

多重複合型超並列クラスター Cygnus について

1. はじめに

Cygnus は筑波大学計算科学研究センターが管理運用する多重複合型超並列クラスターである。同システムは 2019 年 4 月から約 1 ヶ月の試験運用期間を経て、2019 年 5 月から公式運用が開始される。各計算ノードに最先端 GPU である NVIDIA Tesla V100 (Volta) を 4 台搭載するのに加え、全体の 40% のノードにはさらに最先端 FPGA (Field Programmable Gate Array) である Intel Stratix10 を搭載した Nallatech 520N FPGA card を 2 枚搭載する。これらを統合する CPU として Intel Xeon Gold を 2 基搭載する。GPU と FPGA を 1 つのノードに搭載することから、本システムを多重複合型超並列クラスターと呼ぶ。このようなアーキテクチャを持つスーパーコンピュータは国内初の試みであり、極めて先進的かつ挑戦的なシステムである。

Cygnus の構成は 80 台の超高性能ノードと強力な相互結合網、大規模高速共有ファイルシステムからなる。80 台全てのノードに高性能 CPU と GPU を搭載し、さらにうち 32 台にはこれらに加え FPGA が搭載される。GPU と CPU のみのノード (48 台) は Deneb ノードと呼ばれ、FPGA を追加したノード (32 台) は Albireo ノードと呼ばれる。全てのノードは先進の InfiniBand HDR100 x4 ポート (合計 400Gbps) の超高速相互結合網で接続されるが、これに加え、Albireo ノード間は FPGA に直結する 4 ポートの 100Gbps の光インターコネクトで 8x8 の 2 次元トーラス相互結合網で接続される。すなわち、Albireo ノードの全 FPGA はノード当たり 800Gbps の超高性能・低遅延ネットワークで結合され、並列 FPGA システムの高性能計算及び AI 等の先進的アプリケーションの研究を推進する。

2. Cygnus システムの概要

Cygnus のハードウェア諸元について以下にまとめる。

項目	仕様
Deneb ノード構成 (FPGA 非搭載) x48 台	
CPU	Intel Xeon Gold 6126 x2 台
CPU コア数	24
CPU ピーク性能	2 TFLOPS (倍精度)
CPU メモリ	DDR4 192 GiB (255.9 GB/s)
GPU	NVIDIA Tesla V100 x4 台
GPU ピーク性能	28 TFLOPS (倍精度), 56 TFLOPS (単精度),

	112 TFLOPS (半精度)
GPU メモリ	32 GiB/GPU, 合計 128 GiB (3600GB/s)
ノードローカルストレージ	3.2 TB NVMe
Albireo ノード構成 (FPGA 搭載) x32 台	
CPU, GPU, メモリ	Deneb ノードと同じ
FPGA	Nallatech 520N with Intel Stratix10 x2 台
FPGA メモリ	32 GiB/FPGA (64 GiB/ノード, 153.6 GB/s)
FPGA ピーク性能	10 TFLOPS (単精度) /FPGA
並列処理ネットワーク	
全ノード相互結合網	InfiniBand HDR100 x4 ポート
ノード当たりバンド幅	400Gbps
トポロジ	Fat-Tree (full bisection bandwidth)
FPGA 間相互結合網	100Gbps x4 ポート/FPGA (x8 ポート/ノード)
ノード当たりバンド幅	800Gbps (うち 100Gbps x2 はノード内の FPGA 間を接続)
トポロジ	2次元トーラス (8x8)、リンク当たり 100Gbps
全体システム構成	
総ノード数	80 (うち 2 ノードの Deneb ノードはデバッグ用のため、最大並列度は 78 ノード、Albireo ノードの最大並列度は 32 ノード)
ファイルシステム	2.5PB の RAID6 Lustre file system
ファイルシステムへのアクセス	全ノード上のプロセスから任意にアクセス可能
ログインノード	3 台

Cygnus へのアクセスはログインノードへのログインを通じて行う。フロントエンドノードは `cygnus.ccs.tsukuba.ac.jp` の代表ホスト名が与えられており、負荷に応じて自動的に 3 台のうち 1 台にログインサービスが振り分けられる。ジョブ投入及びファイルサーバへのアクセス機能については 3 台のログインノードは均等である。外部とのファイル転送はログインノードに対する `scp` コマンドを通じてのアクセスのみが許される。

計算ノードに対するジョブ実行は NEC Network Queuing System V (NQSV) によるバッチキュー・システムを通じて行う。各計算ノードへの直接ログインは許されていない。計算に必要なファイル及びジョブスクリプト群を予めファイルサーバに置き、ジョブを投入する。ファイルサーバはフロントエンドノード及び全計算ノードから透過にアクセス可能で

ある。計算結果は各計算ノードによってファイルサーバに出力されるようプログラムする。

Cygnus のファイルシステムは全計算ノード及びログインノードから均質にアクセス可能な大規模共有クラスタファイルシステムであり、**2.5PB** の **RAID6 Lustre** ファイルシステムで構成される。

Cygnus システムの構成及び利用方法の詳細については、「Cygnus 利用の手引き」（プロジェクト採択後にアクセス方法を通知）を参照のこと。