令和 5 年度厚生労働行政推進調査事業費補助金（政策科学総合研究事業（政策科学推進研究事業））
臨床疫学に活用可能なNDB等データセットの作成に関する研究（21AA2006）

分担研究報告書

NDB オンサイトリサーチセンターと比較した HIC

研究協力者　松居宏樹　東京大学大学院医学系研究科臨床疫学・経済学准教授
研究分担者　康永秀生　東京大学大学院医学系研究科臨床疫学・経済学教授

研究要旨　この報告書では、NDB オンサイトリサーチセンターにおける大規模医療データを用いた研究の成果について説明する。さらに、NDB データを取り扱う際の利用方法である、NDB オンサイトリサーチセンターと HIC について、その特性の違いを説明する。そのうえで、NDB オンサイトリサーチセンターを基準とした場合に HIC が抱えている課題として、データ記憶領域の不足と、データベース応答速度の問題を明らかにした。それを解消する糸口となる技術として、PostgreSQL の citus 拡張を用いたデータの圧縮、および partitioning について検証を行った。その結果、データの圧縮と partitioning の併用はデータベース応答速度を上げることがわかった。HIC は今後 NDB の利用形態として重要な位置を占めると考えられるものの、NDB オンサイトリサーチセンターと比較すると、現状では HIC 特有の課題が存在する。HIC を用いてデータベース研究を行うためには、ユーザー側の十分な技術的工夫が必要であると考えられる。

A．研究目的

匿名医療保険等関連情報データベース（NDB）が研究目的で国内研究者に提供され始めて 10 年以上が経過した。一部の研究者の間では、NDB を用いた臨床疫学研究が実施され、ハイインパクトジャーナルに掲載されるなど、その存在価値は増している。しかし、多くの研究者にとっては、そのデータ利用に伴う制約の多さから利用が難しい側面があった。提供開始当初、NDB は特別抽出が主要なデータ利用方法として活用されてきた。特別抽出は、ユーザーがデータ利用環境を独自に用意し、その環境でデータを解析する。データ利用環境には、十分なセキュリティ要件を満たす必要がある。しかし、一部の研究者にとってそのような環境を確保することは困難であった。そういった研究者に向けて、NDB オンサイトリサーチセンターが国内数か所に開設された。現在、データ利用環境がクラウド上に構築され、各 NDB オンサイトセンターに向けて利用環境が提供されてい

る。我々は、東京大学 NDB オンサイトリサーチセンターにおいて、複数の臨床疫学研究を実施し、成果を上げてきた。
また、データ利用環境をクラウド上に構築し、各ユーザーに向けて利用環境を提供する Healthcare Intelligence Cloud (HIC)が、2024 年から運用を開始する予定とされている。HIC は、ユーザーがオンサイトセンターに出向かなくとも、ユーザー個人の端末からクラウド上の仮想環境に接続することで、NDB の解析を可能とするものである。しかしながら、それぞれのシステムにおける制約について利用者に十分な理解が得られていないため、その環境を十分に活用できず、成果が上がらない事例も散見される。この報告書では、まず、我々が NDB オンサイトリサーチセンターを活用して挙げてきた成果を紹介する。次に、HIC 本格利用に先立ち HIC の試行利用を行い、NDB オンサイトリサーチセンターでの利用方法を基準として、HIC のシステム上の制約下で成果を得るために必要となる工夫について検討したので報告する。

Ｂ．研究方法

（1）NDB オンサイトセンターでの成果

2022 年以前は、NDB データは厚生労働省のオンプレミス環境に構築された Oracle 社製データベースに保存されていた。ユーザーはオンサイトセンター内から、専用回線を用いて Oracle 社製データベースを操作し、データをハンドリングし解析を行っていた。2022 年以前にオンサイトセンターを用いて実施した研究には、以下がある。

1. Ishimaru M, Matsui H, Ono S, Hagiwara Y, Morita K, Yasunaga H. Preoperative oral care and effect on postoperative complications after major cancer surgery. British Journal of Surgery. 2018 Oct 12;105(12):1688–96.
2. Hashimoto Y, Matsui H, Michihata N, Ishimaru M, Yasunaga H, Aihara M, et al. Incidence of sympathetic ophthalmia after inciting events: a national database study in Japan. Ophthalmology. 2021 Sep 21;S0161-6420(21)00719-3.
3. Takeuchi Y, Kumamaru H, Hagiwara Y, Matsui H, Yasunaga H, Miyata H, et al. Sodium-glucose cotransporter-2 inhibitors and the risk of urinary tract infection among diabetic patients in Japan: Target trial emulation using a nationwide administrative claims database. Diabetes Obes Metab. 2021 Jun;23(6):1379–88.
4. Ishimaru M, Ono S, Morita K, Matsui H, Hagiwara Y, Yasunaga H. Prevalence, Incidence Rate, and Risk Factors of Medication-Related Osteonecrosis of the Jaw in Patients With Osteoporosis and Cancer: A Nationwide Population-Based Study in Japan. Journal of Oral and Maxillofacial Surgery. 2022;80(4):714–27.
5. Kasajima M, Eggleston K, Kusaka S, Matsui H, Tanaka T, Son BK, et al. Projecting prevalence of frailty and dementia and the economic cost of care in Japan from 2016 to 2043: a

microsimulation modelling study. The Lancet Public Health. 2022;7(5):e458–68.

2022 年 4 月から、NDB システムは AWS クラウド上に移行され、オンサイトセンターもクラウド上のシステムへの接続に変更となった。ユーザーの利用形態も、オンサイトセンター内から専用回線を用いてクラウド環境にある Redshift （AWS 社製高速分散 DWH）を操作し、データをハンドリングし解析を行う形へ変更となった。2022 年以降にオンサイトセンターを用いて実施した研究には、以下のようなものがある。

6. Kimura Y, Suzukawa M, Jo T, Hashimoto Y, Kumazawa R, Ishimaru M, et al. Epidemiology of severe childhood asthma in Japan: A nationwide descriptive study. Allergy. 2024 Jan 7;all.16008.
7. Kimura Y, Jo T, Hashimoto Y, Kumazawa R, Ishimaru M, Matsui H, et al. Epidemiology of patients with lymphangioleiomyomatosis: A descriptive study using the national database of health insurance claims and specific health checkups of Japan. Respir Investig. 2024 May;62(3):494–502.

この報告書では特に、上記 6 の研究についてそのデータ特性の観点から紹介する。

（2）HIC と NDB オンサイトセンターの比較軸

研究者が利用可能な NDB の提供方法には、サンプリングデータセット、特別抽出、集計表、オンサイトリサーチセンター利用（i），オンサイトリサーチセンター利用(ii), HIC の 6 種類の方法がある。

サンプリングデータセットは、年間複数時点のレセプト横断データからあらかじめ一定割合でランダムに抽出した個票レベルレセプトを提供する方法である。

特別抽出は、レセプトデータから、利用者の指定する特定条件で抽出した個票レベルレセプトを提供する方法である。

集計表は、レセプトデータを、利用者の指定する特定条件で集計した集計表レベルの情報を提供する方法である。

オンサイトリサーチセンター利用(i)は、オンサイトリサーチセンターで NDB データに直接アクセスし、データの解析をオンサイトセンター内で終える方法である。

オンサイトリサーチセンター利用(ii)は、オンサイトリサーチセンターで NDB データに直接アクセスし抽出した個票レベルレセプトを、オンサイトセンターから取り出し、利用者が用意した自身の環境で解析する方法である。

それぞれの方法で特性が異なっているが、今回は特にオンサイトセンター利用(i),(ii)と HIC に対象を絞り、以下のポイントで整理行う。

1) 利用者に求められるセキュリティ要件の違い
2) データ利用可能になるまでの時間
3) 使用可能データの違い
4) システム上の制約の違い
   ・システムのデフォルトのスペック
   ・システムの拡張性

・ソフトウェア
・利用可能なクラウドリソース

また、上記のポイントで各提供方法を整理したうえで、NDB オンサイトリサーチセンターと HIC のシステム上の制約で特に着目すべき点を挙げる。

（3）HIC のシステムを利用するうえでの工夫とその使用感

我々は当研究班の昨年の報告書において、HIC でデータを解析する場合に必要となる工夫を以下のように提言した。

- Linux 環境が用意できる場合、適切な拡張機能のインストールが望ましい。
- S3 上に保存されたデータを直接 PostgreSQL に load するためには、PostgreSQL 拡張である、s3csv_fdw のインストールが望ましい。
- PostgreSQL において、テーブルデータの圧縮を行うことが望ましい。NDB オンサイトリサーチセンターで実績のある citus database のインストールが望ましい。
- Docker 環境を立ち上げたうえで解析環境を整備することが可能であるならば、そうすべきである。

今年度は、本研究班代表者の尽力により、PostgreSQL の拡張である citus (2)を用いて、テーブルデータを圧縮した NDB データベースが作成された。このデータベースの課題と、その解決のための工夫を提案する。

C．研究結果

（1）NDB オンサイトリサーチセンターでの成果

上記研究 6 において、我々は、NDB を用いて小児重症喘息患者の全国的有病率の記述研究を行い、有病率の経時的変化を明らかにした。

この研究において、診断コードと喘息関連薬の処方状況を組み合わせて喘息児を定義した。

その結果、2019 年には、0～5 歳の喘息児 253,684 人のうち、253,052 人（99.8%）、632 人（0.2%）がそれぞれ軽度～中等度、重度のグループに分類された。

重症小児喘息の全国有病率は、調査期間中に 0～5 歳（10 万人当たり 19.7 人から 11.0 人）および 6～11 歳（44.4 人から 22.8 人）の両年齢層でほぼ半減したが、軽度から中等度の喘息はより緩やかな減少傾向を示した。

（2）NDB オンサイトリサーチセンターでのデータハンドリング

この研究のデータハンドリングにおいて重要なポイントは、診断コードと喘息関連薬の処方状況を組み合わせて喘息児を定義した点である。**図 1** に示すように、症例を検索する際に、「診断コードを有するレセプト」と「喘息関連薬の処方があるレセプト」は単一レセプト内で完結しない場合がある。例えば、患者の診断名と処方状況が同一以下レセプトではなく、医科・調剤で分かれて処方される場合や、診断がついた翌月にたまたま処方薬を調剤した場合である。そのため、この研究においては、まず、条件に該当するレセプトを特定した。

次に、そのレセプトを算定した症例のレセプトデータを、時系列に沿って追跡した。また、時系列の追跡精度を向上させる目的

で、図 2 にあるように、再帰的な ID 追跡を行った。再帰的 ID 追跡は、抽出条件を絞って行う特別抽出データでは実施された報告がない。そのため、条件を絞らずにレセプトデータを取得する特別な研究か、オンサイトリサーチセンターでのみ行うことが可能な抽出方法である。

さらに、我々のチーム内の研究者がデータを取り扱いやすくするために、データを時系列に整理するプログラム（NDB データ抽出プログラム）を作成し、研究者に整理されたデータを共有している。

NDB データ抽出プログラムは、Redshift 上に保存された NDB オンサイトリサーチセンターからアクセス可能な NDB データに対して動作する、Postgresql の Stored Procedure として構築されている。NDB データは規模が大きいため、そのデータを整形する際に適切な処理手順を踏まなければ、データの処理が不可能になる。

NDB オンサイトリサーチセンターからアクセス可能な NDB データは、月単位（PRAC_YM 単位）でパーティショニングがなされている。そのため、抽出プログラムのアルゴリズムを各 PRAC_YM に対して独立して動作するプログラムとして構築した。プログラムのプロセスは以下のとおりである。

1.キーレセプトの取得

(1)特定条件に合致するキーレセプトのレセプト ID(i)を取得

2. ID 追跡

(1)レセプト ID(i)に紐づく ID1N(i)を取得

(2) ID1N(i) とペアとなる ID2(i) を取得

(3) ID1N(i)ID2(i)から紐づ ID1N(ii),ID2(ii) をすべて取得しする（再帰的処理）。

(4) ID1N(ii), ID2(ii)の中で、ID の接続を確認して ID0 と ID1N(ii) の対応を取得する（再帰的処理）。

3.患者詳細情報の取得

(1)任意の PRAC_YM(i) に対して、ID1N(iii) に紐づくレセプト ID(ii) を取得する。

(2)レセプト ID(ii)に紐づく傷病情報・診療情報を取得する。

一連の患者情報を整理取得する際には、上記のプロセスを経て抽出を実施している。特に 3 のプロセスはすべて、任意の PRAC_YM に対して逐次的に実行できる。そのため、対象となる PRAC_YM(n) が増えても処理時間は n に比例して増えるだけで、指数関数的に処理時間が延びることはない。

（4）NDB オンサイトリサーチセンターと HIC の比較

1.利用者に求められるセキュリティ要件の違い

NDB オンサイトリサーチセンターも HIC も、データ本体は AWS 上の閉じられた VPC(https://aws.amazon.com/jp/vpc/) に構築されており、外部からアクセスすることはできない。NDB オンサイトリサーチセンターは HIC に比べ、物理的に隔離されたセキュリティレベルの高い環境からのデータへのアクセスが可能である。これに対して、HIC は物理的に隔離された環境を整備するわけではない。そのため、若干 HIC のほうがセキュリティの厳重さは低く設計がなされている。（医療・介護データ等解析基盤（HIC）の利用に関するガイドライン

https://www.mhlw.go.jp/content/12400000/001174849.pdf）

2.利用開始までの時間

NDB 利用方法によって、承認を受けてからデータ利用可能となるまでの時間が異なる。例えば、特別抽出では利用承諾を受けてからデータ利用可能となるまでに長大な時間が必要となる。令和 5 年 6 月の規制改革実施計画においては、利用申請から申請者が実際にデータの利用を開始するまでに平均で 1 年以上の時間を要することが課題としてあがっている。この原因として、研究に必要な NDB データのみを抽出したデータセットを研究者に提供していることが挙げられる。

NDB オンサイトリサーチセンターでは、承認後、事務処理を経てデータの利用を開始できる。これは、オンサイト向けデータセットがあらかじめ構築されており、利用者がそこから必要となる情報を選択して解析をするためである。

これに対し、HIC では、承認後、研究に必要な NDB データのみを抽出したデータセットを環境内に配置する形でデータ提供が行われる。そのため、NDB オンサイトリサーチセンターのように、あらかじめ構築されたデータセットを環境内に設置しておく対応をとらなければ、特別抽出と同様に利用開始まで長大な時間を要するものと考えられる。

3.使用可能データの違い

利用形態によって、利用可能なデータは異なる。

まず、NDB 本体に含まれるレセプト、特定健診データは、NDB オンサイトリサーチセンターでも HIC でも利用可能である。しかし、NDB 本体データに紐づけることができる介護データや DPC データ等のその他データは NDB オンサイトリサーチセンターでは利用することができない。

次に、NDB 本体データについても、NDB オンサイトリサーチセンターと HIC で利用可能な情報が異なる。NDB オンサイトリサーチセンターでは、直近 10 年分の NDB 本体データがデータベースとして保管されている。これに対し、HIC では、上述のように研究者が指定した条件で抽出した研究用データセットが環境内に設置される。

4.NDB 利用方法によるシステム上の制約の違い

昨年の報告書でもふれたが、NDB オンサイトリサーチセンター、HIC ともに、解析環境として m5.4xlarge（vCPU:16, memory:64GB, Strage: 3TB）か、m5.2xlarge（vCPU:8, memory:32GB, Strage: 1TB）が利用可能である。NDB オンサイトリサーチセンター、HIC ともに、OS は Windows か Linux を選択可能である。

NDB オンサイトリサーチセンター、HIC ともに、仕様で決定しているシステムの拡張は原則としてできない。そのため、NDB オンサイトリサーチセンター利用(ii) では、ユーザーが特別抽出と同等の環境を独自に用意することで、個票レベルのデータを独自システムで解析する仕組みが存在している。

また、システム仕様として決定しているハードウェア仕様に対して、ソフトウェアイ

ンストールは厚生労働省と協議のうえ、柔軟に実施できる。NDB オンサイトリサーチセンターでは、Linux 上に Docker 環境が用意されており、Docker コンテナを環境内に持ち込んで運用することができる。HIC においては、ソフトウェア持ち込みを申請して、厚生労働省の確認を経て、ソフトウェアのインストールをユーザー自身が行うことができる。この際、環境をインターネット上にオンラインにすることはできないため、ソフトウェアによっては利用ができない場合がある。

最後に、利用可能なクラウドリソースが HIC と NDB オンサイトリサーチセンターで異なっている。いずれの利用環境でも解析環境から S3 のバケットを参照できる。しかし、NDB オンサイトリサーチセンターでは、Redshift 上の NDB データにアクセスできるのに対し、HIC 環境では Redshift へのアクセスはできない。そのため、HIC では、大規模なデータをハンドリングするためには十分な工夫が必要である。

（5）HIC のシステムを利用する上での工夫

HIC を利用する上での課題は、データ記憶領域のサイズ問題が第一に挙げられた。そこで、この研究班では、昨年の報告書の提案に基づき Postgresql の拡張である、citus を用いたデータの圧縮を行った。citus はもともと、データの分散配置を行う事で大規模データの高速処理を行う目的で開発が進んでいた拡張である。Citus Data（https://www.citusdata.com/）が開発を行っており、2019 年に Microsoft に買収されたものの、オープンソース版が提供されている。また、分散配置を行う機能以外にも columnar table という機能が追加されており、テーブルを列単位で圧縮することで、データサイズの圧縮と全件検索の高速化を可能としている。今回の研究班においても、テキスト形式でおおよそ 21TB の容量がある 4 年分のレセプトデータを、2.6TB まで圧縮することができた。

しかしながら、columnar table では、INDEX を利用することができない制約がある。そのため、大規模なデータを Full Scan してしまいデータベースの応答速度が致命的に低下することがある。今回作成したデータベースにおいても、特定の診療年月の MED_SI レコードをスキャンし 10 行表示する以下のクエリを実行するだけで 1 時間以上の時間が必要であった。

```
SELECT * FROM med_si WHERE prac_ym =
'202204' LIMIT 10;
```

データベースの検索速度を向上するための工夫として、NDB オンサイトリサーチセンターの Redshift 上のデータベースでは、partitioning を行っていることはすでに述べたとおりである。今回、columnar table と partitioning を併用することで、ディスク記憶領域の課題を回避しながら、データの検索速度を上げることができるか検討した。検討プロセスは以下 4 つテーブルを作成したうえで、同一データを挿入し、ディスク容量と検索速度を比較した。データは 48 ヶ月分の dpc_si テーブルから、レセプト ID(seq2_no),レセプト電算コード (prac_act_code),prac_yo(診療年月)を選択し、格月 25 万行を挿入した。

1. partitioning を行わず、圧縮も行わないテーブル test_nonp_heap
2. prac_ym で partitioning し、圧縮を行わないテーブル test_p_heap
3. partitioning を行わず、圧縮を行うテーブル test_nonp_col
4. prac_ym で partitioning し、圧縮を行うテーブル test_p_col

Table 1 に、データの圧縮効率と検索速度（テストクエリ実行時間）を示した。わずかなオーバーヘッドはあるものの、partitioning を行った圧縮テーブル test_p_col でも、partitioning を行わない圧縮テーブル test_nonp_col と同等のテーブルサイズにデータを圧縮できた。

さらに、prac_ym を用いて検索範囲を絞ったテストクエリでは、Partitioning を行わない圧縮テーブル test_nonp_col に比べ、Partitioning を行った圧縮テーブル test_p_col の検索速度が大幅に向上した。

D. 考察

この報告書では、NDB オンサイトリサーチセンターにおける大規模医療データを用いた研究の成果について説明した。さらに、NDB データを取り扱う際の利用方法である、NDB オンサイトリサーチセンターと HIC について、その特性の違いを説明した。そのうえで、NDB オンサイトリサーチセンターを基準とした場合の HIC が抱えている課題として、データ記憶領域の不足と、データベース応答速度の問題を明らかにした。それを解消する糸口となる技術として、pstgresql の citus 拡張を用いたデータの圧縮、および partitioning について検証を行った。その結果、データの圧縮と partitioning の併用はデータベース応答速度を上げることがわかった。

NDB オンサイトリサーチセンターにおける抽出プログラムは、prac_ym で partitioning されたデータベースへの適用を前提として作成した。今回構築した圧縮データベースは実用に耐えるものではなかったが、適切な partitioning を行うことで既存資産を利用しながら、臨床疫学研究を実施することが可能であると考えられる。

NDB データの提供については、HIC での提供の実現で利用者の初期投資を大幅に削減できることが期待されている。実際に使用可能なデータの種類を考えると、HIC は重要な利用方法であると思われる。しかしながら、技術的な工夫を行わなければ、データが得られても現実的な時間で解析を行うことは不可能であることが、今回の試行利用から明らかとなった。このことから、ユーザー側も十分な技術的工夫を行う能力が求められる。また、ユーザー側の技術的工夫を補うための対応として、厚労省側が検討することができるオプションは複数残されている。

例えば、HIC において Redshift 等のリソースへのアクセスを許可することは、一つの方策であろう。ただし、Redshift は従量課金のシステムであり、ユーザー側の不手際で多大なリソースを消費ことも危惧される。NDB の維持コストとの兼ね合いも検討する必要である。他にも例えば、NDB オンサイトリサーチセンターでデータの抽出処理を行い、その個票レベル生成物を HIC 環境で利用する方策も有用な方法であると考えられる。すでに我々はプロジェクトごとの研究用標準データセットを抽出するプロ

グラムを NDB オンサイトリサーチセンターで運用していることは上述の通りである。これらのプログラムを有償で維持していくような形をとることも今後検討してよいのではないかと考える。
また、今年度は検証ができなかったものの、Postgresql から S3 に保存した csv ファイルを直接参照する s3csv_fdw のような拡張機能が存在しており、適切に利用することで、保存領域の課題を解消できる可能性がある。

Ｅ．結論
HIC は今後 NDB の利用形態として重要な位置を占めると考えられる。しかし、NDB オンサイトリサーチセンターと比較すると、HIC 環境で臨床疫学研究を行うためには、ユーザー側の十分な技術的工夫が必要であることが明らかとなった。

Ｆ．健康危険情報
なし

Ｇ．研究発表
なし

Ｈ．知的財産権の出願・登録状況
なし

図 1：オンサイトセンターでの症例抽出（第 6 回 NDB ユーザー会資料より）

オンサイトセンターでの研究の実際

症例抽出計画の策定

- Inclusion Period の決定
  - 許諾された全期間で症例を検索する。
  - 容量削減や抽出プロセス短縮のために、期間を絞ることはある。
- キーレセプト条件は単一レセプト内で完結する条件にする。
  - 他のレセプトを参照して決定する条件は好ましくない。
  - 例えば、”脳梗塞病名のついた症例のレセプト”や”rfPA実施症例のレセプト”は検索できる。
  - ”脳梗塞発症後”、”リハビリテーション”を行ったレセプトの検索は危険
- 研究に利用する情報を決める。

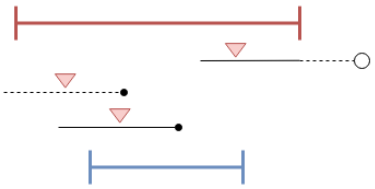

図 2：オンサイトセンターでの症例追跡（第 6 回 NDB ユーザー会資料より）

オンサイトセンターでの症例の実際

症例の追跡

- 症例追跡期間の決定
  - 許諾された全期間でキーレセプト前後情報を追跡する。
- 症例の追跡
  - キーレセプトを基にID1,ID2を再帰的に追跡する。
  - ‘ID1 = OR ID2 = ‘の 条件一回で追跡を打ち切るのはダメ
- 追跡したID1,ID2を基にID0を作成する。
  - 別人を誤って追跡した場合は後でその症例を除くか諦める。
  - ランダム抽出する場合は、ID0単位で抽出が望ましい。

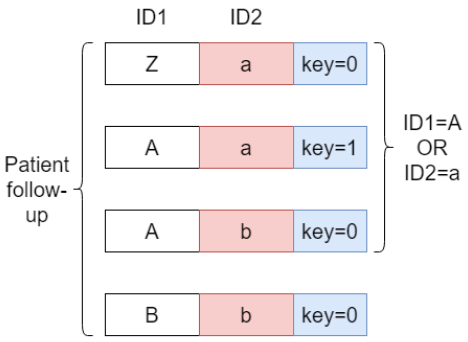

図 3：オンサイトセンターでの抽出プログラム（第 6 回 NDB ユーザー会資料より）

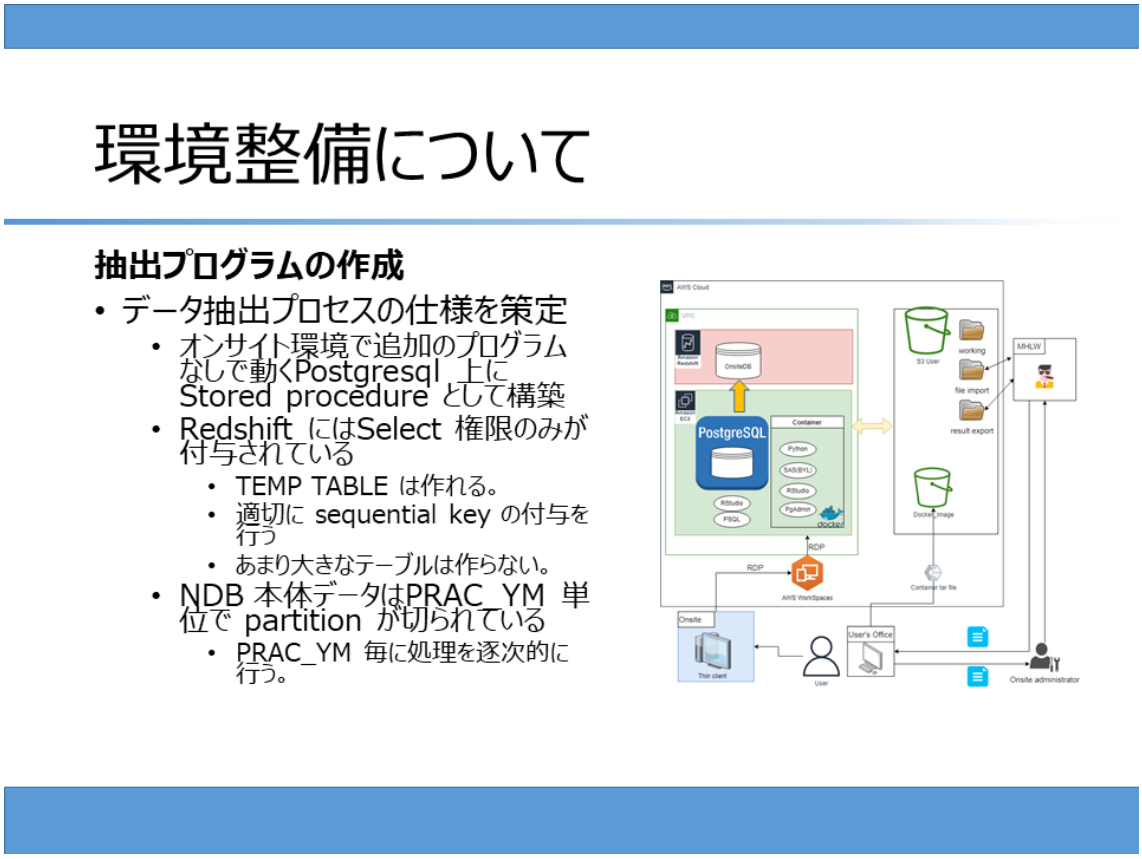

Table 1: 列圧縮と Partitioning test

	テーブルサイズ	テストクエリ 1*	テストクエリ 2**
test_nonp_heap	690Mb	865.776ms	119.406 ms
test_p_heap	691Mb	0.025ms	43,816 ms
test_nonp_col	23Mb	939.189ms	1043.778 ms
test_p_col	24Mb	0.811ms	34.084 ms

** テストクエリ 1 は Select * FROM table where prac_ym = ‘202204’ Limit 10;

** テストクエリ 2 は *Select * FROM table where prac_ym = ‘202204’ and prac_act_cd = ‘190101770’ Limit 10;*

Reference

1. Kimura Y, Suzukawa M, Jo T, Hashimoto Y, Kumazawa R, Ishimaru M, et al. Epidemiology of severe childhood asthma in Japan: A nationwide descriptive study. Allergy. 2024 Jan 7;all.16008.
2. Citus Documentation - Citus 12.1 documentation [Internet]. [cited 2024 May 8]. Available from: https://docs.citusdata.com/en/v12.1/?_gl=1*2oenlh*_ga*MTYyODQ0MTI5NC4xNzE1MDk2MzE2*_ga_DS5S1RKEB7*MTcxNTA5NjMxNi4xLjAuMTcxNTA5NjMxOS4wLjAuMA..

付録: **Postgresql** の操作

Partitioning Table の作成と columnar 圧縮の適用

まずは、以下のように partitioning したテーブルを作成する。

```
CREATE TABLE test_p_col (
column1 data_type,
column2 data_type,
...
prac_ym character(6)
) partitioned by prac_ym;
```

次に、partitioning された子テーブルを各月分用意しデータ圧縮設定(using columnar)を適用する。

```
CREATE Table test_p_col_202204 PARTITION OF test_p_col FOR VALUES IN ('202204') using columnar;
```

test_p_col に対してデータを挿入すれば、partition で分けられた圧縮テーブルが完成する。