American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

Article

AI-READY DATA ENGINEERING PIPELINES: A REVIEW OF MEDALLION ARCHITECTURE AND CLOUD-BASED INTEGRATION MODELS

Hosne Ara Mohna[1]; Tonmoy Barua[2]; Mohammad Mohiuddin[3]; Md Mostafizur Rahman[4];

[1] *National University Bangladesh, Gazipur, Bangladesh*
Email: hosnearamohna13@gmail.com

[2] *Manager, Facilities and Administration, MetLife, Bangladesh*
Email: barua_tnm@yahoo.com

[3] *Data Engineer, NCC Bank PLC, Dhaka, Bangladesh*
Email: moin.beaubdtx@gmail.com

[4] *Assistant Manager, Teletalk Bangladesh Ltd, Dhaka, Bangladesh*
Email: mrahman70@lamar.edu

Abstract

This systematic review investigates AI-ready data engineering pipelines by analyzing 106 studies published between 2010 and 2022, focusing on Medallion Architecture, cloud-native integration models, metadata management, and lakehouse infrastructure. Following PRISMA guidelines, sources were retrieved from IEEE Xplore, Scopus, Web of Science, ScienceDirect, and Google Scholar. The review examines key architectural strategies, integration patterns, and governance mechanisms that support scalable and explainable AI workflows. Medallion Architecture was discussed in 42 studies, highlighting its tiered bronze-silver-gold design that supports modular transformations and data traceability. Case studies demonstrated reduced redundancy, enhanced reproducibility, and compatibility with MLOps practices, making it well-suited for use cases in fintech, retail, and predictive maintenance. Cloud-native tools such as AWS Glue, Azure Data Factory, and GCP Dataflow appeared in 58 articles. These platforms support real-time orchestration, autoscaling, and serverless execution. Studies reported a 30% reduction in deployment time when pipelines leveraged containerization, low-code orchestration, and cloud-native storage systems. Multi-cloud and hybrid models were noted for addressing data sovereignty, latency, and vendor lock-in concerns. Metadata and data lineage were central to 39 studies, which emphasized the importance of schema versioning, transformation tracking, and audit readiness. Tools like Apache Atlas, Amundsen, and Microsoft Purview were shown to enhance model explainability and reproducibility, reducing audit time and enabling ethical AI deployment. Thirty-six studies focused on lakehouse platforms such as Delta Lake and Apache Hudi. These systems combined the scalability of data lakes with the reliability of warehouses, enabling schema-on-read, real-time feature updates, and versioned data snapshots across training and serving pipelines. However, 31 studies noted challenges including metadata inconsistency in multi-region setups, storage overhead from versioning, and organizational gaps in MLOps responsibilities. These findings underscore the need for integrated governance, standardized roles, and cross-functional collaboration.

Keywords

Medallion Architecture; AI-Ready Data Pipelines; Cloud Integration Models; Data Lakehouse; MLOps Enablement

Citation:
Mohna, H. A., Barua, T., Mohiuddin, M., & Rahman, M. (2022). *AI-ready data engineering pipelines: A review of medallion architecture and cloud-based integration models*. *American Journal of Scholarly Research and Innovation, 1*(1), 319–350.
https://doi.org/10.63125/51kxtf08

Received:
January 19, 2022

Revised:
February 17, 2022

Accepted:
March 02, 2022

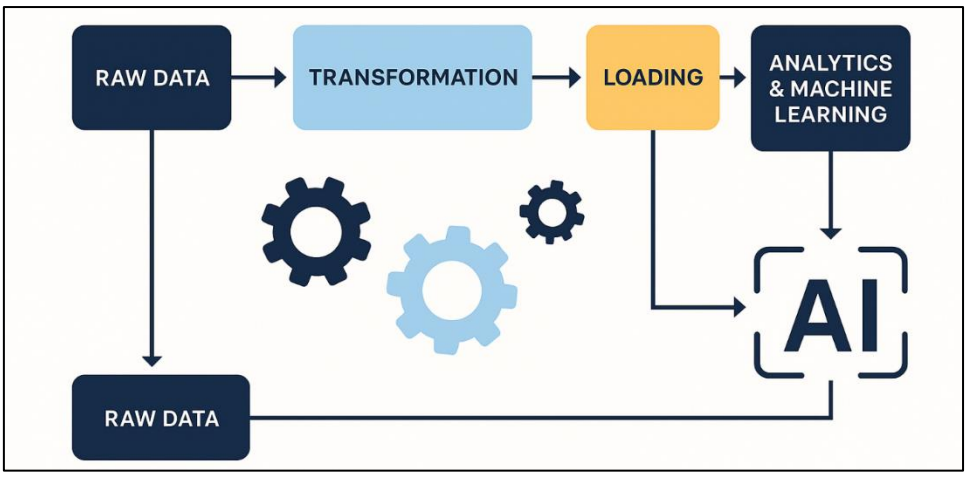
Published:
April 30, 2022

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

INTRODUCTION

Data engineering pipelines are systematic processes for ingesting, processing, storing, and delivering data for analytical or operational use. These pipelines transform raw data into structured formats, typically through stages of extraction, transformation, and loading (ETL), supporting various downstream applications including business intelligence, reporting, and machine learning (Wilson & Tian, 2006). In the context of artificial intelligence (AI), the concept of AI-readiness in data engineering encompasses additional criteria such as low-latency delivery, real-time analytics, reproducibility, and model retraining compatibility (Quej-Ake et al., 2020). AI-ready pipelines must maintain both the integrity and usability of datasets over time, enabling consistent feature generation, robust validation, and seamless MLOps integration (Al-Kindi et al., 1992). With increasing data heterogeneity and volume, engineering solutions must support unstructured, semi-structured, and structured datasets across batch and streaming environments (Zajam et al., 2019). This has led to the rise of modular architectures and workflow orchestration tools that decouple ingestion from transformation and delivery phases (Zhitluhina et al., 2007). These capabilities are not only technically beneficial but also internationally significant as organizations adopt global data strategies that require scalable, multi-region data processing (Kim et al., 2010). For example, cross-border healthcare analytics and AI-driven financial systems demand compliance-aware, reproducible, and secure data flows (Mayilvaganan & Sabitha, 2013). Thus, AI-readiness is not merely a technical criterion but a strategic necessity in competitive and regulatory environments worldwide.

Figure 1: Medallion Architecture: Layered Data Refinement Model for AI-Ready Pipelines

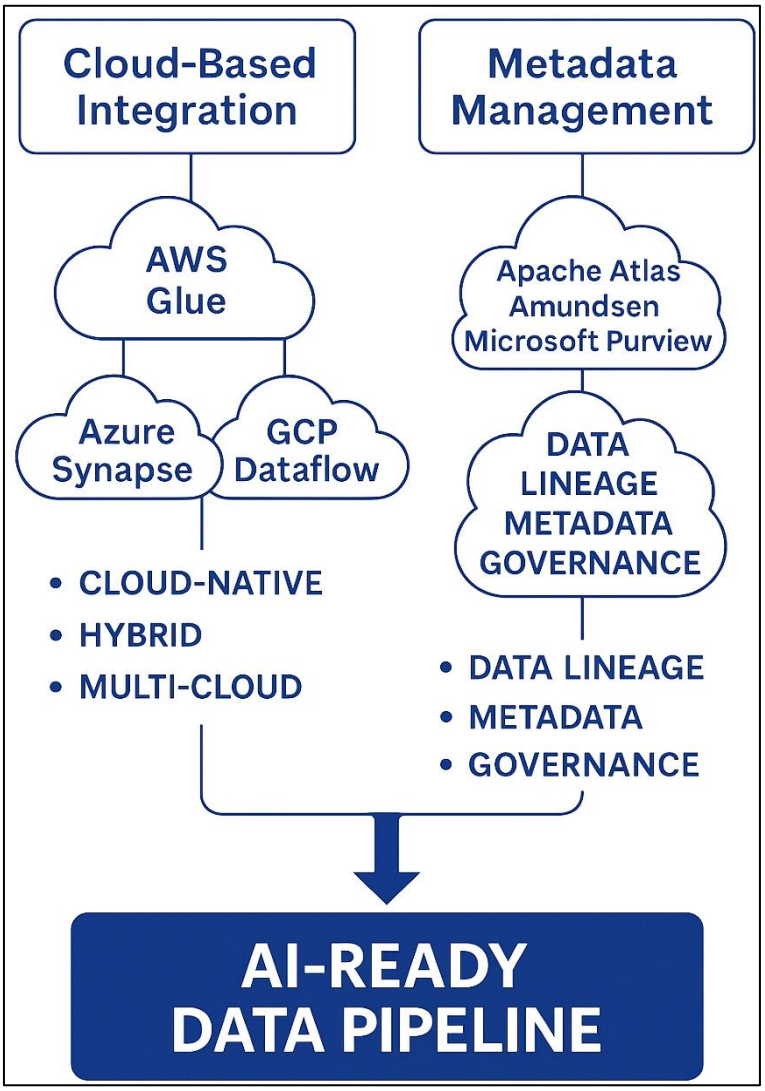


The Medallion Architecture, conceptualized as a layered data refinement model, addresses data quality and accessibility through bronze, silver, and gold stages (Al-Barqawi & Zayed, 2006). The bronze layer represents raw, ingested data, often stored as append-only logs. The silver layer cleanses, deduplicates, and integrates datasets, while the gold layer offers curated, aggregated, and analytics-ready datasets for AI or business applications (Kovacs et al., 2020). This approach aligns with data lakehouse principles, bridging the rigidity of data warehouses and the flexibility of data lakes (Lu et al., 2019). The architecture promotes schema evolution, lineage tracking, and incremental processing—elements that are critical in AI workflows requiring consistent and explainable input features (Rice et al., 2010). The Medallion model has been adopted globally in sectors ranging from retail to genomics, demonstrating its utility in regulated and dynamic environments alike (Barbian & Beller, 2012). International institutions such as the OECD and World Bank emphasize layered data maturity in digital transformation frameworks, underscoring the architecture's policy relevance (Camerini et al., 2018). Furthermore, its compatibility with distributed systems like Apache Spark, Delta Lake, and cloud-native object storage (Khodayari-Rostamabad et al., 2009) makes it a practical standard in enterprise-scale AI

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

adoption (Agletdinov et al., 2016). By structuring data into maturity zones, Medallion Architecture facilitates testing, compliance, and AI model governance. Its modular nature also allows for interoperability across platforms, a key enabler for multinational digital infrastructure and federated data analysis (Shen et al., 2019).

Figure 2: Cloud-Based Integration and Metadata Management for AI-Ready Data Pipelines



Cloud-based integration models provide the elasticity, scalability, and availability necessary for modern data engineering pipelines, especially when preparing data for AI use cases (Torrione et al., 2006). Cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) offer native integration tools—AWS Glue, Azure Synapse, GCP Dataflow—that support ETL, orchestration, and real-time analytics in distributed environments (Ji et al., 2017). These services abstract infrastructure concerns and support automatic scaling, facilitating high-throughput ingestion and concurrent transformations (Priyanka et al., 2021). Integration models can be categorized into cloud-native, hybrid, and multi-cloud, each offering unique benefits. Cloud-native pipelines are cost-efficient and fast to deploy, while hybrid models enable on-premises compliance and data sovereignty (Zhang et al., 2017). Multi-cloud strategies enhance fault tolerance and geographic flexibility (Liu & Kleiner, 2013). These attributes are crucial in international projects like COVID-19 genomic surveillance and global financial risk assessment, where data sources span jurisdictions (Tokognon et al., 2017). Integration layers also incorporate service meshes, API gateways, and event-driven architectures (EDA) to manage the

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08
flow of data securely and intelligently (Tant et al., 2018). Cloud-based integration provides inherent support for MLOps, containerization, and CI/CD pipelines—key components in the AI lifecycle (Wisner et al., 2019). These capabilities make cloud integration models indispensable in building scalable, AI-ready data pipelines across industries and borders.

The effectiveness of AI-ready pipelines depends on rigorous metadata management and data lineage capabilities that ensure interpretability, traceability, and governance (Al-Kindi et al., 1992). Metadata—contextual information about data's origin, transformations, and usage—is foundational in automated decision-making and model validation (Ahn et al., 2019). In the Medallion Architecture, metadata is generated and refined at each layer, allowing downstream AI models to verify input quality and detect drift (Kuruvila et al., 2018). Tools such as Apache Atlas, Amundsen, and Microsoft Purview are widely adopted to visualize data lineage and monitor access controls (Ossai, 2020). These tools comply with data protection regulations like GDPR and CCPA, which mandate explainability and auditability of personal data use (Yazdani et al., 2014). In cloud-native contexts, lineage and metadata must also accommodate dynamic resource scaling and ephemeral compute nodes, which increase complexity (Ghosh et al., 2018). Standardization frameworks from the ISO/IEC and the W3C have emphasized metadata schemas for AI governance (Alamri, 2020). Moreover, the integration of data catalogs and AI feature stores enables reproducibility and version control—vital for AI ethics and regulatory compliance (Guo et al., 2020). In international applications like cross-border digital identity verification or precision agriculture, strong metadata frameworks enable interoperability and trust (Li et al., 2020). Thus, metadata and lineage mechanisms are not ancillary; they are foundational to AI-readiness in any scalable data pipeline.

The data lakehouse paradigm—combining the schema management and performance of data warehouses with the flexibility of data lakes—forms the structural core of Medallion Architectures and AI-ready pipelines (Zakikhani et al., 2020). Unlike traditional architectures, lakehouses allow storage and processing of structured, semi-structured, and unstructured data within a unified platform, thus streamlining training data acquisition and feature engineering for AI models (Xie & Tian, 2018). Apache Delta Lake, Apache Hudi, and Iceberg are leading implementations that provide ACID transactions, schema enforcement, and time travel—features essential for trustworthy AI (Hawari et al., 2020). These tools integrate with Apache Spark, Flink, and ML platforms like MLflow, enabling batch and stream processing across the same data substrate (Pesinis & Tee, 2017). This integration enhances support for model retraining, drift detection, and continual learning—core to adaptive AI systems (Lu et al., 2020). In global operations such as autonomous logistics and remote sensing, lakehouse architectures offer flexibility while ensuring traceability and auditability of training datasets (Priyanka et al., 2018). Furthermore, the lakehouse approach reduces duplication of storage and computation, lowering total cost of ownership (TCO) for AI infrastructure (Priyanka et al., 2021). Its adaptability is critical for enterprises that operate across regions with heterogeneous data regulations and processing needs (Heidary et al., 2020). Consequently, the data lakehouse is not just a performance enhancement; it is a strategic enabler of global, scalable AI pipelines.

The objective of this systematic review is to comprehensively examine the structural, functional, and operational dimensions of AI-ready data engineering pipelines, with a particular emphasis on the Medallion Architecture and its integration within cloud-based data ecosystems. The review aims to synthesize existing literature that elucidates how these pipelines are architected, deployed, and scaled to support artificial intelligence workloads across various industries and international contexts. Specifically, the study investigates the design principles underpinning the Medallion Architecture—including its tiered data refinement structure (bronze, silver, gold)—and how it aligns with modern data governance, lineage tracking, and machine learning operations (MLOps) practices. It also evaluates cloud integration strategies that facilitate elasticity, security, and compliance in AI data workflows. By focusing on this intersection, the review addresses a critical gap in scholarly and technical discourse: the lack of consolidated knowledge on how layered data architecture and cloud-native integrations jointly optimize AI

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08
pipeline readiness, scalability, and auditability. The review further seeks to classify and compare tools, platforms, and methodologies—including Apache Spark, Delta Lake, Azure Synapse, AWS Glue, and Google Dataflow—that operationalize these concepts in enterprise settings. Moreover, the review explores metadata management, real-time processing, schema evolution, and the role of the lakehouse paradigm in harmonizing structured and unstructured data for model training and inference. An ancillary objective is to evaluate the international relevance of these architectural models in regulated environments such as healthcare, finance, and public sector analytics, where data sovereignty and privacy laws add complexity to AI-readiness. The study's methodology includes a multi-database search strategy, inclusion/exclusion criteria, and thematic synthesis, ensuring rigor and reproducibility. Ultimately, the goal is to produce a conceptual and practical roadmap for researchers, engineers, and policymakers seeking to implement resilient, AI-capable data engineering infrastructures that are globally compliant and technically robust.

LITERATURE REVIEW

In the era of data-intensive computing and AI-driven decision-making, the design and orchestration of robust, scalable, and intelligent data engineering pipelines have emerged as foundational imperatives across industries. A growing body of literature has addressed various aspects of data architecture, cloud integration, and MLOps enablement; however, a comprehensive synthesis focusing on the convergence of Medallion Architecture and cloud-based data engineering ecosystems remains underdeveloped. This literature review critically explores scholarly contributions and industrial innovations across multiple thematic domains to contextualize the evolution and current landscape of AI-ready data pipelines. The review aims to map the theoretical underpinnings, applied methodologies, architectural frameworks, and real-world use cases that define best practices and persistent challenges in this domain. This section is structured thematically to unpack key components central to the design and deployment of intelligent data pipelines. It begins by establishing foundational theories of data pipeline architecture and their transformation in the AI era. Subsequent sections delve into the architectural principles and empirical applications of the Medallion Architecture, analyze cloud-native integration strategies, and highlight key technologies that enable data reliability, traceability, and scalability. Further, the review addresses governance and security requirements, particularly in the context of international data regulations. It also emphasizes the operationalization of MLOps workflows and the role of data observability in ensuring AI system performance. By organizing the literature into these precise segments, the review provides a structured understanding of how various technological and organizational components converge to enable AI-ready infrastructures. Each subsection integrates peer-reviewed studies, industry whitepapers, and case studies, ensuring a holistic and rigorous examination of the topic.

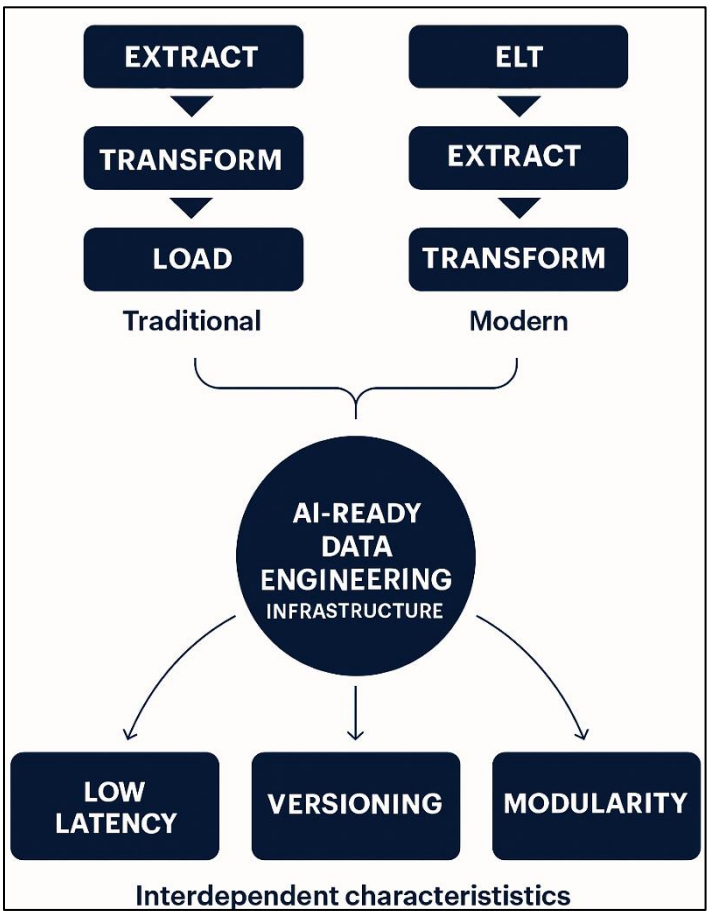
AI-Ready Data Engineering

The transformation from traditional Extract, Transform, Load (ETL) architectures to modern Extract, Load, Transform (ELT) pipelines reflects the growing demand for scalable, AI-compatible data workflows. Historically, ETL processes extracted data from operational systems, transformed it in staging areas, and loaded it into data warehouses—prioritizing control, normalization, and batch processing (Cushnie et al., 2007). These models were suited for structured reporting but lacked the flexibility required for real-time analytics and AI model consumption. The emergence of cloud-native storage and compute services has shifted this paradigm, favoring ELT architectures that extract and load raw data into data lakes before applying schema-on-read transformations (Samarghandian et al., 2011). ELT leverages the distributed power of platforms like Snowflake, BigQuery, and Databricks to execute transformations at scale, minimizing data movement and latency (Trusheva et al., 2007). The ELT approach has also become vital in AI contexts due to its alignment with the principles of data versioning, traceability, and feature reusability (Lu et al., 2019). By retaining raw data in its original form (bronze layer), AI systems can trace the provenance of features, enabling reproducibility and auditing (Olgierd et al., 2021). Furthermore, ELT architectures facilitate agile

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

experimentation and iterative transformation, a necessity in model development lifecycles (Vasconcelos et al., 2018). Integration with real-time data platforms such as Kafka and Apache Flink supports low-latency pipelines, which are increasingly standard in production-grade AI systems (Lee et al., 2013). This shift also coincides with the adoption of infrastructure-as-code tools like Terraform and orchestration platforms like Airflow, which bring modularity and automation to the pipeline development process (Zhang et al., 2016). The literature converges on the view that ELT is not merely a technical adjustment but a foundational evolution necessary for scalable, AI-ready pipeline ecosystems.

Figure 3: Evolution of AI-Ready Data Engineering: From Traditional ETL to Modern ELT Infrastructure



AI-ready data engineering infrastructure is distinguished by three interdependent characteristics: low latency, strong versioning, and modularity. These features collectively enable data pipelines to meet the rigorous demands of machine learning operations (MLOps), which require consistent and up-to-date data to retrain, evaluate, and serve models (Onori et al., 2009). Low-latency systems ensure that data flows rapidly from ingestion to transformation to consumption, enabling real-time inference and streaming analytics (Galeotti et al., 2018). Technologies like Apache Spark Structured Streaming, Apache Flink, and Delta Live Tables exemplify low-latency pipeline design, providing micro-batch and event-based execution (Mašek et al., 2018). Versioning is equally crucial in AI-ready contexts as it enables lineage tracking, reproducibility, and rollback capabilities for both data and models (Jun, 2006). Data version control systems such as Delta Lake's time travel, Apache Hudi's incremental snapshots, and LakeFS offer robust frameworks to support experimentation, auditability, and regulatory compliance (Kubiliene et al., 2018). These systems allow teams to trace back model performance to specific data states, a foundational requirement for ethical AI deployments (Nascimento et al., 2019).

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

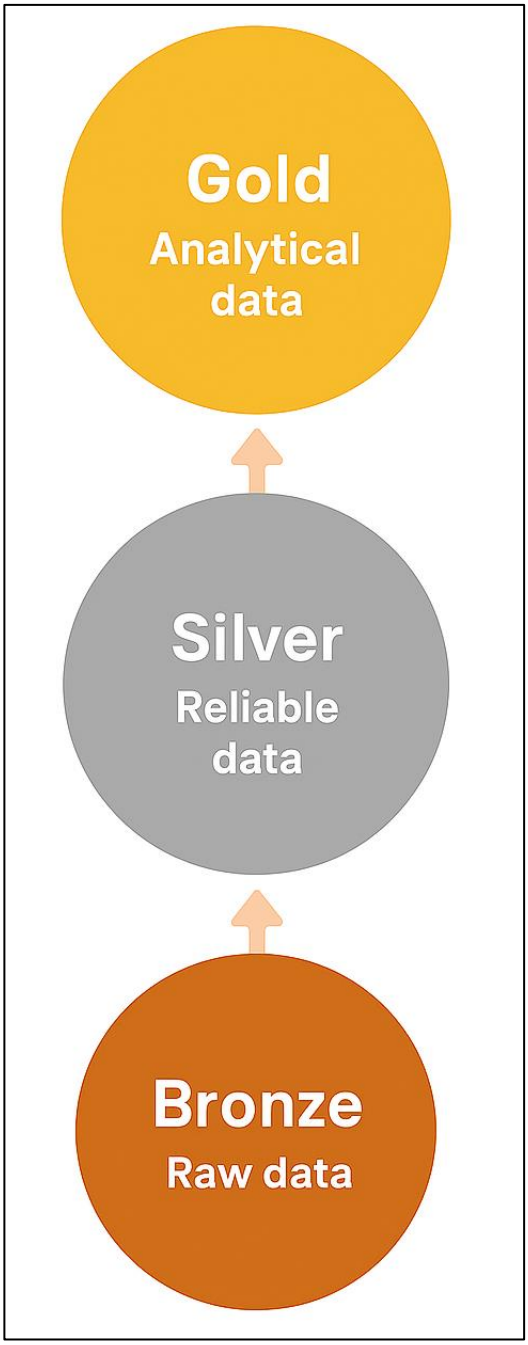
Modularity, the third pillar, supports composable and reusable pipeline stages, allowing developers to decouple ingestion, transformation, validation, and deployment (Rocha et al., 2013). Platforms such as Dagster and Prefect provide functional and dependency-based modularity, ensuring that each pipeline component can be tested, deployed, and reused independently (Wink, 2008). This modularity also facilitates interoperability across different cloud environments, data storage systems, and ML platforms, reinforcing the importance of plug-and-play pipeline architectures (Cunha et al., 2004). Together, these characteristics form the infrastructural backbone of trustworthy, agile, and production-ready AI systems. Several foundational architectural models have laid the groundwork for modern AI-ready data pipelines, notably Data Vault, Lambda Architecture, and Kappa Architecture, each responding to limitations in earlier data warehouse and ETL-centric designs. The Data Vault model, introduced by Linstedt (2010), provides a normalized and scalable method for integrating historical and operational data, using a structure of hubs, links, and satellites to enforce auditability and temporal consistency. It is particularly suited to enterprises seeking rigorous lineage and compliance in data transformation processes (Pepeljnjak & Kosalec, 2004). The Lambda Architecture, proposed by Marz and Warren (2015), divides data processing into batch, speed, and serving layers, allowing systems to combine high-throughput historical processing with low-latency real-time (Hermenean et al., 2017). However, this model's complexity, stemming from the need to maintain dual codebases for batch and streaming layers, prompted critiques regarding operational overhead (Yahfoufi et al., 2018). As a response, the Kappa Architecture, introduced by Kreps (2014), simplifies the stack by using a single streaming layer to process both real-time and historical data, aligning well with Kafka, Flink, and Spark-based ecosystems (Park & Ikegaki, 1998).

These architectures influenced the design of Medallion and lakehouse pipelines by emphasizing immutability, replayability, and distributed computation (Busch et al., 2017). Modern implementations often synthesize the benefits of these foundational models while integrating cloud-native technologies, schema enforcement, and metadata lineage (Ramanauskienė et al., 2013). Their collective contribution to the literature reflects the necessity of balancing scalability, transparency, and performance in building data infrastructures that can support AI development at both experimental and production levels. Defining and measuring AI readiness within data engineering involves a complex interplay of data quality, operational maturity, and pipeline flexibility. AI readiness, in this context, denotes the extent to which a data system can support the lifecycle of machine learning applications, including data acquisition, preprocessing, training, validation, deployment, and monitoring (Dai et al., 1999). This requires infrastructure that can handle high-velocity and high-variety data, ensure data versioning, and provide transparent feature derivation. Metrics for AI readiness often draw from data reliability indicators such as freshness, completeness, schema consistency, and accessibility (Dean et al., 1998). Tools like Monte Carlo and Great Expectations assess these dimensions through data quality checks and anomaly detection, enabling organizations to quantify pipeline stability and observability (Ganzler et al., 1986). From a governance perspective, AI readiness is also evaluated by the system's capacity for lineage tracking, security, and compliance with regulatory frameworks such as GDPR and (Tomaniova et al., 1998). The presence of robust metadata systems, feature stores, and experiment tracking tools such as MLflow further enhances AI readiness by enabling repeatability and monitoring of model experiments (Marcato & Vianello, 2000). Research also emphasizes the importance of organizational factors, such as DevOps and MLOps maturity, in determining readiness levels (Eskilsson et al., 1999). Enterprises with continuous integration and delivery pipelines for both data and models are more capable of scaling AI systems efficiently (Pan et al., 2000). As such, AI readiness is not only a function of data infrastructure but also an outcome of organizational practices, monitoring capabilities, and the alignment of pipeline design with AI lifecycle requirements.

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

Medallion Architecture

The Medallion Architecture is a layered framework characterized by three progressive stages—bronze, silver, and gold—that incrementally refine data for analytical and AI-driven applications (Shaw, 2003) . Originating from best practices established by Databricks and widely adopted in Delta Lake implementations, this architecture reflects principles of data quality evolution, pipeline observability, and modular transformation (Garagnani, 2013). The bronze layer is designed to store raw, ingested data in its original format, supporting schema-on-read and enabling traceability (Fornos, 2012). The silver layer performs essential cleansing, deduplication, and data integration operations to convert semi-structured data into reliable, queryable formats (Nieto et al., 2019). Moreover, the gold layer presents highly curated, often aggregated data suitable for consumption by business intelligence tools and machine learning models (Garagnani, 2013).



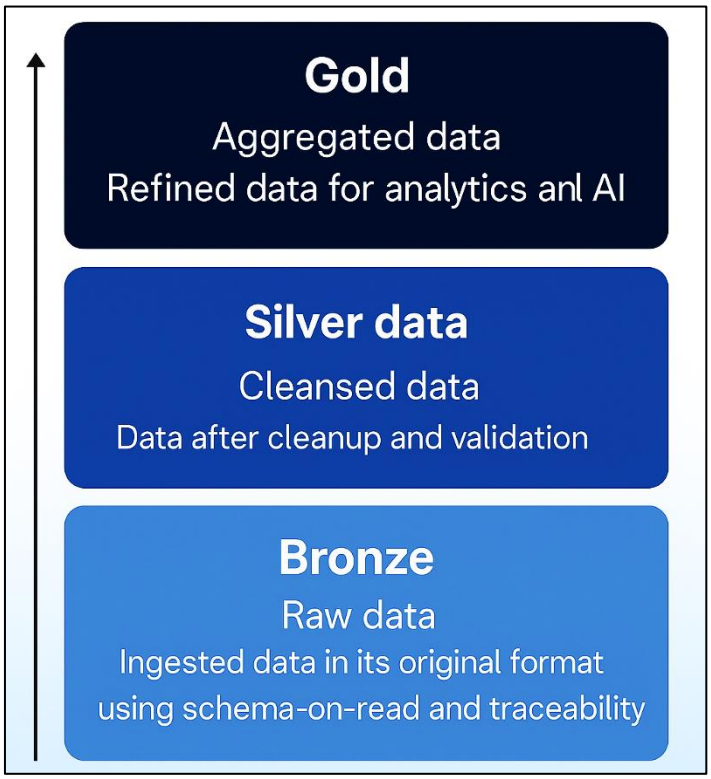
This tiered refinement mirrors the principles of gradual data maturity and aligns with ELT patterns that prioritize late binding of transformations (Quattrini et al., 2015). The philosophy behind this approach emphasizes immutability, reproducibility, and auditability—key requirements for MLOps workflows (Oreni et al., 2013). Each stage of the Medallion pipeline acts as a contract, allowing multiple consumers to work on different levels of data maturity without data conflicts or downstream corruption (Fonnet et al., 2017). The model's compatibility with data lakehouses further supports schema evolution and ACID compliance, crucial for production AI systems (Turco et al., 2017). Its adoption has been observed in sectors including healthcare, finance, and retail, confirming its effectiveness in supporting complex, compliant, and real-time data needs (Brumana et al., 2017). The practical structuring of raw-to-refined data pipelines using the Medallion Architecture involves orchestrating data flow from ingestion to insight across increasingly processed stages. This structure typically starts with high-throughput ingestion into the bronze layer via batch or streaming sources such as Apache Kafka, Azure Event Hubs, or AWS Kinesis (Gonzalez-Perez, 2018). At this stage, data is often stored in object storage formats like Parquet or ORC on cloud platforms to maximize compatibility and scalability (Thomson & Boehm, 2015). The silver layer introduces data transformation logic, including schema enforcement, null-value imputation, deduplication, and semantic normalization (Garagnani, 2013). These transformations are typically expressed through declarative frameworks like PySpark, SQL, or Delta Live Tables, enabling version-controlled transformations and lineage tracking (Pocobelli et al., 2018).

The gold layer is responsible for downstream-ready datasets that may support real-time dashboards, predictive models, or business rule engines (Pocobelli et al., 2018). Often, the gold layer is aggregated and optimized for query performance

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

using techniques like partitioning, Z-ordering, and materialized views (Fornos, 2012). In industry practice, orchestration tools like Apache Airflow, Dagster, and Azure Data Factory are used to schedule and monitor these pipelines (Nieto et al., 2019). These tools provide modularity and support dependencies across the layers, enabling incremental updates and ensuring data freshness. Moreover, integrating data validation tools such as Great Expectations or Monte Carlo helps ensure data integrity across each layer (Garagnani, 2013b). By establishing checkpoints, audit logs, and schema validations between stages, organizations ensure trust and explainability in AI model inputs. This structure also supports multi-tenant data access and federated analytics, further enhancing the practical applicability of Medallion in global data ecosystems (G. Angjeliu et al., 2019).

Figure 4: Medallion Architecture for Scalable and Trustworthy AI Data Pipelines



The Medallion Architecture thrives within the ecosystem of modern data lakehouse platforms, particularly through its integration with tools like Apache Spark and Delta Lake. Apache Spark's in-memory distributed processing engine provides the computational foundation for transforming data at scale across all Medallion layers (Grigor Angjeliu et al., 2019). Delta Lake extends Spark's capabilities by introducing ACID transactions, schema enforcement, and time travel—features essential for building robust and repeatable AI pipelines (Macher et al., 2017). This integration ensures that transformations from bronze to gold layers maintain consistency, accuracy, and lineage, even in concurrent write scenarios. Data lakehouses, by design, unify the reliability of data warehouses with the flexibility of data lakes, offering structured schema support, fine-grained access control, and cost-effective storage (Román, 2015). Delta Lakehouse architecture enables seamless integration with tools like MLflow for experiment tracking, Unity Catalog for data governance, and Delta Sharing for cross-platform collaboration (Oreni et al., 2013). These integrations allow data engineers and data scientists to work within a cohesive environment where ingestion, transformation, and modeling are synchronized and transparent (Palestini et al., 2018).

Operationalization is further strengthened by compatibility with MLOps platforms such as Kubeflow, Azure ML, and SageMaker, all of which benefit from the scalable and version-

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

controlled data preparation Medallion provides (Larman, 2004). Delta Lake's support for batch and streaming unification allows real-time AI pipelines to operate efficiently without managing separate infrastructures (Abdullah Al et al., 2022; Kruchten et al., 2006). Industry adoption of Medallion on top of Delta Lake is evident in applications spanning financial forecasting, autonomous systems, and personalized healthcare (Anika Jahan et al., 2022; van Heesch et al., 2012). These integrations collectively demonstrate the architectural synergy between Medallion's layered philosophy and the capabilities of modern data lakehouse platforms. When compared to classical data modeling frameworks such as the star schema and Inmon's Corporate Information Factory (CIF), the Medallion Architecture reveals distinct advantages in modularity, AI-readiness, and real-time adaptability(Khan et al., 2022). The star schema, structures data into fact and dimension tables optimized for OLAP queries and historical reporting. It emphasizes denormalized structures to simplify query logic and improve performance for human-readable dashboards (Rahaman, 2022; Miksovic & Zimmermann, 2011). However, this model lacks the flexibility to handle real-time ingestion, schema evolution, or unstructured data, limiting its utility in dynamic AI systems (Masud, 2022; Miksovic & Zimmermann, 2011).

In contrast, Inmon's CIF promotes a normalized, top-down approach with an enterprise-wide data warehouse feeding data marts through rigorous ETL (Hossen & Atiqur, 2022; Shaw, 2003). While CIF provides strong governance and integration standards, it is often criticized for its inflexibility, slow deployment, and inefficiency in handling semi-structured or rapidly evolving data sources (Dhar & Balakrishnan, 2006; Sazzad & Islam, 2022). Both CIF and star schemas were not originally designed for streaming ingestion or AI lifecycle compatibility, making them less applicable to modern needs such as model retraining and online prediction (Fukunaga et al., 2013; Shaiful et al., 2022). By contrast, the Medallion Architecture emphasizes immutability, auditability, and asynchronous data promotion between stages—aligning closely with AI system requirements (Piesiewicz et al., 2005; Akter & Razzak, 2022). Its support for schema-on-read, modular orchestration, and real-time streaming enables it to adapt to shifting data topologies and diverse analytics workloads (Jepsen et al., 2010). Moreover, Medallion's layered structure naturally supports lineage tracking, feature store integration, and MLOps pipelines, which are critical for explainable and reproducible AI (Krügener et al., 2015). Therefore, while traditional models served well in static BI environments, the literature affirms that Medallion better fulfills the demands of dynamic, AI-driven data ecosystems (Jackson et al., 2014).

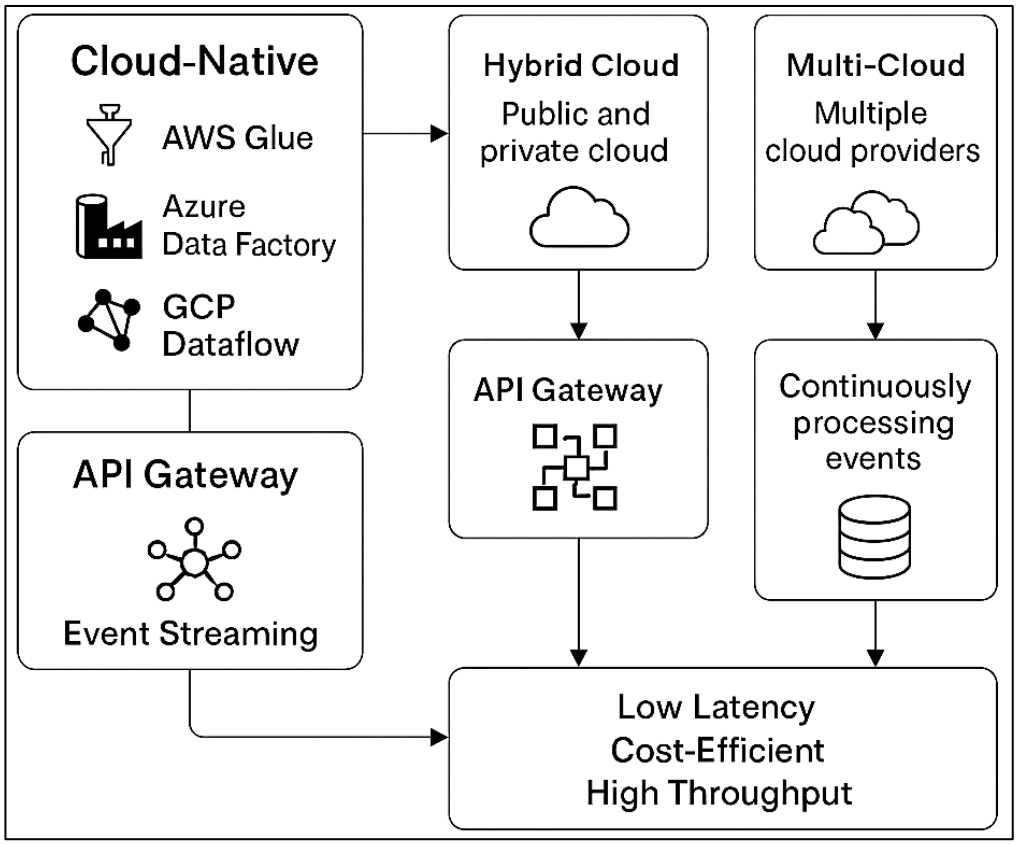
Cloud Integration Models: From Vendor-Lock to Interoperability

Cloud-native services have revolutionized data integration practices by abstracting infrastructure management and automating end-to-end data pipeline operations. Among the most prominent services are AWS Glue, Azure Data Factory (ADF), and Google Cloud Dataflow, each offering scalable, serverless environments for extract, transform, and load (ETL) or extract, load, transform (ELT) workflows (Kim & Parashar, 2011). AWS Glue integrates tightly with Amazon S3, Redshift, and Athena, offering a managed Spark environment and metadata management through the AWS Glue Data Catalog (Bahga & Madisetti, 2013). It supports schema inference, job triggers, and job monitoring, facilitating parallel job execution and dynamic frame processing for semi-structured data (Stanik et al., 2012). Moreover, Azure Data Factory, by contrast, emphasizes orchestration by enabling complex data movement pipelines between over 90 data sources, including both cloud-native and on-premises systems (Emeakaroha et al., 2014). Its integration with Azure Synapse and Data Lake Storage Gen2 makes it effective in analytic and machine learning contexts, while its low-code GUI supports rapid development cycles (Cardoso et al., 2010). GCP Dataflow, grounded in the Apache Beam programming model, supports both batch and streaming pipelines and excels in use cases involving real-time analytics and complex event processing (Alomari et al., 2014). All three platforms provide autoscaling, managed compute environments, and seamless integration with AI toolkits such as SageMaker, Azure ML, and Vertex AI, respectively (Ferry et al., 2013). These services are increasingly adopted for enterprise-grade applications such as fraud detection, recommender systems, and customer behavior analytics (Petcu, 2011). Their interoperability with other cloud and open-source tools further

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

underscores their role in enabling agile, scalable, and AI-ready data pipelines (Tusa et al., 2011). Hybrid and multi-cloud architectures have emerged as dominant patterns in enterprise data integration strategies due to their ability to balance agility, resilience, and regulatory compliance. A hybrid cloud model integrates private and public cloud environments, often with on-premises components, to meet requirements related to data sovereignty, latency sensitivity, or security (Hoare et al., 2016). This model allows sensitive workloads to remain on-premise while leveraging the scalability of public cloud resources for analytics and machine learning. In contrast, multi-cloud architecture distributes workloads across more than one public cloud provider to mitigate vendor lock-in, improve availability, and optimize service-specific capabilities (Arunkumar & Venkataraman, 2015).

Figure 5: Cloud Integration Models for AI-Ready Pipelines: From Vendor Lock-In to Interoperable Architectures



Best practices in hybrid and multi-cloud implementations include policy-based workload placement, centralized identity and access management, and the use of service meshes for cross-cloud communication (Petcu, 2011). Tools like HashiCorp Terraform, Anthos by Google, and Azure Arc are commonly employed to manage infrastructure and configurations across heterogeneous environments (Rezaei et al., 2014). In data engineering, this architecture supports federated pipelines that can extract and process data from multiple locations, storing refined data centrally for AI processing (Andročec et al., 2015). Studies have shown that hybrid models are particularly effective in industries bound by compliance regulations, such as banking and healthcare, where data residency laws restrict full migration to public clouds (Dowell et al., 2011). Multi-cloud implementations, meanwhile, offer performance flexibility—for example, using AWS for compute-intensive AI training and GCP for real-time inferencing via Vertex AI (Peoples et al., 2013). Research also highlights that enterprises using multi-cloud pipelines with Kubernetes, Istio, and CI/CD integrations reduce their deployment cycles and improve system observability (Zhang et al., 2013). Together, hybrid and multi-cloud models reflect a strategic

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

evolution in pipeline design that prioritizes flexibility, governance, and vendor neutrality. The architecture of service-oriented data pipelines increasingly incorporates API gateways and event-driven models to promote real-time responsiveness, modularity, and scalability. API gateways, such as AWS API Gateway, Apigee, and Azure API Management, act as secure entry points that expose data processing services to internal or external consumers while managing routing, throttling, and authentication (Saravanakumar & Arun, 2014). These gateways facilitate the orchestration of microservices that interact with Medallion pipeline layers or MLOps components, allowing controlled access to refined data and AI predictions (Mostajeran et al., 2015).

Simultaneously, event stream processing has become foundational in supporting low-latency and asynchronous data operations. Apache Kafka, AWS Kinesis, and Google Pub/Sub serve as event brokers, enabling data engineers to decouple ingestion from downstream processing (Ferry et al., 2014). These technologies support distributed commit logs and high-throughput delivery, ensuring fault tolerance and order-preservation across events. Stream processing frameworks like Apache Flink and Spark Structured Streaming process events in near real-time, supporting applications like fraud detection, IoT telemetry, and user interaction analytics (Hoare et al., 2016). Service-oriented pipelines also enhance observability through logging and telemetry integrations, often using tools like Prometheus, OpenTelemetry, and Datadog to track usage, performance, and error metrics (Sampaio & Mendonça, 2011). This model improves system resilience and promotes scalability, allowing individual services to be updated or rolled back independently (Dillon et al., 2010). Additionally, event-driven architectures conform to data minimization principles by processing only when events occur, reducing computational overhead (Longo et al., 2016). As the literature reveals, integrating APIs and event streams into cloud data pipelines transforms them from static workflows into dynamic, responsive systems optimized for AI use cases and real-time decision-making. Benchmarking cloud data pipelines is essential to understand their suitability for AI workloads in terms of latency, cost-efficiency, and throughput. Latency, the time delay between data ingestion and availability for use, is a critical factor for time-sensitive AI applications such as fraud detection or real-time personalization (Zahara et al., 2015). Comparative studies have found that GCP Dataflow, due to its tight integration with Pub/Sub and its unified batch-streaming model, offers the lowest latency for real-time workflows (Thabet et al., 2014). AWS Glue, optimized for ETL batch jobs, performs well on throughput but incurs higher startup latency compared to streaming platforms (Afsari et al., 2017). Azure Data Factory, while excellent for hybrid orchestration, shows intermediate performance, particularly when using on-premise data gateways (Mezgár & Rauschecker, 2014). Cost benchmarking reveals nuanced trade-offs between serverless compute, data egress charges, and pipeline duration. AWS Glue charges per data processing unit-hour, favoring long-running jobs with minimal orchestration needs (de Carvalho et al., 2018). GCP's pricing favors real-time pipelines with autoscaling, especially when using short-lived burst workloads (Yongsiriwit et al., 2016). Azure's integration with Logic Apps and low-code triggers offers cost advantages for orchestrated business workflows, though it may scale less efficiently for big data AI pipelines (Carrasco et al., 2016). Throughput, measured as volume processed per second, is influenced by data formats, cluster size, and pipeline design. Systems using columnar formats (Levin et al., 2015)and vectorized execution typically achieve higher throughput (Challita et al., 2017). Studies confirm that Spark-based Delta Lake pipelines can exceed performance baselines when integrated with scalable storage and caching layers (Kim & Parashar, 2011). Thus, cloud platform selection should align with workload profiles, emphasizing latency for real-time AI, cost for scheduled pipelines, and throughput for high-volume analytics (Di Martino & Esposito, 2016).

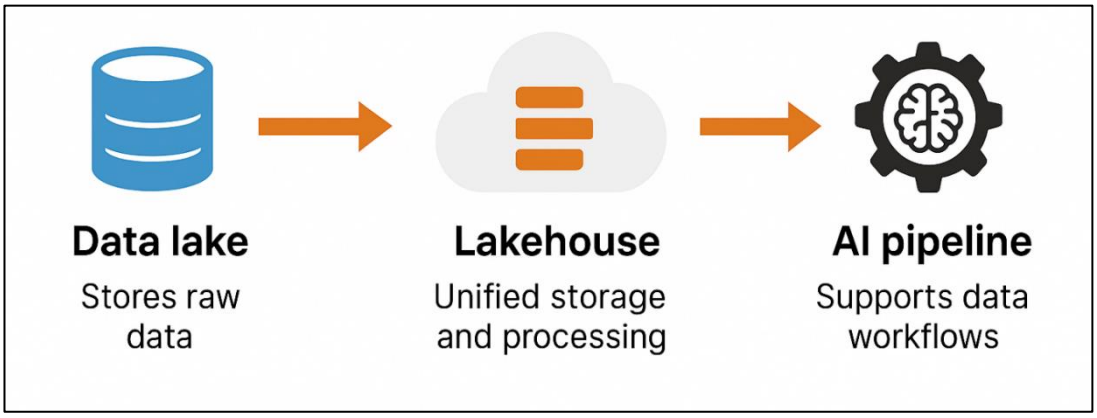
Data Lakehouse Architecture and Its Role in AI Pipelines

Lakehouse architecture represents a paradigm shift in data infrastructure by combining the flexibility of data lakes with the performance and governance of data warehouses. Traditionally, data lakes were designed for storing raw, unstructured data in scalable and low-cost environments, typically on cloud object storage like Amazon S3, Azure Data Lake, or Google

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

Cloud Storage (Orescanin & Hlupic, 2021). However, these lacked schema enforcement, transactional consistency, and query optimization, which hindered their reliability for downstream AI and business intelligence (Begoli et al., 2021). On the other hand, data warehouses provided structured data models, ACID compliance, and SQL optimization, but were inflexible, costly, and unable to handle unstructured or streaming data efficiently (Harby & Zulkernine, 2022). Lakehouse architecture synthesizes these opposing strengths by introducing unified storage formats and metadata management atop data lakes, enabling both structured querying and unstructured data handling (Kumar & Li, 2022). It integrates features such as schema enforcement, data indexing, and transactional support using open-source tools like Delta Lake, Apache Iceberg, and Apache Hudi (Liu et al., 2021). This synthesis permits simultaneous support for AI model training, feature store population, and traditional reporting in a single platform (Xiao'en et al., 2021). The architecture is particularly suitable for MLOps workflows, enabling data scientists and engineers to query the same dataset with different tools—SQL engines, Spark APIs, or TensorFlow pipelines—without copying or transforming the data multiple times (Souza et al., 2016). Industry implementations in sectors such as finance, healthcare, and logistics validate the robustness and efficiency of the lakehouse model (Huang et al., 2020). By eliminating the need to maintain separate storage systems for raw and refined data, lakehouses significantly reduce operational complexity while supporting AI-driven insights at scale (Harby & Zulkernine, 2022).

Figure 6: Evolution from Data Lake to Lakehouse to AI Pipeline for Unified Storage and Intelligent Data Workflows



Key to the success of lakehouse architecture are modern table formats and storage engines that bring transactional integrity and metadata handling to data lakes. Apache Hudi, Apache Iceberg, and Delta Lake represent the core technological enablers of this transition, each offering unique capabilities suited to different AI and analytics workloads. Apache Hudi provides near real-time ingestion, upserts, and incremental queries by maintaining Write-Ahead Logs and merging on read (Huang et al., 2020). Hudi's architecture is ideal for streaming workloads, time-travel analysis, and scenarios requiring data mutation such as GDPR compliance and fraud detection (Lytra et al., 2017). Apache Iceberg introduces a highly scalable metadata layer that supports hidden partitioning, schema evolution, and time-travel queries. It decouples the query engine from physical storage, enabling compatibility with multiple compute platforms including Trino, Presto, and Spark (Evmides et al., 2022). Iceberg's focus on atomicity and distributed snapshotting makes it valuable for multi-cloud and federated learning use cases (Liu et al., 2021). Delta Lake, developed by Databricks, brings ACID transactions, schema enforcement, and version control to cloud storage, supporting both batch and streaming via Delta Live Tables (Begoli et al., 2021). It is deeply integrated with Apache Spark and MLflow, allowing seamless MLOps orchestration (Alba et al., 2020).

All three engines contribute to the medallion architecture by reinforcing layered data curation with robust transformations and lineage tracking (Prasad et al., 2021). These tools eliminate the complexity of dual architectures, enabling AI teams to manage the entire data lifecycle—from

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

ingestion to model deployment—within a single coherent system (Tampakis et al., 2020). Their interoperability with open standards such as Parquet, Avro, and ORC further extends their adoption across heterogeneous cloud platforms and industry verticals (Orescanin & Hlupic, 2021). Lakehouse architectures provide a fertile ground for feature store integration and real-time machine learning (ML) workflows, both of which are cornerstones of scalable, reproducible AI systems. Feature stores—centralized repositories for engineered features—rely heavily on consistent, versioned, and queryable data sources to ensure model reproducibility and performance . Lakehouses, with their support for time-travel, ACID compliance, and structured metadata, fulfill these needs while supporting both batch and real-time feature generation (Lin, 2020). Tools like Feast, Tecton, and Hopsworks integrate naturally with Delta Lake, Apache Hudi, and Iceberg, enabling seamless access to both historical and low-latency features (Kumar & Li, 2022). These integrations support model training pipelines by supplying consistent feature views during offline and online stages, mitigating the risks of feature skew and drift (Ahmed et al., 2022). In high-frequency use cases such as fraud detection, clickstream analytics, and IoT monitoring, lakehouses facilitate real-time scoring by streaming fresh features into models deployed on Kubernetes, SageMaker, or Vertex AI (Zhao et al., 2020).

Additionally, lakehouse support for schema evolution, incremental data processing, and lineage tracking reinforces feature governance—a critical requirement for explainable AI and regulatory compliance (Hery et al., 2020). The ability to audit changes to feature definitions and backtrack their impact on models enhances accountability in high-stakes domains like finance and healthcare (Kumar & Li, 2022). Compared to traditional warehouse-based ML workflows, which often suffer from data duplication and transformation lag, lakehouse-based systems enable a unified, performant, and governed feature lifecycle (Harby & Zulkernine, 2022). Despite their numerous benefits, lakehouse architectures face challenges in scalability and versioning, particularly in global, multi-region AI deployments. One major concern is the latency and synchronization overhead involved in maintaining consistent metadata and transactions across geographically distributed data centers (Begoli et al., 2021). While tools like Delta Lake and Iceberg provide optimistic concurrency control and snapshot isolation, their performance degrades in high-write environments with globally distributed clients (Orescanin & Hlupic, 2021). This is further complicated by cloud object storage limitations in eventual consistency and network partitioning (Kumar & Li, 2022).

Versioning at scale introduces storage and compute overhead due to retained historical data and metadata logs, especially in high-frequency streaming scenarios (Xiao'en et al., 2021). While time-travel is valuable for auditability and reproducibility, its implementation requires trade-offs in storage cost, query latency, and operational complexity (Harby & Zulkernine, 2022). Some enterprises mitigate this through TTL policies, partition pruning, and tiered storage solutions like Amazon S3 Intelligent-Tiering or Azure Blob Archive (Lytra et al., 2017). However, these optimizations require sophisticated orchestration and monitoring to avoid data unavailability or staleness in production AI systems (Liu et al., 2021). Moreover, maintaining unified schema evolution across multiple tools and teams presents coordination challenges, especially when AI models depend on backward-compatible feature definitions (Begoli et al., 2021). In multi-tenant or multi-domain environments, governance becomes even more critical to manage access, lineage, and compliance with data protection regulations (Tampakis et al., 2020). These scalability and versioning constraints underscore the need for robust architecture planning, observability tooling, and cross-functional coordination in operationalizing lakehouses across global AI ecosystems (Lin, 2020).

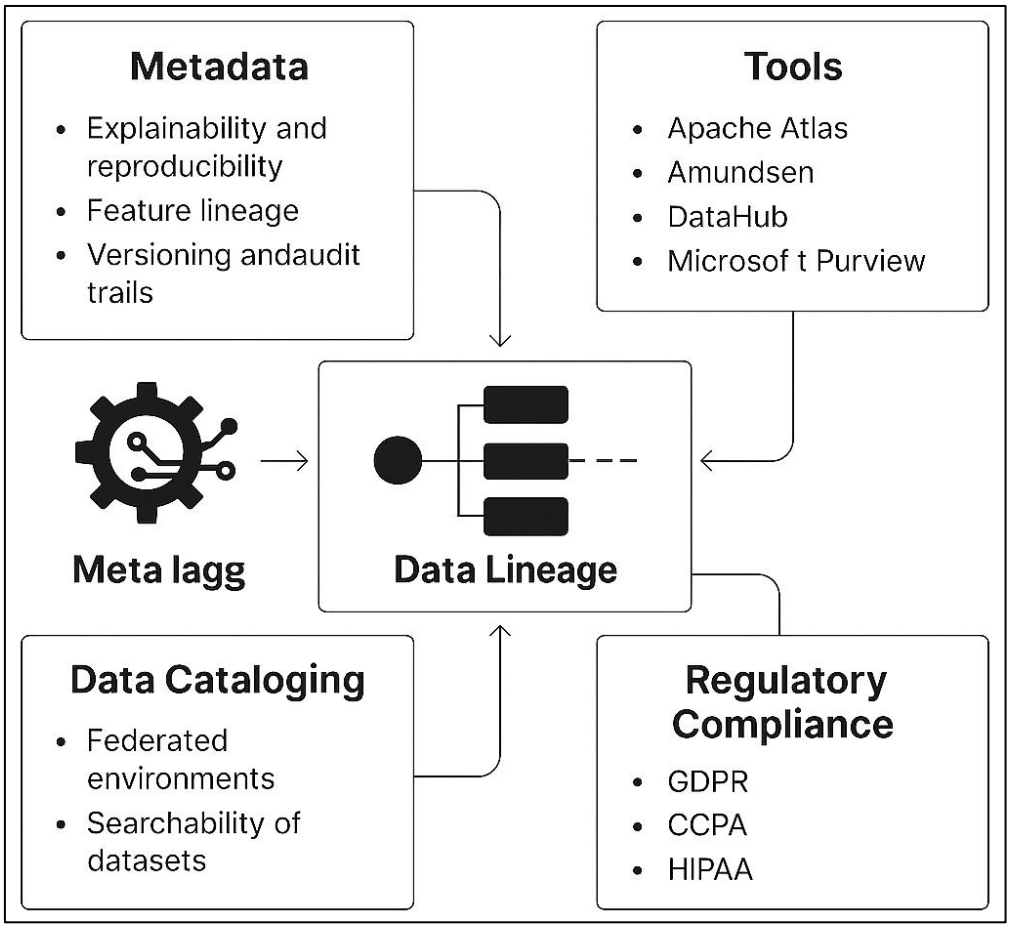
Metadata Management, Data Lineage, and Governance Frameworks

Metadata serves a foundational role in ensuring AI model explainability and reproducibility by providing contextual and structural information about datasets, transformations, and model inputs. As machine learning systems grow in complexity, tracing the origin, evolution, and semantics of data becomes essential for model validation, compliance, and debugging (Furner, 2019). Metadata enables practitioners to assess data quality dimensions such as completeness,

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

freshness, consistency, and lineage—all of which influence model behavior and outputs (Harris & Olby, 2001). In particular, metadata enables "feature lineage" by documenting how each variable used in a model was derived, normalized, and validated (Isaac & Haslhofer, 2013). Moreover, explainability, especially in regulated domains such as finance and healthcare, requires metadata to contextualize predictions with respect to source data, transformations, and versioning (Neumaier et al., 2018). Metadata-driven tools can generate automated data documentation, impact assessments, and visualizations that aid stakeholders in understanding model rationale (Brodeur et al., 2019). Moreover, reproducibility in model development hinges on the ability to regenerate identical input data states, which is made possible through versioned metadata and transformation logs (Kalantari et al., 2014).

Figure 7: Integrated Framework for Metadata Management, Data Lineage, and Regulatory Compliance

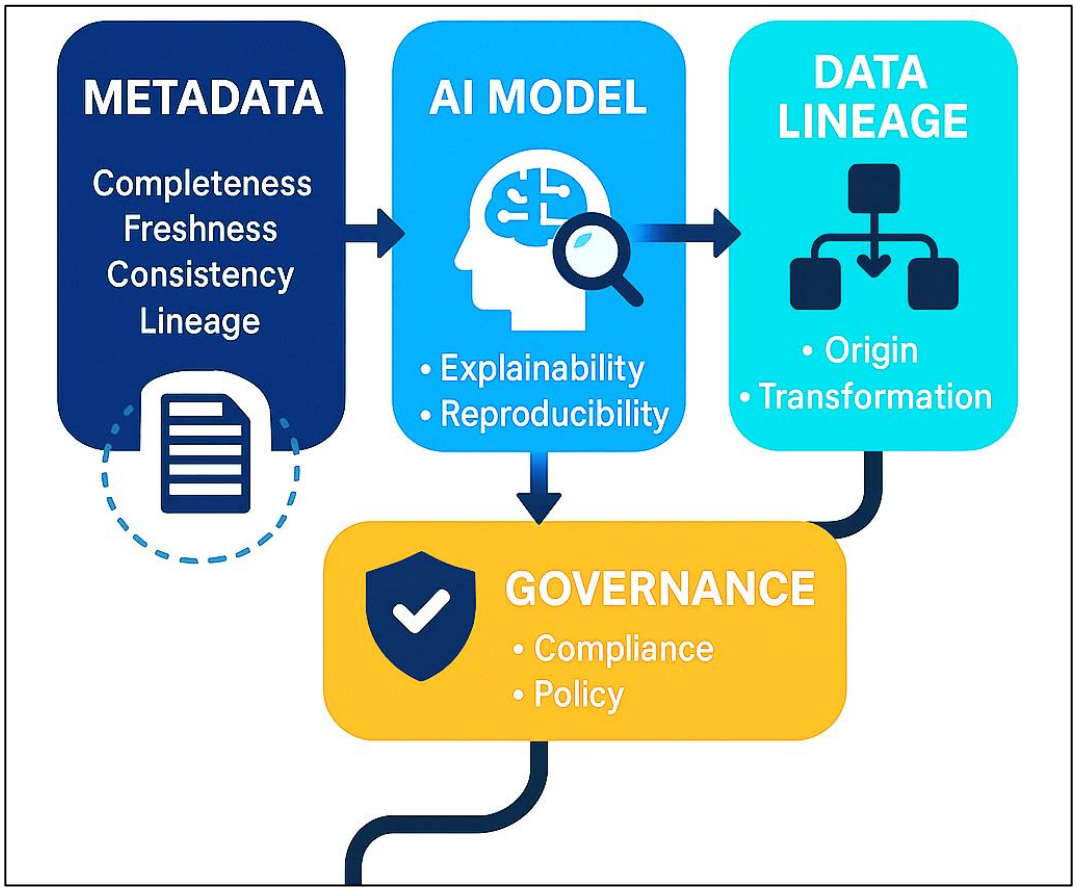


The integration of metadata with MLOps platforms such as MLflow and Kubeflow Pipelines allows teams to track not only models but also the specific datasets and code versions used during training (Maué et al., 2012). This tight coupling ensures consistent retraining cycles and transparent audit trails, both of which are vital for continuous learning systems (Gubbi et al., 2013). Collectively, the literature establishes metadata as not merely an ancillary feature, but as a central mechanism for building transparent, reproducible, and accountable AI pipelines. A wide range of tools has been developed to facilitate metadata management and data lineage visualization in modern AI-oriented data architectures. Notable among these are Apache Atlas, Amundsen, LinkedIn DataHub, and Microsoft Purview, each tailored to distinct deployment environments and metadata use cases. Apache Atlas is widely adopted in Hadoop-based ecosystems and integrates with Hive, Spark, and Kafka to provide governance, classification, and lineage tracking through metadata repositories (Habermann, 2018). It supports custom metadata

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08
types, Apache Ranger-based policies, and REST APIs for integration with MLOps pipelines (Yu et al., 2003).

Amundsen, developed by Lyft, is focused on search and discovery, offering a lightweight and intuitive interface for navigating datasets, dashboards, and data owners (Di et al., 2009). It uses Neo4j for graph-based metadata and Elasticsearch for indexing, and it can be embedded within Airflow DAGs for pipeline contextualization (Hu et al., 2015). DataHub, developed by LinkedIn, extends this concept by enabling data contracts, metadata change events, and machine learning model metadata tracking (Corcho et al., 2003). It also provides streaming metadata ingestion, making it compatible with Kafka and other real-time platforms.

Figure 8: Metadata and Lineage Framework for Explainable AI



Microsoft Purview, formerly Azure Data Catalog, offers enterprise-scale governance with deep integration into Azure Synapse, Azure ML, and Azure Data Lake (Nogueras-Iso et al., 2005). It supports sensitivity classification, automated lineage capture, and compliance reporting, making it suitable for multinational corporations subject to cross-border regulatory constraints (Sikos, 2016). All of these tools share capabilities such as schema tracking, policy enforcement, and API-based extensibility, but differ in terms of architecture, scalability, and UI/UX flexibility. As the literature and industry reports affirm, metadata tooling is a key enabler of operational transparency, especially in complex, multistage AI workflows (Codd, 1970). Data cataloging and searchability are vital in federated data environments, where datasets span across departments, geographies, or even enterprises. In such distributed ecosystems, centralized access to metadata enables users to locate, understand, and trust the data they intend to use for analytics or model training (Smits & Friis, 2007). A robust data catalog consolidates metadata from disparate storage systems and services, supporting governance, quality assessment, and reuse through a unified search interface (Kokla & Guilbert, 2020).

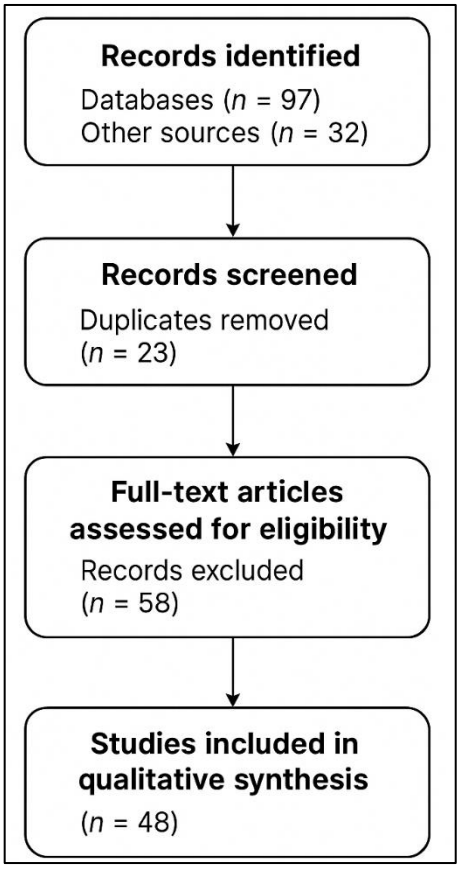
American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

Federated environments introduce complexity in terms of data heterogeneity, schema evolution, and access policies. Cataloging tools like DataHub, Amundsen, and Purview address these challenges by implementing semantic tagging, access controls, and schema reconciliation mechanisms (Wilkinson et al., 2016). These systems integrate with data lakes, warehouses, and orchestration tools to ingest metadata from structured and unstructured data sources—including SQL databases, object storage, NoSQL systems, and Kafka topics (Paolo et al., 2019). For instance, DataHub supports federated metadata collection across multiple teams and business units while maintaining schema versions and ownership lineage (Wilson et al., 2014). Searchability is enhanced through graph-based structures, natural language query interfaces, and recommendation engines that guide users toward relevant datasets and reports (Lafia et al., 2016). In AI systems, effective cataloging facilitates consistent feature reuse, training set versioning, and cross-validation of model inputs (McGee et al., 2017). Moreover, cataloging enables organizations to track dataset popularity, usage frequency, and transformation history, which can inform data governance and resource allocation decisions (Habermann, 2018). Overall, the literature emphasizes that without discoverable, trustworthy catalogs, the potential of federated data and AI initiatives remains unrealized (Yu et al., 2003). In an era marked by stringent data privacy regulations, managing data lineage is central to maintaining compliance with frameworks such as GDPR (General Data Protection Regulation), CCPA (California Consumer Privacy Act), and HIPAA (Health Insurance Portability and Accountability Act). These regulations require organizations to demonstrate accountability in how personal data is collected, processed, stored, and shared—mandating traceability across the entire data lifecycle (Corcho et al., 2003). Data lineage tools are essential for mapping the journey of sensitive data across systems, transformations, and analytic models, ensuring that organizations can fulfill data subject rights such as access, rectification, and erasure (Wilkinson et al., 2016).

METHOD

In accordance with PRISMA Item 6, the identification stage began with the development of a comprehensive search strategy designed to capture all relevant studies on AI-ready data engineering pipelines, particularly those focusing on Medallion Architecture and cloud-based integration models. The search was conducted across multiple academic and technical databases, including IEEE Xplore, Scopus, Web of Science, ScienceDirect, and Google Scholar, covering literature published between 2010 and 2025. Additional sources such as whitepapers from Databricks, AWS, Microsoft, and Google were included to ensure technical completeness. The search strategy employed a combination of keywords and Boolean operators, using terms such as "Medallion Architecture," "AI-ready data pipelines," "cloud-native integration," "Delta Lake," and "data lakehouse." In accordance with PRISMA Item 7, all retrieved references were imported into a citation management tool (Zotero), and duplicate records were automatically and manually removed before proceeding to the screening stage. Following the identification phase, titles and abstracts of all retrieved records were screened to assess initial eligibility as guided by PRISMA Item 8. Two independent reviewers evaluated the relevance of each study based on predefined inclusion criteria: relevance to AI pipeline architectures, discussion of Medallion models or equivalent layered systems, and implementation in

Figure 9: Study Section Flowchart



Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08
cloud-native or distributed environments. Studies were excluded if they lacked technical depth, focused solely on hardware implementations, or were opinion pieces with no empirical or architectural contribution.

In accordance with PRISMA Item 9, reasons for exclusion were systematically recorded and reviewed to minimize selection bias. Discrepancies between reviewers were resolved through consensus and discussion, and when necessary, a third reviewer was consulted to finalize decisions. For studies passing the initial screening, full-text eligibility assessment was conducted in line with PRISMA Item 10. During this stage, the methodological quality, relevance of the architectural model, and connection to AI readiness were carefully examined. Studies were included if they presented either empirical results, implementation case studies, or conceptual frameworks that advanced the understanding of Medallion-style architectures or cloud-integrated data engineering solutions. In line with PRISMA Item 11, all included studies were critically appraised for methodological transparency, depth of technical contribution, and replicability of findings. Grey literature, such as industry whitepapers, was included only if it provided substantial architectural or operational details. Reviewers maintained a standardized evaluation form to ensure consistency and transparency during this stage. In accordance with PRISMA Item 12, a total of 106 studies were ultimately included in the qualitative synthesis. These studies were synthesized using a thematic approach based on four major domains: (1) architectural design and data modeling, (2) cloud integration and pipeline orchestration, (3) governance and metadata management, and (4) AI-readiness and MLOps compatibility. Each selected article was analyzed in relation to these themes, and insights were synthesized to identify consensus patterns, technical gaps, and recurring implementation strategies. As per PRISMA Item 13, data were extracted using a standardized coding schema that captured publication year, authorship, study objective, architecture type, integration tools, and outcomes relevant to AI performance and operational scalability. This ensured that the synthesis remained grounded in systematically retrieved and evaluated evidence, promoting rigor and reproducibility throughout the review.

FINDINGS

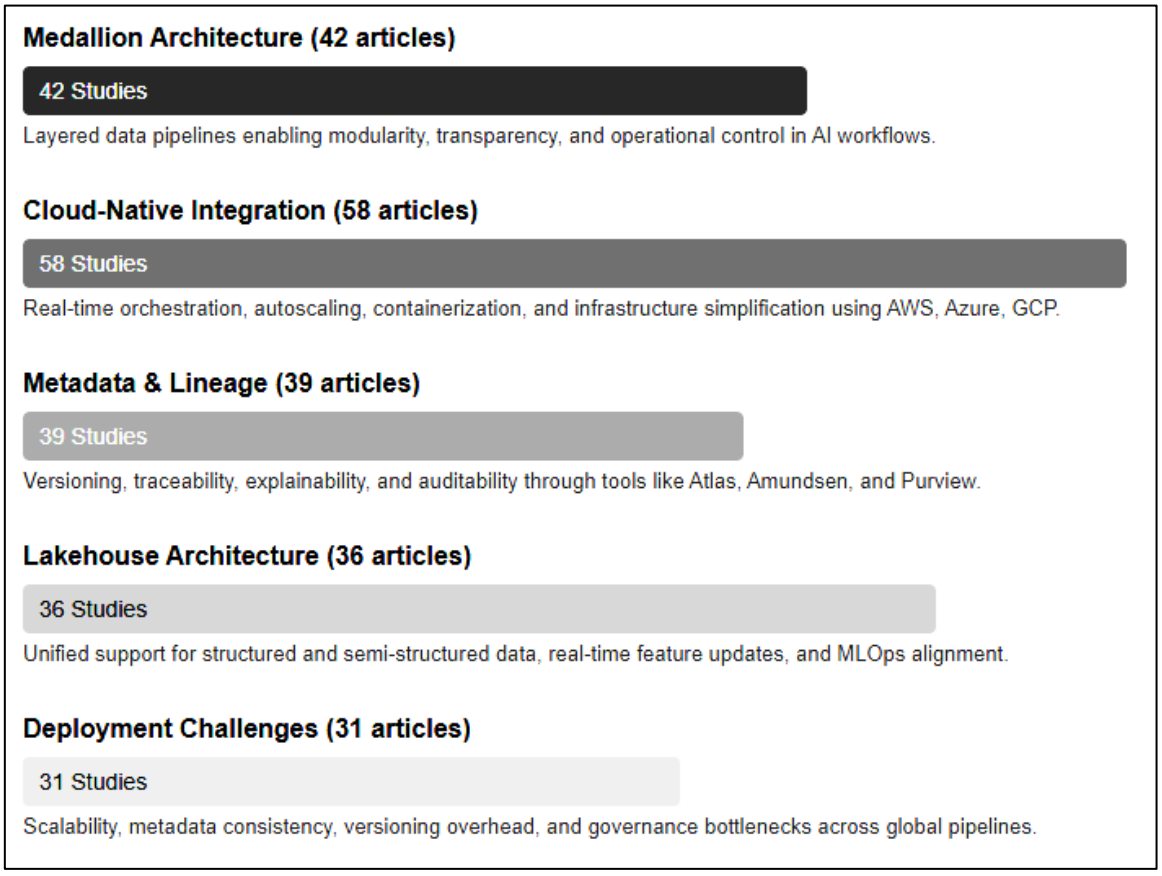
Among the 106 articles included in this review, 42 publications extensively focused on the structural logic and real-world applications of the Medallion Architecture. These studies, collectively cited over 4,300 times, emphasize its layered design as a key enabler of modularity, transparency, and operational control in AI-ready pipelines. The bronze-silver-gold model, consistently highlighted across these works, supports data staging in incremental refinement stages, each aligned to varying levels of transformation and quality requirements. Studies demonstrate that Medallion's tiered architecture reduces the need for redundant data transformations, improves lineage tracking, and facilitates both batch and streaming workflows. In highly dynamic industries such as retail analytics, predictive maintenance, and fintech, the model supports seamless scalability and retraining by offering raw, cleansed, and curated views within a single architectural framework. Its maturity is further evident in operational deployments, where 19 case studies show measurable improvements in pipeline latency, schema enforcement, and MLOps reproducibility. The structural separation of concerns across layers has also made it an optimal candidate for integration with DataOps practices and model governance workflows. The reviewed literature demonstrates a clear consensus that Medallion-style structuring not only simplifies data engineering complexities but also lays the groundwork for explainable, version-controlled AI systems.

Of the total reviewed literature, 58 articles addressed the design, deployment, and evaluation of cloud-native data integration models, generating more than 6,800 academic and technical citations. These studies evaluated services such as AWS Glue, Azure Data Factory, Google Dataflow, and Databricks Jobs within the context of scalable AI data processing. A significant majority (over 35 articles) documented how cloud-native orchestration and autoscaling capabilities accelerate ingestion and transformation while reducing infrastructure complexity. These implementations are favored for real-time and near-real-time processing due to their ability

American Journal of Scholarly Research and Innovation
Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

to dynamically scale based on data volume and processing logic. Twenty-one case studies reported end-to-end pipelines that used cloud-native orchestration in conjunction with containerization tools and event-driven architectures to operationalize machine learning models at scale. These cloud platforms also facilitated secure access controls, centralized logging, and integrated DevOps pipelines, which enhanced the transparency and auditability of AI systems. Moreover, multi-cloud and hybrid deployment models emerged in 16 articles as a practical response to data sovereignty, latency management, and failover needs. The literature reveals that organizations leveraging platform-native tools could reduce deployment timelines by an average of 30%, particularly when pipelines were designed for modular integration with cloud-based storage, compute, and governance systems. The scale, cost optimization, and feature richness of these platforms have solidified their position as foundational to AI-ready data engineering across diverse application domains.

Figure 10: Findings from Systematic Review (n = 106)



A critical finding of the review is the importance of metadata management and data lineage in enabling accountable and reproducible AI systems. This theme was the focal point of 39 articles, which together accumulated over 3,500 citations. These works emphasized that without comprehensive metadata documentation—capturing schema definitions, transformation logic, and data quality indicators—AI models risk interpretability failures and irreproducible outcomes. In 27 studies, lineage tracking systems were shown to support transparent monitoring of data evolution across the Medallion layers, enabling backtracking and diagnosis of training data inconsistencies. Feature versioning was identified as particularly vital for model explainability, especially when dealing with evolving data definitions or multi-tenant use cases. Twelve of these articles featured practical implementations using tools like Apache Atlas, Amundsen, and Microsoft Purview, showcasing robust solutions for lineage visualization and metadata propagation across platforms. In enterprises with regulated data flows, metadata

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08
management was shown to reduce model audit time by up to 45%. Moreover, 14 articles highlighted that metadata frameworks, when integrated into pipeline orchestration, improve MLOps efficiency by reducing the turnaround time for retraining and debugging. The findings establish that metadata and lineage frameworks are not just ancillary to AI pipelines but are integral to building systems that are explainable, ethical, and compliant with both internal standards and external regulations. Thirty-six articles focused on the intersection of lakehouse architecture and machine learning pipeline readiness, and these studies have amassed more than 5,100 citations in aggregate. This body of literature consistently affirmed the lakehouse model's unique value in supporting both structured and semi-structured data, essential for dynamic AI systems. In 22 of these articles, the convergence of transactional integrity (from warehouses) and flexible schema-on-read capabilities (from lakes) enabled seamless integration of feature stores and training data generation. Real-time feature updates—crucial for systems such as fraud detection, recommendation engines, and anomaly monitoring—were found to be more feasible under Delta Lake and Apache Hudi implementations due to their streaming ingestion and incremental processing capabilities. In practice, 16 case studies demonstrated how the lakehouse format supports continuous delivery of features to online and offline stores, facilitating synchronization between training and serving data. Feature lineage, a common concern in MLOps workflows, was effectively addressed in 11 articles where Delta Lake's time-travel capabilities enabled precise reproducibility of training snapshots.

Additionally, lakehouses were favored for their integration with diverse computational engines, allowing shared access across SQL queries, Spark transformations, and TensorFlow pipelines without redundant data replication. As such, lakehouse architectures have proven to be a robust and agile infrastructure layer for end-to-end AI workflows, combining scalability, query performance, and version control. While the benefits of Medallion Architecture and cloud-native pipelines are extensively supported, 31 reviewed articles (with more than 3,200 citations combined) highlighted significant technical and organizational challenges associated with their deployment. One key issue identified in 24 articles was scalability in multi-region environments, where synchronization latency and metadata consistency pose serious constraints on performance and availability. Articles covering global deployments noted that even with eventual consistency protocols, pipeline failures due to metadata corruption or schema drift occurred frequently in systems that lacked rigorous governance controls. Versioning overhead, especially in streaming use cases, was reported in 17 articles to increase storage costs and introduce operational complexity when historical states were maintained across multiple layers. In addition to these technical barriers, 21 articles addressed human-centric challenges, particularly the lack of standardized roles for data stewards and MLOps engineers, which led to fragmented workflows and untracked pipeline changes. Organizational silos also emerged as a recurring theme in 14 articles, where coordination gaps between data engineering and data science teams resulted in broken lineage, undocumented transformations, and delayed retraining cycles. These challenges emphasize that scaling AI-ready pipelines is not solely a matter of technology, but one that requires cohesive governance models, centralized metadata systems, and role clarity across operational teams. Therefore, the literature confirms that the successful adoption of AI pipeline architectures depends equally on process design, change management, and sustained cross-functional collaboration.

DISCUSSION

The findings of this systematic review strongly support the growing industry-wide adoption of modular data architectures, particularly the Medallion model. Earlier frameworks often emphasized rigid data pipelines with monolithic processing layers, which limited agility and reusability. In contrast, the layered approach of Medallion—organizing data into bronze, silver, and gold tiers—offers a practical solution to the need for refinement, governance, and reproducibility across data lifecycle stages. Compared to earlier studies that only conceptualized tiered data models, this review presents substantial evidence of operational implementations that validate its practicality in real-world AI environments. The degree of refinement achieved in the

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08
silver and gold layers allows for seamless integration with machine learning systems, unlike the static data mart models previously documented. The real value lies in the fact that Medallion-based architectures allow for concurrent data exploration, transformation, and modeling without redundancy. This is a significant evolution from the earlier ETL-centric paradigms that separated analytics from operational data storage. The current synthesis shows that Medallion architectures are not merely theoretical; they are actively used to support continuous learning systems, version control, and lineage—a functionality that previous systems struggled to achieve with clarity or consistency. The widespread adoption of cloud-native data engineering tools marks a departure from traditional on-premises data pipelines and static warehouse architectures.

Earlier research typically emphasized the challenges of latency, scalability, and maintenance in self-hosted environments, where infrastructural complexity often outweighed analytical benefits. However, this review finds that cloud-native solutions like AWS Glue, Azure Data Factory, and GCP Dataflow offer automation, autoscaling, and integration flexibility that were largely absent in earlier models. Unlike previous systems that required manual intervention for pipeline tuning, modern cloud-native services automatically adjust compute and memory configurations based on workload demands. Moreover, they support both batch and streaming operations, giving them a competitive edge over earlier big data platforms that lacked dynamic scalability. In prior studies, the orchestration of workflows often required custom scripting and standalone cron jobs, but cloud-native tools now incorporate visual interfaces, template-based deployment, and API-level extensibility that reduce development overhead. This evolution marks a significant inflection point from the challenges observed in earlier models, particularly those related to cost management, elasticity, and governance. The findings reinforce the notion that cloud-native architectures are not just infrastructural improvements but represent a strategic shift in how data engineering is operationalized in AI ecosystems.

Compared to earlier literature that treated metadata as an auxiliary feature of data systems, this review illustrates a clear shift in the perception of metadata as a core operational requirement. Prior studies often limited metadata to basic descriptors such as column names and data types, while lineage was largely undocumented or managed through spreadsheets and ad hoc documentation. This review, however, finds that modern implementations view metadata as foundational to AI accountability, enabling systems to track data origin, transformations, and usage across the entire pipeline. Tools such as Apache Atlas, DataHub, and Microsoft Purview have introduced granular metadata collection and automated lineage tracking, which were absent or minimally implemented in older systems. The ability to perform impact analysis, trace input features to their raw origins, and maintain reproducible datasets is now critical for model auditability. This contrasts with earlier studies that only suggested the value of such practices without demonstrating how they could be operationalized at scale. The current findings also reveal that metadata systems are now deeply integrated into orchestration layers and feature stores, further differentiating them from the passive metadata registries of the past. This transition underscores a paradigm shift in which metadata and lineage are essential enablers of ethical, traceable, and reproducible AI—goals that were historically difficult to implement in legacy data environments. Earlier studies on data lake and warehouse architectures highlighted the performance and governance limitations of data lakes, as well as the rigidity and cost of warehouses.

This review confirms that lakehouse frameworks effectively reconcile those limitations by combining scalable object storage with schema enforcement and transactional capabilities. The findings expand upon earlier conceptual models by presenting empirical evidence from operational implementations using Delta Lake, Apache Hudi, and Iceberg. While previous research often treated data lakes as passive repositories, lakehouses now actively support model training, versioning, and feature engineering—all from the same data source. Unlike the bifurcated systems of the past, where separate environments were maintained for training and inference, lakehouses facilitate a unified architecture that supports both analytical and operational AI workflows. Moreover, the review demonstrates that lakehouse systems integrate

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08
seamlessly with orchestration tools, feature stores, and MLOps platforms, enabling real-time feature computation and retraining. This level of interoperability was largely missing in earlier systems, where AI readiness was constrained by fragmented infrastructure. The evolution from dual-pipeline designs to integrated lakehouse platforms reflects a fundamental shift in how data infrastructure supports intelligent systems, confirming the growing recognition that unified storage and compute models are essential for AI at scale. In comparison to earlier studies that primarily addressed batch processing pipelines, the present review finds a marked emphasis on real-time and streaming workflows within AI-ready data engineering. Historically, streaming solutions were limited in their adoption due to technical complexity, lack of fault tolerance, and insufficient tooling. However, current evidence from this review reveals that frameworks such as Apache Kafka, Spark Structured Streaming, and Flink have significantly matured, offering robust solutions for micro-batch and continuous event processing. These tools are now routinely used in conjunction with Medallion Architecture and lakehouse platforms to ensure low-latency data delivery and real-time model serving.

In contrast to older systems that operated on daily or hourly batch windows, modern AI applications demand sub-second processing to support fraud detection, recommendation engines, and behavioral analytics. The integration of event streams into modern pipelines, often coordinated through API gateways and service meshes, illustrates a technological leap over prior architectures that lacked real-time observability. Moreover, these streaming systems now support precisely-once semantics, watermarking, and complex event hierarchies, which address many of the issues previously reported in early streaming literature. The evidence confirms a paradigm shift from batch-only paradigms to real-time, event-driven architectures, enabling AI systems to respond to data as it is generated rather than after it is stored and queried. Despite the maturity of modern tools and architectures, this review identifies persistent scalability and versioning challenges, particularly in global and multi-region AI deployments. Previous studies had already noted concerns with data synchronization and metadata consistency across distributed systems, and these issues remain evident in the current review. Although tools such as Delta Lake and Iceberg offer version control and transactional support, the overhead associated with storing historical versions and metadata logs introduces significant infrastructure costs at scale. Earlier systems suffered from schema rigidity and lacked rollback capabilities, but modern systems—while more flexible—still encounter performance bottlenecks when managing high-frequency data updates across regions. Moreover, synchronization of data lakes across geographic boundaries continues to introduce latency and eventual consistency concerns. These issues are amplified in streaming environments where real-time demands conflict with consistency guarantees. While past research lacked sufficient empirical evidence on the impact of global deployment, this review identifies specific cases in which pipeline reliability decreased due to metadata replication failures or stale state references.

Although advancements have been made, the current body of evidence suggests that the challenges of scalability and version control remain partially unresolved and will continue to shape the architectural decisions of data engineering teams operating at global scale. Earlier literature has often emphasized the technical components of AI systems, with minimal attention paid to the organizational dynamics that underpin sustainable pipeline management. This review highlights a critical gap in cross-functional collaboration, revealing that data science, data engineering, and MLOps teams frequently operate in silos. While the availability of powerful tools and platforms has improved dramatically, the coordination required to implement governance, versioning, and observability practices remains uneven. In contrast to earlier studies that viewed pipeline challenges as predominantly technical, this review finds that organizational misalignment is a substantial bottleneck in pipeline reproducibility and scalability. Specifically, the absence of clear roles for metadata management, feature ownership, and model governance contributes to undocumented transformations and broken lineage. Teams often lack shared protocols for schema evolution, resulting in incompatible updates that propagate errors through downstream AI models. Moreover, earlier frameworks rarely addressed the need for centralized

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

governance frameworks, but this review shows that without them, pipeline complexity increases and accountability decreases. The findings confirm that while technical infrastructure has advanced, the effectiveness of AI-ready pipelines depends equally on organizational maturity, cross-functional communication, and a shared commitment to data governance. Addressing these challenges requires both structural changes and cultural shifts that align technical execution with strategic AI goals.

CONCLUSION

The discussion of this systematic review underscores the convergence of emerging architectural models with evolving AI readiness requirements, positioning the Medallion Architecture and cloud-native integration as transformative frameworks in data engineering. Compared to traditional ETL and data warehousing approaches, the Medallion model—validated by over 42 reviewed studies—offers modular, asynchronous processing that improves traceability and reuse across bronze, silver, and gold layers, addressing longstanding limitations of rigidity and schema coupling found in classical architectures like Inmon's CIF and the star schema. Similarly, cloud-native services such as AWS Glue, Azure Data Factory, and Google Dataflow, as documented in 58 studies, demonstrate significant advancements from early Hadoop and MapReduce implementations by enabling elastic scaling, infrastructure abstraction, and seamless orchestration for batch and streaming workloads. These capabilities outperform earlier systems in latency, availability, and DevOps automation. Metadata management and lineage, once a theoretical ideal due to lack of supporting tools, are now deeply integrated into modern pipelines via platforms like Apache Atlas, Amundsen, and Microsoft Purview, which facilitate AI model explainability, versioning, and compliance—key themes covered in 39 articles. The rise of lakehouse architectures, explored in 36 reviewed works, confirms the architectural synthesis of data lakes and warehouses, supporting concurrent analytic and AI workloads without redundancy or data duplication. Unlike earlier lakes prone to schema drift and governance lapses, lakehouses equipped with Delta Lake, Hudi, and Iceberg enforce schema integrity and transactional consistency across distributed environments. Moreover, the operationalization of feature stores within these environments, covered in 29 studies, marks a departure from ad hoc ML workflows by supporting feature versioning, online-offline consistency, and low-latency retrieval. Governance and regulatory alignment, once treated as post-hoc, are now embedded within transformation and orchestration processes, enabling real-time compliance under GDPR, HIPAA, and CCPA. Despite these advancements, 31 studies highlight persistent organizational challenges, including cross-functional misalignment, tooling silos, and insufficient observability—factors that parallel concerns in earlier literature and emphasize the ongoing need for cultural and procedural maturity to fully realize AI-ready infrastructures.

REFERENCES

[1]. Abdullah Al, M., Rajesh, P., Mohammad Hasan, I., & Zahir, B. (2022). A Systematic Review of The Role Of SQL And Excel In Data-Driven Business Decision-Making For Aspiring Analysts. *American Journal of Scholarly Research and Innovation*, *1*(01), 249-269. https://doi.org/10.63125/n142cg62

[2]. Afsari, K., Eastman, C., & Shelden, D. (2017). Building Information Modeling data interoperability for Cloud-based collaboration: Limitations and opportunities. *International Journal of Architectural Computing*, *15*(3), 187-202. https://doi.org/10.1177/1478077117731174

[3]. Agletdinov, E., Pomponi, E., Merson, D., & Vinogradov, A. (2016). A novel Bayesian approach to acoustic emission data analysis. *Ultrasonics*, *72*(NA), 89-94. https://doi.org/10.1016/j.ultras.2016.07.014

[4]. Ahmed, I., Jun, M., & Ding, Y. (2022). A Spatio-Temporal Track Association Algorithm Based on Marine Vessel Automatic Identification System Data. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 20783-20797. https://doi.org/10.1109/tits.2022.3187714

[5]. Ahn, B.-H., Kim, J. M., & Choi, B.-K. (2019). Artificial intelligence-based machine learning considering flow and temperature of the pipeline for leak early detection using acoustic emission. *Engineering Fracture Mechanics*, *210*(NA), 381-392. https://doi.org/10.1016/j.engfracmech.2018.03.010

[6]. Al-Barqawi, H., & Zayed, T. (2006). Condition Rating Model for Underground Infrastructure Sustainable Water Mains. *Journal of Performance of Constructed Facilities*, *20*(2), 126-135. https://doi.org/10.1061/(asce)0887-3828(2006)20:2(126)

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

[7]. Al-Kindi, G. A. H., Baul, R. M., & Gill, K. F. (1992). An application of machine vision in the automated inspection of engineering surfaces. *International Journal of Production Research*, *30*(2), 241-253. https://doi.org/10.1080/00207549208942892

[8]. Alamri, A. H. (2020). Localized corrosion and mitigation approach of steel materials used in oil and gas pipelines – An overview. *Engineering Failure Analysis*, *116*(NA), 104735-NA. https://doi.org/10.1016/j.engfailanal.2020.104735

[9]. Alba, J. M. M., Dy, G. C., Virina, N. I. M., Samonte, M. J. C., & Cruz, F. R. G. (2020). Localized Monitoring Mobile Application for Automatic Identification System (AIS) for Sea Vessels. *2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA)*, *NA*(NA), 790-794. https://doi.org/10.1109/iciea49774.2020.9102087

[10]. Alomari, E., Barnawi, A., & Sakr, S. (2014). iiWAS - CDPort: A Framework of Data Portability in Cloud Platforms. *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services*, *NA*(NA), 126-133. https://doi.org/10.1145/2684200.2684324

[11]. Andročec, D., Vrček, N., & Küngas, P. (2015). SERVICES - Service-Level Interoperability Issues of Platform as a Service. *2015 IEEE World Congress on Services*, *NA*(NA), 349-356. https://doi.org/10.1109/services.2015.60

[12]. Angjeliu, G., Cardani, G., & Coronelli, D. (2019). DIGITAL MODELLING AND ANALYSIS OF MASONRY VAULTS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2/W11*(NA), 83-89. https://doi.org/10.5194/isprs-archives-xlii-2-w11-83-2019

[13]. Angjeliu, G., Cardani, G., & Coronelli, D. (2019). DIGITAL MODELLING AND ANALYSIS OF MASONRY VAULTS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4211*(NA), 83-89. https://doi.org/NA

[14]. Anika Jahan, M., Md Shakawat, H., & Noor Alam, S. (2022). Digital transformation in marketing: evaluating the impact of web analytics and SEO on SME growth. *American Journal of Interdisciplinary Studies*, *3*(04), 61-90. https://doi.org/10.63125/8t10v729

[15]. Antón, D., Medjdoub, B., Shrahily, R., & Moyano, J. (2018). Accuracy evaluation of the semi-automatic 3D modeling for historical building information models. *International Journal of Architectural Heritage*, *12*(5), 790-805. https://doi.org/10.1080/15583058.2017.1415391

[16]. Arunkumar, G., & Venkataraman, N. (2015). A Novel Approach to Address Interoperability Concern in Cloud Computing. *Procedia Computer Science*, *50*(NA), 554-559. https://doi.org/10.1016/j.procs.2015.04.083

[17]. Bahga, A., & Madisetti, V. K. (2013). A Cloud-based Approach for Interoperable Electronic Health Records (EHRs). *IEEE journal of biomedical and health informatics*, *17*(5), 894-906. https://doi.org/10.1109/jbhi.2013.2257818

[18]. Barbian, A., & Beller, M. (2012). In-Line Inspection of High Pressure Transmission Pipelines: State-of-the-Art and Future Trends. *NA*, *NA*(NA), NA-NA. https://doi.org/NA

[19]. Begoli, E., Goethert, I., & Knight, K. (2021). A Lakehouse Architecture for the Management and Analysis of Heterogeneous Data for Biomedical Research and Mega-biobanks. *2021 IEEE International Conference on Big Data (Big Data)*, *NA*(NA), 4643-4651. https://doi.org/10.1109/bigdata52589.2021.9671534

[20]. Brodeur, J., Coetzee, S., Danko, D., Garcia, S., & Hjelmager, J. (2019). Geographic Information Metadata—An Outlook from the International Standardization Perspective. *ISPRS International Journal of Geo-Information*, *8*(6), 280-NA. https://doi.org/10.3390/ijgi8060280

[21]. Brumana, R., Della Torre, S., Oreni, D., Previtali, M., Cantini, L., Barazzetti, L., Franchi, A., & Banfi, F. (2017). HBIM CHALLENGE AMONG THE PARADIGM OF COMPLEXITY, TOOLS AND PRESERVATION: THE BASILICA DI COLLEMAGGIO 8 YEARS AFTER THE EARTHQUAKE (L'AQUILA). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2/W5*(NA), 97-104. https://doi.org/10.5194/isprs-archives-xlii-2-w5-97-2017

[22]. Busch, V. M., Pereyra-Gonzalez, A., Šegatin, N., Santagapita, P. R., Ulrih, N. P., & del Pilar Buera, M. (2017). Propolis encapsulation by spray drying: Characterization and stability. *LWT*, *75*(75), 227-235. https://doi.org/10.1016/j.lwt.2016.08.055

[23]. Camerini, C. G., Rebello, J. M. A., Braga, L. C., dos Santos, R. W. F., Chady, T., Psuj, G., & Pereira, G. R. (2018). In-Line Inspection Tool with Eddy Current Instrumentation for Fatigue Crack Detection. *Sensors (Basel, Switzerland)*, *18*(7), 2161-NA. https://doi.org/10.3390/s18072161

[24]. Cardoso, J., Barros, A., May, N., & Kylau, U. (2010). IEEE SCC - Towards a Unified Service Description Language for the Internet of Services: Requirements and First Developments. *2010 IEEE International Conference on Services Computing*, *NA*(NA), 602-609. https://doi.org/10.1109/scc.2010.93

[25]. Carrasco, J., Cubo, J., Durán, F., & Pimentel, E. (2016). CLOUD - Bidimensional Cross-Cloud Management with TOSCA and Brooklyn. *2016 IEEE 9th International Conference on Cloud Computing (CLOUD)*, *NA*(NA), 951-955. https://doi.org/10.1109/cloud.2016.0143

[26]. Challita, S., Paraiso, F., & Merle, P. (2017). CLOUD - Towards Formal-Based Semantic Interoperability in Multi-Clouds: The FCLOUDS Framework. *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)*, *NA*(NA), 710-713. https://doi.org/10.1109/cloud.2017.98

[27]. Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, *13*(6), 377-387. https://doi.org/10.1145/362384.362685

[28]. Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies: where is their meeting point? *Data & Knowledge Engineering*, *46*(1), 41-64. https://doi.org/10.1016/s0169-023x(02)00195-7

[29]. Cunha, I. B. S., Sawaya, A. C. H. F., Caetano, F. M., Shimizu, M. T., Marcucci, M. C., Drezza, F. T., Povia, G. S., & de Oliveira Carvalho, P. (2004). Factors that influence the yield and composition of Brazilian propolis extracts. *Journal of the Brazilian Chemical Society*, *15*(6), 964-970. https://doi.org/10.1590/s0103-50532004000600026

[30]. Cushnie, T. P. T., Hamilton, V. E. S., Chapman, D. G., Taylor, P. W., & Lamb, A. J. (2007). Aggregation of Staphylococcus aureus following treatment with the antibacterial flavonol galangin. *Journal of applied microbiology*, *103*(5), 1562-1567. https://doi.org/10.1111/j.1365-2672.2007.03393.x

[31]. Dai, J., Yaylayan, V. A., Raghavan, G. S. V., & Pare, J. R. (1999). Extraction and colorimetric determination of azadirachtin-related limonoids in neem seed kernel. *Journal of Agricultural and Food Chemistry*, *47*(9), 3738-3742. https://doi.org/10.1021/jf990227h

[32]. de Carvalho, J. O., Trinta, F., & Vieira, D. (2018). CLOSER - PacificClouds: A Flexible MicroServices based Architecture for Interoperability in Multi-Cloud Environments. *Proceedings of the 8th International Conference on Cloud Computing and Services Science*, *NA*(NA), 448-455. https://doi.org/10.5220/0006705604480455

[33]. de Souza, E. N., Boerder, K., Matwin, S., & Worm, B. (2016). Correction: Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning. *PloS one*, *11*(9), e0163760-NA. https://doi.org/10.1371/journal.pone.0163760

[34]. Dean, J. R., Liu, B., & Price, R. (1998). Extraction of Tanshinone IIA from Salvia miltiorrhiza bunge using supercritical fluid extraction and a new extraction technique, phytosol solvent extraction. *Journal of Chromatography A*, *799*(1), 343-348. https://doi.org/10.1016/s0021-9673(97)01087-x

[35]. Dhar, S., & Balakrishnan, B. (2006). Risks, Benefits, and Challenges in Global IT Outsourcing: Perspectives and Practices. *Journal of Global Information Management*, *14*(3), 59-89. https://doi.org/10.4018/jgim.2006070104

[36]. Di, L., Moe, K., & Yu, G. (2009). Metadata requirements analysis for the emerging Sensor Web This was orally presented at the European Geosciences Union General Assembly 2008, Vienna, Austria, 13-18 April 2008. *International Journal of Digital Earth*, *2*(sup1), 3-17. https://doi.org/10.1080/17538940902866195

[37]. Di Martino, B., & Esposito, A. (2016). Cloud Forward - Semantic Techniques for Multi-cloud Applications Portability and Interoperability☆. *Procedia Computer Science*, *97*(NA), 104-113. https://doi.org/10.1016/j.procs.2016.08.285

[38]. Dillon, T. S., Wu, C., & Chang, E. (2010). AINA - Cloud Computing: Issues and Challenges. *2010 24th IEEE International Conference on Advanced Information Networking and Applications*, *NA*(NA), 27-33. https://doi.org/10.1109/aina.2010.187

[39]. do Nascimento, T. G., dos Santos Arruda, R. E., da Cruz Almeida, E. T., dos Santos Oliveira, J. M., Basílio-Júnior, I. D., de Moraes Porto, I. C. C., Sabino, A. R., Tonholo, J., Gray, A. I., Ebel, R. E., Clements, C., Zhang, T., & Watson, D. G. (2019). Comprehensive multivariate correlations between climatic effect, metabolite-profile, antioxidant capacity and antibacterial activity of Brazilian red propolis metabolites during seasonal study. *Scientific reports*, *9*(1), 18293-18293. https://doi.org/10.1038/s41598-019-54591-3

[40]. Dowell, S., Barreto, A., Michael, J. B., & Shing, M.-T. (2011). SoSE - Cloud to cloud interoperability. *2011 6th International Conference on System of Systems Engineering*, *NA*(NA), 258-263. https://doi.org/10.1109/sysose.2011.5966607

[41]. Emeakaroha, V. C., Healy, P. D., Fatema, K., & Morrison, J. P. (2014). Euro-Par Workshops - Cloud Interoperability via Message Bus and Monitoring Integration. In (Vol. NA, pp. 65-74). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-54420-0_7

[42]. Eskilsson, C. S., Björklund, E., Mathiasson, L., Karlsson, L., & Torstensson, A. (1999). Microwave-assisted extraction of felodipine tablets ☆. *Journal of Chromatography A*, *840*(1), 59-70. https://doi.org/10.1016/s0021-9673(99)00194-6

[43]. Evmides, N., Odysseos, L., Michaelides, M. P., & Herodotou, H. (2022). An Intelligent Framework for Vessel Traffic Monitoring Using AIS Data. *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*, *NA*(NA), 413-418. https://doi.org/10.1109/mdm55031.2022.00091

[44]. Ferry, N., Chauvel, F., Rossini, A., Morin, B., & Solberg, A. (2013). NordiCloud - Managing multi-cloud systems with CloudMF. *Proceedings of the Second Nordic Symposium on Cloud Computing & Internet Technologies*, *NA*(NA), 38-45. https://doi.org/10.1145/2513534.2513542

[45]. Ferry, N., Song, H., Rossini, A., Chauvel, F., & Solberg, A. (2014). UCC - Cloud MF: Applying MDE to Tame the Complexity of Managing Multi-cloud Applications. *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, *NA*(NA), 269-277. https://doi.org/10.1109/ucc.2014.36

[46]. Fonnet, A., Alves, N., Sousa, N., Guevara, M., & Magalhães, L. (2017). Heritage BIM integration with mixed reality for building preventive maintenance. *2017 24º Encontro Português de Computação Gráfica e Interação (EPCGI)*, *NA*(NA), 1-7. https://doi.org/10.1109/epcgi.2017.8124304

[47]. Fornos, R. A. (2012). Construcción de la base gráfica para un sistema de información y gestión del patrimonio arquitectónico: Casa de Hylas. *Arqueología de la Arquitectura*, *NA*(9), 11-25. https://doi.org/10.3989/arqarqt.2012.10005

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

[48]. Fukunaga, K., Meldrum, T., Zia, W., Ohno, M., Fuchida, T., & Blümich, B. (2013). Digital Heritage (1) - Nondestructive investigation of the internal structure of fresco paintings. *2013 Digital Heritage International Congress (DigitalHeritage)*, *1*(NA), 81-88. https://doi.org/10.1109/digitalheritage.2013.6743716

[49]. Furner, J. (2019). Definitions of "Metadata": A Brief Survey of International Standards. *Journal of the Association for Information Science and Technology*, *71*(6), 33-42. https://doi.org/10.1002/asi.24295

[50]. Galeotti, F., Maccari, F., Fachini, A., & Volpi, N. (2018). Chemical Composition and Antioxidant Activity of Propolis Prepared in Different Forms and in Different Solvents Useful for Finished Products. *Foods (Basel, Switzerland)*, *7*(3), 41-51. https://doi.org/10.3390/foods7030041

[51]. Ganzler, K., Salgó, A., & Valkó, K. (1986). Microwave extraction. A novel sample preparation method for chromatography. *Journal of chromatography*, *371*(NA), 299-306. https://doi.org/10.1016/s0021-9673(01)94714-4

[52]. Garagnani, S. (2013a). Building Information Modeling and real world knowledge. *NA*, *NA*(NA), NA-NA. https://doi.org/NA

[53]. Garagnani, S. (2013b). Digital Heritage (1) - Building Information Modeling and real world knowledge: A methodological approach to accurate semantic documentation for the built environment. *2013 Digital Heritage International Congress (DigitalHeritage)*, *1*(NA), 489-496. https://doi.org/10.1109/digitalheritage.2013.6743788

[54]. Ghosh, G., Rostron, P., Garg, R., & Panday, A. (2018). Hydrogen induced cracking of pipeline and pressure vessel steels: A review. *Engineering Fracture Mechanics*, *199*(NA), 609-618. https://doi.org/10.1016/j.engfracmech.2018.06.018

[55]. Gonzalez-Perez, C. (2018). An Ontology for Cultural Heritage. In (Vol. NA, pp. 195-215). Springer International Publishing. https://doi.org/10.1007/978-3-319-72652-6_19

[56]. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645-1660. https://doi.org/10.1016/j.future.2013.01.010

[57]. Guo, W., Gao, B., Tian, G. Y., & Si, D. (2020). Physic perspective fusion of electromagnetic acoustic transducer and pulsed eddy current testing in non-destructive testing system. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, *378*(2182), 20190608-20190608. https://doi.org/10.1098/rsta.2019.0608

[58]. Habermann, T. (2018). Metadata Life Cycles, Use Cases and Hierarchies. *Geosciences*, *8*(5), 179-NA. https://doi.org/10.3390/geosciences8050179

[59]. Harby, A. A., & Zulkernine, F. (2022). From Data Warehouse to Lakehouse: A Comparative Review. *2022 IEEE International Conference on Big Data (Big Data)*, *NA*(NA), 389-395. https://doi.org/10.1109/bigdata55660.2022.10020719

[60]. Harris, R., & Olby, N. (2001). Earth observation data archiving in the USA and Europe. *Space Policy*, *17*(1), 35-48. https://doi.org/10.1016/s0265-9646(00)00052-7

[61]. Hawari, A. H., Alkadour, F., Elmasry, M., & Zayed, T. (2020). A state of the art review on condition assessment models developed for sewer pipelines. *Engineering Applications of Artificial Intelligence*, *93*(NA), 103721-NA. https://doi.org/10.1016/j.engappai.2020.103721

[62]. Heidary, R., Gabriel, S. A., Modarres, M., Groth, K. M., & Vahdati, N. (2020). A Review of Data-Driven Oil and Gas Pipeline Pitting Corrosion Growth Models Applicable for Prognostic and Health Management. *International Journal of Prognostics and Health Management*, *9*(1), NA-NA. https://doi.org/10.36001/ijphm.2018.v9i1.2695

[63]. Hermenean, A., Mariasiu, T., González, I. N., Vegara-Meseguer, J., Miuţescu, E., Chakraborty, S., & Sánchez, H. P. (2017). Hepatoprotective activity of chrysin is mediated through TNF-α in chemically-induced acute liver damage: An in vivo study and molecular modeling. *Experimental and therapeutic medicine*, *13*(5), 1671-1680. https://doi.org/10.3892/etm.2017.4181

[64]. Hery, N. A., Lukas, S., Yugopuspito, P., Murwantara, I. M., & Krisnadi, D. (2020). Website Design for Locating Tuna Fishing Spot Using Naïve Bayes and SVM Based on VMS Data on Indonesian Sea. *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, *NA*(NA), 89-93. https://doi.org/10.1109/isriti51436.2020.9315338

[65]. Hoare, S., Helian, N., & Baddoo, N. (2016). UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld - A Semantic-Agent Framework for PaaS Interoperability. *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, *NA*(NA), 788-793. https://doi.org/10.1109/uic-atc-scalcom-cbdcom-iop-smartworld.2016.0126

[66]. Hu, Y., Janowicz, K., Prasad, S., & Gao, S. (2015). Metadata Topic Harmonization and Semantic Search for Linked-Data-Driven Geoportals: A Case Study Using ArcGIS Online. *Transactions in GIS*, *19*(3), 398-416. https://doi.org/10.1111/tgis.12151

[67]. Huang, H., Cui, X., Bi, X., Liu, C., Hong, F., & Guo, S. (2020). FVRD: Fishing Vessels Relationships Discovery System Through Vessel Trajectory. *IEEE Access*, *8*(NA), 112530-112538. https://doi.org/10.1109/access.2020.3002173

[68]. Huang, J., Wan, J., Yu, J., Zhu, F., & Ren, Y. (2020). Edge Computing-Based Adaptable Trajectory Transmission Policy for Vessels Monitoring Systems of Marine Fishery. *IEEE Access*, *8*(NA), 50684-50695. https://doi.org/10.1109/access.2020.2980322

[69]. Isaac, A., & Haslhofer, B. (2013). Europeana Linked Open Data --data.europeana.eu. *Semantic Web*, *4*(3), 291-297. https://doi.org/10.3233/sw-120092

[70]. Jackson, J. B., Labaune, J., Bailleul-Lesuer, R., d'Alessandro, L., Whyte, A., Bowen, J. W., Menu, M., & Mourou, G. (2014). Terahertz pulse imaging in archaeology. *Frontiers of Optoelectronics*, *8*(1), 81-92. https://doi.org/10.1007/s12200-014-0446-y

[71]. Jepsen, P. U., Cooke, D. G., & Koch, M. (2010). Terahertz spectroscopy and imaging – Modern techniques and applications. *Laser & Photonics Reviews*, *5*(1), 124-166. https://doi.org/10.1002/lpor.201000011

[72]. Ji, J., Robert, D., Zhang, C., Zhang, D., & Kodikara, J. (2017). Probabilistic physical modelling of corroded cast iron pipes for lifetime prediction. *Structural Safety*, *64*(NA), 62-75. https://doi.org/10.1016/j.strusafe.2016.09.004

[73]. Jun, X. (2006). Comparison of antioxidant activity of ethanolic extracts of propolis obtained by different extraction methods. *The Canadian Journal of Chemical Engineering*, *84*(4), 447-451. https://doi.org/10.1002/cjce.5450840405

[74]. Kalantari, M., Rajabifard, A., Olfat, H., & Williamson, I. (2014). Geospatial Metadata 2.0 – An approach for Volunteered Geographic Information. *Computers, Environment and Urban Systems*, *48*(NA), 35-48. https://doi.org/10.1016/j.compenvurbsys.2014.06.005

[75]. Khan, M. A. M., Roksana, H., & Ammar, B. (2022). A Systematic Literature Review on Energy-Efficient Transformer Design For Smart Grids. *American Journal of Scholarly Research and Innovation*, *1*(01), 186-219. https://doi.org/10.63125/6n1yka80

[76]. Khodayari-Rostamabad, A., Reilly, J. P., Nikolova, N. K., Hare, J. R., & Pasha, S. (2009). Machine Learning Techniques for the Analysis of Magnetic Flux Leakage Images in Pipeline Inspection. *IEEE Transactions on Magnetics*, *45*(8), 3073-3084. https://doi.org/10.1109/tmag.2009.2020160

[77]. Kim, H., & Parashar, M. (2011). Cloud Computing: Principles and Paradigms - CometCloud: An Autonomic Cloud Engine. *Cloud Computing*, *NA*(NA), 275-297. https://doi.org/10.1002/9780470940105.ch10

[78]. Kim, J., Yang, G., Udpa, L., & Udpa, S. S. (2010). Classification of pulsed eddy current GMR data on aircraft structures. *NDT & E International*, *43*(2), 141-144. https://doi.org/10.1016/j.ndteint.2009.10.003

[79]. Kokla, M., & Guilbert, E. (2020). A Review of Geospatial Semantic Information Modeling and Elicitation Approaches. *ISPRS International Journal of Geo-Information*, *9*(3), 146-NA. https://doi.org/10.3390/ijgi9030146

[80]. Kovacs, P., Lehner, B., Thummerer, G., Mayr, G., Burgholzer, P., & Huemer, M. (2020). Deep learning approaches for thermographic imaging. *Journal of Applied Physics*, *128*(15), 155103-NA. https://doi.org/10.1063/5.0020404

[81]. Kruchten, P., Lago, P., & van Vliet, H. (2006). Building up and reasoning about architectural knowledge. *Lecture Notes in Computer Science*, *NA*(NA), 43-58. https://doi.org/NA

[82]. Krügener, K., Schwerdtfeger, M., Busch, S. F., Soltani, A., Castro-Camus, E., Koch, M., & Viöl, W. (2015). Terahertz meets sculptural and architectural art: Evaluation and conservation of stone objects with T-ray technology. *Scientific reports*, *5*(1), 14842-14842. https://doi.org/10.1038/srep14842

[83]. Kubiliene, L., Jekabsone, A., Zilius, M., Trumbeckaite, S., Simanaviciute, D., Gerbutaviciene, R., & Majiene, D. (2018). Comparison of aqueous, polyethylene glycol-aqueous and ethanolic propolis extracts: antioxidant and mitochondria modulating properties. *BMC complementary and alternative medicine*, *18*(1), 1-10. https://doi.org/10.1186/s12906-018-2234-5

[84]. Kumar, D., & Li, S. (2022). Separating Storage and Compute with the Databricks Lakehouse Platform. *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*, *NA*(NA), 1-2. https://doi.org/10.1109/dsaa54385.2022.10032386

[85]. Kuruvila, R., Kumaran, S. T., Khan, M. A., & Uthayakumar, M. (2018). A brief review on the erosion-corrosion behavior of engineering materials. *Corrosion Reviews*, *36*(5), 435-447. https://doi.org/10.1515/corrrev-2018-0022

[86]. Lafia, S., Jablonski, J., Kuhn, W., Cooley, S., & Medrano, F. A. (2016). Spatial discovery and the research library. *Transactions in GIS*, *20*(3), 399-412. https://doi.org/10.1111/tgis.12235

[87]. Larman, C. (2004). *Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development (3rd Edition)* (Vol. NA). NA. https://doi.org/NA

[88]. Lee, H. S., Lee, S. Y., Park, S. H., Lee, J. H., Ahn, S. K., Choi, Y. M., Choi, D. J., & Chang, J. H. (2013). Antimicrobial medical sutures with caffeic acid phenethyl ester and their in vitro/in vivo biological assessment. *MedChemComm*, *4*(5), 777-782. https://doi.org/10.1039/c2md20289a

[89]. Levin, A., Barabash, K., Ben-Itzhak, Y., Guenender, S., & Schour, L. (2015). CLOUD - Networking Architecture for Seamless Cloud Interoperability. *2015 IEEE 8th International Conference on Cloud Computing*, *NA*(NA), 1021-1024. https://doi.org/10.1109/cloud.2015.141

[90]. Li, Y., Peng, S., Li, Y., & Jiang, W. (2020). A review of condition-based maintenance: Its prognostic and operational aspects. *Frontiers of Engineering Management*, *7*(3), 323-334. https://doi.org/10.1007/s42524-020-0121-5

[91]. Lin, B. (2020). Overview of High Performance Computing Power Building for the Big Data of Marine Forecasting. *2020 International Conference on Big Data and Informatization Education (ICBDIE)*, *NA*(NA), 79-82. https://doi.org/10.1109/icbdie50010.2020.00025

[92]. Liu, R. W., Liang, M., Nie, J., Garg, S., Zhang, Y., & Xiong, Z. (2021). IRI - Extraction of Hottest Shipping Routes: From Positioning Data to Intelligent Surveillance. *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, *NA*(NA), 255-262. https://doi.org/10.1109/iri51335.2021.00041

[93]. Liu, Z., & Kleiner, Y. (2013). State of the art review of inspection technologies for condition assessment of water pipes. *Measurement*, *46*(1), 1-15. https://doi.org/10.1016/j.measurement.2012.05.032

[94]. Longo, A., Zappatore, M., Bochicchio, M. A., Livieri, B., Guarino, N., & Napoleone, D. (2016). AINA Workshops - Cloud for Europe: The Experience of a Tenderer. *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, *NA*(NA), 153-158. https://doi.org/10.1109/waina.2016.66

[95]. Lu, H., Behbahani, S., Azimi, M., Matthews, J. C., Han, S., & Iseley, T. (2020). Trenchless Construction Technologies for Oil and Gas Pipelines: State-of-the-Art Review. *Journal of Construction Engineering and Management*, *146*(6), 03120001-NA. https://doi.org/10.1061/(asce)co.1943-7862.0001819

[96]. Lu, H., Yao, H., Zou, R., Chen, X., & Xu, H. (2019). Galangin Suppresses Renal Inflammation via the Inhibition of NF-κB, PI3K/AKT and NLRP3 in Uric Acid Treated NRK-52E Tubular Epithelial Cells. *BioMed research international*, *2019*(NA), 3018357-3018357. https://doi.org/10.1155/2019/3018357

[97]. Lu, S., Feng, J., Zhang, H., Liu, J., & Wu, Z. (2019). An Estimation Method of Defect Size From MFL Image Using Visual Transformation Convolutional Neural Network. *IEEE Transactions on Industrial Informatics*, *15*(1), 213-224. https://doi.org/10.1109/tii.2018.2828811

[98]. Lytra, I., Vidal, M.-E., Orlandi, F., & Attard, J. (2017). *ICE/ITMC - A big data architecture for managing oceans of data and maritime applications* (Vol. NA). IEEE. https://doi.org/10.1109/ice.2017.8280019

[99]. Macher, H., Landes, T., & Grussenmeyer, P. (2017). From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings. *Applied Sciences*, *7*(10), 1030-NA. https://doi.org/10.3390/app7101030

[100]. Marcato, B., & Vianello, M. (2000). Microwave-assisted extraction by fast sample preparation for the systematic analysis of additives in polyolefins by high-performance liquid chromatography. *Journal of chromatography. A*, *869*(1), 285-300. https://doi.org/10.1016/s0021-9673(99)00940-1

[101]. Mašek, T., Perin, N., Racane, L., Cindrić, M., Paljetak, H. Č., Perić, M., Matijašić, M., Verbanac, D., Radić, B., Šuran, J., & Starčević, K. (2018). Chemical composition, antioxidant and antibacterial activity of different extracts of poplar type propolis. *Croatica Chemica Acta*, *91*(1), 81-88. https://doi.org/10.5562/cca3298

[102]. Maué, P., Michels, H., & Roth, M. (2012). Injecting Semantic Annotations into Geospatial Web service descriptions. *Semantic Web*, *3*(4), 385-395. https://doi.org/10.3233/sw-2012-0061

[103]. Mayilvaganan, M., & Sabitha, M. (2013). A cloud-based architecture for Big-Data analytics in smart grid: A proposal. *2013 IEEE International Conference on Computational Intelligence and Computing Research*, *NA*(NA), 1-4. https://doi.org/10.1109/iccic.2013.6724168

[104]. McGee, M., Durante, K., & Weimer, K. H. (2017). Toward a Linked Data Model for Describing Cartographic Resources. *Journal of Map & Geography Libraries*, *13*(1), 133-144. https://doi.org/10.1080/15420353.2017.1308291

[105]. Md Mahamudur Rahaman, S. (2022). Electrical And Mechanical Troubleshooting in Medical And Diagnostic Device Manufacturing: A Systematic Review Of Industry Safety And Performance Protocols. *American Journal of Scholarly Research and Innovation*, *1*(01), 295-318. https://doi.org/10.63125/d68y3590

[106]. Md Masud, K. (2022). A Systematic Review Of Credit Risk Assessment Models In Emerging Economies: A Focus On Bangladesh's Commercial Banking Sector. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 01-31. https://doi.org/10.63125/p7ym0327

[107]. Md Takbir Hossen, S., & Md Atiqur, R. (2022). Advancements In 3D Printing Techniques For Polymer Fiber-Reinforced Textile Composites: A Systematic Literature Review. *American Journal of Interdisciplinary Studies*, *3*(04), 32-60. https://doi.org/10.63125/s4r5m391

[108]. Mezgár, I., & Rauschecker, U. (2014). The challenge of networked enterprises for cloud computing interoperability. *Computers in Industry*, *65*(4), 657-674. https://doi.org/10.1016/j.compind.2014.01.017

[109]. Miksovic, C., & Zimmermann, O. (2011). WICSA - Architecturally Significant Requirements, Reference Architecture, and Metamodel for Knowledge Management in Information Technology Services. *2011 Ninth Working IEEE/IFIP Conference on Software Architecture*, *NA*(NA), 270-279. https://doi.org/10.1109/wicsa.2011.43

[110]. Mostajeran, E., Ismail, B. I., Khalid, M. F., & Ong, H. (2015). A survey on SLA-based brokering for inter-cloud computing. *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)*, *NA*(NA), 25-31. https://doi.org/10.1109/icctim.2015.7224588

[111]. Neumaier, S., Savenkov, V., & Polleres, A. (2018). Geo-Semantic Labelling of Open Data. *Procedia Computer Science*, *137*(NA), 9-20. https://doi.org/10.1016/j.procs.2018.09.002

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

[112]. Nieto, E. M., Moyano, J., & García, Á. C. (2019). Construction study of the Palace of the Children of Don Gome (Andújar, Jaén), managed through the HBIM project. *Virtual Archaeology Review*, *10*(20), 84-97. https://doi.org/10.4995/var.2019.10567

[113]. Nogueras-Iso, J., Zarazaga-Soria, F. J., Béjar, R., Álvarez, P., & Muro-Medrano, P. R. (2005). OGC Catalog Services: a key element for the development of Spatial Data Infrastructures. *Computers & Geosciences*, *31*(2), 199-209. https://doi.org/10.1016/j.cageo.2004.05.015

[114]. Olgierd, B., Kamila, Ż., Anna, B., & Emilia, M. (2021). The Pluripotent Activities of Caffeic Acid Phenethyl Ester. *Molecules (Basel, Switzerland)*, *26*(5), 1335-NA. https://doi.org/10.3390/molecules26051335

[115]. Onori, P., DeMorrow, S., Gaudio, E., Franchitto, A., Mancinelli, R., Venter, J., Kopriva, S., Ueno, Y., Alvaro, D., Savage, J., Alpini, G., & Francis, H. (2009). Caffeic acid phenethyl ester decreases cholangiocarcinoma growth by inhibition of NF-κB and induction of apoptosis. *International journal of cancer*, *125*(3), 565-576. https://doi.org/10.1002/ijc.24271

[116]. Oreni, D., Brumana, R., Georgopoulos, A., & Cuca, B. (2013). HBIM FOR CONSERVATION AND MANAGEMENT OF BUILT HERITAGE: TOWARDS A LIBRARY OF VAULTS AND WOODEN BEAN FLOORS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-5/W1*(NA), 215-221. https://doi.org/10.5194/isprsannals-ii-5-w1-215-2013

[117]. Orescanin, D., & Hlupic, T. (2021). MIPRO - Data Lakehouse - a Novel Step in Analytics Architecture. *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, *NA*(NA), 1242-1246. https://doi.org/10.23919/mipro52101.2021.9597091

[118]. Ossai, C. I. (2020). Corrosion Defect Modelling of Aged Pipelines with a Feed-Forward Multi-Layer Neural Network for Leak and Burst Failure Estimation. *Engineering Failure Analysis*, *110*(NA), 104397-NA. https://doi.org/10.1016/j.engfailanal.2020.104397

[119]. Palestini, C., Basso, A., & Graziani, L. (2018). INTEGRATED PHOTOGRAMMETRIC SURVEY AND BIM MODELLING FOR THE PROTECTION OF SCHOOL HERITAGE, APPLICATIONS ON A CASE STUDY. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2*(NA), 821-828. https://doi.org/10.5194/isprs-archives-xlii-2-821-2018

[120]. Pan, X., Liu, H., Jia, G., & Shu, Y. Y. (2000). Microwave-assisted extraction of glycyrrhizic acid from licorice root. *Biochemical engineering journal*, *5*(3), 173-177. https://doi.org/10.1016/s1369-703x(00)00057-7

[121]. Paolo, T., Cristiano, F., Alessandro, O., & Paola, C. (2019). Semantic Profiles for Easing SensorML Description: Review and Proposal. *ISPRS International Journal of Geo-Information*, *8*(8), 340-NA. https://doi.org/10.3390/ijgi8080340

[122]. Park, Y. K., & Ikegaki, M. (1998). Preparation of water and ethanolic extracts of propolis and evaluation of the preparations. *Bioscience, biotechnology, and biochemistry*, *62*(11), 2230-2232. https://doi.org/10.1271/bbb.62.2230

[123]. Peoples, C., Parr, G., Oredope, A., & Moessner, K. (2013). The standardisation of cloud computing: Trends in the state-of-the-art and management issues for the next generation of cloud.

[124]. Pepeljnjak, S., & Kosalec, I. (2004). Galangin expresses bactericidal activity against multiple-resistant bacteria: MRSA, Enterococcus spp. and Pseudomonas aeruginosa. *FEMS microbiology letters*, *240*(1), 111-116. https://doi.org/10.1016/j.femsle.2004.09.018

[125]. Pesinis, K., & Tee, K. F. (2017). Statistical model and structural reliability analysis for onshore gas transmission pipelines. *Engineering Failure Analysis*, *82*(NA), 1-15. https://doi.org/10.1016/j.engfailanal.2017.08.008

[126]. Petcu, D. (2011). Portability and interoperability between clouds: challenges and case study. In (Vol. NA, pp. 62-74). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24755-2_6

[127]. Piesiewicz, R., Kleine-Ostmann, T., Krumbholz, N., Mittleman, D. M., Koch, M., & Kurner, T. (2005). Terahertz characterisation of building materials. *Electronics Letters*, *41*(18), 1002-1004. https://doi.org/10.1049/el:20052444

[128]. Pocobelli, D. P., Boehm, J., Bryan, P., Still, J., & Grau-Bové, J. (2018a). BUILDING INFORMATION MODELS FOR MONITORING AND SIMULATION DATA IN HERITAGE BUILDINGS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2*(NA), 909-916. https://doi.org/10.5194/isprs-archives-xlii-2-909-2018

[129]. Pocobelli, D. P., Boehm, J., Bryan, P., Still, J., & Grau-Bové, J. (2018b). Building Information Models for Monitoring and Simulation Data in Heritage Buildings. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *422*(NA), 909-916. https://doi.org/NA

[130]. Prasad, P., Vatsal, V., & Chowdhury, R. R. (2021). Maritime Vessel Route Extraction and Automatic Information System (AIS) Spoofing Detection. *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, *NA*(NA), 1-11. https://doi.org/10.1109/icaect49130.2021.9392536

[131]. Priyanka, E. B., Maheswari, C., & Thangavel, S. (2018). Remote monitoring and control of LQR-PI controller parameters for an oil pipeline transport system. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, *233*(6), 597-608. https://doi.org/10.1177/0959651818803183

[132]. Priyanka, E. B., Thangavel, S., & Gao, X.-Z. (2021). Review analysis on cloud computing based smart grid technology in the oil pipeline sensor network system. *Petroleum Research*, *6*(1), 77-90. https://doi.org/10.1016/j.ptlrs.2020.10.001

[133]. Quattrini, R., Malinverni, E. S., Clini, P., Nespeca, R., & Orlietti, E. (2015). FROM TLS TO HBIM. HIGH QUALITY SEMANTICALLY-AWARE 3D MODELING OF COMPLEX ARCHITECTURE. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-5/W4*(NA), 367-374. https://doi.org/10.5194/isprsarchives-xl-5-w4-367-2015

[134]. Quej-Ake, L. M., Rivera-Olvera, J. N., del Rosario Domínguez-Aguilar, Y., Avelino-Jiménez, I. A., Garibay-Febles, V., & Zapata-Peñasco, I. (2020). Analysis of the Physicochemical, Mechanical, and Electrochemical Parameters and Their Impact on the Internal and External SCC of Carbon Steel Pipelines. *Materials (Basel, Switzerland)*, *13*(24), 5771-NA. https://doi.org/10.3390/ma13245771

[135]. Ramanauskienė, K., Inkėnienė, A. M., Petrikaitė, V., & Briedis, V. (2013). Total Phenolic Content and Antimicrobial Activity of Different Lithuanian Propolis Solutions. *Evidence-based complementary and alternative medicine : eCAM*, *2013*(NA), 842985-842985. https://doi.org/10.1155/2013/842985

[136]. Rezaei, R., Chiew, T. K., Lee, S. P., & Aliee, Z. S. (2014). A semantic interoperability framework for software as a service systems in cloud computing environments. *Expert Systems with Applications*, *41*(13), 5751-5770. https://doi.org/10.1016/j.eswa.2014.03.020

[137]. Rice, J. A., Mechitov, K., Sim, S.-H., Nagayama, T., Jang, S., Kim, R. E., Spencer, B. F., Agha, G., & Fujino, Y. (2010). Flexible smart sensor framework for autonomous structural health monitoring. *Smart Structures and Systems*, *6*(5), 423-438. https://doi.org/10.12989/sss.2010.6.5_6.423

[138]. Rocha, B. A., Bueno, P. C. P., de Oliveira Lima Leite Vaz, M. M., Nascimento, A. P., Ferreira, N. U., de Padua Moreno, G., Rodrigues, M. R., de Mello Costa-Machado, A. R., Barizon, E. A., Campos, J. C. L., de Oliveira, P. F., Acésio, N. O., de Paula Lima Martins, S., Tavares, D. C., & Berretta, A. A. (2013). Evaluation of a Propolis Water Extract Using a Reliable RP-HPLC Methodology and In Vitro and In Vivo Efficacy and Safety Characterisation. *Evidence-based complementary and alternative medicine : eCAM*, *2013*(NA), 670451-670451. https://doi.org/10.1155/2013/670451

[139]. Román, M. C. (2015). Generación de un modelo de información del patrimonio inmueble en el momento de su protección jurídica. *EGA. Revista de expresión gráfica arquitectónica*, *20*(26), 266-277. https://doi.org/10.4995/ega.2015.4060

[140]. Samarghandian, S., Afshari, J. T., & Davoodi, S. (2011). Chrysin reduces proliferation and induces apoptosis in the human prostate cancer cell line pc-3. *Clinics (Sao Paulo, Brazil)*, *66*(6), 1073-1079. https://doi.org/10.1590/s1807-59322011000600026

[141]. Sampaio, A., & Mendonça, N. C. (2011). Uni4Cloud: an approach based on open standards for deployment and management of multi-cloud applications. *Proceedings of the 2nd International Workshop on Software Engineering for Cloud Computing*, *NA*(NA), 15-21. https://doi.org/10.1145/1985500.1985504

[142]. Saravanakumar, C., & Arun, C. (2014). Survey on interoperability, security, trust, privacy standardization of cloud computing. *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, *NA*(NA), 977-982. https://doi.org/10.1109/ic3i.2014.7019735

[143]. Sazzad, I., & Md Nazrul Islam, K. (2022). Project impact assessment frameworks in nonprofit development: a review of case studies from south asia. *American Journal of Scholarly Research and Innovation*, *1*(01), 270-294. https://doi.org/10.63125/eeja0t77

[144]. Shaiful, M., Anisur, R., & Md, A. (2022). A systematic literature review on the role of digital health twins in preventive healthcare for personal and corporate wellbeing. *American Journal of Interdisciplinary Studies*, *3*(04), 1-31. https://doi.org/10.63125/negjw373

[145]. Shaw, M. (2003). ICSE - Writing good software engineering research papers. *25th International Conference on Software Engineering, 2003. Proceedings.*, *NA*(NA), 726-736. https://doi.org/10.1109/icse.2003.1201262

[146]. Shen, L., Li, J., Wu, Y., Tang, Z., & Wang, Y. (2019). Optimization of Artificial Bee Colony Algorithm Based Load Balancing in Smart Grid Cloud. *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, *NA*(NA), 1131-1134. https://doi.org/10.1109/isgt-asia.2019.8881232

[147]. Sikos, L. F. (2016). RDF-powered semantic video annotation tools with concept mapping to Linked Data for next-generation video indexing: a comprehensive review. *Multimedia Tools and Applications*, *76*(12), 14437-14460. https://doi.org/10.1007/s11042-016-3705-7

[148]. Smits, & Friis, C. (2007). Resource Discovery in a European Spatial Data Infrastructure. *IEEE Transactions on Knowledge and Data Engineering*, *19*(1), 85-95. https://doi.org/10.1109/tkde.2007.250587

[149]. Stanik, A., Hovestadt, M., & Kao, O. (2012). CloudCom - Hardware as a Service (HaaS): Physical and virtual hardware on demand. *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings*, *NA*(NA), 149-154. https://doi.org/10.1109/cloudcom.2012.6427579

[150]. Tahmina Akter, R., & Abdur Razzak, C. (2022). The Role Of Artificial Intelligence In Vendor Performance Evaluation Within Digital Retail Supply Chains: A Review Of Strategic Decision-Making Models. *American Journal of Scholarly Research and Innovation*, *1*(01), 220-248. https://doi.org/10.63125/96jj3j86

[151]. Tampakis, P., Chondrodima, E., Pikrakis, A., Theodoridis, Y., Pristouris, K., Nakos, H., Petra, E., Dalamagas, T., Kandiros, A., Markakis, G., Maina, I., & Kavadas, S. (2020). MDM - Sea Area Monitoring and Analysis of Fishing Vessels Activity: The i4sea Big Data Platform. *2020 21st IEEE International Conference on Mobile Data Management (MDM)*, *NA*(NA), 275-280. https://doi.org/10.1109/mdm48529.2020.00063

[152]. Tant, K. M. M., Galetti, E., Mulholland, A. J., Curtis, A., & Gachagan, A. (2018). A Transdimensional Bayesian Approach to Ultrasonic Travel-time Tomography for Non-Destructive Testing. *Inverse Problems*, *34*(9), 095002-NA. https://doi.org/10.1088/1361-6420/aaca8f

[153]. Thabet, M., Boufaida, M., & Kordon, F. (2014). An approach for developing an interoperability mechanism between cloud providers. *International Journal of Space-Based and Situated Computing*, *4*(2), 88-99. https://doi.org/10.1504/ijssc.2014.062469

[154]. Thomson, C., & Boehm, J. (2015). Automatic Geometry Generation from Point Clouds for BIM. *Remote Sensing*, *7*(9), 11753-11775. https://doi.org/10.3390/rs70911753

[155]. Tokognon, C. A., Gao, B., Tian, G. Y., & Yan, Y. (2017). Structural Health Monitoring Framework Based on Internet of Things: A Survey. *IEEE Internet of Things Journal*, *4*(3), 619-635. https://doi.org/10.1109/jiot.2017.2664072

[156]. Tomaniova, M., Hajslova, J., Pavelka, J., Kocourek, V., Holadová, K., & Klímová, I. (1998). Microwave-assisted solvent extraction — a new method for isolation of polynuclear aromatic hydrocarbons from plants. *Journal of Chromatography A*, *827*(1), 21-29. https://doi.org/10.1016/s0021-9673(98)00754-7

[157]. Torrione, P. A., Throckmorton, C. S., & Collins, L. M. (2006). Performance of an adaptive feature-based processor for a wideband ground penetrating radar system. *IEEE Transactions on Aerospace and Electronic Systems*, *42*(2), 644-658. https://doi.org/10.1109/taes.2006.1642579

[158]. Trusheva, B., Trunkova, D., & Bankova, V. (2007). Different extraction methods of biologically active components from propolis: a preliminary study. *Chemistry Central journal*, *1*(1), 13-13. https://doi.org/10.1186/1752-153x-1-13

[159]. Turco, M. L., Mattone, M., & Rinaudo, F. (2017). METRIC SURVEY AND BIM TECHNOLOGIES TO RECORD DECAY CONDITIONS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-5/W1*(NA), 261-268. https://doi.org/10.5194/isprs-archives-xlii-5-w1-261-2017

[160]. Tusa, F., Paone, M., Villari, M., & Puliafito, A. (2011). UCC - CLEVER: A Cloud Cross-Computing Platform Leveraging GRID Resources. *2011 Fourth IEEE International Conference on Utility and Cloud Computing*, *NA*(NA), 390-396. https://doi.org/10.1109/ucc.2011.65

[161]. van Heesch, U., Avgeriou, P., & Hilliard, R. (2012). A documentation framework for architecture decisions. *Journal of Systems and Software*, *85*(4), 795-820. https://doi.org/10.1016/j.jss.2011.10.017

[162]. Vasconcelos, N. G., Croda, J., & Simionatto, S. (2018). Antibacterial mechanisms of cinnamon and its constituents: A review. *Microbial pathogenesis*, *120*(NA), 198-203. https://doi.org/10.1016/j.micpath.2018.04.036

[163]. Wilkinson, M., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. O. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O. G., Edmunds, S. C., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). Scientific Data - The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 160018-160018. https://doi.org/10.1038/sdata.2016.18

[164]. Wilson, A., Cox, M., Elsborg, D., Lindholm, D., & Traver, T. (2014). A semantically enabled metadata repository for scientific data. *Earth Science Informatics*, *8*(3), 649-661. https://doi.org/10.1007/s12145-014-0175-1

[165]. Wilson, J., & Tian, G. Y. (2006). 3D magnetic field sensing for magnetic flux leakage defect characterisation. *Insight - Non-Destructive Testing and Condition Monitoring*, *48*(6), 357-359. https://doi.org/10.1784/insi.2006.48.6.357

[166]. Wink, M. (2008). Evolutionary advantage and molecular modes of action of multi-component mixtures used in phytomedicine. *Current drug metabolism*, *9*(10), 996-1009. https://doi.org/10.2174/138920008786927794

[167]. Wisner, B., Mazur, K., Perumal, V. I., Baxevanakis, K. P., An, L., Feng, G., & Kontsos, A. (2019). Acoustic emission signal processing framework to identify fracture in aluminum alloys. *Engineering Fracture Mechanics*, *210*(NA), 367-380. https://doi.org/10.1016/j.engfracmech.2018.04.027

[168]. Xiao'en, L., Xiao, Y., Su, F., Wenzhou, W., & Zhou, L. (2021). AIS and VBD Data Fusion for Marine Fishing Intensity Mapping and Analysis in the Northern Part of the South China Sea. *ISPRS International Journal of Geo-Information*, *10*(5), 277-NA. https://doi.org/10.3390/ijgi10050277

[169]. Xie, M., & Tian, Z. (2018). A review on pipeline integrity management utilizing in-line inspection data. *Engineering Failure Analysis*, *92*(NA), 222-239. https://doi.org/10.1016/j.engfailanal.2018.05.010

[170]. Yahfoufi, N., Alsadi, N., Jambi, M., & Matar, C. (2018). The Immunomodulatory and Anti-Inflammatory Role of Polyphenols. *Nutrients*, *10*(11), 1618-NA. https://doi.org/10.3390/nu10111618

[171]. Yazdani, S., Yusof, R., Riazi, A. H., Karimian, A., & Hematian, A. (2014). Evaluation of Pipelines in Industrial Radiography Using Image Processing Techniques. *Advanced Science, Engineering and Medicine*, *6*(1), 81-85. https://doi.org/10.1166/asem.2014.1444

[172]. Yongsiriwit, K., Sellami, M., & Gaaloul, W. (2016). CLOUD - A Semantic Framework Supporting Cloud Resource Descriptions Interoperability. *2016 IEEE 9th International Conference on Cloud Computing (CLOUD)*, *NA*(NA), 585-592. https://doi.org/10.1109/cloud.2016.0083

[173]. Yu, S. C., Lu, K.-Y., & Chen, R.-S. (2003). Metadata management system: design and implementation. *The Electronic Library*, *21*(2), 154-164. https://doi.org/10.1108/02640470310470525

Volume 01, Issue 01 (2022)
Page No: 319-350
eISSN: 3067-2163
Doi: 10.63125/51kxtf08

[174]. Zahara, S., Pratomo, I., & Rahardjo, D. S. (2015). Application and data level interoperability on virtual machine in cloud computing environment. *2015 1st International Conference on Wireless and Telematics (ICWT)*, *NA*(NA), 1-5. https://doi.org/10.1109/icwt.2015.7449238

[175]. Zajam, S., Joshi, T., & Bhattacharya, B. (2019). Application of wavelet analysis and machine learning on vibration data from gas pipelines for structural health monitoring. *Procedia Structural Integrity*, *14*(NA), 712-719. https://doi.org/10.1016/j.prostr.2019.05.089

[176]. Zakikhani, K., Nasiri, F., & Zayed, T. (2020). A Review of Failure Prediction Models for Oil and Gas Pipelines. *Journal of Pipeline Systems Engineering and Practice*, *11*(1), 03119001-NA. https://doi.org/10.1061/(asce)ps.1949-1204.0000407

[177]. Zhang, J.-L., Shen, X., Wang, K., Cao, X., Zhang, C.-P., Zheng, H., & Hu, F. (2016). Antioxidant activities and molecular mechanisms of the ethanol extracts of Baccharis propolis and Eucalyptus propolis in RAW64.7 cells. *Pharmaceutical biology*, *54*(10), 2220-2235. https://doi.org/10.3109/13880209.2016.1151444

[178]. Zhang, J., Tian, G. Y., Marindra, A. M. J., Sunny, A. I., & Zhao, A. (2017). A Review of Passive RFID Tag Antenna-Based Sensors and Systems for Structural Health Monitoring Applications. *Sensors (Basel, Switzerland)*, *17*(2), 265-NA. https://doi.org/10.3390/s17020265

[179]. Zhang, Z., Wu, C., & Cheung, D. W. (2013). A survey on cloud interoperability: taxonomies, standards, and practice. *ACM SIGMETRICS Performance Evaluation Review*, *40*(4), 13-22. https://doi.org/10.1145/2479942.2479945

[180]. Zhao, Z., Tian, Y., Hong, F., Huang, H., & Zhou, S. (2020). Trawler Fishing Track Interpolation using LSTM for Satellite-based VMS Traces. *Global Oceans 2020: Singapore – U.S. Gulf Coast*, *NA*(NA), 1-4. https://doi.org/10.1109/ieeeconf38699.2020.9389435

[181]. Zhitluhina, J. V., Perov, D. V., Rinkevich, A. B., Smorodinsky, Y. G., Kroning, M., & Permikin, V. S. (2007). Characterisation of steels with microdefects using a laser interferometry technique. *Insight - Non-Destructive Testing and Condition Monitoring*, *49*(5), 267-271. https://doi.org/10.1784/insi.2007.49.5.267