データカタログをベースとしたアクセス制御管理支援機構の提案

増田　博亮† 馬場　恒彦† 谷川　桂子†

†株式会社日立製作所研究開発グループ　中央研究所 〒185-8601 東京都国分寺市東恋ヶ窪 1-280
E-mail: hiroaki.masuda.vh@hitachi.com

あらまし　データ利活用の重要性が増す一方で，データの扱いに関する規制遵守義務が高まっており，セキュアなデータアクセスの実装に対する重要度が増している．アクセス制御に関し，データの所在は多岐に渡り，マルチクラウド環境や，種々データストアに跨ったデータに対し設定する必要がある．しかし，アクセス制御定義はクラウド・ストア毎で異なる上に，格納データの変遷に追従して，管理者が個々に設定する工数が課題となる．本稿では，データの属性情報を活用し，異種クラウド・ストアに跨るデータのアクセス制御を一元的に管理するフレームワークを提案した．提案手法による動的なアクセス制御により管理者工数を削減可能である．

キーワード　データレイク，データカタログ，アクセス制御，メタデータ

1. はじめに

様々なモノが通信によって繋がる IoT システムの普及に伴い，様々な事象のデータを収集することが可能となっている．収集された大量のデータを分析することによって，生産効率の向上といった新たな価値の創出が可能となった．そうした中で，データ利活用の幅を広げるためにデータレイクの構築が注目されている．

データレイクは，組織内の横断的なデータ利活用を目的として，多様なデータを種々の形式で論理的に格納・管理している．また，データカタログと連携することによって，データの所在等のメタデータを視覚化することが可能となり，組織内での横断的なデータ利活用に貢献している．データの所在は，同一の組織内であっても管理する部門等によって異なる．組織内システムのクラウド化も進んでおり，オンプレミス環境やクラウド環境，もしくはそのハイブリッド環境やマルチクラウド環境といった可能性が考えられる．

一方で，法制度やコンプライアンス遵守義務の観点から，セキュリティに関する対策が組織にとって必要不可欠な項目となっている．アクセス制御の集中管理が運用面と管理面から注目されており，アクセス制御の手法としてロールベースアクセス制御(RBAC:Role-Based Access Control)モデル[1][2][3]により設定されたアクセス制御ポリシーが，組織の情報システムで実践されている．しかし，アクセス制御ポリシーの設定に際し，データ管理者あるいはデータ管理者の指示の下，データ運用者がアクセス制御ルールの設計と適用等を行う必要があり，管理工数の増大が課題となっている．加えて，前述のようにセキュリティ管理対象がハイブリッド/マルチクラウド環境となった場合，各クラウドが提供するアクセス制御ポリシーとの統合が必要となるため，データ管理者の工数はさらに大きくなると考えられる．

本論文では，ハイブリッド環境のデータレイクに対し，データを視覚化するデータカタログを起点として統合的なデータアクセス制御を行うことで，アクセス制御管理に関する工数を削減することを目指す．2 章では，ハイブリッド/マルチクラウド環境下におけるアクセス制御管理モデルの関連技術について示す．3 章では，データレイク実装によるデータ利活用のユースケースとその課題について示す．4 章では，提案手法であるデータカタログをベースとしたアクセス制御管理支援機構と構成する機能について示す．5 章では，提案手法を適用した際の試算ベースの評価を示す．6 章で，本論文を纏める．

2. 関連研究

ハイブリッド/マルチクラウド環境における統合的なアクセス制御方式に関する研究が進んでおり [4][5]．各ベンダの提供するアクセス制御機構に適応したフレームワークの研究が行われている．

また，各ベンダのアクセス制御機構に依存せず，ハイブリッド/マルチクラウド環境のデータに対するアクセス制御方式についても研究されている．マルチクラウドのソフトウェアが稼働するサーバ統合環境に対し，アクセス制御設定の統合管理基盤が開発されている[6]．制御対象のリソース情報の操作制限等が記載されたリソース情報を収集し，共通形式で書かれたアクセス制御設定を，対象依存のアクセス制御リストに自動変換・付与することで，多様なソフトウェアで統合的にアクセス制御できるとしている．

Banyal らは，マルチクラウドのセキュリティとプライバシー問題に対処するための柔軟かつ効率的なアクセス制御フレームワークを提案している[7]．提案フレームワークは，静的および動的な信頼要素を用いて信頼値を算出し活用することで不正な操作を保護し，承

認されたユーザがアクセス可能としている．

Komninos らは，種々の属性からなるセットに対しアクセス制御ポリシーを紐づけることで，データアクセスを制御するフレームワークを提案している[8]．Li らは，プロキシベースのマルチクラウド環境用に属性ベースのアクセス制御システムを構築し[9]，分散アクセス制御を実現する MACPABE スキームを提案している．

いずれの研究に関しても制御対象のデータに対し，アクセス制御ポリシーの適用方式に関しての研究であり，アクセス制御ポリシーに関してはデータ管理者によって定義済みであることを前提としており，データ管理者の工数については触れていない．また分散アクセス制御方式は，セキュリティ問題への対応が困難となる場合がある．

Demchenko らは，組織内のマルチクラウド環境の統合アクセス制御に関して，ゲートウェイ方式を用いたアーキテクチャを研究している[10]．組織内のデータアクセスを集中的に管理することによる運用面での価値を主張している．しかし，本論文で課題として捉えているデータ管理者の管理工数に関しては考慮されておらず，依然として実ケースとのギャップが存在する．

3. データレイク実装によるデータ利活用ケースにおける課題

本提案では，企業の部門間を跨ったデータ管理システムのデータレイク管理における課題を述べる．

企業における業務や分析は，各部門内で閉じて行われる場合が多い．収集されたデータは自部門内のオンプレミス環境に格納，管理されるため，同部門に所属する者は，内容と格納場所について把握可能であるが，関係者以外はデータの存在についても知ることができない．データカタログを利用することにより，関係者以外の分析者のデータを発見･活用を支援する一方で，機密度の高いデータに関しては，存在を知るユーザを絞りたい場合や，開示できる範囲を設定する場合等のアクセス制御を管理できることが望ましい．しかし，アクセス制御ルールは静的なため，人やデータの流動が生じる場合，それらに追従してルールを更新する必要がある．

これらを踏まえ，データレイク管理の課題を述べる．アクセス制御方法の一つとして，データ管理上のロールを定義し，ロールに対してアクセス制御ルールを付与する RBAC がある．各ユーザの属性情報からロールを定義し，データの属性情報と実行可能なアクションを紐づけることで，アクセス制御ルールとする．データ管理者とデータ運用者は，ハイブリッド環境に適応しつつ，人やデータの流動に追従してルールを更新する必要があり，これらを手作業で行う場合，管理工数が大きくなる課題がある．そのため，アクセス制御管理業務を支援する機能が必要である．

4. 提案手法

本章では，3 章の課題を解決するための方針と，提案するデータカタログベースのアクセス制御管理アーキテクチャ，及びその機能について述べる．

データカタログは，データの属性情報や，関係者外に対してデータ開示範囲に関するデータ所有者の意向を管理可能のため，データカタログをベースに関連するアクセス制御ルールやデータアクセス API の情報をデータカタログと紐づけて管理できると考える．ユーザに付与するロールは，ユーザの属性情報から判断する必要があるため，ユーザが所属する会社の持つ認証機構に登録されているユーザ情報を参照して，ロールを自動的に生成，付与を行うことによって，組織内の人の流動に対して，動的にアクセス制御を提供する．

データカタログに含まれるアクセス制御に関する情報には，データカタログとして公開する対象，範囲及び，アクセス可能な対象データと操作可能アクションを示すとする．これにより，データカタログにデータアクセス API やロールを紐づけることで，データカタログをベースとしてアクセス制御ルールを生成，及び管理を行う．

また，連携する外部ソフトウェアであるデータカタログ，データアクセス API，認証認可ツール等に依存しないアクセス制御機能を提供する．外部ソフトウェアとのインターフェイスをマイクロサービスとして定義することにより，外部ソフトウェアの環境に依らず，要件を満たす API で必要情報を取得する．

4.1. Access Control Management Achetechture

ハイブリッドクラウド環境のデータレイクにおけるアクセス制御アーキテクチャを図 1 に示す．Access Control Management(以下 ACM)は，データ管理者のアクセス制御管理業務を支援するため，関連する各モジュールと API にて連携する．

ACM は，組織共通のプラットフォームとして機能し，随時更新されるユーザの属性情報，データカタログ，データアクセス API に適応してアクセス制御ルールを更新，適用することにより，データカタログとデータソースへのアクセス制御を管理する．

4.2. 機能詳細

ACM が提供する機能の詳細について述べる．図 1 中の各番号と提供機能が対応関係にある．

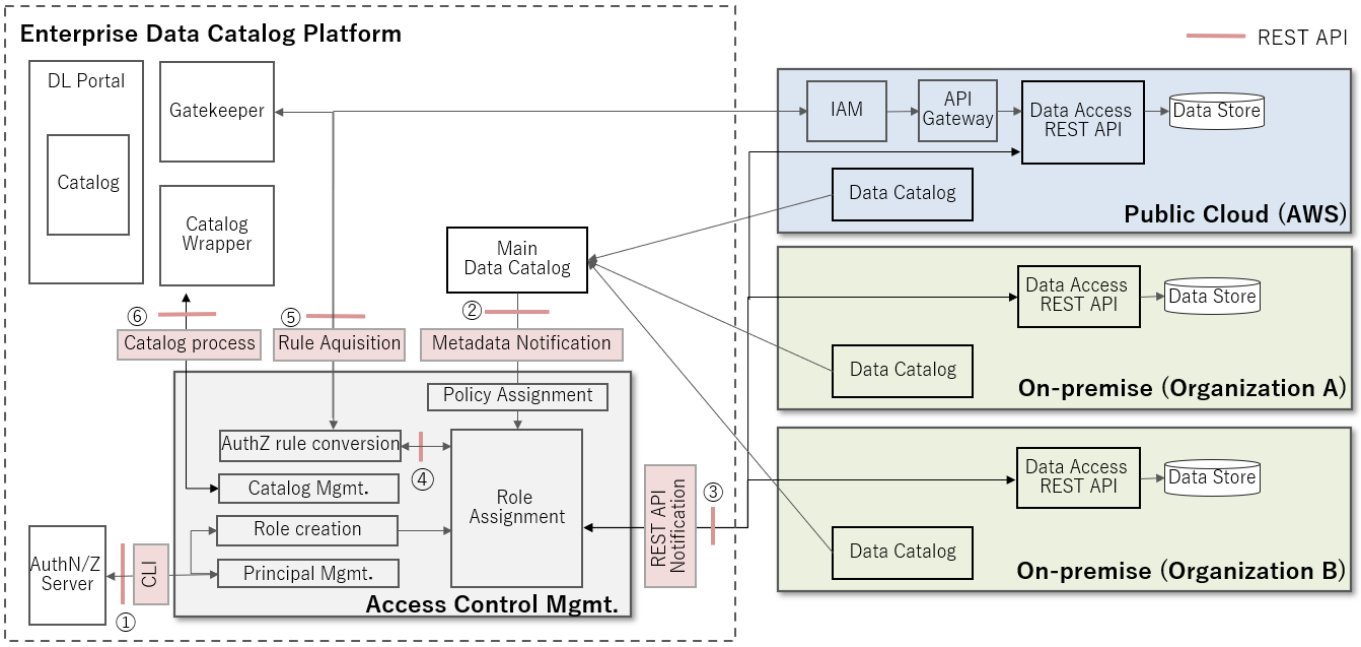

図 1 データカタログベースのアクセス制御アーキテクチャ

① 認証認可に用いるユーザ情報の取得機能

ユーザ属性管理サービスに登録されている既存のユーザ属性情報を取得し，認証認可機構と連携して管理を行う．認証認可機構では，ユーザ登録情報を用いてデータカタログ，データソースへのアクセスを要求するユーザのアクセス可否を判断する．図 1 の API-①では，ユーザ属性情報を基に，連携する認証認可機構上でアクセス制御に用いるロールをユーザに対して付与する．

② データカタログ連携アクセス制御情報管理機能

データカタログの変更が行われた際，変更内容を受け ACM 内で管理するアクセス制御ルールを更新する．

一つのデータカタログに登録される情報が，同部門に属する複数の課の営業データである場合，一つのデータカタログに複数のデータ所有者が存在することが予想される．データカタログにおいて上記の情報をデータ所有者毎に複数持たせることにより，データアクセス制御情報を適宜抽出し，アクセス制御ルールへと反映することができる．

③ データストアへの統合アクセス API 管理機能

データカタログからアクセス先のデータソースのメタデータを参照し，アクセス制御ルールを生成を行う．

本機能では，データアクセス API から，アクセス先のデータのアクセス制御情報を参照して，個別にアクセス制御ルールを生成する．例えば，データアクセス API としては一つのテーブルの読み込みであるが，テーブル内の A カラムと B カラムで別のアクセス制御情報が存在する場合，A カラムに対応するルールと B カラムに対応するルールを個別に生成し，アクセス API の結果をユーザに返す際にルールを参照して加工する．これにより，同一テーブルのカラム内のデータに対して，同クエリ API を実行した場合であっても，実行したユーザのロールに準拠して異なるアクセス制御を実現できることを意味し，粒度の細かいアクセス制御を提供する．

④ 認可サーバーのサポートするルール表記への変換機能

機能②と③により，生成したアクセス制御ルールを認証認可ツールのルール表記への変換を行う．本機能はプラグイン形式であり，連携する認証認可ツールのルール表記へ変換方式に応じてプラグインを追加することが可能である．ACM が管理するアクセス制御ルールは，本システムにおけるアクセス制御機能に必要な最低限の情報を持っており，変更が起こる度に本機能が呼び出され，変換を行う．

⑤ 認可機構へのアクセス制御ルール反映機能

機能④による，認可サーバーの表記方式への変換内容を認可サーバーに書き込みを行う．書き込み先の認可サーバーは，管理者が事前に設定を行うことにより定義されるものとする．本機能は，プラグイン形式であり，連携するアクセス制御機構に応じてプラグインを追加することにより，対応可能なアクセス制御機構のスケーリングが可能である．

⑥ データカタログへのアクセス管理機能

ユーザのロールから，必要に応じてデータカタログに表示して公開する情報を加工し，その結果を返すことによってデータカタログに関するアクセス制御を行う機能である．

4.2.1. アクセス制御処理フロー

図 2 に，ACM の処理フローを示す．ACM は，データ管理に係る以下の場面において，データ管理者を支援する．

・ACM 起動時

認証認可ツールからユーザ情報を取得し，ユーザに対して付与する．データカタログからアクセス制御情報を取得し，アクセス制御ルールを生成/反映する．

・データカタログ更新時

更新内容を ACM が管理するアクセス制御ルールへ反映する．認可サーバーとカタログアクセス制御ルールを更新する．

・データアクセス API 更新時

更新内容を ACM が管理するアクセス制御ルールへ反映する．認可サーバーのアクセス制御ルールを更新する．

5. 提案方式の評価

本章では，データカタログのアクセス制御の評価について述べる．アクセス制御管理の容易化に対し，提案方式であるデータカタログベースのアクセス制御アーキテクチャの効果検証のため，管理者の工数観点で評価を行った．

図 3 にデータ収集からアクセス制御の適用，及びデータカタログとして運用するフローを示す．図 3 の左部は，新しいデータを格納する際のフローを，図 3 の右部は，既にデータカタログに登録済みのデータに関する変更後のフローを示す．各作業における管理者の工数を時間として試算し評価する．試算に利用する各数値は，DB 内に存在するテーブル 1400 件，テーブルのカラム 35000 件とし，それぞれのデータをデータカタログに登録し，アクセス制御ルールを設定する．各定量値の計算根拠を示す．

・アクセス制御ルールの設定に必要な要素の登録

データ所有者の意向に沿ってアクセス制御ルールの要素となるユーザ・ロール・リソースを登録する．GUI 等の環境がある場合，所要時間を短縮可能であるが，基本的には全て管理者の手作業によって実現される．また,他部門のユーザに対してロールを付与する場合，ロールの付与が適切かどうかについて，組織内の人材データを通じて調査する必要がある．そのため，12 時間かかるものとした．提案手法である ACM を活用する場合では，ユーザ情報を認証認可ツールか得し，ロールを自動的に生成・ユーザに付与するため，3 分かかるものとした．

・データカタログへの反映の依頼

データカタログの管理を行うデータ運用者にデータカタログ上への反映依頼に関して，3 分かかるものとした．

・データカタログへの反映

格納されたデータのデータカタログの登録を行う必要がある．複数のデータ運用者が行う作業であり，情報のやり取りが必要となるため，12 時間かかるものとした．

・アクセス制御ルール設定・適用の依頼

データ運用者にアクセス制御ルールの設定・適用の依頼に関して，3 分かかるものとした．

・アクセス制御ルール設定・適用

ユーザ・ロール・リソースの 3 要素を基に，アクセ

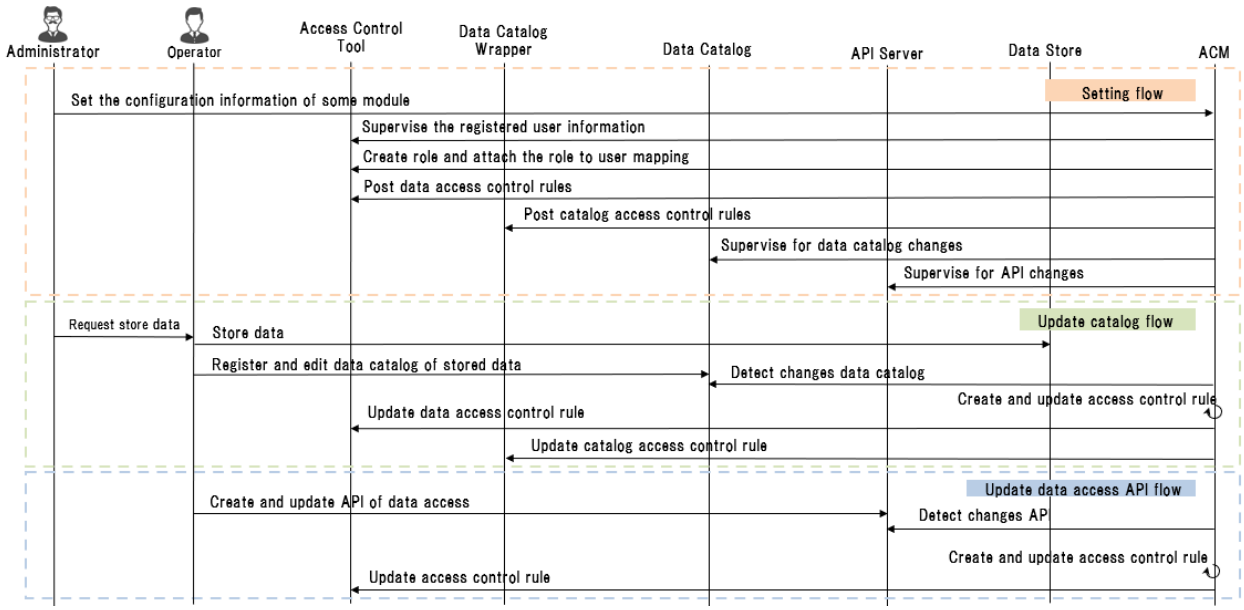

図 2　ACM のアクセス制御管理フロー

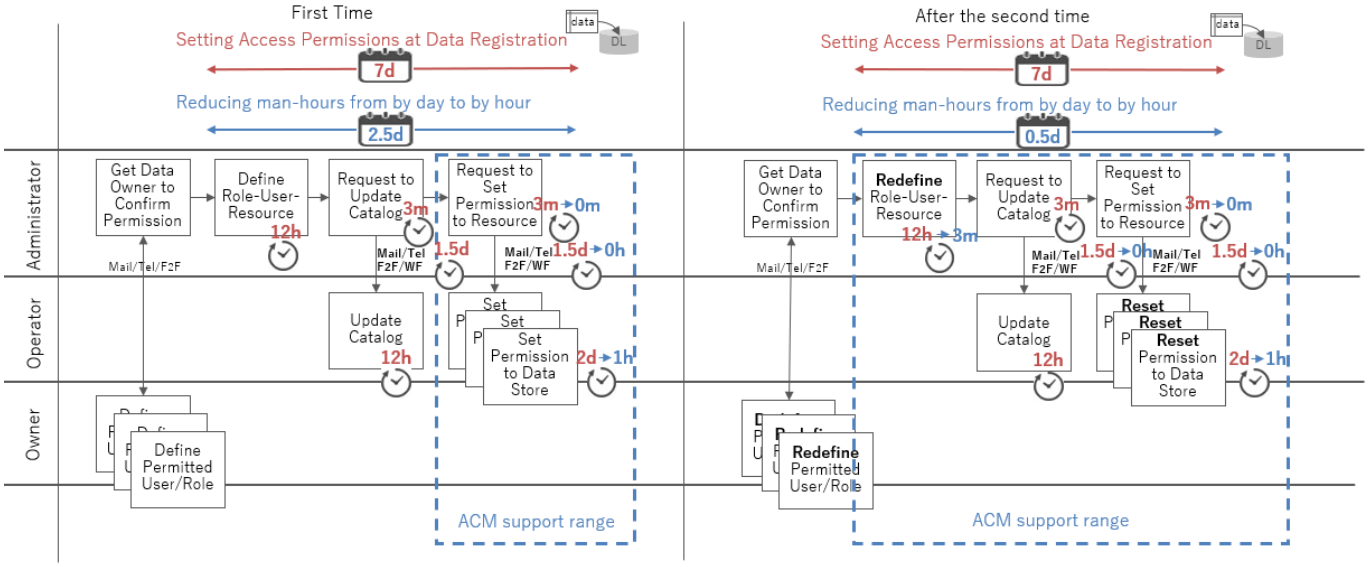

図 3 提案手法によるアクセス制御管理の工数削減効果

ス制御ルールを設定，適用する．適用に際しルールの変更の有無等の確認事項があった場合，データ管理者に問い合わせする必要があることから，2 日かかるものとした．ACM を活用する場合では，アクセス制御ルールの設定と適用を自動化するため，1 時間とした．

データ所有者の意向に沿ったアクセス制御ルールの生成・適用に至る，上記項目の工程を完了するためには，データ管理者間の会話が必要である．各機会毎にリードタイムとして 1.5 日かかるとし，ユーザ・ロール・リソースの 3 要素の定義から，アクセス制御ルール適用まで全体で 7 日間かかるものとした．ACM を活用する場合，データカタログの初回登録時は，全体工数が 2.5 日間となり，管理工数は 65%削減する見込みである．また，データカタログが既登録の場合，ACM のサポート範囲が拡張され，全体工数が 0.5 日間となり，データ管理者の工数は 93%削減する見込みである．

6. 結論

ハイブリッドクラウド環境のデータレイク利活用ケースにおいて，データ管理者及びデータ運用者のアクセス制御管理に関する工数削減のため，アクセス制御管理支援機構が必要である．本論文では，データカタログとアクセス制御の紐づけにより，データ管理者の管理業務の支援として，人とデータの流動に追従して動的にアクセス制御ルールを生成し，反映するデータカタログをベースとしたアクセス制御管理支援機構を提案した．提案システムでは，データ管理者および運用者のアクセス制御管理に関する工数の試算を行い，比較評価の結果 93%の工数削減効果を見込んでいる．

参 考 文 献

[1] Ferraiolo, D. and Kuhn, R.:Role-Based Access Control, Communications of the 15th NIST-NSA National Computer Security Conference (1992).

[2] Ferraiolo, D.,Sandhu, R., Gavrila, S. and Kuhn, R.:Proposed NIST standard for Role-Based Access Control, *ACM Trans. Information and System Security*, Vol.4 No.3 (2001).

[3] Ferraiolo, D.,Kuhn, R.and Chandramouli, R.:Role-Based Access Control Second Edition,Computer Security Series, ARTECH HOUSE (2007).

[4] Lee, Craig A. "Cloud federation management and beyond: Requirements, relevant standards, and gaps." *IEEE Cloud Computing* 3.1 (2016): 42-49.

[5] Sette, Ioram S., David W. Chadwick, and Carlos AG Ferraz. "Authorization policy federation in heterogeneous multicloud environments." *IEEE Cloud Computing* 4.4 (2017): 38-47.

[6] 芦野佑樹, and 中江政行. "統合アクセス制御モデルの標準化について." *第 73 回全国大会講演論文集* 2011.1 (2011): 261-262.

[7] R. Banyal, V. Jain, and P. Jain, “Dynamic Trust Based Access Control Framework for Securing Multi-Cloud Environment,” International Conference on Information & Communication Technology for Competitive Strategies, 2014, pp. 1-8

[8] N. Komninos and A. Junejo, “Privacy Preserving Attribute Based Encryption for Multiple Cloud Collaborative Environment,” Ieee/acm International Conference on Utility

[9] M. Singhal, S. Chandrasekhar, T. Ge, and et al. “ Collaboration in Multicloud Computing Environments: Framework and Security Issues,” Computer, vol. 46, 2013, pp. 76-84.

[10] Demchenko, Yuri, et al. "Federated access control in heterogeneous intercloud environment: Basic models and architecture patterns." *2014 IEEE International Conference on Cloud Engineering*. IEEE, 2014.