

計算科学研究センター
スーパーコンピュータ
Cygnusの概要

Cygnusについて

計算科学研究センターではこれまでGPU、many-core等のアクセラレータ付きクラスタによる超高速演算加速スーパーコンピュータを導入して来ました。many-coreスーパーコンピュータとしてIntel Knights Cornerカードを搭載したクラスタCOMAは2019年3月末をもって運用を終了します。このプログラムや利用技術は筑波大学・東京大学共同運用によるOakforest-PACS (Intel Knights Landingプロセッサによる超並列クラスタ) に引き継がれています。また、GPUクラスタHA-PACS及びHA-PACS/TCAはそれぞれ2018年3月、2018年10月をもって運用が終了しました。

同センターでは、次世代のスーパーコンピュータとして、GPUクラスタの技術を踏襲しつつ、さらなる新技術の開発とアプリケーションプログラム開発を目指し、GPUに加え超高性能FPGAを併せて搭載する計算ノードを持つ新スーパーコンピュータCygnusを導入することになりました。Cygnusの全ノードは4台の最新型GPUを備え、さらに全ノードの半分弱のノードにはこれらに加え2台の最新型FPGAを搭載します。

GPUとFPGAの混載により、これまでGPUだけでは効率的な計算が難しかったアプリケーションの性能を高め、またこれからの時代のstrong scaling並列処理に対応する演算加速技術の確立を目指すことを目的とし、Cygnusクラスタの運用を行います。

Cygnusの構成(1)

- 筑波大学計算科学研究センターが平成31年4月より運用を開始するスーパーコンピュータ
- システム構成
 - システム構成：GPU搭載ノード“Deneb node”とGPU・FPGA搭載ノード“Albireo part”からなる演算加速クラスター
 - Deneb 及び Albireo 共通仕様
 - CPU: Intel Xeon Gold 6126 2.6GHz (12 core) x 2 socket
 - GPU: NVIDIA Tesla V100 PCIe x 4台
 - Memory (CPU): 192 GiB, 255.9 GB/s
 - Memory (GPU): 32 GiB x 4 = 128 GiB, 900 GB/s x 4 = 3600 GB/s
 - Network: InfiniBand HDR100 x 4 channel (400Gbps x 4)
 - Node Storage: NVMe 3.2TB
 - Albireo ノード仕様
 - FPGA: Nallatech 520N (Intel Stratix10 with 100GBps x 4 channel ext. port)
 - Inter-FPGA Network: 100Gbps 2D Torus (8x8)

Cygnusの構成(2)

- Albireo nodeのFPGA
 - Intel Stratix10 GX2800 H-Tile
 - LE: 2753 Kgates (8-in 4-out Lookup Table + 1bit register)
 - Memory
 - BRAM: M20K 229Mbit
 - MLAB: 15Mbit
 - External memory: 32 GiB, 76.8 GB/s
 - External network: 100Gbps x 4 channel

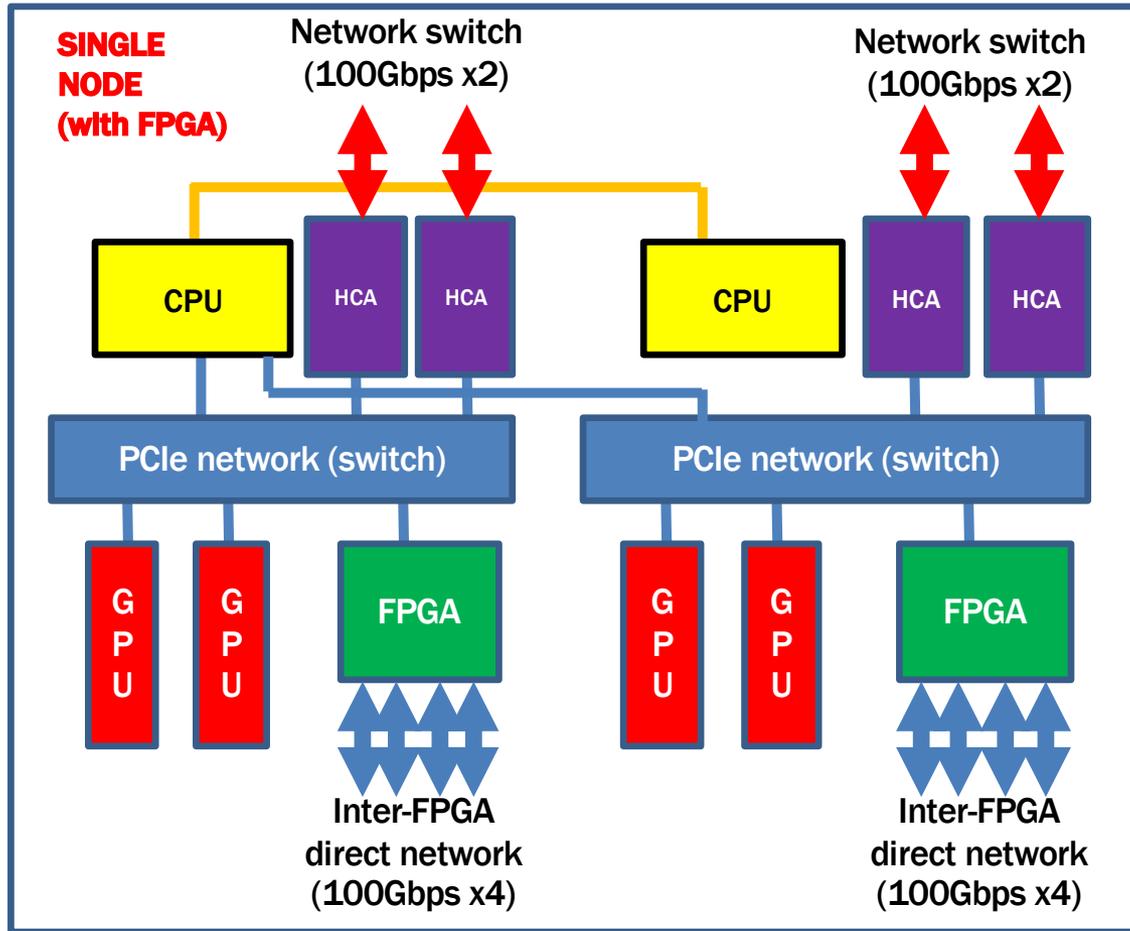
Cygnusの構成(3)

- 共有ファイルシステム
 - 全ノードから均一にアクセス可能な2.5 PB (user space)の大容量高速アクセス可能なLustre共有ファイルシステム
 - RAID6構成
 - “/work”として参照可能、計算処理全般及びデータアーカイブとして利用
- システム理論ピーク性能
 - ノード台数： Deneb x 48 nodes, Albireo x 32 nodes = 80 nodes
 - CPU性能 (double precision)： 2 TFLOPS x 80 = 160 TFLOPS
 - GPU性能 (double precision)： 28 TFLOPS x 80 = 2240 TFLOPS
 - FPGA性能 (single precision): 10 TFLOPS x 32 = 320 TFLOPS
 - システム合計性能 (double precision): 2.4 PFLOPS
システム合計性能 (single precision): 5.12 PFLOPS
- システムNetwork理論ピーク性能
 - 100Gbps x 4 x 80 = 4GB/s (full bisection bandwidth)

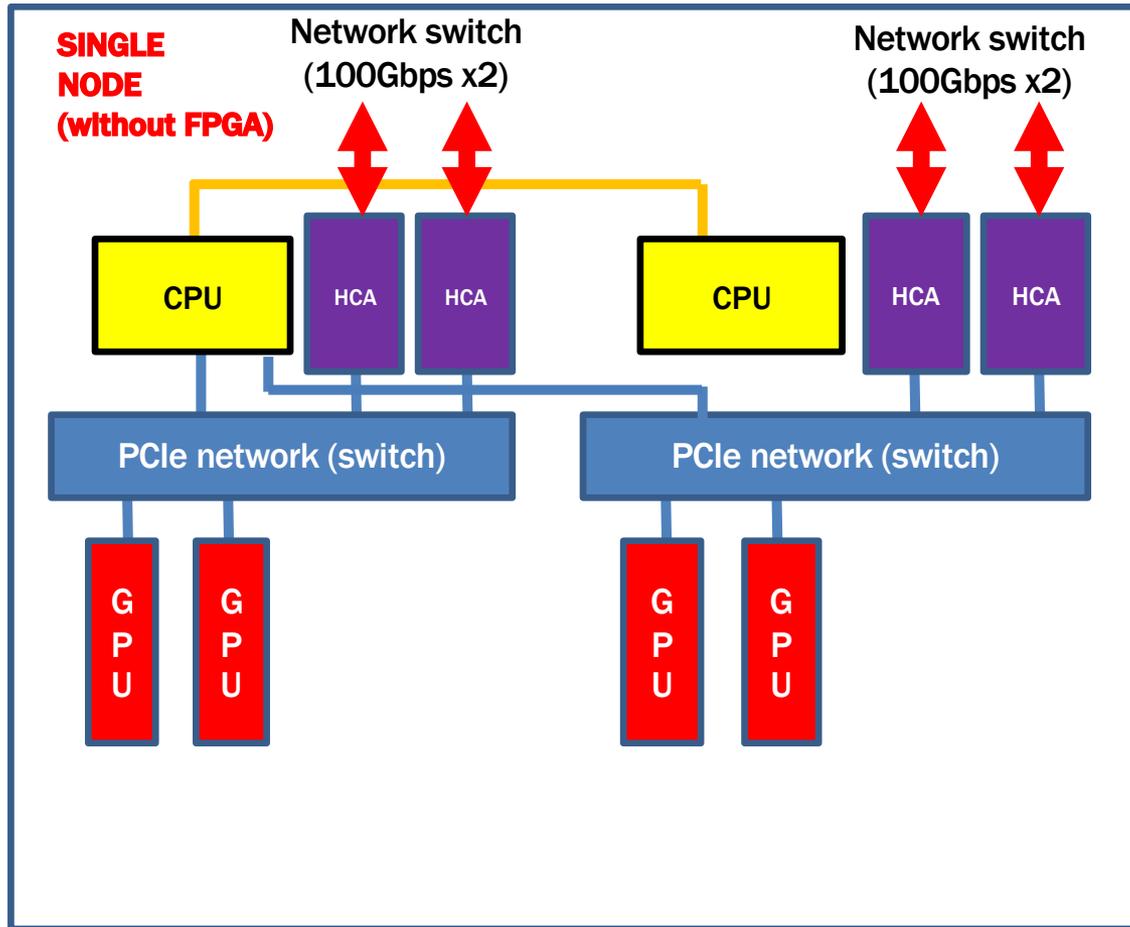
Cygnusの構成(4)

- ホーム領域
 - 全ノードからアクセス可能なNFSファイルサーバ (InfiniBand EDR結合)
 - “/home” として参照可能、ログインノードにログインした際のユーザーホームディレクトリとして利用
- 外部ネットワーク
 - 10Gbps x 4 port のL3ルータ
- ログインノード
 - 外部からssh接続可能なログインノードを3台設置
 - FPGAコンパイル用の専用ノードを別途用意

Albireo nodeのブロックダイアグラム

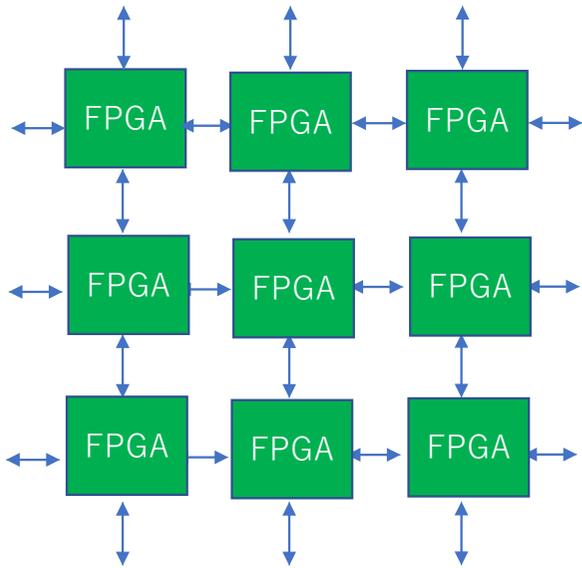


Deneb nodeのブロックダイアグラム



2種類の相互結合網

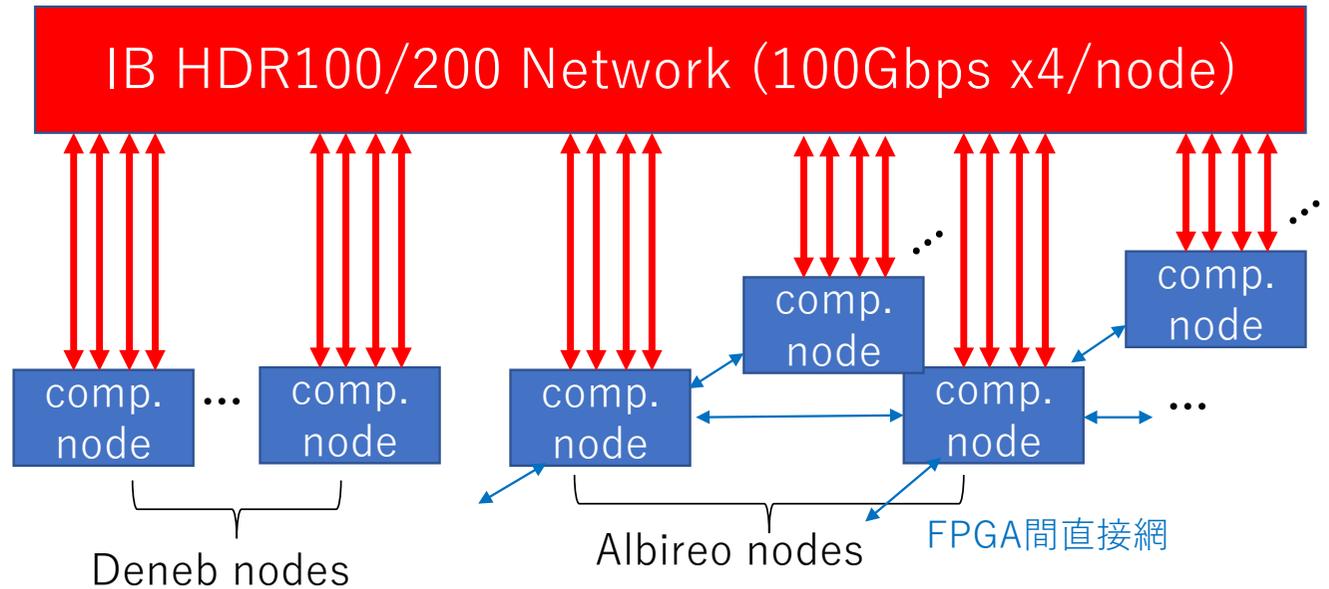
FPGA間直接網 (Albireoのみ)



Albireoノードの64台のFPGA (2 FPGA/node) は2次元トーラスネットワークによりスイッチなしで結合される

2018/12/03

全ノードに共通する並列処理/ストレージアクセス用相互結合網



全ノード共通の並列処理ネットワークはInfiniBandによる通常のネットワークで、MPI等による並列処理が可能。CPU及びGPUの通信に用いるが、FPGAからCPUを介しての通信も可能。

EPCC-CCS-Workshop2018@Edinburgh

Cygnusの性能・機能諸元

Item	Specification
Peak performance	2.4 PFLOPS DP (GPU: 2.2 PFLOPS, CPU: 0.2 PFLOPS, FPGA: 0.6 PFLOPS SP) ⇒ enhanced by mixed precision and variable precision on FPGA
# of nodes	78 (32 Albireo (GPU+FPGA) nodes, 46 Deneb (GPU-only) nodes) ⇒ 2 additional nodes will come, in total 80
Memory	192 GiB DDR4-2666/node = 256GB/s , 32GiB x 4 for GPU/node = 3.6TB/s
CPU / node	Intel Xeon Gold (SKL) x2 sockets
GPU / node	NVIDIA V100 x4 (PCIe)
FPGA / node	Intel Stratix10 x2 (each with 100Gbps x4 links/FPGA and x8 links/node)
Global File System	Lustre, RAID6, 2.5 PB
Interconnection Network	Mellanox InfiniBand HDR100 x4 (two cables of HDR200 / node) 4 TB/s aggregated bandwidth
Programming Language	CPU: C, C++, Fortran, OpenMP, GPU: OpenACC, CUDA FPGA: OpenCL, Verilog HDL
System Vendor	NEC