



Hospital Network Data Engineering

Alexander Miller
Asst Professor

Abstract

Healthcare is increasingly recognized as a data-intensive industry. Multi-hospital networks, among other organizations, face mounting operational and governance challenges because of rigid data-integration pipelines that support all data sources and destinations in the network. These pipelines have become difficult to modify, causing them to lag behind the changing needs of the clinical operation. Scalable data-pipeline architectures better support clinical decision making, optimize hospital operations, ease data quality and compliance concerns, and contribute to improved patient outcomes. Meeting scalability goals requires breaking up monolithic data-integration pipelines into smaller decoupled components and aligning service-level agreements of pipeline components and source systems.

Parallelization and adoption of distributed data-warehouse technology mitigate the burden of ingesting data into a multi-hospital network. However, latency requirements still warrant the construction of separate pipelines for data ingress from clinical devices, electronic health records, and external laboratory-information systems. Healthcare associations recommend near real-time data availability for a growing list of clinical and operational applications. Mishandling the real-time ingestion of data from clinical devices, in particular, compromises availability and performance. Scalable architectural patterns for real-time streaming Ingestion from heterogeneous data sources, transport processes, and back-end processing structures are detailed.

Keywords : Health Data Analytics, Healthcare Data Pipeline, Clinical Decision-Making Support, Data Governance, Healthcare Data Silos, HTAP.

1. Introduction

Instead of merely accumulating large data lakes in a distributed manner, a healthcare data pipeline must provide core administrative, governance, and integration capabilities across hospitals. To meet the needs of healthcare networks at scale, design goals for a data pipeline include modularity, elasticity, observability, security-by-design, and standards-driven observability.

Healthcare organizations in the United States and around the world are striving to overcome clinical decision-support silos that inhibit coordinated care across multiple hospitals. Health outcomes worsen with patient transfers among hospitals as caregivers lose situational awareness. Latency requirements demand a new approach that allows near real-time monitoring and analytics for data science, operational management, and clinical decision support. In practice, hospitals continuously ingest and process information generated by the electronic health record (EHR) system, devices, imaging, social determinants, and external laboratories. Although wide-area communication is relatively slow, data volumes and velocity have reached the point where bundling and managing these workloads through a central pipeline is no longer feasible. Consequently, patient monitoring is often limited to the hospital exhibiting clinical responsibility, and early warning systems that could alert external caregivers remain unrealized.

Healthcare data pipelines must evolve beyond passive, centralized data lake architectures toward intelligent, federated platforms that embed governance, security, and interoperability as foundational capabilities rather than afterthoughts. As healthcare networks grow more interconnected, pipelines must be modular and elastic to support heterogeneous workloads, while providing consistent administrative control, policy enforcement, and standards-based observability across institutions. The persistent fragmentation of clinical data silos directly undermines coordinated care, particularly during patient transfers, where incomplete situational awareness contributes to adverse outcomes. Meeting stringent latency requirements for near real-time analytics and clinical decision support necessitates a distributed processing model that brings computation closer to data sources, enabling continuous ingestion from EHR systems, medical devices, imaging platforms, social determinants sources, and external laboratories without reliance on slow wide-area aggregation. By adopting a security-by-design, interoperable, and observability-rich architecture, healthcare organizations can unlock cross-hospital early warning systems, shared patient monitoring, and collaborative analytics—transforming raw data streams into actionable intelligence that supports timely, high-quality, and coordinated patient care.

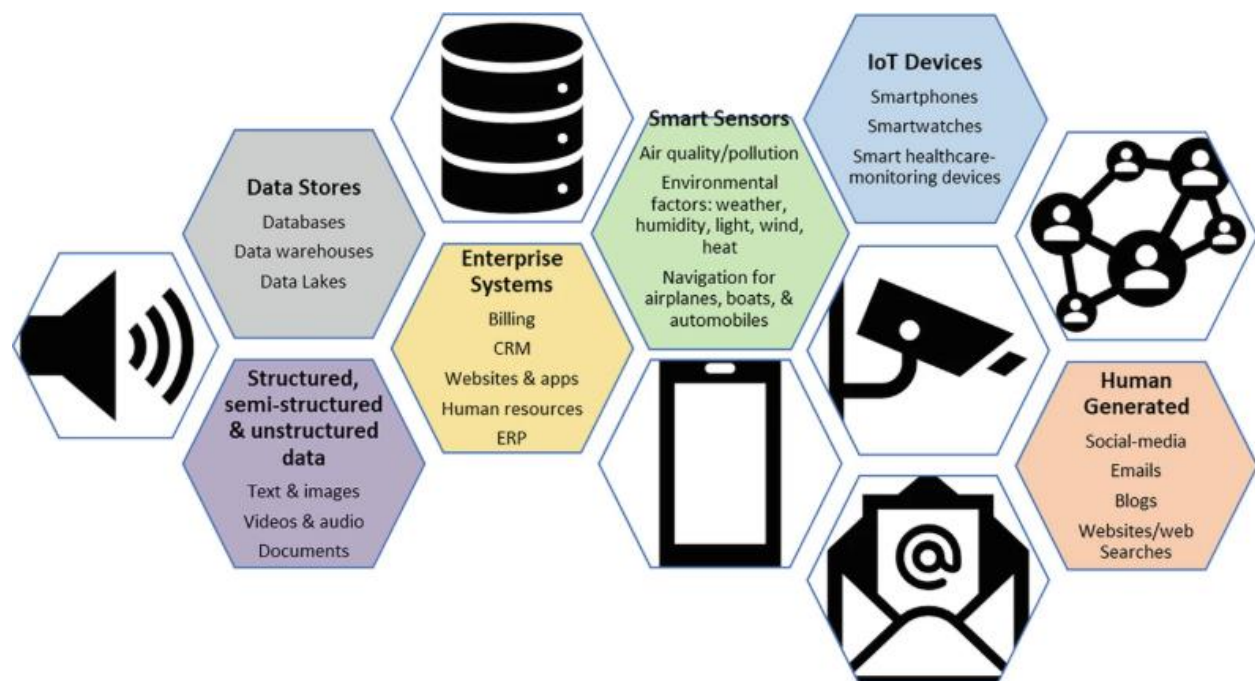


Fig 1: Health Data Pipelines

1.1. Background and Significance

Isolated data silos in healthcare networks impede timely access to consolidated datasets for business and data science applications. Suppliers and vendors focused on real-time monitoring and digital transformation mandate low-latency data movement between hospital systems, capitalizing on operational uptime. In some instances, data-fusion and decision-support



tasks are becoming more complex, as seen in multi-hospital networks where patient queries can span multiple places of care. These inefficiencies are often linked to a lack of integration. Traditional approaches that bolt-on an enterprise platform add yet another data silos for sensitive healthcare information unless a rigorous data stewardship program is followed. They also require a specialized team for implementation that is expensive to hire and hard to retain. And even the smaller scale implementation may include hiring cloud resources with deep cloud-specific expertise.

A scalable architecture provides a different approach. Rather than bolt a “one-size-fits-all” solution onto a data ecosystem, it visually draws architectural principles based on the quality of service expectations of various data use cases. By decoupling the various data pipeline components, it become possible to align execution with the needs, performance profile and depth or breadth of analysis. Using different hosting strengths to provide transaction-based services and batch-based analysis improves the time and cost of delivery and support where the assets are deployed. Such an architecture is also easier to adapt and maintain over time since specialized teams can support the individual components without having to understand the nitty-gritty of the entire data ecosystem.

Equation 1: SLA fundamentals: latency, availability, completeness

Define per-event latency components:

- T_{ingest} : time to capture/normalize at source (or edge)
- T_{queue} : waiting time in message broker / buffer
- T_{proc} : compute time in stream processor
- T_{persist} : time to write to storage/warehouse/index

Equation (sum of stage latencies):

$$T_{\text{e2e}} = T_{\text{ingest}} + T_{\text{queue}} + T_{\text{proc}} + T_{\text{persist}}$$

2. Background and Rationale

Evidence-based analysis highlights scalability in designing data pipelines that support multi-hospital networks. Scalability is critical as service quality, patient safety, and costs depend on timely access to sensitive clinical information. Scalability considerations help maintain tight service level agreements for latency, completeness, and availability. A scalable pipeline supports data governance, provenance, and compliance with regulations such as HIPAA. Scalability is also important for AI-based predictions and data quality monitoring. Shifting architectural principles from monolithic systems to scalable distributed designs enables greater modularity and elasticity, promotes observability, enforces security-by-design, and enables interoperability using established standards.

Healthcare delivery in the United States has experienced a paradigm shift in recent years, leading patient care to be centralized in multi-hospital networks. A scalable data engineering pipeline is essential to meet health systems' growing information needs; evidence demonstrates that timely access to sensitive clinical information improves patient outcomes and reduces costs, while data compliance with regulations such as HIPAA minimizes risks to patients and health systems. However, panic-induced inaction has largely converted the data and analytics landscape into a series of disparate, localized data silos supporting specific projects with limited returns and repeated costs. A considerable gap exists between the original growth ambitions and the actual state of the infrastructure, which is neither scalable nor fault-tolerant. Indeed, especially during the COVID-19 pandemic, data-informed prediction of healthcare system utilization has proven timely availability of relevant information to decision-makers a paramount factor in ensuring security, quality, and safety in service delivery. Therefore, providing healthcare operators with near-real-time advanced notifications of service latency, quality, and completeness failures could greatly improve both the quality of service delivery and clinical decision-making.

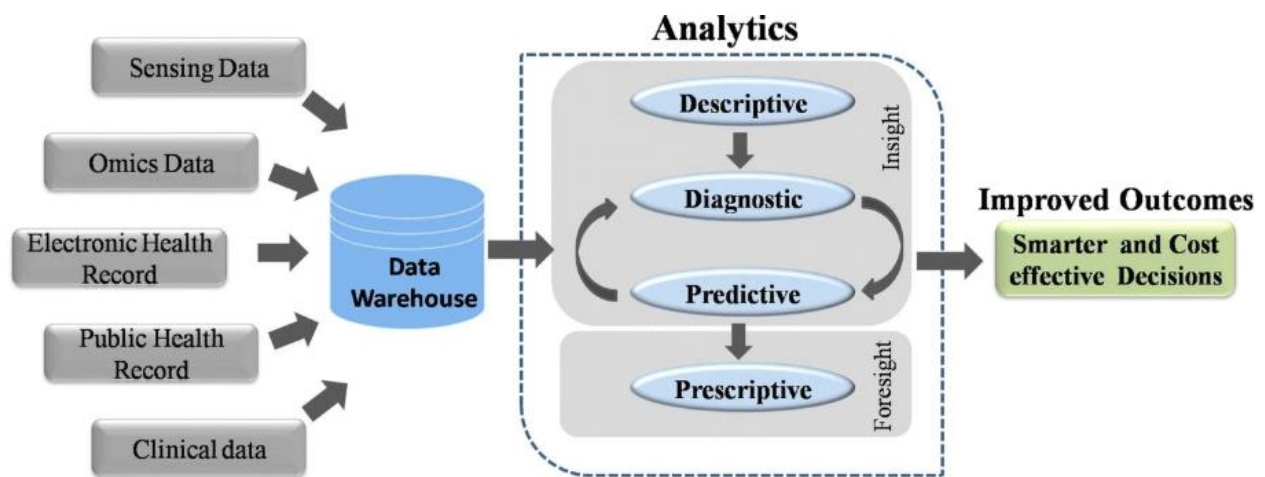


Fig 2: Background and Rationale of Designing Scalable Healthcare

2.1. Research design

The research design constitutes a comparative analysis evaluating architectural patterns in the expansion, adaptation, and integration of data-processing pipelines within multi-hospital networks. It defines the desired characteristics of a scalable pipeline, formulates four hypotheses concerning the choice of components and design principles, identifies critical factors for governance and clinical decision support, specifies criteria for comparing traditional and newly proposed solutions, and describes the simulated data scenarios employed to validate the evaluation framework. Data partitions serve as the experimental setting, while the availability of the Data Pipeline Patterns for Healthcare scale and the ongoing Google Data Analytics Capstone project for Coursera provide the empirical foundation.

A multi-hospital network management team has requested recommendations on the pipeline design option to pursue for new data sources with different latency requirements. Scalability has been defined as the system's ability to accommodate concurrent users



and data volume growth with manageable cost; quality is often an implicit requirement. However, criteria such as responsiveness, tolerance of unanticipated changes, intrusion into standard operations, and compliance with data-governance policies are also of strategic significance. In the context of a multi-hospital network, quality has been explicitly linked to the capacity to consolidate, cleanse, and harmonize data in close to real time.

Equation 2: Availability (simple definition used in ops)

Let:

- MTBF = mean time between failures
- MTTR = mean time to repair

Then:

$$A = \frac{MTBF}{MTBF + MTTR}$$

If a pipeline has multiple required components in series (all must be up), a simple approximation is:

$$A_{\text{series}} \approx \prod_{i=1}^n A_i$$

3. Architectural Principles for Scalability

Defining core architectural principles serves two purposes. First, identifying common qualities in scalable healthcare pipelines normalizes future comparisons and charts potential deviations. The focus is on modularity, elasticity, observability, security-by-design, and standards-driven interoperability. The second aim is guidance for concrete architectures. A clear articulation of the desired properties informs the design of specific pipeline components and their interconnections. Individual pipelines can then adopt differentiated architectural choices, provided that the resulting components satisfy the principles.

At a high level, the key requirement is that each pipeline can respond independently to increments in data volume and velocity. Synchronous requests from clinical applications must remain responsive even when data is being ingested from multiple sources or the volume of back-end processing grows. Modular component design is essential to achieving this decoupling, together with observability tooling that monitors the load on different stages and supports dynamic scaling decisions. Hospitals are often required to follow strict security and governance processes on data usage. Therefore, it makes sense to manage the ingestion workload independently and schedule these tasks according to local policies, without incurring excessive delays. For this purpose, an event-driven architecture decouples the source data from its destination through a messaging system. Such an architecture further facilitates the integration of healthcare devices and third-party data sources.

Defining core architectural principles establishes a foundation for building scalable, resilient healthcare data pipelines while enabling consistent evaluation across diverse implementations. By emphasizing modularity, elasticity, observability, security-by-

design, and standards-driven interoperability, organizations can normalize how future architectures are compared and identify meaningful deviations from best practice. At the same time, these principles provide actionable guidance for concrete system design: individual components and their interconnections can vary in technology or topology, as long as they collectively uphold the desired properties. Central to this approach is the requirement that each pipeline scale independently in response to changes in data volume and velocity, ensuring that synchronous clinical workloads remain responsive even as ingestion rates or downstream processing demands increase. Modular, loosely coupled components—supported by comprehensive observability and dynamic scaling mechanisms—enable this decoupling, while strict security and governance considerations motivate independent management of ingestion workflows in accordance with local policies. An event-driven architecture, built around messaging systems, further reinforces these goals by separating data producers from consumers, enabling asynchronous processing, and simplifying the integration of heterogeneous healthcare devices and third-party data sources without introducing excessive latency.

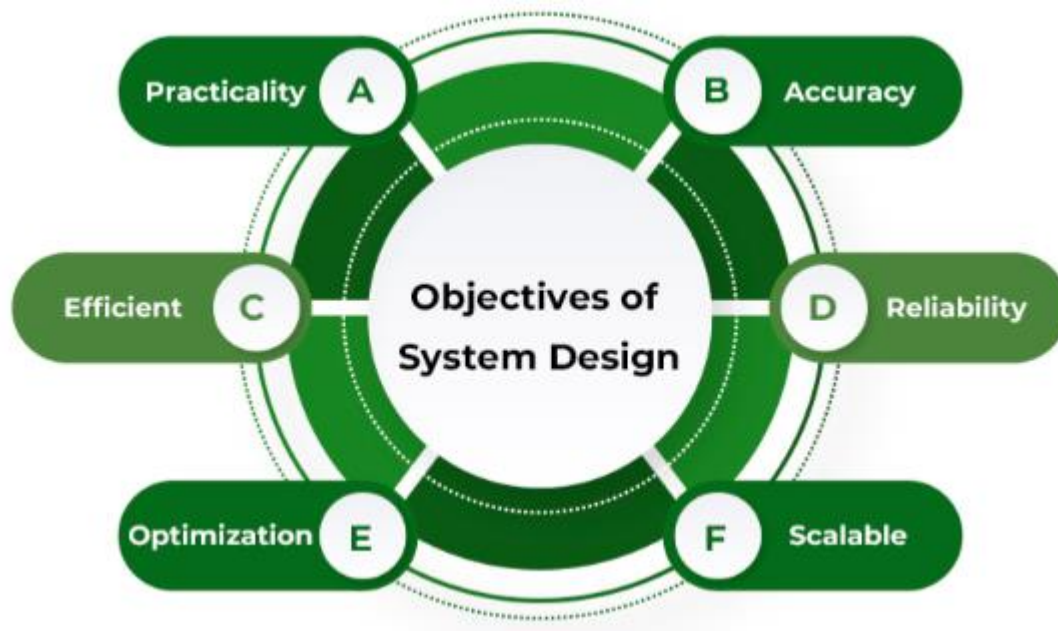


Fig 3: Essential System Design Principles for Scalable Architectures

3.1. Data Partitioning and Sharding

Architectural scalability is a multidimensional property, and there are many ways to approach it. Regardless of the strategy chosen, it is often desirable to reduce cross-location traffic, since different hospitals within a healthcare network will usually be in different geographical regions, with the associated latency, internet egress costs, and data exfiltration risk. Data partitioning at the pipeline level can help achieve this goal. By orchestrating the routing of data flows according to logical partitioning strategies, data ingress points in individual hospitals are able to ingest data only for the local environment. In the case of a clinical



notification pipeline that sends real-time alerts by integrating the events emitted from streaming data models per hospital, partition keys could be designed to choose the correct endpoint based on the patient context.

Modifying the traffic direction and the way message queues are used in a data pipeline is a classic method for ensuring elasticity. For example, when patient entities are assigned to hospital groups that serve as clusters, a pool of application server clusters can be deployed for each group. A load balancer manages incoming traffic from users and dispatches onto the cluster. As hospitals migrate to more important parts of the architecture, be it ingestion or consumption, a moving average can be applied to key monitoring parameters such as resource consumption or traffic rate. A smooth increase or decrease will be controlled by the pools of containers running inside Kubernetes or App Engine.

Equation 3: Completeness (data quality / missingness rate)

Let:

- N_{expected} = events expected in a time window
- N_{received} = events actually ingested (after dedupe rules)

Then:

$$\text{Completeness} = \frac{N_{\text{received}}}{N_{\text{expected}}}$$

and missingness:

$$\text{Missing rate} = 1 - \text{Completeness}$$

4. Data Models and Schema Management

Flexible data models accommodate the inevitable evolution of upstream business systems while supporting data quality, traceability, and regulatory compliance. Everything should change so that everything can stay the same. Or so argues the character Tancredi in Tomasi di Lampedusa's novel *The Leopard*, a tale of the decline of the Sicilian aristocracy in the 19th century. This notion is frequently whimsically rendered as If you want things to stay the way they are, things will have to change.

Changing things is certainly the case when it comes to the Healthcare industry, with new regulations, consolidation pressures, heightened patient experience demands, and emerging technology reshaping virtually every aspect of service delivery. These changes reverberate both upstream and downstream of an organization. For example, to support a heightened focus on patient outcome, Electronic Health Records (EHRs) now often require far more associated metadata to be meaningful and useful than was the case just a few years ago. Driving these changes is the growing realisation that a robust machine learning based data-driven approach can provide solutions to many of the challenges faced by Healthcare organisations today. Common examples include predictive staffing, preventive medicine, and risk determination to name just a few. However, the foundation of any data-

driven approach is high quality data. To be useful, such data must also be timely, and yet these changes are introducing delays and silos in the creation of new data.

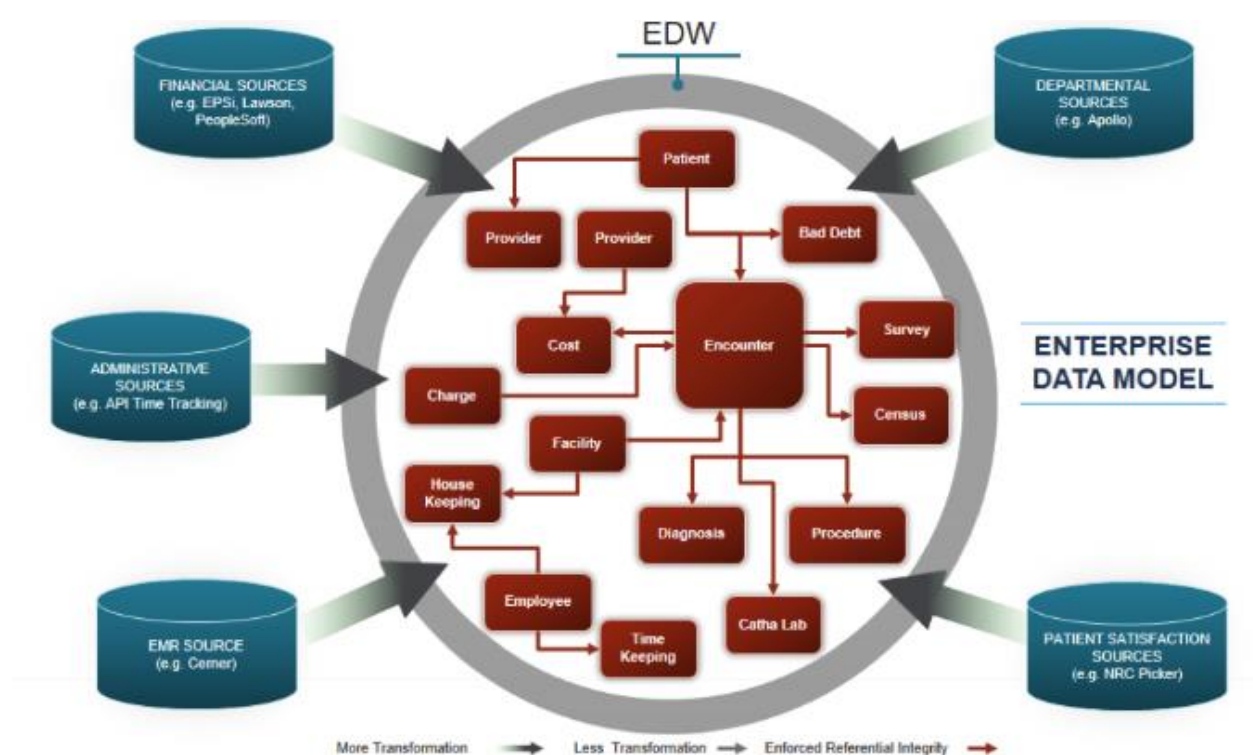


Fig 4: Healthcare Data Warehouse Models

4.1. Domain-Driven Modeling in Healthcare

Healthcare data can be divided into five major domains, each with specific concerns and regulations. Compliance with the respective regulations is important to guarantee data quality before the information is reflected on data models and used for further analysis. The most critical domains include clinical, administrative, billing, and operating. Furthermore, the U.S. Government and international relations created and still support two main standards for data and information interchange and transmission among and between healthcare providers, agencies, and organizations. The last important component of the domain dashboard is the glossary, which presents the terms used and their meanings, constituting a ubiquitous language for the healthcare domain model.

A bounded context describes an explicit agreement between all partners concerning the meaning of a term; in the healthcare environment, it refers to the definitions presented by the main standards that rule the respective domain. Clinical information corresponds to the established standard for clinical information interchange and transmission exchange format. Information for an



operation is the standard dedicated to information interchange, which also concerns major clinical equipment in the healthcare system. Billing information corresponds to the billing systems and processes in a healthcare organization, while administration refers to the administration of patients' data in a healthcare system. The glossary uses terms from those domains to connect and explain the data. In addition to the definitions of external standards, it also contains explanations for other terms.

Equation 4: Queueing + back-pressure for real-time ingestion

Let:

- λ = arrival rate (events/second)
- μ = service rate (events/second)

Utilization:

$$\rho = \frac{\lambda}{\mu}$$

Stability condition (no unbounded backlog):

$$\lambda < \mu \quad \Leftrightarrow \quad \rho < 1$$

5. Data Ingestion and Integration Strategies

Designing Healthcare Data Pipelines for Multi-Hospital Networks: Data Ingestion and Integration Strategies across Heterogeneous Sources and Formats. The proposed designs for a network of healthcare data pipelines address not only the volume and velocity of data arrival but much more, including the variety of data formats and structures from different software applications, the quality demands of the data, auditability and lineage tracking, and so forth. Such elements are expected to be crucial for scalable data pipelines. An analysis of the gap between the desired properties of the data and those imposed by the sources leads to the choice of using data pipelines rather than simply data lakes for data storage.

Data pipelines are an attractive option for responding to the assimilation of heterogeneous and complex data generated within a hospital context thanks to their properties of accessibility, auditability, reliability, and governance; these properties arise from the layered architecture that pipelines follow, allowing the construction of dedicated components responsible for these aspects. Above all, the concept of data governance is capital for organizations that work with data about patients and it includes aspects of importance for the ownership of the data, de-duplication, consistency, relationships with several other components, and much more. A complete data governance, taking care of all the elements above, is usually tackled by projects of dedicated MDF. Nevertheless, when the data is processed in a pipeline, the governance aspects can already be introduced with the corresponding data quality checks, because the modern trend is that the raw data should be kept as long as possible and usually in the level 0 of the data lake. Therefore, it is enough that the data quality framework only takes care of the data generated by the pipeline since they should be already governance compliant. Indeed, the additional challenge consists of taking care of the missing data in an external provenance system.



The design should ensure that a proper level of quality is given to the data when inserted in the data lake; the clear definition of markers in the data can play a big part in the data quality, allowing the detection of duplicate records for example.

Equation 5: Little's Law (very common in pipeline sizing)

Let:

- L = average number of events "in the system"
- W = average time an event spends in the system

Little's Law:

$$L = \lambda W$$

5.1. Real-Time Streaming Ingestion

Real-time streaming architectures address the need to minimize event latency by capturing high-update-rate events immediately at source. A sequential data-flow composed of ingest orchestration, persist-and-forget messaging, and stream processing steps allows decoupling of the components, enabling independent scaling with respect to back-pressure handling, horizontal elasticity, and capability alignment with the service level agreements for event read access defined by the consumers. Depending on the source of the stream, additional quality and consistency mechanisms may need to be integrated into the design.

In healthcare environments, devices are the most straightforward source of real-time streams, typically emitting well-formed events heralding not only the physical measures but also the context of the measurements (e.g., patient ID, creation timestamp). Stream ingestion from Electronic Healthcare Records continues to be a challenge, as clinical workflows do not permit interruption upon individual event generation. The event sources are typically append-derived tables, where a record (orm) has been inserted or modified in the last transaction. Messages emitted from other system components (external laboratories, billing) are customarily unaffected by these limitations.



other jurisdictions should clarify concepts around data collation, sharing for societal benefit, and use of identifiable and quasi-identifiable information.

Design-science research could practically contribute to these future trends. Exploratory work could investigate the promise of generic middleware systems and components, such as those for data access and quality, support for Latin American and European health standards and data collation in the clinical domain. Design work could provide a real-time stream-based reader component that supports advanced back-pressure management and low-latency usage, together with support for EHR digitization and the scrubbing and tagging of free-text clinical documents. A simple validation procedure could also strengthen it by measuring latency and data volume for a defined period, job performance, resource footprint, and operational issues.

Hospital	Arrival λ (events/s)	Instances	Service μ (events/s)
H1	1200	8	1600
H2	900	6	1200
H3	1500	10	2000
H4	700	5	1000
H5	1100	7	1400

Table: Illustrative per-hospital queueing/bandwidth metrics

7. References

1. Aerts, H., Kalra, D., Sáez, C., Ramírez-Angueta, J. M., Mayer, M.-A., Garcia-Gomez, J. M., Durà-Hernández, M., Thienpont, G., & Coorevits, P. (2021). Quality of hospital electronic health record (EHR) data based on the International Consortium for Health Outcomes Measurement (ICHOM) in heart failure: Pilot data quality assessment study. *JMIR Medical Informatics*, 9(8), e27842.
2. Tian, Q., Han, Z., Yu, P., An, J., Lu, X., & Duan, H. (2021). Application of openEHR archetypes to automate data quality rules for electronic health records: A case study. *BMC Medical Informatics and Decision Making*, 21, 113.
3. Razzaghi, H., Greenberg, J., & Bailey, L. C. (2021). Developing a systematic approach to assessing data quality in secondary use of clinical data based on intended use. *Learning Health Systems*, 6(1), e10264.



4. Mattaparthi, R. (2024). Transformer-Based Fault Diagnosis for Large-Scale Standby Power Generators: Partial Discharge Pattern Recognition at Hyperscale Data Center Installations. *Journal of Computational Analysis and Applications (JoCAAA)*, 33(08), 8781-8799.
5. Li, E., Clarke, J., Neves, A. L., Ashrafian, H., & Darzi, A. (2021). Electronic health records, interoperability and patient safety in health systems of high-income countries: A systematic review protocol. *BMJ Open*, 11(7), e044941.
6. Ozonze, O., Scott, P. J., & Hopgood, A. A. (2023). Automating electronic health record data quality assessment. *Journal of Medical Systems*, 47(1), 23.
7. Bernstam, E. V., Warner, J. L., Krauss, J. C., Ambinder, E., Rubinstein, W. S., Komatsoulis, G., Miller, R. S., & Chen, J. L. (2022). Quantitating and assessing interoperability between electronic health records. *Journal of the American Medical Informatics Association*, 29(5), 753–760.
8. Mangalampalli, B. M. (2024). Transparent Intelligence Explainability Frameworks for AI-Driven Clinical Decision Support in Healthcare Business Intelligence. *International Journal of Research Publications in Engineering, Technology and Management (IJPETM)*, 7(3), 10566-10579.
9. Cheng, K. Y., Pazmino, S., & Schreiweis, B. (2022). ETL processes for integrating healthcare data—Tools and architecture patterns. *Studies in Health Technology and Informatics*, 294, 355–359.
10. Romanchikova, M., & Hippolyte, J.-L. (2022). A framework for user-configurable data quality assurance of electronic patient records. *Studies in Health Technology and Informatics*, 289, 89–92.
11. Schmeelk, S., Kanabar, M., Peterson, K., & Pathak, J. (2022). Electronic health records and blockchain interoperability requirements: A scoping review. *JAMIA Open*, 5(3), ooac068.
12. Goldstein, N. D., Kahal, D., Testa, K., Gracely, E. J., & Burstyn, I. (2022). Data quality in electronic health record research: An approach for validation and quantitative bias analysis for imperfectly ascertained health outcomes via diagnostic codes. *Harvard Data Science Review*, 4(2).
13. Loganathan, R., & Kolla, S. H. (2024). Harnessing generative AI for adaptive risk policy generation in multi-regulatory data center environments. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 15(3), 505-515.



14. Diaz-Garelli, F., Long, A., Bancks, M. P., Bertoni, A. G., Narayanan, A., & Wells, B. J. (2022). Developing a data quality standard primer for cardiovascular risk assessment from electronic health record data using the DataGauge process. *AMIA Annual Symposium Proceedings*, 2021, 388–397.
15. Wiley, K. K., Mendonca, E., Blackburn, J., Menachemi, N., De Groot, M., & Vest, J. R. (2022). Quantifying electronic health record data quality in telehealth and office-based diabetes care. *Applied Clinical Informatics*, 13(5), 1172–1180.
16. Avendano, J. P., Gallagher, D. O., Hawes, J. D., Boyle, J., Glasser, L., Aryee, J., & Katt, B. M. (2022). Interfacing with the electronic health record (EHR): A comparative review of modes of documentation. *Cureus*, 14(6), e26330.
17. Reddy, V. A. R. (2024). Generative Intelligence for Healthcare Claims Processing and Personalized Benefits Management. *International Journal of Advanced Engineering Science and Information Technology (IJAESIT)*, 7(3), 14099.
18. Mihăilescu, A. I., et al. (2022). Semantic interoperability in health records standards: A systematic literature review. *Health and Technology*, 12, 1009–1024.
19. Kaspar, M., Liman, L., Morbach, C., Dietrich, G., Seidlmayer, L. K., Puppe, F., & Störk, S. (2023). Querying a clinical data warehouse for combinations of clinical and imaging data. *Journal of Imaging Informatics in Medicine*, 36, 715–724.
20. Kohler, S., Boscá, D., Kärcher, F., Haarbrandt, B., Prinz, M., Marschollek, M., & Eils, R. (2023). Eos and OMOCL: Towards a seamless integration of openEHR records into the OMOP Common Data Model. *Journal of Biomedical Informatics*, 144, 104437.
21. Hechtel, N., Apfel-Starke, J., Köhler, S., Fradziak, M., Schönfeld, N., Steinmeyer, J., & Oeltze-Jafra, S. (2023). Harmonisation of German health care data using the OMOP Common Data Model—A practice report. *Studies in Health Technology and Informatics*, 305, 287–290.
22. Kolla, S. H., & Peddi, R. K. (2024). Designing Governance-Aligned GenAI Pipelines Using Small Language Models for Enterprise Workflow Intelligence. *International Journal of Science, Research and Technology*, 7(6), 13256-13268.
23. Oja, M., Tamm, S., Mooses, K., Pajusalu, M., Talvik, H.-A., Ott, A., Laht, M., Malk, M., Lõo, M., Holm, J., Haug, M., Šuvalov, H., Särg, D., Vilo, J., Laur, S., Kolde, R., & Reisberg,



- S. (2023). Transforming Estonian health data to the Observational Medical Outcomes Partnership Common Data Model: Lessons learned. *JAMIA Open*, 6(4), ooad100.
24. Williams, N. (2023). Building the Observational Medical Outcomes Partnership's T-MSIS Analytic File common data model. *Informatics in Medicine Unlocked*, 39, 101251.
 25. Uwase, M., Iradukunda, J. D., Rusa, D. U., Ndahimana, R., Mvuyekure, B., Birindabagabo, P., Ruton, H., Mpunga, T., Mugisha, M., Twizere, C., & Tumusiime, D. (2023). Towards improving data quality in electronic medical records: An investigation of data completeness in a tertiary hospital in Rwanda. *Studies in Health Technology and Informatics*, 305, 390–393.
 26. Mangala, N. (2022). Real-Time Data Quality Monitoring and Gating Frameworks in Cloud-Based Data Pipelines. *International Journal of Research and Applied Innovations*, 5(6), 8197-8219.
 27. Oh, S. W., Ko, S. J., Im, Y. S., Jung, S., Choi, B. Y., Kim, J. Y., Choi, W., & Choi, I. Y. (2023). Data quality assessment for Observational Medical Outcomes Partnership Common Data Model of multi-center. *Studies in Health Technology and Informatics*, 302, 322–326.
 28. Madandola, O. O., Bjarnadottir, R. I., Yao, Y., Ansell, M., Dos Santos, F., Cho, H., Dunn Lopez, K., Macieira, T. G. R., & Keenan, G. M. (2023). The relationship between electronic health records user interface features and data quality of patient clinical information: An integrative review. *Journal of the American Medical Informatics Association*, 31(1), 240–255.
 29. Gude, F., et al. (2023). Interoperability of heterogeneous health information systems: A systematic literature review. *Journal of Medical Systems*, 47, 1–15.
 30. Anand, G., & Sadhna, D. (2023). Electronic health record interoperability using FHIR and blockchain: A bibliometric analysis and future perspective. *Perspectives in Clinical Research*, 14(4), 187–194.
 31. Stausberg, J., et al. (2023). An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. *International Journal of Medical Informatics*, 169, 104925.
 32. Zhang, S., et al. (2024). Cross-standard health data harmonization using semantics of data elements. *Scientific Data*, 11, 1407.



33. Papapostolou, G., Chytas, A., Rekkas, A., Bigaki, M., Zeimpekis, D., Dermentzoglou, L., Tortopidis, G., & Natsiavas, P. (2024). Real-world data in Greece: Mapping the Papageorgiou General Hospital data to the OMOP Common Data Model. *Studies in Health Technology and Informatics*, 316, 1324–1325.
34. Fraser, H. S. F., Mugisha, M., Bacher, I., Ngenzi, J. L., Seebregts, C., Umubyeyi, A., & Condo, J. (2024). Factors influencing data quality in electronic health record systems in 50 health facilities in Rwanda and the role of clinical alerts: Cross-sectional observational study. *JMIR Public Health and Surveillance*, 10, e49127.
35. Förstel, S., Förstel, M., Gallistl, M., Zanca, D., Eskofier, B. M., & Rothgang, E. M. (2024). Data quality in hospital information systems: Lessons learned from analyzing 30 years of patient data in a regional German hospital. *International Journal of Medical Informatics*, 192, 105636.
36. Mahadevan, S. (2024). Clouding the Future: Innovating Towards Net-Zero Emissions. *International Journal of Computing and Engineering*, 6(2), 17-23.
37. Tabari, P., Costagliola, G., De Rosa, M., & Boeker, M. (2024). State-of-the-art Fast Healthcare Interoperability Resources (FHIR)-based data model and structure implementations: Systematic scoping review. *JMIR Medical Informatics*, 12, e55062.
38. Nițulescu, A., & Stoicu-Tivadar, L. (2024). Data standardization in the medical field through FHIR and FAIR implementation: A systematic review. *Studies in Health Technology and Informatics*, 316, 1128–1132.
39. Hoffmann, N., Nesterow, S., Peng, Y., Henke, C., Barnett, N., Klengel, A., Gruhl, M., Bartos, S., Nüßler, J., Gebler, F., Grummt, S., Seim, J., Bathelt, M., Reinecke, S., Wolfien, M., Weidner, S., & Sedlmayr, M. (2024). Streamlining intersectoral provision of real-world health data: A service platform for improved clinical research and patient care. *Journal of Medical Internet Research*, 26, e53547.
40. Nesterow, S., et al. (2024). Uncovering harmonization potential in health care data through iterative refinement of Fast Healthcare Interoperability Resources profiles based on retrospective discrepancy analysis: Case study. *JMIR Medical Informatics*, 12, e57005.
41. Zhang, X., et al. (2024). Development and initial validation of a data quality evaluation tool in obstetrics real-world data through HL7-FHIR interoperable Bayesian networks and expert rules. *JAMIA Open*, 7(3), ooae062.



42. Kleesiek, J., et al. (2024). Establishing medical intelligence—Leveraging Fast Healthcare Interoperability Resources to improve clinical management: Retrospective cohort and clinical implementation study. *Journal of Medical Internet Research*, 26, e55148.