



# Pacific Teck 過去 8年半と未来

設立20周年記念PCクラスタシンポジウム

December 2021

Howard Weiss ワイスハワード



# Agenda

- 会社概要と自己紹介
- 過去の8年半の実績
- 将来に向けての課題



Managing Director

# Howard Weiss

ハワードワイス

- アメリカ ミシガン州出身
- 1990年日本移住
- 英語・日本語
- Phoenix Technologies / Sales Director, Japan
- BakBone Software / VP APAC
- Cofio Software (acquired by HDS) / co-founder BOD member
- Voltaire / Mellanox / VP APAC
- DataDirect Networks / VP APAC
- Incorporated Pacific Teck Limited in July 2013



**PacificTeck**  
HPC and Machine Learning Experts

# 世界中の最先端技術製品にフォーカス

- 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支援が可能
- ハードウェアには依存しないソフトウェアソリューションを提供
- APAC最大のスパコンでの採用実績多数
- ストレージ/コンテナ/ジョブ管理のエキスパート

# Pacific Teck in the TOP500! Nov 2021 list



- 1<sup>st</sup> Riken/Fugaku** (SingularityPRO, iRODS, ARM Forge)
- 16<sup>th</sup> AIST/ABCI** (Grid Engine, SingularityEnterprise, BeeGFS)
- 54<sup>th</sup> NCHC/Taiwania 2** (Grid Engine, SingularityPRO)
- 59<sup>th</sup> TiTech/TSUBAME3.0** (Grid Engine, BeeGFS)
- 79<sup>th</sup> Osaka U/Squid** (SingularityPRO)
- 98<sup>th</sup> Nagoya U/Flow** (BeeGFS, NVMesh)
- 110<sup>th</sup> Tokyo U/Oakbridge-CX** (BeeGFS)
- 433<sup>th</sup> Tohoku U/AFI-NITY** (Grid Engine)

# 取扱製品カテゴリー

- ジョブ管理システム



- HPC仮想コンテナシステム



- ストレージソフトウェア

- 並列ファイルシステム



- NVMe高速ストレージ



- 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



- ティアリングソフト・データ移動のツール



- クラスターマネジメントシステム



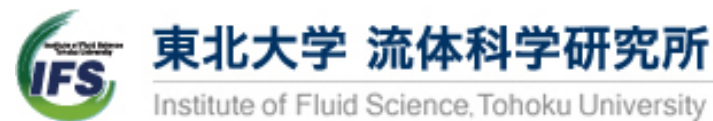
- プログラム開発者用ツール



# 過去 8 年で達成できたこと

- GPU-NUMA bus aware ノード分割 job scheduling
- Promoting the containerization of applications to improve portability
- Usage of a compute node as a high performance storage scratch system
- Creation of high speed storage systems from commodity hardware

CPU・GPUの構成を考慮した最適なジョブ実行が可能



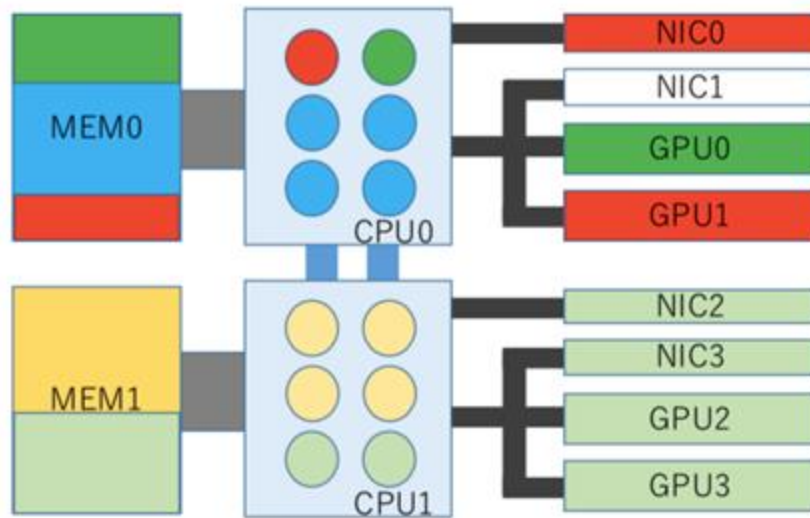


# Altair (Univa) Grid Engine - TSUBAME3.0



東京工業大学  
Tokyo Institute of Technology

## TSUBAME3.0 Container-Based Fine-grained Spatial Resource Allocations of Fat Nodes

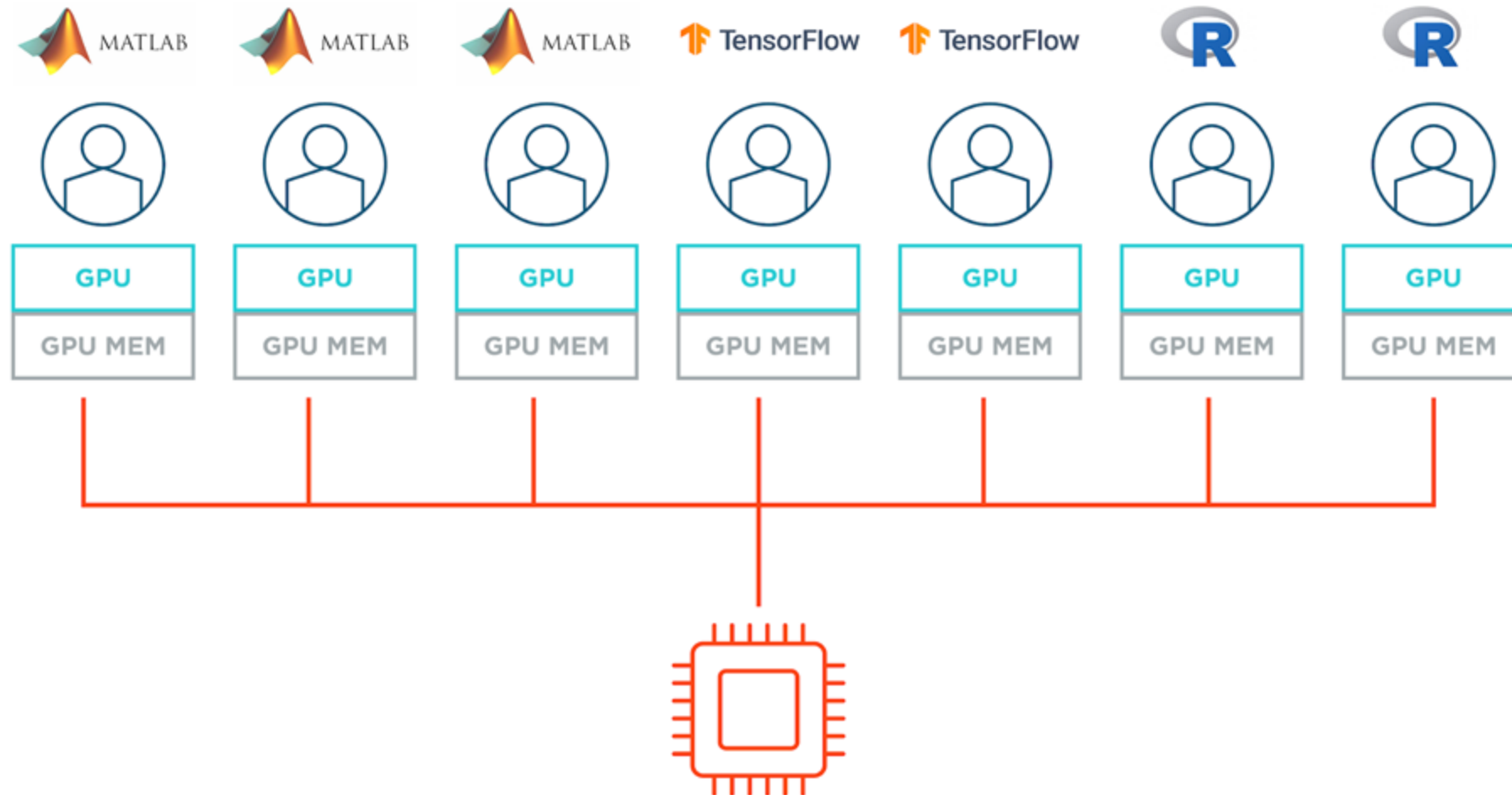


Resource Isolation via UGE  
Containers (future Docker etc.)

Job	Allocated Resource
1	CPU 2Cores, NIC0, GPU1, 32GB Mem
2	CPU 8 Cores, 64GB Mem
3	CPU 4 Cores, GPU0, 16GB Mem
4	CPU 8 Cores, 64GB Mem
5	CPU 4 Cores, NIC2&3, GPU2&3, 48G Mem

Container configuration and deployment tied to Univa Grid Engine

## Realize the full potential of the A100 GPU



# 過去8年で達成できたこと

- GPU-NUMA bus aware job scheduling
- Promoting the containerization of applications to improve portability
- Usage of a compute node as a high performance storage scratch system
- Creation of high speed storage systems from commodity hardware



# SingularityPRO · Singularity Enterprise Sample Endusers





## Singularity Enterpriseの特徴

- PGP公開鍵を追加
- コンテナライブラリを提供（オンプレ版・Cloud版）
- Remote Build Services を提供（オンプレ版・Cloud版）



## SingularityPROの特徴

- Sylabs社により技術確認されたSingularityPROのバイナリ版を提供
- 同じバージョンのロングランサポート
- お客様のご要望を次期バージョンに反映

# 過去 8 年で達成できたこと

- GPU-NUMA bus aware job scheduling
- Promoting the containerization of applications to improve portability
- Usage of a compute node as a high performance storage scratch system
- Creation of high speed storage systems from commodity hardware



# BeeOND Parallel File System

## Sample End users

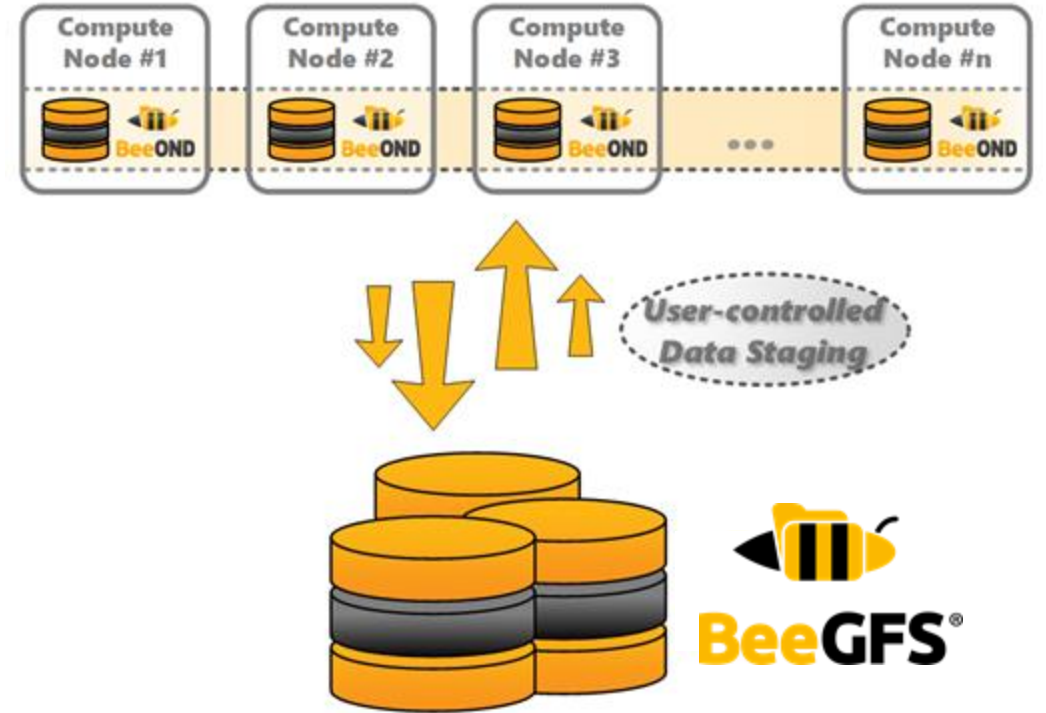


# BeeOND - BeeGFS On Demand

- パラレルファイルシステムをその場で作成
- 簡単な起動/停止コマンド
- Jobスクリプトに記述しジョブスケジューラによる作成・削除の制御が可能（ジョブスケジューラ: Altair Grid Engine）

<< これから流行 >>

計算ノードの中のファイルシステムは  
パーマントユースとして使用



東京工業大学  
Tokyo Institute of Technology



東京大学情報基盤センター  
INFORMATION TECHNOLOGY CENTER, THE UNIVERSITY OF TOKYO



©2020 - Pacific Teck Japan G.K.



# 過去 8 年で達成できたこと

- GPU-NUMA bus aware job scheduling
- Promoting the containerization of applications to improve portability
- Usage of a compute node as a high performance storage scratch system
- Creation of high speed storage systems from commodity hardware



# CSIRO Storage Architecture

## 2.048 PB usable capacity all NVMe

Metadata: Dell PowerEdge R740XD



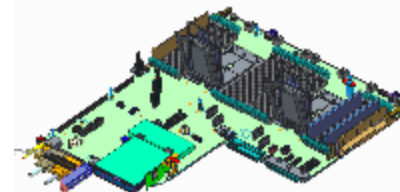
x 4

Storage: Dell PowerEdge R740XD



x 32

3.2 TB NVME



x 24 per server

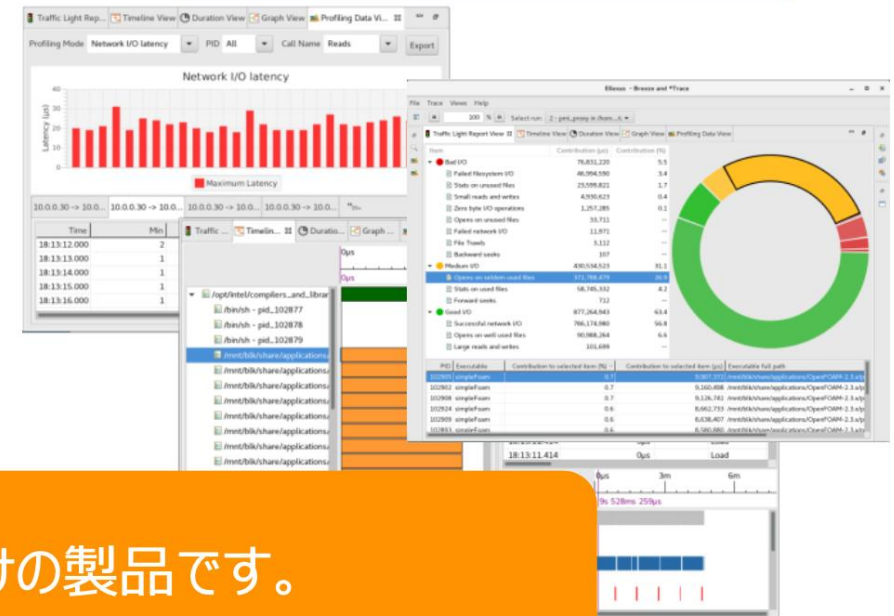
# 将来に向けての課題

- How to support storage, containers, debugging tools on upcoming architectures such as ARM and accelerators
- Moving of the parallel permanent file system to the compute node
- Seamless operation of on premise and cloud resources with job schedulers, containers, etc.
- How to move / archive data on an S3 compliant cloud
- Various use cases for S3
- Application specific accelerator eco system support

# Altair Breeze™ の概要

- Altair Breeze™（以後、Breezeと略します。）は、Mistralより詳細に、全てのファイルとプロセスについてのI/Oを監視する事が出来ます。I/Oエキスパートユーザー、管理者に大変役に立つツールです。
- 内製アプリの最適化のために利用したり、レガシーアプリケーションのマイグレーションのために利用出来ます。

Mount Point	File Type	Filename	Full Path	Package	Read # Call	Small (< 32kB) Read # Call	Large (≥ 100MB) Read # Call	Write # Call
/	Data File	dumb	/usr/share/terminfo/d/dumb	terminfo-base-6.1-1p150.4.3...	7	7	0	0
/	Data File	xterm	/usr/share/terminfo/x/xterm	terminfo-base-6.1-1p150.4.3...	1	1	0	0
/	Data File	bindkey.tchsh	/etc/profile.d/bindkey.tchsh	tchsh-6.20.00-1p150.1.9.x86_64	5	4	0	0
/	Data File	complete.tchsh	/etc/profile.d/complete.tchsh	tchsh-6.20.00-1p150.1.9.x86_64	8	7	0	0
/	Data File	hosts.equiv	/etc/hosts.equiv	netcfg-11.6-1p150.1.1.noarch	2	1	0	0
/	Shared Library	libz.so.1	/lib64/libz.so.1	libz1-1.2.11-1p150.2.3.1.x86...	0	0	0	0
/	Shared Library	libxm2.so.2	/usr/lib64/libxm2.so.2	libxm2-2-2.9.7-1p150.2.6.1...	0	0	0	0
/	Shared Library	libselinux.so.1	/lib64/libselinux.so.1	libselinux1-2.6-1p150.2.14.x...	0	0	0	0
/	Shared Library	libgthread-2.0.so.0	/usr/lib64/libgthread-2.0.so.0	libgthread-2.0-0-2.54.3-1p1...	0	0	0	0
/	Shared Library	libgobject-2.0.so.0	/usr/lib64/libgobject-2.0.so.0	libgobject-2.0-0-2.54.3-1p15...	0	0	0	0
/	Shared Library	libglib-2.0.so.0	/usr/lib64/libglib-2.0.so.0	libglib-2.0-0-2.54.3-1p150.3...	0	0	0	0
/	Shared Library	libattr.so.1	/lib/libattr.so.1	libattr1-32bit-2.4.47-1p150.2...	0	0	0	0
/	Shared Library	libattr.so.1	/lib64/libattr.so.1	libattr1-2.4.47-1p150.2.16.x...	0	0	0	0
/	Shared Library	libacl.so.1	/lib/libacl.so.1	libacl1-32bit-2.2.52-4p150.3...	0	0	0	0
/	Shared Library	libacl.so.1	/lib64/libacl.so.1	libacl1-2.2.52-4p150.3.3.1.x8...	0	0	0	0
/	Shared Library	librt.so.1	/lib/librt.so.1	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	libutil.so.1	/lib/libutil.so.1	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	libcrypt.so.1	/lib/libcrypt.so.1	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	ld-linux.so.2	/lib/ld-linux.so.2	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	libnsl.so.1	/lib/libnsl.so.1	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	libdl.so.2	/lib/libdl.so.2	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	libpthread.so.0	/lib/libpthread.so.0	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	libc.so.6	/lib/libc.so.6	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	libm.so.6	/lib/libm.so.6	glibc-32bit-2.26-1p150.11.9...	0	0	0	0
/	Shared Library	libnss_dns.so.2	/lib64/libnss_dns.so.2	glibc-2.26-1p150.11.9.1.x86...	0	0	0	0
/	Shared Library	ld-linux-x86-64.so.2	/lib64/ld-linux-x86-64.so.2	glibc-2.26-1p150.11.9.1.x86...	0	0	0	0
/	Shared Library	libutil.so.1	/lib64/libutil.so.1	glibc-2.26-1p150.11.9.1.x86...	0	0	0	0



Breezeは、I/Oエキスパートの方向けの製品です。

# 将来に向けての課題

- How to support storage, containers, debugging tools on upcoming architectures such as ARM and accelerators
- **Moving of the parallel permanent file system to the compute node**
- Seamless operation of on premise and cloud resources with job schedulers, containers, etc.
- How to move / archive data on an S3 compliant cloud
- Various use cases for S3
- Application specific accelerator eco system support

# コンピュータノード上にストレージを実現する



- タイプIIサブシステム100ノードはBeeGFSを使用
- 50ノードがNVMeShを使用
- 50ノードでRAID 1とイレイシャーコーディング使用パーマネントファイルシステム
- 2020年7月1日に運用開始
- 永続ファイルシステムとして使用可能

# 将来に向けての課題

- How to support storage, containers, debugging tools on upcoming architectures such as ARM and accelerators
- Moving of the parallel permanent file system to the compute node
- Seemless operation of on premise and cloud resources with job schedulers, containers, etc.
- How to move / archive data on an S3 compliant cloud
- Various use cases for S3
- Application specific accelerator eco system support



NEW

Altair Control  
navops

- Migrates workloads to the cloud
- Helps organizations control spending

*"We gained practically infinite capacity with Univa's hybrid cloud solution in a very cost-effective manner"*  
Enterprise Computing Director, HPC Mellanox

Sample  
clients





# 将来に向けての課題

- How to support storage, containers, debugging tools on upcoming architectures such as ARM and accelerators
- Moving of the parallel permanent file system to the compute node
- Seamless operation of on premise and cloud resources with job schedulers, containers, etc.
- How to move / archive data on an S3 compliant cloud
- Various use cases for S3
- Application specific accelerator eco system support

POSIXファイル  
システム



GPFS, Lustre etc

POSIX

オンプレミス  
オブジェクト  
ストレージ



Scality etc

S3

クラウドストレージ



Oracle Cloud, AWS, Azure etc

S3 / API

データカタログ・データ移動



# 将来に向けての課題

- How to support storage, containers, debugging tools on upcoming architectures such as ARM and accelerators
- Moving of the parallel permanent file system to the compute node
- Seamless operation of on premise and cloud resources with job schedulers, containers, etc.
- How to move / archive data on an S3 compliant cloud
- Various use cases for S3
- **Application specific accelerator, eco system support**



**Pacific Teck**  
HPC and Machine Learning Experts

# Thank you

[sales@pacificteck.com](mailto:sales@pacificteck.com)

