



**Pacific Teck**

HPC and Machine Learning Experts

# 変革を支える技術・踏襲を支える技術

2025/02/03 Supercomputing Japan 2025 於: タワーホール船堀

Pacific Teck Japan G.K. Senior Engineer 森本 賢治



**PacificTeck**  
HPC and Machine Learning Experts

# 世界中の最先端の次を作る技術にフォーカス

- HPC/AI 環境におけるストレージ/コンテナ/ジョブ管理に習熟
- 日本を拠点にインドを含むアジア太平洋エリア(APAC)に製品を提供
- 英語 / 中国語 / 日本語のグローバルな言語でのサポートが可能
- 各スパコンや民間企業研究所等での採用実績多数

# カテゴリー別 取扱製品

## ■ ジョブ管理システム



## ■ HPC向けコンテナ



## ■ セキュリティ分析プラットフォーム



## ■ ストレージソフトウェア

### ■ 並列ファイルシステム



BeeGFS®

BeeOND®

VAST

### ■ S3 オブジェクトストレージ



## ■ S3クラウドストレージサービス



## ■ データマネジメント



## ■ クラスターマネジメントシステム



## ■ I/O Profiling 開発者用ツール



linaroforge



スパコンを作るのに必要なミドルウェア群をワンストップで調達  
ライセンス販売だけでなく、技術サポートを展開



# 変革期

Supercomputingは  
常に変革の最前線では？

- 様々な技術の実験場
- 生まれては消えていく技術や製品を数多く目にしてきた。

今の最先端が  
次の最先端を生む

- 先行投資と回収のジレンマ
- 研究？ 実用？

この数年のトピック

- AIと量子
- AI「で」何を計算したいのか？
- AI「が」何を計算しているのか？

一番変わっているのは

- 「欲しい正解」へのアプローチ方法？
- 「欲しい正解」の定義？

適用範囲の拡大

- 実用上の「欲しい正解」と、アカデミックでの「欲しい正解」の関係。
- 投資と回収のサイクルの短縮化。

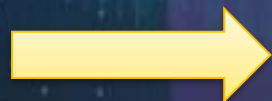
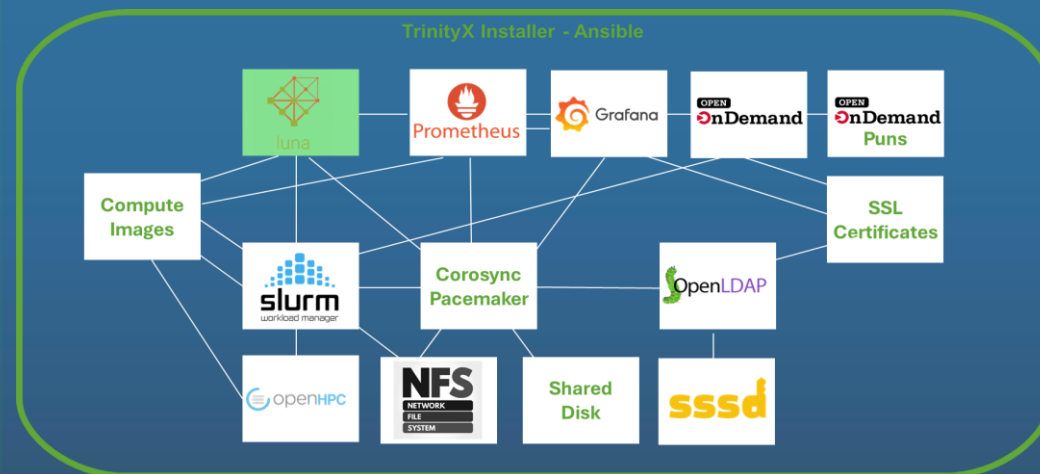
天気予報？  
生成AIでイラスト描画？  
小説やアニメ作成、  
作曲？ 論文執筆？

# システム構築について

# Supercomputing環境を構築方法の変革

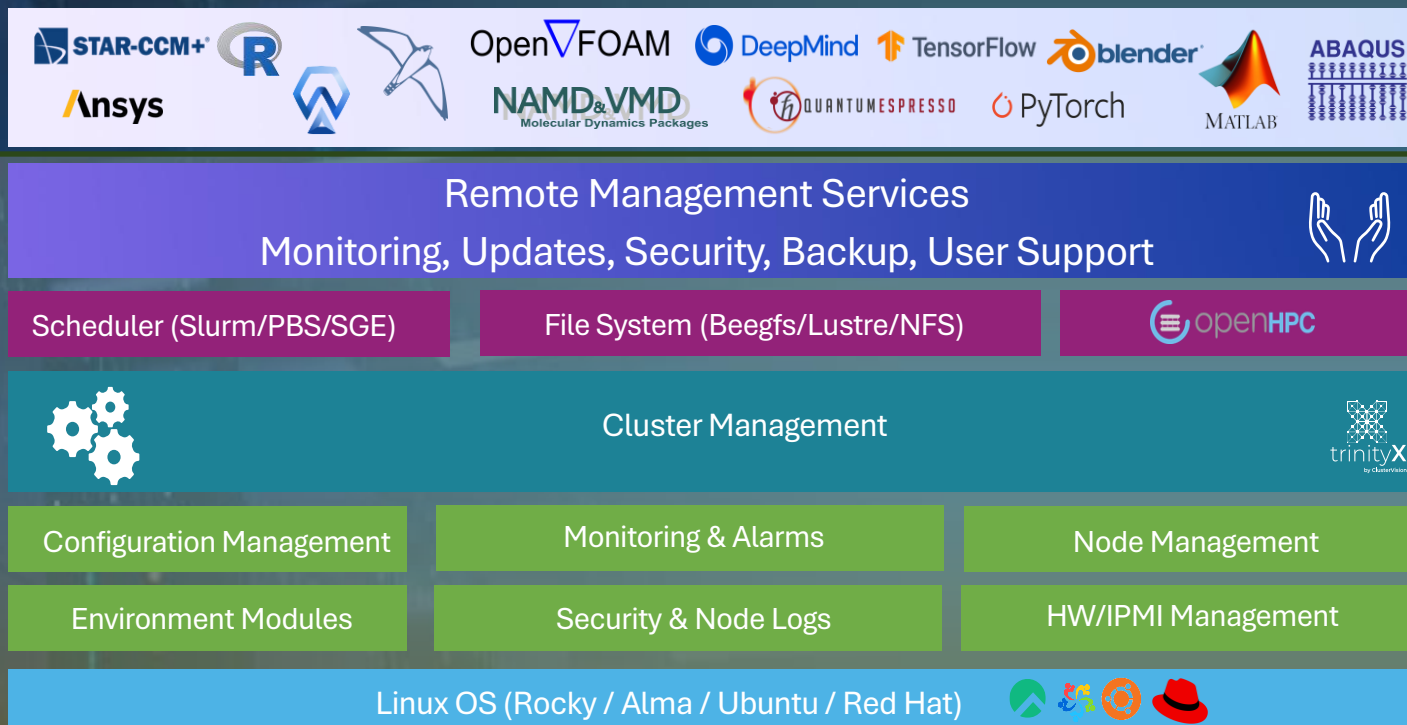
サーバーを購入するだけではクラスターは作れません。  
システム構築・運用には膨大な工数がかかります。

1. 全台にOSインストール、設定
2. GPU・インターコネクタ用ソフトウェア導入
3. 個別サーバーのネットワーク設定
4. 共有ストレージ作成、マウント設定
5. 共有認証基盤(LDAP等)作成、接続設定
6. ジョブスケジューラの導入、キュー設計
7. 並列計算ソフトウェアのインストール、設定
8. アプリケーションのインストール、設定
9. GUI のインストール、設定
10. 監視・モニタリングのインストール、設定
11. ユーザー管理、教育、トラブルシュートなど
12. システムアップデート、セキュリティ対応など



コンポーネントはOSSで揃うが、システムに組み上げることは膨大な工数とノウハウ、経験が必要。もはや力業は限界に近く、自動化するツールが必須。

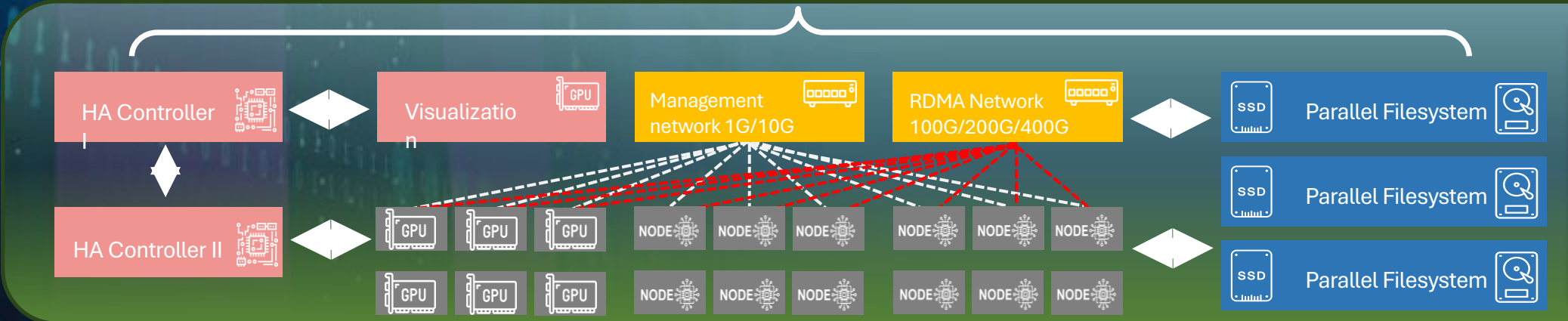
# HPC・AI クラスタ(スパコン)の構造



ここを整備してハードウェアを統合し、システムを構築・運用

様々なコンポーネント群が個々のシステムを特徴づけてきた

⇒ 統合管理ツールを使いたい



ハードウェア群  
管理ノード  
計算ノード  
ストレージ  
ネットワーク  
.....



# 柔軟なシステム構築・マネジメント

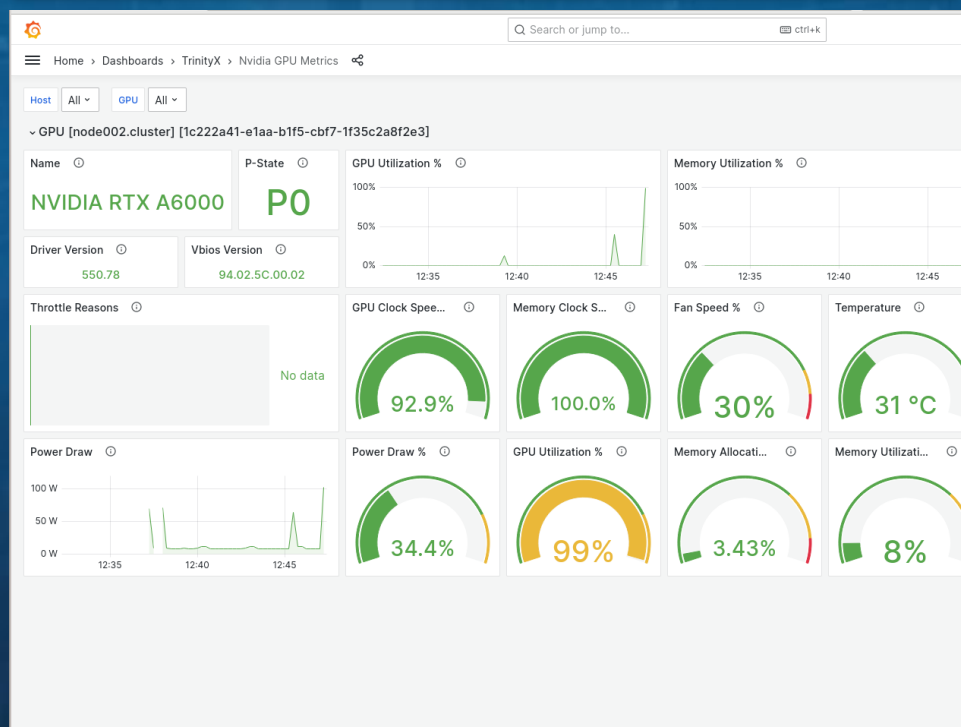


最初に Bright Cluster Manager を作った、ClusterVision社による、新設計のHPCクラスター構築・管理ツール。

OpenOnDemand の上にユーザーポータルのみならず、管理ツールUIも統合。

充実のGPUモニタリング、ジョブ実行状況分析に寄与する様々な管理画面。

計算ノードOSイメージを数千ノードまで一斉に展開可能。システムのOS更新も円滑に。





# 管理者にも利用者にも直感的なUI

Home / infiniband

# Switches 6 # Nodes 96 # Links 104 Simulation All

Nodes Links Save

Name	UID	Port	Target
SwitchB Mellanox Technologies	S-ec0d9a0300ec8240		
SwitchB Mellanox Technologies	S-ec0d9a0300ec8480		
SwitchB Mellanox Technologies	S-ec0d9a0300f548e0		
SwitchB Mellanox Technologies	S-ec0d9a0300ec8200		
SwitchB Mellanox Technologies	S-ec0d9a0300ec81e0		
SwitchB Mellanox Technologies	S-ec0d9a0300f548c0		
node019 HCA-1	H-506b4b03003fa98a		
node017 HCA-1	H-506b4b03003fa9a2		
node015 HCA-1	H-506b4b030043f9b2		
node013 HCA-1	H-506b4b030043f9aa		
node011 HCA-1	H-506b4b030043fa72		
node009 HCA-1	H-506b4b03003fa952		
node007 HCA-1	H-506b4b030041d02a		
node005 HCA-1	H-506b4b03003fa9a6		
node003 HCA-1	H-506b4b030043fa6e		
node001 HCA-1	H-506b4b03003fa9e2		
gpu002 HCA-1	H-506b4b030043fa3e		
gpu003 HCA-1	H-506b4b030043fa86		
node046 HCA-1	H-08c0e0b300ddc0f6		
gpu001 HCA-1	H-506b4b030043f9f6		
node004 HCA-1	H-506b4b030043f9de		

Home / Rack

Rack Representation

- Manage Rack Frames
- Manage Inventory
- Device Pool EMPTY

Select Metric

0 0.59

mydemorack Front Back

mydemorack Back Front

U 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16

U 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16

Select Metric

- No Scale
- Temperature
- System Load
- Power Consumption

node005 node006 node007 node008 node009 node010

Home / Dashboards / TrinityX / Thermals and Power

Host node003.cluster

Power Status: Powered On

Available Sensors: dmc Yes chassis Yes dcim Yes ipmi Yes

Machine Info: manufacturer\_id GIGA-BYTE TECHNOLOGY CO., LTD (15370) firmware\_revision 12.61

IPMI Sensors State

name	type	state
Watchdog	Watchdog 2	OK
SEL	Event Logging Disabled	OK
PS2_Status	Power Supply	Critical
PS1_Status	Power Supply	OK
CPU1_Status	Processor	OK
CPU0_Status	Processor	OK

Power (3 panels)

Temperatures

name	temperature	state
CPU0_TEMP	47 °C	Ok
CPU1_TEMP	50 °C	Ok
DIMMG0_TEMP	36 °C	Ok
DIMMG1_TEMP	35 °C	Ok
DIMMG2_TEMP	36 °C	Ok
DIMMG3_TEMP	35 °C	Ok
HDD_TEMP_0	27 °C	Ok
HDD_TEMP_1	28 °C	Ok
INLET_AIR_TEMP	27 °C	Ok
MB_TEMP1	44 °C	Ok
MB_TEMP2	42 °C	Ok
VR_DIMMG0_TEMP	40 °C	Ok
VR_DIMMG1_TEMP	45 °C	Ok

name	speed	state
BPB_FAN_1A	6450 rpm	Ok
BPB_FAN_1B	5250 rpm	Ok
BPB_FAN_2A	6600 rpm	Ok
BPB_FAN_2B	5250 rpm	Ok
BPB_FAN_3A	6600 rpm	Ok
BPB_FAN_3B	5250 rpm	Ok
BPB_FAN_4A	6450 rpm	Ok
BPB_FAN_4B	5250 rpm	Ok
BPB_FAN_5A	6600 rpm	Ok
BPB_FAN_5B	5250 rpm	Ok
BPB_FAN_6A	6600 rpm	Ok
BPB_FAN_6B	5250 rpm	Ok
BPB_FAN_7A	6600 rpm	Ok

name	voltage	state
P_12V	12.1 V	Ok
P_1V0_AUX_LLAN	990 mV	Ok
P_3V3	3.25 V	Ok
P_5V	4.96 V	Ok
P_5V_STBY	4.93 V	Ok
P_VBAT	3.05 V	Ok
PO_VDD_1B	1.82 V	Ok
PO_VDDCR_CPU	730 mV	Ok
PO_VDDCR_SOC	680 mV	Ok
P1_VDD_1B	1.82 V	Ok
P1_VDDCR_CPU	730 mV	Ok
P1_VDDCR_SOC	680 mV	Ok
VR_DIMMG0_VOUT	1.25 V	Ok

Home / Job Composer

Jobs Templates

Jobs

New Job

Edit Files Job Options Open Terminal Submit Stop Loading

Show 25 entries

Created	Name	ID	Cluster	Status
December 6, 2023 2:34pm	(default) Simple Sequential Job	667	TrinityX	Completed
December 6, 2023 2:34pm	(default) Simple Sequential Job	21	TrinityX	Completed
October 31, 2023 2:59pm	(default) Simple Sequential Job	20	TrinityX	Completed

Showing 1 to 3 of 3 entries

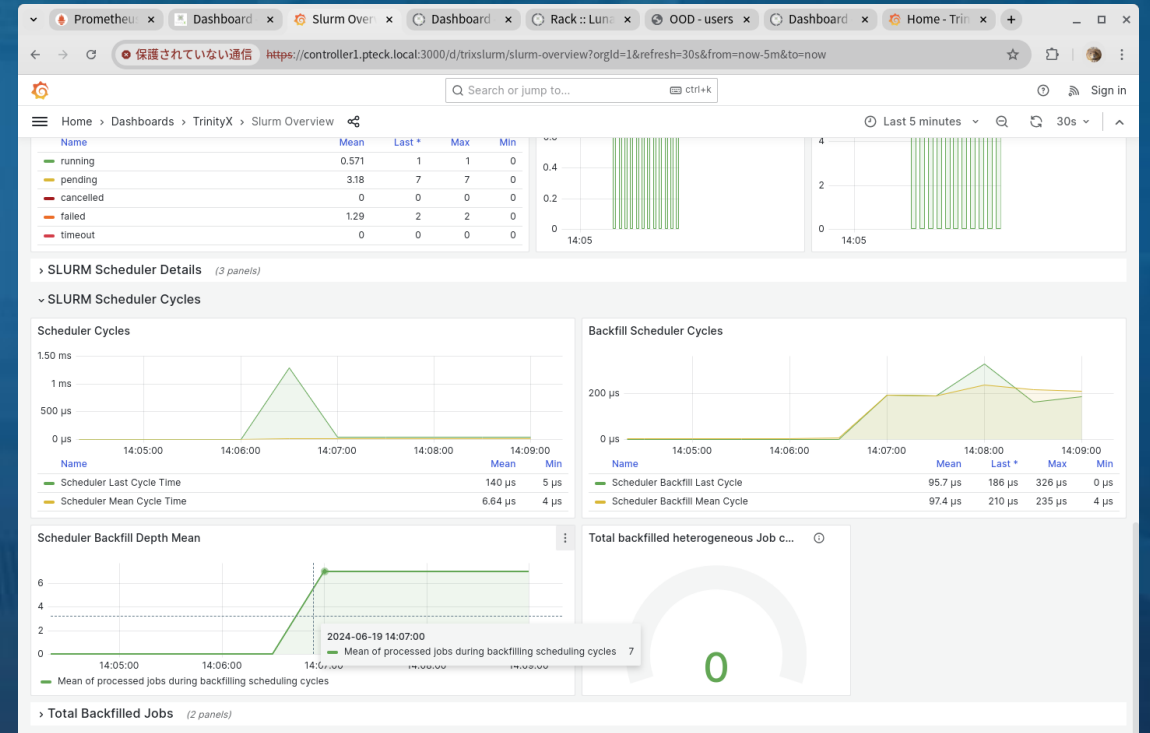
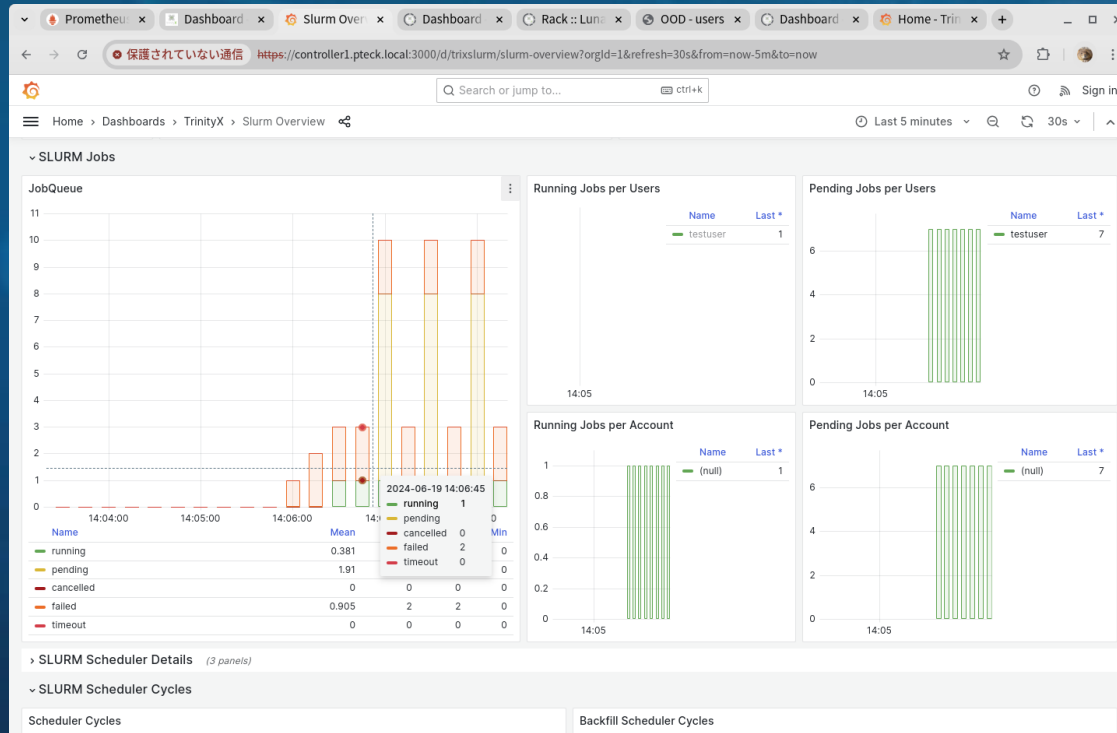
Job Details

slurm-15.out  
slurm-17.out  
slurm-19.out  
slurm-20.out

# ジョブスケジューラの稼働状況の分析

ジョブの待ち時間等のデータを分析するダッシュボードがデフォルトで作成されており、利用状況に合わせたスケジューリング設定の調整だけでなく、将来的なリソース追加の計画策定にも寄与します。

slurm 設定の構築・検証ツールも付属しているため、GUIから必要な作業が完了します。



# アプリケーション実行環境について

# アプリケーション環境構築の変革

## 準備

- ユーザーが使うアプリケーションを事前に募り、必要なランタイムをリストアップして、共存できるようにすり合わせを行う。



## 導入

- リストアップしたアプリケーション、ランタイムに対し、それぞれ必要なバージョンを并存させ、切り替えて使う機構を用意する。



## 運用

- アップデートが必要になった際は、管理者に要望を出し、メンテナンスのタイミングで追加してもらう。



- 多様化する用途に合わせて、多彩なアプリが登場。高頻度でアップデートを繰り返す。
- システム全体でそれらの環境をすり合わせるのは、既に現実的でない。
- NGCをはじめ、コンテナを用いて実装をデリバリーする仕組みの充実。
- ユーザーの方が管理者より、アプリの実行環境の詳細に詳しい。







## SingularityCE

- オープンソースで提供され、有償サポートはありません。
- 新機能は最初に実装されますが、セキュリティ対応や修正は最新版にのみです。



## SingularityPRO

- Sylabs社による検証・署名済みバイナリを提供します。
- 長期サポート(リリースから2年)・機能追加・プラグイン開発を依頼できます。
- 脆弱性の公開前に対応済みバイナリが提供され、バックポートにも対応します。



## Singularity Enterprise

- コンテナランタイムとしてはSingularityPROを提供します。
- Kubernetes上に、RemoteBuilder, KeyStore, Libraryを提供します。EKS, GKEもサポート。ただし、Kubernetesの運用はユーザー責任です。
- 自社・自サイトの認証基盤を統合可能です

SingularityにはKubernetesやOpenShiftのようなオーケストレーター実装がありません。  
ただし、OCIイメージを取り込んだり、SIFをOCI bundle として見せたり、そこからcrun/runcを使って動作可能。

# SIF利用による優位性

Singularityという概念は、「ソフトウェアとしてのSingularity」と「規格としてのSIF」、両方からなります。コンテナエンジンとしてのSingularityのみならず、SIF自体にも際立った優位性があります。

## 高い可搬性

- シングルファイルで保存するため、持ち運びが用意。
- 通常のNASでの管理も可能。
- リポジトリに依存せず、電子署名による同一性を保証。

## 低フットプリント

- 内部はSquashFS。tar+gzipを直接マウントしているイメージ。
- オンメモリでアクセスでき、事前に展開する必要がない。メモリやCPU負荷も極めて小さい。

## ハイパフォーマンス

- ファイルアクセスはローカルのオンメモリで行われ、共有ファイルシステムへの、メタデータアクセス負荷を極小化できる。通信負荷も下げられる。

特に昨今、AIのワークロードがPythonベースで実行されるため、個別環境がユーザーホームディレクトリに展開。共有ファイルシステムに高負荷を生み、大きな問題となっている。  
例) PyTorchはimportするだけで1000回もファイルをオープンする。

- ホストの環境とは独立にランタイム環境を固定できる。
- 古い実行環境の保存だけでなく、既存システムが対応できない、最新の環境を試すという、踏襲・変革どちらの方向にも有用。



参考: 弊社ブログ

Singularity利用による共有ファイルシステムの負荷軽減

Pros

SIFを使うメリットは前項の通り、直感的な可搬性やI/O負荷の低減など、極めて高い。

ファイル名で管理されるため、誤選択やコピーミス等で、再現性トラブルだけでなく、セキュリティリスクとなり得てしまう。

Cons



## リポジトリ+タグ指定による実行で解決

- SIFファイルの直接指定ではなく、リポジトリのイメージを指定する。
- `$ singularity exec ./sample.sif appli` → ×
- `$ singularity exec library://myrepo/smp:2.1 appli` → ○
- ローカルにSIFでキャッシュされ、1ファイルのメリットを失わない。

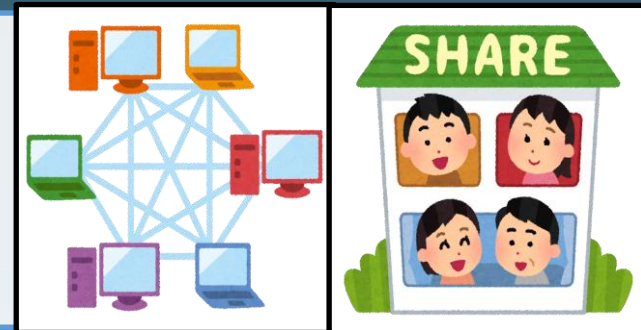
# 環境の統合と分割



# コンテナを利用したリソースの切り出し

## リソースの統合と分割

- とにかく巨大な1システムが欲しかったHPC
- 大きくなりすぎたマシンリソースを分割し多目的に利用可能に
- マルチテナントの要求



## コンテナによるリソース分割

- dockerとKubernetesの台頭
- バッチジョブスケジューリングとの競合



## 落としどころ

- Hadoop ブームの時代から試行錯誤が続いている
- TrinityX の K3s(ライトな K8s 実装)対応を予定。
- SchedMD の Slinky による slurm との共存。



- これまで培われた蓄積をさらに発展させ、従来を踏襲しつつ、これからの変革に耐える環境を作る。そうした流れに有用と思われるプロダクトをご紹介しました。
- 皆様の「欲しい正解」を、もう一度考え直すきっかけとなりましたら幸いです。

# Thank you

お問い合わせはこちらまで  
[sales@pacificteck.com](mailto:sales@pacificteck.com)