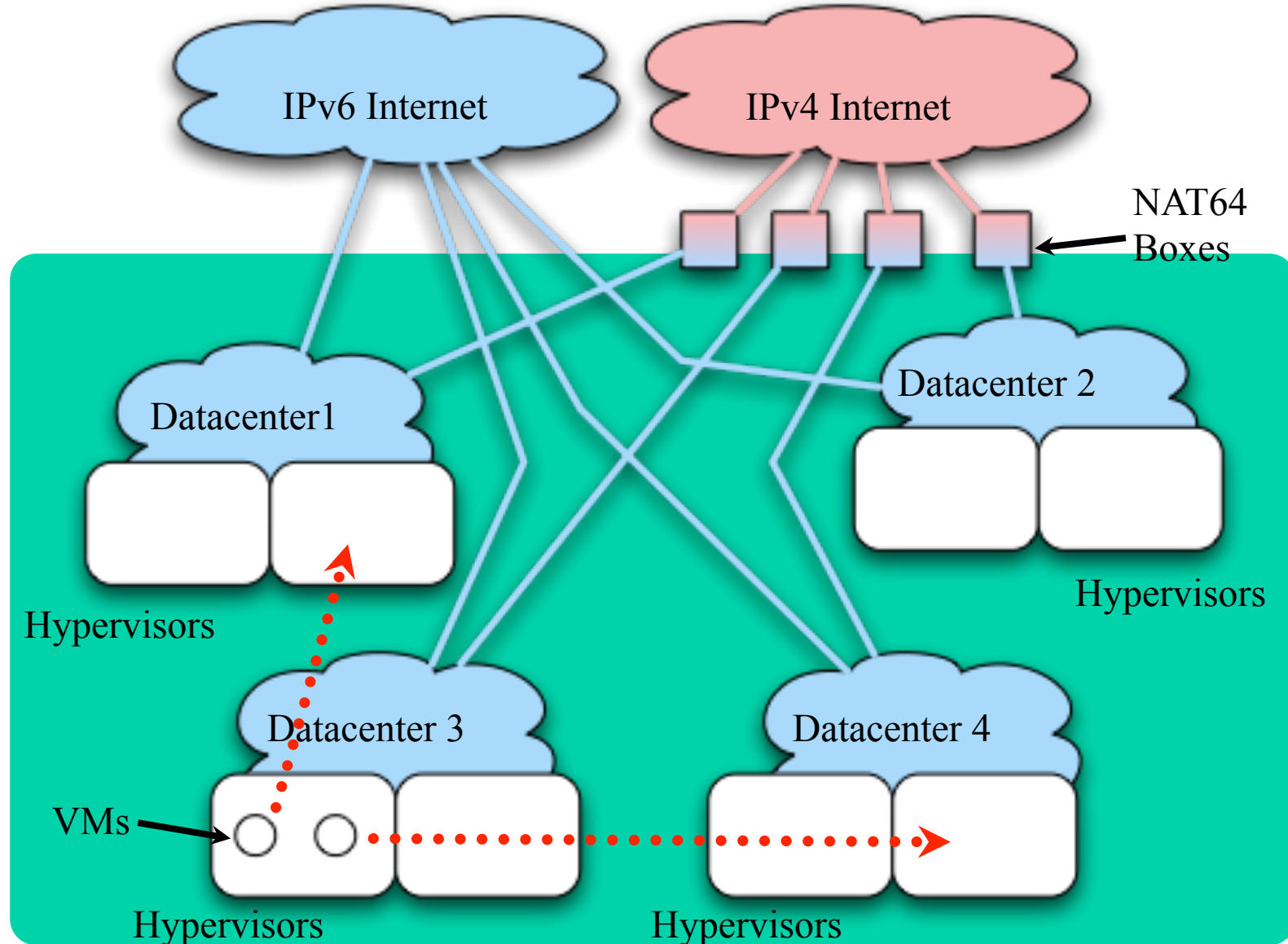
拠点単位の障害回避



ネットワークの柔軟性



ネットワーク移動透過性の実現



データセンター間ネットワーク

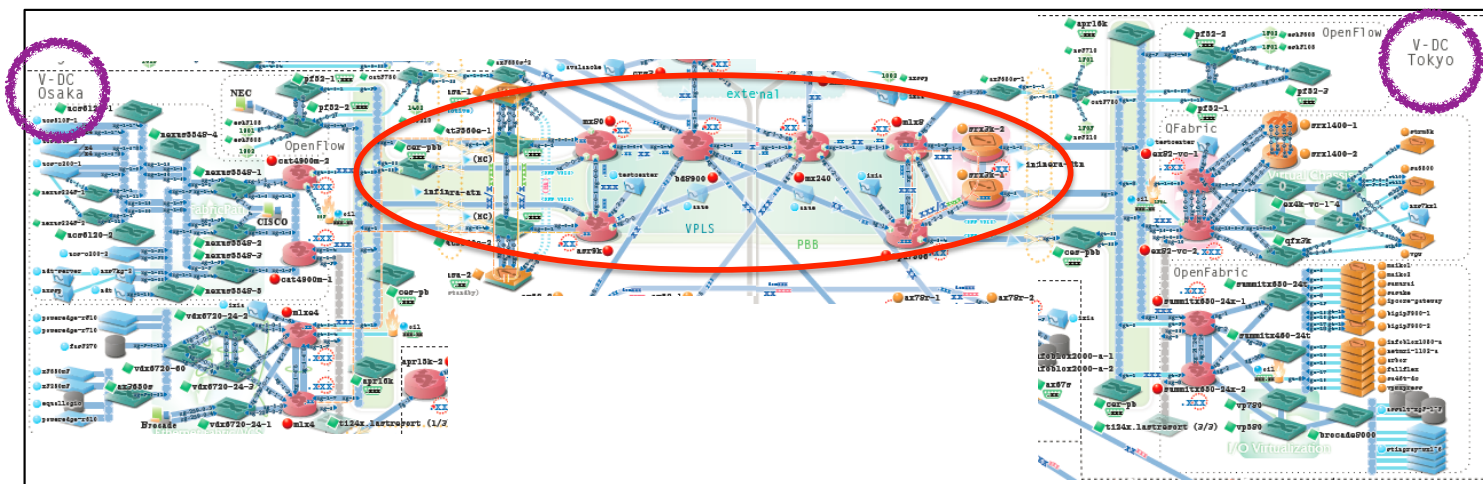


- データセンター間接続技術
 - 専用線 (VLAN)
 - Q-in-Q
 - Mac-in-Mac
 - ネットワーク仮想化
 - VPLS
 - VXLAN
 - OVT
 - SDN (OpenFlow)
- Interop Tokyo における実証実験
 - Interop ShowNet 2011
 - Interop ShowNet 2012
- IETF による標準化
 - nvo3 WG

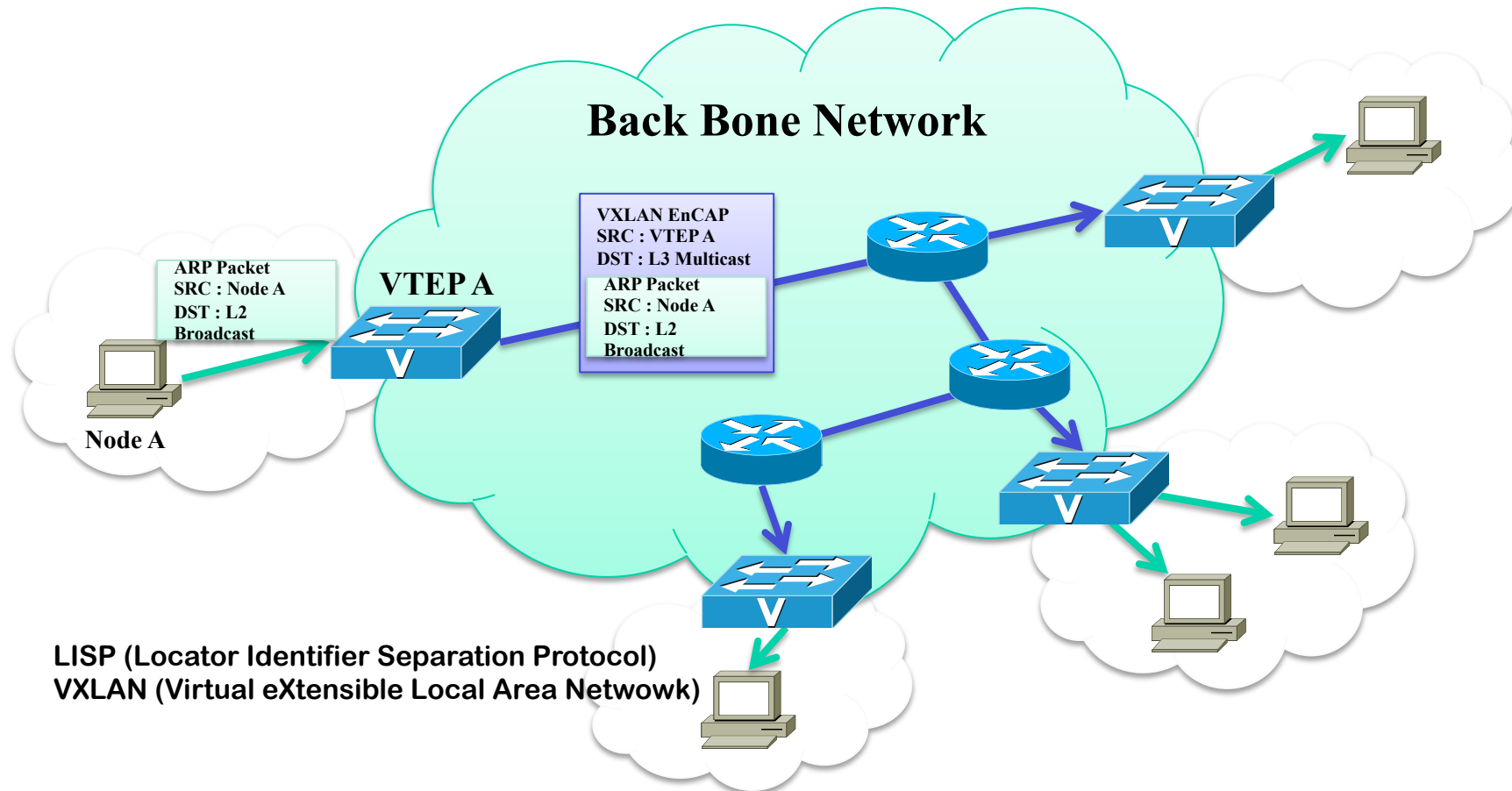


Interop ShowNet におけるクラウド技術

- Interop Tokyo
 - 毎年6月上旬に幕張メッセで開催
 - ネットワーク技術の祭典
- ShowNet
 - 最新の機器・技術を使って生きたネットワークを構築
 - NOC Generalist (2010年から)
 - 2011年は VPLS + PBB による広域 L2 マネージメント技術を検証



LISP+VXLAN によるネットワーク仮想化



AaaS (AS as a Service) の実現



クラウドストレージ技術

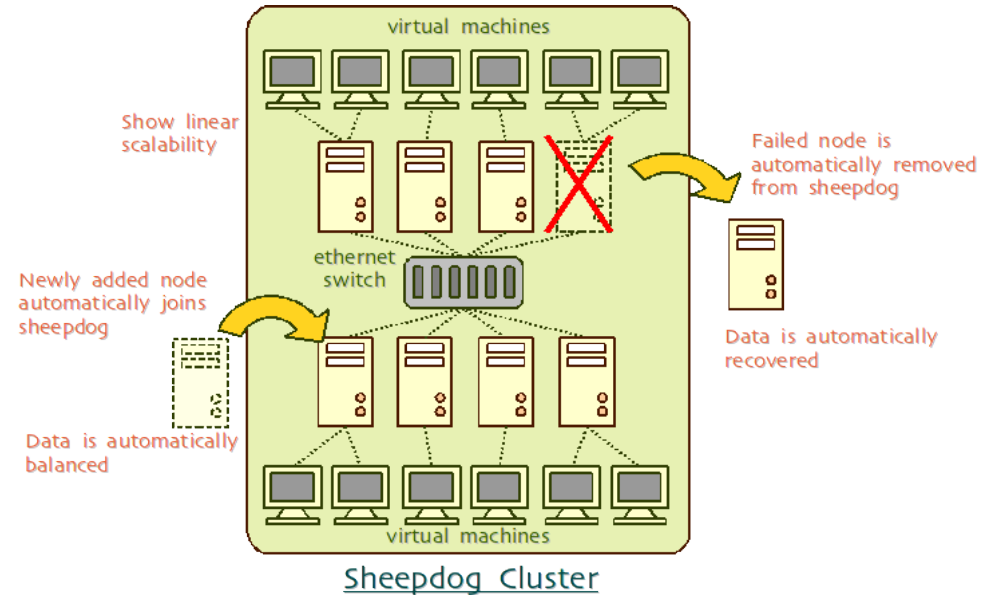


- レプリケーション v.s. 分散ストレージ
 - ストレージがボトルネックになる
 - 特に IaaS
- レプリケーション
 - ストレージアプライアンス
 - DRBD
- 分散ストレージ
 - Ceph
 - Sheepdog
 - GlusterFS
 - Gfarm



分散ファイルシステムの可能性

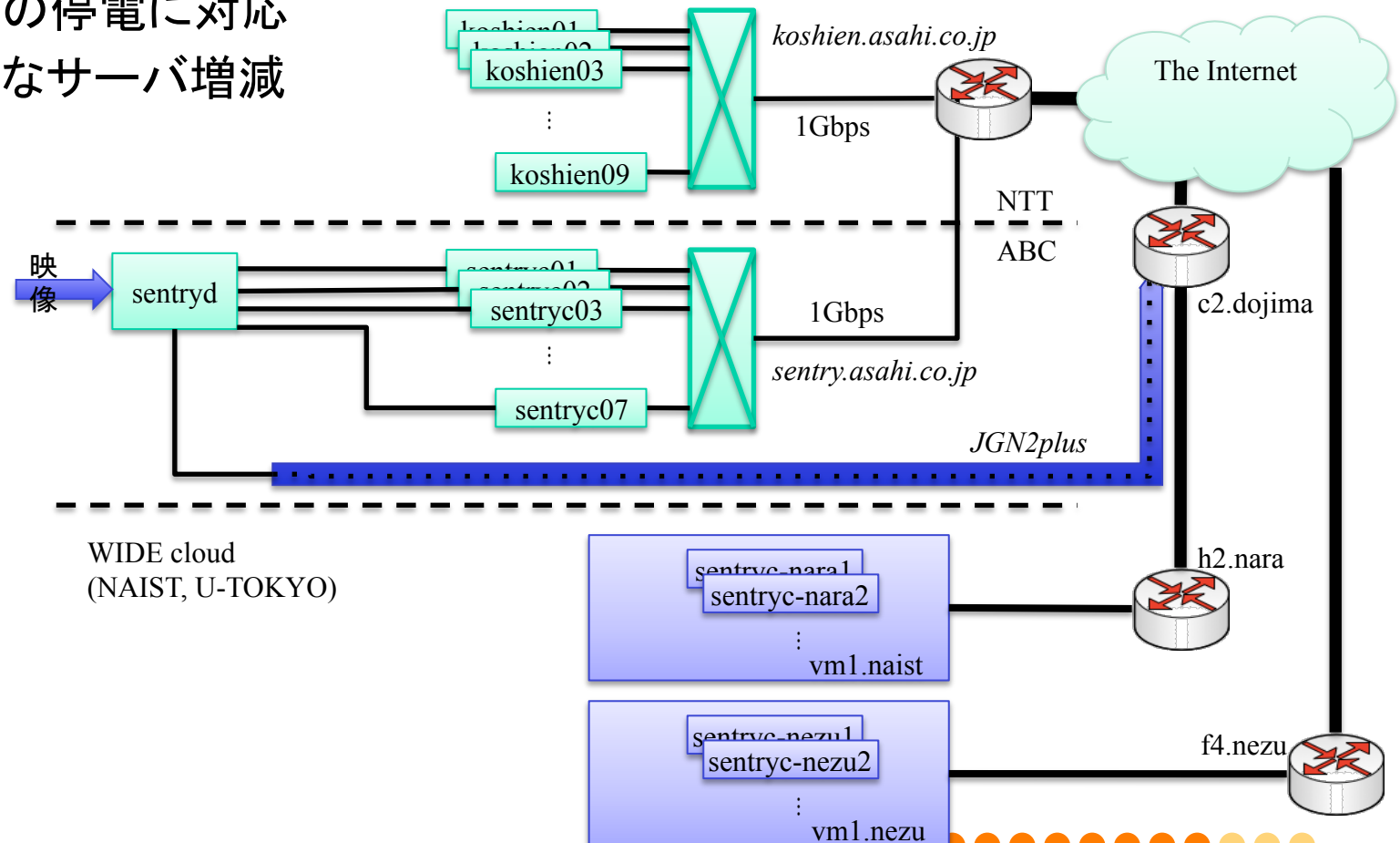
- 100台以上のサーバを用いて実験
 - StarBED @ NICT 北陸リサーチセンター
 - 遅延を入れて広域環境をエミュレーション
- 障害が発生した場合のリカバリー
 - ナイーブな技術



Sheepdog
<http://www.osrg.net/sheepdog/>
GlusterFS
http://gluster.org/community/documentation/index.php/Main_Page

実証実験事例

- 2010年, 2011々夏の甲子園中継
(ABC放送との共同研究)
- 拠点の停電に対応
- 動的なサーバ増減





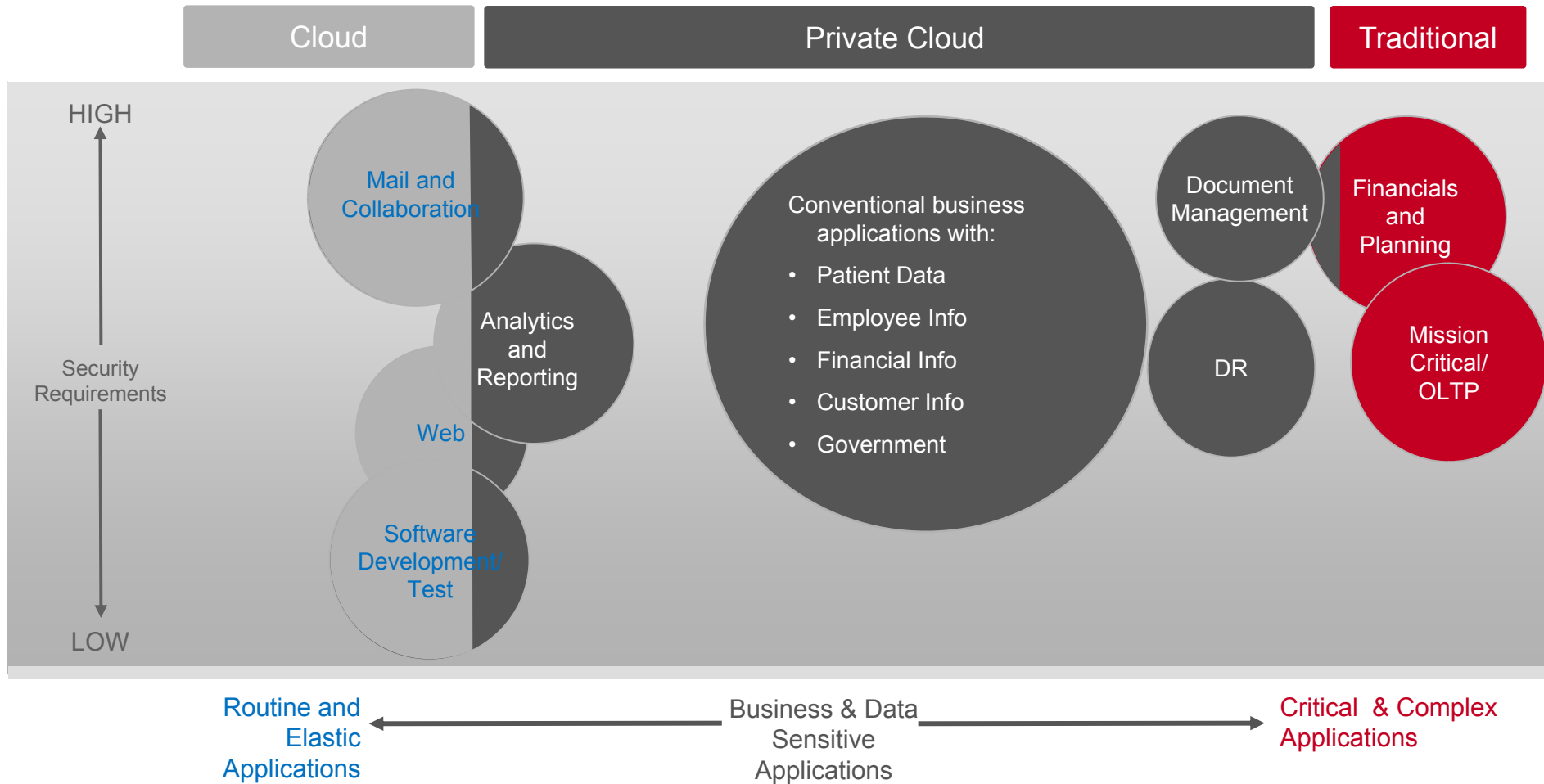
適切にクラウドを利用する



用途別クラウドの利用法



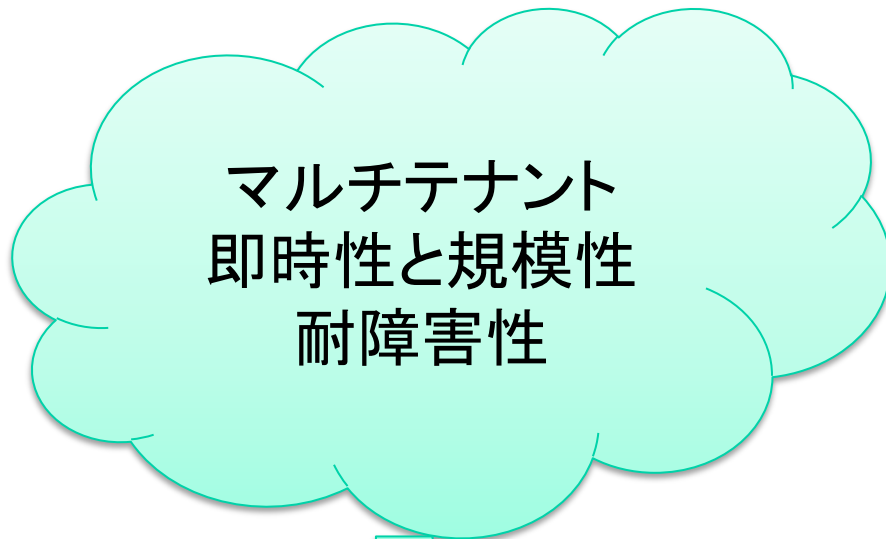
米国 rackspace 社による定義



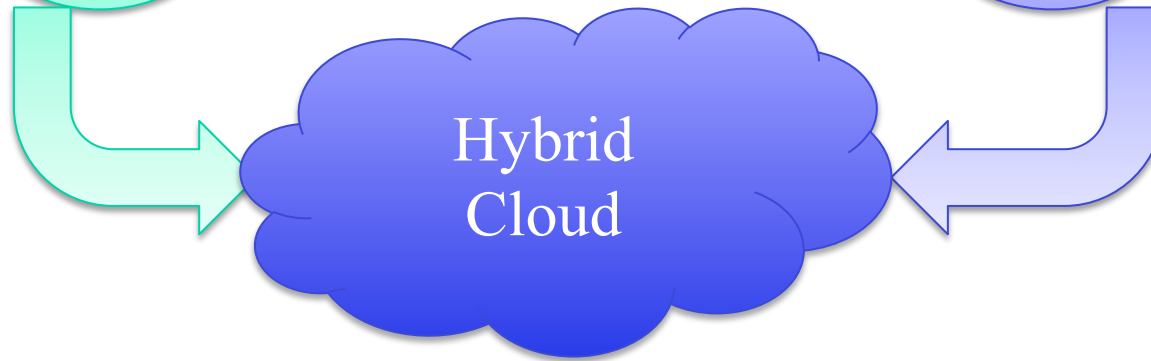
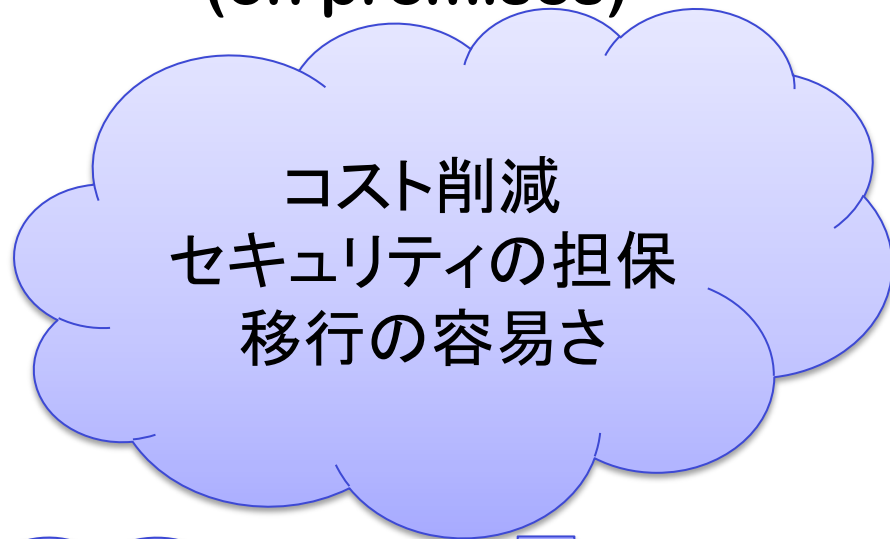
Hybrid Cloud の可能性



Public Cloud



Private Cloud (on premises)



Hybrid Cloud の可能性

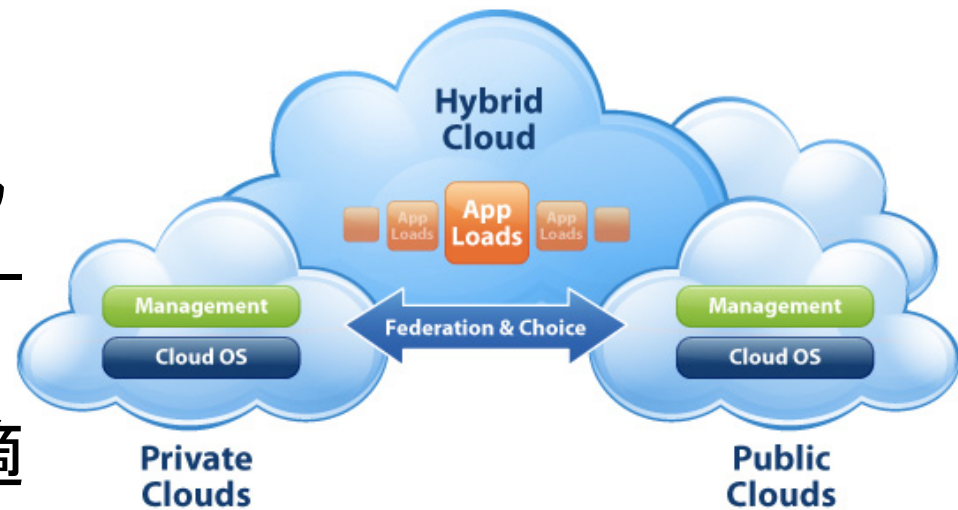


- Public Cloud と Private Cloud (on premises) との融合による利点取り
 - 法的 (言い訳的) に社内に置いて管理しなければならないデータは Private Cloud
 - 計算資源や外部サービスサーバは Public Cloud
- リソースの融合
 - どのクラウドからどれだけのリソースを？
 - クラウド間の相互接続性問題
 - VPN or 専用線
 - データ形式や VM イメージの相互利用



その先に。。。

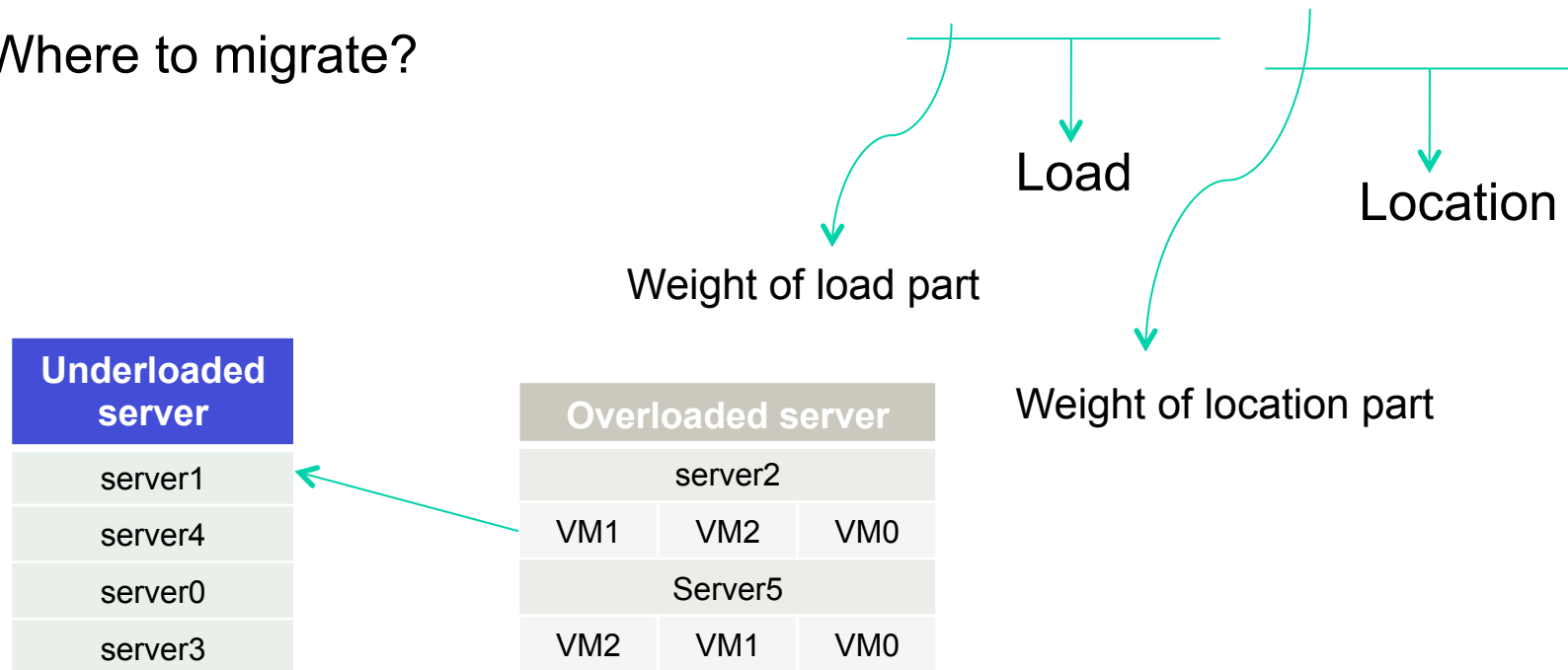
- Federated Cloud (連邦型クラウド)
 - リソースをポリシーと組み合わせて管理
 - 異なった性格のクラウドを組み合わせてサービスを生み出す
- 大学間クラウドには適した概念
 - WIDE Cloud においても実験中



WIDE Cloud における広域リソース計算



Where to migrate?



Initiate migration when $Migration\ duration = \frac{MEMu}{NETthr} < T$



本日のまとめ



- クラウドをクラウドとして利用する
 - 「責任のなすりつけ」では有効利用できない
 - コストダウンは魔法でできているわけではない
 - 囲い込み、危険度が上がっている
- クラウドを支える技術
 - 過去の技術のインテグレーション
 - 「クラウド」としての新しい技術
 - まだ発展途上段階
- 何ができて何ができないのか
 - 見極めて使えば効率的

