

パブリッククラウドを活用したデータ連携基盤の Infrastructure as Code による自動構築方式の検討

南 豪介[†] 長谷 亮[†] 松浦 陽平[†]

[†]三菱電機株式会社 情報技術総合研究所

1. はじめに

近年、行政・民間組織におけるデータ活用の拡大が期待されている[1]。一例として、複数の場所に保存されるデータの統合・分析を目的に、データを一元管理する機能を有したデータ基盤が多分野に適用されつつある[1]。また、上記の基盤構築にパブリッククラウド(以降、クラウド)の活用が広まっている。ここでは、クラウドを利用してデータを連携する基盤を、データ連携基盤(以降、本基盤)と記載する。

データ活用を抑制する要因に、データの処理・分析を取り扱う人材の不足がある[1]。本基盤では、上記人材の役割としてシステム運用担当者が挙げられる。システム運用担当者は本基盤の構築・設定作業において通常、UI を用いて IT リソースの構築・設定を手作業で行う。上記 UI は、容易な構築・設定作業を支援するが、構築・設定対象が増加すると構築・設定作業量も増加する。そのため、構築・設定作業を自動化し、システム運用担当者の作業労力を低減する手法が求められている。上記手法の一つに、本基盤の構築・設定項目をコードとして記述し、自動構築する技術である Infrastructure as Code (以降、IaC) を用いた手法が挙げられる。

本稿では、IaC を用いた本基盤の構築・設定作業労力の低減手法を検討した結果を記載する。また、UI や Command Line Interface (以降、CLI) を用いる手法との作業状況による使い分けについて検討した結果も記載する。

2. 関連技術と課題

本基盤では、データの収集、前処理、分析を行う受け渡しフローを定め、一部にデータウェアハウスを用いてデータ連携を実現する [2]。上記のデータフローをデータパイプラインと呼び、収集データをデータウェアハウスの形式に変換する一連の処理の手法を Extract-Transform-Load (以降、ETL) / Extract-Load-Transform (以降、

ELT) と呼ぶ[2]。一般的に、連携用途に対応するデータパイプラインを設定し、データ連携機能を構築する。したがって、連携用途が増加すると、データパイプラインが複雑化し、システム運用担当者の作業労力が増加すると考えられる。

本基盤の運用を支援する手法として Platform as a Service (以降、PaaS) の利用が挙げられる。PaaS とは、クラウドベンダが OS やミドルウェアを含めて、利用者にシステムの実行環境を提供するサービスである。PaaS により、データの分析・連携システムを容易に構築できるが、設定項目は依然として多い。そのため、多くの設定項目を一度に設定し、構築する手法が求められる。

本基盤の構築・設定自動化手法として、IaC や CLI がある。IaC では、コードの記述通りに IT リソースを自動構築・設定する。CLI では、コマンドで IT リソースを構築・設定するため、複数コマンドを一度に実行するスクリプトを作成すれば、構築・設定作業を自動化できる。IaC は一般的に、構築・設定内容の追記・変更部分のみが実行される冪等性という性質を持つ。一方、CLI では、冪等性を考慮したスクリプトを作成する必要がある。

本基盤では、運用中の機能変更・追加が考えられるため、冪等性が重要となる。そのため本研究では、IaC を用いて本基盤の自動構築・設定を行う手法の検討を行った。

3. 本基盤の概要

複数サーバ内のデータをクラウド上のストレージに収集し、データ連携を行う環境を構築した。本基盤の構成を図 1 に示す。本基盤に構築した機能の概要を以下に示す。

・ データ連携機能

複数サーバ内の収集対象データを、クラウド上に構築したストレージに収集することが可能である。また、収集したデータを他のストレージや分析アプリに転送することが可能である。

・ 構築・設定自動化機能

IT リソースの設定項目を記述したコード(以降、IaC テンプレート)を実行して、構築対象の IT リソースの自動構築・設定が可能である。また、複数 IT リソースの自動構築・設定も可能である。

A Study on a Method of Automated Resource Provisioning using Infrastructure as Code for a Data Integration Platform in a Public Cloud

[†]Information Technology R&D Center, Mitsubishi Electric Corporation.

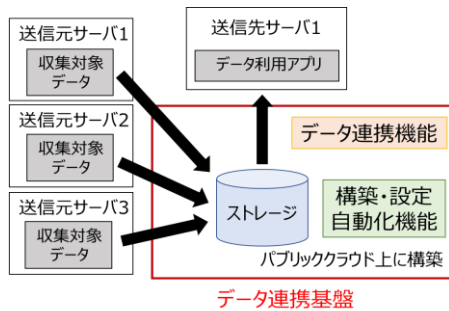


図1 本基盤の構成

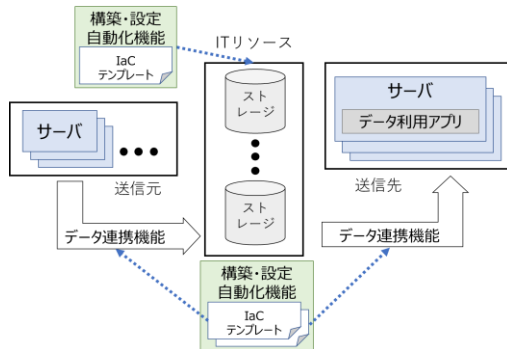


図2 IaCテンプレートによる自動構築・設定

上記の機能に関して、IaCテンプレートと構築するITリソースの関係を図2に示す。

4. IaCによる基盤、連携機能の構築・設定

PaaSを利用する観点から、Microsoft Azureを用いて本基盤を試作した。そのため、Microsoft Azureで利用可能なIaCであるAzure Resource Managerテンプレート（以降、ARMテンプレート）を用いて、ITリソース、データ連携機能の構築・設定自動化を行った。以下に、IaCを利用したITリソース、データ連携機能の構築・設定手法の概要を示す。

・ 基盤を構成するITリソースの構築

ITリソースの設定項目を記述したARMテンプレートを実行して、本基盤を構成するITリソースの構築を行った。ARMテンプレートに記述した設定項目には、ITリソースの種類、名称、ネットワーク設定などが挙げられる。本基盤ではオブジェクトストレージ、リレーショナルデータベースなど様々なストレージを用いる。これらのITリソースを構築するIaCテンプレートを作成しておくことで、必要時にストレージを迅速に自動構築・設定することができる。

・ データ連携機能の設定

データ連携機能、通信設定などをARMテンプレートに記述することで設定を行った。これにより、送信元・送信先のストレージや、転送データ形式、転送頻度などの拡張を容易に行える。また、データ連携機能の複製も、過去のIaCテンプレートを流用することで、容易に行える。

既存のデータ連携機能の複製や、機能拡張を行う場面でUIを用いる場合、機能変更・追加の度にUIを操作し、データ連携機能の設定を行う必要があるため、作業の労力が大きい。一方、IaCを用いた手法では、IaCテンプレートの記述を必要部分のみ変更することで機能変更・追加を実現できる。したがって、データ連携機能の設定変更・追加の際、IaCを用いることで作業労力を低減できる可能性が高いことが分かった。さらに、人手による作業では作業ミスなどのリスクが大きくなるが、IaCによる自動構築・設定ならば、作業ミスのリスクを低減できる。

しかし、IaCを用いてITリソースやデータ連携機能を構築する手法は、UIやCLIを用いる手法と手順が大きく異なる。そのため、システム運用担当者がIaCの仕様を理解し、利用するための学習コストが大きく、結果的に初期構築時の労力が大きくなることが考えられる。それゆえに、ITリソースやデータ連携機能などのうち、繰り返し設定しないものは、UIやCLIを利用して構築・設定する方が望ましいことがわかった。なお、構築済みのITリソースやデータ連携機能からは、PaaSの機能を用いてIaCテンプレートを出力できる。本基盤の運用で、繰り返し設定するITリソースやデータ連携機能は、初回構築時はUIを使い、それ以降の構築時は出力したIaCテンプレートを利用することで、IaCに関する学習コストを抑えることも可能である。

5. おわりに

本研究では、クラウドを活用し、データの連携・分析を行うデータ連携基盤のシステム運用担当者の労力を低減することを目的として、PaaSやIaCを用いた、本基盤の自動構築・設定手法について研究を行った。その結果、データ連携基盤を構築・設定した後、長期にわたって機能追加・変更を行っていく場面において、上記手法がシステム運用担当者の労力を低減する可能性を示唆した。また、繰り返し設定しないITリソースやデータ連携機能については、UIやCLIを用いて構築・設定を行う手法が有効である可能性を示唆した。今後の課題として、実環境への適用に向けた検証を進めていく必要がある。

参考文献

- [1] 総務省、情報通信白書令和2年版、<https://www.soumu.go.jp/johotsusintokei/whitepaper/r02.html> (閲覧：2020年12月17日)。
- [2] 西田圭介、『ビッグデータを支える技術-刻々とデータが脈打つ自動化の世界-』、技術評論社、2017。