

jh180077-ISJ

Software-Defined IT インフラストラクチャにおける オーケストレーションに向けた資源管理システム

渡場 康弘（大阪大学）

概要 本研究では、組織間での連携を想定した災害管理（Disaster Management）の支援を目的としたアプリケーション基盤として、障害などへの高い耐久性と拠点間での継続的な情報共有を可能とする広域連携大規模可視化システムの実現を目指しており、本可視化システムを提供可能とする Software-Defined IT インフラストラクチャ技術の研究開発を行っている。前年度までの研究開発では、計算機サーバおよび大規模可視化装置が被災した場合に柔軟に移動することを想定したシナリオに対し、計算機資源の仮想化技術および Software-Defined Networking（SDN）技術を組み合わせたプロトタイプシステムを PRAGMA-ENT 上に構築してきた。しかし、これらのシナリオでは仮想計算機イメージやデータ保管ストレージの被災などは考慮されておらず、被災によるさまざまな状況に対して自律的にサービスを継続可能な基盤には達していない。そこで、本課題ではプロトタイプシステムで使用している Docker コンテナのイメージを、割り当てた計算資源および仮想ネットワーク環境を包括して管理制御する資源管理システムについて研究開発を行う。

1. 共同研究に関する情報

(1) 共同研究を実施した拠点名

大阪大学

(2) 共同研究分野

- ☐ 超大規模数値計算系応用分野
- ☐ 超大規模データ処理系応用分野
- ☐ 超大容量ネットワーク技術分野
- ☒ 超大規模情報システム関連研究分野

(3) 参加研究者の役割分担

渡場 康弘（奈良先端科学技術大学院大学）：

代表者

José A.B. Fortes（University of Florida）：

副代表者・国際連携基盤の整備

下條 真司（大阪大学）：

副代表者・国際連携基盤の整備

木戸 善之（大阪大学）：

副代表者・広域可視化テストベッドの整備

市川 晃平（奈良先端科学技術大学院大学）：

副代表者・PRAGMA-ENT の整備・運用

伊達 進（大阪大学）：

阪大拠点の整備・運用

阿部 洋丈（筑波大学）：

広域可視化テストベッドの整備

山中 広明（情報通信研究機構）：

PRAGMA-ENT の整備・運用

Jason Haga（産業技術総合研究所）：

可視化ツール開発

高野 了成（産業技術総合研究所）：

広域可視化テストベッドの整備

Fang-Pang Lin（National Center for High-performance Computing）：

国際連携基盤の整備

遠藤 新（大阪大学）：

ストレージシステムの導入設計

石田 和也（大阪大学）：

TDW 基盤切り替えシステムの開発

2. 研究の目的と意義

Information-as-a-Service（InfaaS）というコンセプトは、複数の組織間で連携した災害管理（Disaster Management）を支援するアプリケーション基盤、具体的には災害に対するリスク管理、災害時における避難支援、救

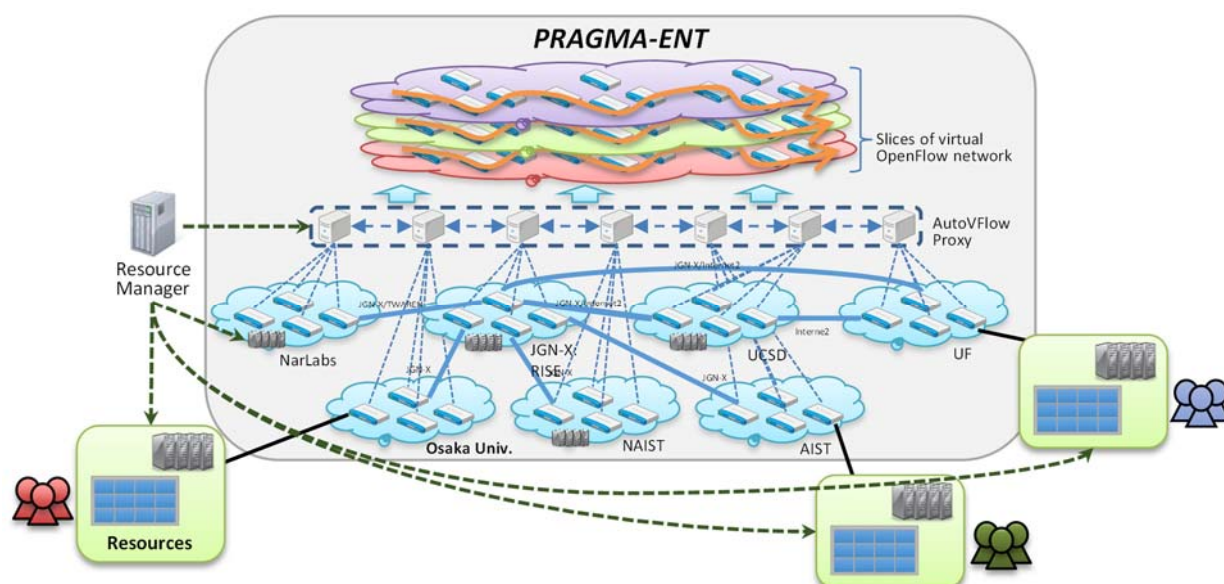


図 1 PRAGMA-ENT 上における Software-Defined IT インフラストラクチャと連携した広域連携大規模可視化システムの構成図

助支援などの活動方針の決定作業を支援する災害管理システムを構築する上で重要な考えである。複数の組織による災害管理における対策方針を決定するプロセスでは、収集されたデータ、およびその解析結果などから得られた情報をさまざまな組織の災害管理担当者に同時かつ分かりやすく提供することが重要である。前述の観点において、広域連携大規模可視化システムは意思決定を支援する災害対策アプリケーションの基盤の 1 つとして注目されている。

複数の組織間での連携を支援する災害対策アプリケーションでは、組織間における情報の伝達および同期は必要不可欠であるが、一般的に構築される広域連携大規模可視化システムではこれらを保証する仕組みは備えていない。それゆえ、被災により利用するリソースの一部が使用できない状況に対する高い耐久性と、情報伝達を維持するために代替のリソース構成に迅速に切り替えることが可能である IT インフラストラクチャが、複数組織の連携を支援する災害対策アプリケーション基盤として利用可能な広域連携大規模可視化システムを実現する上で必要

不可欠である。

障害などへの高い耐久性と拠点間での継続的な情報共有を可能とする IT インフラストラクチャの実現に向け研究開発を進めてきた。2016 年度は、情報を配信するアプリケーションサーバの被災を想定したシナリオに基づき、近年注目されている Software-Defined 技術と計算機の仮想化技術を IT インフラストラクチャに適用した Software-Defined IT インフラストラクチャの研究開発を行い、図 1 に示す Software-Defined IT インフラストラクチャと連携した広域連携大規模可視化システムのプロトタイプを構築を行った。

また、2017 年度は、2016 年度のシナリオでは対象としていなかった大規模可視化装置が配備された拠点での被災を想定し、大規模可視化装置で実行されている可視化アプリケーションを他の拠点の可視化装置に柔軟に移動する場合のシナリオに取り組んだ。本研究で利用している大規模可視化装置には、複数のディスプレイを並べて 1 つのスクリーンとして扱う Tiled Display Wall (TDW) を採用している。TDW では、複数のディスプ

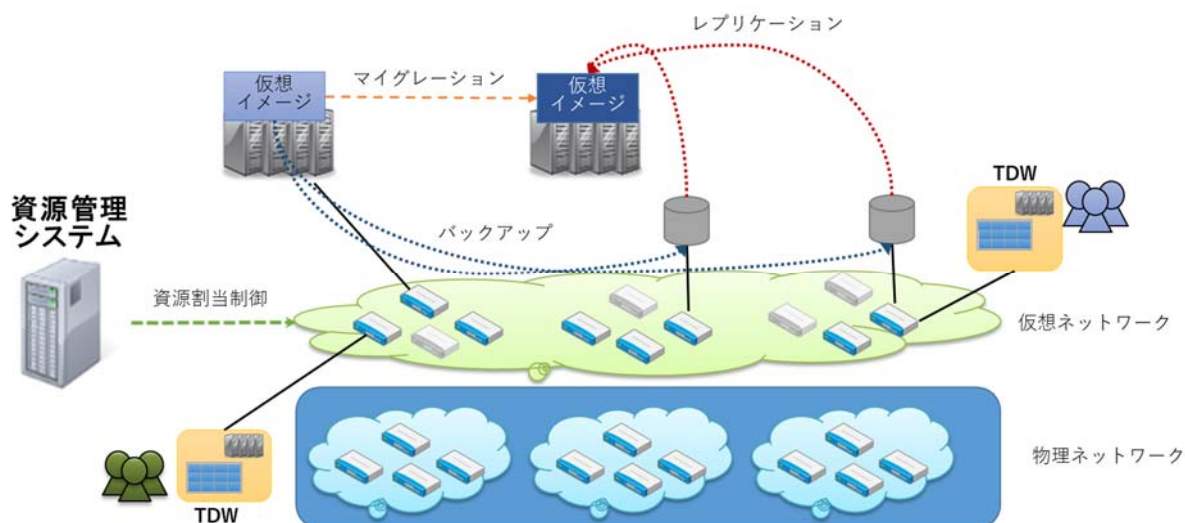


図2 Software-Defined IT インフラストラクチャにおける資源管理システムの概念図

レイを1つのスクリーンとして制御して可視化結果を表示するため、専用の TDW 制御ミドルウェアの利用や TDW 表示に対応した専用可視化アプリケーションが必要となる。これらの基盤ソフトウェアはそれらを共存させたプラットフォーム環境を構築することは困難であることから、TDW 大規模可視化装置に対する柔軟な環境切替機構の研究開発を行い、広域連携大規模可視化システムのプロトタイプと統合した環境の構築を行った。

しかし、これまでの取り組みで扱ってきた想定シナリオでは、仮想計算機イメージのストレージ、データ保管用のストレージなど、一部の災害管理システムに必要な資源が被災する状況は考慮されていない。そのため、これまでの課題で開発した Software-Defined IT インフラストラクチャ技術における資源管理システムでは、被災によって生じるさまざまな状況に対応するために必要となる機能を自律的に管理および制御を行うオーケストレーションとしての仕組みが備わっていない。

そこで、本課題では、図2に示す Software-Defined IT インフラストラクチャと連携した広域連携大規模可視化システムのプロト

タイプで使用している Docker コンテナのイメージを、割り当てた計算資源および仮想ネットワーク環境を包括して管理制御する資源管理システムについて研究開発を行う。

3. 当拠点公募型共同研究として実施した意義

本研究課題で目指す広域連携大規模可視化システム実現のための Software-Defined IT インフラストラクチャ技術の構築・検証を行うためには、大規模データに対する解析を行う計算資源、TDW などに代表される大規模可視化装置、そしてこれらの各種資源が Software-Defined Networking (SDN) に対応した広域拠点間ネットワークで接続された環境を想定している。大阪大学拠点ではこれらのリソースを提供できる環境が整備されているため、当拠点公募型共同研究の枠組みを利用することで本課題の研究を推進することが可能であると考えます。

4. 前年度までに得られた研究成果の概要

2016 年度は、各可視化拠点にデータから生成された可視化結果を配信する可視化アプリケーションが実行されている計算機サーバが配置されている拠点で被災した場合を想定したシナリオに基づき、Software-

Defined IT インフラストラクチャと連携した広域連携大規模可視化システムのプロトタイプ的设计・構築を行った。本プロトタイプの構築にあたり、(1-1) **Multi-site** 可視化ツールの配備、(1-2) **SDN-based** インフラストラクチャの生成、(1-3) 障害発生時における可視化ツールのロバスト性の検証の3つのステージを設けて研究を進めた。(1-1)では **Multi-site** 可視化ツールとして、多くの災害管理システムで求められる仮想ホワイトボードおよびワークスペースを **Scalable Amplified Group Environment (SAGE2)** を利用して開発・構築した。(1-2)では、可視化処理を行う計算サーバに割り当てる資源、可視化サーバと大規模可視化装置間のネットワークの両方を管理・制御を行う資源管理システムについて取り組んだ。(1-3)ではこれらを用いたプロトタイプを実際の広域環境で検証するため、仮想ネットワーク技術および **SDN** 技術を利用可能な広域ネットワークテストベッドである **PRAGMA-ENT (PRAGMA Experimental Network Testbed)** 上での配備を行った。

2017 年度は、大規模可視化装置が配備された拠点での被災を想定し、可視化装置上で実行されている可視化アプリケーションを他の拠点の可視化装置に柔軟に移動する場合のシナリオに対し、大規模可視化装置のプラットフォーム環境を柔軟に切り替え可能な **Software-Defined IT** インフラストラクチャのプロトタイプ of 構築について取り組んだ。本プロトタイプ of 研究開発は、(2-1)可視化プラットフォーム of 切替機能 of 開発、(2-2) 割当計算・ネットワーク資源も含めた制御機構 of 開発、(2-3)**Software-Defined IT** インフラストラクチャ of プロトタイプ of と of 統合および検証 of 3 つ of ステージに of 分けて進めた。(2-1)では、計算機仮想化技術を用いて **OS** レベルで可視化プラットフォーム環境を切り分け、利用アプリケーションにあわせた

TDW 環境 of 管理・切替を行う手法について研究開発を行った。(2-2)では、一昨年度開発したプロトタイプ of 資源管理システムに対し、(2-1) of **TDW** 環境 of 管理・切替機構 of 制御を行うための拡張を行った。(2-3)では、**PRAGMA-ENT** 上における **Software-Defined IT** インフラストラクチャ of プロトタイプに対して(a)、(b) of 成果を配備して動作 of 検証を行った。

5. 今年度 of 研究成果 of 詳細

本課題 of 目的である **Software-Defined IT** インフラストラクチャと連携した広域連携大規模可視化システムのプロトタイプで使用している **Docker** コンテナ of イメージを、割り当てた計算資源および仮想ネットワーク環境を包括して管理制御する資源管理システム of 実現に向け、まず初めに既存プロトタイプ of における資源管理システム of 機能 of 整理および再設計を行った。既存プロトタイプ of 資源管理システムは、これまでのシナリオ下 of 制御に着目して設計されていたため、各要素 of 管理・制御は限定的な動作しかできない。しかし、災害によって生じる様々な状況に柔軟に対処することを可能とするためには、**Software-Defined IT** インフラストラクチャ of さまざまな機能を包括的な管理・制御を行えることが必要不可欠である。そこで、既存プロトタイプ of における資源管理システム of 構成・制御機構を抜本から見直した。

これまでに取り扱ってきた2つのシナリオに基づき、資源管理システムが被災による広域分散可視化システムのサービス停止から復元を実施するために必要な処理について分析を行い、以下の5つの機能が必要であると考えた。

- (1) 各コンポーネント of 状態管理機能
- (2) サービス停止 of 検出機能

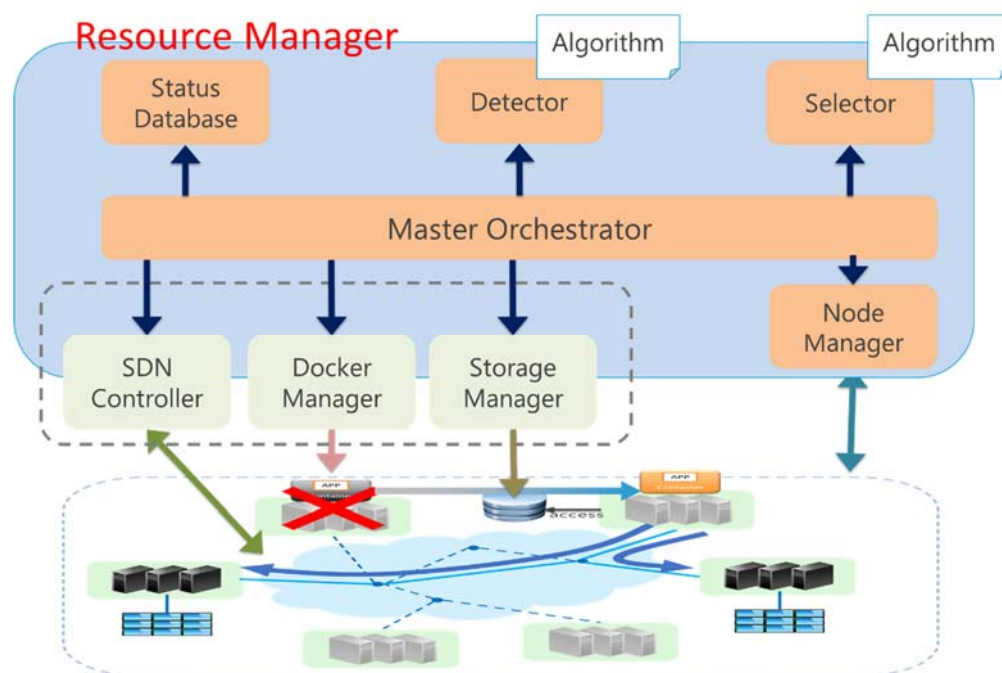


図3 提案資源管理システムのシステム構成図

(3) 割当資源の決定機能

(4) 各コンポーネントへの制御機能

機能(1) は **Software-Defined IT** インフラストラクチャの構成要素である、計算機、ネットワーク、ストレージ、および **Docker** コンテナの利用状況などのステータス情報の収集機能である。本研究では **IT** インフラストラクチャは共用資源であると仮定しているため、広域分散可視化システム内だけでなく、全体の利用状況に基づきサービスを行う必要があるため、各コンポーネントのステータス情報を収集する必要がある。

広域分散可視化システムのサービスを自律的に継続させるためには、機能(2) のサービス停止の検知機能を資源管理システムが備えている必要がある。サービス停止の検出にはさまざまなアプローチが考えられるため、特定の手法を資源管理システムが備えるのではなく、検出手法の追加・選択を可能とする機能も求められる。

機能(3) は広域分散可視化システムのサービスを復旧する際に、利用可能な資源の中か

らどの資源を割り当てるのかを決定する機能である。本研究では、広域分散可視化システムのサービスは使用する計算ノード、可視化装置だけでなく、使用する通信経路も **L2** ネットワークのプレーンとしてサービスに割り当てるため、計算機、ネットワーク、ストレージの状況から適切な資源を選択する仕組みが必要となる。この処理を実行するアルゴリズムも状況や運用方針によってさまざまな手法が考えられるため、切替が可能である必要がある。

機能(4)は機能(3) で決定された割当に基づき、**SDN** および **Docker** を制御する機能である。これらは既存のコントローラまたはマネージャにより独立で管理・制御が行われているため、資源管理システムはこれらのコンポーネントに指示を出すための仕組みが必要である。また、通常運用時におけるストレージにおけるレプリケーション・バックアップなどの制御機能も本機能に含まれる。

以上の機能要件に基づき、提案する広域分散可視化システムのサービス維持を保証する **Software-Defined IT** インフラストラク

チャにおける資源管理システムのプロトタイプ設計を行った。提案資源管理システムのシステム構成を図 3 に示す。本資源管理システムは 8 つの要素で構成される。これらの構成要素のうち、資源管理システムのコアとなるのが *Master Orchestrator* モジュールであり、他の 7 つの要素は基本的に本モジュールからの指示に基づき動作するよう設計した。

前述の機能(1) は、*Master Orchestrator* モジュールから各コンポーネントを管理するモジュールに指示を出して行う。計算機資源の利用状況は *Node Manager* モジュールから、ネットワーク、*Docker* コンテナ、ストレージに関してはそれぞれの既存制御機構を介してステータス情報の収集を行い、*Master Orchestrator* モジュールで整形処理を行い *Status Database* に格納する。なお、計算機資源についても各拠点のローカル資源管理システムにより制御されている状況も想定されるが、本プロトタイプではすべての計算機資源を本資源管理システムが管理している仮定の下で実装を行う。

機能(2) は *Detector* モジュールで実行される。本モジュールはサービス停止の検出処理で必要とされる基本的な機能群で構成しており、具体的な処理内容は別に作成したアルゴリズムをプラグイン形式で利用できる構成を取っている。このプラグイン形式のモジュール構成は機能(3) を実現する *Selector* モジュールにおいても採用しており、任意の資源探索アルゴリズムを採用できる仕組みを提供する。

機能(4) については機能(1) でも使用している *SDN Controller*、*Docker Manager*、*Storage Manager* に対する制御機構として実装している。なお、*SDN Controller* 制御機構については、使用するコントローラを *Trema* から *Ryu* に変更したため、従来の処理の構成を踏襲しつつ、制御機構を新規に構

築した。

6. 今年度の進捗状況と今後の展望

今年度は、本課題で目指している災害管理アプリケーションのための広域分散可視化システムを実現する *Software-Defined IT* インフラストラクチャの実現に向け、*Software-Defined IT* インフラストラクチャにおけるオーケストレーションを実現するための資源管理システムの構築に取り組んだ。本課題申請における研究計画では、オーケストレーションに向けてこれまでの資源管理システムで扱われていなかった *Docker* イメージや可視化処理の対象データなどのストレージ資源に対する機能を中心に研究開発を行う予定であったが、*Software-Defined IT* インフラストラクチャにおけるオーケストレーションを実現する今後の研究発展の観点から、資源管理システムを抜本的に再構成する必要があると判断し、資源管理システム全体の設計の見直しから本課題を進めた。その結果、現時点では提案資源管理システムは開発のために構築したローカル環境での動作検証段階であり、これまでに構築してきた *PRAGMA-ENT* 上のプロトタイプシステムでの配備までは達成できておらず、研究計画からは遅れている状況となった。一方で、新しく構成した資源管理システムは、サービス停止検出および割当資源決定処理でさまざまなアルゴリズムを適用可能なプラグイン形式の実装を実現したことで、実際の広域連携災害管理への適用だけでなく、新しい状況やアルゴリズムを開発・検証することを可能とする基盤に向けても活用が可能となった点は大きな進捗であると考ええる。

今後は *PRAGMA-ENT* 上のプロトタイプシステムに配備し、資源管理システムの動作検証および各処理の性能について評価を行う予定である。また、実現したプラグイン機

構を利用した新たなアルゴリズムの研究開発 12 月.
発に取り組む予定である。

(5) その他（特許，プレス発表，著書等）

- なし

7. 研究成果リスト

(1) 学術論文

- なし

(2) 国際会議プロシーディングス

- Kazuya Ishida, Daiki Asao, Arata Endo, Yoshiyuki Kido, Susumu Date, Shinji Shimojo, “High-Resolution Streaming Functionality in SAGE2 Screen Sharing”, Proceedings of the 2019 Future of Information and Communications Conference (FICC 2019), Advances in Information and Communication, Lecture Notes in Networks and Systems, vol. 70, pp.384-399, San Francisco, U.S., March 2019.

(3) 国際会議発表

- Yasuhiro Watashiba, Yuki Matsui, Yoshiyuki Kido, Susumu Date, Shinji Shimojo, “Toward Orchestration on Software-Defined IT Infrastructure for Disaster Management Application”, PRAGMA 36 Workshop, Apr. 2019.
- Yuki Matsui, Yasuhiro Watashiba, Yoshiyuki Kido, Susumu Date, Shinji Shimojo, “Architecture of Resource Manager for Software-Defined IT Infrastructure”, ISGC 2019, Taipei, Taiwan, Apr. 2019.

(4) 国内会議発表

- 松井祐希，渡場康弘，伊達進，木戸善之，下條真司，“広域連携型災害管理アプリケーション基盤を提供する資源管理システムの検討”，日本ソフトウェア科学会 第 16 回ディペンダブルシステムワークショップ，金沢，2018 年