

社内クラウドを活用した研究開発環境の構築

桑田 喜隆¹ 武田 健太郎 岩谷 正広 飛内 拓弥

NTT データ

Building Research and Development Environment on Private Cloud

Yoshitaka KUWATA Kentaro TAKEDA Masahiro IWATANI Takuya TOBINAI

NTT DATA CORPORATION¹

概要

近年、プライベートクラウドを活用して既存の情報インフラを統合化し、運用を効率化する動きが活発に行われている。筆者らは、研究開発活動を活性化するため、社内でも手軽に利用可能な研究開発用のプライベートクラウドを構築し、その試行運用を1年半以上行っている。本稿では、研究開発用クラウドのコンセプトおよびその活用状況を報告する。

Abstract: It becomes very popular to make use of private cloud for the integration of legacy IT infrastructure. In order to accelerate our activities of Research and Development, we built a private cloud for R&D purpose, on which all of employees can use virtual servers. We had run the private cloud for more than 18 months. We report the design concept and activities of the private cloud.

Key Words: Cloud Computing, Private Cloud, Research and Development on Cloud.

1. はじめに

NIST のクラウドコンピューティングの定義¹⁾によると、クラウドはその提供される形態からプライベートクラウド、コミュニティクラウド、パブリッククラウド、ハイブリッドクラウドに大別される。プライベートクラウドは、社内等で占有して利用するために独自に構築したクラウドである。一方、パブリッククラウドは一般利用者に対してサービスを提供するために構築したクラウドである。コミュニティクラウドは複数の団体が共同で利用する目的で構築したクラウドで、団体の中でのみ利用される。ハイブリッドクラウドはプライベートクラウドとパブリッククラウドとをうまく組み合わせて、お互いの長所を活かす設計を行うクラウドを指す。

近年、企業や政府・地方公共団体等においてはプライベートクラウドを活用して既存の情報インフラを統合化することで、その構築および運用を効率化する動きが活発に行われている。これまで複数のサーバから構成されていたシステムを、

そのまま仮想化して一台の物理サーバに集約することで、サーバの購入コストや電気代などの運用コストを下げることも期待される。

他方、Amazon EC2²⁾に代表されるパブリッククラウドの活用も広がりを見せている。例えば、Amazon EC2では、利用料金の決済の方法さえ持っていれば、誰でも世界中から Amazon のデータセンター上の仮想マシンを自由に借りてサーバを構築することが可能である。特に人手を介さないセルフサービスの機能によって、ほとんど待たずにサーバを借りることが可能である点は非常に重要であると考えられる。負荷に応じて借り出すマシン数を増減したり、サーバを一時的に借りる等のこれまで不可能であった利用形態が可能になる。パブリッククラウドの持つこれらの特性を活かした新たな形態の情報インフラの構築が模索されている。

現在、プライベートクラウドは、その利用目的として仮想化技術を利用したシステムの集約化によるコスト削減および効率向上の観点が目まぐるしく注目されている。しかしながら、筆者らはクラウドの特性を活かすことで、プライベートクラウドにおいても柔軟性や利便性の高いサービスを実現することが可能になると考えている。

例えば、研究開発の分野では、新しいソフトウェアを試してみるために、サーバを用意する必要

¹ Yoshitaka Kuwata
NTT データ 技術開発本部
東京都江東区豊洲 3-3-9 豊洲センタービルアネックス
kuwatay@nttdata.co.jp

があることも多い。また、ソフトウェア開発においては、製造したソフトウェアの試験の段階で試験用のサーバが必要になる。どちらの場合も、一時的な利用であり、試行や試験のフェーズが完了すれば用意した環境は不要になる。クラウド上に一時的にサーバを借りることで、サーバの調達にかかる時間やコストが大きく削減可能である。

そこで、筆者らは社内の研究開発の業務に注目し、社内の試行サービスとしてプライベートクラウドを実際に構築し、研究開発目的で利用することでその有効性や課題等を抽出することとした。

本論文では、まず評価のために構築した「研究開発用クラウド」の概要について述べる。次に、プライベートクラウドの利用仮説について述べた後、実際の利用状況の説明をする。また、プライベートクラウド上のサービスの例をあげ、最後にまとめと今後の課題を述べる。

2. 研究開発用クラウドの概要

図 2.1 に研究開発用クラウドの概要図を示す。

人手をかけずに、プライベートクラウドを運用するため、Amazon EC2と同様のセルフサービス型のクラウド基盤を採用した。前述の NIST の定義では、本サービスは Infrastructure As A Service (IaaS) に相当する。利用者側の操作だけでなく複数台の仮想サーバを利用することか可能となる。

クラウド基盤ソフトウェアとしては、OSS である Eucalyptus を採用し、ニーズに応じて Windows のサポート等の機能の拡張を行ったものを利用した。

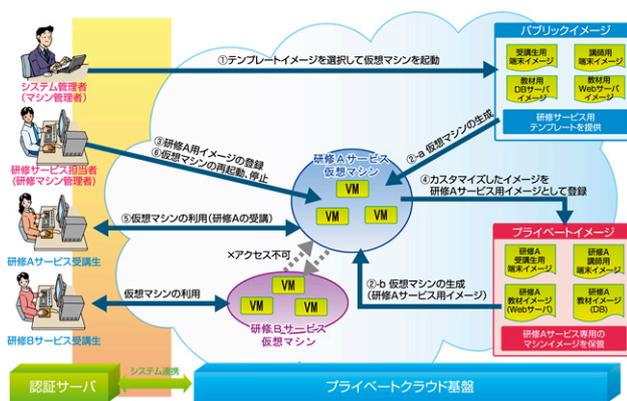


図 2.1 研究開発用クラウドの概要図

研究開発用クラウド構築の目的は以下の2点である。

- (1) 研究開発用途への適合性の評価
- (2) クラウド基盤ソフトウェアの評価

なお、本論文では、(1) の評価についてのみ説明することとし(2) クラウド基盤ソフトウェアの評価に関しては言及しない。

2. 1 試行サービスの概要

以下の表 2.1 に、試行サービスの要件をまとめた

表 2.1 研究開発用クラウドの概要

項目	要件/事項
利用目的	社内での一般の研究開発用途 直接顧客案件に利用することは禁止
試行期間	2009年11月5日から継続中
利用方法	社内認証システム経由での認証
利用者	社内システムの利用権限を持つ社員 および協働者
利用可能ネットワーク	クラウドは開発部門のLANに接続 社内のLANからクラウド上のマシンへはアクセスすることが可能 インターネットへの接続は社内ファイアウォールを経由することで可能
制限事項	クラウド基盤ソフトウェアの評価用でもあるため、事前連絡の上で評価のために停止する場合がある。 試行サービスのためクラウド側でバックアップは行わない。
その他	利用開始時アンケートへの回答

試行サービスのために用意した研究開発用クラウドの主な仕様を表 2.2 に示す

表 2.2 研究開発用クラウドの仕様

項目	事項
ハードウェア	サーバ (13 台) Intel Xeon 2.5/2.6GHz Memory 18GB/256GB Disk 13TB (トータル 192 コアを提供可能)
クラウド基盤ソフトウェア	Eucalyptus 1.6.2 ³⁾ / Xen ・ Amazon EC2 / Amazon S3 互換のオープンソースソフトウェア ・ EBS 起動 Windows 仮想マシンの機能を拡張済み
仮想マシン上で利用可能な OS	Cent OS 5.4 Microsoft Windows XP Microsoft Windows 7 Microsoft Windows Server 2003R2 Microsoft Windows Server 2008
クラウド制御ソフトウェア	専用クライアントソフトウェア (クラウドクライアント)

2. 2 利用者から見た研究開発用クラウド

研究開発用クラウドでは Amazon EC2 と同様に、ユーザは起動時に払い出しを受ける仮想マシンタイプ（仮想マシンの仕様）を選択することが可能である。選択可能な仮想マシンタイプは表 2.3 の 5 種類である。

表 2.3 研究開発用クラウドの仮想マシンタイプ

マシンタイプ	CPU コア数	メモリ (MB)	ディスク (GB)
m1.small	1	768	10
c1.medium	2	1536	15
m1.large	2	1536	20
m1.xlarge	3	2304	30
c1.xlarge	4	3072	40

例えば、研究開発用クラウドにおいて利用者が m1.xlarge の仮想マシンを指定した場合、3 個の CPU コア、2304MB のメモリ、30GB のディスクを持つ仮想マシンが払い出される。

また、ユーザはクラウドクライアントと呼ばれる専用クライアントソフトウェアを利用して仮想マシンの操作を行う。

図 2.2 はクラウドクライアントの仮想マシンの操作画面の一例である。仮想マシンに対して、利用者の実施する主な操作は以下の通りである。

① 仮想マシンイメージの一覧

仮想マシンイメージの一覧から、起動する仮想マシンのひな形を選択する。例えば、Windows XP のマシンを利用したい場合には、該当する仮想マシンのテンプレートを選択する。一度利用して変更を加えた仮想マシンのイメージもこのエリアに一覧表示される。

② 起動中の仮想マシンインスタンス一覧

自分の起動した仮想マシンインスタンスが一覧表示される。このエリアから選択して対象となる仮想マシンにログインしたり、仮想マシンの停止を指示したりすることが可能である。

③ セキュリティグループ

セキュリティグループとは、ある仮想マシンインスタンスへアクセスすることのできるポート番号である。このタブによってセキュリティグループを指定することが可能である。

④ 仮想ボリューム

このタブによって仮想マシンから利用する仮想ボリューム（外部ディスク）を操作する。仮想マシンに対して仮想ボリュームの払い出し、取り付け（マウント）および取り外し

（アンマウント）の操作を行う。

⑤ パブリック IP アドレス

このタブによって仮想マシンに対してパブリック IP を付与することで、外部からはパブリック IP アドレスを指定してアクセス可能となる。

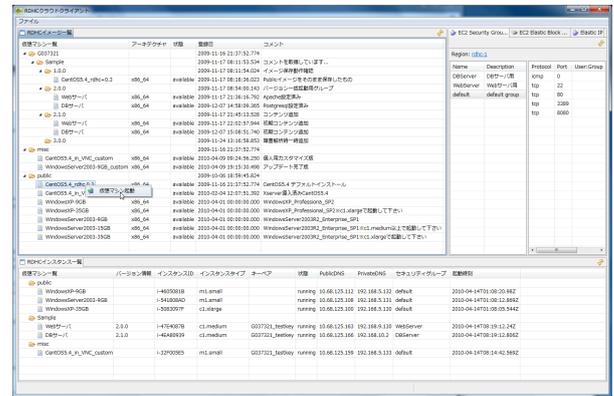


図 2.2 クラウドクライアントの操作画面

3. 利用仮説

本試行サービスを開始するにあたり、研究開発用クラウドの利用用途として以下が想定された。

1. 社内情報共有
社内 WEB ページやファイル共有等の社内での情報共有を目的とした利用。
2. ソフトウェア試行用
OS やミドルウェア、アプリケーションなどの試行目的での利用。
3. ソフトウェア試験用
自分達で開発したソフトウェアの試験環境としての利用。
4. その他

4. 利用状況の評価

利用仮説の検証のため、以下の 2 種類の方法で利用状況の分析評価を実施した。

1. 利用者アンケート
利用開始時にクラウドの運用担当者から各利用者にメール等でアンケートを依頼している。そこで、第一ステップとしてアンケートの分析を行う。
2. 利用状況ログ
クラウド基盤ソフトウェアは利用状況のログを取得する機能を有する。そこで、利用者がどのようにクラウドを実際に利用したかの分析を行う。

以下に、分析結果を述べる。

4.1 利用者アンケート分析

利用者アンケートの概要を表 4.1 に示す。

表 4.1 利用者アンケートの概要

項目	
評価期間	2010年9月10日から 2011年6月26日
登録利用者数	128
有効回答数	30
質問の項目	<ul style="list-style-type: none"> ・ 利用予定期間 ・ 利用目的 ・ 利用予定の仮想マシン数

なお、登録利用者に対して有効回答率が低い、登録済みグループメンバーにはアンケートを実施しなかったこと、回答は任意としたためである。

4.1.1 利用予定期間

利用規程で3ヶ月以内を原則とするとアナウンスしているため、アンケートもほとんど3ヶ月以内の回答となっていた。

4.1.2 利用目的

利用目的は様々であるが、概ね利用仮説で想定していた利用目的であった。表 4.2 に利用目的と回答数を示す。

表 4.2 利用目的の集計結果

分類	回答数
社内情報共有用	8
ソフトウェア試行用	14
クラウドそのものの試行	7
ソフトウェア試験用	1
合計	30

当初想定されていなかった利用としては、クラウドそのものを試したいという回答が7人で、全体の約1/4あったことが興味深い。主な利用者の所属が技術開発を行う部署であるため、新しい技術に興味を持っている人が多いことが考えられる。

クラウドそのものの試行を含めたソフトウェアの試行用途としては21人の回答があり、全体の60%を占める。ソフトウェアの試行が研究開発用クラウドの最も多い用途であると言える。

他方、ソフトウェア試験用に利用すると回答した人は1名と少なく、クラウドでのソフトウェア開発があまり定着していないことが分かる。

4.1.3 利用予定の仮想マシン数

多くのユーザは利用目的に合わせて1-3台の仮

想マシンを利用するとの回答であった。例外的にソフトウェアの試行で、分散処理ミドルウェアである Hadoop⁴⁾の試行を行うユーザが最大10個程度を利用したいと記述している。

4.2 運用ログの分析

今回実施した運用ログの分析の概要について、以下に示す。

分析期間：2010年9月10日～2011年8月9日
分析方法：クラウド基盤の運用記録の分析
クラウド基盤ソフトのログ分析

なお、4.1で実施したアンケート分析と運用分析との相関を取る分析までは、今回は実施していない。

4.2.1 利用率の時系列分析

研究開発用クラウド全体の利用率の分析を行った。図 4.1 は CPU リソースの利用状況を日単位に時系列で示した物である。

図では横軸が日付を、縦軸が利用中の CPU コア数を示している。

例えば、利用者が m1.xlarge の仮想マシンを立ち上げた場合、このグラフ上では CPU リソースとして3個分を利用することとなる。なお、利用可能な CPU リソースは全部で192個であり、グラフ上では最も上にあたる。

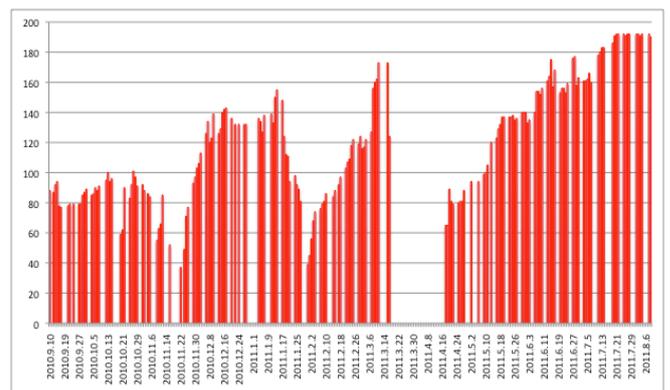


図 4.1 CPU リソースの利用状況

CPU リソースの利用状況のグラフの中で、特徴的な事象は次の通りである。

- (1) 2010年9月の計測開始から2011年3月の6ヶ月間は、変動はあるものの CPU リソースの約半数が利用された状態にある。
- (2) 2011年7月以降は、ほぼ全ての CPU リソース(90%以上)を使い切った状態にある。
- (3) クラウド停止後の再開時には、CPU リソー

スが急激に増加する。

(1) はクラウド導入期であり、利用者がいろいろな設定を試した時期であることが考えられる。利用数も徐々に広がっていることが分かる。

(2) については、東日本大震災とそれに起因する電力危機の影響が大きい。まず、計画停電への対応およびその後の節電のため、研究開発用クラウドについても計画停止を行った。

(2011/3/18-2011/4/17)運用再開後は、順調に利用が増加しており、CPU リソースの利用数も増加している。これは、節電のための集約サーバとして利用され始めたためである。

(3) については、計画停止後に直ぐに利用を開始する必要のあるユーザが仮想マシンを一斉に起動するためである。社内向けとはいえ、情報共有用のサーバなどなるべくサービス停止時間を短くしたいニーズがある。このため、サービス開始後には、研究開発用クラウドの負荷が通常時よりも上がる傾向にある。クラウド基盤の設計時には、仮想マシンの同時起動などの設計はこの数値を元にする必要があるとの知見が得られた。

4.2.2 利用者グループごとの総 CPU 時間

利用者のクラウドの利用形態を分析するため、仮想マシンの総 CPU 時間を分析した。利用実態のヒアリングから、利用者グループ単位で同じ仮想マシンを立ち上げている事例があるため、本分析は、利用者個人単位ではなく、利用者グループごとの分析とした。

分析対象として、4.2.1 の分析と同じ期間内で一度以上仮想マシンを起動した利用者グループとした。対象となった利用者グループは 89 組であった。

図 4.2 に利用者グループごとの仮想マシンの総 CPU 時間の頻度分布および累積パーセンテージを示す。

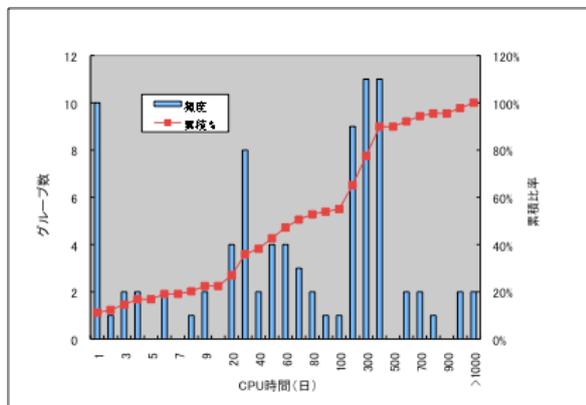


図 4.2 利用者グループごとの仮想マシンの総 CPU 時間の分布

仮想マシンの総 CPU 時間からは、ユーザを以

下の 3 つのカテゴリに分類することが可能であるとえられる。

(1) ライトユーザ (総 CPU 時間が 10 日未満)

CPU リソースの利用が少ないことから、このカテゴリの利用者はソフトウェアやクラウド基盤そのものの使い勝手等を試すことが目的と想定される。全体の 30%程度の利用者がこのカテゴリに含まれる。

(2) ミドルユーザ (総 CPU 時間が 10 日以上 100 日未満)

全体の 30%のユーザがこのカテゴリに含まれる。想定利用形態として、仮想マシンを立ち上げたままの場合、3 ヶ月以下の利用になる。業務時間中だけ仮想マシンを立ち上げて利用した場合、CPU 時間は一ヶ月で 160 時間 (6.6 日) になるため、2 ヶ月から 15 ヶ月利用した場合もこのカテゴリに含まれる。

総 CPU 時間だけでは、利用用途の想定は難しいため、その利用実態についてはより詳細な分析を行うことが必要であると考えられる。

(3) ヘビーユーザ (総 CPU 時間が 100 日以上)

このカテゴリはリソースを多量に消費しているヘビーユーザと位置づけられ、全体の 40%のユーザが含まれる。継続的にサーバを立ち上げて情報共有用のサーバとして利用している場合や、同時に多くのサーバを立ち上げて、計算用途に利用しているなどの用途が想定される。

いずれのカテゴリに関しても、ヒアリングなどによって内容の確認が必要であると考えられる。

4.2.3 仮想マシンタイプごとの利用状況

4.2.1 の分析期間中に起動された仮想マシンの総数は 2772 であった。仮想マシンタイプの利用状況を図 4.3 に示す。

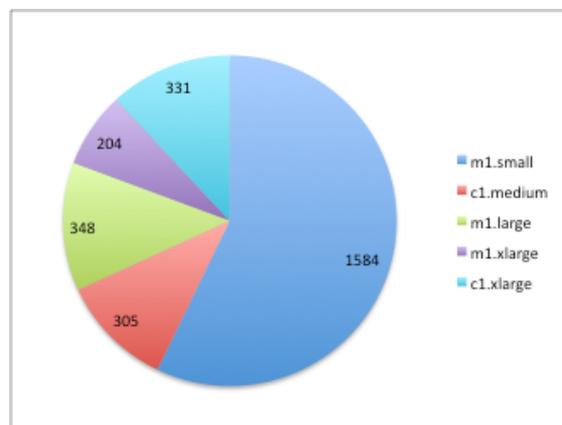


図 4.3 仮想マシンタイプと利用状況

利用された仮想マシンの 6 割は最も小さな m1.small であった。その他のマシンタイプは、ほぼ同程度の利用頻度である。

大きなマシンの利用が比較的少ないことから、必要最小限の仮想マシンを選択して利用していることが分かる。

5. 応用事例

本章では、研究開発用のクラウドの利用イメージを説明するため、特徴的な応用事例を取り上げてサービスの概要を説明する。

情報共有用としてはホームページ用に WEB サーバを利用する事例が多いが、サービス開始の初期から社内 Twitter が利用されている。

Twitter⁵⁾は 140 文字以内で「今、していること」をつぶやく (Tweet する) ことにより、不特定多数の人に対して今自分がしていることを共有することができるサービスである。

社内 Twitter は、社内の情報をリアルタイムで共有する目的等で利用されている。社内の有志が立ち上げたサービスで、社内の正式な情報システムという位置づけではない。名称として社内 Twitter という用語を使っているが、Twitter のサービスとは全く関係のない。投稿された情報も社内限定で共有されるため、どちらかと言えば、知り合い同士の会話メディアとして使われているようである。有志が立ち上げた、社内のソーシャルメディアという意味では、社内 SNS である Nexti⁶⁾に近い使われ方であると考えられる。

利用形態も様々で、日常の情報やカジュアルな情報の交換やプレゼンスの確認などのほかに、社内幹部のつぶやきなども投稿されて話題になった。

図 5.1 に社内 Twitter の画面例を示す。



図 5.1 社内 Twitter の画面例

6. まとめと今後の課題

本論文では、社内のソフトウェア開発や情報共有などの業務のためにセルフサービス型のプライベートクラウドを活用することを提案し、その利用仮説の検証を実施した。

筆者らは研究開発用のクラウドを実際に構築し、1年半の期間でその運用を行うことで評価を行った。アンケートの分析の結果、半数の利用者はソフトウェアの試行のための環境としてうまく活用していることが分かった。また、東日本大震災後の電力危機への対応などの影響で、研究開発用のクラウドの利用が広がり、利用率が 90%以上となっている。

今回の分析はクラウド基盤から観察可能な範囲のログを利用しており、利用内容まで踏み込まない範囲とした。今後の課題として、利用者により詳細なアンケートを実施し、その利用の詳細を分析することが必要になると考える。

参考文献

- 1) Peter Mell and Tim Grance, The NIST Definition of Cloud Computing, National Institute of Standards and Technology (2009) Volume 53, Issue 6, Pages 50
- 2) Amazon Web Services, Amazon Elastic Computing Cloud (Amazon EC2), <http://aws.amazon.com/jp/ec2/>
- 3) Eucalyptus Systems, Eucalyptus Open Source, <http://open.eucalyptus.com/>
- 4) Apache Hadoop Project, Hadoop, <http://hadoop.apache.org/>
- 5) Twitter 社, Twitter, <http://twitter.com/>
- 6) Nexti 運営メンバー有志, NTT データ流ソーシャルテクノロジー, リックテレコム, 2010 年 2 月

※ 記載されている会社名、商品名、又はサービス名は、各社の商標又は登録商標です。