Pluribus と Dell EMC に搭載される Apache Hadoop の高速化

Pluribus Open Netvisor® Linux と Dell EMC Open Networking による Big Data 処理の利点

2016年２月





概要

企業が日々作り出し、時間とともに蓄積していく大規模なデータの中にビジネス価値が隠れているということは、ますます明らかになってきています。 その結果、Apache Hadoop のようなインフラストラクチャを使用する Big Data の処理は、一般的な企業のデータセンターにおいて主流となっています。ほとんどの場合、Hadoop のジョブ時間はビジネスの結果に直接関係しており、Big Data の処理は、リアルタイムもしくはリアルタイムに近い形で実行される事が期待されています。残念ながら Hadoop のジョブ結果は、多数のタスクがすべて完了した場合にのみ利用可能であり、並列したタスクの実行における不均衡に対して非常に敏感です。さらに、Hadoop の処理速度に影響を与えるその他の要因がいくつかあります、例を挙げるとデータの摂取率、マッパーデータアクセス、中間マッパーの結果整理、データへのアクセス回数低減、HDFS の初期複製、複製リカバリなど、ネットワークに依存するものです。これらのすべてがネットワークインフラに特有の要求をもたらし、同様に対処的なコントロールを伴うきめ細かな性能パラメーターを可視化する必要があります。

Dell EMC Open Network スイッチ上で稼働する Pluribus OpenNetvisor® Linux (ONVL) は、シングルポイントでネットワークファブリック全体の管理アクセスを使用して、この優れたレベルの可視性と制御を提供します。 一元化されたネットワークのファブリック全体の管理アクセスを使用して、優れたレベルの可視性と制御を提供します。ファブリック全体の統合されたビューにアクセスすることにより、CLI ベースの管理と API のプログラムによるアクセスが大幅に簡素化され、使いやすさとシンプルなアプリケーションインテグレーションを実現します。

Pluribus Open Netvisor Linux (以下 ONVL) によって提供される詳細な可視性により、Hadoop クラスタによくみられる危機的な状況を明らかにする事が出来ます。これらの結果は対応処置のためにリアルタイムで、もしくは遅延したジョブの根本的な原因を特定するために事後にも取り込むことが可能です。 Pluribus ONVL のテレメトリは、ネットワークリソースに起因するパフォーマンスの低下とアプリケーションレベルの不平衡を判別する便利なツールです。Hadoop ジョブの全体的な実行時間は、複数のフェーズに依存します。ONVL のアプリケーションフローの可視性は、クラスタ内のあらゆるフローのパフォーマンスの特性を提供します。

オープンなネットワーク基準と Dell EMC Open Networking マーチャントシリコンスイッチラインに基づいて、Pluribus/Dell EMC のソリューションが共同で、業界最高レベルの経済性を備えた Hadoop インフラストラクチャにおける前例のないファブリックの可視性と制御を提供します。

目次

はじめに 4

Dell EMC Open Networking スイッチ 4

Pluribus Open Netvisor® Linux の 革 新 4

ビジネスで要求される Big Data の処理 6

Apache Hadoop の ネットワーク設計に関する留意事項 6

Big Data の処理に関する 一般的なネットワークアーキテクチャ 6

Hadoop クラスタにおける一般的なデータパス 7

Dell EMC 向け Pluribus ONVL による Apache Hadoop ネットワーク運用の強化 8

Dell EMC 向け Pluribus ONVL による Apache Hadoop ネットワークチャレンジへの取組み 8

Pluribus ONVL の導入オプション 9

グリーンフィールドラックの Dell EMC 向け Pluribus ONVL によるファブリック設計 10

Dell EMC 向け Pluribus ONVL による リーフスイッチ導入 11

Pluribus ONVL 仮想化中心型 ファブリック (VCF™) 11

プログラマビリティ 12

可視性と制御：vPort と vFlow 12

統合ネットワークおよびアプリケーションフローテレメトリ 13

ホストロケーター 13

アプリケーションフローテレメトリ 13

結論 14

膨大なデータフローを交換、監視、または制御するニーズをサポートするため、Dell EMC Open Networking (以下 ON) 向け Pluribus Open Netvisor® Linux (以下 ONVL) は、サーバクラスタテクノロジーに一部基づいた、仮想化中心型ファブリックアーキテクチャを開発しました。

はじめに

Big Data とは、従来のデータ処理アプリケーションでは不十分な、大規模または複雑なデータ集合の処理のことです。目標は、データから価値ある情報を抽出することにより、予測分析や他の高度な用途に使用することです。これには、情報のプライバシーを保護しながら大量のデータを収集、転送、分析、検索、可視化、保存、または共有することが必要となります。 Big Data の処理における正確性はより効率的な意思決定につながり、運用効率の向上、コストの削減、リスクの低減を可能にします。データアプリケーションには気象学、ゲノミクス、生物学、環境研究の分野における科学的研究を促進するだけでなく、ビジネスや犯罪防止の動向を把握するために新しい相関関係を発見する可能性なども含まれています。

Apache Hadoop のような一般的な Big Data のフレームワークでは、多階層処理向けに整理された一連のサーバを利用しています。一般的な Map-Reduce アルゴリズムでは、サーバマップを予備的な "mapping" タスクに専念させています (例えば、各ネーム毎に1つのキューを割当て、ファーストネーム毎にスチューデントをソート)、次に "reduce" タスクを実行し、最終結果を計算するために (この場合、各キュー内のスチューデント数を数え、ネームの頻度を算出)、中間結果を第2のサーバセットに送信します。

パイプラインとして実行されるとジョブデータはそのソースから取り込まれ、複数の複製が分散ファイルシステムに保存されます。マッピング後、中間結果は分散ファイルシステムに保存され、reduce フェーズに備えます (この操作はシャッフルと呼ばれます)。 Reduce 後、実行される分析によっては出力が大量のデータ (純粋なソートタスクの場合のように) になる場合もあれば、比較的小さなデータセット (統計分析の場合のように) になる場合もあります。

膨大なデータフローを交換、監視、または制御するニーズをサポートするために、Pluribus は、Dell EMC Open Networking (以下 ON) スイッチ上で作動する Open Netvisor® Linux (以下 ONVL) を開発しました。Pluribus ONVL を搭載した Dell EMC ON イーサネットスイッチは、外部コントローラを必要とせずに、前例のない洞察力、機敏性、またはセキュリティを提供する単一の仮想化中心型ファブリックに統合されています。 Pluribus ONVL と Dell EMC ON スイッチは協業して、ネットワーク OS をベースにした従来の Linux の役割をはるかに超えた Apache Hadoop の導入を進めていきます。

このホワイトペーパーでは、Pluribus ONVL と Dell EMC ON スイッチが協業して、ネットワーク OS をベースにした従来の Linux の役割をはるかに超えた Apache Hadoop の導入を、どのように進めていくのかを説明しています。

Dell EMC Open Networking スイッチ

Broadcom の最新のチップセットをベースにした Dell EMC ON スイッチは費用対効果に優れ、ノンブロッキング、そして予測可能な低レイテンシを備えたスパイン/リーフ型アーキテクチャを実装するように設計されています。これにより、ワークロード管理とネットワークの俊敏性が劇的に向上します。Dell EMC ON のスパイン/リーフ型インフラストラクチャは移行、big data の分析、クラウド、および VDI アプリケーションを備えた大規模な仮想マシンプールに特に適しています。

Pluribus Open Netvisor® Linux の革新

Pluribus Networksは、業界で最もプログラマブルで、ネットワーク OS をベースにしたオープンソースである Open Netvisor Linux (以下 ONVL) を通して、ソフトウェア定義型ネットワーク (SDN) を発展させています。ONVL はフローの理解、ビジネスニーズに迅速に対応、そしてデータを保護するための実績のあるアプローチである Pluribus 仮想化中心型ファブリック (VCF™) アーキテクチャに基づいています。

リーフ型のトラフィックは常に2つのリンクを横断し、全てのスパイン型にわたって均等に配置されており、従来の3階層設計と比較した場合、性能、可用性、および予測可能なエンド・ツー・エンドのネットワークの待機時間という点において優れています。

Pluribus ONVL は、Linux の利点と分散コントローラファブリックとを同時に実現します：

- ファブリック全体のコマンドラインインターフェイス（CLI）は、管理の一元化によって俊敏性と自動化を実現するために、プログラマビリティ（C、RESTful API)、OpenStack インテグレーション、DevOps ツール (Anabilities など) と組み合わされています。
- 統合されたタップレステレメトリおよびアプリケーション性能の監視は、ネットワークアプリケーションフローとクライアント・サーバ接続をトラックする強力なファブリック全体の性能に基づいています。
- トラブルシューティングとポリシーの適用を簡素化するために、ファブリック全体のエンドポイントデータベースまたは vPort テーブルは、すべてのスイッチで認識されアクセス可能です
- QoS とセキュリティにより、vFlow プログラミングによる精度の高いフローフィルタリングとセキュリティ処理が可能です

Pluribus ONVL と Dell EMC ON スイッチのポートフォリオの組み合わせは、クラス最高のハードウェアスイッチにおける経済性を提供します。 導入の柔軟性は CLI を備えた Pluribus ONVL の レイヤ2/レイヤ3 (L2/L3) スタックによって保証され、既存のネットワークインフラストラクチャとの完全な相互運用性を提供します。

Pluribus ONVL の革新	利点
分散型コントローラファブリック	過去20年間、人々は管理上のオーバーヘッドとエラーを経験することにより、CLI 経由でボックス毎にルータとスイッチを設定しました。Pluribus ONVL は、Pluribus ONVL スイッチ間のクラスタ、つまりサードパーティ製のネットワークデバイス間でもファブリッククラスタを作成します。 1つの論理管理ポイントを介して単一の安全なログインを使用することにより、単一のスイッチであるかのようにネットワークを構成および修正する機能が提供されます
アプリケーションフローの可視性	IT 企業はアプリケーションの導入と操作の容易さ、アプリケーションの性能、およびアプリケーションが計算、ストレージ、ネットワーク、および仮想化のリソースを最適に活用できる 'アプリケーションの品質' を確保する必要があります。Pluribus ONVL のファブリックは分析機能を統合し、ホストおよび仮想マシンレベルでのあらゆるアプリケーションフローをリアルタイムで表示するために分析機能を統合します。完全な画像を作成するために、待機時間、パケットロス、および接続間隔をすべて使用できます。
vPort	レガシーなイーサネットスイッチは、自身のポートに接続されているデバイスのみを認識しますが、Pluribus ONVL は、各エンドポイントを vPort にマッピングします。vPort は、ファブリック全体のデータベース内のレコードで、特定の時間範囲にわたってファブリック全体のポート属性を追跡する機能を提供します。vPort レコードは、IP アドレスから MAC アドレスへのマッピングのように、単純な情報を保存することは出来ますが、エンド・ツー・エンドのアプリケーショントラブルシューティングを容易にするために、ホスト名、OS、およびハードウェア構成として追加のエンドポイントメタデータを保存して使用することが可能です。仮想マシンがあるサーバから別のサーバに移行すると、vPort レコードが更新され、関連付けられたポリシーはエンドポイントに従います。
vFlow	セキュリティを強化し QoS を適用するには、物理ポート、L2、L3、および L4 情報の任意の組み合わせに対して、ネットワークフローを制御する必要があります。 Pluribus ONVL vFlow は、ネットワークトラフィックの精度の高い可視性と制御を提供します：コントローラなしでも、スイッチごとではなくファブリック全体に設定が適用されます。フロープログラミングは整合条件に基づいており、例えば送信元/宛先IP、プロトコル、ポートなどで、ドロップ、有限帯域幅、リダイレクト、ミラー、トラップ、VLAN/VXLAN の変更、カプセル化または CPU へのコピーを含みます

ビジネスで要求される Big Data の処理

近年各企業は、Big Data として知られ、実用的なビジネスインテリジェンスが含まれている膨大な量のデータにアクセスできるようになっており、一定の時間内に合理的なコストでデータを処理することが可能になっています。戦略的価値を高めるために、実行時間をリアルタイム操作に近いレベルにまで短縮する傾向があります。

Apache Hadoop は、Big Data 処理に広く採用されているオープンソースのアプリケーションスイートです。既存のコンピューティングリソースを使用する Map-Reduce アルゴリズムを実装しています。また、計算ノード自体にデータを保存する、高度に並列化し分散されたファイルシステムである Hadoop のファイルシステムを含みます。

Hadoop の導入は一般的に組織内の複数の部門をまたがって共有されるため、非常にビジーな状態です。予約済みの時間に Hadoop のタスク結果を利用できるようになるにつれて、各部門の依存度はますます高まります。 例を挙げると、大型流通店で突然必要になった商品リストを決定するために出力結果が使用される場合です。 実行スピードを速くするため、Hadoop のインフラストラクチャはバッチを多数のジョブに分割して並列実行し、並列処理された各ジョブの結果を統合して最終結果を計算します。

Hadoop インフラストラクチャによって提供される高度な並列性は、タイムリーに実用的な結果を導き出しますが、ほんの一部のジョブの処理が他のピアよりも著しく長くかかる場合にでも問題が発生します。 処理が開始されると、最後のジョブが完了した時にのみ処理タスクが完了します。 したがって、システムはタスク全体の継続時間に著しく影響を及ぼす処理の遅いジョブを待たなくていいように、バランスをとる必要があります。 これらの理由から、Hadoop インフラストラクチャは、あらゆるパフォーマンス上の問題を特定するために継続的な監視下に置かれています。

Apache Hadoop のネットワーク設計に関する留意事項

ビジネス要件は、アーキテクチャから運用および管理機能に至るまで、ネットワーク設計の留意事項を向上します。

Big Data の処理に関する一般的なネットワークアーキテクチャ

現代のデータセンターネットワーク POD アーキテクチャは、図1に示すようにスパイン型とリーフ型トポロジに基づいています。トップオブラック (ToR) スイッチはリーフスイッチで、それらはスパインスイッチとしてコアスイッチに接続されています。リーフスイッチは互いに接続されておらず、スパインスイッチのみがリーフスイッチ (もしくはアップストリームのコアスイッチ) に接続しています。このアーキテクチャでは、リーフスイッチからのアップリンク数は、スパインスイッチの数に等しくなっています。

リーフ型のトラフィックは常に2つのリンクを横断し、全てのスパイン型にわたって均等に配置されており、従来の3階層設計と比較した場合、性能、可用性、および予測可能なエンド・ツー・エンドのネットワークの待機時間という点において優れています。可用性は、計算処理ノードとリーフスイッチ間、およびリーフスイッチとスパインスイッチ間のリンクアグリゲーションによって提供されています。リーフスイッチとスパインスイッチは、マルチシャーシリンクアグリゲーション (MLAG) テクノロジーを使用しており、ペアで配置することによりあらゆる単一障害点を取り除くことができます (図1には示されていません)。

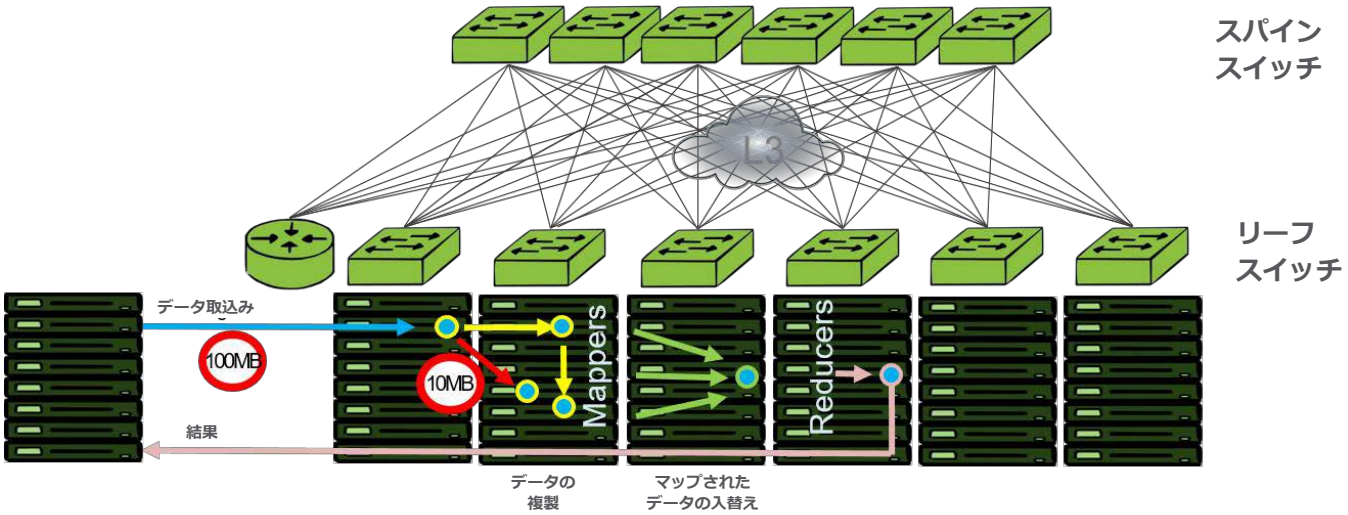


図1 - Apache Hadoop クラスタのスパイン/リーフ型ネットワークアーキテクチャ

ほとんどの場合、Apache Hadoop は1つ以上のラックのクラスタとして導入されています。各ラックは IP サブネットにマップされます。拡張性のある導入では、ラック間でのデータの取り込みと出力、およびファイルシステムの動作はレイヤ3の境界にまたがって実行されます。重要な要件はシステム全体の性能を最適化し、同じサーバ内もしくは同じラック内でのデータ処理をローカライズすることです。

Hadoop クラスタにおける一般的なデータパス

クラスタは複数のジョブを同時に処理することが可能です。各ジョブはデータの取り込み、結果の出力、map と reduce の処理ステージ間のデータの入れ替え、もしくは多数の map と reduce タスクの実行を異なるフェーズ内に置くことが可能です。

ネットワークトラフィックは、プライマリトラフィック (ジョブを完了するために不可欠) とセカンダリトラフィックに分かれており、Hadoop ファイルシステム (HDFS) のような Hadoop コアインフラストラクチャの完全性を維持するために必要なものとなっています：

- North-South トラフィックとして知られるクラスタ内外のプライマリトラフィックは、膨大な量のデータを取り込み結果を出力します。
- East-West トラフィックとして知られるクラスタ内のトラフィックは、mapper から reducer への中間結果を入れ替えるために、プライマリトラフィックのコンポーネントを持っています。このコンポーネントは、タスクの持続時間に重要です。これは実行される分析に依存し、多対多、もしくは多対少の mapper と reducer に作用します。
- East/West セカンダリトラフィック (例えば、新しいデータの複製を作成するのか、複製障害後に作成するのか) は、ジョブの実行 (ロングテール) に支障をきたす可能性があります。セカンダリトラフィックは、特権ジョブの実行を監視する必要があります。

最適なパフォーマンスを得るために、Hadoop インフラストラクチャでは、ネットワークトポロジとステータスに関する信頼できる情報が必要となります：

- ローカリティが重要 - ジョブのワークロードとジョブデータを同じ場所に配置する必要があります (同じサーバが理想的；同じラックでも許容範囲)。したがって信頼性の高い、ネットワーク全体におよぶトポロジ情報の可用性が必要となります。
- セカンダリトラフィックは、プライマリトラフィックの性能を維持するために調整しなければなりません。したがって、トラフィックフローに関するリアルタイム情報が必要となります

Dell EMC 向け Pluribus ONVL による Apache Hadoop ネットワーク運用の強化

Dell EMC Open Networking スイッチ向けの Pluribus ONVL は、データセンターのネットワーク設計者と管理者に特有の機能を提供します：

特有機能付き Dell EMC 向け Pluribus ONVL	Apache Hadoop 導入の利点
機能豊富な L2/L3 およびマルチキャスト	新しい Hadoop POD を導入することにより、柔軟な設計オプションを提供するレガシーデバイスとの相互運用性を保証
コントローラ不要の CLI および/または API を介しファブリック全体の管理一元化：ファブリック全体がシングルスイッチとして認識	各エンドポイントの管理はアタッチメントポイントの知識を必要としないため、最大限の機敏性を実現 Hadoop スイートとの統合による自動化は、インフラストラクチャの変更に関して独立
統合された、またはタップレスのファブリックテレメトリにより、エンド・ツー・エンドのアプリケーションフローを可視化	ダイナミックな環境に適した強力なトラブルシューティングツールはネットワークテレメトリーの拡張セットが内蔵されており、大規模なデータフロー、ファイルシステム動作、および結果の入れ替えに関するインジケータを提供
vPort ホスト/エンドポイント ID およびロケーションデータベース	トラブルシューティングとアンカーポリシーを簡素化するための、ファブリック全体にわたるエンドポイントの統合されたビュー
データフローを選択、および制御するように設定された vFlow コマンド	管理者は個々のフローまたはフローのクラスを識別して、QoS およびセキュリティポリシーをファブリック全体に適用

Dell EMC 向け Pluribus ONVL による Apache Hadoop ネットワークチャレンジへの取組み

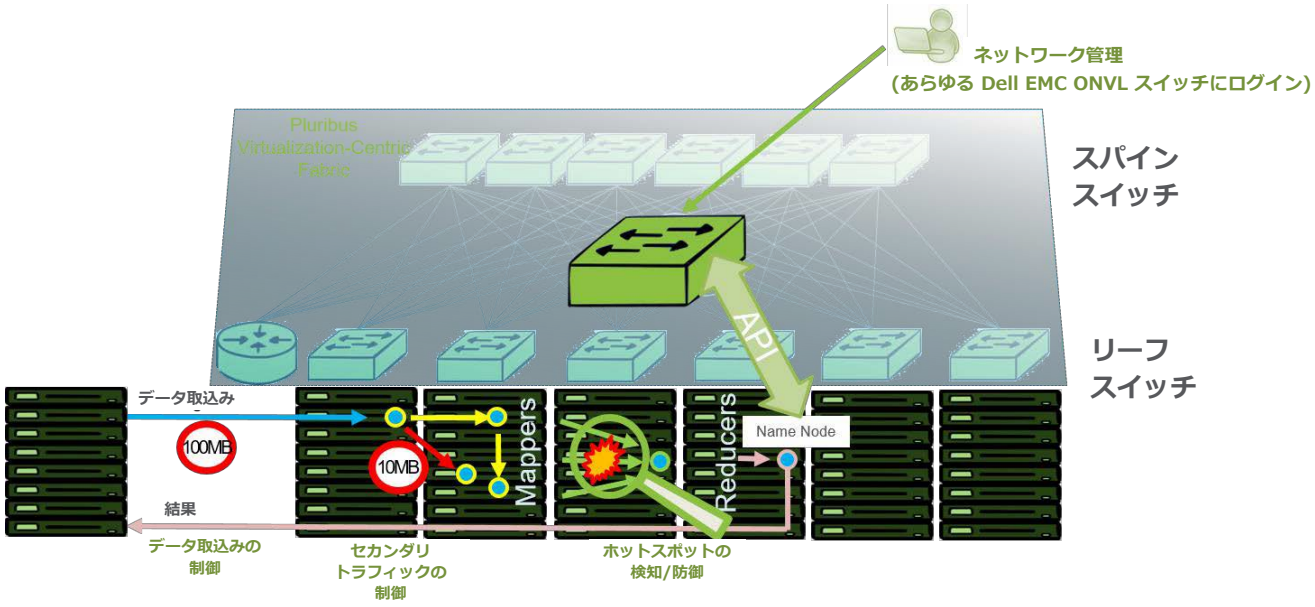


図 2 - Dell EMC ONVL ファブリックに導入されている Apache Hadoop

下記の表は Apache Hadoop 動作上の Pluribus ONVL の利点を要約したものです：

Hadoop の特徴	挑戦	Dell EMC ソリューション向け Pluribus ONVL
新しいクラスタの導入	異機種混在インフラストラクチャのデータセンター、レガシーコア/スパイン型で通信する Hadoop ラック	Pluribus ONVL は新機種または異機種混在の導入において、ファブリック全体の可視性と制御を提供 (レガシー ToR または Pluribus ONVL ToR を搭載したラックの組み合わせを含む)。CLI の使用により、従来のネットワーク環境に簡単に挿入可能
データの取込みと出力結果	アクティブなジョブ性能を維持しながら、データフローを最大化	データ取り込みと出力結果に、ファブリック全体で使用される許容帯域幅を設定
ジョブの配置	タスクの実行とデータのローカライゼーションが古いトポロジ情報に基づいている可能性があるため、一部のジョブが他のジョブ完了よりも時間がかかる可能性 (ロングテール)	最新のトポロジを読み込み、同じラック内のサーバを特定する Hadoop 向け vPort API
中間結果の入れ替え	多くの mapper が少数の reducer と通信する場合の輻輳の防止および/または検出	中間結果を運ぶエンド・ツー・エンドのデータフローを監視するための輻輳分析
HDFS プライマリ	ローカライズされていないトラフィックを検出するためのトラフィック量の監視	ファブリック全体のアプリケーションフローおよびクライアント・サーバ接続テレメトリ
HDFS セカンダリ	セカンダリトラフィック量を許容レベルに保つ	ファブリック全体の帯域幅管理により、セカンダリトラフィックを動的に制御
全段階	ネットワーク管理者は、ノード名とロールの可視性を持たない	ONVL vPort の "ディレクトリ" ノード名とロールにプログラミングするためのファブリック全体のプログラム Hadoop 向けAPI

次項ではこれらの機能と、それらが Apache Hadoop クラスタ導入の利点にどのように利用出来るかを説明します。

Pluribus ONVL の導入オプション

ファブリック内の各スイッチは、従来の L2/L3 ネットワークプロトコル (例えば、STP、BGP、OSPF) を用いてネットワークトポロジの状態を局部的に計算し、レガシーインフラストラクチャとの相互運用性を提供します。Pluribus ONVL を搭載した Dell EMC Open Networking スイッチは、全てのラックまたはラックのサブセットに TOR としても導入、もしくはスパインスイッチとしても導入出来ます。これにより、ネットワーク設計に最大限の柔軟性がもたらされます：

- 新しい ToR：スパインスイッチとリーフスイッチを含む Pluribus ONVL を搭載した Dell EMC ON スイッチから構成されるネットワーク POD を導入
- 異機種混在/サードパーティ製 TOR：各ラックに Pluribus ONVL を搭載した ToR Dell EMC ON スイッチを導入
- 異機種混在/サードパーティ製 POD：サードパーティ製デバイスで構築されたレガシーネットワーク内の Hadoop のような特定のアプリケーションをサポートするために、Pluribus ONVL ToR を搭載した Dell EMC ON スイッチ内蔵のラックセットを挿入

サードパーティ製のレガシースイッチが存在する場合、全ての Pluribus ONVL スイッチは、管理インターフェイスまたはインバンドを介して ONVL ノード間に L2 または L3 接続を提供することにより、シームレスなファブリックを形成できます。

Pluribus ONVL は、Hadoop クラスタを導入する際に、設計オプションを提供します：Pluribus ONVL ToR を採用している各ラックは、データセンターインフラストラクチャと相互運用する一方で、全てのクラスタノードに対して制御ポイントの一元化とエンド・ツー・エンドのテレメトリを提供します。

管理者は、ファブリック全体のポリシーを導入することもできます。たとえば、クラスタ内のセカンダリトラフィックのスループットを制限し、ジョブの性能を維持するような場合です。

グリーンフィールドラックの Dell EMC 向け Pluribus ONVL によるファブリック設計

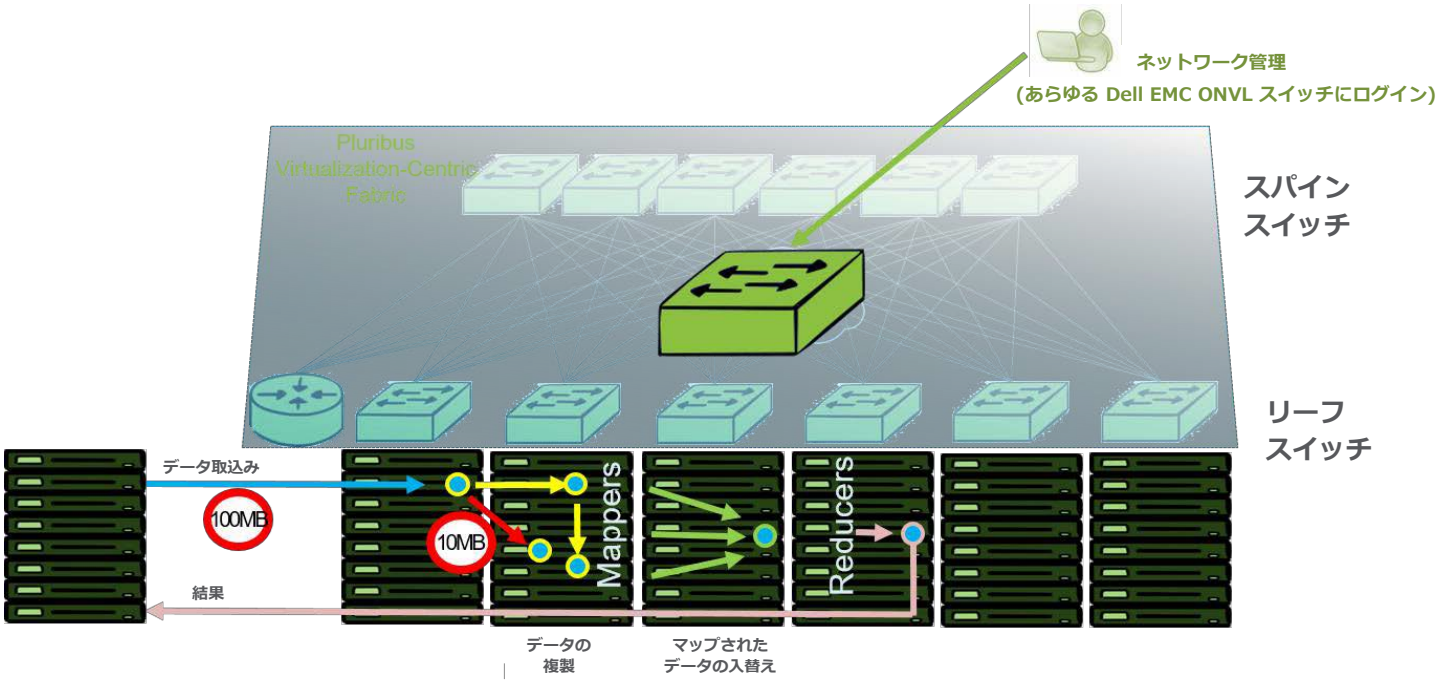


図 3 – Dell EMC 向け Pluribus ONVL 搭載の Hadoop クラスタの導入：ファブリックはシングルスイッチのように制御の一元化提供

Dell EMC 向け Pluribus ONVL による リーフスイッチ導入

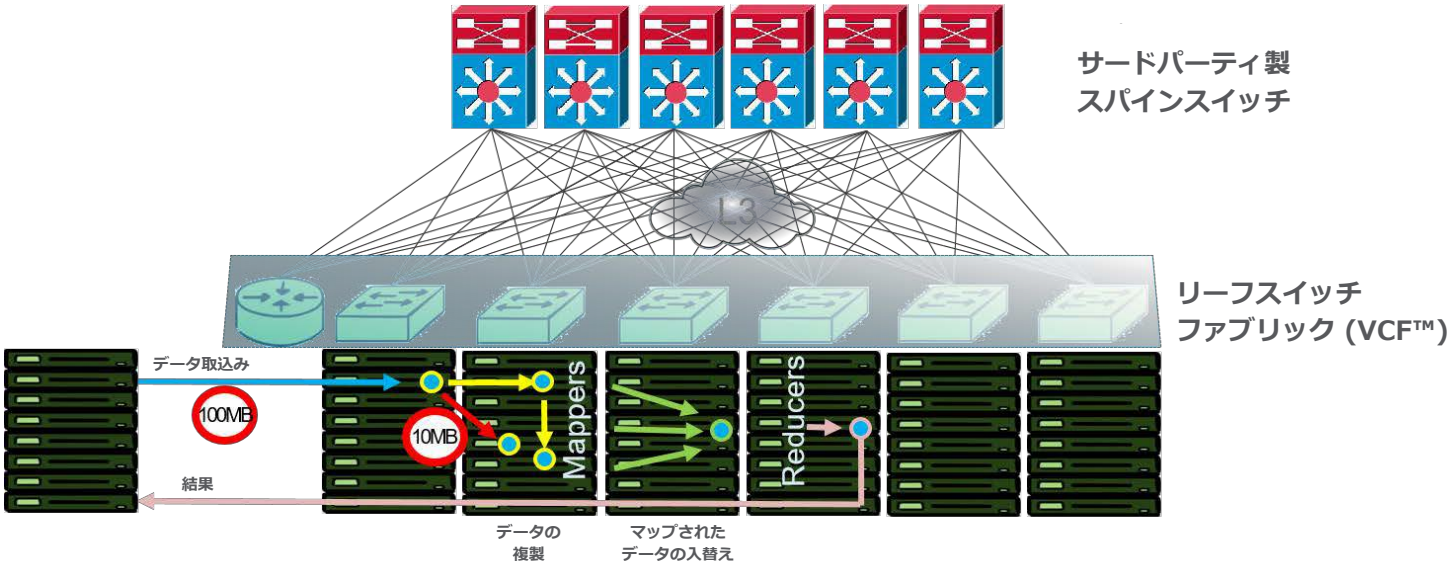


図4 – ブラウンフィールドラック環境における Hadoop クラスタの導入：スパインスイッチはレガシー、ToR は Dell EMC スイッチ向け Pluribus ONVL のファブリック

Pluribus ONVL 仮想化中心型 ファブリック (VCF™)

Pluribus ONVL 仮想化中心型ファブリック (VCF™) は分散型アーキテクチャであり、1つの論理エンティティとしてオープンで標準ベースのイーサネットファブリックであるクラスタリング技術計算に基づいており、ネットワークファブリックの洞察力、機敏性、セキュリティを強化します。Pluribus ONVL は、シングル論理スイッチ下において Dell EMC Open Networking スイッチのあらゆる組合せをクラスタすることを可能にします。

各スイッチは従来の L2/L3 ネットワークプロトコルを使用してローカルにネットワークトポロジの状態を計算し、各スイッチはクラスタ内の他の全てのピアスイッチにネットワークファブリックの "ビュー" を複製するよう、クラスタリングアルゴリズムに依存します。その結果、クラスタ内の全てのスイッチはファブリック全体の状態を知ることができます：例を挙げるとホスト MAC アドレス、ホスト IP アドレス、ポート、接続フロー、ネットワークリソースなどです。クラスタは、ネットワーク管理を大幅に簡素化する "マルチボックス仮想スイッチ" を形成します。ネットワーク内のあらゆるノードが、クラスタ全体の管理と制御の中心点として機能します。

Pluribus ONVL 仮想化中心型ファブリックは、ネットワークファブリックを構築するためのオープンで標準ベースのネットワークプロトコルの代わりではなく、L2/L3 接続に加えてネットワーク機能を向上させる補完的なソフトウェアアーキテクチャです。

ネットワーク管理者は、ファブリック全体のテレメトリ情報に容易にアクセスして、Hadoop のプライマリトラフィックやセカンダリトラフィックフローの識別、ポリシーの適用、およびトポロジの確認をすることが可能です。Hadoop 管理者は、単一のプログラムインターフェイスを介して、Pluribus ONVL のファブリックからファブリック全体の情報に容易にアクセスすることが可能です。

プログラマブル API を使用すると、アプリケーションフローをプログラミングし、ファブリック全体のネットワークリソースを制御することが出来ます：ネットワークは、単一でプログラム可能なスイッチとして扱うことが可能です。

プログラマビリティ

Pluribus ONVL は、シンプルで効果的な API を使用して、ネットワークリソース (ポート、VLAN、MAC など)、ネットワークサービス、および個々のアプリケーションフローを監視および制御する機能を開発者に提供します。C や RESTful API を含む CLI 以外のプログラミングオプションは、アプリケーション層との容易な統合を提供します。

プログラマブル API を使用するとアプリケーションフローをプログラミングし、ファブリック全体のネットワークリソースを制御することが可能となります。ネットワークは単一のプログラム可能なスイッチとして扱うことができます。このアプローチはアプリケーションの統合をさらに簡素化します。

Pluribus ONVL API は可視化情報、テレメトリおよびトラブルシューティングを簡素化する Hadoop クラスタに関する情報を、ファブリックに挿入するためのツールとして提供します：

- サーバ名と役割 (クライアント、ジョブトラッカー、ノード名、データノード、タスクトラッカーなど)
- ジョブに参加しているサーバ
- 同時に Pluribus ONVL API は、以下のようなファブリック全体で信頼できる情報を Hadoop クラスタに提供することが出来ます：
    - ローカリティを維持するラックの認識度
    - 重要な領域にジョブが配置されるのを避けるネットワークホットスポット

ファブリックは制御の一元化を提供するため、シンプルなスクリプトはファブリック全体の情報にアクセスして、クラスタ全体に効果的なリアルタイム処理を実行することが可能です。一般的な例としては、ビジネスの重要なジョブの実行中にセカンダリ HDFS トラフィックを速度制限することが挙げられます。

プログラム可能な Pluribus ONVL 仮想化中心型ファブリックは、ファブリック全体の TopTalker や FlowTrace のようなアプリケーションの初期セットを提供します。TopTalker は、トラフィック量によって順序付けられたエンドポイントのリストを提供します。FlowTrace は、特定のデータフローが通過するすべてのスイッチの入力/出力ポートのリストと詳細な輻輳状態を提供し、中間結果の入れ替え中に輻輳を識別し、防止する重要な情報を提供します。Pluribus アプリケーションは、CLI を介してネットワーク管理者にアクセス、およびプログラマブル API を介して Hadoop にアクセスすることが可能です。

可視性と制御：vPort と vFlow

Pluribus ONVL は、ネットワークファブリック内のトラフィックを直接監視かつ制御する高度な機能を提供します。

- エンドポイント情報は、基本的な物理トポロジ情報を含む単一のファブリック全体の “ディレクトリ” に整理された vPort オブジェクトとして表されます。vPort のレコードスキーマは、サーバタイプや OS のようなアプリケーション固有のメタデータを含むようにプログラムで拡張することが可能です。
- トラフィックは、vFlow と呼ばれるファブリック全体のオブジェクトを参照して識別、監視、制御されます。
- 平均レイテンシのようなテレメトリに関連する性能は、“接続” および “クライアント・サーバ” オブジェクトを参照すると容易にアクセスすることが可能です。

トラフィックの可視性と制御機能は、アプリケーションロジックとの容易な統合のためにプログラムのインターフェイス、および CLI を介して利用することが可能です。Dell EMC 向けの ONVL は、あらゆるネットワークイベントの監視機能を備えており、ファブリック全体のトラフィック制御と可視性を提供しています。トラフィック制御には帯域制限、アクセス制御、および出力先変更が含まれています。Hadoop クラスタに適用する場合、Pluribus ONVL の可視性と制御機能には、以下のようなアプリケーションがあります：

洗練されたテレメトリをネットワークスイッチの中に直接埋め込む事により、Pluribus ONVL は冗長監視インフラストラクチャの必要性を排除し、独立した監視タップ、プローブ、または専用の可視性ファブリックのコストを削減します。

- vPort テーブルをプログラミングすることにより、Hadoop の管理者はノード名やロールのような共有情報をネットワーク管理者に提供し、性能分析とトラブルシューティングを容易に行うことができます。
- vFlow 機能は、データの取り込みと出力結果に関連するフローを制御し、実行中のジョブにスムーズに対応する帯域幅を割り当てることが出来るワイルドカードをサポートします。 同じアプローチを使用して HDFS セカンダリトラフィックを制御し、時間的制約があるジョブの迅速な実行を可能にします。
- "connection-show" コマンドと "client-server-stats-show" コマンドを使用すると、リモートデータやファイルシステムのデータ複製にアクセスする map タスクや reduce タスクのような HDFS のプライマリおよびセカンダリトラフィックを識別できます。

統合ネットワークおよびアプリケーションフローテレメトリ

Pluribus ONVLは、全てのトランザクションの交差点にある物理ネットワーク内で実行され、インフラストラクチャを監視する理想的な場所です。ONVL の統合されたタップレスネットワークの可視性は、以下の複数のケースにおいて有益です：

- ネットワークの課題であるトラブルシューティングにかかる平均時間の削減
- 潜在的なセキュリティ問題にフラグを立て、監査をサポートするために標準からの逸脱を特定
- キャパシティプランニングのベースラインを作成

洗練されたテレメトリをネットワークスイッチの中に直接埋め込む事により、Pluribus ONVL は冗長監視インフラストラクチャの必要性を排除し、独立した監視タップ、プローブ、または専用の可視性ファブリックのコストを削減します。

Pluribus ONVL の分析機能の例は以下の通りです：

ホストロケーター

- 物理的および仮想的ないずれのノードからもファブリック全体のホストを配置ことが可能
- あらゆるノードからファブリック全体の仮想マシン移行をトレース
- ロケーションベースのロジックを有効にするために、ホストが接続されている特定のスイッチと ToR ポートを報告

アプリケーションフローテレメトリ

- アプリケーションフローの輻輳統計、ポート統計、タイムスタンプ、フローレイテンシ、さらにはクラスタ内のあらゆるノードからのファブリック全体のフローパスにいたるまで追跡
- サンプリングに基づくものではなく、ファブリック全体のネットフローは非常に正確

Hadoop のジョブの全体的な実行時間は、複数のフェーズに依存しています。アプリケーションフローテレメトリは、クラスタ内すべてのフローのパフォーマンス特性を提供します：

- データの取り込み
- マッパーデータアクセス
- mapper から reducer への中間結果の入れ替え
- データアクセスの削減
- 出力結果
- HDFS 初期複製
- HDFS 複製リカバリ

ONVL によって提供される詳細なテレメトリは、以下のような重大な状態を明らかにします：

- Mapper/Reducer データアクセスと競合する HDFS 複製リカバリデータ
- 多数の mapper から少数の reducer への出力結果の入れ替え中の輻輳/ホットスポット
- 準最適データ配置によるタスクトラッカーとデータノード間の過度のトラフィック

結果として対応処置のためにリアルタイムで、もしくは遅延したジョブの根本的な原因を特定するために事後にも取り込むことが可能です。Pluribus ONVL のテレメトリは、ネットワークリソースに起因するパフォーマンスの低下とアプリケーションレベルの不平衡を判別する便利なツールです。

結論

企業が日々作り出し、時間とともに蓄積していく大規模なデータの中にビジネス価値が隠れているということは、ますます明らかになってきています。その結果、Apache Hadoop のようなインフラストラクチャを使用する Big Data の処理は、一般的な企業のデータセンターにおいて主流となっています。ほとんどの場合、Hadoop のジョブ時間はビジネスの結果に直接関係しており、Big Data の処理は、リアルタイムもしくはリアルタイムに近い形で実行される事が期待されています。残念ながら Hadoop のジョブ結果は、多数のタスクがすべて完了した場合にのみ利用可能であり、並列したタスクの実行における不均衡に対して非常に敏感です。これにより、ネットワークインフラに特有の要求をもたらし、同様に対処的なコントロールを伴うきめ細かな性能パラメーターを可視化する必要があります。

Dell EMC Open Network スイッチ上で稼働する Pluribus OpenNetvisor® Linux (ONVL) は、シングルポイントでネットワークファブリック全体の管理アクセスを使用して、この優れたレベルの可視性と制御を提供します。一元化されたネットワークのファブリック全体の管理アクセスを使用して、優れたレベルの可視性と制御を提供します。ファブリック全体の統合されたビューにアクセスすることにより、CLI ベースの管理と API のプログラムによるアクセスが大幅に簡素化され、使いやすさとシンプルなアプリケーションインテグレーションを実現します。

オープンなネットワーク基準と Dell EMC Open Networking マーチャントシリコンスイッチラインに基づいて、Pluribus/Dell EMC のソリューションが共同で、界業界最高レベルの経済性を備えた Hadoop インフラストラクチャにおける、前例のないファブリックの可視性と制御を提供します。

Pluribus Networks について

Pluribus Networks は、御社のビジネスを制約しないデータセンターソリューションを提供します。当社のソフトウェア定義型オープンネットワーキングおよびファブリックベースのソリューションは、既存のネットワークインフラストラクチャを、柔軟で戦略的な最新のデジタルビジネスニーズに完全に対応させます。

当社の仮想化中心型ファブリック (VCF™) アーキテクチャは、操作を簡素化し、費用対効果の高い運用を実現し、新しいアプリケーションをより迅速にオンラインでお届けするために、これまでにない洞察力、機敏性、およびセキュリティを提供します。

お問い合わせ先
http://www.pluribusnetworks.com/dell and @pluribusnet.

Pluribus Networks, Inc.
2455 Faber Place, Suite 100, Palo Alto, CA 94303
1-855-GET-VNET / +1 650-289-4717

2016年２月

Copyright© 2016 Pluribus Networks, Inc. All rights reserved.