

資料2 NCDA システムの外部化に関する検証結果報告書

1. 目的

国立病院機構様(以降、NHOとする)では、現在、国立病院機構診療情報集積基盤(NHO Clinical Data Archives:以降、NCDAとする)および診療情報データバンク(Medical Information Analysis databank : 以降、MIAとする)を運営している。NCDA は厚労省より平成 26 年度地域診療情報連携推進補助金「電子カルテデータ標準化等のための IT 基盤構築事業」の成果を引き継ぐ形で平成 28 年(2016 年)1 月から医療機関からのデータ収集および MIA との連携運用を開始している。運用開始時点では 41 医療機関の参加でスタートし、令和 2 年 3 月時点では 66 病院まで拡大しており、今後も基盤活用が拡大していく見込みである。

このたび、NCDA を国立病院機構以外でも使うシステムとして拡張していくことを検討し、現状のオンプレ環境からパブリッククラウド環境への移行を 1 案として検討することとなった。パブリッククラウド環境にはアマゾンウェブサービス(以降、AWSとする)を Direct Connect により専用ネットワーク接続することが検討されている。

本検証では、オンプレミス環境で稼働している NCDA を AWS サービスに移行する際の懸念事項である、「アプリケーション動作確認」「簡易性能評価」を実施したので報告する。

2. AWS 検証作業概要

2.1 NCDA の概要

NCDA は、センタ側システム、協力病院側システムを連携ネットワークで接続する構成となっている。
NCDA のシステム全体概要図を図 2.1 に示す。

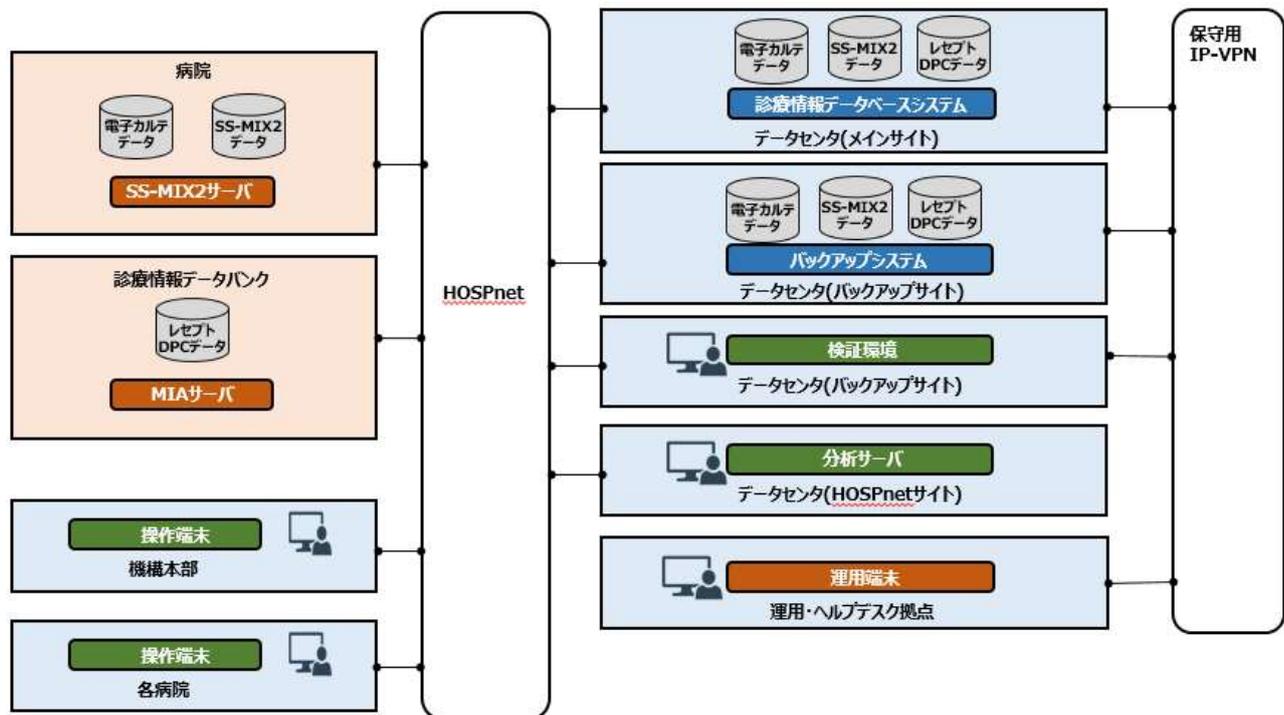


図 2.1 NCDA システム全体概要図

2.2 NCDA の全体運用フロー

NCDA は、協力医療機関から統一標準コードを用いた臨床情報(SS-MIX2 トランザクションストレージ)を日次で取得する。取得した臨床情報について、SS-MIX2 規約(Ver1.2d)との整合性をセンタ側でチェックし、チェックを通過したデータがデータベースに登録される。このようにして構築されたデータベースを用い、医療の質向上と、病院経営状況の把握と改善につなげている。NCDA の全体フローを図 2.2 に示す(拡大版は別添 1 に掲載)。

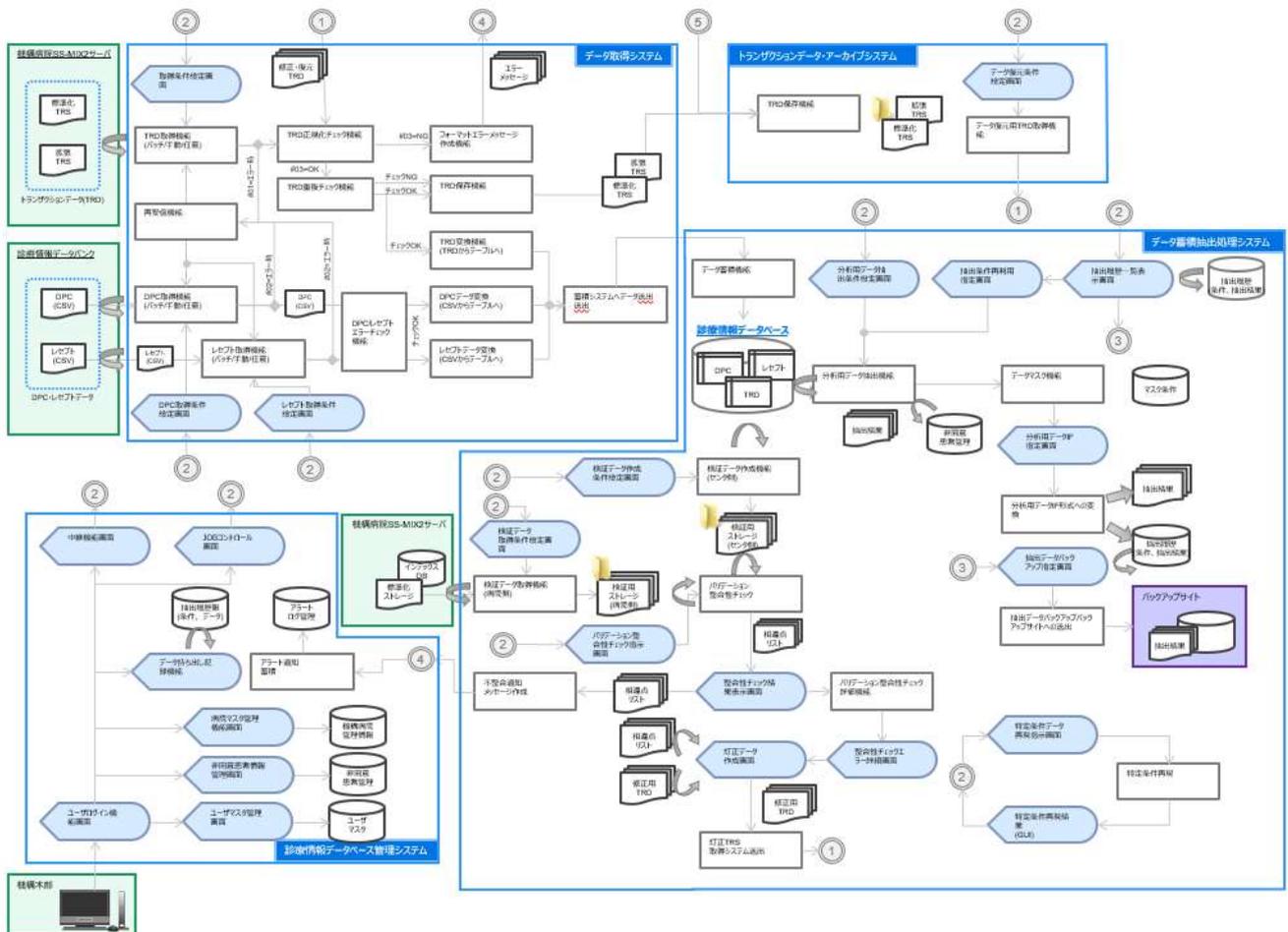


図 2.2 NCDA システム全体フロー

2.3 検証観点について

NCDA は、表 2.3 に示すサブシステムから構成されている。本検証では、業務運用の主要機能であるデータ取得システム(項番 1)、トランザクションデータアーカイブシステム(項番 2)、データ蓄積抽出処理システム(項番 3)、診療情報データベース管理システム(項番 5)について検証を行った。

表 2.3 NCDA を構成するサブシステム

#	サブシステム	役割・設計内容
1	データ取得システム	<ul style="list-style-type: none"> ・機構各病院の SS-MIX2 サーバ上のトランザクションデータファイルの定期的な収集 ・データ受信時の整合性の判定 ・診療情報データバンク(MIA)に蓄積されている DPC・レセプトデータの定期的な収集
2	トランザクションデータアーカイブシステム	<ul style="list-style-type: none"> ・トランザクションデータを加工することなく蓄積、保存
3	データ蓄積抽出処理システム	<ul style="list-style-type: none"> ・特定の用途に限定することなくトランザクションデータ、DPC・レセプトデータを蓄積 ・蓄積データを分析用に様々な形式でデータ抽出し、バリデーションで蓄積データの整合性を確認できる
4	データ分析システム	<ul style="list-style-type: none"> ・抽出データを分析し、医療の質向上と、病院経営状況の把握と改善に繋げる
5	診療情報データベース管理システム	<ul style="list-style-type: none"> ・データベース監視、JOB 管理、アラートメール送信、各種ログ確認、バックアップ管理など、システム全体の管理及び運用
6	バックアップシステム	<ul style="list-style-type: none"> ・災害発生時を考慮し、重要データをバックアップサイトに保管(BCP 対応)
7	開発システム	<ul style="list-style-type: none"> ・本システムの全機能を有する開発システム
8	検証システム	<ul style="list-style-type: none"> ・本システムの全機能を有する検証システム

本検証にあたり、評価を行った該当フローを図 2.3 に示す(図中の黄色部分)。

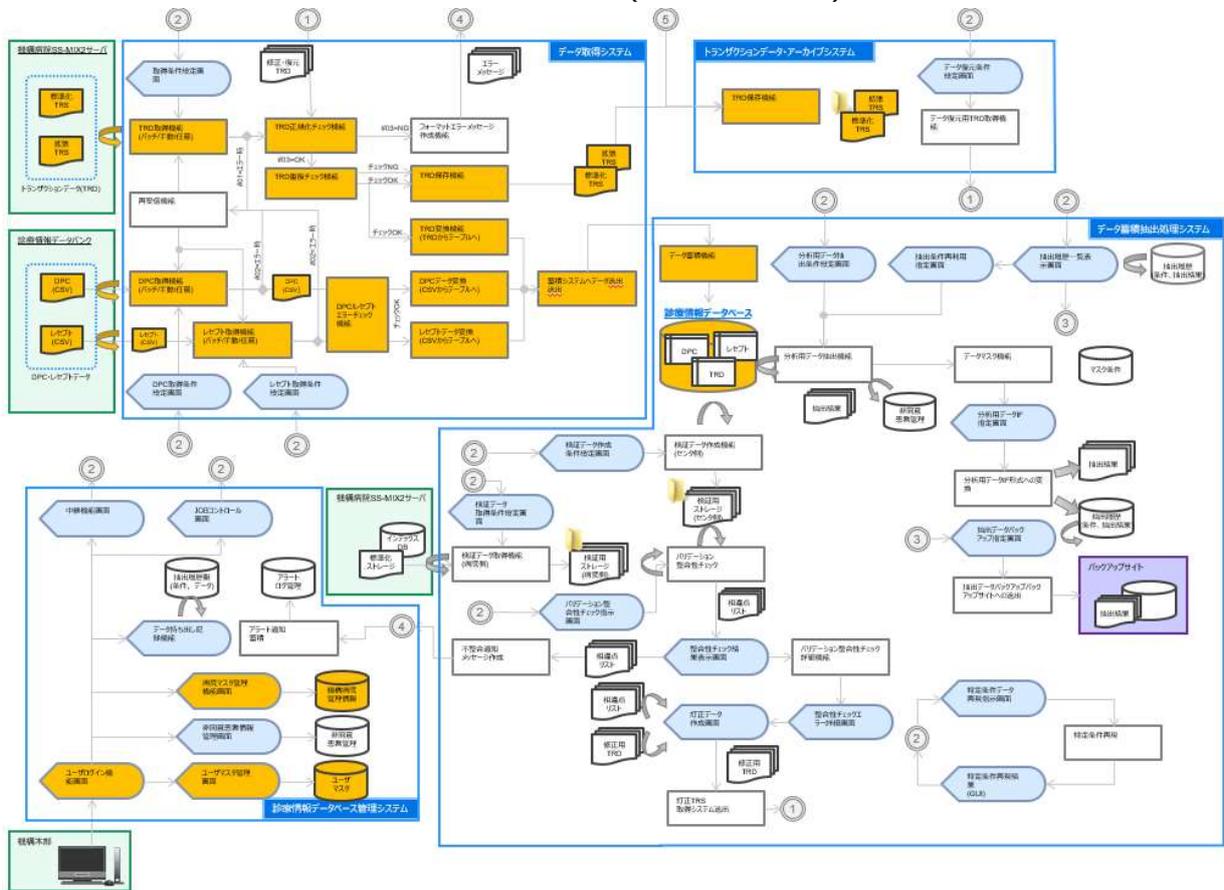


図 2.3 検証対象の処理フロー(黄色部分)

【主な検証観点】

- ・現在、オンプレミス環境で稼働する NCDA を AWS に構築するために必要な設計内容の整理と手順整備
- ・AWS に構築した NCDA の動作検証
- ・AWS 環境におけるシステム動作性能の評価(インスタンス、ストレージを複数種類に変更し差異を評価する)

2.4 スケジュール

本検証では、AWS 環境設計と環境構築、サブシステムの組み込みと動作検証、サブシステムの性能評価を行った。
2月1日に作業着手し、3月29日に作業完了した。

作業は図 2.4 のとおりスケジュールを計画し、遅延なく作業完了した。

	2/1~	2/15~	3/1~	3/15~	3/29
AWS 環境設計構築 サブシステム組み込み 起動確認					
検証・評価 報告書作成					
					

図 2.4 作業スケジュール

3. AWS 検証環境

3.1 AWS 構成

AWS 上の VPC に現状のオンプレ環境相当のインスタンスを構築した。構築した構成を図 3.1 に示す。

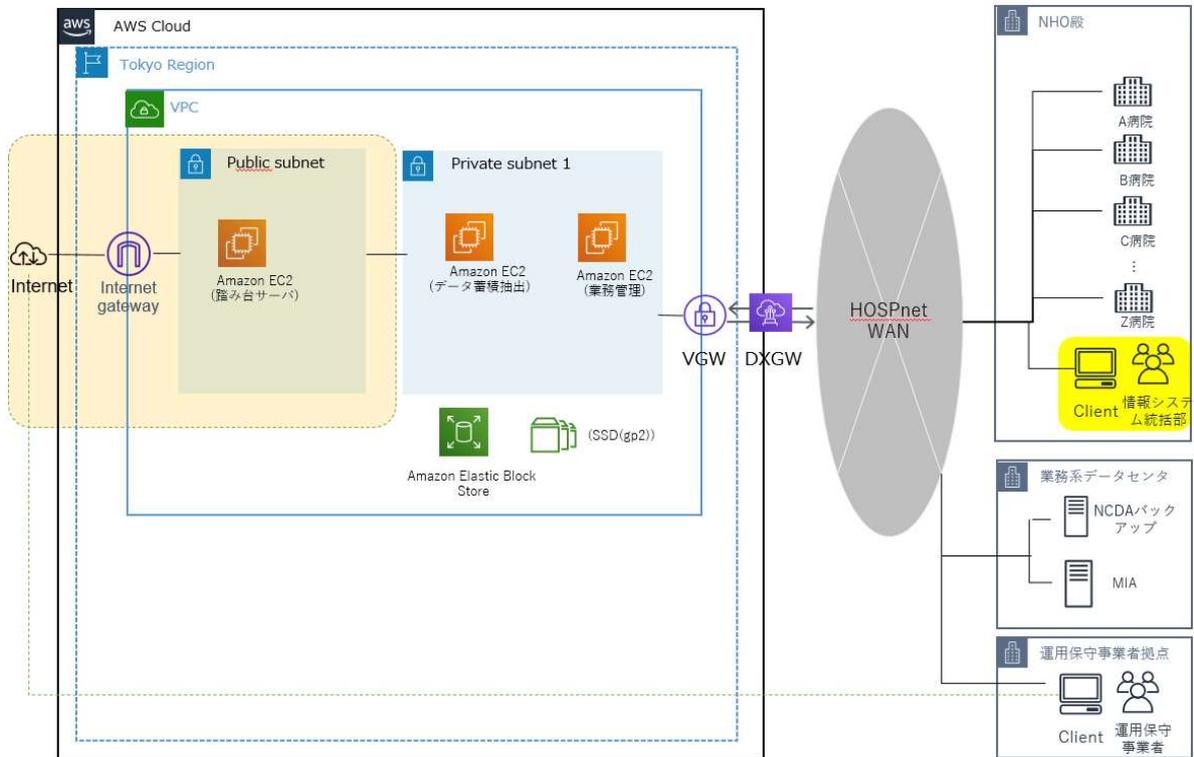


図 3.1 AWS 構成

3.2 ハードウェア構成

本作業では、AWS 検証用サーバとして、表 3.2 に示すサーバを構築した。なお、性能検証用に複数のインスタンスとストレージを使用しており、詳細は「4.2 性能検証」に示す。踏み台サーバは現行 NCDA 環境では設置されていないが、環境準備作業のため一時的に使用した。

表 3.2 ハードウェア構成

項番	対象サーバ	PoC用OS	現NCDA リソース			PoC用EC2スペック			PoC用EBS				
			CPU (ソケット×プロセッサ)	メモリ (GB)	ディスク 必要総量 (GB)	インスタンス タイプ	CPU (コア×スレッド)	メモリ (GB)	合計 (GB)	C 汎用 SSD (GB)	D 汎用 SSD (GB)	E スループット 最適化 HDD (GB)	F コールド HDD (GB)
1	データ蓄積サーバ	WindowsServer2019	14	232	-	r5.xlarge	4	32	1400	100	300	500	500
2	業務管理サーバ	WindowsServer2019	4	64	-	r5.xlarge	4	32	150	100	50		
3	踏み台サーバ	AmazonLinux2	-	-	-	t3.small	2	2	0	-	-	-	-

3.3 ソフトウェア構成

現状の NCDA 稼働環境相当のソフトウェア構成とした。表 3.3 にソフトウェア構成を示す。

表 3.3 ソフトウェア構成

項番	種別	ソフトウェア名	バージョン等	備考
1	Operation System	WindowsServer2019	Windows_Server-2019-Japanese-Full-Base-2020.12.09	
2	ウイルス対策	Microsoft Defender		
3	Web-AP	Java	amazon-corretto-8.282.08.1-windows-x64	
4		Tomcat	apache-tomcat-8.5.61	
5		Apache	httpd-2.4.46-win64-VS16	
6		Microsoft Visual C++ 2015-2019 Redistributable (x64)		
7	DB	IRIS (評価版)	IRISHealth_Community-2020.1.0.217.1-win_x64	

3.4 ネットワーク構成

本作業では、主にルーティング設計を実施した。結果を図 3.4.1 及び図 3.4.2 に示す。

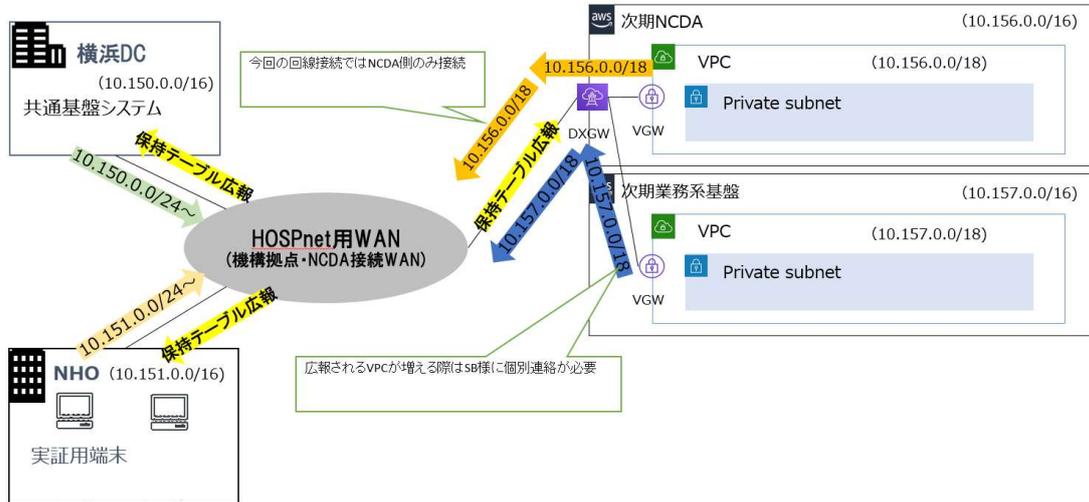


図 3.4.1 ネットワーク構成(接続時のルート広報の確認)

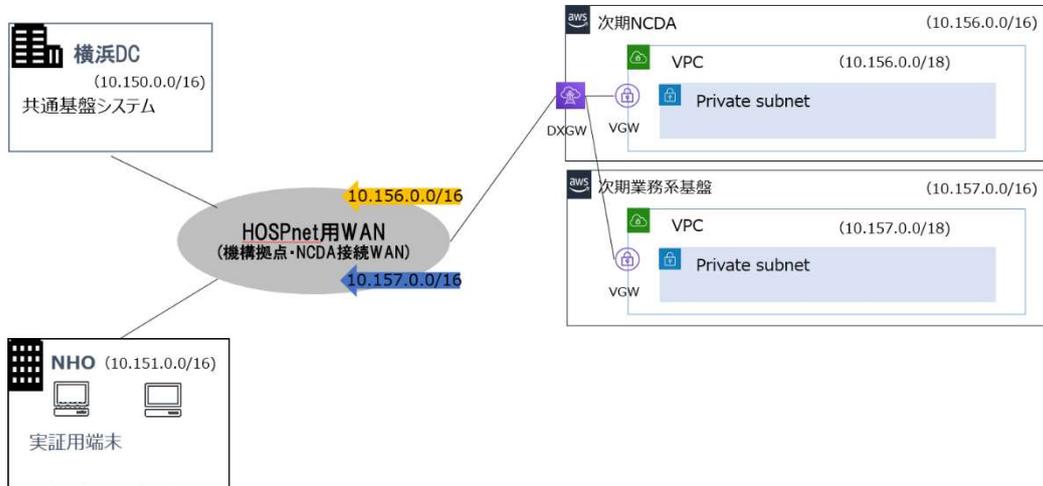


図 3.4.2 ネットワーク構成(接続時のアクセスリスト更新の確認)

4. 検証と評価

検証および評価について以下のとおり報告する。

4.1 アプリケーション動作確認

NCDA のアプリケーションを AWS 環境へ移行し、アプリケーション機能の動作確認を実施した。結果を以降に示す。

【動作確認の評価結果】

表 4.1 に示す機能について、問題なく動作することを確認した。

表 4.1 アプリケーション評価項目及び評価結果

#	サブシステム名	#	機能名称	評価内容	結果	備考
1	データ取得システム	1-1	トランザクションデータ取得機能	医療機関用トランザクションストレージよりトランザクションファイルが取得できること	○	【対象データ】 ・検体検査 ・バイタル ・退院サマリ(CDA) ・災害時記録(JSPEED)
		1-2	DPC・レセプトデータ取得機能	MIA サーバより DPC・レセプトデータが取得できること	○	【対象データ】 ・レセプトデータ ・DPC 入院データ ・DPC 外来データ
		1-3	データチェック機能	取得したデータのチェックが行えること	○	
2	トランザクションデータアーカイブシステム	2-1	トランザクションデータ保存機能	トランザクションデータが保存できること	○	
3	データ蓄積抽出処理システム	3-1	データ蓄積機能	正規化されたデータがデータベースへ格納できること	○	データが書き込めたことを、検索クエリー実行により確認
4	データ分析システム	検証対象外(本番環境相当の規模でデータベースを準備するのが困難なため)				
5	診療情報データベース管理システム	5-1	ログイン機能	画面表示され、ログインできること	○	
		5-2	ユーザマスタ管理機能	画面表示され、登録データが表示できること	○	
		5-3	病院マスタ管理機能	画面表示され、登録データが表示できること	○	
6	バックアップシステム	検証対象外				
7	開発システム					
8	検証システム					

4.2 性能評価

NCDA の運用保守において稼働分析を行うため、協力病院からのデータ取得とデータベース格納に要する時間を取得している。これと比較検討が行いやすいため、性能評価の対象機能は、上述の機能を有しているデータ取得システムおよびデータ蓄積抽出処理システムとした。データ取得システムではデータチェック機能、データ蓄積抽出処理システムではデータ蓄積機能を担っている。

4.2.1 評価項目

インスタンスおよびストレージによる性能差異を確認することを目的とし、本作業では以下を評価した。

表 4.2.1 性能評価項目と評価方法

#	評価項目	評価方法
1	E C 2 インスタンスタイプ	SS-MIX データ、C D A データの取込チェックを実施し処理時間測定
2	E B S ストレージ	SS-MIX データ、C D A データのファイル変換・格納を実施し処理時間測定

4.2.2 評価対象インスタンスタイプ

評価対象とするインスタンスタイプについて、以下に示す。

表 4.2.2 評価対象インスタンスタイプ

インスタンス	c5.xlarge コンピューティング 最適化	m5.xlarge 汎用	r5.xlarge メモリ 最適化	r5.2xlarge メモリ 最適化
vCPU	4	4	4	8
メモリ	8 GiB	16 GiB	32 GiB	64 GiB
ネットワーク パフォーマンス	最大 10Gbps			
EBS 帯域幅	最大 4,750Mbps			
AWS が提示する ユースケース	<ul style="list-style-type: none"> ・小規模/中規模のデータベース ・メモリ追加が必要なデータ処理タスク ・キャッシングフリート、 ・SAP ・Microsoft SharePoint ・クラスターコンピューティング ・その他エンタープライズアプリケーションのためのバックエンドサーバーの実行 	<ul style="list-style-type: none"> ・高性能のウェブサーバ ・科学的モデリング ・バッチ処理 ・分散分析 ・高性能コンピューティング(HPC)、 ・機械学習/深層学習推論 ・広告サービス ・拡張性の高いマルチプレイヤーゲーム ・動画エンコーディング 	<ul style="list-style-type: none"> ・高パフォーマンスデータベース ・分散型ウェブスケールインメモリキャッシュ ・中規模のインメモリデータベース、 ・リアルタイムのビッグデータ分析、 ・その他エンタープライズアプリケーション 	※メモリ集約型アプリケーションに最適
コスト(USD) ※1 か月あたり 想定値	302.54	315.36	368.24	724.48

4.2.3 評価対象ストレージタイプ

評価対象とするストレージタイプについて、以下に示す。

表 4.2.3.1 評価対象 EBS

ストレージタイプ	sc1	st1	gp2
	コールド HDD	スループット最適化 HDD	EBS 汎用 SSD
簡単な説明	アクセス頻度の低いワークロード向けに設計された極めて低コストの HDD ボリューム	高いスループットを必要とするアクセス頻度の高いワークロード向けの低コストの HDD ボリューム	幅広いトランザクションワークロードに対応できる価格とパフォーマンスのバランスが取れた汎用 SSD ボリューム
AWS が提示するユースケース	・1 日のスキャン必要回数が少ない コールドデータ	・ビッグデータ ・データウェアハウス ・ログ処理	・ブートボリューム ・インタラクティブで低レイテンシーのアプリケーション ・開発およびテスト環境
ボリュームサイズ	125 GB～16 TB	125 GB～16 TB	1 GB～16 TB
最大 IOPS/VOL	250	500	16,000
最大スループット/VOL	250 MB/秒	500 MB/秒	250 MB/秒
最大スループット/インスタンス	7,500 MB/秒	7,500 MB/秒	7,500 MB/秒
コスト(1 か月あたり)	0.015 USD/GB	0.045 USD/GB	0.10 USD/GB

【参考】今回評価のための性能測定に使用するテストデータは以下の通り。

表 4.2.3.2 使用テストデータ情報

テストデータ	容量	メッセージ数	備考
SS-MIX データ	81,485KB	8000	中規模医療機関の 1 日分のデータ量想定
CDA データ	400,000KB	55000	中規模医療機関の 1 か月分のデータ量想定

4.2.4 性能測定結果

以下に性能測定結果を示す。

4.2.4.1 インスタンスタイプ別

評価対象の各インスタンスタイプ環境にて、SS-MIX データ及び CDA データの処理を実行し、処理時間を測定した。結果を以下に示す。

表 4.2.4.1 インスタンスタイプ別 測定結果

インスタンスタイプ	ストレージ				測定結果 (h:mm:ss)			
					<下段はチェック+DB 書き込みの合算値>			
	OS	プログラム管理 DB	データ DB	トランザクション一時領域	SS-MIX		CDA	
				チェック	DB 書き込み	チェック	DB 書き込み	
c5.xlarge (4コア,8GB)	gp2 (汎用 SSD)				0:01:20	0:05:45	4:00:14	0:32:26
					0:07:05(合計)		4:32:40(合計)	
m5.xlarge (4コア,16GB)					0:01:25	0:05:27	4:09:43	0:31:44
					0:06:52(合計)		4:41:27(合計)	
r5.xlarge (4コア,32GB)					0:01:27	0:05:42	4:25:39	0:49:24
					0:07:09(合計)		5:15:03(合計)	
r5.2xlarge (8コア,64GB)					0:01:23	0:05:23	4:12:12	0:52:34
					0:06:46(合計)		5:04:46(合計)	

4.2.4.2 ストレージタイプ別

評価対象の各ストレージ環境にて、SS-MIX データ及び CDA データの処理を実行し、処理時間を測定した。測定結果を以下に示す。

表 4.2.4.2 ストレージタイプ別 測定結果

インスタンスタイプ	ストレージ				測定結果 (h:mm:ss)			
					<下段はチェック+DB 書き込みの合算値>			
	OS	プログラム管理 DB	データ DB	トランザクション一時領域	SS-MIX		CDA	
				チェック	DB 書き込み	チェック	DB 書き込み	
r5.2xlarge (8コア,64GB)	汎用 SSD	汎用 SSD	sc1 (コールド HDD)	0:01:23	0:05:30	4:10:26	0:50:11	
				06:53(合計)		5:00:37(合計)		
			st1 (スループット最適化 HDD)	0:01:25	0:05:25	4:11:56	0:30:50	
				06:50(合計)		4:42:46(合計)		
r5.xlarge (4コア,32GB)			gp2 (汎用 SSD)	0:01:23	0:05:23	4:12:12	0:52:34	
				06:46(合計)		5:04:46(合計)		
			st1 (スループット最適化 HDD)	0:01:24	0:05:45	4:24:24	0:32:26	
				0:07:09(合計)		4:56:50(合計)		

5. 総合評価

当初計画した検証について全て実施完了した。以下にまとめを記載する。

5.1 アプリケーション動作確認

4.1 章に示したとおり、当初評価対象とした機能について動作確認を行い、全て動作した。

5.2 性能評価

SS-MIX、CDA 別にインスタンスタイプとストレージタイプについて処理時間を整理した。

5.2.1 SS-MIX に関する性能評価

表 5.2.1 に SS-MIX に関する処理を整理した。SS-MIX 取込み処理では、インスタンスタイプについてメモリ量よりも CPU クロックを高めることでチェック処理にかかる時間が減少するトレンドとなった。ストレージタイプは、有意なトレンドは見当たらなかった。

表 5.2.1 SS-MIX に関する処理

SS-MIX の処理時間		データ DB 領域、トランザクション一次領域のストレージ (項目上段左がチェック時間、右が DB 書き込み時間、下段が合計)					
		sc1 (コールド HDD)		st1 (スループット最適化 HDD)		gp2 (汎用 SSD)	
インスタンス	c5.xlarge (4 コア,8GB)	N/A		N/A		0:01:20	0:05:45
						07:05(合計)	
	m5.xlarge (4 コア,16GB)	N/A		N/A		0:01:25	0:05:27
						06:52(合計)	
	r5.xlarge (4 コア,32GB)	N/A		0:01:24	0:05:45	0:01:27	0:05:42
				0:07:09(合計)		07:09(合計)	
	r5.2xlarge (8 コア,64GB)	0:01:23	0:05:30	0:01:25	0:05:25	0:01:23	0:05:23
		06:53(合計)		06:50(合計)		06:46(合計)	

5.2.2 CDA に関する性能評価

表 5.2.2 に CDA に関する処理を整理した。CDA 取込み処理でも、インスタンスタイプについてメモリ量よりも CPU クロックを高めることでチェック処理にかかる時間が減少するトレンドとなった。ストレージタイプは、sc1、gp2 において r5 インスタンスと組み合わせると処理性能が低下するトレンドが見られた。

表 5.2.2 CDA に関する処理

CDA の処理時間		データ領域、トランザクション一次領域のストレージ (項目上段左がチェック時間、右が DB 書き込み時間、下段が合計)					
		sc1 (コールド HDD)		st1 (スループット最適化 HDD)		gp2 (汎用 SSD)	
インスタンス	c5.xlarge (4 コア,8GB)	N/A		N/A		4:00:14	0:32:26
						4:32:40(合計)	
	m5.xlarge (4 コア,16GB)	N/A		N/A		4:09:43	0:31:44
						4:41:27(合計)	
	r5.xlarge (4 コア,32GB)	N/A		4:24:24	0:32:26	4:25:39	0:49:24
				4:56:50(合計)		5:15:03(合計)	
	r5.2xlarge (8 コア,64GB)	4:10:26	0:50:11	4:11:56	0:30:50	4:12:12	0:52:34
		5:00:37(合計)		4:42:46(合計)		5:04:46(合計)	

6. 提言

1. 目的で示したように、NCDA は外部の病院を含めた運用について検討をした結果、AWS 環境への移行がひとつの案として検討されている。今回の検証において、アプリケーションの移行性には大きな問題がないこと、データ取込み処理において AWS のインスタスタブ、ストレージタイプの組み合わせによる性能変化トレンドを簡易的に確認できた。

しかし、NCDA は「医療の質向上と、病院経営状況の把握と改善に繋げる」ことを目的とした基盤であり、今回検証対象外とした機能のうち、性能観点で重要なのはデータベースの検索であると言える。今回の検証では、本番相当のデータ件数規模でデータベースを準備するための時間的、費用的制約があり、検証未実施となった。

実際の AWS 移行にあたっては、ここまで示してきた検証の継続としてデータベース検索機能の評価を行ったうえで、NCDA 外部化の際の新構成を決定するプロセスをとっていくことを提言したい。