2-2　大規模情報処理基盤

Large Scale Information Processing Systems

菊地能直　夏目貴史

Abstract

多様な産業の付加価値創造の支援や産業間連携等による新ビジネス創出に向け，今後の情報処理基盤には，爆発するデータ量に対応する拡張性に加え，アプリケーションやデータの自在な共有並びにステークホルダとの多様なコラボレーションを可能とする柔軟性が求められる．これを実現する今後のソフトウェアプラットホームの在り方，ICT 資源を柔軟に運用する技術として普及・標準化が進むクラウドコンピューティング技術及びそれを支えるネットワーク技術の今後の進化の方向性，主な技術課題等を議論する．

キーワード：クラウドコンピューティング，OpenStack，SDN，オーバレイネットワーク

1. 大規模情報処理基盤を実現するクラウドコンピューティング技術

1.1　背景

近年，ソーシャルメディアやスマートフォンの普及及び IoT の拡大により，大量のデータを処理する基盤が必要とされている．その基盤，つまり大規模情報処理基盤としてクラウドコンピューティングサービスが用いられる．

それは以下のような理由による．

① データを大量に蓄積できる．
② 処理能力を柔軟に拡張できる．
③ 処理を行わない場合はリソースを開放することができ，リソースを効率的に利用することができる．

また，大量のデータを蓄積でき，そのデータの利用・処理を間近で行うことが可能であることもクラウドコンピューティングサービスが大規模情報処理基盤に用いられる理由である．

1.2　大規模情報処理基盤としての OpenStack

クラウドコンピューティングサービスを構築するためのオープンソースソフトウェア（以下「OSS」と呼ぶ）に OpenStack[(1)] がある．OpenStack では以下のようなコンポーネントや機能により，大規模情報処理基盤を実現する．

（1） Heat[(2)]

オーケストレーションサービスである．リソース（仮想マシン，仮想ネットワーク等）のライフサイクルの全体を管理する．オートスケーリングという機能で，負荷やデータ量等に応じて仮想マシン等のリソースを自動的にスケールさせる（増やす）ことができる．

（2） Sahara[(3)]

Apache Hadoop[(4)] や Apache Spark[(5)] 等のクラスタ管理を行い，それらのクラスタにおいてジョブを実行できる．

菊地能直　正員　日本電信電話株式会社ソフトウェアイノベーションセンタ
E-mail　kikuchi.yoshinao@lab.ntt.co.jp
夏目貴史　正員　日本電信電話株式会社ソフトウェアイノベーションセンタ
E-mail　natsume.takashi@lab.ntt.co.jp
Yoshinao KIKUCHI and Takashi NATSUME, Members (Software Innovation Center, NIPPON TELEGRAPH AND TELEPHONE CORPORATION, Musashino-shi, 180-8585 Japan).
電子情報通信学会誌　Vol.100 No.8 pp.767-770 2017 年 8 月

（3） Senlin[6]
クラスタリングサービスを提供する．ポリシーによりスケーリングやロードバランシングを行うことができる．

（4） Magnum[7]
Docker Swarm[8]，Kubernetes[9]や Apache Mesos[10]を利用するコンテナ管理サービスである．

（5） Storlets[11]
オブジェクトストレージ（後述する Swift[12]）を格納しているノードにおいて安全に独立した分散処理を行うコンポーネントである．また，Docker[13]コンテナを利用する．

また，以下の大容量のデータを格納するためのサービスも提供されている．

（1） Swift
いわゆるオブジェクトストレージサービスを提供する．一つのオブジェクトとして格納できるサイズは，デフォルトでは 5 GByte が上限であるが，分割してアップロードしたファイルを一つのファイルとしてダウンロードできる機能がある．また，複数のレプリカを持つことができ，信頼性も向上させている．

（2） Trove[14]
データベースをサービスとして提供する（Database as a Service）．データベースにはリレーショナルデータベースだけでなく，非リレーショナルデータベースも含まれる．

上述したコンポーネントや機能を活用して，大規模情報処理基盤を実現できる．

1.3 OpenStack における大規模処理基盤を支える仮想ネットワーク機能

OpenStack の第 1.2 節にあるコンポーネントで処理を行うには，柔軟に拡張を行うことができる仮想ネットワーク機能が必須である．そのため，OpenStack には以下のコンポーネントがあり，スケーラブルで柔軟な大規模情報処理基盤を支えている．

（1） Neutron[15]
仮想ネットワーク機能を提供する．仮想ネットワーク機能とは具体的には L2 ネットワーク（ブリッジ），L3 ネットワーク（ルータ）等である．また，ロードバランサ，ファイヤウォールもある．（ただし，別のコンポーネントである．）

（2） Kuryr[16]
コンテナのためのネットワーキングサービスを提供する．コンテナサービスから Kuryr を経由して Neutron の API を呼び出すことができるようにする．

1.4 OpenStack の技術課題

OpenStack には大規模情報処理基盤を実現する上での技術課題として以下の課題が存在する．

（1） スケーラビリティ
管理のために用いるデータベースやメッセージングミドルウェア等がボトルネックとなり，スケーラビリティに制限がある場合やスケーラビリティを上げると機能制限がある場合がある．

（2） 自律的な管理
スケーラビリティが上がり，より多くのリソースを利用するようになると，利用者の管理の労力が大きくなる．利用者の管理の労力を小さくする必要がある．

1.5 大規模情報処理基盤を実現するクラウドコンピューティング技術の将来

1.4 で述べた課題を解決するため，現在 OpenStack では以下のような方向で開発が進められている．例えば，以下のような機能やコンポーネントの開発が続けられている．

（1） Cell v2[17]
仮想マシンの管理サービスを提供する Nova[18]において，よりスケーラブルに仮想マシンを配置できる機能の開発が行われている．物理ホストを複数のセルに分けて，そのセルごとに Nova が利用する管理情報格納用のデータベースや Nova のプロセスが連携するためのメッセージングミドルウェアを分けることにより，負荷を分散させてスケーラビリティを確保する．

（2） Congress[19]
ポリシーを定義し，そのポリシーに適合しているかチェックして，適合していなければ通知を行う，または是正のためのアクションを取ることができるコンポーネントである．

現在はルールベースによる管理機能が取り入れられて開発が進められているが，将来的には人工知能（機械学習など）によって，より人間の手を介さない管理を実現する方向で進化するものと考えられる（図 1）．

（夏目貴史）

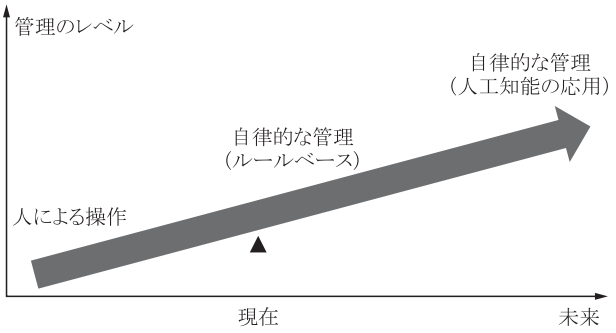

図1　OpenStack の今後の進化の方向性

2. 大規模情報処理基盤を支えるネットワーク技術

大規模情報処理基盤としてのクラウドコンピューティングサービスを支えるネットワークには，処理データ量の増大に伴う広帯域化が必要なことや，耐故障性の向上，ネットワーク機器の増設が容易であること等の様々な要件が求められる．本稿では誌面の都合上，利用者の観点から仮想マシンの物理サーバ間の移動に伴う柔軟なネットワークの構築，運用者の観点から帯域利用の効率性や電力消費量に影響しデータセンターの構築・運用コストに直結するネットワークトポロジーの2点について動向をまとめ，それを踏まえ今後の進化の方向性を示す．

2.1　柔軟なネットワークの構築に関する技術の動向

クラウドコンピューティングサービスでは，物理サーバの故障時や電力削減向けに，仮想マシンを別の物理サーバに移動させることがある．この際に，仮想マシンの移動に合わせてネットワークを短時間で切り換えることがサービスの断時間短縮の観点から求められる．また，切換に伴う不具合発生の防止が求められる．この迅速かつ間違いのない作業という観点から，ネットワーク構築作業の自動化に対するニーズが高まっている．

Software Defined Network（SDN）は，スイッチやルータ等のネットワーク機器をソフトウェアで動的に制御する技術であり，仮想マシンの移動に伴うネットワーク構築作業の自動化への適用も可能である．

2.1.1　OpenFlow

OpenFlow は，Open Networking Foundation（ONF）[20]で標準化されたプロトコルであり，SDN を実現するための実装の一つとして利用されている．対応したネットワーク機器や OSS も増えており，OpenFlow を利用したデータセンターのサービスも開始されている．OpenFlow ではコントローラがネットワーク機器を集中的に管理し，ヘッダ情報の組合せによりネットワーク経路を設定する．仮想マシンの移動の際には，対応するヘッダ情報を基にソフトウェアで動的にネットワーク経路を切り換えることで，ネットワーク構築作業の自動化が可能となる．

2.1.2　オーバレイネットワーク

OpenFlow には，既存のネットワーク機器の置き換えが必要になり導入へのハードルが高いという課題がある．オーバレイネットワークは，VLAN，VXLAN や NVGRE 等のトンネリング技術を利用して，物理的なネットワーク機器を意識せずに仮想的なネットワークを構築する技術であり，既存のネットワーク機器をそのまま利用することが可能である．これらトンネリング技術を API 経由で制御可能な製品も増えてきており，API を活用したネットワーク機器の制御の取組みも増えてきている．

現状の SDN の適用例としては，物理サーバや利用者の端末の近傍のネットワーク機器のみを OpenFlow 対応の機器に置き換えヘッダ情報を基に細かなネットワーク経路の制御を行い，OpenFlow 対応の機器間はオーバレイネットワークで接続し，それらの機器を API で制御する取組み等が見られる．

2.2　ネットワークトポロジーに関する技術の動向

近年，データセンターの大規模化が進んでおり，帯域利用の効率化や電力消費量削減，それらを踏まえたデータセンターの構築・運用コストの削減が求められている．ここでは，それらに大きく影響を与えるデータセンターのネットワークトポロジーの動向について述べる．

2.2.1　3層トポロジー

3層トポロジーは，従来，データセンターで用いられてきたトポロジーである．エッジ層は物理サーバを集約し，複数のエッジ層は上位のアグリゲーション層で集約される．アグリゲーション層は，上位のコア層で集約され3層でトポロジーを構築する．このトポロジーは，物理サーバ数が増えてきた際に，十分な帯域の確保が難しくなることに加え，コア層で使われるコアスイッチは高い処理能力が必要なため高価なスイッチが必要となり構築コストの面でも課題があることが知られている[21]．

2.2.2　Fat トリートポロジー

Fat トリートポロジーは，比較的安価なスイッチのみを用いて，十分な帯域を確保できるトポロジーである．物理サーバは，リーフスイッチに接続され，リーフスイッチは全てのスピンスイッチに接続されている．スピンスイッチ同士は互いに接続されず並列な構成となる．異なるリーフスイッチに接続された物理サーバ間の通信は，スピンスイッチ経由となる．この構成により3層トポロジーでは不足しがちであった帯域の有効活用が可能

となっている．大規模なデータセンターでもFatトリートポロジーを踏まえ，改良を加えたトポロジーの適用が報告[22]されている．

2.2.3 DCell

物理サーバのNetwork Interface Card（NIC）を複数用いて，物理サーバ同士を接続し，大規模データセンターを構築するトポロジーである[23]．このトポロジーは，高価なコアスイッチなしでスケールアップが可能なメリットがある．しかし，トラヒック量が増えた際には，物理サーバ間の通信がボトルネックになりやすい課題があることが知られている．

2.2.4 BCube

BCube[24]はDCellと同様に物理サーバの複数のNICを利用する構成であるが，物理サーバ同士の接続にスイッチを用いる構成であり，DCellよりも帯域の有効活用が可能なトポロジーとなっている．

以上が基本的なトポロジーとなるが，アプリケーションの要件や構築・運用コスト等を勘案し，データセンターに適用するネットワークトポロジーを選択する必要がある．

2.3 今後の進化の方向性

以上，クラウドコンピューティングサービスを支えるネットワークについて，柔軟なネットワークの構築とトポロジーの観点から現在の動向をまとめたが，以下に今後の進化の方向性について示す．

クラウドコンピューティングサービスに関連する今後のトレンドとしては，IoT機器が収集したセンサ等の情報を，データセンターに集約し大量に処理するようなケースが考えられる．この際に，処理結果に基づきIoT機器を制御するような場合には，低遅延の処理が必要になることが想定され，ネットワークの観点では，従来よりも低遅延なネットワークが求められる．また，機械学習向けに利用者にGPUを提供することが必要となるが，必要な際に必要な量のGPUを利用者に提供可能なネットワークの構成が必要となる．これらの新しいトレンドからの要件に対応可能なネットワーク技術の確立が望まれる．

（菊地能直）

文　　献

（1） OpenStack Foundation, “OpenStack open source cloud computing software,” https://www.openstack.org/（2017年2月1日閲覧）
（2） OpenStack Foundation, “Heat-OpenStack,” https://wiki.openstack.org/wiki/Heat（2017年2月13日閲覧）
（3） OpenStack Foundation, “Sahara-OpenStack,” https://wiki.openstack.org/wiki/Sahara（2017年2月13日閲覧）
（4） OpenStack Foundation, “Welcome to apache hadoop!,” http://hadoop.apache.org/（2017年2月13日閲覧）
（5） OpenStack Foundation, “Apache spark-lightning-fast cluster computing,” http://spark.apache.org/（2017年2月13日閲覧）
（6） OpenStack Foundation, “Senlin-OpenStack,” https://wiki.openstack.org/wiki/Senlin（2017年2月13日閲覧）
（7） OpenStack Foundation, “Magnum-OpenStack,” https://wiki.openstack.org/wiki/Magnum（2017年2月13日閲覧）
（8） OpenStack Foundation, “Docker swarm | docker,” https://www.docker.com/products/docker-swarm（2017年2月13日閲覧）
（9） OpenStack Foundation, “Kubernetes-production-grade container orchestration,” https://kubernetes.io/（2017年2月13日閲覧）
（10） OpenStack Foundation, “Apache mesos,” http://mesos.apache.org/（2017年2月13日閲覧）
（11） OpenStack Foundation, “Storlets-OpenStack,” https://wiki.openstack.org/wiki/Storlets（2017年2月13日閲覧）
（12） OpenStack Foundation, “Swift-OpenStack,” https://wiki.openstack.org/wiki/Swift（2017年2月13日閲覧）
（13） OpenStack Foundation, “Docker-build, ship, and run any app, anywhere,” https://www.docker.com/（2017年2月13日閲覧）
（14） OpenStack Foundation, “Trove-OpenStack,” https://wiki.openstack.org/wiki/Trove（2017年2月13日閲覧）
（15） OpenStack Foundation, “Neutron-OpenStack,”https://wiki.openstack.org/wiki/Neutron（2017年2月13日閲覧）
（16） OpenStack Foundation, “Kuryr-OpenStack,” https://wiki.openstack.org/wiki/Kuryr（2017年2月13日閲覧）
（17） OpenStack Foundation, “Nova-cells-v2-OpenStack,” https://wiki.openstack.org/wiki/Nova-Cells-v2（2017年2月13日閲覧）
（18） OpenStack Foundation, “Nova-OpenStack,” https://wiki.openstack.org/wiki/Nova（2017年2月13日閲覧）
（19） OpenStack Foundation, “Congress-OpenStack,” https://wiki.openstack.org/wiki/Congress（2017年2月13日閲覧）
（20） Open Network Foundation, http://www.opennetworking.org/
（21） M. Al-Fares, A. Loukissas, and A. Vahdat, “A scalable, commodity data center network architecture,” Proc. SIGCOMM, pp. 63-74, Seattle, Washington, USA, Aug. 2008.
（22） A. Singh, J. Ong, A. Agarwal, G. Anderson, A. Armistead, R. Bannon, S. Boving, G. Desai, B. Felderman, P. Germano, A. Kanagala, J. Provost, J. Simmons, E. Tanda, J. Wanderer, U. Holzle, S. Stuart, and A. Vahdat, “Jupiter rising : A decade of clos topologies and centralized control in Google’s datacenter network,” Proc. SIGCOMM, pp. 183-197, London, United Kingdom, Aug. 2015.
（23） C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu, “Dcell : A scalable and fault-tolerant network structure for data centers,” Proc. SIGCOMM, pp. 75-86, Seattle, Washington, USA, Aug. 2008.
（24） C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, C. Tian, Y. Zhang, and S. Lu, “Bcube : A high performance, server-centric network architecture for modular data centers,” Proc. SIGCOMM, pp. 63-74, Barcelona, Spain, Aug. 2009.

（平成29年2月28日受付）

菊地　能直（きくち　よしなお）（正員）

平11東大・工・物工卒．平13同大学院修士課程了．同年NTT入社．以来，標準化活動，クラウドサービス構築のグローバルプロジェクトマネジメント等を推進．現在は，ソフトウェアイノベーションセンタにて製造業向けIoT基盤に関する研究開発を推進中．

夏目　貴史（なつめ　たかし）（正員）

平9東大・工・電子情報卒．平11同大学院修士課程了．現在，NTTサービスイノベーション総合研究所ソフトウェアイノベーションセンタ勤務．OpenStackの開発及びコミュニティ活動，OpenStackを利用したクラウドコンピューティングシステムの研究・開発を行っている．