

# Academic Baremetal Cloud の実現

長久 勝<sup>1)</sup>, 政谷 好伸<sup>1)</sup>, 横山 重俊<sup>2)</sup>, 谷沢 智史<sup>3)</sup>, 中川 晋吾<sup>4)</sup> 合田 憲人<sup>1)</sup>

- 1) 国立情報学研究所 クラウド基盤研究開発センター
- 2) 群馬大学 総合情報メディアセンター
- 3) 株式会社ボイスリサーチ
- 4) 有限会社カラビナシステムズ

mnagaku@nii.ac.jp

## Academic Baremetal Cloud

Masaru Nagaku<sup>1)</sup>, Yoshinobu Masatani<sup>1)</sup>, Shigetoshi Yokoyama<sup>2)</sup>, Satoshi Yazawa<sup>3)</sup>,  
Shingo Nakagawa<sup>4)</sup>, Kento Aida<sup>1)</sup>

- 1) Center for Cloud Research and Development, National Institute of Informatics
- 2) Library and Information Technology Center, Gunma University
- 3) Voice Research Inc.
- 4) Carabiner Systems, Inc.

### 概要

国立情報学研究所（以下、NII）では、2009年度以降、4世代のプライベートクラウド基盤を構築・運用してきた。この間、一貫して、NIIの教職員が主導的に関わり、経験を積み重ねてきた。最新の第4世代のプライベートクラウド基盤「Academic Baremetal Cloud（以下、ABC）」は、構築作業自体の再現性、構成拡張の容易性を、その特徴の一つとしており、最小限の互換性を満たすハードウェア機器が揃えば、短時間で基盤を配備可能、構成変更可能となっている。本稿では、NIIの10年の経験から生まれた「ABC」について、歴史的経緯も踏まえながら紹介する。

## 1 はじめに

NIIでは、2009年度以降、10年にわたり、「edubase Cloud」「Academic Community Cloud（以下、ACC。別の呼称として"gunnii"とする場合もあるため、文献にあたる場合には注意が必要）」「Academic Inter Cloud（以下、AIC）」「ABC」と、4世代のプライベートクラウド基盤を構築・運用してきた。この間、一貫して、NIIの教職員が主導的に関わり、運用や活用の経験を積み重ね、次世代の設計に反映し、より先進的で使いやすいクラウドサービスの実現を目指してきた。本稿では、現在の最新版である第4世代のプライベートクラウド基盤「ABC」について、以前の世代からの経緯も含め特徴を紹介する。

「ABC」は、その設計指針として、構築作業自体の再現性と、運用作業を容易化するための物理構成を追求している。物理構成を単純化する一方、複雑な管理ネットワーク構成はSDN化する、構築作業自体を機械的に再現可能とするなどの方策を

採用している。その結果、例えばクラスタに対するノードの追加や、同等なクラスタの増設を行う際に、構築ベンダに依頼することなく運用チームの業務として構築作業を「再現」する対応が可能である。一般的に、情報システムの構築には、固有のノウハウや煩雑な構築作業自体に外部発注による高いコストがかかることが多いが、「ABC」では、運用作業範囲内での対応を選択できるぐらい、容易に基盤を配備可能とした。これは、NII同様のユースケース、組織背景を持つ、多くの研究教育機関に、参考になるものと考えている。

## 2 「edubase Cloud」から「AIC」まで

以下に、「edubase Cloud」から「AIC」までのプライベートクラウド基盤の設計と機能について、時系列で述べる。あわせて、クラウド基盤の監視についても述べる。「edubase Cloud」から「AIC」までのプライベートクラウド基盤の機能とユースケースの関連については、「教育・研究クラウドサービスのためのパターンランゲージ」[1]にまと

めがある。

## 2.1 edubase Cloud

NII における最初のプライベートクラウド基盤となった「edubase Cloud」[2]は、Amazon EC2[3]と互換性を持った OSS「Eucalyptus」[4]を 15 セット配備したものであった。複数の研究グループの計算機実験が干渉しないように、基盤のレベルで分離しており、研究グループのユースケースに応じて、基盤単位でカスタマイズを可能としていた。例えば、当初構築時には、仮想マシンの OS として Linux しかサポートしていなかったが、FreeBSD や WindowsXP の利用要望があり、これに対応するため、特定の基盤にカスタマイズを施すなどした。また、仮想マシンを管理するクラウド基盤では、リソース効率を高めるため、物理的な CPU コア数よりも多くの仮想 CPU を設定するオーバーコミットを行うが、計算機実験でベンチマークを取った際の結果が安定しないなどの問題があるため、「edubase Cloud」ではオーバーコミットを行わなかった。

「edubase Cloud」は最初の取り組みであったこともあり、利用者から想定外のような要望を受けた。その中で最も大きかったのは、仮想マシンではなくベアメタル（物理マシン）を借りられるようにしてほしいという声であった。仮想マシンイメージを作成し、それを必要なだけオンデマンドで起動できるクラウドの特徴は、演習環境などの教育目的利用のユースケースには相性が良いものの、研究目的利用で計算機実験を行う際に、仮想化のオーバーヘッドを気にする利用者が多かった。

「edubase Cloud」は、2009 年度に構築が行われ、2010 年度から利用を開始、2014 年度末をもってサービスを終了した。

## 2.2 ACC

「edubase Cloud」の運用経験を踏まえて、より研究目的利用を意識して設計・構築されたのが「ACC」[5]である。仮想マシンではなくベアメタルの貸し出しを行え、研究グループの環境分離に SDN を採用した基盤である。OSS のクラウド基盤である「OpenStack」[6]を改修して採用した。

「edubase Cloud」の運用経験からは、研究グループの環境分離において、個々のリソース量を動的に変更可能にするべき、であるとか、利用開始までの準備時間を短縮するために、アプリケーション配備機能が必要、といった課題を設定し、

「ACC」で実現した。研究グループ個々のリソース量を動的に変更可能、については、ベアメタルの計算機間のネットワークを SDN で制御し、通信できる計算機グループを作り、グループへの計算機の追加・削除を柔軟に行えるようにした。利用開始までの準備時間の短縮、については、クラスタ管理フレームワーク「Dodai」[7]を用いて対応できるようにした。「Dodai」は、「OpenStack」管理下でベアメタルの IaaS を実現する「Dodai-compute」、更に複数のノードをクラスタとして扱い、アプリケーションテンプレートに従って、クラスタシステムを一括構築する「Dodai-deploy」から構成されている。これにより「Hadoop」[8]などの分散アプリケーションを配備可能とした。

「ACC」は、2011 年度に構築が行われ、2012 年度から利用を開始、2017 年度末をもってサービスを終了した。

## 2.3 AIC

「ACC」に続く「AIC」[9]では、基本設計は「ACC」を踏襲し、ネットワークの強化、SINET を活用した機能強化、運用に考慮した予備設備の保持を行った。SINET を活用した機能強化では、多拠点で L2 接続する SINET L2VPLS サービスを活用し、ベアメタルのサーバが L2VPLS のネットワーク内に払い出される仕組みを実現した。これにより、共同研究における L2VPLS サービス利用に、計算機資源の調達オプションが加わることとなり、より素早い研究環境構築が行えるようになった。また、運用の中で、基盤の改修を行う際に、事前に挙動を確認したい場合、予備設備がないと何が起こるか確認できない。これに対応するために、小規模な事前検証環境を保持し、本番環境構築後も、基盤改修時の挙動確認などに使えるようにした。

「AIC」は、2013 年度に構築が行われ、2014 年度から利用を開始、現在もサービス中で、2020 年度末をもってサービス終了の予定である。

## 2.4 クラウド基盤の監視

ここで、NII におけるクラウド基盤の監視についてもふれておく。オンプレミスの機器運用において機器故障への対応が必要となるが、クラウド基盤では、スイッチやサーバといった構成機器が数百台にも上るため、頻繁な故障対応を迫られる。また、同型の機器が大量にあるため、どの機器の

故障がどのユーザに影響を与えるのか、間違わないように把握する注意が必要となる。このため、「edubase Cloud」から一貫して、ラックの収容図をベースにクラウド基盤を視覚化し、故障情報や稼働情報を表示する方針を採っている。「ACC」以降は、「マップ」[10]と呼ぶシステムを運用している。「マップ」では、どのベアメタルのサーバが、どの研究グループのネットワークに払い出されているか、確認できるようにしており、故障対応時の影響範囲が把握しやすいように考慮している。サーバの負荷や、ネットワークの流量もグラフで確認できるため、システム不調時の初動調査などにも活用される。

### 3 Academic Baremetal Cloud

先に述べた経緯を踏まえ、NIIにおける第4世代プライベートクラウド基盤として設計・構築されたのが「ABC」である。「ABC」は、2016年度に構築が行われ、2017年度から利用を開始している。「OpenStack」ベースのクラウド基盤であることは、前の世代と同じであるが、「OpenStack」が製品として成熟し、Ironic という標準モジュールでベアメタル対応が行えるようになり、「Dodai」によるIaaS実現は不要となった。また、「ACC」「AIC」では計算機間のネットワークを「OpenStack」から自動的に制御するために、当時はAPIを備えたスイッチが一般的ではなかったため、ベンダー製品のSDNコントローラと、それに対応する「OpenFlow」スイッチ[11]を選定していたが、「ABC」ではVLANセグメンテーションなど単純な構成制御はRest APIを備える汎用スイッチを用いる設計へと変更している。

「OpenFlow」については利用目的を見直し、SDNコントローラもベンダー製品からOSSの「OpenDaylight」[12]に変更、カスタマイズすることで、管理・制御ネットワークのセキュリティ制御に特化した非対称な経路制御を実現している[13]。「OpenFlow」を利用して非対称な経路制御を実現することで、ベアメタルのサーバが悪意を持ってIPアドレスやMacアドレスを詐称したとしても、管理マシンとの通信しか行えない。従来は物理スイッチや物理セグメントを分割することでしか実現できなかった、管理・制御ネットワーク内のセキュリティ区分を、単一セグメント上で厳格かつ簡素に実現できるようになった。

また、「ABC」全体に関わる変更、「OpenStack」のバージョンアップなどの検証が必要となった場合、本番環境の管理・制御ネットワークをSDNの経路制御により論理分割することで、本番環境と等価な検証環境を切り出すことができる

クラウド基盤関連OSSが成熟期に入ったこととAPI制御できる物理装置が一般化したことで、OSSのみでも十分な品質のクラウド基盤が構築できるようになったと考えられ、NIIのクラウド運用に関する知見をより多くの機関に提供可能となったと言える。

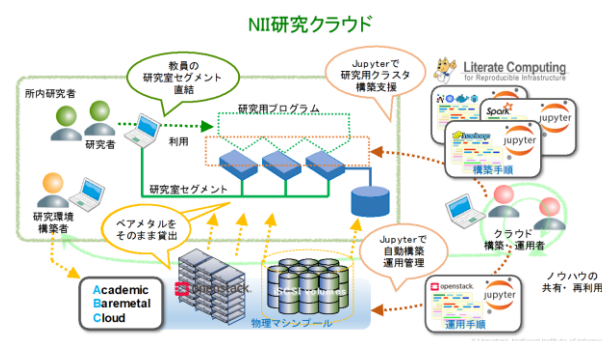
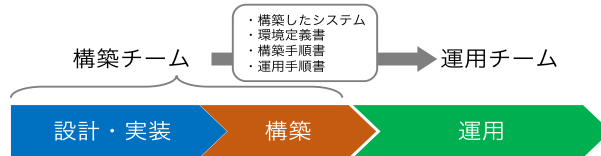


図1: 「ABC」の構成

edubase CloudからAICまで



ABC



図2: クラウド基盤の設計から運用までの流れ

### 4 「ABC」の構築方法

NIIクラウド運用チームでは、2014年度から、情報システムの構築・運用の方法論「Literate Computing for Reproducible Infrastructure（以下、LC4RI）」[14][15]を実践しており、「ABC」は、構築作業自体の再現性の向上と、運用作業を容易化するため、「LC4RI」ネイティブなクラウド基盤として開発した。「LC4RI」では、Notebookと呼ばれる実行可能な手順書を用いて情報システムの構築・運用を行うが、「ABC」の構築手順もNotebookにより記述した。従来と「ABC」のクラウド基盤構築・運用の違いを図2に示す。従来は構築チー

ムにクラウド基盤構築までを依頼し、運用チームが構築チームから引き継いだ資料に基づいて運用を行っていたのに対し、「ABC」では構築チームはクラウド基盤構築作業自体を行わず、構築・運用に関わる作業を Notebook として運用チームに対して引き継ぎ、クラウド基盤構築は運用チームが行う形式を採った。

運用チームがクラウド基盤構築を経験することで、構成機器の物理配線などの事前作業を済ませさえすれば、運用管理サーバ上で Notebook に従い作業することで、「ABC」を構築できる。実際、「ABC」の納品に関しては、クラウド基盤を初期化した後、Notebook で再構築できるかを確認している。また、NII では、「ABC」を 2 セット保持しており、同様の Notebook で構築を行っており、今後の追加調達においては、外部への作業発注を縮小ないしはゼロとし、運用チームでの構築を行う予定である。

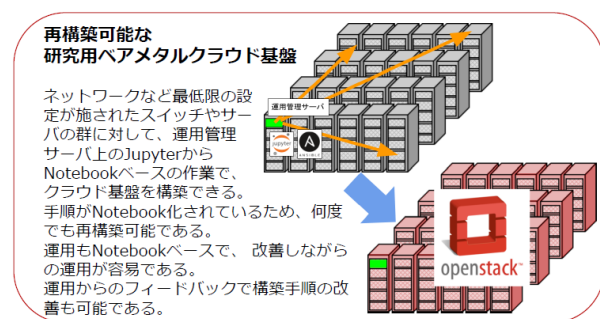


図 3：構築作業の概念図

## 5 「LC4RI」と、「ABC」の改善

NII のクラウド運用チームでは、故障対応や障害調査などを課題管理している。クラウド基盤の操作は Notebook を介して行っているため、故障対応や障害調査を行うと、作業内容と結果が Notebook に残る。このため、課題管理は、対応作業の Notebook を紐づけて行っている。

運用を行っていると、設計・構築時には分からなかったクラウド基盤の不具合を見つけることがあるが、「LC4RI」では、その対応作業が Notebook にコード化された状態で残っているため、それらを手掛かりに、クラウド基盤の改善を行うことができる。実際に、コンソール接続不具合の解消や、ストレージネットワークの運用手順の改善を、そうした手掛かりをもとに実施した。

## 6 さいごに

本稿では、NII でのクラウド運用を振り返り、その中で得られた知見から生まれた「ABC」について述べた。「ABC」は、OSS クラウド基盤の進化と、「LC4RI」に依拠しコンパクトな運用体制で、プライベートクラウド基盤としてサービスを行っている。これは、NII 同様のユースケース、組織背景を持つ、多くの研究教育機関に、参考になるものと考えている。

## 参考文献

- [1] 教育・研究クラウドサービスのためのパターンランゲージ  
<https://www.slideshare.net/mnagaku/ss-91096038>
- [2] 横山 他、edubase Cloud: Cloud Platform for Cloud Education、EduRex '12、17-20、2012
- [3] Amazon EC2 <https://aws.amazon.com/jp/ec2/>
- [4] Eucalyptus  
<https://github.com/eucalyptus/eucalyptus/wiki>
- [5] 横山 他、アカデミッククラウドアーキテクチャの提案と評価、情報処理学会誌、54、2、688-698、2013
- [6] OpenStack <https://www.openstack.org/>
- [7] Dodai <https://github.com/nii-cloud/dodai>
- [8] Hadoop <http://hadoop.apache.org/>
- [9] 合田 他、アカデミックインタークラウドの構想、信学技報、113、376、1-6、2014
- [10] 谷沢 他、広大な「マップ」を活用したクラウド監視システムの実運用への適用、信学技報、113、492、161-165、2014
- [11] OpenFlow  
<https://www.opennetworking.org/technical-communities/areas/specification/open-datapath/>
- [12] OpenDaylight <https://www.opendaylight.org/>
- [13] 三宅 他、オープンフローを活用したベアメタル・クラウド向けインタークラウド・ネットワーク機能の実現について、信学技報、115、72、SC2015-4、19-23、2015
- [14] 長久 他、Literate Computing for Reproducible Infrastructure による研究・教育環境、AXIES2017、2017
- [15] 政谷、NII での計算機環境の運用及び、Literate Computing for reproducible infrastructure について、HPC OPS 研究会、<https://www.slideshare.net/nobu758/hpcops-2018-0702pub-106398220>

※参考文献で URL が示されたものについては、全て、2018 年 8 月 31 日に確認