R&D ホットコーナー ソリューション

クラウド環境のリソース設計を効率化するIntent-Based Service Management（IBSM）フレームワーク

NTTアクセスサービスシステム研究所

呉（ご）　超（ちょう）　/堀内（ほりうち）　信吾（しんご）　/田山（たやま）　健一（けんいち）

クラウドサービスの普及に伴い，サービス事業者がWebサービス，機械学習等のさまざまな機能をクラウド環境により実装し，エンドユーザにサービス提供する形態が増加しています．クラウド提供事業者が，迅速かつ効率的にこれらのクラウドサービスを提供・運用することを目的に，サービス事業者のサービス要件に応じたクラウドリソース設計を効率化する技術に取り組んでいます．ここでは，サービスの機能性・安全性・信頼性・パフォーマンス等のサービス要件と合わせて，クラウド環境・運用ポリシーを考慮し，サービス要件を満たすために必要なリソースを導出するIntent-Based Service Management（IBSM）フレームワークについて紹介します．

現状のクラウドサービス提供形態と課題

近年，さまざまな業界，規模のサービス事業者が，クラウド環境を用いて自社またはお客さま向けのサービスを構築しています．クラウドサービスの提供において，サービス事業者はサービス要件を提供し，一方でクラウド提供事業者はクラウドサービスの実装に必要となるリソース要件を必要とします．そのためサービス要件を満足するリソース設計が必要となります（**図1**）．サービス要件とは，サービス事業者が求めるサービス特性に関する要件であり，機能性，安全性，信頼性，ワークロード，パフォーマンス等があります．リソース要件とは，リソース特性に関する要件であり，リソースの構成（リソースの種類，数，接続関係）とそれらのリソース量（vCPU数，メモリ量，ストレージサイズ）等があります．　このため，サービス提供の際には，サービス事業者のサービス要件に応じて，クラウド提供事業者はリソース構成およびリソース量を決定することが必要です．

また，クラウドサービスの運用時に，サービス要件を継続して満足するため，サービス要件や環境・運用状況等が変化した場合に，リソース構成および量の再調整が必要となります．例えば，10000 requests per secondのワークロードを収容でき，高安全・高信頼のWebサービスを実現するために必要なリソース構成（ファイアウォール，Webサーバ，データベースなどの機能とこれら機能の冗長化），およびリソース量を決定する必要があります．さらにサービス要件の変動，例えば，ワークロード量の増加，環境の変化（ホストの混雑度上昇等）に応じて，リソース構成またはリソース量の調整が必要となります．

現状，サービス要件に応じたリソース設計方式として，次の2種類があります．

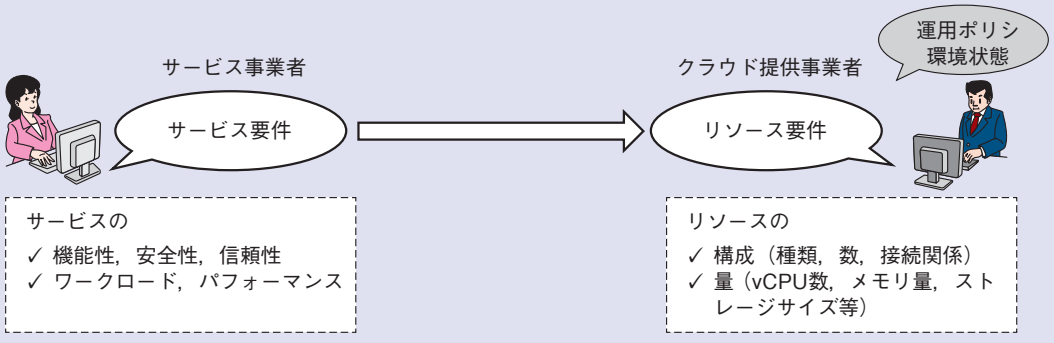

図1　サービス要件を満足するためのリソース設計

⑴　Cloud-Consultant方式

クラウド提供事業者の営業担当者がサービス事業者にヒアリングを行い，サービス事業者のサービス要件を把握，そのうえでクラウド提供事業者のエンジニア等がこれらの要件に応じてリソースを決定します．

⑵　Self-Management方式

クラウド提供事業者がサービス事業者向けのリソース管理インタフェースを用意し，サービス事業者が自らサービスレイヤ要件をリソース要件に変換し，インタフェースを通して必要なリソースを要求・調整します．

いずれの方式も，クラウド提供事業者側あるいはサービス事業者側にクラウド環境の運用経験と専門知識を有する人材が必要であるため，クラウド設計コストの増大等の問題が生じます．

NTTアクセスサービスシステム研究所（AS研）では，クラウドサービスの設計・運用の効率化とクラウドサービスの満足度向上のため，サービス要件に応じたリソース設計を実現するIntent-Based Service Management（IBSM）フレームワークを検討しています．IBSMフレームワークは，自然言語あるいはGUI選択で表現したサービス事業者のサービス要件を解析し，リソース構成の決定とリソース量の算出を可能とします．さらに，IBSMをリソース管理システム（例えばOpenstack）と連携させることにより，サービス要件受領からリソース設定までの一連の業務の自動化も実現可能です．

IBSMフレームワーク

IBSMフレームワーク[1]は主に3つの機能部（要件解析部，リソース構成決定部，リソース量算出部）によって構成されます[2]（**図2**）．以下に，各機能部の原理および役割を紹介します．

■要件解析部

サービス事業者はさまざまなチャネルを通じて，自然言語やGUI選択等でサービス要件を表現します．要件解析部は，これらのチャネルから収集したサービス要件を自然言語分析等の手法を用い，要件API（Application Programming Interface）の指定形式に基づいて，サービスの機能性，安全性，信頼性，ワークロード，パフォーマンス要件に関する要件を抽出します．抽出した個々のサー

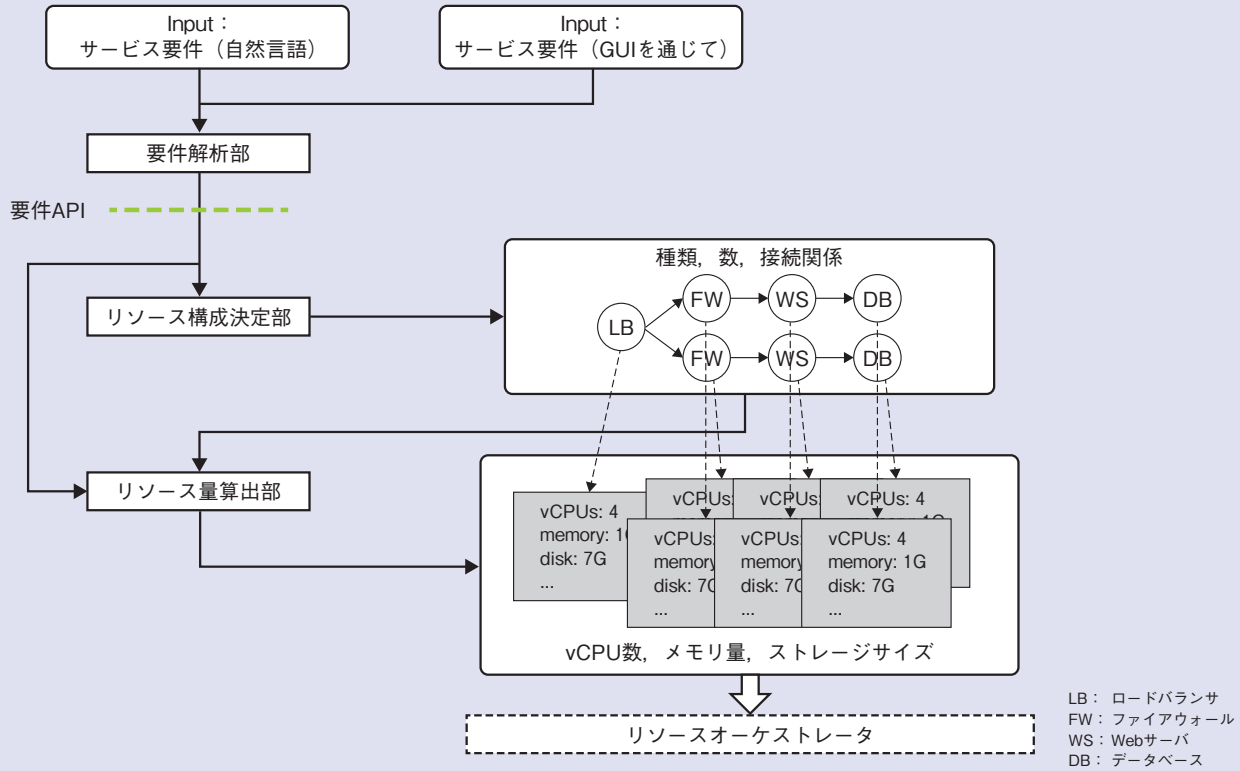

図2　IBSMフレームワーク

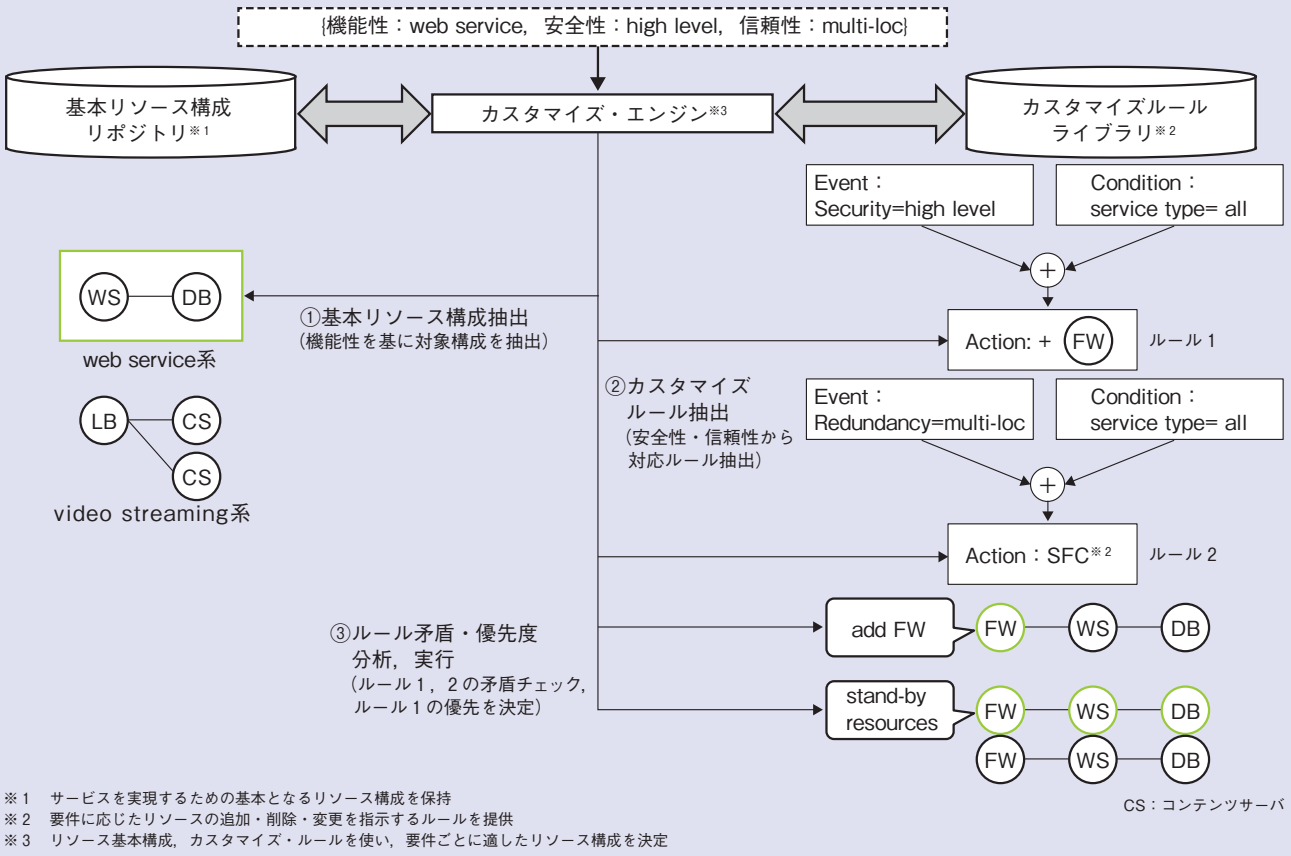

図 3　リソース構成決定部および決定例

ビス要件を，リソース構成決定部およびリソース量算出部に要件APIを通じて渡します．

■リソース構成決定部

リソース構成決定部は，要件APIから受領したサービス要件（特にサービス機能性，安全性，信頼性に関する要件）に応じて，リソース構成を決定します．これらの決定においては，一般的に，サービスごとに定義されたリソース構成テンプレートを参照して構成を確定する方法があります．しかし，この方法は，サービスバリエーションの増加に伴い，テンプレート数が増加する問題があります．この問題の解決方法を**図 3** に示します．少数の基本リソース構成のテンプレートをカスタマイズ可能とすることで，テンプレート数の増加を防ぐことが可能です．

■リソース量算出部

リソース量算出部はサービス開通時と運用時のそれぞれにおいて以下の役割があります．

・開通時：パフォーマンス要件を満たすため，ワークロード，環境状態，運用ポリシーを考慮して，VM（Virtual Machine）に配置するリソース量を決定します
・運用時：パフォーマンス要件を継続して満たすため，ワークロード，環境状態，運用ポリシーの変動に応じて，VMに配置するリソース量の調整に関する決定をします

以下では，「ワークロードとパフォーマンス要件」「環境状態（サーバの運用状態）」「運用ポリシー」の各要素がリソース量の決定に影響を与える原理を説明します．

⑴　ワークロードとパフォーマンス要件

クラウド提供事業者はワークロードを処理するとともに，その処理に対して，一定のパフォーマンスを求めます．

ワークロードとは，サービス事業者がクラウドを使って実行する処理の種類・特徴および処理量の総称であり，以下の例があります．

①　ワークロード種類：Webサービストラフィック，特徴：Webページサイズ20 KB，処理量：10000 requests

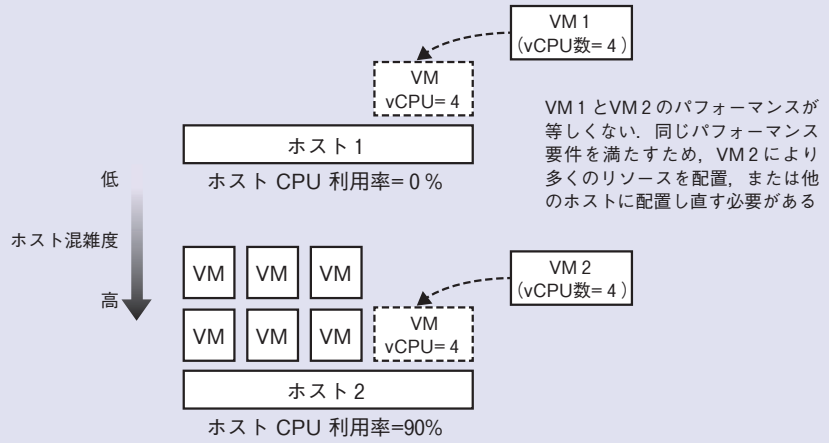

図 4　環境状態のパフォーマンスへの影響例

per second

② ワークロード種類：機械学習訓練，特徴：神経網の層数，各層のノード数および関数，処理量：1 batchの訓練データのサイズ（30枚の32×32 pixel 256色の画像）

パフォーマンス要件には，主にワークロード処理能力と遅延の 2 つの要件があります．ワークロード処理能力要件の例として，①のようなワークロードに対して，95 %以上のrequestsを処理可能，90 %以上のユーザを収容可能等があります．遅延要件は，ワークロードの処理時間への要求を表し，例えば，①のワークロードに対して平均requests処理時間が100 ms以下であること，②のワークロードに対して1batchの訓練データを130 ms内に学習が完了すること等があります．

ワークロードに対してVMに配置するvCPU数やメモリ量は，VMのワークロード処理能力や処理時間に直接影響します．このため，リソース量を決定する際に，ワークロードとパフォーマンス要件は重要な影響要素となります．

⑵ 環境状態

環境状態は，サービスを構成するVMを収容する物理サーバ（ホスト）の状態です．動的な環境状態の例として，ホストのリソース利用率（ホストの混雑度）があります．静的な環境状態の例として，物理サーバのCPU種類，メモリのアーキテクチャ等があります．リソース量の決定には，環境状態を考慮する必要があります．ホスト混雑度を例として，環境状態のパフォーマンスへの影響例を**図 4** に示します．ホスト混雑度の変動によって，VMの処理能力が変動する可能性があります．したがって，パフォーマンス要件を満たし続けるためには，ホスト混雑度を考慮して，VMに配置するリソース量を決定・調整する必要があります．

⑶ 運用ポリシー

クラウドサービスの運用ポリシーは，サービス事業者またはクラウド提供事業者が定めます．運用ポリシーを定めることで，クラウドサービスの運用状態を望ましい状態に維持することができます．VM内のリソース利用率制限の運用ポリシーを定めることにより，パフォーマンス要件達成，ボトルネック防止，およびリソースの効率的配置等の効果が期待できます．以上より，該当ワークロードとパフォーマンス要件に対してリソース量を決定する際には，運用ポリシーを考慮することが重要となります．

リソース量算出部は，ワークロード，リソース量および環境状態から導いたパフォーマンスと運用状況の関係性を表すモデルを用いて，リソースの算出を可能とします．モデルにリソース量，ワークロード，環境状態を入力すると，パフォーマンスと運用状況の推測が可能です．

モデルを生成するために，クラウドサービスを提供・運用する際に収集した過去の監視データを用います．蓄積された監視データを整形し，機械学習の訓練データとして学習させます．学習の結果として，ワークロード，パフォーマンス，運用状況，環境状態とリソース量の関係性を表すモデルが生成されます．

モデルを用いたリソース量算出部の構成および算出ロジックを**図 5** に示します．モデルは事前に教師データを用いて学習します．リソース量算出部は，まず，ワークロード（サービス事業者の設定あるいは監視データから抽出），環境状態（リアルタイムに監視データから抽出），vCPU数およびメモリ量等を組み合わせてモデルの入力とします．それぞれのリソース量の組合せに対して，モデルを用いてパフォーマンスおよび運用状態を推測します．モデルから算出した結果をパフォーマンス要件（サービス事業者が設定），運用ポリシー（サービス事業者あるいはクラウド提供事業者が設定）と比較し，要件を満足するリソース量の組合せを抽出し，リソース量算出部の出力とします．

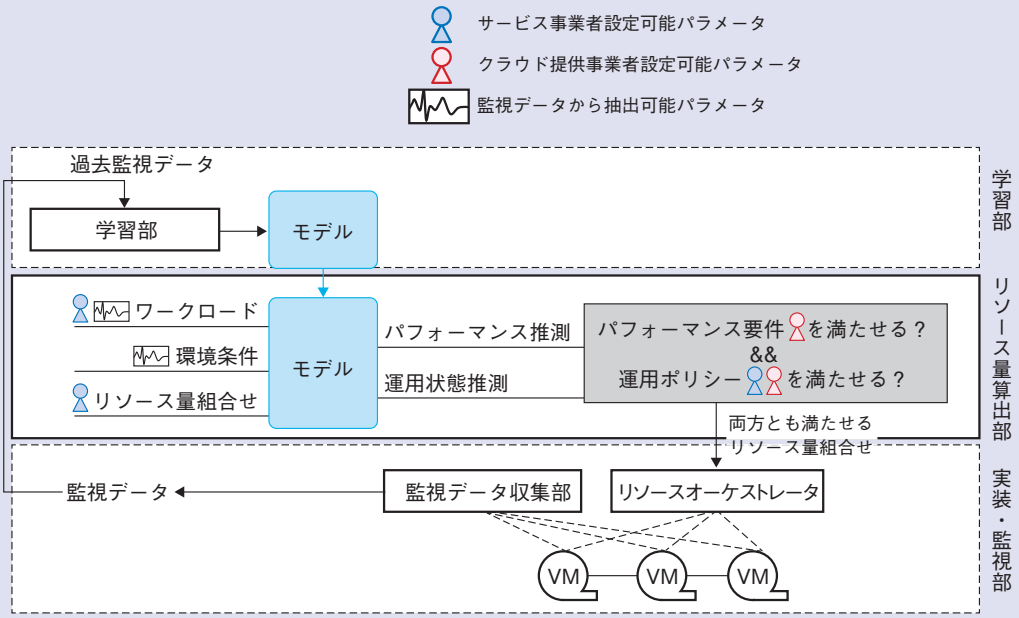

図5　リソース量算出部の構成

技術の活用例

IBSMフレームワークはクラウドサービスの提案・開通・運用業務において，いろいろな場面での支援を可能とします．例えば，既存のオンプレミス型のサーバから，クラウド環境への移行を検討しているお客さまに対して，クラウド環境にサービスを移行した場合のパフォーマンスを予測し，提示することが可能です．また，期待するパフォーマンスの達成に必要なリソースを提示することにより，適切なリソース・費用の提案・提示に活用可能と考えています．クラウドサービス提供時のリソース構成の設計およびリソース量の算出を自動化することにより，サービス提供時間の短縮や開通の自動化および有スキル人材の稼働抑制などが期待できます．さらに，サービス運用時における運用環境の変動に応じたリソース構成およびリソース量の適正化により，サービス要件の維持を可能とし，サービス事業者およびエンドユーザの満足度の向上にも貢献します．

今後の展開

ここではAS研で取り組んでいる，クラウドサービスにおけるリソース設計技術について紹介しました．今後は，クラウドサービス環境での代表的なワークロードを対象に，監視データを用いた本技術の有効性の確認，サービス提案，開通，運用等の異なる状況に応じた必要機能の精査等を行い，トライアルを通じて実運用でのフィージビリティを高め，実用化を進めていく予定です．

■参考文献

⑴ 呉・堀内："仮想化ネットワーク管理の複雑性に対するIBSM（Intent-Based Service Management）の検討，" 信学技報，Vol.117，No.305，pp.41-46，2017.
⑵ C. Wu and S. Horiuchi："Intent-based Service Management，" Proc. of ICIN 2018，Paris，France，Feb. 2018.

（左から）堀内　信吾／ 呉　　超／
田山　健一

クラウドサービスの迅速な提供と運用の効率化を実現する技術開発をとおして，クラウドサービスの普及・拡大に貢献していきます．

◆問い合わせ先
NTTアクセスサービスシステム研究所
アクセスオペレーションプロジェクト
オペレーション方式SEグループ
TEL　0422-59-3030
E-mail　fteam-ml@hco.ntt.co.jp