

WIDE クラウド WG 2014 年度活動報告

島 慶一

2014 年 12 月 10 日

1 はじめに

WIDE クラウドワーキンググループは、今後のクラウド技術の研究開発を推進するために 2010 年 1 月に設立された。複数の WIDE 組織間に渡って運用される広域連邦型クラウドシステムである WIDE クラウドシステムの運用と、それをを用いた研究開発を行っている。

2014 年度の主な活動は以下の通りである。

- OpenStack ワークショップ (第 2 章)
- 仮想計算機用ストレージシステム技術 (第 3 章)

以降、それぞれの活動の詳細を報告する。

2 OpenStack ワークショップ

2014 年春合宿研究会にて、OpenStack をローカル環境へインストール、動作させることを目的としたワークショップを開催した。

OpenStack は半年に一度のタイミングでリリースされている。2014 年春の時点で利用できる公式リリースは Havana であり、ワークショップも Havana を対象とした。OpenStack の導入はハードルが高いことで知られており、なかなか動作させるまで至らないという声を聞くことも多い。このワークショップでは、自分の仮想環境に OpenStack をクリーンインストールすることを目標に、合宿期間中に有志メンバーを集めて実施した。

OpenStack のインストール手順は、公式ドキュメントとしても公開されており、リリース時期に合わせた形で最新のインストール手順が参照できる¹。ただし、このドキュメントも完全ではなく、手順通りに進めて

もうまくいかなかったり、必要な手順が抜けていることもあり、OpenStack と関連した技術研究開発を実施している研究者、開発者にとってノウハウの共有は重要である。また、OpenStack が半年に一度リリースされるということもあり、一度獲得したノウハウは短期間で陳腐化してしまう傾向がある。その意味でも手数を増やして知識を共有することに大きな意味があった。

ワークショップで使った資料を付属 USB に添付する。ただし、対象バージョンが Havana (報告書発行時点の最新版から数えると 2 つ前のバージョン)であることを考慮すると、その資料自体の価値は大きくない。参考程度に参照していただきたい。

OpenStack の認知度が上がり、OpenStack が公式にサポートするイベント、OpenStack ユーザー会、有志による勉強会も多く目にするようになった。WIDE 内でのワークショップ開催の意味も薄れつつあるが、要望があれば研究会などのタイミングで最新の OpenStack をフォローする時間を設けるなどしていきたい。

3 実データの分散配置場所を柔軟に制御できる仮想計算機用ストレージシステムの提案

3.1 仮想計算機用ストレージの必要性

仮想化技術の著しい発展により、実用に耐える仮想計算機の運用が可能になった。ネットワーク上で提供されるサービスにおいても仮想機材を用いて構成する場合が増えている。仮想化により、サービスの迅速な展開や高負荷時の緊急設備拡張、また利用減に伴う設備縮退など、物理計算機では手間がかかる運用が比較的容易に実現可能になった。今後は、仮想資源の配置や再配置を柔軟に実現する仕組みが重要となる。

¹<http://docs.openstack.org/>

CPU やメモリ資源の再配置技術はすでに実用化され、オープンソース、商用を問わず多くのハイパーバイザーに実装されている [1, 2, 3, 4]。また、SDN 技術を用いたネットワーク資源の再配置技術の実用化も近い。残る問題はストレージ資源の再配置である。

データセンターは効率を求めて大規模化している。しかし、我が国のような国土が狭い環境で諸外国にみられる規模のデータセンターを建設することは困難である。そのため、地理的に独立した中規模のデータセンターを仮想的に大規模データセンターとみなす分散型データセンター運用が検討されている。分散したデータセンターを横断して基盤サービスを構成する場合、地理条件や遅延を考慮して仮想資源を効率的に配置運用する技術が重要となる。

本書では、仮想計算機基盤運用者の視点に立ち、分散型データセンターにおいて仮想ディスクの実データの配置を柔軟に制御できるストレージシステム「UKAI」[5]を解説する。UKAIシステムを用いることで、仮想ディスクイメージを構成する実データの冗長性や配置場所を、運用者が自在に制御できるようになる。

3.2 UKAI の設計と実装

UKAI の仮想ディスクは、実データへのポイントの集合として構成される。仮想ディスクは複数の論理ブロックに分割され、それぞれのブロックに対応する実際のデータは地理的に分散された外部のストレージノードに保持される。ストレージノードは複数運用することができ、必要とされる冗長度に応じて複数のブロックの複製を保持できる (図 1)。仮想ディスクの構成操作は仮想計算機の運用とは独立に実行できるようになっており、運用者は仮想計算機を動作させたまま、仮想ディスクの実データをより適切な場所、例えば遅延の短い場所に移動させたり、地理的に離れた場所に複製を作ったりできる。既存の分散ストレージ [6, 7, 8] ではブロックの自律配置に主眼をおいているが、UKAI では運用者に配置決定の権限と責任が委ねられる。

図 2 に FUSE を用いた参照実装² のモジュール関連図を示す。仮想ディスクの構成を記録したメタデータは ZooKeeper クラスタで保持され、仮想ディスクが利用される際にハイパーバイザーから動的に参照される。KVM などに代表される仮想計算機の実行プログ

²<https://github.com/keiichishima/ukai/>

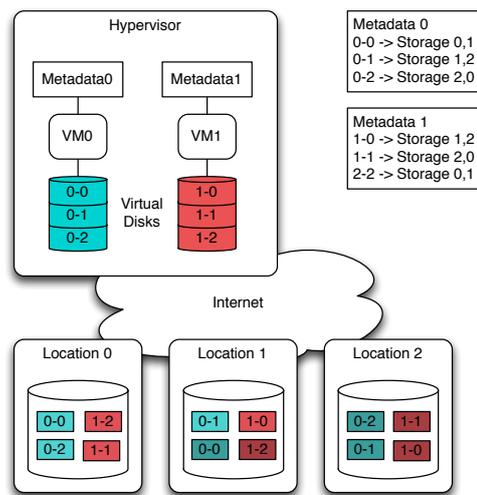


図 1: UKAI 仮想ディスクの構成図

ラム、および OpenStack などのクラウドコントローラは、FUSE 経由で提供されるファイルを仮想ディスクの単位として操作する。

3.3 OpenStack 対応

基盤システムの運用はその大部分がシステム化されており、できる限り人の手がかからないような構造を目指して設計されてきた。近年、クラウド化の流れが進んだことにより、これまで想定されていなかった規模での計算機資源を取り扱う必要がでてきている。クラウド基盤運用のための統合システムとしては OpenStack [9] や CloudStack [10] が広く認知されており、特に OpenStack は多くのベンダー、開発者の注目を集めている。

OpenStack はシステムを構成する機能がモジュール化されており、外部サービスとの連携や機能拡張がある程度考慮されている。本活動でも、UKAI を OpenStack のブロックストレージとして利用するための拡張を実装し、その動作を確認した。UKAI 拡張されたモジュールは github にて公開している³。

3.4 規模拡張性の検証に向けて

UKAI は、クラウド基盤運用事業者に向けた技術であり、数万台規模の仮想計算機のディスクイメージを

³Nova: <https://github.com/keiichishima/nova/>, Cinder: <https://github.com/keiichishima/cinder/>

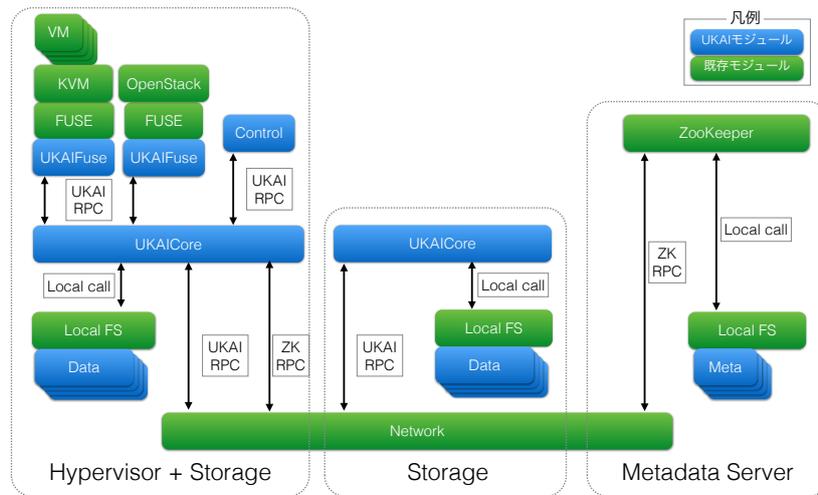


図 2: UKAI モジュール関連図

取り扱うことを目標としている。規模拡張性の検証には StarBED などの実環境テストベッドを利用する。今回、予備実験として 16 台の実計算機上にそれぞれ 10 台仮想計算機を構築し、都合 160 台の仮想計算機を動作させる環境を構築した。この環境で動作したものの、仮想ディスクイメージのメタデータ操作に伴う問題が発生することが確認されたため、その対応に向けた作業を予定している。実計算機の数をもさらに増やし、大幅に規模を拡張した大規模実験は来年実施予定である。

3.5 まとめ

仮想基盤の活用が進むにつれ、その効率的な運用のために仮想資源の柔軟な配置制御が重要な要素となる。本書では、仮想計算機基盤運用者の視点に立ち、分散型データセンターでの基盤運用に必要な地理位置情報を考慮できるストレージシステムの提案を行った。また、その参照実装を概説するとともに、広く普及しつつある OpenStack との統合動作を紹介した。

4 まとめ

今年度は OpenStack のノウハウ共有、また昨年から引き続いての分散仮想ディスクシステムの研究開発を実施した。なお、クラウドワーキンググループの関連活動として、Network Function Virtualization (NFV) の機能や動作検証を実施する目的でコンソーシアムの

運用を開始している。こちらの活動に関しては、別途特集報告としてまとめられているので、そちらを参照して欲しい。

参考文献

- [1] R.A. Harper, A.N. Aliguori, and M.D. Day. KVM: Kernel-based Virtual Machine, 2014. <http://www.linux-kvm.org/>.
- [2] R.A. Harper, A.N. Aliguori, and M.D. Day. KVM: The Linux Virtual Machine Monitor. In *Proceedings of the Linux Symposium*, pages 225–230, 2007.
- [3] Citrix Systems, Inc., 2014. <http://www.xen.org/>.
- [4] Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, and Andrew Warfield. Xen and the art of virtualization. In *SOSP'03: Proceedings of the nineteenth ACM symposium on Operating systems principles*, pages 164–177. ACM, 2003.
- [5] Keiichi Shima. UKAI: Centrally Controllable Distributed Local Storage for Virtual Machine Disk Images. In *Proceedings of Globecom 2013*

Workshop - Cloud Computing Systems, Networks, and Applications (CCSNA), pages 432–438, December 2013.

- [6] Sage A. Weil, Andrew W. Leung, Scott A. Brandt, and Carlos Maltzahn. RADOS: A Scalable, Reliable Storage Service for Petabyte-scale Storage Clusters. In *Proceedings of the 2th international Petascale Data Storage Workshop (PDSW'07)*, pages 35–44, November 2007.
- [7] Gluster Inc. Gluster File System Architecture. Technical report, Gluster Inc., 2010.
- [8] Kazutaka Morita. Sheepdog: Distributed Storage System for QEMU. KVM Forum 2010, August 2010.
- [9] The OpenStack Foundation. OpenStack, 2014. <http://www.openstack.org/>.
- [10] The Apache Software Foundation. Apache CloudStack™, 2014. <http://cloudstack.apache.org/>.