



IOWN Global Forum が推進するDCIのアーキテクチャについて

IOWN Global Forumは、IOWN APN (All-Photonics Network) の高速、低遅延といった特徴を活用した新たなICT基盤、「Data Centric Infrastructure (DCI)」を提案しています。本稿では、2023年3月に発表された「DCI機能アーキテクチャ2.0版」を基にIOWN Global Forumが定義するDCIのアーキテクチャを解説し、今のICT基盤が抱えるスケーラビリティ、パフォーマンス、低消費電力といった課題に対するDCIの優位性を説明します。IOWNが提案する新たなICT基盤、DCIへの理解を深めてください。

キーワード：#データセントリック、#IOWN Global Forum、#DCI

よしだ ひろし

吉田 寛

NTT 研究企画部門

はじめに

本誌2023年11月号では、『IOWN時代のデータ処理を支えるデータセントリック基盤とそのコンセプト実証』⁽¹⁾と題して、Data Centric Infrastructure (DCI) のデータ処理基盤やアクセラレータの概説と、映像解析をユースケースとしたコンセプトの実証について紹介しました。

本稿では、先月号に続かたちで、IOWN Global Forumが2023年3月に公開した「DCI機能アーキテクチャ2.0」⁽²⁾を中心に、DCIの全体像を解説します。ぜひ先月号の特集とともに読んでいただき、IOWNが提案する新たなICT基盤、DCIへの理解を深めてください。

現在のICT基盤の課題

■スケーラビリティに対する課題

ICT基盤は、大量の問い合わせに対する応答を必要とするオンライントランザクション処理や、大規模なデータ処理を実施するデータバッチ処理など、要求の異なるさまざまなデータ処理に対応する必要があります。また、例えばセール等のイベントに伴ってインターネット通販サイトに大量のアクセスが発生するなど、ICT基盤のデータ処理に対する要求が急激に変化し、数分または数時間の間に対応を迫られることもあります。現在のICT基盤にはこうした要求の変化に対して柔軟に対応するスケーラビリティが求められます。

■パフォーマンスに対する課題

例えば、仮想空間において各参加者の動きを周囲の人間に伝えたり、金融において高速なトランザクションを処理したりする場合など、応答時間について厳しい要件を持つデータ処理が存在します。現在のICT基盤では、こうした高い要求条件に対して、特にデータ転送の観点で十分対応することができません。

■エネルギー消費に対する課題

現在のICT基盤では、さまざまなボトルネックが存在します。例えばデータ転送がボトルネックとなっている場合、CPUはその処理能力の大半をデータ転送の待ち合わせのために消費することになります。また、最近では画像処理に特化することで高

速な処理が可能なGPU (Graphics Processing Unit) や、ネットワーク処理、セキュリティ処理等特定の計算に特化したDPU (Data Processing Unit) など、CPUと得意分野が異なるアクセラレータに処理を分担させることが一般化しています。計算リソースを効率活用し、かつ適切に役割分担させることは、ICT基盤全体としての電力効率の最適化を実現します。

DCIについて

DCIは、こうした課題に対処するためにIOWN Global Forumが提案する新しいICT基盤のアーキテクチャです。まず、図1をご覧ください。この図は、DCIを中心

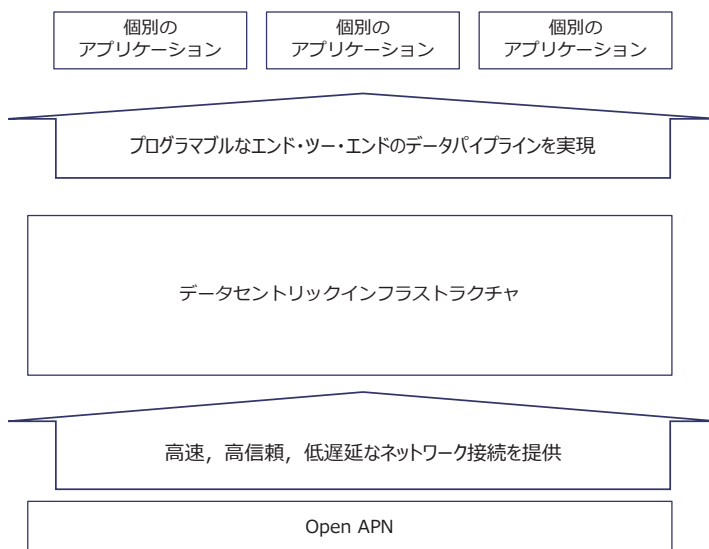


図1 DCIを中心としたICT基盤の全体像

としたICT基盤の全体像を表しています。このICT基盤の階層構造では、一番下にOpen APN (All-Photonic Network) が位置します。そして最上位には複数のアプリケーションが存在します。DCIは、Open APNが提供する大容量かつ低遅延なネットワーク環境を前提として、各アプリケーションに対して「プログラマブルなエンド・ツー・エンドのデータパイプライン」を提供し、各アプリケーションの機能を実現させます。

本稿では、「Open APN」「データパイプライン」の概念について説明したうえで、DCIがどのようにアプリケーションの機能を実現させるのかについて説明します。

■Open APNとは

Open APNとは、その全区間を光通信によって構成された高速なネットワークです。その特徴としては、大容量かつ低遅延のネットワーク、特に遅延が予測可能なネットワークを提供できることにあります。Open APNは、その大容量と低遅延を活かし、ネットワークに接続されたさまざまな機器、装置の間で、実際の距離や設置場所によらず「すぐ近く」にあるかのようにデータを転送することができます。Open APNの詳細、およびNTTの取り組みは、本特集記事『Open APNの詳細化、実用化に向けた取り組み』⁽³⁾でも紹介していますので、ぜひご参照ください。

■データパイプラインとは

次に、DCIが実現する「データパイプライン」の概念を説明します。データパイプラインとは、さまざまなアプリケーションが要求するデータの取得、処理、変換、提示等の機能を実現する統一されたストリームです。データパイプラインの例を図2に示します⁽⁴⁾。図2は、監視カメラの映像を、メタデータを基に集計・分析するシステムのデータパイプラインの例を示しています。この例では、監視カメラの映像が解析サーバに送られ、写っているものの情報などのメタデータが解析されます。メタデータについては集計のために分析サーバに送られ

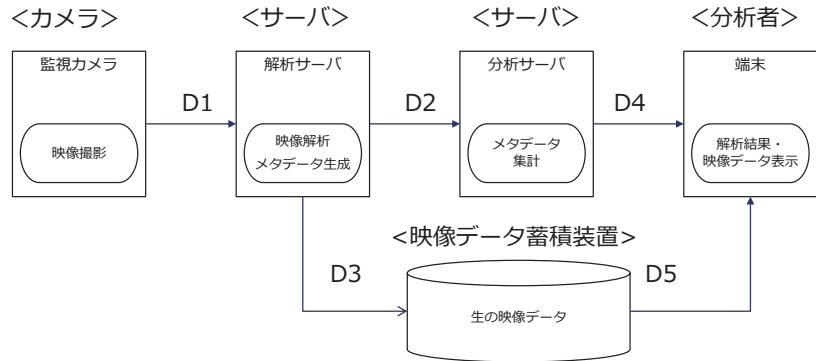


図2 監視カメラの映像をメタデータを基に集計・分析するシステムのデータパイプラインの例

るとともに、生の映像データがストレージに蓄積されます。分析サーバでメタデータについての集計が行われ、分析者は端末を用いてメタデータで解析を行うとともに、実際の映像データをストレージから参照することができます。丸みを帯びた楕円形は「データプロセッシング」、つまりデータの収集や解析といった自律的な動作の集合を意味します。プロセスを実行するために、プロセスの外側に長方形で示される「機能ノード」が存在します。ドラムで表現される「データベース/ストレージ」は典型的にはファイルシステムやデータベースであり、データを蓄積する機能を意味します。これらをつないだものが「データフロー」であり、「データプロセッシング」や「ストレージ」にデータを転送する機能を意味します。IOWN GFが「データパイプラインダイアグラム」としてこのデータパイプラインの表現方法を定義しました⁽⁴⁾。データパイプラインは、アプリケーションの機能を実現するためのリソースを抽象化したものになります。

DCIの各種リソース

では、DCIはどのように「データパイプライン」を実現するのでしょうか。それについて説明する前に、まずDCIが各種リソースをどのように管理しているかを説明します。

■機能カード、物理ノードとDCIクラスタ

DCIが管理する各種リソースの概念図を図3に示します。

「データプロセッシング」は、前述の通りデータパイプラインのうち、データの収集や解析を実施する機能を意味し、典型的にはCPUによって実現されています。また最近では、画像処理に特化することで高速な処理が可能なGPU (Graphics Processing Unit) や、ネットワーク処理、セキュリティ処理等特定の計算に特化したDPU (Data Processing Unit) なども注目されています。同様に「ストレージ」は、SSDやHDDによって実現されます。

DCIでは「データプロセッシング」と「ストレージ」を「ホストボード」と「機能カード」の2つに分類します。「ホストボード」は典型的にはCPUとメモリから構成され、一般的なサーバにおけるマザーボードに相当します。「機能カード」はそれ以外の部品、例えばGPU、DPUやネットワークインタフェース、あるいはストレージなどを指し、一般的なサーバにおけるNIC (Network Interface Controller) やグラフィックカード等の拡張ボードやSSD等のストレージに相当します。「ホストボード」と複数の「機能カード」を組み合わせたものが「DCI物理ノード」であり、DCI物理ノード内でホストボードと機能カードを接続するのが「ノード内インタコネクト」となります。現在サーバ等で広く採用されているPCI Expressはノード内インタコネクトの一例

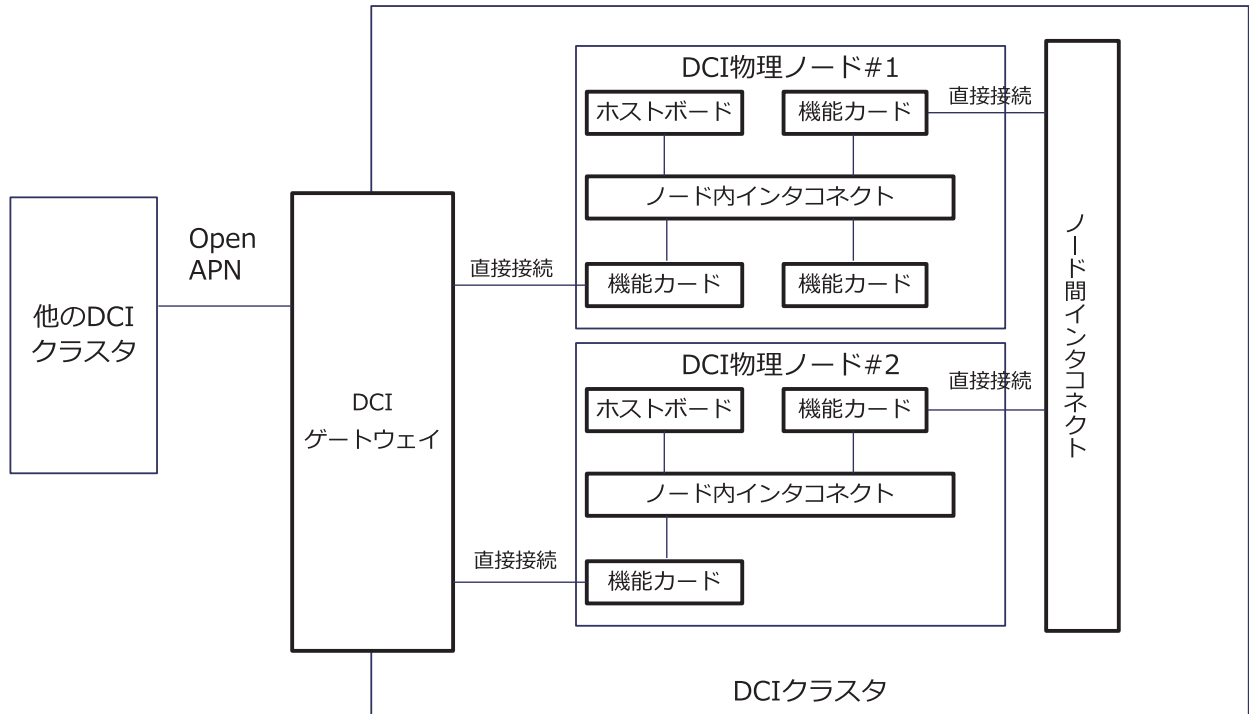


図3 DCIが管理する各種リソースの概念図

となりますが、CXL (Compute Express Link) など、より柔軟性の高い規格も検討されています。

さらに、このようなDCI物理ノードを複数集めて構成される計算インフラが「DCI クラスタ」です。DCI クラスタは複数の物理ノードから構成されますが、特徴的な機構として「ノード間インタコネク」 と「DCI ゲートウェイ」を持ちます。

ノード間インタコネクはDCI物理ノード間を接続するネットワークです。一般的なクラスタにおいてもクラスタを構成するサーバ間を接続するネットワークは存在しますが、ノード間インタコネクの特徴として、一部の機能カードは直接このノード間インタコネクにアクセスできます。このようにノード間インタコネクに直接アクセスできる機能カードを「ネットワーク対応機能カード」と呼びます。ネットワーク対応機能カードは、所属する物理ノードのノード内インタコネクを経由することなく、他の物理ノード内の機能カードやホ

ストボードと直接データ転送が可能です。NICは当然ネットワーク対応機能カードですが、DCIではGPUやストレージなどの機能カードの中にもネットワーク対応機能カードがあることを想定しています。

DCIゲートウェイは、DCIクラスタ間を接続するゲートウェイにあたり、他のDCIクラスタとOpen APNで接続されます。ネットワーク対応機能カードは、DCIゲートウェイに対しても直接アクセス可能であり、他のDCIクラスタに属する機能カードともOpen APNを経由して高速なデータ伝送を行うことができます。

■ LSNとデータパイプラインの実現

次に、DCIクラスタによって、アプリケーションごとにデータパイプラインを実現する仕組みについて解説します。データパイプラインを実現するためには、DCIクラスタ内の機能カードやホストボードをピックアップして、論理サービスノード (LSN: Logical Service Node) を構成します。

もしLSNが1つの物理ノード内の要素で

構成されるのであれば、データパイプラインのデータフローをすべてその物理ノードのノード内インタコネクによって実現することも可能です。しかしLSNが複数の物理ノードにまたがる場合、データフローを実現するためにはノード間インタコネクが必要になります。また、データパイプラインは所属するDCIクラスタが異なる複数のLSNによって構成することも可能です。この場合、LSNをまたがったデータフローはDCIゲートウェイを経由する必要があります。

データパイプラインの構成例を図4示します。今回DCIクラスタからは黄色く塗った機能カードA、B、CをピックアップしてLSNを構成することとしました。機能カードAはネットワーク対応機能カードではないので、ノード間インタコネクに直接接続することができず、一度ノード内インタコネクおよび機能カードXを経由してノード間インタコネクに接続します。

一方DCI物理ノード#2の機能カードB

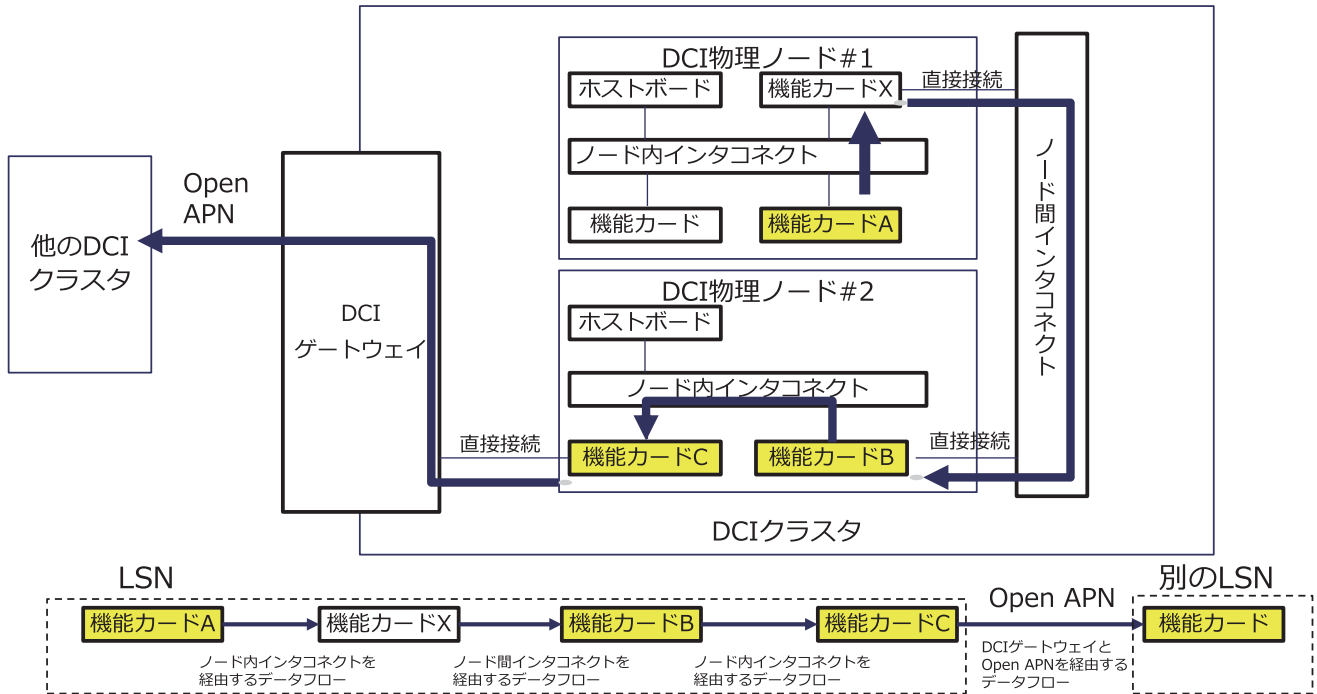


図4 データパイプラインの構成例

はネットワーク対応機能カードなので、ノード間インタコネクトから直接機能カードBに接続することができます。また機能カードCもネットワーク対応機能カードでDCIゲートウェイにつながっているため、機能カードCから直接DCIゲートウェイを経由して他のDCIクラスタにある別のLSNに接続することが可能となります。

このようにDCI物理ノードやDCIクラスタをまたがった柔軟なデータパイプラインを設定できることがDCIの特徴となります。

DCIの課題と解決

ここまで、DCIがデータパイプラインを提供する仕組みについて説明しました。ではDCIによって、前述したスケーラビリティ、パフォーマンス、低消費電力の課題がそれぞれどのように解決されるのかについて次に述べます。

■スケーラビリティに対する課題の解決

従来のICT基盤では、計算を行うコン

ピューティングと、データを転送する広域ネットワークが独立して進化してきたため、物理ノードやクラスタをまたがった柔軟なスケールが困難でした。しかしDCIではネットワーク対応機能カードがノード間インタコネクトを経由して直接相互接続されるため、データ転送がボトルネックとなることがなくなります。さらにDCIゲートウェイを経由して異なるDCIクラスタのLSNまでを容易に統合できるため、物理ノードやクラスタをまたがったリソースを統合によるスケーラビリティを実現します。

■パフォーマンスに対する課題の解決

DCIではネットワーク対応機能カードがノード間インタコネクトを経由して直接相互接続されることから、それらの間のデータ転送はCPUを経由することなく極めて高速に実行されます。そしてDCIクラスタをまたがるデータ転送もOpen APNを介することで高速な実行が実現します。このことから、応答時間について厳しい要件を求められるアプリケーションに対しても、

十分な応答性能を提供することが可能となります。

■エネルギー消費に対する課題

DCIではデータ転送におけるボトルネックが少ないため、CPUがデータ転送を無駄に待ち続けることがありません。また、計算の要求に応じて、GPUやDPUといったアクセラレータを柔軟に選択することができます。このことから、計算リソースの利用効率が改善するとともに、システム全体としてのエネルギー消費を最適化することができます。

KEY VALUES AND TECHNOLOGY EVOLUTION ROADMAP と今後のDCIの発展

最後に、2023年8月に発表されたKEY VALUES AND TECHNOLOGY EVOLUTION ROADMAP⁽⁵⁾を参照しつつ、DCIに関連する各技術が今後の発展を示すロードマップ上にどのように表現されたかにつ

表 IOWN コンピューティングの発展ロードマップ

要素	Phase 1 (現在)	Phase 2 (3-5年)	Phase 3 (by 2030)
計算リソースの割り当て管理単位	サーバレベル	計算デバイスレベル	計算デバイスレベル
エネルギー効率	測定を実施 ワークロード/イベント/トラフィックに依存しない	正確に測定され、制御されるワークロード/イベント/トラフィックと密接に一致	フェーズ2と比較して大幅な削減を目指す
データ転送	ホストボードによって支援される	XPU (GPU, DPU) 間の直接接続	XPU (GPU, DPU) 間の直接接続
キャッシュコヒーレンスメモリスペース	COTSスケール	ラックスケール (CXL 経由)	ラックスケール (CXL 経由)
RDMA	データセンタ内での適用	データセンタをまたがった適用 (by RDMA over APN)	データセンタをまたがった適用 (by RDMA over APN)
論理サーバの構成	サーバレベル	モジュールカードレベル (PCIe カード, CXL カードなど)	チップレベル
セキュリティ	現在の暗号 (例: RSA, DH)	ポスト量子暗号	ポスト量子暗号にくわえ、そのゼロトラスト概念による管理
計算リソースの構成時間	数時間	数分	数秒
非常に厳しい時間的制約や決定論的処理時間を必要とするアプリケーション	実現不可	ms レベルで決定論的に処理時間を担保可能	ms レベルで決定論的に処理時間を担保可能
ストレージ/データ移動	ローカルレベルでの移動	地域レベルでの移動	国家レベルでの移動

いて説明します。

まず現在サーバレベルとなっている計算リソース割当管理単位を、Phase2で要素単位、つまりホストボードや機能カード単位にすることを実現します。また、Phase2では、前述したネットワーク対応機能カードにより、GPUやDPU間で直接伝送を実現します。

また、計算リソースの構成時間について現在は1時間単位で掛かっていますが、本稿で述べたLSNの実現により、Phase2では分単位、Phase3では秒単位での構成を可能としています(表)。

結論とまとめ

本稿では、DCI機能アーキテクチャについて概要を説明しました。

IOWNがデータ駆動型社会の高度なITを活かすためには、「QoSが管理された高速ネットワーク」を前提として、計算基盤とネットワークを融合させた「新たなICT基盤」が求められており、IOWNのDCI

はそれにあたるものです。

なお、本稿では文字数の都合から、DCIを支えるいくつかの重要な概念、例えば「インフラストラクチャオーケストレータ」や「機能別ネットワーク」「リソースプール」について説明することができませんでした。これらについては、IOWN Global Forumが公開しているDCI機能アーキテクチャ2.0版に記載がありますので、ぜひご参照ください。

参考文献

- (1) 樽林・石崎・Sampath・Christoph・水野：“IOWN時代のデータ処理を支えるデータセントリック基盤とそのコンセプト実証,” NTT技術ジャーナル, Vol. 35, No. 11, pp. 47-52, 2023.
- (2) https://iowngf.org/wp-content/uploads/2023/04/IOWN-GF-RD-DCI_Functional_Architecture-2.0.pdf
- (3) 海沼・武田・可児・西沢：“Open APNの詳細化、実用化に向けた取り組み,” NTT技術ジャーナル, Vol. 35, No. 12, pp. 15-19, 2023.
- (4) Figure B.2-2: An Example of Data Pipeline Diagram (Big Mail System): <https://iowngf.org/wp-content/>

- uploads/formidable/21/IOWN-GF-RD-RIM-for-AM-Use-Case-1.0.pdf
(5) https://iowngf.org/wp-content/uploads/formidable/21/IOWN-GF-RD-KVaTE_Roadmap-1.0-1.pdf



吉田 寛

DCIについてはまだまだIOWN Global Forum内で議論中の概念であり、併せてフォーラムメンバがPoCを通じた評価を進めております。DCIがもたらす新たなICT基盤にぜひご期待ください。

◆問い合わせ先

NTT 研究企画部門
IOWN 推進室
E-mail iown-info@ntt.com