International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

Cloud-Based Data Fabric Architectures for National Healthcare Interoperability

Ganesh Pambala

Independent Researcher, India

ganeshpambal@gmail.com

ABSTRACT: Cloud-based services are revolutionizing storage and processing, breaking the strong coupling between applications, data, and infrastructures within a single organization. Different components can now be deployed in different geographic areas, in a flexible and cost-effective manner. A cloud-based architecture for data virtualization, sensitive-data-oriented, based on a fabric model, is presented. The architecture optimizes operational and system-recovery costs by leveraging a cloud resources-market-driven processing environment. A sensitive-data protection mechanism prevents sensitive data from being exposed even during data re-use. A system-cache approach based on a monitoring and statistical model mechanism improves system utilization and performance by offloading data from the system cache to cloud storage. Concepts, principles, phases, and building blocks of the fabric support any type of resource and information service from the cloud, and any resource-market-driven virtualization requirements that improve application-provisioning flexibility and processing cost.

The presentation of National Cloud Platform for Data Sharing shows that Public and Private sectors maintain distributed data resources in Taiwan, but most national data cannot be integrated and shared easily, due to different forms, contents, and access mechanisms. Recently, strategy of Cloud Computing was proposed to solve resource-usage optimization and fault-tolerant problems. A Cloud Platform connecting the Information-resource-provide agencies by Cloud links is developed, so that their resources can be shared without unnecessary data-refining and –duplication processes. The platform optimizes last-mile data-sharing processes for data consumers and build a Cloud Sharing Marketplace for Information Resources.

KEYWORDS : Cloud-native data fabric,Healthcare interoperability frameworks,HL7 FHIR integration,National health information exchange (HIE),Multi-cloud healthcare architecture,Federated health data platforms,Healthcare API gateway,Real-time clinical data streaming,Master patient index (MPI),Secure health data governance,HIPAA-compliant cloud infrastructure,Semantic data harmonization,AI-enabled health data analytics,Zero-trust security in healthcare IT,Cross-border healthcare data exchange.

I. INTRODUCTION

Cloud-Based Data Fabric Architectures for National Healthcare Interoperability

Electronic Health Records (EHRs) contain vast amounts of healthcare data but use incompatible formats and standards. These differences prevent integrated data analysis and make it difficult to maintain analytics-based next-best action services. Consequently, large-scale government-initiated homelessness, economics, tourism, travel, and accident prevention analytics remain elusive. Diverse cloud-based platforms operated by Public-Sector Units (PSUs) are used for a range of healthcare services. A cloud-based national health data fabric that serves as an interoperability layer will help derive useful economy-wide analytics from EHRs. Such a healthcare-data fabric must implement data-fabric principles using specialized data-virtualization capabilities.

Cloud-based multi-cloud and hybrid approaches are emerging as popular deployment models for various businesses, with development efforts gaining support from Global System Integrators (GSIs) and various Cloud Service Providers

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

(CSPs). Therefore, fabric-data architectures are becoming critical for establishing a hybrid-shared or multi-cloud-based data layer to provide a seamless experience for customers using different services. Data governance, access, policies, and legal-compliance mechanisms must also be in place because enterprise-wide or multi-enterprise data access is the key analytics-enabler requirement for breach detection and response for organizations or enterprises in the national healthcare service ecosystem. Data-driven Next-Best-Action service capabilities become highly useful for related parties in the ecosystem, viz. Police, Revenue and Disaster Management, Airports Authorities and Airlines, Transport and Travel Industry, and Local Governments.

Recent research identified gaps in voice, video, and language bias in existing large language models. The health industry has been a laggard adopter of these models, but Open AI has announced plans for a health domain-focused version of its renowned ChatGPT AI language tool.

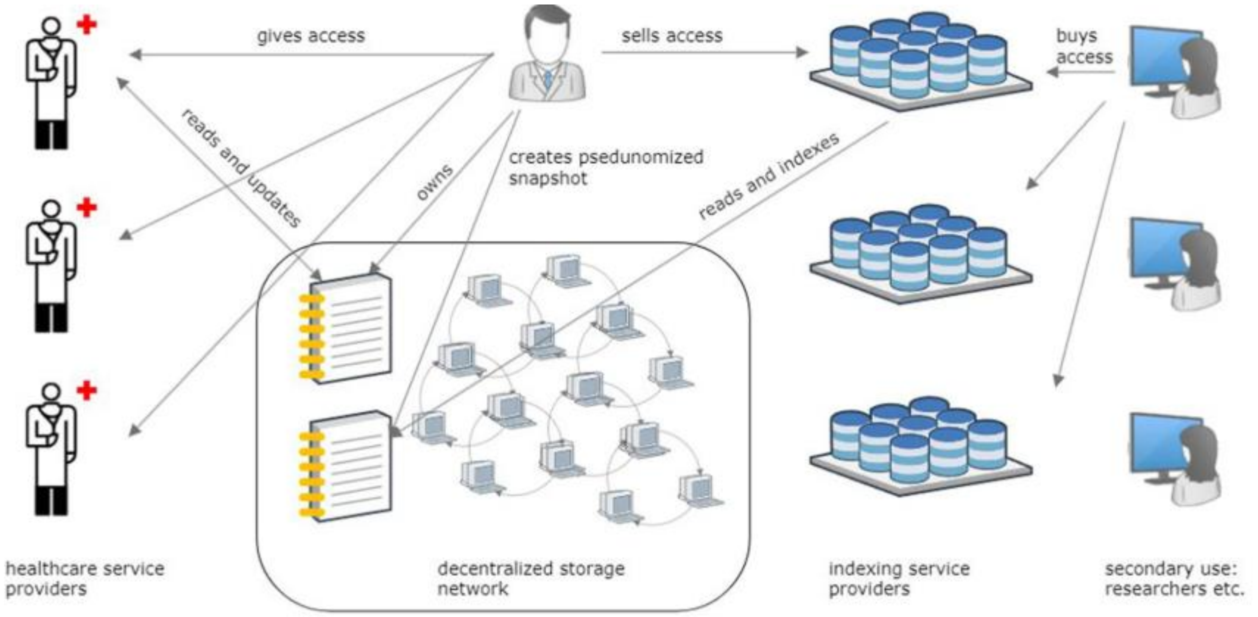


Fig 1: Architecture for personal health data

1.1. Background and Significance

The United States health system fails to deliver on one of our nation's founding principles: equality of access and opportunity for all, despite major investments of time and resources and decades of effort by many motivated people. Poor health and shorter lives cost our nation dearly—not just in lost productivity and unnecessary suffering, but in the funds required to care for the unhealthy. A major contributing factor to ineffective use of the country’s healthcare expenditure appears to be the lack of interoperable electronic health records data. The establishment of a trusted national healthcare data exchange service that would allow healthcare providers to transmit confidential information about their patients' health histories for treatment purposes is proposed as a targeted solution.

Over the past four decades, decades more than a half-dozen major federal initiatives have attempted to remedy this situation. The most recent and longest-lasting effort, the 21st Century Cures Act, aims to provide patients with access to their electronic health information. Much of the needed infrastructure is in place, but progress in realizing this vision has been slow, and extensive further effort is required to encourage or require healthcare organizations to comply.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

Equation 1: Define workload and decision variables

- Let $W(t)$= workload demand (compute units/hour).
- Let there be $m$clouds.
- Let $x_i(t) \in [0,1]$be the fraction of workload placed on cloud i, with:

$$\sum_{i=1}^{m} x_i\,(t) = 1$$

II. FOUNDATIONS OF CLOUD-BASED DATA FABRICS

Data virtualization and fabric principles

Data virtualization is an integration and abstraction technology that provides a single addressable logical view of data from multiple and heterogeneous sources without requiring physical data movement. Its key features include: data federation; a unified and consistent view of data that may be semantically heterogeneous; a data abstraction layer that uses a subset of SQL; agile integration, enabling the decoupling of data consumption from data provision; on-the-fly business transformation; real-time access; and cross-source join capabilities.

The adoption of data virtualization solutions is growing in areas where faster time to value, cross-source agility, reduced data copies, data-lake augmentation, and architectural simplification are valued. However, organizations also recognize challenges related to the performance overhead of on-the-fly processing of complex transformations, the performance trade-offs of cross-source joins, and addressing analytics events where a data copy is needed for performance reasons. As such, data virtualization organizations adopt a “data virtualization where it makes sense” approach. Industry analysts have reported that enterprises are growing increasingly interested in data fabric concepts—virtualization capabilities in a broader data-exchange context that adds automated data-curation capabilities and external data-service leverage for connected consumption.

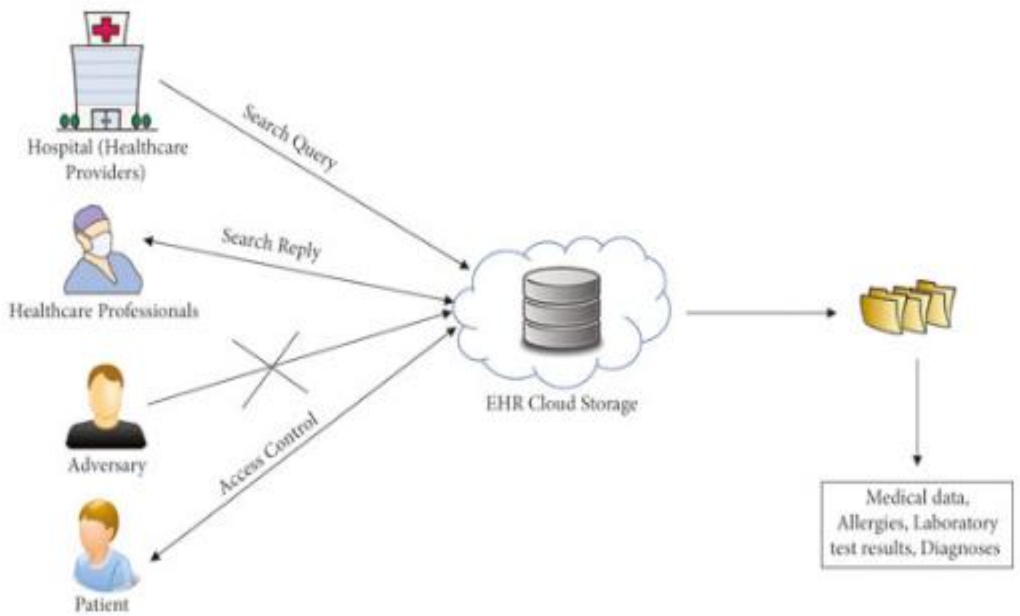


Fig 2: Cloud-based electronic health data architecture

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

2.1. Data virtualization and fabric principles

Data virtualization defines a technology and an approach for assembling and presenting data from disparate and often geographically distributed data sources, in real-time and in a unified view, without requiring substantial data movement or duplication.

A data fabric is an architectural approach, enabled by data virtualization, that provides a unified data model, along with the required tooling, best practices, and governance, to support responsive, business-driven, self-service data access, and sharing across heterogeneous data sources. This extends beyond internal enterprise resources to encompass third-party, community, and public data sources, and so it supports Data as a Service (DaaS) across the enterprise supply chain.

Equation 2: Write each cost term

Let:

- p_i= compute price per unit on cloud i
- e_i= egress price per GB on cloud i
- g= GB transferred per compute unit
- l_i= typical latency on cloud i, $l_{\backslash *}$= SLA target latency
- κ= SLA penalty ($/ms/hr)
- q_i= failure probability per hour on cloud i
- R= recovery cost per failure event ($)

Compute:

$$C_{\text{comp}}(t) = \sum_i p_i \ x_i(t)\, W(t$$

Egress:

$$C_{\text{egr}}(t) = \sum_i e_i \ g\, x_i(t)\, W(t)$$

Latency penalty:

$$C_{\text{sla}}(t) = \sum_i \kappa \cdot \max(0, l_i - l_{\backslash *}) \cdot x_i(t)$$

Expected recovery:

$$C_{\text{rec}}(t) = \sum_i q_i \ R\, x_i(t) C_{\text{tot}}(t) = C_{\text{comp}} + C_{\text{egr}} + C_{\text{sla}} + C_{\text{rec}}$$

Because $C_{\text{tot}}$is **linear** in x_i, the minimum is achieved by putting all workload on the cloud with smallest unit cost:

$$u_i(W) = p_i W + e_i g W + \kappa \text{MAX}(0, l_i - l_{\backslash *}) + q_i R$$

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

III. REFERENCE ARCHITECTURES FOR NATIONAL HEALTHCARE INTEROPERABILITY

The architecture provides a multi-cloud data-fabric infrastructure, including a data lake and data repository for a national healthcare information-exchange organization. The emphasis is on making multi-level data from disparate sources available to different consumers while addressing governance, risk, regulatory, and compliance concerns of electronic health records interoperability.

Cloud-based data architectures are an enabler for national healthcare interoperability, considering different use-case domains and architectural offerings. Data sources may consist of sources with open APIs, such as data.gov, and data sources that may require ingestion via an agent/microservice approach. A common data-lake platform can eliminate redundancy and be an on-premises deployment hosting sensitive data that should not be shared in a public cloud, yet still support a multi-cloud architecture. A data-fabric infrastructure enables the creation of a data-repository space that governs sharing across clouds, logically grouping data sources according to regulatory, risk, and governance requirements for the respective consumers.

Data ingestion across clouds should be decoupled from the consumer views. Data accessed from public clouds must comply with levels of data sensitivity and implementing jurisdictional access control over the data sources. Interoperability needs to be considered not just within the country, but also with neighboring countries and global public-health incidents such as the COVID-19 pandemic.

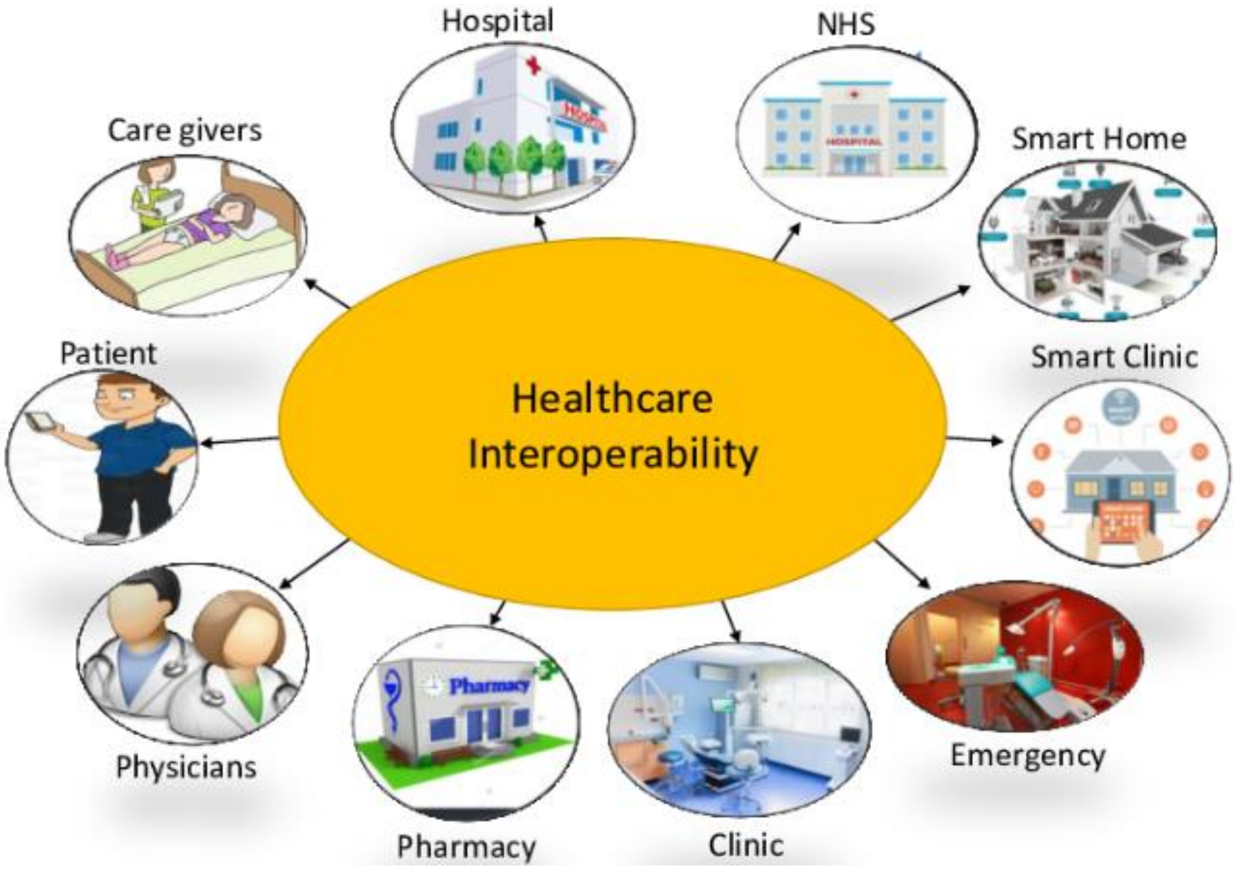


Fig 3: An illustration of Healthcare interoperability

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

3.1. Data sources and ingestion

Data sources are defined by the platform services that enable it. Ingestion services allow data flow in one or more directions between the platform and individual sources such as databases, APIs, message queues, and file shares; push events or change notifications from sources may trigger these operations. External sources managed by third parties do not require dedicated platform locations by default but those operated by the provider, where dedicated services are enabled, take precedence.

Direct access to data sources requires a data access control model that supports both authorization of underlying interactions at source level and distribution of identities across platforms. Onboarding or configuration services assist external source administrators in their initial setup. Although the administrator-level access control needed can be minimal, continuous operations benefit from an additional intermediate-privilege access level. Cloud or cloud-like data sources fully functional for real production workloads justify further services tuning for automatic cloning, scaling, and cleaning processes, data sharding, and schema evolution. The latter ensures compatibility for analytics processes potentially requiring older data states. Given evolution, source read-only status, and supported multi-cloud deployments, data virtualization is a needed additional feature, as it alleviates most long-term source management burdens.

Equation 3: Monitoring + statistical cache offloading model Step 1 — Item popularity distribution Items ranked $r = 1..N$:

$$p(r) = \frac{r^{-a}}{\sum_{k=1}^{N} k^{-a}}$$

Step 2 — Cache hit rate for cache size C

If cache holds top Citems:

$$H(C) = \sum_{r=1}^{C} p(r)$$

Miss rate:

$$M(C) = 1 - H(C)$$

Step 3 — Latency model

If cache latency is $L_{\text{cache}}$and cloud retrieval latency is L_{cloud}:

$$L(C) = H(C)L_{\text{cache}} + M(C)L_{\text{cloud}}$$

Step 4 — Monitoring-based offload rule

Define a "hotness threshold" $q$from monitoring:

- keep items with $p(r) \geq q$
- offload the rest to cloud storage

Let $C_{\text{eff}} = \min(C, \mid \{r: p(r) \geq q\} \mid)$then:

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

$$H_{\text{off}}(C) = \sum_{r=1}^{C_{\text{eff}}} p(r)$$

Step 5 — Choose an optimal cache size

If cache costs $a_{\text{cost}} \cdot C$($/hr) and misses cost $b_{\text{pen}} \cdot M(C)$:

$$J(C) = a_{\text{cost}} C + b_{\text{pen}} M(C)$$

IV. GOVERNANCE, POLICY, AND COMPLIANCE

Cloud data fabrics have untapped potential for supporting healthcare interoperability. Issues such as privacy and trust are a primary concern in the realization of truly national interoperable cloud-based healthcare systems. By establishing a fabric containing patient data, organizations could adhere to principles of access control models based on data owner restrictions and assign guarantees of compliance with national health service regulations by legal authorities.

Identity management follows the distributed multi-cloud nature of the data fabric architecture, offers configuration by each organization of any of its identity management clouds, can be ruled by data-owning organizations, and follows principles of distribution of trust adoption. The foundation of any data fabric is based on the distributed network of clouds and organizations supporting it. Despite the need for online authentication, the trust on data access is established among organizations. The final step needs to ensure user authorization fulfillment when they need access to these resources.

Interoperability solutions based on data fabric technologies have gained incentives recently due to their innovative capacity to respond to complex enterprise data requirements and have obtained some momentum particularly in the healthcare domain. Telecommunication providers have explored interoperability solutions supporting the implied technological requirements, including a trusted environment with data always re-sided locally, control models providing on-going user consent agreements, user management by any service actor, a focus upon the embodied privacy concept of the user, cost associated with data control and consent infrastructures, user control of its consent at any moment, and trust in the consent management service. Cloud data fabric architectures tackling other interoperability paradigms have been explored, without focusing on guaranteed timely complex-event detection.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

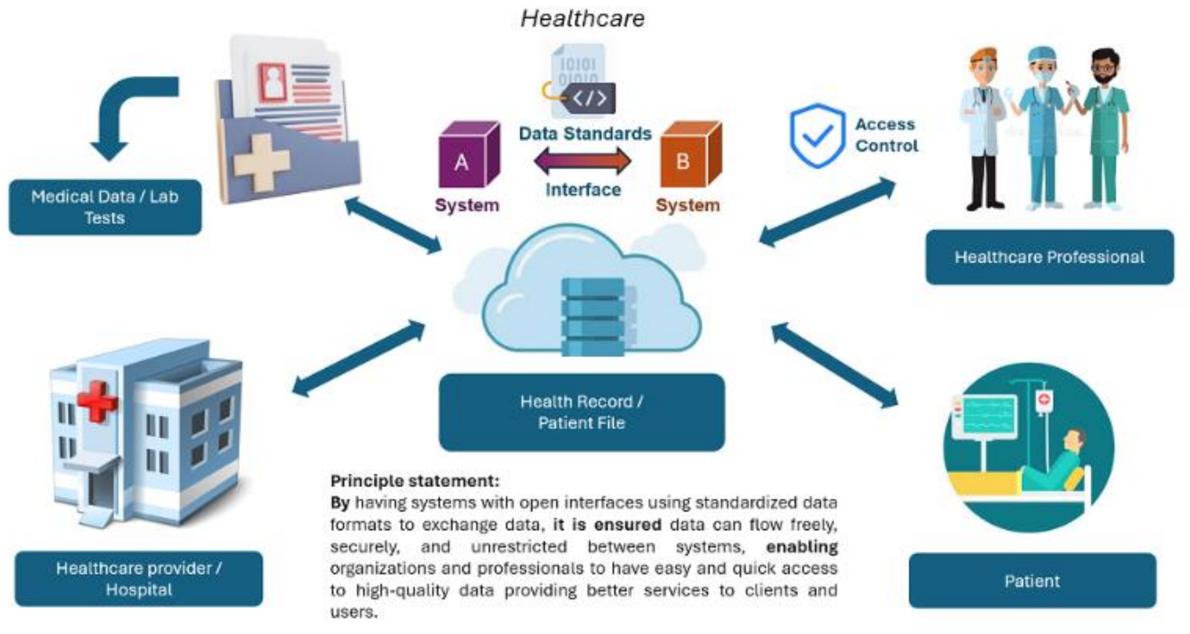


Fig 4: Governance, Policy, and Compliance

4.1. Access control models and identity management

Access control refers to the set of policies determining how health information can be accessed, shared, modified, deleted, and created through the infrastructure. The design of access control models aims to fulfil these policies while allowing secure system operation and maintaining user privacy. Models must also support inter-organizational contexts at varying granularity levels, whether without explicit user consent—e.g., exfiltration of children vaccines in a disaster scenario—or requiring alternative trust mechanisms, e.g., a centralized healthcare authority.

Healthcare Data Fabrics are a federated architecture enabling the computation of data queries across different authentication domains. A central Trusted Third Party (TTP) stores the mapping between identity established by each domain—e.g., through Kerberos and Privacy-ABAC API—and the TTP identifier, enabling federated queries between external identity providers. The system supports fine-grained data-sharing policies without the need for explicit notifications or consent when sharing across domains. While a federated architecture alleviates the need for centralized trust, it becomes difficult to maintain when the number of domains becomes larger.

Identity management encompasses the life cycle of service users and their credentials, covering the creation detection and update of entities, service registration, credential update, and revocation. Privilege delegation aims to prevent knowledge leaks that allow service impersonation, either through granting users the ability to issue temporary proxy credentials or through a designated identity management domain. The latter provides users with proxy credentials indirectly associated with their identity, allowing the design of algorithms that limit information exposure during the delegation process.

Equation 4: Access control + federated identity mapping (TTP)

ABAC decision rule (formalized)

Permit a request if all attribute predicates pass:

$$\mathrm{Permit}(u,r,a) = 1 \iff \forall k: f_k(u,r,a) \geq \theta_k$$

(where predicates can represent role, purpose-of-use, consent, jurisdiction, sensitivity level, etc.)

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

TTP mapping for cross-domain joins

For domain d, local identity $id_d$is mapped by TTP:

$$\text{map}_d(id_d) = ID_{\text{TTP}}$$

Federated query join key:

$$\text{Join}(\text{data}_{d1} \text{ on } \text{map}_{d1}(id_{d1}) = ID_{\text{TTP}}, \text{data}_{d2} \text{ on } \text{map}_{d2}(id_{d2}) = ID_{\text{TTP}})$$

V. INTEROPERABILITY SCENARIOS AND USE CASES

Seven use cases are identified and categorised according to the primary component of the architecture involved: data sources and ingestion, data fabric, and data consumers. The roles of the data fabric and its governance policy are explored in greater detail through two cross-cutting scenarios, one focused on electronic health records interoperability and the other on data access and use. Most use cases could be implemented with existing commercial services and products, albeit with some limitations, while at least two require further development. The scenarios and use cases inform a set of design principles needed to ensure the architectural solution can be successfully built.

Most healthcare data management systems implement only a subset of the components required for a cloud-based data fabric. For those components to interoperate, additional connectors, wrappers, or translation services are usually required. Healthcare governance models often do not have the appropriate cyberinfrastructure support to ensure that data can be easily accessed and shared across disparate business domains. Indeed, constructing a cloud-based data fabric for use within an entire nation requires extending data access policies beyond those imposed for local operational control and enabling data exchange and usage in accordance with an expanding set of use case scenarios.

5.1. Electronic Health Records interoperability

While the exchange of electronic health records (EHRs) has a high priority, interoperability remains technically very difficult due to the many diverse systems involved. While interoperability approaches based on data exchange can solve many needs, they cannot meet all use cases. Further complications arise from identity management and the sensitivity of protected health information. Use Case A considers an alternative approach using electronic health records provided as a virtualized data service rather than exchanged between the systems involved. The need for a patient's information is expressed in a SQL-like language, but data privacy rules are automatically taken into account. The integration of appropriate access control models in cloud-based data fabric platforms enhances the possibility of solving these needs behind-the-scenes.

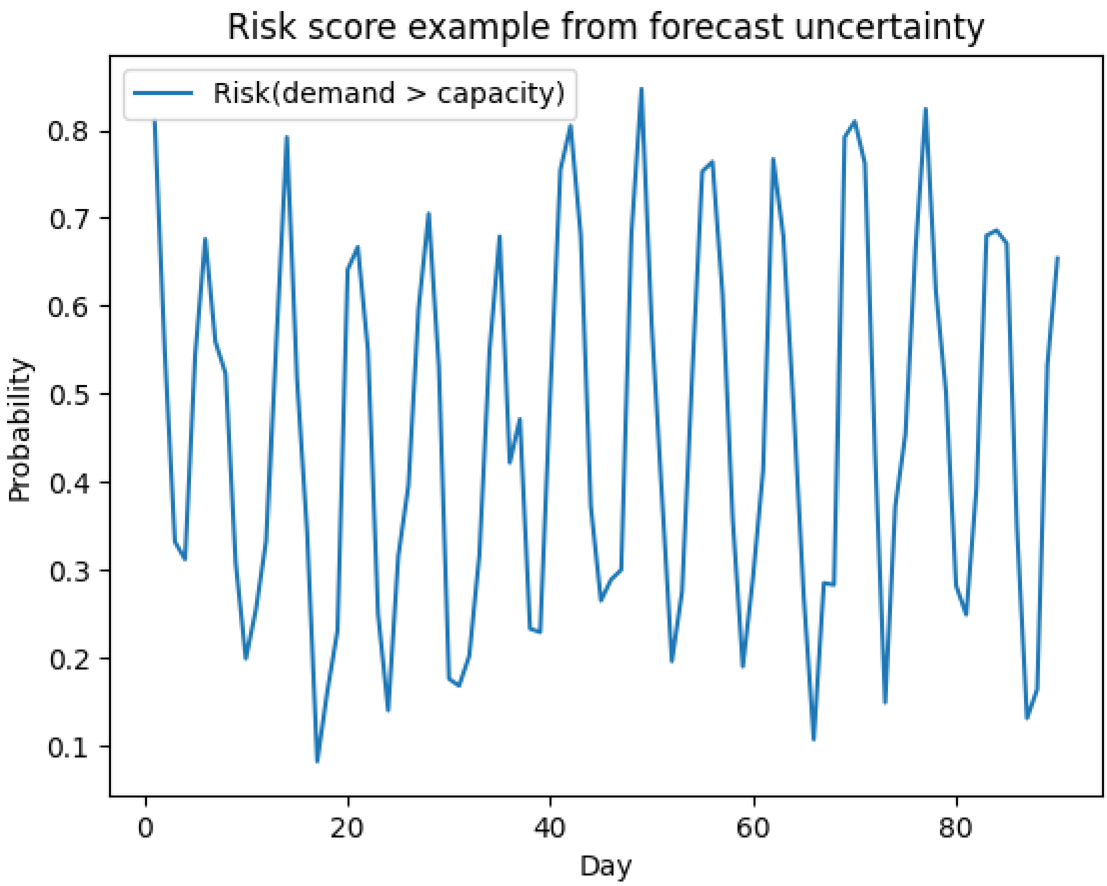
It is well established that the greatest priority in the drive toward nationwide healthcare data interoperability is the ability to share patient health information among healthcare providers, ideally to facilitate patient treatment. Data interoperability, however, presents a multitude of technical challenges. For example, the Integrating the Healthcare Enterprise movement has defined many use-case scenarios for health information exchange; these invariably have adopted an approach based on the underlying exchange of a patient's electronic health record between the participants involved in the specific use case scenario. Moreover, while the use of EHR exchange indeed covers many assimilation needs, especially for commonly used tangible items, these data-exchange techniques are inadequate for other needs, especially where patient information has to be inferred from complementary point-in-time health records among different patients. While the exchange of complete electronic health records (EHRs) has become the dominant model for advancing healthcare data interoperability, this document-centric approach does not fully address the complexity of modern clinical decision-making. In many real-world scenarios, meaningful insights are derived not solely from a

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

single patient's longitudinal record, but from synthesizing complementary, point-in-time data across multiple patients, populations, and care settings. For example, identifying emerging treatment patterns, adverse drug reactions, or early indicators of disease progression often requires comparative analysis and inference beyond the boundaries of an individual EHR. Traditional exchange frameworks, including those promoted through Integrating the Healthcare Enterprise (IHE) use cases, focus primarily on structured record transmission rather than on semantic harmonization, cross-patient analytics, or real-time data aggregation. As a result, while EHR exchange supports fundamental continuity of care, it remains insufficient for advanced interoperability needs such as predictive modeling, population health management, and clinical intelligence generation. Achieving true nationwide interoperability will therefore require not only standardized data transport, but also shared data models, semantic consistency, robust governance frameworks, and analytical infrastructures capable of transforming distributed clinical data into actionable knowledge.



VI. PLATFORM AND SERVICE CONSIDERATIONS

Selecting platforms and services for a national cloud-based data fabric is usually a complex task or not bilevel but operational-level decision-making. Cost, performance, usability, maintainability, and service-level agreements must all be considered, along with technical debts and risks. A harmonized set of public cloud services is seldom chosen. Therefore, functionally similar cloud services must be realized as microservices and deployed in the cloud environment following the multi-cloud cloud pattern. Some cloud functions often cannot be realized or are not cost-effective in public cloud offerings. The hybrid cloud cloud pattern can be applied to determine the on-premise fabric segments and the data-in-cloud or data-in-place cloud-delivery models. Cloud function selected for the data fabric platform should support multiple deployment models (public, private) as well as delivery models (data-in-cloud, data-in-place).

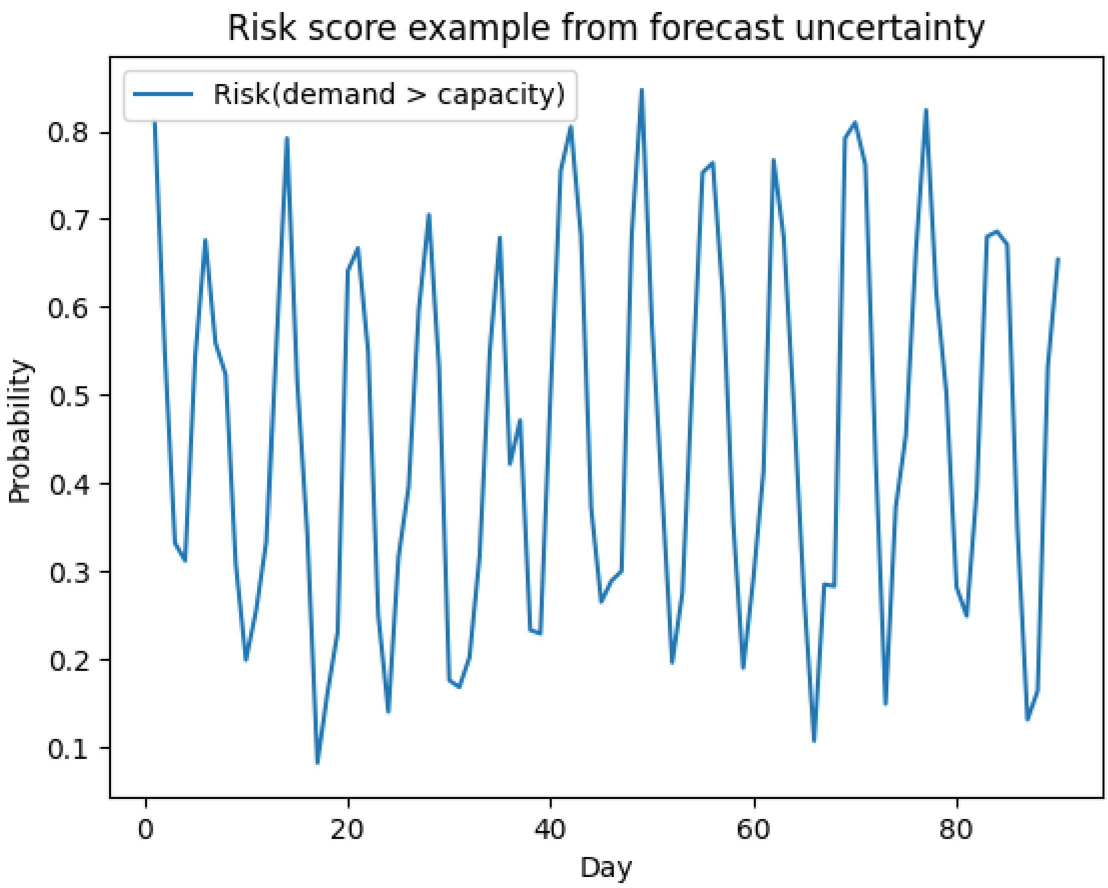
International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

Different cloud offerings can also be required simultaneously for distinct data fabrics in a hybrid or federation scenario. Different compliance requirements might require such solutions for data fabrics at a large transnational destination, yet the data fabric needs to operate as a harmonized service from a service perspective. Therefore, support for a data fabric solution at the cloud service level must be carefully aligned with such requirements. When a data fabric needs to be offered as an integrated and seamless cloud service across GAFAM within a single country, a private national cloud offering has to be pursued with appropriate controls, regulations, and governance around that cloud serving a single country.



6.1. Multi-cloud and hybrid deployments

Digital business environments en masse are evolving from dependent relationships with a single cloud to interconnected multi-cloud or poly-cloud relationship patterns. In either context, the data fabric principle becomes essential to seamless and secure interoperability. Hybrid clouds blend on-premises systems and public clouds to fulfil regulatory or security requirements. In federated authentication schemes, a distributed architecture with services located in diverse clouds assuages such concerns by deploying the function elsewhere and retaining only audited interfaces in sensitive premises.

Such architectural patterns enhance security resilience by isolating sensitives systems and data; loading authentication and access control services that apply survivability and restoration rules, and instinctively degrading capabilities under attack; and distributing residual, supportive systems and data across clouds, with virtualized control and real-time consolidation. The ability of heterogeneous clouds to evolve independently and respond to local business demands further strengthens resilience. These patterns support ad hoc utilities that transcend the multi-cloud safeguard in the contrast between necessity and choice by demand-based, architectural demand-supply envelopes.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

VII. CONCLUSION

Computing, storage, and bandwidth capabilities are no longer constraining factors, even for large-scale data collection and processing. Citizens' privacy must be scrupulously respected, but legislative measures protecting confidential data are generally available. Data has become the most valuable of all assets; the knowledge of how to use it is equally important. The missing piece is a platform to allow the various data holders to share, analyze, and exploit data at national level while maintaining complete control over their data assets.

Cloud-based data fabric architectures can provide all the building blocks needed for fulfilment of the above requirements. National healthcare interoperability scenarios and use cases demonstrate the ability to virtualize and federate services, abstract service accessibility through layer components that encapsulate data quality, constraints, and security policies, and enable cloud-based development of new services based on existing ones. National healthcare authorities are well positioned to take advantage of cloud-based data fabric architectures, which can grant them the interoperability required for research, development, and innovation while driving the economy and improving citizens' well-being. These architectures are cloud-agnostic and thus equally applicable to other sectors, such as energy, finance, and transport. More cloud middleware in the form of services would boost their development and pave the way to a more mature digital economy.

Cache_items	Hit_rate_plain	Hit_rate_with_offload
100	0.5256832631983532	0.5256832631983532
300	0.6203876543461493	0.6203876543461493
1000	0.7185846926294566	0.7185846926294566
3000	0.8032334414509303	0.8032334414509303
10000	0.8908322757213876	0.8908322757213876
20000	0.9389265661042377	0.9389265661042377
40000	0.9853834698081831	0.9853834698081831

Table : Cache hit rate table

7.1. Future Directions

Future research will focus on data flow orchestration, exposing cloud application-level and infrastructure-level services for access, and extending Cloud Data Fabric functionalities.

Insight into cloud and data fabric service deployments will then help align cloud services across the national healthcare system to support the Data-as-a-Platform model of the Data Sharing Consortium. Matching the growing adoption of electronic services in all economic sectors, the strategy aims to create a historical and permanent Data Space, generating secure and high-quality open datasets as well as living API interfaces for different sources within the healthcare ecosystem. Eventually, the Data Sharing Consortium’s mission is to promote the establishment of a complete ecosystem, offering technological, methodological, and financial support to the entire Italian healthcare service, in line with relevant national policies.

Real-world deployment will test the proposed reference architecture, which reflects a specific cloud-based National Health Service Data-as-a-Platform project. The Data Sharing Consortium includes the Ministry of Health and the main healthcare system providers, who form the National Health Service of the Italian Republic. Data-as-a-Platform aims to

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

favour the the creation of Data Spaces at the national level, in collaboration with the European Commission and other European Member States.

REFERENCES

[1] Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. Nature Reviews Microbiology, 13(1), 42–51. https://doi.org/10.1038/nrmicro3399.

[2] Davuluri, P. N. (2020). Improving Data Quality and Lineage in Regulated Financial Data Platforms. Finance and Economics, 1(1), 1-14.

[3] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734–749.

[4] Kothapalli Sondinti, L. R., & Syed, S. (2022). The Impact of Instant Credit Card Issuance and Personalized Financial Solutions on Enhancing Customer Experience in the Digital Banking Era. Universal Journal of Finance and Economics, 1(1), 1223. Retrieved from https://www.scipublications.com/journal/index.php/ujfe/article/view/1223

[5] Arasu, A., & Kaushik, R. (2014). Data cleansing: A context dependent approach. Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, 135–146.

[6] Vadisetty, R., Polamarasetti, A., Guntupalli, R., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2021). Privacy-Preserving Gen AI in Multi-Tenant Cloud Environments. Sateesh kumar and Raghunath, Vedaprada and Jyothi, Vinaya Kumar and Kudithipudi, Karthik, Privacy-Preserving Gen AI in Multi-Tenant Cloud Environments (January 20, 2021).

[7] Dai, J., Xu, H., Chen, T., Huang, T., Liang, W., et al. Artificial intelligence for medicine Navigating the endless frntier. The Innovation Medicine, 3, 100120.

[8] Gottimukkala, V. R. R. (2021). Digital Signal Processing Challenges in Financial Messaging Systems: Case Studies in High-Volume SWIFT Flows.

[9] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50–58.

[10] Chava, K., Chakilam, C., & Recharla, M. (2021). Machine Learning Models for Early Disease Detection: A Big Data Approach to Personalized Healthcare. International Journal of Engineering and Computer Science, 10(12), 25709–25730. https://doi.org/10.18535/ijecs.v10i12.4678

[11] Babcock, J., Chaudhuri, S., & Das, G. (2004). Dynamic sample selection for approximate query processing. Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, 539–550.

[12] Sriram, H. K. (2022). Advancements in Credit Score Analytics using Deep Learning and Predictive Modeling Techniques. Available at SSRN 5255128.

[13] Bifet, A., & Gavaldà, R. (2007). Learning from time-changing data with adaptive windowing. Proceedings of the 2007 SIAM International Conference on Data Mining, 443–448.

[14] Muthusamy, S., Kannan, S., Lee, M., Sanjairaj, V., Lu, W. F., Fuh, J. Y., ... & Cao, T. (2021). Cover Image, Volume 118, Number 8, August 2021. Biotechnology and Bioengineering, 118(8), i-i.

[15] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[16] Aitha, A. R. (2022). Cloud Native ETL Pipelines for Real Time Claims Processing in Large Scale Insurers. Available at SSRN 5532601.

[17] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171–209.

[18] Segireddy, A. R. (2021). Containerization and Microservices in Payment Systems: A Study of Kubernetes and Docker in Financial Applications. Universal Journal of Business and Management, 1(1), 1–17.

[19] Gangula, R. (2021). Leveraging the Hyperledger Fabric for enhancing the efficacy of clinical decision support systems. Blockchain in Healthcare Today.

[20] Davuluri, P. S. L. N. (2021). Event-Driven Compliance Systems: Modernizing Financial Crime Detection Without Machine Intelligence. Journal of International Crisis and Risk Communication Research , 339–354. https://doi.org/10.63278/jicrcr.vi.3636

[21] Das, T., Zhu, A., Li, S., Narayanamurthy, S., & Bhat, P. (2013). Distributed and fault-tolerant streaming computation in Spark. Proceedings of the ACM Symposium on Cloud Computing, 1–12.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

[22] Siva Hemanth Kolla. (2022). Knowledge Retrieval Systems for Enterprise Service Environments. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 495–506. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8037

[23] Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. Communications of the ACM, 51(1), 107–113.

[24] Paleti, S. (2022). Financial Innovation through AI and Data Engineering: Rethinking Risk and Compliance in the Banking Industry. Available at SSRN 5250726.

[25] DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Vosshall, P., & Vogels, W. (2007). Dynamo: Amazon's highly available key-value store. Proceedings of the 21st ACM Symposium on Operating Systems Principles, 205–220.

[26] Sriram, H. K., ADUSUPALLI, B., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.

[27] Dwork, C. (2008). Differential privacy: A survey of results. Proceedings of the 5th International Conference on Theory and Applications of Models of Computation, 1–19.

[28] Varri, D. B. S. (2021). Cloud-Native Security Architecture for Hybrid Healthcare Infrastructure. Available at SSRN 5785982.

[29] Elmagarmid, A. K., Ipeirotis, P. G., & Verykios, V. S. (2007). Duplicate record detection: A survey. IEEE Transactions on Knowledge and Data Engineering, 19(1), 1–16.

[30] Dwaraka Nath Kummari,. (2022). Machine Learning Approaches to Real-Time Quality Control in Automotive Assembly Lines. Mathematical Statistician and Engineering Applications, 71(4), 16801–16820. Retrieved from https://philstat.org/index.php/MSEA/article/view/2972

[31] Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). "Counting your customers" the easy way: An alternative to the Pareto/NBD model. Marketing Science, 24(2), 275–284.

[32] Inala, R. (2022). Engineering Data Products for Investment Analytics: The Role of Product Master Data and Scalable Big Data Solutions. International Journal of Scientific Research and Modern Technology, 155-171.

[33] Davuluri, P. N. (2020). Improving Data Quality and Lineage in Regulated Financial Data Platforms. Finance and Economics, 1(1), 1-14.

[34] Meda, R. Enabling Sustainable Manufacturing Through AI-Optimized Supply Chains.

[35] Ghemawat, S., Gobioff, H., & Leung, S. T. (2003). The Google file system. Proceedings of the 19th ACM Symposium on Operating Systems Principles, 29–43.

[36] Varri, D. B. S. (2022). A Framework for Cloud-Integrated Database Hardening in Hybrid AWS-Azure Environments: Security Posture Automation Through Wiz-Driven Insights. International Journal of Scientific Research and Modern Technology, 1(12), 216-226.

[37] Yandamuri, U. S. (2021). A Comparative Study of Traditional Reporting Systems versus Real-Time Analytics Dashboards in Enterprise Operations. Universal Journal of Business and Management, 1(1), 1–13. Retrieved from https://www.scipublications.com/journal/index.php/ujbm/article/view/1357

[38] Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. International Journal of Scientific Research and Modern Technology, 1(12), 177-186.

[39] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

[40] Vadisetty, R., Polamarasetti, A., Guntupalli, R., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2022). AI-Driven Cybersecurity: Enhancing Cloud Security with Machine Learning and AI Agents. Sateesh kumar and Raghunath, Vedaprada and Jyothi, Vinaya Kumar and Kudithipudi, Karthik, AI-Driven Cybersecurity: Enhancing Cloud Security with Machine Learning and AI Agents (February 07, 2022).

[41] Hellerstein, J. M., Haas, P. J., & Wang, H. J. (1997). Online aggregation. Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data, 171–182.

[42] Garapati, R. S. (2022). Web-Centric Cloud Framework for Real-Time Monitoring and Risk Prediction in Clinical Trials Using Machine Learning. Current Research in Public Health, 2, 1346.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

[43] Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. Proceedings of the 2008 IEEE International Conference on Data Mining, 263–272.
[44] Amistapuram, K. (2022). Fraud Detection and Risk Modeling in Insurance: Early Adoption of Machine Learning in Claims Processing. Available at SSRN 5741982.
[45] Meda, R. (2021). Digital Infrastructure for Predictive Inventory Management in Retail Using Machine Learning. International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), DOI, 10.
[46] Meda, R. (2022). Integrating Edge AI in Smart Factories: A Case Study from the Paint Manufacturing Industry. International Journal of Science and Research (IJSR), 1473-1489.
[47] Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. Communications of the ACM, 57(7), 86–94.
[48] Segireddy, A. R. (2020). Cloud Migration Strategies for High-Volume Financial Messaging Systems.
[49] Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148–152.
[50] Amistapuram, K. (2021). Digital Transformation in Insurance: Migrating Enterprise Policy Systems to .NET Core. Universal Journal of Computer Sciences and Communications, 1(1), 1–17.
[51] Kleppmann, M. (2017). Designing data-intensive applications. O'Reilly Media.
[52] Nagabhyru, K. C. (2022). Bridging Traditional ETL Pipelines with AI Enhanced Data Workflows: Foundations of Intelligent Automation in Data Engineering. Available at SSRN 5505199.
[53] Lahiri, M., & Venkatasubramanian, S. (2013). Robust record linkage. Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, 101–112.
[54] Avinash Reddy Aitha. (2022). Deep Neural Networks for Property Risk Prediction Leveraging Aerial and Satellite Imaging. International Journal of Communication Networks and Information Security (IJCNIS), 14(3), 1308–1318. Retrieved from https://www.ijcnis.org/index.php/ijcnis/article/view/8609
[55] Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). Mining of massive datasets (2nd ed.). Cambridge University Press.
[56] Rongali, S. K. (2022). AI-Driven Automation in Healthcare Claims and EHR Processing Using MuleSoft and Machine Learning Pipelines. Available at SSRN 5763022.
[57] Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1), 76–80.
[58] Keerthi Amistapuram , "Energy-Efficient System Design for High-Volume Insurance Applications in Cloud-Native Environments," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), DOI 10.17148/IJIREEICE.2020.81209
[59] Lin, J., Kolcz, A., & Szymanski, B. K. (2012). Large-scale machine learning at Twitter. Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, 793–804.
[60] Sheelam, G. K. Power-Efficient Semiconductors for AI at the Edge: Enabling Scalable Intelligence in Wireless Systems. International Journal of Innovative Research in Electrical, Elec-tronics, Instrumentation and Control Engineering (IJIREEICE), DOI, 10.
[61] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
[62] Vadisetty, R., Polamarasetti, A., Guntupalli, R., Rongali, S. K., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2021). Legal and Ethical Considerations for Hosting GenAI on the Cloud. International Journal of AI, BigData, Computational and Management Studies, 2(2), 28-34.
[63] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. Proceedings of the International Conference on Learning Representations, 1–12.
[64] Ramesh Inala. (2022). Cross-Domain MDM Integration Using AI-Driven Data Governance: A Case Study In Financial Technology Architecture. Migration Letters, 19(2), 280–304. Retrieved from https://migrationletters.com/index.php/ml/article/view/11982
[65] Montoya, D. Y., Neto, A. M., & da Silva, A. S. (2016). A survey of entity resolution in big data. Journal of Big Data, 3(1), 1–22.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|
||Volume 5, Issue 6, November - December 2022||
DOI:10.15662/IJRPETM.2022.0506019
[66] Aitha, A. R. (2021). Optimizing Data Warehousing for Large Scale Policy Management Using Advanced ETL Frameworks.
[67] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster computing with working sets. Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing, 1–7.
[68] Varri, D. B. S. (2022). AI-Driven Risk Assessment and Compliance Automation in Multi-Cloud Environments. Available at SSRN 5774924.
[69] Zaharia, M., Das, T., Li, H., Shenker, S., & Stoica, I. (2012). Discretized streams: Fault-tolerant streaming computation at scale. Proceedings of the 24th ACM Symposium on Operating Systems Principles, 423–438.
[70] Rongali, S. K. (2021). Cloud-Native API-Led Integration Using MuleSoft and .NET for Scalable Healthcare Interoperability. Available at SSRN 5814563.
[71] Zhai, C., & Massung, S. (2016). Text data management and analysis: A practical introduction to information retrieval and text mining. ACM & Morgan Claypool.
[72] Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. Online Journal of Engineering Sciences, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/ojes/article/view/1360
[73] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135–146.
[74] Hasselgren, A. (2022). Decentralized identity management for e-health applications: State-of-the-art and guidance for future work. Blockchain in Healthcare Today.
[75] Goutham Kumar Sheelam. (2022). Reconfigurable Semiconductor Architectures For AI-Enhanced Wireless Communication Networks. Kurdish Studies, 10(2), 1027–1040. https://doi.org/10.53555/ks.v10i2.3867
[76] Batarseh, F. A., & Yang, R. (2019). Federal data science: Transforming government and society. Academic Press.
[77] Dwaraka Nath Kummari. (2022). Fiscal Policy Simulation Using AI And Big Data: Improving Government Financial Planning. Kurdish Studies, 10(2), 934–945. https://doi.org/10.53555/ks.v10i2.3855
[78] Bhasin, H., & Bhatia, P. (2020). Clickstream data mining for web analytics and customer behavior modeling: A review. ACM Computing Surveys, 53(6), 1–34.
[79] Goutham Kumar Sheelam, "Semiconductor Innovation for Edge AI: Enabling Ultra-Low Latency in Next-Gen Wireless Networks," International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), DOI: 10.17148/IJARCCE.2022.111258
[80] Davuluri, P. N. (2020). Event-Driven Architectures for Real-Time Regulatory Monitoring in Global Banking.
[81] Abedjan, Z., Golab, L., & Naumann, F. (2016). Profiling relational data: A survey. The VLDB Journal, 24(4), 557–581.
[82] Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 1(12), 227–237. https://doi.org/10.38124/ijsrmt.v1i12.1111
[83] Dwaraka Nath Kummari. (2022). AI-Driven Audit Frameworks For Enhancing Compliance In Modern Manufacturing Systems. Migration Letters, 19(S8), 2150–2177. Retrieved from https://migrationletters.com/index.php/ml/article/view/11912
[84] Davuluri, P. N. Event-Driven Compliance Systems: Modernizing Financial Crime Detection Without Machine Intelligence.
[85] Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2021). Fraud analytics using descriptive, predictive, and social network techniques: A guide to data science for fraud detection (2nd ed.). Wiley.
[86] Uday Surendra Yandamuri. (2022). Cloud-Based Data Integration Architectures for Scalable Enterprise Analytics. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 472–483. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/8005
[87] Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. fairmlbook.org (Book manuscript).
[89] Garapati, R. S. (2022). AI-Augmented Virtual Health Assistant: A Web-Based Solution for Personalized Medication Management and Patient Engagement. Available at SSRN 5639650.

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 5, Issue 6, November - December 2022||

DOI:10.15662/IJRPETM.2022.0506019

[90] Gottimukkala, V. R. R. (2020). Energy-Efficient Design Patterns for Large-Scale Banking Applications Deployed on AWS Cloud. power, 9(12).

[91] Ahmad, M. A., Eckert, C., & Teredesai, A. (2018). Interpretable machine learning in healthcare. Proceedings of the ACM Conference on Health, Informatics, and Data Science, 1–10.

[92] Inala, R. Advancing Group Insurance Solutions Through Ai-Enhanced Technology Architectures And Big Data Insights.

[93] Avinash Reddy Segireddy. (2022). Terraform and Ansible in Building Resilient Cloud-Native Payment Architectures. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 444–455. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/7905

[94] Kolla, S. K. (2021). Architectural Frameworks for Large-Scale Electronic Health Record Data Platforms. Current Research in Public Health, 1(1), 1–19. Retrieved from https://www.scipublications.com/journal/index.php/crph/article/view/1372

[94] Rongali, S. K. (2020). Predictive Modeling and Machine Learning Frameworks for Early Disease Detection in Healthcare Data Systems. Current Research in Public Health, 1(1), 1-15.

[95] Bani-Salameh, N., & Olariu, C. (2019). Efficient healthcare big data processing using cloud computing. Journal of Parallel and Distributed Computing, 132, 47–55.