International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

Scalable Cloud Storage Models for National Digital Identity Systems

Mallesham Goli

Independent Researcher, India

mallesham.goli.research01@gmail.com

ABSTRACT: Cloud-based identity services are evolving to meet the needs of National Digital Identity systems, which function as the main trust anchors underpinning key government services, including Social Security, taxation, and healthcare. To bridge the gap between existing data-protection regulations and the design of federated National Digital Identity systems, specific attributes need to be articulated—these should be privacy-protective, allow for custom content, include preventive and remedial compliance measures, and make auditing obligatory. A trusted credential repository serves as the basis for a Digital Identity platform-as-a-service provider that integrates supporting identity-management-as-a-service solutions.

The Digital Identity Infrastructure must include suitable data storage models allowing for protected privacy, security, integrity, and reliability, together with defined incident-reporting and auditing procedures. The models should encompass architecture choices supporting operation and resilience, including availability, regulatory obligations, jurisdiction, data localization, and mitigation of single-point-of-failure issues. Two cloud-storage models for national identity services are examined—centralized storage managed by a single entity and federated storage operated by multiple identity providers—as well as a hybrid combination.

KEYWORDS: National Digital Identity Systems, Cloud-Based Identity Services, Federated Identity Architecture, Digital Identity Platform-as-a-Service, Identity-Management-as-a-Service (IDaaS), Trusted Credential Repositories, Privacy-Preserving Identity Design, Data Protection and Regulatory Compliance, Federated versus Centralized Storage Models, Data Localization and Jurisdictional Governance, Identity Infrastructure Resilience, Security and Integrity in Digital Identity, Incident Reporting and Auditability, Single-Point-of-Failure Mitigation, Government Digital Trust Frameworks.

I. INTRODUCTION

The growing adoption of digital identities offers considerable economic and social value. However, the worldwide proliferation of Privacy-Enhancing and Privacy-By-Design® technologies supporting these emerging digital identities underscores the established need for trustworthy identity data management. Cryptographic technology can certainly establish the identity behind an identity verifier but not necessarily the identity data itself. Therefore, reliable Cloud Storage is essential for national digital identity systems, and Sepura, a proven application ecosystem specifically for identity systems, delivers such Cloud Storage. In tandem, Sepura delivers a Governance Group that ensures local lawfulness, international compliance and risk management best practices.

Can national digital identity systems store identity data in a trustworthy manner? The answer determines which Cloud Storage model constitutes a benchmark for these identity systems. A Centralized Cloud Storage model stores all identity data in a single Cloud Server under an entity's control. Augmented by an integrated Risk Assessment process meeting established standards, this model supports the economies and analytics of centralized data mining. A Federated Cloud Storage model stores identity data in multiple independently operated Cloud Servers, each resident within the jurisdiction of the individual concern supplying the data: these Cloud Servers constitute the Identity Data Sovereign's Cloud Storage. When implemented with full-privacy principles, such systems ensure trustworthy data residency, jurisdictional compliance and governance privacy-exposure minimization.

1.1. Overview of the Study

What are the implications for a country's national digital identity system cloud storage model if it is to support a secure migration to electronic residency? From country to country, considerations differ for the long-term storage and distribution of digital identities. During the design phase for the national digital identity system, some countries use a

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

fully on-premises, fully centralized model, while others prefer a hybrid one that combines trusted third-party nodes with a multi-cloud strategy. These decisions must take into account the countries' specific prerequisites, e-residency service infrastructure, storage implementation, identity assurance level and the ability to cooperate with neighboring countries. If these elements are well designed, they can support cross-border data flow and the nondiscriminatory access and operation of the identity service, which is often seen as a government monopoly.

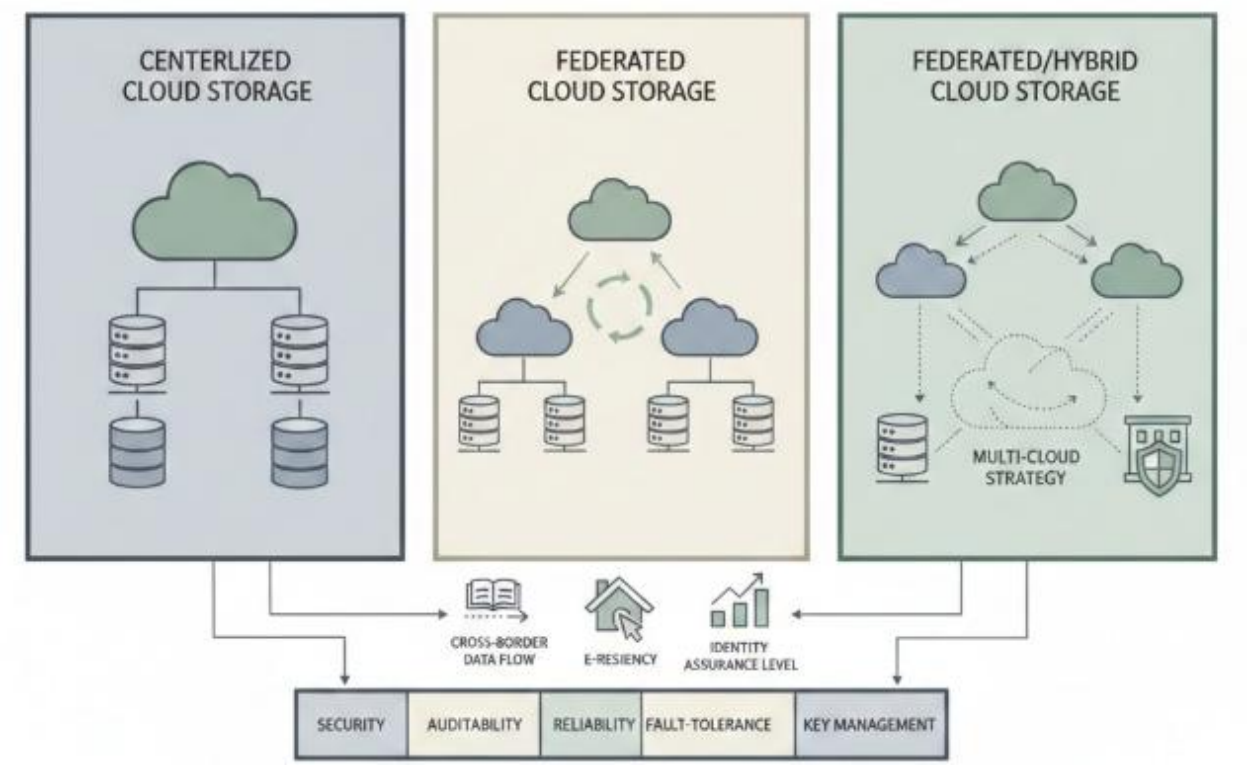


Fig 1: Architecting Sovereign Identity: A Comparative Analysis of Cloud Storage Models for Secure E-Residency Migration and Cross-Border Interoperability

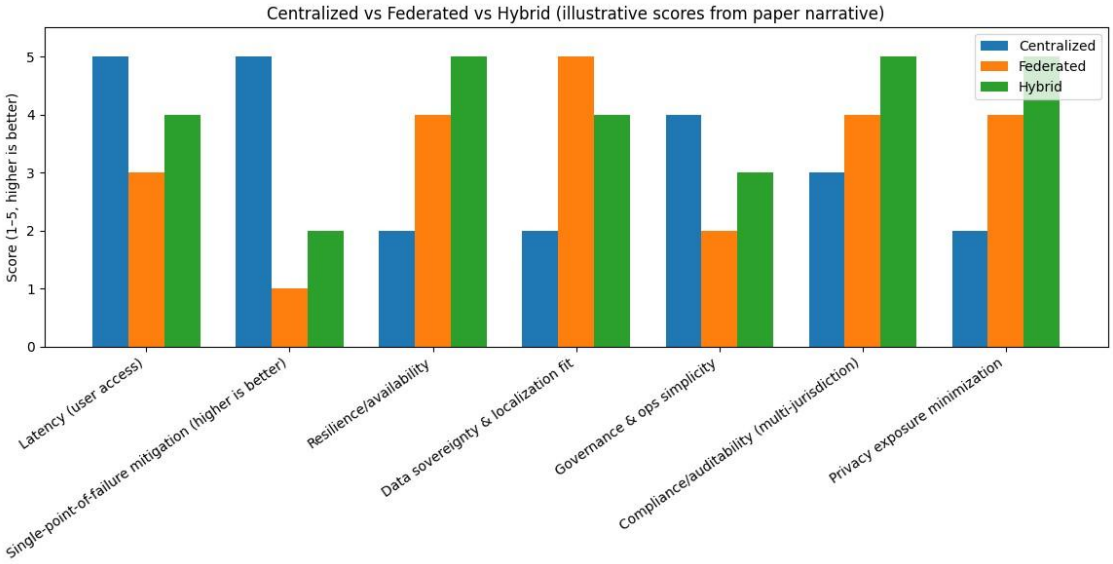
Therefore, suitable cloud storage models for a national digital identity system's long-term implementation and production are examined and analyzed on the basis of openness, privacy, security, reliability and auditability. Three types of cloud storage architecture are distinguished: a centralized cloud storage model, a federated cloud storage model and a federated/hybrid cloud storage model. Four aspects must be taken into account: the permanent storage of identity evidence; the ability to manage and distribute keys; the need for a fault-tolerant architecture that can ensure identity evidence integrity and identity system availability; and the ability to carry out audits and make use of third-party audits.

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004



II. BACKGROUND AND TERMINOLOGY

National Digital Identity Systems (NDIS) are primarily deployed to provide individuals with a digital form of identity that can be used for authentication and verification while accessing services over the Internet. In many countries, such systems are used for proof of life verification during pension disbursements, provision of banking services, verification of age for government services, financial assistance during natural calamities, and digital vaccination certificates. However, researchers and law enforcement services are also looking at these databases as a potential tool for surveilling individuals for a large number of applications that infringe upon privacy. Hence, employing a privacy-by-design approach while ensuring compliance with existing laws of the country and governing the NDIS in a manner acceptable to all stakeholders is paramount.

To ensure that the NDIS has a model for storage that can be scaled and that can be implemented, operated, maintained, and governed by multiple entities, comparison of the Centralized, Federated, and Hybrid cloud storage models with respect to parameters such as data availability, privacy, compliance, and governance has been conducted, concluding that NDIS data can be hosted either in a Federated or Hybrid model. Scalable infrastructural suggestions are provided for Federated or Hybrid models, along with a description of models that provide storage and services from a single entity – Infrastructure as a Service, Platform as a Service, Software as a Service – and the impact on the NDIS implementation and operation of the decision to host it in a single entity or in a multi-entity cloud."

Equation 1) Data integrity using hashing + digital signatures (PKI)

Step 1 — Hash the data

Let the identity record (or evidence) be a byte string m. Choose a cryptographic hash function $H(\cdot)$ (e.g., SHA-256).

$$h = H(m)$$

- h is a fixed-length digest (fingerprint) of m.

Step 2 — Sign the hash (private key)

Let Sign$(\cdot)$ be a digital signature algorithm (RSA/ECDSA), with keypair (sk, pk).

$$s = \text{Sign}(sk, h)$$

Now store/transmit (m, s) or store m and keep s alongside it (as the paper describes).

Step 3 — Verify later (public key)

When a verifier receives m and s:

1. Recompute the hash:

$$h' = H(m)$$

2. Verify signature:

$$\text{Verify}(pk, h', s) \stackrel{?}{=} \text{true}$$

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

- If true, then (with high confidence) m **was not modified** and the signer possessed sk.

Step 4 — Tamper detection logic

If the stored data becomes m^* (tampered), then:

$$H(m^*) \neq H(m) \Rightarrow \text{Verify}(pk, H(m^*), s) = \text{false}$$

2.1. Key Concepts and Definitions

Digital Identity

A digital identity comprises the information and attributes associated with an individual or entity, which can be used to establish and verify the person's or entity's identity in the digital space. Identity assurance levels characterize the degree of confidence in the identity validation process. Higher levels provide greater assurance and therefore allow access to more sensitive data and systems. The various identity assurance levels mapped for different identity transactions according to the identity assurance levels framework of India, which is one of the commonly referenced frameworks, are listed. Data sovereignty refers to the concept that data is subject to the laws and governance structures in the nation where it is collected. In the cloud storage context, the meaning of data sovereignty and the implications for data localization and cross-border data flows are also covered.

Interoperability pertains to the technical and nontechnical arrangements that enable entities doing business or conducting transactions with each other and sharing data to do so with minimal friction. The various interoperability arrangements relevant to national digital identity systems include data format standards, application programming interface (API) standards, authentication standards, and the governance of these arrangements. Unified national standards, when established for such arrangements, minimize the cost of connection for all stakeholders involved. Reliability in the context of a national digital identity system operates at two levels: the reliability of the underlying cloud infrastructure where identity attribute data is hosted, and the reliability of the identity data itself. All reliability metrics associated with the cloud infrastructure are also relevant, including availability, fault tolerance, disaster recovery, and survivability. Reliability in terms of identity attribute data centers around ensuring data integrity so that only authentic and accurate identity attributes about individuals are held in the system; and preventing data loss or unavailability for any period, thereby combining aspects of integrity and availability. Within the realm of data reliability, the various technical and operational controls and tools implemented in the longitudinal and horizontal dimensions (over time and across the distributed system) of data attribute integrity and availability are relevant.

Criterion	Centralized	Federated	Hybrid
Latency (user access)	5	3	4
Single-point-of-failure mitigation (higher is better)	5	1	2
Resilience/availability	2	4	5
Data sovereignty & localization fit	2	5	4

III. REQUIREMENTS OF NATIONAL DIGITAL IDENTITY SYSTEMS

The requirements of national digital identity systems can be understood through eight questions, describing aspects including system security, scalability, availability, privacy, and governance. Six of these questions are rooted in FIPPs and the principle of minimality, while the remaining two address data availability and integrity, key factors in sustaining citizen trust. Asked separately, they clarify the role of identity data and reflect on requirements for data hosting mechanisms, contributing to the establishment of evaluation criteria.

1. Are privacy-sensitive attributes stored in the national digital identity solution?
2. Is the personal data minimization principle applied at all stages of the life cycle of the national digital identity solution?
3. Is the risk of unauthorized disclosure of identity data managed and mitigated at all times?
4. Is access to identity data only granted to entities that require it in order to fulfil their responsibilities?
5. Are the provisions of national and international privacy, data protection, supervision, and audit regulations respected?
6. Do mechanisms ensure that identity data remain correct and trustworthy even in the face of failure or attack?
7. Is the personal data availability attribute protected at all stages of the life cycle of the national digital identity solution?

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

8. Are mechanisms in place for effective, timely identification, management, and notification of any incident affecting the national digital identity solution?

The first five of these questions relate to FIPPs and the concept of privacy-by-design; the last three address important trust drivers for citizens engaging with a national digital identity solution. Identity data location and availability are both determining factors in hosting architecture, whether centralized, federated, or hybrid.

3.1. Data Integrity and Availability

Maintaining data integrity and availability, key requirements of any reliable cloud storage model, is of particular concern in the context of a national digital identity system. Mechanisms are needed that provide assurance that user data have not been tampered with or deliberately altered by an unauthorized party (intentional data corruption) and that data stored in a cloud service are accessible as expected and can be used by authorized parties (data availability). Given the volume of data typically stored in cloud environments, the architecture should also ensure that the risk model provides transparency, robustness, and accountability.

The assumption of data integrity within cloud storage models is commonly achieved through the use of hashing techniques combined with digital signatures, providing assurance that data have not been modified during storage or transit. For data integrity, some cloud vendors apply checksums to all data-at-rest, while others use secret-shared hashing over all blocks of data. The framework should also provide tamper-resistant data. A signature created by the private key of the PKI is attached to the user data stored in the cloud repository, and maintenance of the signature is undertaken by the owner. Multiple copies of the user-generated hashed code are stored and periodically checked for tampering. Data integrity is also assured by the replication factor and placement of the data of the cloud service.

Data availability is supported through multiple replication of user data, usually in different geographical zones. Control data are created based on the user data and also replicated following a similar replication factor policy. Typically, the consistency model applied aims to achieve strong consistency for authentication and access control and eventual consistency for user data. To ensure disaster recovery, backup must be performed sufficiently frequently to minimize loss of service or data (targeted by RPO) and, where possible, a failover strategy should be in place that directs consumers to use a different service for all needs for as long as that is required, with service returned once the primary has been restored (targeted by RTO). Given the volume of data, testing of the DR management plan should be undertaken periodically to prove that the actions are effective.

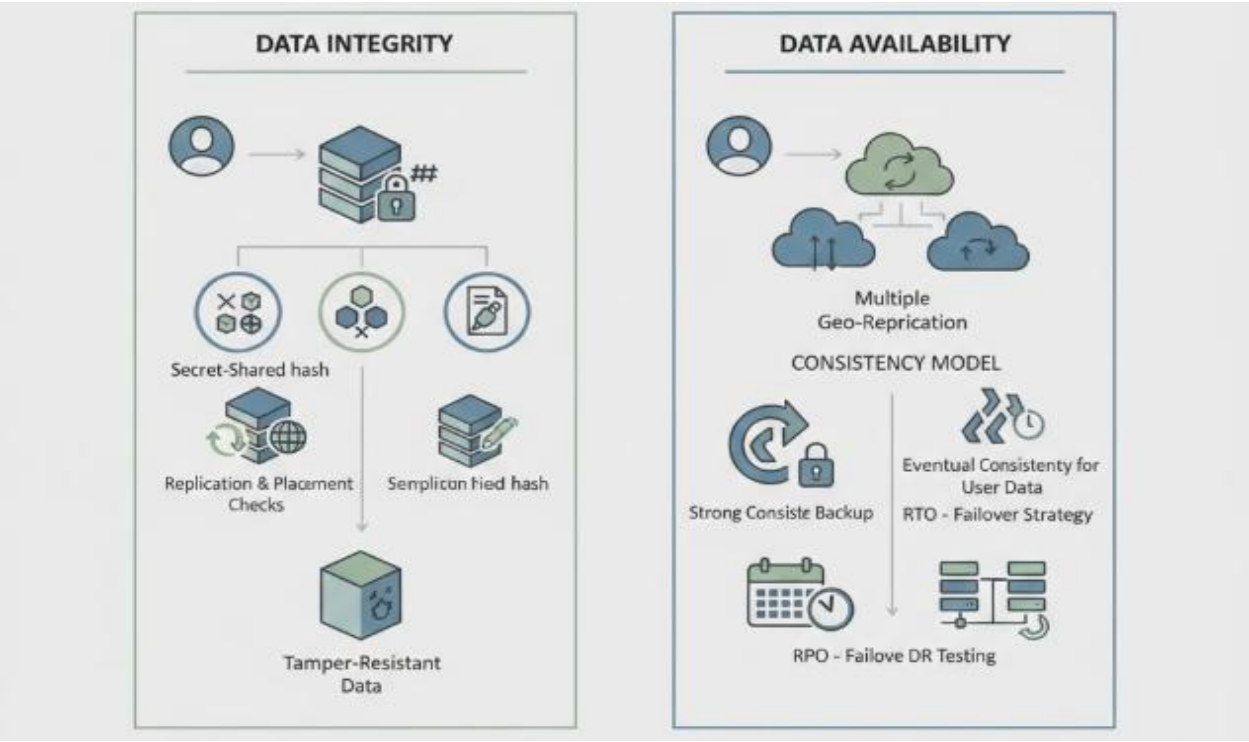


Fig 2: Architecting Trust: Integrity and Availability Frameworks for National Digital Identity Systems

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

3.2. Privacy, Compliance, and Governance

Recognizing that privacy is an important piece in the management of national digital identity systems, it is vital that privacy protections are built in and not added after the fact. A "privacy by design" approach is critical for trusting and protecting identity systems and participants. The principle focuses on ensuring that any identity information collected or used is essential to the stated purpose. Adherence to privacy laws in different jurisdictions that also focus on the principle of data minimization is important. Always, full accountability of identity system authorities — and especially private organizations involved — is critical. Privacy cannot be an afterthought.

Auditable compliance with applicable legal frameworks is a strong obligation that contributes to the assurance of the systems. Legal compliance involves issues of data protection and privacy, related to the data considered in the identity management process. Depending on jurisdiction, these issues range from secure storage, use, and disposal of identity data to incident-reporting obligations; from facial recognition to data locality requirements; and from cross-border data flow provisions to data residency concepts.

Equation 2) Integrity over large objects: block hashing and periodic checks

Step 1 — Split into blocks

Split m into B blocks:

$$m = (b_1, b_2, \dots, b_B)$$

Step 2 — Hash each block

$$h_i = H(b_i) \quad \text{for } i = 1..B$$

Step 3 — Aggregate (simple concatenation hash)

A simple integrity root can be:

$$h_{\text{root}} = H(h_1 \parallel h_2 \parallel \cdots \parallel h_B)$$

(Practical systems often use a Merkle tree, but the above shows the idea step-by-step.)

Step 4 — Periodic verification

On audit cycle t, re-fetch blocks, recompute $h_i^{(t)}$ and compare:

$$\forall i,\ h_i^{(t)} \stackrel{?}{=} h_i^{(0)}$$

IV. CLOUD STORAGE MODEL ARCHITECTURES

Cloud storage for national identity data may employ centralized, federated, or hybrid models. Centralized storage is managed by a single legal entity, while federated storage involves multiple organizations and jurisdictions. Hybrid solutions combine both approaches. Centralized storage is burdened by the inherent risks associated with a single point of failure. Latency is typically minimized, but governance and compliance-related risks increase. A clearly defined single authority facilitates the enforcement of privacy, security, and other compliance requirements, but also increases the risk of operational failure. Latency and replication–consistency trade-offs must be considered for cloud storage services located in separate jurisdictions. Federated storage registers copies of identity data, which may be used to enable or augment related services. Ownership of the data is distributed among multiple legal entities in different jurisdictions. Cross-border data flow regulations may affect service implementation, but localized data repositories reduce the likelihood of transgression. Data minimization principles further mitigate exposure. The absence of a single point of failure enhances operational resilience, but disaster recovery planning must address indirect access and jurisdictional challenges. Hybrid solutions may combine the strengths of both approaches, but jurisdictional data sovereignty requirements must be respected.

The overall storage model can directly impose data management capabilities and associated requirements. Decentralized identity storage solutions are designed to enable data localism or data minimization principles. Data localism demands that identity data be stored as close as possible to the identity data subject. Data minimization requires that identity data in cloud storage services have the minimum volume and access life necessary. When encryption keys are separated from the encrypted data, neither the storage nor the access service can be trusted with greater authority than the key management organization. It may therefore be considered on an independent basis.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

	Overall average score
Centralized	3.2857142857142856
Federated	3.2857142857142856
Hybrid	4.0

4.1. Centralized Cloud Storage

The simplest cloud storage model comprises a single repository controlled by a single entity, and it is already in use in some national digital identity systems. This approach offers potentially lower latency in access and service provisioning since identity data are stored and accessed from a centralized location. However, having a single point of storage introduces increased risk, not only because it makes the complete dataset more attractive as a target but also because breach of the repository can lead to debilitating consequences. As a result, user data and user interactions may be exposed, identity assurance levels may not be met or may be lowered, and operation of the identity system may be disrupted. Recovery from an incident may also be complicated. For national identity systems with statutory mandates and centralized identity data repositories, a breach of identity data can be a matter of trust. Cloud storage is susceptible to four principal types of threats: threat actors, attack methods, cloud service weaknesses, and vulnerabilities. An incident in a cloud service provider can potentially affect all users of that service, for example, through service disruption, a data breach, data loss, or account hijacking. Prevention through effective redundancy and risk mitigation remains a goal.

Consequently, the governance structure for such deployments can be clearly defined. However, jurisdictions with data-sovereignty concerns may be hesitant to place their identity data, and potentially the identity data of their citizens, in a single repository. Such concerns highlight the importance of establishing legal frameworks that govern cross-border data flows and enable assured reciprocity in processing the identity data of citizens of one nation by a cloud service provider in another nation. Policymakers must also remain aware that it may not be possible to eliminate physical access to identity data stored in a foreign jurisdiction. Therefore, data localization, and compliance with the privacy requirements of the national jurisdiction in question, remain important considerations. The centralization model also raises questions of resilience, as single points of failure necessitate thorough disaster recovery and business continuity planning and increased focus on resiliency testing.

4.2. Federated and Hybrid Storage

Federated and hybrid cloud storage architectures facilitate governance by multiple entities and help meet data localization and sovereignty requirements. Data from other jurisdictions may be accommodated within the cloud on a case-by-case basis, subject to adequate privacy and security safeguards. Cross-border data flows can be governed with mechanisms such as the EU-U.S. Privacy Shield, which attests compliance with the European General Data and Protection Regulation's data protection principles. In all cases, cross-border flows should follow the principle of least privilege and be limited to the minimum necessary to achieve the intended purpose.

Referring back to the cloud reliability metrics discussed earlier, having data under the control of a plurality of entities makes the cloud more resilient to catastrophic events. Both the federated and the hybrid model, by virtue of their geographic and administrative distribution, significantly reduce the risks associated with the central storage model. However, they also introduce a level of complexity that must be carefully monitored and managed. For example, although identities are checked and deleted by different authorities, identity assurance primarily remains the responsibility of the federated cloud. Consequently, if a country fails to comply with incident reporting requirements, others relying on the federated cloud for identity verification could suffer reputational, operational, and possibly legal damage.

V. DATA MANAGEMENT AND SECURITY MECHANISMS

Maintaining user trust warrants diligent execution of data management processes and comprehensive implementation of information security controls tailored for cloud environments. These include data encryption — both at rest and during transmission, access control mechanisms to ensure proper identity and entitlements for all individuals involved with the data (security and privacy administrators) — as well as reliable key management strategies.

At-rest data encryption safeguards stored information from unauthorized access, while in-transit encryption protects it while being uploaded to or downloaded from the cloud storage service. The information is encrypted before being fed to the underlying storage service — and its corresponding key managed by a dedicated key management service — to

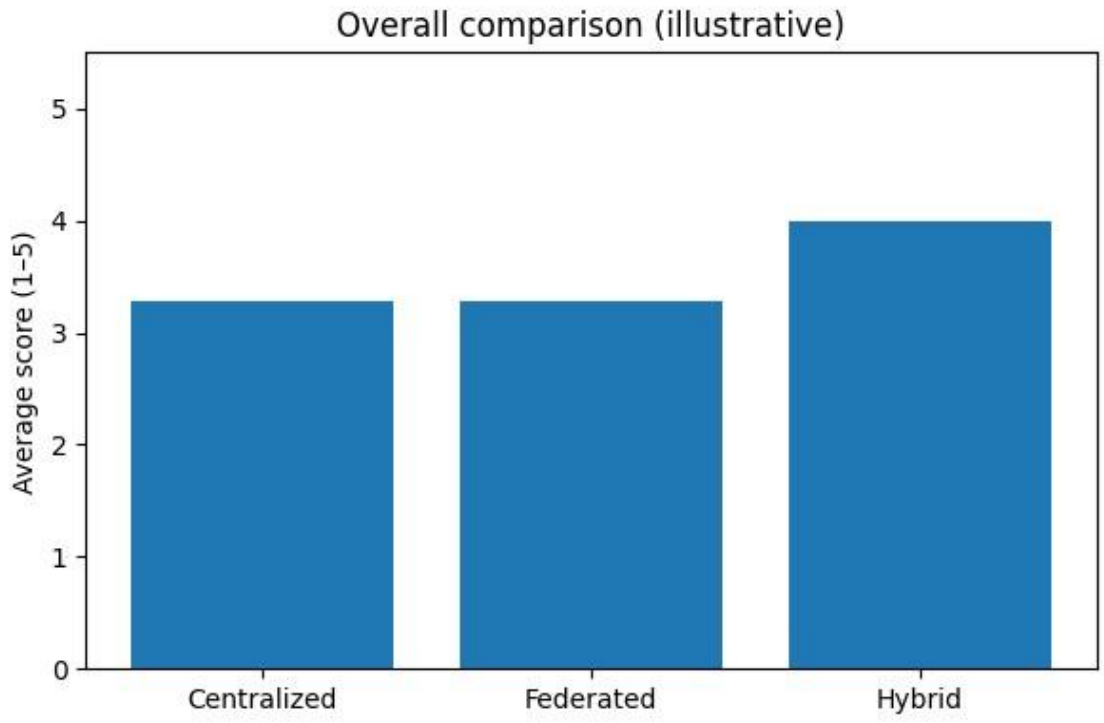
| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

achieve at-rest encryption of user data. The central key management service generates a random key for each encryption operation, stores the encryption metadata, encrypts the metadata with a dedicated key-hierarchy key set on the HSM and then securely stores the metadata in a database for audit purposes. To jointly encrypt and upload the data, the key-provider service uses a specialized key encryption key (KEK) residing only at the key provider to encrypt each random key. These KEKs are securely stored in the underlying key-management infrastructure provided by an HSM, allowing for secure data-at-rest encryption. Periodic rotation of both KEK and HSM keys ensures that all historical data are not compromised if a KEK or HSM key is leaked.

Communication between the services is encrypted using TLS. The CDN provider encrypts data before sharing it on the edge platform by using a customer-set data encryption key (DEK) that is securely stored in an HSM. Generally, the service provider security administrator sets long-lived DEKs, typically for the duration of the product in production, and encrypts with HSM the DEK before associating it with its respective product. The DEK is deleted from the security components after the encryption process is complete. The DEK is stored in the CDN in encrypted form and is decrypted only before the communication to the CDN platform is done.



Equation 3) Availability from replication (multi-zone copies)
Let:

- n = number of replicas (replication factor)
- p = probability a single replica is available at a random moment (assume independent failures for a first-order model)

Step 1 — Probability all replicas are down
A single replica is down with probability $(1-p)$.
All n are down:

$$P(\text{all down}) = (1-p)^n$$

Step 2 — Probability at least one replica is up (system availability)

$$A = 1-(1-p)^n$$

5.1. Data Encryption and Key Management
Data Erasure and Hardware Security Modules (HSM) ensure adequate security, proper Access Control and Audit mechanisms cover authentication, detection of malicious data access and action forensics, intrusion alerts, Audit

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

mechanism should be integrated with rules at various levels. For At-Rest Encryption, Encryption of plain/clear data before being saved in storage.

Data Encryption is a critical mechanism posteriority identity and sensitive information in Cloud Storage. Data can be encrypted before being stored using Asymmetric Encryption Data Encryption Standard (AES-DES) keys. The performance overhead and latency can be reduced by caching the keys in Memory for small size data otherwise in HSM. Data In-Transit Encryption must be enforced as an end-to-end encryption mechanism to ensure that Data is not readable by the service provider, it must be encrypted on the User side and decrypted on the Identity Service Side by DES Cryptographic routines; while the data is in transit, the encryption will take care of the confidentiality preservation, Integrity and confidentiality will be taken care by HMAC are hash based message authentication code. HSM must be used for the management of server keys (At-Rest and In-Transit) by having a Reliable Key Management Policy for creation, storage, updating, distribution and deletion of AES keys.

Encryption is a critical mechanism behind Cloud Security, "Encrypt all Data" must be the mantra for all Cloud Service Providers storage Model. The principle is simple: As Data is stored in Forum of Plain/clear text in Cloud Service Provider therefore it may be physically accessible to malicious cloud service providers, malicious insider, accidental breach by cloud service provider, potential servers and data centre administrators accessing Data over-the-wire and external hackers, if they want to. Thus without Encryption, Cloud Data Security could never be enforced. Encrypting Data before storing in the cloud ensures that Data Stored into Cloud is never readable nor usable.

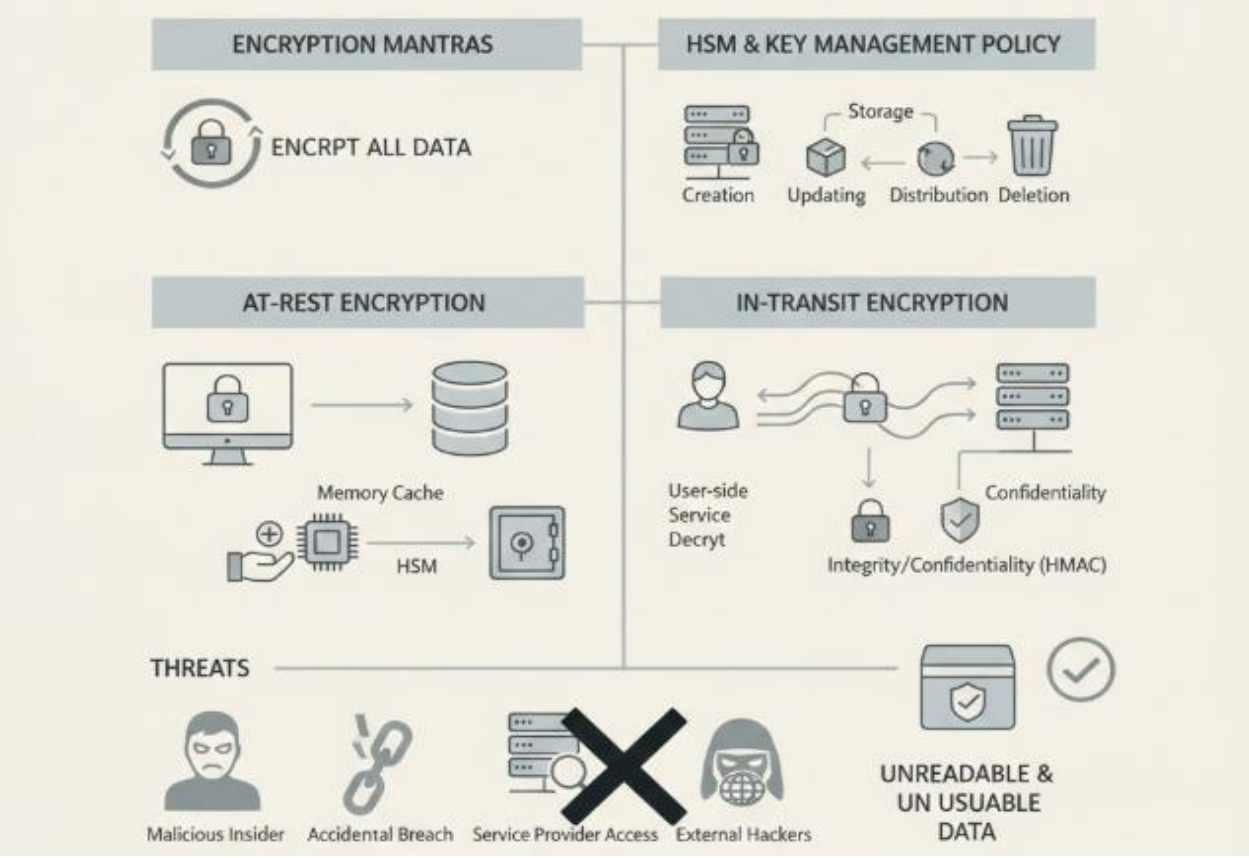


Fig 3: End-to-End Cryptographic Governance: Integrating HSM-Backed Key Management and Multi-Layered Encryption for Robust Cloud Data Security

5.2. Access Control and Identity Verification

Multi-factor authentication must be mandated for all identity data access, validating users with a set of distinct queries. Access must be role-based, constrained to rules that are appropriate and minimised for every authority in conjunction with the privacy-preserving principle.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

To counteract risks related to privileged access, the segregation of duties principle should be applied. This involves delegating different security tasks to separate users so that no single user has complete authority, thus diminishing the risk of unqualified actions. In addition, special surveillance should be undertaken for users with elevated privileges, with auditing, meaningful notification, justification details, and periodic review of the use of such privileges.

VI. OPERATIONAL AND COMPLIANCE IMPLICATIONS

Considerations for the ongoing operation of cloud storage services must cover aspects of resilience to ensure data remains available in potential disasters affecting operations. Such considerations relate to the data management mechanisms in place, which must support the defined reliability and continuity objectives when assessed with the applicable threat and risk profile. A well-understood compliance framework—such as that provided by the Cloud Security Alliance Cloud Controls Matrix or ISO/IEC 27001—supports the establishment of requirements that address these needs in a consistent and auditable manner.

The reliability of any cloud service is defined by the capability of the provider to ensure the expectations placed upon it with regards to data availability. These expectations are commonly expressed in terms of Recovery Time Objective (RTO) and Recovery Point Objective (RPO). The former relates to how quickly a service must be restored following an interruption, while the latter is the maximum period through which data loss is acceptable. For any cloud service, these objectives must be matched with a suitable data-management technology that supports meeting these goals within the required cost constraints. To ensure confidence that these objectives can be met, a provider should maintain and periodically test a formal disaster-recovery plan, with the results subject to appropriate and ongoing review.

Equation 4) Disaster recovery metrics: RPO and RTO (with backup frequency)

4.1 RPO (Recovery Point Objective)

Let:

- Backups occur every Δ time units (e.g., every 1 hour).
- Failure happens at an arbitrary time between backups.

Worst-case data loss occurs if failure happens right before the next backup:

$$\text{DataLoss}_{\max} \approx \Delta$$

So to satisfy an RPO target:

$$\Delta \leq \text{RPO}$$

Expected data loss (if failure uniformly random in the interval) is:

$$\mathbb{E}[\text{DataLoss}] = \frac{\Delta}{2}$$

4.2 RTO (Recovery Time Objective)

Let:

- T_{detect} = time to detect incident
- T_{decide} = time to decide/authorize failover
- T_{restore} = time to restore services/data
- T_{validate} = time to validate system correctness

Then:

$$\text{RTO} = T_{\text{detect}} + T_{\text{decide}} + T_{\text{restore}} + T_{\text{validate}}$$

To meet an RTO requirement:

$$T_{\text{detect}} + T_{\text{decide}} + T_{\text{restore}} + T_{\text{validate}} \leq \text{RTO}_{\text{target}}$$

6.1. Reliability, Disaster Recovery, and Continuity

To ensure data reliability and availability, identity data is regularly backed up to prevent unintentional deletion and to provide for recovery in the event of data corruption or loss due to natural disasters, system/component failures, or other incidents. Identity-management service providers or organizations should define necessary Recovery Time Objectives (RTOs) and Recovery Point Objectives (RPOs) for identity data, which are generally based on the criticality of the data and the cost of maintaining the backup. The RTO defines the maximum allowable time within which an identity service or data should be restored following an interruption (e.g., due to disaster recovery), while the RPO indicates the maximum tolerable amount of data loss. Identity data backups must be regularly tested to validate the backup process and the ability to restore data within the predefined RTO and RPO.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

While RTO and RPO are generally most critical for primary identity-management services running against transactional data stores, organizations must also define failover plans for operational systems and non-transactional data stores. These plans typically identify a secondary geographic location, the services hosted there, and the process to failover to that location. Periodic failover exercises are then conducted to validate the failover plan. These reviews generally focus on the viability of the failover destination and the fidelity of the failover plan itself; the RTO and RPO for a failover plan may not need to be defined or tested if the failover is considered a worst-case scenario. Resilience testing validates the ability of the identity services to operate directly against a backup copy of the data and without the assistance of the primary production facilities (e.g., sandbox environment).

6.2. Compliance Frameworks and Risk Management

Identity data storage in national digital identity service deployments must comply with applicable privacy and security regulations. Regulations may emerge from the jurisdictional or operational location of the service (i.e., national policymakers), or from cross-border flows of identity data between nations that enforce legislative frameworks (e.g., a nation X's identity data stored in a nation Y service provider). To mitigate emerging risks, compliance frameworks, privacy-by-design and -by-default principles, and a stratified risk assessment process advocate for the identification and management of risks associated with data processing within cloud infrastructure storage.

Identity data storages and service interfaces should comply with established standards (where available) or suitably aligned industry best practices (e.g., with respect to security, data localization, incident detection and reporting, and privacy). It is good practice for the compliance framework to be stratified into distinct security/incident impact categories (e.g., low, medium, high) and to incorporate a risk assessment methodology that guides the regular assessment of risks and potential impacts on the data and services in the relevant category. Furthermore, the risk assessment process should regularly incorporate representative stakeholders, including relevant national authorities, scrutiny bodies, affected entities, and trusted third parties. Regulatory bodies overseeing the implementation of a risk assessment framework should regularly audit whether the management strategy of each identified risk during the assessment is commensurate with the impact of the risk.

Equation 5) Encryption + key management equations (DEK/KEK, HSM-backed hierarchy)

A clean step-by-step math model is:

Step 1 — Generate a random Data Encryption Key (DEK)

$$K_d \leftarrow \text{Random}(\lambda)$$

(λ is key length, e.g., 256 bits)

Step 2 — Encrypt plaintext data *M*

Let symmetric encryption be Enc(·):

$$C = \text{Enc}(K_d, M)$$

Step 3 — Wrap the DEK with a Key Encryption Key (KEK)

KEK K_k is protected in an HSM.

$$W = \text{Enc}(K_k, K_d)$$

Step 4 — Store ciphertext + wrapped key

Store (C, W) (and audit metadata).

Step 5 — Decrypt workflow

Unwrap DEK:

$$K_d = \text{Dec}(K_k, W)$$

Recover plaintext:

$$M = \text{Dec}(K_d, C)$$

VII. CONCLUSION

Three fundamental principles guide the analysis: (1) national digital identity systems and services, from both user and provider perspectives, require hardened cloud operational models that ensure the ongoing trustworthiness of stored identity data; (2) three principal cloud storage architectures—centralized, federated, and hybrid—are envisaged for national digital identity systems, each addressing data storage, access, jurisdictional, and control requirements with differing strengths and weaknesses; and (3) the choice among the aforementioned architectures should depend on considerations of data integrity, privacy, regulatory compliance, and governance, rather than the optimal location to minimize latency. Centralized systems have latency advantages for users and can deploy encryption key management in tempos of single-entity ownership and role delimitation. Should identity service providers therefore favour

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

centralized models, a roadmap supporting interconnection—together with the jurisdictions addressing issues of data sovereignty, privacy, and security—should also be delineated.

The key determinant of national identity cloud storage architecture is the specific cross-border data flow requirements for identity service ecosystems; subsequently, resilience becomes critical. By anticipating turbulent or unforeseen times, such as natural disasters, civil unrest, or geopolitical tensions, the digital service delivery infrastructures of the ecosystem partners, whether national or non-nationals, are endowed with full business continuity. By easing these boundaries into cross-border service interconnection, cost-savings and optimised latencies are attainable. Future research needs to address how to build such interconnections in a manner that allows for regional or worldwide use of the identity services, while also defining these interconnections in a manner that meets the jurisdictional concerns of the specific implementing agencies.

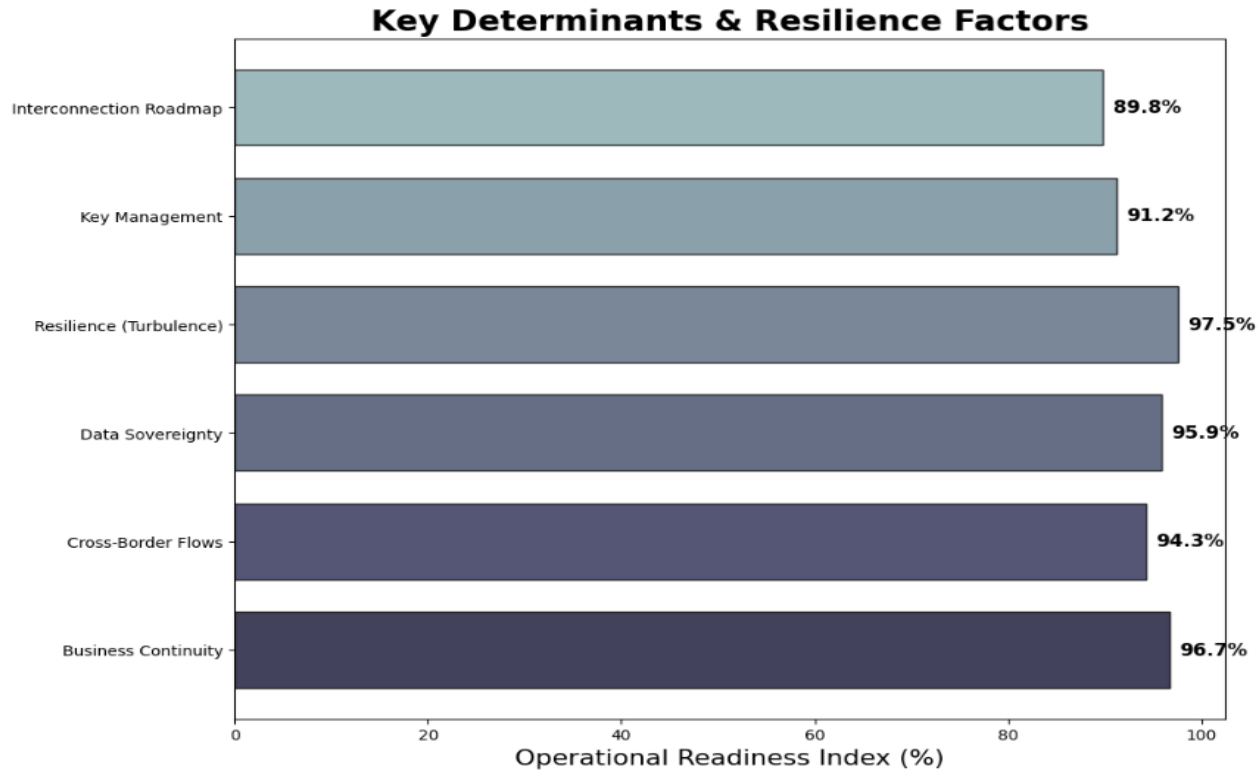


Fig 4: Key Determinants & Resilience Factors

7.1. Key Takeaways and Future Directions

A concise set of key takeaways narrates the benchmarking criteria, identifies cloud storage models recommended for scalable national digital identity systems, and outlines topics meriting further examination.

Cloud storage models that underpin national digital identity systems should aim to balance privacy and risk. The resiliency and trustworthiness of national digital identity systems can be supported through federated and hybrid architectures. These models offer the potential for more direct control over locally stored data while enabling cross-border data flows for processing and authentication.

Further investigation should examine techniques for supporting federated storage of citizen data in a manner that is efficient, intuitive, and cost-effective for citizen credentialing authorities operating at a low level of assurance. Factors that enable identity credentials issued at partner organizations to be migrated, renewed, and revoked in a secure and user-friendly manner may warrant particular attention.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

REFERENCES

[1] Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. National Institute of Standards and Technology.
[2] Buyya, R., Broberg, J., & Goscinski, A. (2011). Cloud computing: Principles and paradigms. Wiley.
[3] 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.
[4] Zhang, Q., Chen, M., Li, L., & Li, Y. (2010). A survey on cloud computing. Journal of Grid Computing, 8(4), 449–472.
[5] Dean, J., & Ghemawat, S. (2008). MapReduce. Communications of the ACM, 51(1), 107–113.
[6] Inala, R. (2020). Building Foundational Data Products for Financial Services: A MDM-Based Approach to Customer, and Product Data Integration. Universal Journal of Finance and Economics, 1(1), 1-18.
[7] Newman, S. (2015). Building microservices. O'Reilly Media.
[8] Pahl, C. (2015). Containerization and the PaaS cloud. IEEE Cloud Computing, 2(3), 24–31.
[9] Bass, L., Clements, P., & Kazman, R. (2013). Software architecture in practice (3rd ed.). Addison-Wesley.
[10] Fielding, R. T. (2000). Architectural styles and the design of network-based software architectures. Doctoral dissertation, University of California.
[11] Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). Design patterns. Addison-Wesley.
[12] Singireddy, S., & Adusupalli, B. (2019). Cloud Security Challenges in Modernizing Insurance Operations with Multi-Tenant Architectures. International Journal of Engineering and Computer Science, 8, 12.
[13] Kreps, J., Narkhede, N., & Rao, J. (2011). Kafka. Proceedings of NetDB Workshop.
[14] Stonebraker, M., Çetintemel, U., & Zdonik, S. (2005). The 8 requirements of real-time stream processing. ACM SIGMOD Record, 34(4), 42–47.
[15] Akidau, T., Chernyak, S., & Lax, R. (2018). Streaming systems. O'Reilly Media.
[16] Zaharia, M., et al. (2016). Apache Spark. Communications of the ACM, 59(11), 56–65.
[17] Davuluri, P. N. (2020). Improving Data Quality and Lineage in Regulated Financial Data Platforms. Finance and Economics, 1(1), 1-14.
[18] Bernstein, P. A., & Newcomer, E. (2009). Principles of transaction processing (2nd ed.). Morgan Kaufmann.
[19] Gray, J., & Reuter, A. (1993). Transaction processing. Morgan Kaufmann.
[20] Amistapuram, K. 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.
[21] Bender, D., & Sartipi, K. (2013). HL7 FHIR. Proceedings of CBMS, 326–331.
[22] Mandel, J. C., et al. (2016). SMART on FHIR. Journal of the American Medical Informatics Association, 23(5), 899–908.
[23] Benson, T., & Grieve, G. (2016). Principles of health interoperability. Springer.
[24] Lehne, M., et al. (2019). The use of FHIR in digital health. Studies in Health Technology and Informatics, 267, 52–58.
[25] Vadisetty, R., Polamarasetti, A., Guntupalli, R., Rongali, S. K., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2020). Generative AI for Cloud Infrastructure Automation. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 1(3), 15-20.
[26] Allen, C. (2016). The path to self-sovereign identity. Internet Identity Workshop.
[27] Hardt, D. (2012). The OAuth 2.0 authorization framework. IETF RFC 6749.
[28] Sakimura, N., Bradley, J., Jones, M., de Medeiros, B., & Mortimore, C. (2014). OpenID Connect core 1.0. OpenID Foundation.
[29] Jøsang, A., Fabre, J., Hay, B., Dalziel, J., & Pope, S. (2005). Trust requirements in identity management. Proceedings of ACSW Frontiers.
[30] Bertino, E., & Takahashi, K. (2011). Identity management: Concepts, technologies, and systems. Artech House.
[31] Pandiri, L., Singireddy, S., & Adusupalli, B. (2020). Digital Transformation of Underwriting Processes through Automation and Data Integration. Global Research Development (GRD) ISSN, 2455-5703.
[32] Cavoukian, A. (2011). Privacy by design. Information and Privacy Commissioner of Ontario.
[33] Solove, D. J., & Schwartz, P. M. (2018). Information privacy law (6th ed.). Wolters Kluwer.
[34] European Union. (2016). General Data Protection Regulation. Official Journal of the European Union.
[35] OECD. (2013). The OECD privacy framework. OECD Publishing.
[36] Inala, R. Designing Scalable Technology Architectures for Customer Data in Group Insurance and Investment Platforms.
[37] ISO/IEC. (2018). ISO/IEC 27018: Protection of PII in public clouds.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 3, Issue 6, November- December 2020 ||

DOI: 10.15680/IJCTECE.2020.0306004

[38] Cloud Security Alliance. (2017). Cloud controls matrix (v3.0).
[39] NIST. (2018). Framework for improving critical infrastructure cybersecurity.
[40] Anderson, R. (2008). Security engineering (2nd ed.). Wiley.
[41] Kahn Academy & Diffie, W., & Hellman, M. (1976). New directions in cryptography. IEEE Transactions on Information Theory, 22(6), 644–654.
[42] Rivest, R. L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures. Communications of the ACM, 21(2), 120–126.
[43] Menezes, A. J., van Oorschot, P. C., & Vanstone, S. A. (1996). Handbook of applied cryptography. CRC Press.
[44] Kahn Academy & Merkle, R. (1988). A digital signature based on a conventional encryption function. Advances in Cryptology, 369–378.
[45] Boneh, D., & Franklin, M. (2001). Identity-based encryption. Advances in Cryptology—CRYPTO.
[46] Gottimukkala, V. R. R. (2020). Energy-Efficient Design Patterns for Large-Scale Banking Applications Deployed on AWS Cloud. power, 9(12).
[47] Dwork, C. (2006). Differential privacy. ICALP.
[48] Chaum, D. (1985). Security without identification. Communications of the ACM, 28(10), 1030–1044.
[49] Sabahi, F. (2011). Cloud computing security threats and responses. IEEE.
[50] Behl, A., & Behl, K. (2017). Cyberwar. Oxford University Press.
[51] Laudon, K. C., & Laudon, J. P. (2018). Management information systems (15th ed.). Pearson.
[52Varri, D. B. S. (2020). Automated Vulnerability Detection and Remediation Framework for Enterprise Databases. Available at SSRN 5774865.
[53] McAfee, A., & Brynjolfsson, E. (2012). Big data. Harvard Business Review, 90(10), 60–68.
[54] Fox, A., & Patterson, D. (2012). Engineering long-lived software. Communications of the ACM, 55(5), 71–79.
[55] Davuluri, P. N. (2020). Event-Driven Architectures for Real-Time Regulatory Monitoring in Global Banking.
[56] Brewer, E. A. (2000). Towards robust distributed systems. PODC.
[57] Gilbert, S., & Lynch, N. (2002). Brewer's conjecture and feasibility of consistent, available, partition-tolerant web services. ACM SIGACT News, 33(2), 51–59.
[58] Hohpe, G., & Woolf, B. (2003). Enterprise integration patterns. Addison-Wesley.
[59] Segireddy, A. R. (2020). Cloud Migration Strategies for High-Volume Financial Messaging Systems.
[60] Adusupalli, B., Pandiri, L., & Singireddy, S. (2019). DevOps Enablement in Legacy Insurance Infrastructure for Agile Policy and Claims Deployment. risk, 7(12).
[61] Bernstein, P. A. (2011). Middleware: A model for distributed system services. Communications of the ACM, 39(2), 86–98.
[62] Weiser, M. (1991). The computer for the 21st century. Scientific American, 265(3), 94–104.
[63] Haux, R. (2010). Medical informatics: Past, present, future. International Journal of Medical Informatics, 79(9), 599–610.
[64] Meystre, S. M., et al. (2017). Clinical data reuse. Yearbook of Medical Informatics, 26(1), 38–52.
[65] Greenhalgh, T., et al. (2017). Beyond adoption. Journal of Medical Internet Research, 19(11), e367.
[66] Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records. Nature Reviews Genetics, 13(6), 395–405.
[67] Shickel, B., et al. (2018). Deep EHR. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589–1604.
[68] Rajkomar, A., et al. (2018). Scalable and accurate deep learning with EHRs. npj Digital Medicine, 1, 18.
[69] Esteva, A., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.