

理研データ科学基盤におけるシステム構築支援

検崎 博生¹⁾, 石井 学²⁾, 實本 英之¹⁾, 舂本 現¹⁾

1) 理化学研究所 情報統合本部

2) 株式会社ゲノムアナリティクスジャパン

hiroo.kenzaki@riken.jp

Support for system construction in RIKEN Data Science Infrastructure

Hiroo Kenzaki¹⁾, Manabu Ishii²⁾, Hideyuki Jitsumoto¹⁾, Gen Masumoto¹⁾

1) RIKEN Information R&D and Strategy Headquarters, RIKEN

2) Genome Analytics Japan Inc.

概要

理化学研究所では、研究データの保存・解析・公開のためにデータ科学基盤 HOKUSAI SailingShip を 2020 年 6 月から運用している。利用者はプライベート IaaS としてシステムを構築して利用することになるが、さまざまな研究用のシステムを簡単に構築できるように Ansible を用いたシステム構築支援を利用者に提供している。本稿では、データ科学基盤とその上でのシステム構築支援について紹介する。

1 はじめに

近年、データ処理の必要性の増大、アプリケーション実行環境の柔軟性が求められている。そこで、理化学研究所（以下、理研という）ではさまざまなデータの保存・処理・公開を行えるように、データ科学基盤 HOKUSAI SailingShip (HSS)^[1]を導入した。HSS は利用者毎にシステムを構築できるプライベート IaaS 環境であり、OpenStack を用いて計算リソース/共有ストレージ/ネットワークを仮想化することにより提供している。利用者は、用途に合わせた Virtual Machine (VM) やネットワークを構築することができ、共用ストレージも多様なプロトコルで利用できるようになっている。

HSS の大きな問題点は、利用者自身でシステム構築を行う必要があるため、利用のハードルが高いことである。対策としてあらかじめ必要な環境構築を行ったイメージを配布することが考えられるが、研究分野や用途はさまざまなものがあるので現実的ではない。

そこで、さまざまなシステムを構築できるように構成管理ツールを用いることが考えられる。構成管理ツールには Ansible、Puppet、Chef などがある。この中で Ansible はエージェントレスであり、python が使用可能で ssh が疎通すれば利用できるため、他の構成管理ツールに比べて非常にシンプルである。そこで本利用支援では Ansible を用いる

こととした。

Ansible ではサーバ構成やソフトウェア構成を YAML 形式の `playbook` ファイルに記述しておけば、システムを構築することが可能である。本システム構築支援では、多くの利用者に利用されるであろう各種機能の構成情報を記述した `playbook` を提供している。そのため、利用者は必要な機能 (role) を選択すれば、用途に合ったシステムを構築可能となっている。また、各種設定ファイルもあらかじめひな型などを用意しておき、利用者は最低限必要な情報を入力すれば、環境設定も含めてシステム構築が完了するようにしている。

本稿では、まず HSS の構成や利用方法について簡単に説明を行った後、システム構築支援の構成要素や利用方法について紹介し、用意した各機能の詳細を述べる。



図 1 HSS の外観

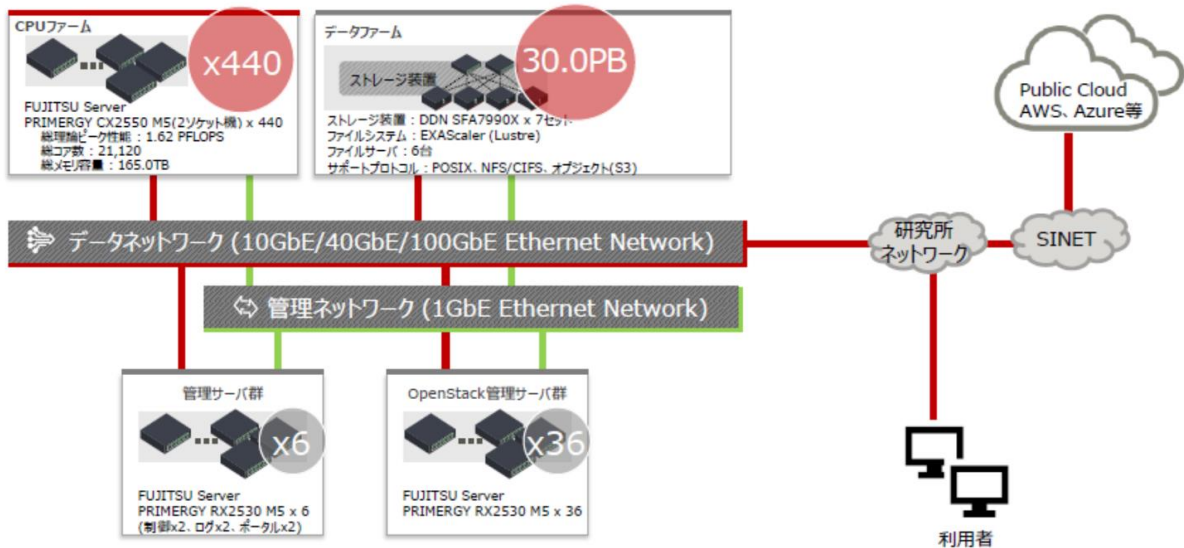


図 2 HSS の概念図

2 理研データ科学基盤 HSS の概要

理研では、導入時期を数年ずらした 2 系統のスーパーコンピュータを運用していたが、その 1 系統をデータ科学基盤として導入することとなった。HSS は富士通株式会社によって開発され、2020 年 6 月から運用を開始している (図 1)。HSS は CPU ファーム、データファーム、仮想化基盤からなり、利用者にプライベート IaaS 環境を提供している (図 2)。

2.1 HSS の利用方法

HSS の利用単位はプロジェクトであり、利用者はプロジェクトの申請やシステムを構築するプロジェクトメンバーと、構築されたシステムを利用するテナント利用者に分かれる。プロジェクトメンバーは基本的に理研の身分を持っている必要があるが、テナント利用者は各プロジェクトで管理するので理研外の利用者も利用することがある。

2.2 CPU ファーム

CPU ファームの 440 ノードで構成され、各ノードは CPU2 基 (Intel Xeon Platinum 8260) を搭載している。メモリ (DDR4-2933) は 384 GB、ローカルストレージ (SSD) は 1.92 TB、インターコネクトは 10GBASE-T x 2 という構成となっている。

CPU ファームは 220 ノードずつの 2 つのクラスタに分かれていて、片方は計算リソースを利用者が常時確保しているプール型であり、もう片方は計算リソースを利用者が空いているときだけ利用するオンデマンド型である。プロジェクト開始時にどちらのクラスタを使用するか選択する必要がある。

2.3 データファーム

データファームは 30PB の Lustre ファイルシステムで、NFS/CIFS/S3 ゲートウェイを持ち、多様なアクセスが可能となっている。

データファームは共有ストレージとして利用可能で、VM を作成した場合は Lustre ファイルシステムをマウントして利用可能である。理研研究情報管理サービス^[3]では VM を作成せずに、CIFS プロトコル経由で利用することもできる。共有ストレージはそれぞれの利用方法で領域は共通であるが、お互いのデータを参照することはできない。

2.4 仮想化基盤

仮想化基盤としては OpenStack の商用ディストリビューションである Mirantis Cloud Platform (MCP) を使い、データファーム、CPU ファーム、ネットワークを仮想化し、利用者毎にテナント (システムインフラ) を構築できるようにしている。

VM を構築する際のイメージは Ubuntu や Rocky Linux などいくつか用意しており、共有ストレージの利用の有無で 2 種類のイメージに分かれている。共有ストレージはプロジェクト毎に領域が分かれていて、他のプロジェクトからアクセスすることはできない。また、/home 領域を Lustre マウントするので、複数 VM から利用する場合は UID や GID を揃える必要がある。

ネットワークはプロジェクト内通信のためのプライベートなプロジェクトネットワークと共有ストレージと通信するためのストレージネットワークの 2 つがある。プロジェクト外から通信するためには、理研のイントラアドレスを VM に Floating IP アドレス (FIP) として設定する必要がある。

ある。必要に応じて理研のグローバルアドレスを設定することもできるが、理研外から通信するためには、理研の監査を受ける必要がある。

2.5 HSS の利用手順

HSS の利用手順は以下のようになる。

1. オンライン申請システムでプロジェクトを作成し、プロジェクトメンバーを登録。
2. プロジェクトメンバーがオンライン申請システムでポイントを購入し、ポイントで計算リソースを確保。
3. プロジェクトメンバーが OpenStack 管理コンソールでテナントの構築（VM の作成やネットワークの設定など）を行い、VM にテナント利用者のアカウントを作成。
4. テナント利用者が、テナント内の VM や共有ストレージを利用。

なお、研究情報管理サービスを利用する場合は、共有ストレージを確保した後で、テナントを構築せずに、研究情報管理サービスの利用申し込みを行うとデータ置き場（ストレージインフラ）として利用できる。

3 システム構築支援について

ここではシステム構築支援の構成要素とシステム構築の流れを説明する。

3.1 システム構築支援の概要

共通利用イメージ（スナップショット）、構成管理ツール Ansible の playbook、各種スクリプトやドキュメントを提供している。利用者は共通利用イメージから VM を作成して、作成した VM 上で Ansible を用いてシステムを構築する。

3.2 構築できる機能概要

多くの利用者に必要なユーザ管理機能としては LDAP を採用した。Samba によるファイル共有機能や RDP を用いたリモートデスクトップも用意している。

計算機クラスタ環境を構築できるようにするために、ジョブ管理として slurm を提供し、その上で singularity でのジョブ実行をできるようにしている。

バイオインフォマティクス向けの基本的な解析環境として Jupyter と R studio server を用意した。もちろん、他の分野での解析環境としても利用することも出来る。

今回提供した機能を表 1 に示す。一部機能は依存関係があるが、基本的には必要な機能だけを利

用できる。

表 1：構築できる機能一覧

機能	詳細
ユーザ管理	LDAP
ファイルサーバ	Samba, NFS（非推奨）
リモートデスクトップ	RDP, VNC（非推奨）
ジョブ管理	slurm, singularity
バイオインフォマティクス	jupyter や R をリバースプロキシ経由で利用

3.3 共通利用イメージを用いた VM 作成

まず、利用者は OpenStack 管理コンソールで共通利用イメージから必要な数の VM を作成する。VM として管理ノードが 1 台必要であり、必要な数だけ計算ノードを作成する。管理ノードには理研のイントラアドレスを FIP として、理研内の利用者の環境からアクセスできるようにする。

共通利用イメージの特徴は以下となる。

- ベースイメージは Ubuntu-22.04
- 共有ストレージを Lustre マウント
- システム構築に利用する Ansible をインストール済
- gcc、openmpi、docker、singularity CE、python-openstackclient をインストール済

3.4 Ansible を用いたシステム構築

VM を作成後、管理ノードに ssh でログインし、前準備の作業を行い、Ansible の playbook を適用することでシステムを構築できる。用途に合わせて必要な role をコメントアウトすると、必要な機能のインストールと環境構築が行える。用意している role は表 2 に示す。また、各種スクリプトを用意して、システム構築前の設定ファイル作成や、システム構築後の設定を簡単に行えるようにしている。

表 2：role 一覧

role	説明
common	/etc/hosts を全ノードに配る
ldapsrvr	LDAP サーバを構築
ldapclient	LDAP クライアントを構築
userandgroup	ユーザとグループを LDAP に登録
sambasrvr	Samba サーバを構築
nfssrvr	NFS サーバを構築
rdp	RDP サーバを構築

vnc	VNC サーバを構築
slurmserver	Slurm サーバを構築
slurmclient	Slurm クライアントを構築
apache_ReverseProxy	Apache のリバースプロキシを構築
bioinfomatis	Jupyter と R studio server の構築

3.5 構築後の運用

システム構築支援はソフトウェアのインストールと環境設定まで行うので、システムのアップデートは利用者が行う必要がある。ソフトウェアのインストールは基本的にパッケージ管理システムである apt 経由でインストールしているので、アップグレードも apt 経由で行える。また、他の方法でソフトウェアをインストールしても基本的に問題ない。

4 各機能の詳細

システム構築支援で利用できる各機能について Ansible でどのように環境構築を行っているか、設定などをどのようにするか、注意点などを説明する。

4.1. ユーザ管理

HSS では利用者が VM のアカウント管理を行う必要があるが、特に共有ストレージを利用する場合は UID や GID を全ての VM で共通にする必要がある。本サービスでは LDAP を使ってアカウント管理を行うこととした。

Ansible の LDAP に関する role を適用すると、必要なパッケージのインストールと LDAP の設定変更、LDAP 管理者のパスワード変更などを行う。ユーザとグループの追加や変更などは、スクリプトを用意して簡単に行えるようにしている。

4.2. ファイルサーバ

ファイル共有としては Samba サーバと NFS サーバを構築できるようにした。Samba の利用は LDAP の利用が前提となっている。

Ansible の Samba の role を適用すると必要なパッケージのインストールと Samba の設定ファイルの配置や LDAP を利用するための設定変更などを行う。構築後は管理者がユーザの Samba パスワードを設定すれば利用可能となる。なお、Samba パスワード変更で UNIX アカウントのパスワードも同時に変更するようにしている。Windows や Mac から、簡単な手順でファイル共有を行うことができるようになる。

NFS については認証の問題があるので、現時点では非推奨としている。プロジェクト内部の利用や IP アドレスで制限できる場合は問題なく利用できる。

4.3. リモートデスクトップ

リモートデスクトップは RDP サーバと VNC サーバを構築できるようにした。

RDP サーバの role を適用すると必要なパッケージのインストールが行われる。構築後、すぐに Windows や Mac から簡単な手順で接続でき、リモートデスクトップ環境を利用することが出来るようになる。

VNC サーバについては利用のための手順が煩雑なので非推奨としている。

4.4. ジョブ管理

ジョブ管理システムとして slurm を構築できるようにし、slurm 上で singularity イメージが実行できるようにしている。

Slurm の role を適用すると、必要なパッケージのインストールと設定ファイルの配置などが行われる。slurm.conf を編集することで、計算に利用できるノードの情報や partition (キュー) を変更できる。

構築後すぐに、gcc と openmpi を用いた MPI ジョブや singularity イメージのジョブを実行することができる。Intel コンパイラなど他の開発環境は、別途インストールすれば利用可能となる。

4.5. バイオインフォマティクス

バイオインフォマティクスの計算環境として、Jupyter と R studio server が構築できる。同時にリバースプロキシも利用することで、管理ノードだけでなく計算ノード上で Jupyter と R studio server を立ち上げた場合も、手元の PC などから利用することができるようになる。

bioinformatics role を適用することで、jupyter と R studio server の構築に必要なパッケージがインストールされる。jupyter はユーザ毎に動かしたいノードで起動すれば利用可能となる。R studio server はインストールしたノードに立ち上がる。なお、jupyter と R studio server はお手軽に試すため用に用意しているので、ユーザが個別に jupyter や R の環境を構築することを推奨している。

apache_reverse_proxy role を適用すると、必要なパッケージのインストールと設定ファイルの配置が行われる。リバースプロキシの設定は reverse_proxy.csv を編集して、Ansible を適用する

と設定が反映される。jupyter と R studio server 用の設定を用意しているが、設定ファイルを書き換えると他の用途にも利用できる。

- [2] mdx: データ活用社会創成プラットフォーム
<https://mdx.jp/>
- [3] 理研研究情報管理サービス
<https://dmsgrdm.riken.jp:5000/8mq9/>

5 システム構築支援の特徴と課題

本システム構築支援では、構成管理ツール Ansible を用いたシステム構築を可能としたが、同様の試みはデータ活用社会創生プラットフォーム (mdx)^[2]でもされている。mdx が全国的なシステムであるのに対して、HSS は理研の研究者用のシステムなので、理研の研究者の用途により適した機能の実装を行っている。

特に、HSS は多くの場合理研のプライベート空間で構築することになるので、簡単な操作で基本的なネットワーク設定が完了できるようになっている。また、本支援でもイントラネットでの利用を前提として各種機能を提供している。今後、グローバル環境での web サーバや理研外とのファイル共有などの機能提供を行う際は、セキュリティ面でのケアがより必要になってくる。

本支援では、あらかじめ OpenStack 管理コンソールで VM の作成を行っておく必要がある。VM の作成を OpenStack コマンドなどで行えるとより柔軟なシステムとなるので、今後の大きな課題となっている。

また、HSS はデータの取扱いが主要な用途となるが、研究の現場ではデータ転送が問題になることが多い。データ転送はさまざまな要素が関係するので HSS だけで解決する問題ではないが、転送ツールや性能測定ツール、さまざまなノウハウ共有などが有用であり、今後支援の内容に含めたいと考えている。

6 まとめ

現在までに用意したシステム構築支援で、必要最低限のシステムを構築することができるようになったと考えている。今後は広報や講習会の開催など利用者への案内が重要になると考えている。また、OS やソフトウェアのアップデートへの対応や新たな機能の導入も必要であり、利用者からのフィードバックを元に、今後さらに有用な支援となるべく開発を継続する予定である。

参考文献

- [1] 理研データ科学基盤
<https://i.riken.jp/data-sci/>