

# コンテナ技術を用いた 教育情報システムの構築

175733E 宮平 賢  
指導教員：河野 真治

## Abstract

ab,ab,ab,ab, abstract

## 1 教育向けの情報システム

情報通信技術の普及に伴い学生が学ぶ学習環境が必要となる。その学習環境として VM やコンテナにより、手軽に開発し試せる技術が普及している。だが、手元の PC 上で VM やコンテナを立ち上げ、開発を行うことはできるが、VM やコンテナの使用には高性能 PC や有料のクラウドサービスが必要になる場合がある。これらの負担を IT 技術を学ぶ学生に負わせない新たな仕組みが必要である。

本コースでは希望する学生に学科の汎用サーバから仮想環境を貸出するサービスを行っている。貸出をする VM の基本スペックとして CPU1 コア、メモリ 1GB、ストレージ 10GB である。基本スペックでは不足する場合は要望に応じてスペックの変更を行っている。しかし、機械学習などの演習では CPU より GPU が求められる場合がある。VM 上で GPU を共有するには PCI パススルーを利用することで可能である。だが、PCI パススルーでは GPU と VM は 1 対 1 の関係となり、GPU を希望する利用者すべてに割り当てることはできない。

本研究では、学生が貸出 VM だけでなく、学科の汎用サーバのリソースを効率的に利用できる教育情報システムを提案する。教育情報システムには複数の汎用サーバと大容量ストレージサーバが存在する。複数のサーバを利用するにあたり、分散ストレージが必要となる。また、学習環境として利用されることから、複数の並列なアクセスに耐えられ、信頼性の高いファイルシステムが必要である。この要件を満たすストレージソフトウェアとして Ceph を採用した。汎用サーバのリソースを効率的に利用するために、コンテナエンジンである Podman, Singularity, ジョブスケジューラである Slurm を採用した。これらのソフトウェアを合わせ教育情報システムの構築を行った。

## 2 Podman

Podman は RedHat 社が開発、提供する Linux 上で OCI コンテナを開発、管理、実行するためのデーモンレスコンテナエンジンである [1]。Podman は OCI 準拠のコンテナラ

ンタイムに依存するため、前述した Docker など他のコンテナエンジンと互換性を持つ。また、Podman CLI は Docker CLI と同じ機能を提供する。Podman はコンテナとイメージストレージ、コンテナランタイムを介して Linux カーネルと直接対話することで、デーモンレスで実行される。Podman の制御下にあるコンテナは、特権ユーザ又は非特権ユーザのいずれかによって実行することができる。

## 3 Singularity

Singularity[2] とは、HPC 環境向けに設計されたコンテナプラットフォームである。Singularity はマルチユーザに対応しており、コンテナ内での権限は実行ユーザの権限を引き継ぐため、ユーザに特別な権限の設定が必要ない。またデフォルトで、\$HOME、/tmp、/proc、/sys、/dev がコンテナにマウントされ、サーバ上の GPU を簡単に利用できる。コンテナイメージは Singularity Image Format(以下、sif)と呼ばれる単一ファイルベースのため、アーカイブや共有が容易である。

## 4 Slurm

Slurm[3] は Linux クラスタ向けのフォールトトレラント設計のジョブスケジューリングシステムである。ジョブスケジューラではサーバ上で実行される処理を「Job」という単位で管理する。ユーザはプログラムの実行手順、実行に必要なとするリソースを記した batch ファイルを作成し、ジョブスケジューラに Job の実行を依頼する。ジョブスケジューラは要求するリソース、実行時間を考慮し、複数の計算ノードから Job を実行するノードを決定する。このようにサーバ上でのプログラム等の実行、サーバのリソースを管理するのがジョブスケジューラである。Slurm には以下の 3 つの主要機能を提供する。

- 計算を実行するユーザに対してリソースへの排他的、非排他的なアクセスを割り当てる

- 割り当てられたノード上のジョブの開始, 実行, モニタリングを行う
- 待機中のジョブキューを管理することにより, リソースの競合を解決する

## 5 Ceph

Ceph は, RedHat 社が開発, 提供する分散ファイルシステムである. Ceph は分散オブジェクトストレージである RADOS(Reliable Autonomic Distributed Object Storage) がベースとなっている. RADOS では, Object Storage Daemon にデータ格納する. オブジェクトの配置には, クラスタマップを元に Controlled Replication Under Scalable Hashing(以下, CRUSH) アルゴリズムによりオブジェクトの格納先を選択する. 配置の計算に必要な情報はごくわずかであるため, Ceph クラスタ内のすべてのノードは保存されている位置を計算できる. そのため, データの読み書きが効率化される. また, CRUSH はデータをクラスタ内のすべてのノードに均等に分散しようとする.

RODOS はクラスタに保存されるデータの管理を待ち受け, 保存オブジェクトへのアクセス方法として Object Gateway, RADOS Block Device(以下, RBD), CephFS がある. Object Gateway は HTTP REST 経由でクラスタに保存されるオブジェクトへ直接アクセスが可能である. RBD はブロックデバイスとしてアクセスが可能で, libvirt を組み合わせることで VM のディスクとして使用できる. また, RBD ドライバを搭載した OS にマップし ext4 や XFS などでフォーマットして利用できる. CephFS は POSIX 互換のファイルシステムである. 複数のアクセス方法を提供することで, 用途に合わせ柔軟に変更することができる.

## 6 教育情報システムの構築

旧システムでは, 学生が演習などで利用できる環境として貸出 VM のみであった. そのため以下のような問題が生じた.

- 仮想環境の貸出サービスにおいて, 新しく仮想環境を作成するにはシステム管理チームへ申請が必要であった. そのため, 一部学生は申請の方法が分からなかったり, 貸出サービスがあることが周知されていなかったため, 旧システムのリソースが余っていた.
- 機械学習の演習では GPU が求められる. だが, 旧システムには GPU が搭載されていないため, 要求されるリソースを提供できない. そのため, 貸出サービスではなく研究室ごとの機器が多く利用された.
- 旧システムのクラスタファイルシステムである GFS2 のロックマネージャーを担当するサーバが停止すると,

ファイルシステムにアクセスができなくなった. そのため, 学科のサービスを提供できなくなった.

## 7 教育情報システムの利用

## 8 教育情報システムの評価

### 8.1 ie-podman の評価

### 8.2 ファイルシステムの評価

## 9 まとめ

## 10 今後の課題

## 参考文献

- [1] Podman, <https://podman.io/>, 2021/1/4.
- [2] Singularity, <https://sylabs.io/singularity/>, 2021/1/8.
- [3] Slurm, <https://slurm.schedmd.com/overview.html>, 2021/1/14.
- [4] Ceph, <https://docs.ceph.com/en/latest/>, 2021/1/12.
- [5] 平良 太貴 and 河野 真治, OS 授業向けマルチユーザ VM 環境の構築, 研究報告システムソフトウェアとオペレーティング・システム (OS)(2014).
- [6] 城戸翔太, 安里悠矢, 城間政司, 長田智和, 谷口祐治, ”情報系学科における教育情報システムの構築及び運用管理に関する取り組み”, 研究報告インターネットと運用技術 (IOT)(2016).