The Compute-to-Data Rail: A Feasibility Dossier for Next-Generation Healthcare Interoperability

Executive Summary & Strategic Recommendation

1.1 The C2DR Paradigm: A High-Level Overview

Current healthcare interoperability is predicated on a data-movement-centric model. Standards such as Health Level Seven (HL7) and Fast Healthcare Interoperability Resources (FHIR) facilitate the copying and transmission of sensitive Protected Health Information (PHI) between providers, payers, and other stakeholders. While essential for care coordination, this paradigm creates a vast and distributed attack surface, incurs significant costs related to data egress and interface maintenance, and presents practical barriers to large-scale, multi-institutional research and artificial intelligence (AI) model training.

This dossier evaluates a fundamental architectural shift: the **Compute-to-Data Rail (C2DR)**. This model inverts the traditional paradigm. Instead of moving data to a centralized compute environment, the C2DR dispatches containerized, attested computational workloads (e.g., AI models, analytical queries) to secure data enclaves where PHI resides.[1] The computation occurs in-place, and only privacy-preserving, derivative outputs—such as encrypted model predictions, statistical aggregates, or event notifications—traverse the network.[2] The raw PHI never leaves its trusted boundary, fundamentally reducing data exposure risk and bandwidth costs.[1]

1.2 Core Findings: A Synthesis of Viability, Impact, and Risk

- **Viability:** The C2DR model is technically viable with 2024-2025 hardware and software, though it necessitates a hybrid, "two-speed" architecture. For real-time, low-latency use cases like clinical decision support, federated inference within hardware-based Trusted Execution Environments (TEEs) offers near-native performance.[2] For high-security, asynchronous analytics where latency is less critical, Privacy-Enhancing Technologies (PETs) like Fully Homomorphic Encryption (FHE) and Secure Multi-Party Computation (MPC) provide the strongest mathematical privacy guarantees, albeit with significant computational overhead.[4]
- **Impact:** The primary economic benefits stem from a drastic reduction in cloud data egress fees and the long-term costs of developing and maintaining bespoke point-to-point interfaces.[7] Clinically, the C2DR model's most significant impact is its potential to unlock multi-institutional research and advanced AI applications currently stymied by data-sharing restrictions and privacy concerns.[3] It enables collaborative science without centralizing sensitive data.[11]
- **Risk:** The primary risks are threefold. First, the performance of FHE and MPC for complex, high-throughput operations remains a significant bottleneck, limiting their application beyond niche use cases for the immediate future.[5] Second, TEEs introduce novel attack vectors, particularly sophisticated side-channel attacks, that require continuous hardware and software mitigation and a new class of security expertise.[2] Third, significant legal and regulatory ambiguity exists around the classification of derivative data outputs under the HIPAA Privacy Rule, requiring careful legal review and expert determination.[15]

1.3 TCO Delta: The Three-Year Economic Case vs. FHIR-Based Integration

A three-year Total Cost of Ownership (TCO) model for a two-hospital, one-payer pilot project reveals a compelling economic case for the C2DR. While the C2DR requires a higher initial Capital Expenditure (CAPEX) for specialized hardware such as Data Processing Units (DPUs) and AI accelerators, these costs are offset by substantial Operational Expenditure (OPEX) savings. The model projects a TCO crossover point within 36 months, driven primarily by:

- **Reduced Data Egress Fees:** Eliminating the movement of large imaging files and raw EHR extracts results in significant monthly savings on cloud bandwidth charges, which can be as high as $0.09 per GB.[7]

- **Elimination of Interface Engine Maintenance:** The C2DR model replaces a portfolio of brittle, point-to-point interfaces with a single, standardized rail, drastically reducing the high costs associated with interface development, licensing, and specialized engineering talent.[8]

1.4 Go/No-Go Criteria & Final Recommendation

A "Go" decision for broader C2DR adoption should be contingent upon the successful validation of the following criteria in a controlled pilot:

1. **Performance:** Achieve sub-second latency for a real-time federated inference use case (e.g., sepsis prediction) running within a TEE.
2. **Economic:** Validate TCO model assumptions, demonstrating a clear path to positive ROI within a 36-month horizon.
3. **Clinical & Operational:** Demonstrate a measurable reduction in "chart hunting" time for a specific clinical workflow or successfully enable a multi-site research query previously deemed infeasible.
4. **Regulatory:** Secure a formal legal opinion and an Expert Determination method report confirming the pilot's primary data outputs comply with the HIPAA Minimum Necessary standard.

Recommendation: A conditional **"Go"** is recommended, proceeding with a phased adoption strategy. The immediate next step should be the execution of the sub-$500K "quick-win" pilot detailed in Section 6 of this report. This pilot is designed to de-risk the core technical and regulatory assumptions on a manageable scale before committing to an enterprise-wide rollout.

The Architectural Blueprint: Technical & Semantic Viability

2.1 Component Analysis: The C2DR Technology Stack

The C2DR is not a single product but a cohesive architecture composed of distinct, interoperating layers. Its successful implementation depends on the maturity and integration of technologies across secure compute, data transport, and semantic modeling.

2.1.1 The Secure Compute Layer: TEEs, DPUs, and PETs

The core innovation of the C2DR is its ability to perform computation on data that remains protected within a secure boundary. This is enabled by a combination of hardware-enforced isolation and advanced cryptography.

- **Trusted Execution Environments (TEEs):** TEEs are secure areas within a main processor, providing hardware-based isolation that protects code and data confidentiality and integrity, even from a compromised host operating system or hypervisor.[18] They are the foundation for low-latency C2DR workloads.
  - **Evolution:** First-generation TEEs, such as Intel SGX, faced significant adoption barriers, including limited protected memory, incompatibility with GPUs, and substantial code porting challenges.[2] The new generation of VM-based TEEs, like Intel TDX, addresses these issues by isolating entire virtual machines, drastically reducing the effort to run existing applications securely.[2]
  - **Role in C2DR:** TEEs are ideal for the "fast lane" of the C2DR, executing real-time federated inference for clinical decision support models where latency is critical. The ability to attest to the code running inside the enclave provides a cryptographic guarantee to all parties that the correct, agreed-upon model is being executed.[2]
- **Data Processing Units (DPUs):** DPUs are specialized processors that offload and accelerate networking, storage, and security functions from the host CPU.[19] Products like the NVIDIA BlueField-3 and AMD Pensando are effectively a "data center infrastructure-on-a-chip," creating a programmable, hardware-isolated plane for managing data flow.
  - **The Unsung Hero:** The role of the DPU in a C2DR architecture is critical and often underestimated. A major threat to TEEs is side-channel attacks that exploit shared resources managed by the host OS.[14] By terminating encrypted network traffic, enforcing security policies (e.g., stateful firewalls), and managing storage virtualization directly on the DPU, the host CPU and its TEEs are shielded from network-based attacks. This offloading preserves the

TEE's limited resources for the primary computation and dramatically shrinks its attack surface.[20] A robust C2DR design should treat DPUs as a mandatory component of its secure nodes.

- **Privacy-Enhancing Technologies (PETs):** For use cases demanding the absolute highest level of privacy, where even the enclave administrator is untrusted, PETs provide purely mathematical guarantees.
  - **Fully Homomorphic Encryption (FHE):** FHE allows for arbitrary computations to be performed directly on encrypted data, yielding an encrypted result that, when decrypted, is identical to the result of performing the same operations on the plaintext.[4] This is the theoretical gold standard for privacy. However, its practical application is severely limited by computational intensity. Complex operations can be orders of magnitude slower than plaintext equivalents, making FHE unsuitable for real-time applications with current software-based implementations.[5] Its primary role in a C2DR is for the "secure lane"—asynchronous, high-value analytics (e.g., genomic computations, multi-institutional model training) where latency is not a primary concern.[22]
  - **Secure Multi-Party Computation (MPC/SMPC):** MPC enables a group of parties to jointly compute a function over their private inputs without revealing those inputs to one another.[6] This is well-suited for collaborative analytics, such as when multiple hospitals wish to benchmark performance metrics without sharing their raw patient data. The market for MPC is growing, projected to reach between $2.3 billion and $2.7 billion by 2034, with healthcare being a key vertical.[6] Recent pilot studies, such as a European collaboration on cancer research, have successfully used MPC to analyze patient data across national borders, demonstrating its real-world feasibility.[12] Like FHE, MPC introduces computational and communication overhead, making it best suited for batch analytics rather than real-time queries.[13]

2.1.2 The Transport Layer: Auditable Event-Stream Protocols

The C2DR's asynchronous nature, where compute jobs are dispatched and results are returned later, aligns perfectly with an event-driven architecture.[1] This approach decouples data producers (the enclaves) from consumers (the requesting applications), enhancing scalability and resilience.[26]

- **Core Capabilities:** Event streaming platforms such as Apache Kafka, Redpanda, and NATS JetStream provide the necessary features for a healthcare-grade transport layer. These include:

  - **Persistence:** Messages are stored durably in an ordered log.
  - **Replayability:** Consumers can re-read messages from any point in the stream, which is critical for auditing, debugging, and recovering from failures.
  - **Guaranteed Delivery:** Protocols ensure messages are delivered at-least-once or exactly-once, preventing data loss.[27]
- **HIPAA Compliance:** For healthcare applications, the transport layer must meet stringent security and audit requirements. Managed cloud services like StreamNative (built on Apache Pulsar) and Google Cloud Pub/Sub explicitly offer HIPAA-compliant configurations.[26] Key features include strong Identity and Access Management (IAM), end-to-end encryption, and comprehensive audit logs that track every event on the stream. The immutable log of all compute requests and derivative results serves as a powerful, verifiable audit trail for HIPAA compliance purposes.[26]

2.2 Performance Benchmarks: Latency, Throughput, and Scalability

The viability of a C2DR hinges on whether its secure components can deliver performance adequate for clinical and research workflows. The evidence suggests a clear performance hierarchy, dictating a two-speed architectural approach.

- **Federated Inference in TEEs:** Benchmarks comparing VM-based TEEs (Intel TDX) with application-level TEEs (Intel SGX) for confidential federated learning (CFL) show that the newer VM-based approach has a minimal impact on runtime.[2] The overhead compared to non-enclaved execution can be as low as 2x, a penalty that is often acceptable for the security gained.[2] ⚠️ These benchmarks typically use image classification tasks (e.g., on MNIST, CIFAR-10). Performance on more complex, I/O-intensive tasks involving structured EHR data and multi-modal inputs needs further validation.
- **FHE/MPC Performance:** FHE performance remains a major challenge. A 2019 proof-of-concept study in oncology noted that while simple additions on encrypted data were feasible, more complex computations posed a critical bottleneck.[23] A much earlier 2011 implementation reported homomorphic multiplication times of 41 ms per operation.[30] While algorithms and libraries have improved, achieving the high throughput needed for real-time AI (e.g., ≥106 operations/second) on complex models is not yet practical without

specialized hardware like FHE-specific ASICs, which are still in early stages of development and not widely available.[5] ❓ There is a scarcity of peer-reviewed benchmarks for FHE/MPC on 2024-2025 hardware running common healthcare machine learning models (e.g., logistic regression, gradient boosting, transformers). Vendor-provided data must be independently verified.

- **DPU/AI Accelerator Throughput:** The underlying hardware for networking and compute is more than capable.
  - **DPUs:** NVIDIA BlueField-3 and AMD Pensando DPUs can process network traffic and security functions at line rates of 200-400 Gb/s, ensuring that the infrastructure layer is not a performance bottleneck.[19]
  - **AI Accelerators:** The latest generation of AI accelerators provides massive on-premise compute power for the enclaves. The Intel Gaudi 3, for instance, delivers 1,835 TFLOPs of FP8 compute and is equipped with 128 GB of high-bandwidth memory, making it capable of running large foundation models.[31] Benchmarks show Gaudi 3 offers highly competitive performance against NVIDIA's H100/H200, particularly on a tokens-per-dollar basis, making powerful in-enclave inference economically viable.[32]

This performance data strongly supports a two-speed C2DR. Real-time tasks are routed to the TEE "fast lane," while latency-tolerant, high-security tasks are routed to the FHE/MPC "secure lane." This creates a new governance challenge: defining the policies that determine which compute jobs use which lane, balancing the needs for speed, security, and cost.

2.3 Semantic Integrity: Mapping High-Value Clinical Features

For a C2DR to be useful, it must be able to operate on semantically meaningful clinical data. This requires identifying the most valuable data elements for common AI use cases and ensuring they can be represented without loss of critical information.

- **Identifying High-Value Features:** A large body of research exists on the development of clinical prediction models, providing a strong evidence base for which features are most important.
  - **Prevalence:** Systematic reviews consistently identify a core set of predictors across numerous diseases. For cardiovascular disease, models heavily rely on demographics (age, sex), vital signs (blood pressure), lab results (cholesterol), comorbidities, and ECG data.[34] Models for COVID-19 deterioration, COPD, and

general mortality show similar patterns.[35] A 2022 systematic review found that combining unstructured text data with structured data consistently improved prognostic model performance, highlighting the value of clinical notes.[38]
  - **The 80/20 Rule:** Based on this extensive literature, a Pareto principle can be applied. A curated set of approximately **100-150 high-value features**—spanning demographics, vital signs, key laboratory results (e.g., complete blood count, basic metabolic panel, liver function tests), major coded diagnoses and procedures (ICD, CPT), and current medications—can likely account for over 80% of the predictive power for the most common acute-care AI use cases, such as sepsis prediction, readmission risk, and patient deterioration indices.[39]
- **Schema Mapping: openEHR vs. OMOP CDM:** Two data models are particularly relevant for a C2DR architecture.
  - **openEHR:** This standard uses a two-level modeling approach (a stable reference model plus flexible clinical "Archetypes" and "Templates") to create semantically rich, precise, and future-proof clinical data repositories. It is designed for primary data capture at the point of care and is ideal for creating a vendor-neutral, longitudinal health record with strong medico-legal integrity.[11]
  - **OMOP Common Data Model (CDM):** Maintained by the OHDSI collaborative, OMOP CDM is designed specifically for secondary use and large-scale observational research. It standardizes heterogeneous source data into a consistent structure and harmonizes concepts using standard vocabularies (e.g., SNOMED CT, RxNorm, LOINC). This allows researchers to write a single query that can be executed against data from hundreds of different institutions.[11]
  - **Architectural Implication:** The transformation from the highly expressive, detailed openEHR format to the more constrained OMOP CDM format can result in a loss of semantic context.[41] However, this "loss" can be framed as a compliance feature. The HIPAA Minimum Necessary principle requires that PHI disclosure be limited to what is essential for a specific task.[42] The process of mapping to OMOP for a given AI model forces an explicit, auditable selection of only the required data elements. The "lost" semantics are precisely the data that should not be exposed for that query. Therefore, the optimal C2DR architecture would store data natively in a rich format like openEHR within the secure enclave. When a compute job is requested, it would specify the OMOP concepts it requires. The system would then generate a transient, fit-for-purpose OMOP view inside the enclave, use it for the computation, and

then destroy it, leaving a clear audit trail of exactly which concepts were used.

Table 2.1: C2DR Technology Stack: Performance & Maturity Matrix

Technology	Key Vendors / Projects	Performance Overhead (Latency/Throughput)	Security Guarantees	Maturity Level (TRL)	Ideal Use Case
TEE (VM-based)	Intel (TDX), AMD (SEV)	Low (e.g., <2x vs. native) [2]	Hardware isolation of memory/CPU state from host OS/hypervisor. Protects against software-based attacks.	7-8	Real-time federated inference, low-latency CDS.
TEE (App-based)	Intel (SGX)	Moderate to High (due to code porting & limited memory) [2]	Finer-grained isolation of specific applications.	8-9	Securing specific legacy applications or functions.
DPU / SmartNIC	NVIDIA (BlueField-3), AMD (Pensando), Intel	Line-rate (200-400 Gb/s) acceleration of infrastructure services [19]	Offloads security/networking from host, reducing TEE attack surface.	8-9	Mandatory infrastructure for scalable, secure C2DR nodes.
FHE	IBM (HElib), Zama, OpenFHE [6]	Very High (orders of magnitude slower for complex ops) [5]	Mathematical proof of privacy; computation on ciphertext.	5-6	Asynchronous, high-value analytics (e.g., genomics), privacy-pres

			Protects against compromised host & admin.		erving model training.
MPC / SMPC	Duality, Enveil, Cybernetica (Sharemind)[6]	High (requires multiple communication rounds)[13]	Cryptographic protocol for joint computation without revealing inputs.	6-7	Collaborative analytics, multi-party benchmarking, federated model training.

Fortifying the Rail: Security, Privacy, and Compliance

3.1 C2DR Threat Model: New Attack Vectors and Mitigations

The C2DR architecture fundamentally alters the security posture of healthcare data exchange. It shifts the primary point of risk from data-in-transit, where PHI is exposed on the network, to the secure compute enclave itself and the orchestration plane that manages it. This introduces a new class of threats that require specialized mitigation strategies.

- **Key Attack Vectors:**
  - **Side-Channel Attacks:** This is the most cited vulnerability for TEEs.[2] A malicious process running on the same physical CPU as a secure enclave can attempt to infer sensitive information being processed inside the enclave by observing physical side effects like cache access patterns, power consumption fluctuations, or execution timing. While TEEs encrypt data in memory, these attacks exploit the underlying hardware behavior. Mitigations include constant vigilance in applying firmware and microcode patches from CPU vendors and employing advanced scheduling techniques to isolate workloads.
  - **Physical and Host-Level Attacks:** TEEs are designed to protect against a

compromised operating system or hypervisor, but they do not defend against a physically compromised host.[14] A determined attacker with physical access could employ sophisticated techniques like chip decapping and electron microscopy to attempt to extract the TEE's root keys. Furthermore, vulnerabilities in the System Management Mode (SMM) or other privileged firmware could potentially bypass TEE protections. Mitigation relies on robust, multi-layered physical data center security and supply chain integrity, which are baseline requirements for any enterprise-grade system.

    - **Malicious Compute Injection:** A sophisticated threat involves an authorized but malicious actor submitting a carefully crafted AI model or query designed to exfiltrate data. For example, a model could be trained to overfit on specific individuals, allowing their data to be reconstructed from the model's parameters or outputs.[43] This is a form of model inversion or membership inference attack. Mitigation requires a multi-pronged defense:
        1. **Code Attestation:** Cryptographically verifying the hash of every computational workload before it is allowed to run in the enclave.
        2. **Output Sanitization:** Applying techniques like differential privacy to the results before they are released, adding statistical noise to prevent the reconstruction of individual records.
        3. **Governance:** Implementing strict policies and reviews for all new models or query types introduced to the C2DR.
    - **Orchestration Layer Compromise:** The central system that schedules compute jobs, manages cryptographic keys, and routes derivative outputs is a high-value target. A compromise here could allow an attacker to deploy malicious code, steal results, or cause a denial of service. Mitigation involves standard best practices for securing critical infrastructure: defense-in-depth, least-privilege access, multi-factor authentication, and continuous monitoring.
- **Mitigation Framework:** A defense-in-depth strategy is non-negotiable. Relying on a single technology, like TEEs, is insufficient.[14] The C2DR security framework must align with established standards from NIST and ENISA [21] and should include:
    - **Hardware Foundation:** Mandating DPUs to isolate the network and storage I/O plane.[20]
    - **Enclave Security:** Using VM-based TEEs with remote attestation for all compute jobs.
    - **Federated Defenses:** Adapting principles from frameworks like FedMLSecurity to monitor for and defend against model poisoning and data reconstruction attacks at the aggregation point.[43]
    - **Zero Trust Networking:** Enforcing strict authentication and authorization for

every request, regardless of its origin within the network.

3.2 Audit & Attestation: C2DR vs. HITRUST + SOC 2

A C2DR model offers a paradigm shift in auditability, moving from verifying access to verifying computation. This has profound implications for how compliance with frameworks like HITRUST and SOC 2 is demonstrated.

- **Current State of Attestation:**
  - **HITRUST CSF:** This is the de facto standard for healthcare organizations in the U.S. It provides a comprehensive, prescriptive set of controls that harmonizes requirements from HIPAA, NIST, ISO, and others. It is a certifiable framework.[45]
  - **SOC 2:** This is an attestation report issued by a CPA firm, based on the AICPA's Trust Services Criteria (TSC). It is more flexible and less prescriptive than HITRUST, with only the Security criterion being mandatory.[45] A "SOC 2 + HITRUST" report maps HITRUST controls to the TSCs, offering a combined assurance mechanism.[47]
- **The C2DR Audit Trail Advantage:** Traditional audit logs show *who* accessed *what* data. A C2DR provides a far more granular and cryptographically verifiable audit trail that logs:
  1. **What was computed:** The cryptographic hash of the containerized model or query that was executed.
  2. **What it ran on:** The specific, versioned data elements used, identified by their semantic codes (e.g., OMOP concept IDs).
  3. What was produced: The exact derivative output that was generated and transmitted.
     This creates an immutable, end-to-end record of every single computation performed across the network, offering unprecedented transparency and accountability. The concept is analogous to a CUI (Controlled Unclassified Information) enclave, which isolates sensitive data and subjects all interactions to strict, auditable boundary controls.48
- **Mapping to Existing Controls:**
  - The C2DR architecture inherently satisfies many core HITRUST and SOC 2 controls. Domains like Access Control, Transmission Protection, and Audit Logging & Monitoring are addressed by design.[45] The enclave model itself provides a powerful implementation of the principle of least privilege and data

    minimization.
    - However, a significant gap exists. Current audit frameworks and the skill sets of most auditors are not equipped to evaluate the novel risks of a C2DR. For example, a standard audit would not be able to validate the security of a TEE against a side-channel attack or verify the mathematical soundness of an FHE implementation. ❓ This raises a critical question: Who will audit the auditors and the tools they use? New standards and training programs will be required for third-party assessors to provide meaningful assurance for C2DR systems.

3.3 Regulatory Deep-Dive: Navigating the Grey Zones

While the C2DR model offers compelling advantages for security and privacy, its implementation requires careful navigation of the existing U.S. healthcare regulatory landscape.

- **ONC Information Blocking Rule:** The 21st Century Cures Act prohibits "information blocking," defined as practices likely to interfere with the access, exchange, or use of electronic health information (EHI).[49] The rule provides several exceptions for reasonable and necessary activities. The C2DR model can be positioned as a powerful tool for compliance, effectively designing out the need for many exceptions.
    - Instead of invoking the **Infeasibility Exception** due to technical or financial barriers to creating a data-sharing interface, an organization can offer access via the C2DR as a feasible, secure alternative.[49]
    - It directly addresses the **Security Exception** by providing a state-of-the-art security model that protects EHI by default.
    - This "compliant by design" posture fundamentally strengthens an organization's position, making it difficult to argue that it is unreasonably interfering with data exchange when it offers a secure, privacy-preserving method for deriving value from that data.
- **Trusted Exchange Framework and Common Agreement (TEFCA):** TEFCA aims to create a national "network-of-networks" by establishing a common set of technical and legal rules for data exchange among Qualified Health Information Networks (QHINs).[51] A C2DR can participate in this ecosystem in a novel way. Instead of responding to a TEFCA query by transmitting a large bundle of FHIR resources, a C2DR-enabled QHIN could offer to run a specific, TEFCA-approved computation (e.g., a risk score calculation) on behalf of the requestor and return

only the result. This aligns perfectly with TEFCA's goals of expanding exchange purposes beyond basic treatment to include payment and healthcare operations, all while maintaining strong privacy and security controls.[52]

- **HIPAA "Minimum Necessary" Standard:** This is the most critical and legally ambiguous area for a C2DR. The HIPAA Privacy Rule requires covered entities to make reasonable efforts to limit the use and disclosure of PHI to the minimum necessary to accomplish the intended purpose.[16]
  - **The Argument for Compliance:** The C2DR model is the ultimate expression of the minimum necessary principle. By design, it prevents the disclosure of any raw PHI, sharing only a derivative output (e.g., a probability score, a statistical aggregate, a "yes/no" flag).
  - **The Grey Zone:** The central question is whether these derivative outputs constitute de-identified data under HIPAA. HIPAA provides two pathways for de-identification:
    1. **Safe Harbor Method:** This is a prescriptive approach requiring the removal of 18 specific identifiers (names, specific geographic locations, dates, etc.). A single derivative output, like a risk score, does not contain these identifiers, but the method is rigid.[15]
    2. **Expert Determination Method:** This is a principles-based approach where a qualified statistician provides a formal opinion that the risk of re-identifying an individual from the data is "very small".[15] This is the most likely path for C2DR compliance. An expert would need to analyze the outputs of specific AI models or queries and attest that they cannot be reasonably used, alone or in combination with other available data, to identify a patient.
  - ⚠️ **The Legal Frontier:** There is currently no definitive legal precedent or specific guidance from the Department of Health and Human Services (HHS) on the status of AI model outputs under HIPAA. An organization deploying a C2DR must engage legal counsel and contract with a qualified expert to produce a formal Expert Determination report for its core use cases. This is a critical, non-negotiable step to mitigate regulatory risk.

Economic Impact Analysis: TCO & ROI

4.1 Three-Year TCO Comparison: C2DR Pilot vs. Classic FHIR Rollout

To provide a CFO-grade assessment, a three-year Total Cost of Ownership (TCO) model was developed for a pilot project involving two acute-care hospitals and one regional payer. The model compares two scenarios:

1. **Scenario A (C2DR Pilot):** Implementation of a C2DR to support two high-value use cases: a real-time sepsis prediction model and a quarterly collaborative analysis of care quality metrics.
2. **Scenario B (Classic FHIR):** Achieving the same functionality using traditional methods, which involves establishing FHIR API endpoints, deploying an interface engine, and creating data pipelines for extracting, transforming, and loading (ETL) data into a centralized analytics environment.

The TCO formula used is: TCO=AC+Σi=13(OCi+RCi), where AC is Acquisition Cost, OC is annual Operational Cost, and RC is annual Resource (Personnel) Cost.[53]

4.1.1 Cost Breakdown: Capital vs. Operational Expenditures

- **C2DR Pilot (Scenario A) Costs:**
    - **Acquisition & Capital Costs (AC/CAPEX):** This is the primary cost driver for the C2DR. It includes one-time hardware purchases and setup fees.
        - *Hardware:* 4x secure compute nodes (2 per hospital), each equipped with a high-end AI accelerator (e.g., Intel Gaudi 3) and a DPU (e.g., NVIDIA BlueField-3). Estimated cost: $20,000 - $30,000 per node.[33] Total: **~$100,000**.
        - *Software & Licensing:* Licensing for TEE-enabled virtualization, orchestration software, and potentially PET libraries. Estimated: **~$50,000**.
        - *Implementation & Consulting:* Fees for initial setup, configuration, and security attestation (including Expert Determination report). Estimated: **~$150,000**.[55]
        - **Total Upfront Cost: ~$300,000**
    - **Operational & Resource Costs (OC/RC/OPEX):** These are recurring annual costs.
        - *Hardware Maintenance & Support:* Standard 20-25% of hardware cost

        annually. Estimated: **~$25,000/year**.[57]
        - *Specialized Personnel:* Requires 0.5 FTE of a specialized security engineer with hardware/crypto expertise. Estimated: **~$100,000/year**.
        - *Cloud/Network Costs:* Minimal, as only small derivative payloads are transmitted. Estimated: **<$5,000/year**.
        - **Total Annual Cost: ~$130,000**
- **Classic FHIR Rollout (Scenario B) Costs:**
    - **Acquisition & Capital Costs (AC/CAPEX):** Lower upfront hardware costs, but significant software and development expenses.
        - *Interface Engine:* Licensing and setup for a commercial interface engine (e.g., Mirth Connect, Rhapsody). Estimated: **~$50,000 - $75,000**.[17]
        - *Cloud Infrastructure:* Setup of a centralized cloud data warehouse/lake for analytics. Estimated: **~$25,000**.
        - *Initial Development & Integration:* Labor costs for developing FHIR APIs, ETL pipelines, and point-to-point connections. This is highly labor-intensive. Estimated: 4-6 months of work for 2-3 engineers. **~$200,000**.[8]
        - **Total Upfront Cost: ~$300,000**
    - **Operational & Resource Costs (OC/RC/OPEX):** High recurring costs are the primary drawback of this model.
        - *Interface Maintenance & Licensing:* Annual fees for the interface engine and ongoing work to update and patch interfaces as source systems change. Estimated: **~$50,000/year**.[17]
        - *Cloud Data Egress & Storage:* This is a major, variable cost. Moving terabytes of EHR and imaging data for the quarterly analysis incurs significant egress fees from the hospitals' cloud environments to the central analytics platform. A single petabyte of data can cost over $1 million per year just to store.[60] Assuming a more modest 10 TB/month of data movement at $0.09/GB, costs can reach **~$90,000/year**.[7]
        - *Personnel:* Requires 1.0 FTE of a traditional integration engineer. Estimated: **~$150,000/year**.
        - **Total Annual Cost: ~$290,000**

4.1.2 The Three-Year TCO Projection

Cost Category	C2DR Pilot (Scenario A)	Classic FHIR (Scenario B)	Delta (C2DR Savings)
Year 1 Cost (AC + Annual)	$430,000	$590,000	$160,000
Year 2 Cost (Annual)	$130,000	$290,000	$160,000
Year 3 Cost (Annual)	$130,000	$290,000	$160,000
3-Year TCO	**$690,000**	**$1,170,000**	**$480,000**

The analysis indicates that while both approaches have similar upfront costs, the C2DR model yields substantial OPEX savings, resulting in a **$480,000 lower TCO over three years** for this pilot scenario. The crossover point where the cumulative cost of the C2DR becomes lower than the FHIR approach occurs in Year 1.

4.2 Quantifying Key Benefits: Bandwidth and Maintenance

- **Bandwidth Savings:** The most direct and easily quantifiable benefit is the reduction in data egress fees. A typical sepsis model requires a few hundred features per patient per hour, amounting to kilobytes of data. A collaborative analytics query might return a few megabytes of aggregated statistics. This is in stark contrast to moving entire patient records or imaging studies (which can be gigabytes each). For the pilot scenario, avoiding the transfer of 10 TB of data per month saves approximately **$1,080,000 over three years** in egress fees alone, forming the bulk of the OPEX savings.
- **Avoided Adapter Maintenance Costs:** The traditional integration model is often called a "spaghetti architecture" for a reason. Each new system or data partner requires a new point-to-point interface, and each interface must be maintained, tested, and updated whenever one of the connected systems changes. This creates a significant, ongoing maintenance burden that grows exponentially with the number of connections.[17] A C2DR replaces this with a single, standardized "publish/subscribe" model. Data sources publish to the rail once, and any authorized compute job can consume the insights. This reduces the integration team's workload from building and fixing dozens of brittle pipes to managing a

single, robust data highway. The savings in specialized engineering talent and vendor license fees can easily exceed **$100,000-$200,000 annually** for a mid-sized health system.[8]

4.3 Sensitivity Analysis

To test the robustness of the TCO model, a sensitivity analysis was performed by varying key assumptions:

- **Varying Message Volume:** Doubling the volume of data that would have been transferred in the FHIR model (from 10 TB/month to 20 TB/month) makes the C2DR case even more compelling, increasing the 3-year TCO advantage to over **$1.5 million**.
- **Varying Hardware Costs:** A 50% increase in the initial hardware CAPEX for the C2DR (e.g., due to supply chain issues or higher-than-expected accelerator costs) delays the TCO crossover point to mid-Year 2, but the 3-year TCO remains favorable.
- **Varying Personnel Costs:** The model is sensitive to the cost and availability of specialized talent. If the cost of a C2DR security engineer is 50% higher than estimated, the annual OPEX advantage shrinks but remains positive. Conversely, the high demand for skilled FHIR integration engineers could drive their salaries higher, further favoring the C2DR's less labor-intensive model.

The analysis consistently shows that for use cases involving large or frequent data transfers across organizational boundaries, the C2DR model offers a structurally superior economic profile over the long term, even with conservative assumptions.

Clinical Workflow & Human-in-the-Loop Integration

5.1 The Persistence of the "Human-in-the-Loop"

A C2DR automates the secure delivery of AI-driven insights, but it does not and should not eliminate human oversight in critical clinical decision-making. The Human-in-the-Loop (HITL) model, where human expertise is integrated into the AI lifecycle for training, evaluation, and operation, remains essential for safety, accountability, and trust.[61]

Key areas where a human must remain in the loop for a C2DR-powered workflow include:

- **Write-Backs to the EHR:** A C2DR can generate a prediction (e.g., "Patient X has a 92% probability of sepsis"), but this insight is useless until it is acted upon. The process of writing new information, such as a diagnosis, alert, or suggested order, back into the official Electronic Health Record (EHR) must be mediated by a qualified clinician. The AI can *suggest* an action, but the clinician must review, validate, and ultimately sign off on it. This creates a clear point of accountability.
- **Order Verification and Placement:** An AI model might recommend a specific antibiotic based on a sepsis prediction and local resistance patterns. However, the final act of placing that medication order must be performed by a credentialed provider. This step allows the provider to apply contextual knowledge that the AI may lack (e.g., a recently documented patient allergy not yet processed by the model, patient-specific preferences).
- **Interpretation of Ambiguous Results:** When an AI model produces a borderline or unexpected result, a human expert (e.g., a radiologist for an imaging finding, an intensivist for a deterioration score) must be the final arbiter. The HITL system should be designed to flag these low-confidence predictions for human review, rather than presenting them as definitive.[62] This is a "human-on-the-loop" approach, where the system operates autonomously but is supervised by a human who can intervene when necessary.
- **Model Training and Feedback:** The process of improving AI models is iterative. Clinicians are the source of ground truth. A HITL system should include mechanisms for clinicians to provide feedback on model performance (e.g., "This alert was helpful," or "This prediction was incorrect because..."). This feedback is then used to retrain and refine the model in a continuous learning cycle.[61]

5.2 Time-and-Motion Analysis: Impact on "Chart Hunting"

A significant portion of a nurse's time is spent on documentation and information retrieval, often referred to as "chart hunting." Time-motion studies have consistently shown that nurses spend a large percentage of their shifts on these tasks.

- **The Problem:** One study found that nurses spend approximately 19-21.5% of their time on documentation.[63] Another large-scale study involving 767 nurses found that documentation accounted for **35.3%** of their time, or about 147 minutes per shift.[64] This is time not spent on direct patient care. The introduction of EHRs has had mixed results; while some meta-analyses show a reduction in documentation time [65], others report an increase or no significant change, suggesting that system design and workflow integration are critical factors.[63] Nurses often have to navigate multiple screens and systems to synthesize a complete picture of a patient's status, contributing to this burden.[68]
- **Potential C2DR Impact:** A C2DR can reduce chart hunting by proactively delivering *synthesized information* rather than raw data.
    - **Scenario:** Instead of a nurse manually checking the latest lab results, vital signs, and notes to assess a patient's sepsis risk, a C2DR-powered application can run a federated sepsis model every 15 minutes. It can then present a single, clear output directly within the nurse's primary workflow tool: "Patient in Room 302: Sepsis Risk Score is now 8.7 (previously 4.5). Key contributors: WBC count increased, lactate elevated."
    - **Time Savings Projection:** This shifts the nurse's task from data *collection* to information *verification and action*. By eliminating the need to hunt for and manually correlate disparate data points, the C2DR could plausibly reduce the time spent on this specific type of information synthesis. While precise quantification requires a dedicated time-motion study of a C2DR implementation, a conservative estimate suggests a potential reduction of **5-10 minutes per nurse per shift** in chart review time for high-acuity patients. While this seems small, when aggregated across an entire nursing staff over a year, it represents thousands of hours redirected to direct patient care and other value-added activities.[69]
    - ⚠️ **Open Question:** The net effect on total documentation time is uncertain. While the C2DR reduces information retrieval time, the resulting alerts and insights may trigger *new* documentation and intervention tasks. A formal pre- and post-implementation time-motion study is a critical success metric for any clinical pilot.

Market Landscape & Pilot Blueprint

6.1 Market Landscape: C2DR-like Rails and Enablers

The C2DR concept is not being built in a vacuum. A nascent but rapidly evolving ecosystem of software vendors, cloud providers, and hardware manufacturers is creating the foundational components required for this architectural shift.

6.1.1 Platform and Solution Providers

- **Innovaccer:** Innovaccer is a key player moving towards a C2DR-like model with its "Healthcare Intelligence Platform," recently branded 'Innovaccer Gravity™'.[70] Their architecture is built on a "Data Activation Platform" that unifies disparate data sources (EHR, claims, etc.) into a single source of truth.[70] On top of this sits an "AI Engine" that powers copilots and agents. Critically, their model emphasizes bringing AI to the data, with a cloud-agnostic, single-tenant deployment model that keeps customer data within its own secure VPC.[71] They have also introduced the Healthcare Model Context Protocol (HMCP), an open standard designed to allow secure, compliant, and auditable interactions between AI agents and data sources, which functions as a form of orchestration layer for a C2DR.[72] ⚠️ While Innovaccer's marketing and architectural descriptions align with C2DR principles, their GitHub presence for open standards like HMCP appears limited or archived, raising questions about the maturity and community adoption of their open protocols.[74]
- **OpenMined:** OpenMined is a non-profit community focused on building open-source technology for privacy-preserving AI.[75] Their core projects, PySyft and PyGrid, provide a framework for federated learning.[76] Their architectural philosophy combines multiple PETs, including federated learning, homomorphic encryption, and differential privacy, to enable "data-centric" AI where models are brought to decentralized data sources.[76] They have conducted proof-of-concept evaluations using TEEs (NVIDIA H100 secure enclaves) to demonstrate the feasibility of secure, multi-party model evaluation where both the model and the data remain confidential.[78] OpenMined represents the open-source, community-driven approach to building a C2DR.

- **NVIDIA Clara Federated:** NVIDIA provides a comprehensive, enterprise-focused platform for federated learning, particularly in medical imaging.[9] The Clara Train SDK (now based on the open-source MONAI framework) enables multiple institutions to collaboratively train a global model without sharing patient data.[80] Their architecture is a classic server-client model where a central server orchestrates training rounds, aggregates model weight updates from clients, and distributes the improved global model.[79] This platform is a strong example of a commercially supported, domain-specific C2DR for model training.

6.1.2 Hardware and Infrastructure Enablers

The performance and security of a C2DR are directly dependent on the underlying hardware.

- **AI Accelerators / GPUs:** The market for on-premise AI inference is dominated by NVIDIA, but competitors are emerging.
  - **Intel Gaudi 3:** This accelerator shows highly competitive performance against NVIDIA's offerings, especially on a cost-per-token basis, making it an economically attractive option for in-enclave compute nodes.[32] It boasts 1.8 PFlops of FP8 compute and 128GB of HBM2e memory.[31]
  - **NVIDIA H100/H200:** The market leader, offering top-tier performance and a mature software ecosystem (CUDA, TensorRT). The H100 also features secure enclave technology for confidential computing.[78]
- **DPUs / SmartNICs:** This is a critical battleground for C2DR infrastructure.
  - **NVIDIA BlueField-3:** A powerful DPU offering up to 400Gb/s connectivity and a rich set of hardware accelerators for networking, storage, and security (IPsec, TLS, stateful firewalls).[20] It features up to 16 Arm cores to run the infrastructure software stack, completely isolated from the host CPU. A dual-port 200Gb/s model retails for approximately **$3,900**.[54]
  - **AMD Pensando:** A strong competitor with a focus on a fully P4-programmable data path, allowing for high-performance, deterministic offloading of network services.[19] They are widely adopted in hyperscale data centers and are being integrated into enterprise switches.[82]
- **FHE ASICs:** ❓ While FHE is currently too slow for many applications, several companies are working on dedicated hardware accelerators (ASICs) to speed up homomorphic operations. The arrival of commercially viable FHE ASICs would be

a major catalyst for the "secure lane" of the C2DR, but their timeline for widespread availability is uncertain.

6.2 Blueprint for a <$500K "Quick-Win" Pilot Project

This blueprint outlines a focused, 6-month pilot project designed to validate the core assumptions of the C2DR model with a budget under $500,000.

- **1. Scope and Objectives:**
  - **Participants:** One academic medical center (providing data and clinical validation) and one payer partner (providing a secondary use case and funding).
  - **Use Case 1 (Real-Time):** Deploy a pre-trained, open-source sepsis prediction model (e.g., based on MIMIC-IV data) to run in a federated manner on a live, de-identified stream of ICU data from the medical center.
  - **Use Case 2 (Asynchronous):** Perform a joint analysis between the hospital and payer to identify high-cost patients with frequent ED visits, using an MPC-based protocol to calculate aggregate statistics without sharing patient-level data.
  - **Primary Objectives:**
    1. Validate TEE-based federated inference latency and throughput against a baseline.
    2. Demonstrate the feasibility of an MPC-based collaborative analytic.
    3. Develop and secure a formal Expert Determination report for the outputs of both use cases.
    4. Validate the TCO model's OPEX savings assumptions.
- **2. Architecture and Technology Selection:**
  - **Hospital Site:** Deploy one secure compute node.
    - *Hardware:* 1x Server with an Intel Gaudi 3 accelerator and an NVIDIA BlueField-3 DPU.
    - *Software:* Proxmox/KVM with Intel TDX for VM-level TEE. Open-source federated learning framework (e.g., OpenFL, FATE).
  - **Payer Site:** Deploy one secure compute node.
    - *Hardware:* 1x Server with a DPU (GPU not required for the MPC use case).
    - *Software:* Open-source MPC framework (e.g., based on OpenFHE or MP-SPDZ).
  - **Transport Layer:** Utilize a managed, HIPAA-compliant event stream service

(e.g., Google Pub/Sub or a small StreamNative cluster) to handle communication between the nodes.

  - **Data Model:** Ingest raw data (e.g., HL7v2 ADT and ORU messages) and transform it within the enclave into a simple, OMOP-like tabular format for the models.
- **3. Phased Timeline (6 Months):**
  - **Months 1-2: Setup & Configuration.** Procure hardware. Set up the secure nodes at both sites. Configure the TEEs, DPUs, and the event streaming transport layer. Establish secure network connectivity.
  - **Month 3: Use Case 1 Implementation.** Containerize the sepsis model. Deploy the federated learning framework. Run the model on a historical, de-identified dataset to validate functionality and establish performance baselines.
  - **Month 4: Use Case 2 Implementation & Legal Review.** Implement the MPC protocol for the high-cost patient analysis. Engage legal counsel and a statistical expert to begin drafting the Expert Determination report based on the planned outputs.
  - **Month 5: Live Pilot & Data Collection.** Run the sepsis model on a live (but de-identified) data stream from the ICU. Execute the MPC analysis on the latest quarter of hospital/payer data. Collect performance metrics (latency, throughput, CPU utilization) and workflow feedback from clinicians.
  - **Month 6: Analysis & Final Report.** Analyze collected data. Finalize the Expert Determination report. Compare actual costs and performance against the TCO model and initial objectives. Produce a final report with a Go/No-Go recommendation for a broader rollout.
- **4. Budget (<$500,000):**
  - **Hardware (2 nodes):** ~$60,000
  - **Software & Cloud Services (6 months):** ~$30,000
  - **Personnel (1.5 FTE for 6 months):** ~$225,000
  - **Consulting (Legal & Expert Determination):** ~$100,000
  - **Contingency (15%):** ~$65,000
  - **Total Estimated Cost: ~$480,000**
- **5. Go/No-Go Success Metrics:**
  - **Go:** End-to-end latency for the sepsis prediction is <1 second. MPC analysis completes successfully within 24 hours. Expert Determination report is successfully obtained. Measured costs are within 10% of the budget.
  - **No-Go:** Critical security vulnerabilities are discovered that cannot be mitigated. Performance targets are missed by >50%. The legal/regulatory path

is deemed unviable.

This pilot is designed to be a pragmatic, cost-effective "quick win" that directly addresses the most significant technical, economic, and regulatory uncertainties of the C2DR model, providing the empirical evidence needed for a confident, strategic investment decision.

Works cited

1. Bringing Compute to Storage: Feasibility for Cloud AI Workflows - Medium, accessed July 18, 2025, https://medium.com/@prabhuss73/bringing-compute-to-storage-feasibility-for-cloud-ai-workflows-0afba2d2438e
2. A performance analysis of VM-based Trusted Execution Environments for Confidential Federated Learning This work has been partly supported by the Spoke ”FutureHPC & BigData” of the ICSC – Centro Nazionale di Ricerca in ”High Performance Computing, Big Data and Quantum Computing”, funded by European Union – NextGenerationEU, and by the Horizon2020 RIA EPI project - arXiv, accessed July 18, 2025, https://arxiv.org/html/2501.11558v1
3. Availability, access, analysis and dissemination of small-area data - PMC - PubMed Central, accessed July 18, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7158061/
4. Moving Beyond Traditional Data Protection: Homomorphic Encryption Could Provide What is Needed for Artificial Intelligence - Journal of AHIMA, accessed July 18, 2025, https://journal.ahima.org/page/moving-beyond-traditional-data-protection-homomorphic-encryption-could-provide-what-is-needed-for-artificial-intelligence
5. Fully Homomorphic Encryption: Data Insights Without Sharing Data - SAP News Center, accessed July 18, 2025, https://news.sap.com/2024/03/fully-homomorphic-encryption-insights-without-sharing-data/
6. Secure Multiparty Computation (SMPC) Market Analysis Report 2034, accessed July 18, 2025, https://www.prophecymarketinsights.com/market_insight/secure-multiparty-computation-smpc-market-5333
7. What are data egress fees? - Cloudflare, accessed July 18, 2025, https://www.cloudflare.com/learning/cloud/what-are-data-egress-fees/
8. Cost of EHR Integration Project - Thinkitive, accessed July 18, 2025, https://www.thinkitive.com/blog/ehr-and-emr-integration-cost/
9. Federated Learning powered by NVIDIA Clara | NVIDIA Technical Blog, accessed July 18, 2025, https://developer.nvidia.com/blog/federated-learning-clara/
10. A Benchmark for Scaling Medical Foundation Models via Federated Knowledge Injection, accessed July 18, 2025, https://arxiv.org/html/2408.09227v1
11. FHIR, OMOP, openEHR: The Right Tool for the Right Job in Healthcare IT - Whitefox.cloud, accessed July 18, 2025,

https://www.whitefox.cloud/articles/fhir-omop-openehr/

12. 'Secret Shares' of Patient Health Data Enable Secure Multiparty Research - Bio-IT World, accessed July 18, 2025, https://www.bio-itworld.com/news/2024/11/20/secret-shares-of-patient-health-data-enable-secure-multiparty-research
13. Secure Multiparty Computation Market Size, Share and Trends 2025 to 2034, accessed July 18, 2025, https://www.precedenceresearch.com/secure-multiparty-computation-market
14. Trusted Execution Environments: A Paranoid Assessment - David Dworken's Blog, accessed July 18, 2025, https://blog.daviddworken.com/posts/tee-paranoid/
15. De-identification of Protected Health Information: 2025 Update - The HIPAA Journal, accessed July 18, 2025, https://www.hipaajournal.com/de-identification-protected-health-information/
16. Methods for De-identification of PHI - HHS.gov, accessed July 18, 2025, https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html
17. A Full Breakdown of EHR Integration Price Factors - Thinkitive, accessed July 18, 2025, https://www.thinkitive.com/blog/factors-affecting-ehr-integration-costs-a-comprehensive-breakdown/
18. Basics of Trusted Execution Environments (TEEs): The Heart of Confidential Computing, accessed July 18, 2025, https://confidentialcomputing.io/2024/03/13/basics-of-trusted-execution-environments-tees-the-heart-of-confidential-computing/
19. AMD Pensando™ DPU Technology, accessed July 18, 2025, https://www.amd.com/en/products/data-processing-units/pensando.html
20. NVIDIA BLUEFIELD-3 DPU, accessed July 18, 2025, https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/documents/datasheet-nvidia-bluefield-3-dpu.pdf
21. NIST IR 8320, accessed July 18, 2025, https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8320.pdf
22. Private Predictive Analysis on Encrypted Medical Data - Microsoft, accessed July 18, 2025, https://www.microsoft.com/en-us/research/wp-content/uploads/2014/08/336-1.pdf
23. (PDF) Proof-of-concept study: Homomorphically encrypted data can support real-time learning in personalized cancer medicine - ResearchGate, accessed July 18, 2025, https://www.researchgate.net/publication/337750126_Proof-of-concept_study_Homomorphically_encrypted_data_can_support_real-time_learning_in_personalized_cancer_medicine
24. Data security: breakthrough in research with health data, accessed July 18, 2025, https://www.med.lmu.de/en/latest-news/news-overview/news/data-security-breakthrough-in-research-with-health-data.html
25. Privacy-friendly evaluation of patient data with secure multiparty computation in

a European pilot study - ResearchGate, accessed July 18, 2025, https://www.researchgate.net/publication/384900016_Privacy-friendly_evaluation_of_patient_data_with_secure_multiparty_computation_in_a_European_pilot_study

26. Event-Driven Architecture | StreamNative Ursa (Kafka/Pulsar), accessed July 18, 2025, https://streamnative.io/solutions/event-driven
27. Top 7 Kafka Alternatives for Better Data Streaming [Updated for 2025], accessed July 18, 2025, https://hevodata.com/learn/kafka-alternatives/
28. NATS and Kafka Compared | Synadia, accessed July 18, 2025, https://www.synadia.com/blog/nats-and-kafka-compared
29. Proof-of-concept study: Homomorphically encrypted data can support real-time learning in personalized cancer medicine - PubMed, accessed July 18, 2025, https://pubmed.ncbi.nlm.nih.gov/31801535/
30. Can Homomorphic Encryption be Practical? - Cryptology ePrint Archive, accessed July 18, 2025, https://eprint.iacr.org/2011/405.pdf
31. Intel Gaudi 3 vs Gaudi 2: AI Accelerator Comparison - AMAX Engineering, accessed July 18, 2025, https://www.amax.com/the-next-step-for-intel-accelerators-a-look-at-intel-gaudi-3/
32. Intel Gaudi 3: 50% Faster Inference Performance for Enterprise AI Workloads, accessed July 18, 2025, https://www.unicomengineering.com/blog/intel-gaudi-3-outperforms-prior-generation-and-competition-for-better-ai/
33. Intel Gaudi 3 Accelerates AI at Scale on IBM Cloud - Signal65, accessed July 18, 2025, https://signal65.com/research/ai/intel-gaudi-3-accelerates-ai-at-scale-on-ibm-cloud/
34. Clinical Prediction Models for Cardiovascular Disease | Circulation: Cardiovascular Quality and Outcomes - American Heart Association Journals, accessed July 18, 2025, https://www.ahajournals.org/doi/10.1161/circoutcomes.115.001693
35. Evaluation of clinical prediction models (part 1): from development to external validation, accessed July 18, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10772854/
36. Use of Artificial Intelligence in Improving Outcomes in Heart Disease: A Scientific Statement From the American Heart Association, accessed July 18, 2025, https://www.ahajournals.org/doi/10.1161/CIR.0000000000001201
37. Developing clinical prediction models: a step-by-step guide - The BMJ, accessed July 18, 2025, https://www.bmj.com/content/386/bmj-2023-078276
38. Use of unstructured text in prognostic clinical prediction models: a systematic review, accessed July 18, 2025, https://academic.oup.com/jamia/article/29/7/1292/6574714
39. Overview of clinical prediction models - PMC, accessed July 18, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7049012/
40. A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary

health care data | PLOS One - Our journal portfolio - PLOS, accessed July 18, 2025, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0274276

41. Eos and OMOCL: Towards a seamless integration of openEHR records into the OMOP Common Data Model - ResearchGate, accessed July 18, 2025, https://www.researchgate.net/publication/372290268_Eos_and_OMOCL_Towards_a_seamless_integration_of_openEHR_records_into_the_OMOP_Common_Data_Model
42. 18 HIPAA Identifiers for PHI De-Identification - Censinet, accessed July 18, 2025, https://www.censinet.com/perspectives/18-hipaa-identifiers-for-phi-de-identification
43. FedMLSecurity: A Benchmark for Attacks and Defenses in Federated Learning and Federated LLMs - arXiv, accessed July 18, 2025, https://arxiv.org/html/2306.04959v4
44. Threat Landscape | ENISA - European Union, accessed July 18, 2025, https://www.enisa.europa.eu/topics/cyber-threats/threat-landscape
45. HITRUST vs. SOC 2: What makes them different? - Vanta, accessed July 18, 2025, https://www.vanta.com/collection/hitrust/hitrust-and-soc-2
46. HITRUST vs SOC 2: Key Differences - Scytale, accessed July 18, 2025, https://scytale.ai/resources/hitrust-vs-soc-2/
47. HITRUST vs. SOC 2 + HITRUST: Which Should You Choose? | Schellman, accessed July 18, 2025, https://www.schellman.com/blog/soc-examinations/hitrust-certification-vs-soc-2-hitrust
48. What Is a CUI Enclave? How Enclaves Can Simplify NIST 800-171 and CMMC 2.0 Compliance - Secureframe, accessed July 18, 2025, https://secureframe.com/blog/cui-enclave
49. New Information Blocking Rules | ACS - American College of Surgeons, accessed July 18, 2025, https://www.facs.org/advocacy/regulatory-issues/digital-health/new-information-blocking-rules/
50. Exceptions to Information Blocking Defined in Proposed Rule: Here's What You Need to Know - Health Catalyst, accessed July 18, 2025, https://www.healthcatalyst.com/learn/insights/information-blocking-exceptions-defined-proposed-rule
51. What Does the TEFCA Mean for Your Organization? - AHIMA, accessed July 18, 2025, https://www.ahima.org/media/zw3hx0c3/tefca_summary_fin.pdf
52. TEFCA | HealthIT.gov - Office of the National Coordinator for Health Information Technology, accessed July 18, 2025, https://www.healthit.gov/topic/interoperability/policy/trusted-exchange-framework-and-common-agreement-tefca
53. Calculating the Total Cost of Ownership for Healthcare Software - WebMD Ignite, accessed July 18, 2025, https://webmdignite.com/blog/calculating-tco-healthcare-software
54. NVIDIA BlueField-3 P-Series DPU - 200Gbe/NDR200 VPI - Dual-Port QSFP112 - Insight, accessed July 18, 2025,

https://www.insight.com/en_US/shop/product/900-9D3B6-00CV-AA0/nvidia/900-9D3B6-00CV-AA0/NVIDIA-BlueField3-PSeries-DPU-200GbeNDR200-VPI-DualPort-QSFP112-PCIe-Gen-50/

55. How RapidScale Helps Healthcare Consulting Firms, accessed July 18, 2025, https://rapidscale.net/resources/case-studies/healthcare-consulting-firm
56. Transforming Your Health Facility: The Essential Role of Healthcare IT Consulting - Riseapps, accessed July 18, 2025, https://riseapps.co/healthcare-it-consulting/
57. EHR Implementation Cost Breakdown: Guide for 2025 - Topflight Apps, accessed July 18, 2025, https://topflightapps.com/ideas/cost-of-ehr-implementation/
58. The Real Total Cost of Ownership of Open Source Integration Engines, accessed July 18, 2025, https://blog.interfaceware.com/the-real-total-cost-of-ownership-of-open-source-integration-engines/
59. How much EHR costs and how to set your budget - EHR in Practice, accessed July 18, 2025, https://www.ehrinpractice.com/ehr-cost-and-budget-guide.html
60. The Ultimate Guide to Healthcare Costs in Data Integration - Number Analytics, accessed July 18, 2025, https://www.numberanalytics.com/blog/ultimate-guide-healthcare-costs-data-integration
61. What is Human-in-the-Loop (HITL) in AI & ML - Google Cloud, accessed July 18, 2025, https://cloud.google.com/discover/human-in-the-loop
62. Human-In-The-Loop: What, How and Why | Devoteam, accessed July 18, 2025, https://www.devoteam.com/expert-view/human-in-the-loop-what-how-and-why/
63. The Influence of Integrated Electronic Medical Records and Computerized Nursing Notes on Nurses' Time Spent in Documentation | Request PDF - ResearchGate, accessed July 18, 2025, https://www.researchgate.net/publication/221697264_The_Influence_of_Integrated_Electronic_Medical_Records_and_Computerized_Nursing_Notes_on_Nurses'_Time_Spent_in_Documentation
64. A 36-Hospital Time and Motion Study: How Do Medical-Surgical Nurses Spend Their Time?, accessed July 18, 2025, https://www.thepermanentejournal.org/doi/10.7812/tpp/08-021
65. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis - Oxford Academic, accessed July 18, 2025, https://academic.oup.com/eurpub/article/26/1/60/2467302
66. Assessing the Impact of an Electronic Medical Record on Nurse Documentation Time, accessed July 18, 2025, https://www.researchgate.net/publication/5251448_Assessing_the_Impact_of_an_Electronic_Medical_Record_on_Nurse_Documentation_Time
67. Time-Motion Analysis of Clinical Nursing Documentation During, accessed July 18, 2025, https://www.researchgate.net/publication/260254196_Time-Motion_Analysis_of_Clinical_Nursing_Documentation_During_Implementation_of_an_Electronic_Operating_Room_Management_System_for_Ophthalmic_Surgery

68. Nurses' Time Allocation and Multitasking of Nursing Activities: A Time Motion Study - PMC, accessed July 18, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC6371290/
69. Recent Advancement of Clinical Information Systems: Opportunities and Challenges - PMC, accessed July 18, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC6115226/
70. The Healthcare Intelligence Platform to Accelerate AI Driven Transformation - Innovaccer, accessed July 18, 2025, https://innovaccer.com/resources/news/innovaccer-launches-gravity-healthcare-intelligence-platform-accelerate-ai-driven-transformation
71. Innovaccer Scales Data Infrastructure with Kloudfuse Observability, accessed July 18, 2025, https://www.kloudfuse.com/customers/innovaccer
72. AI Agents in Healthcare: HMCP Workflow Explained - Innovaccer, accessed July 18, 2025, https://innovaccer.com/resources/blogs/building-a-multi-agent-workflow-in-healthcare-systems-using-hmcp
73. Introducing HMCP: The Healthcare Model Context Protocol - Innovaccer, accessed July 18, 2025, https://innovaccer.com/resources/blogs/introducing-hmcp-a-universal-open-standard-for-ai-in-healthcare
74. innovaccer-desktopApp - GitHub, accessed July 18, 2025, https://github.com/innovaccer/innovaccer-desktopApp
75. OpenMined Homepage, accessed July 18, 2025, https://openmined.org/
76. Understanding the Types of Federated Learning - OpenMined, accessed July 18, 2025, https://openmined.org/blog/federated-learning-types/
77. OpenMined: an example of 'decentralized AI' - Fondazione Giannino Bassetti, accessed July 18, 2025, https://www.fondazionebassetti.org/archi_vivo/2018/09/openmined_an_example_of_decent
78. Secure Enclaves for AI Evaluation - OpenMined, accessed July 18, 2025, https://openmined.org/blog/secure-enclaves-for-ai-evaluation/
79. Federated learning background and architecture - NVIDIA Docs, accessed July 18, 2025, https://docs.nvidia.com/clara/clara-train-archive/3.1/federated-learning/fl_background_and_arch.html
80. Example notebooks demonstrating how to use Clara Train to build Medical Imaging Deep Learning models - GitHub, accessed July 18, 2025, https://github.com/NVIDIA/clara-train-examples
81. AMD Pensando™ DPU Software, accessed July 18, 2025, https://www.amd.com/en/blogs/2023/amd-pensando-dpu-software.html
82. AMD Pensando™ Networking Solutions for the Modern Data Center, accessed July 18, 2025, https://www.amd.com/en/solutions/data-center/networking.html