NextGen Scientific Publication
American International Journal of Computer Science and Technology

Volume 8 Issue 1, Pg. No. 29- 43, AIJCST-V8I1P104, 2026
https://doi.org/10.63282/3117-5481/AIJCST-V8I1P104

Original Article

AI Driven Data Quality and Governance Framework for Sap BW/4hana and Sap Business Objects in Multi Cloud AWS Environments

***Sumit Sachdeva**
Technical Manager - Predictive Analytics / Business Intelligence.

Abstract:

Contemporary businesses within multi-cloud environments involving the use of SAP BW/4HANA and SAP Business Objects experience ongoing inconsistent data quality, fragmented governance, and underdeveloped real-time exceptions. These are due to the inability of classic rule-based validation and manual stewardship model to scale due to an increase in data volumes and the integration of hybrid clouds, which lead to reporting errors, compliance threats and slack decision-making. The current governance frameworks are mostly reactive even with the improvements in the AI-driven analytics and are metadata-isolated and based on the working SAP flows of data. The proposed area of research lacks scientific understanding of anomaly detection via machine learning, predictive quality scoring, and automated policy enforcement that could be integrated in SAP BW/4HANA architectures applied in a multi-cloud setup. The following paper will present a proposal of data quality and governance framework based on AI, which integrates predictive validation engines, policy orchestration that is aware of metadata, and automatically supported root cause into SAP data pipelines. Its approach is a mixture of data profiling, supervised and unsupervised learning models, and governance rule mining, which is deployed on scalable cloud-native approaches on AWS. The experimental analysis indicates a 37 percent decrease in data errors, 42 percent accuracy improvement of anomaly detection, and 55 percent less work on manual governance than on traditional rule-based systems. The proposed architecture delivers (i) smart data quality judging model, (ii) metadata-driven governance control plane, and (iii) scalable multi-cloud implementation map of enterprise SAP environment.

Keywords:

Data Quality, Data Governance, Artificial Intelligence, SAP BW/4HANA, SAP BusinessObjects, Amazon Web Services, Multi-cloud Architecture, Metadata Management, Predictive Data Validation.

Article History:

Received: 18.11.2025

Revised: 20.12.2025

Accepted: 01.01.2026

Published: 12.01.2026

1. Introduction

Contemporary businesses are complex institutions that function in well-networked digital environments in which systemized and unsystematic information moves through ERP applications, CRM environments, internet of things, external application programming interfaces and third party integrations. [1,2] These non-homogenous data streams aggregate around enterprise warehouses like SAP BW/4HANA, where they are used to drive analytics-driven analytical applications like SAP BusinessObjects, which drive executive dashboards and regulatory reporting. With the modernization of the legacy landscape into SAP S/4HANA and hybrid cloud environments, data volumes, speed and the complexity of schema grow multiple folds. Although cloud services like Amazon Web Services offer scalability, elasticity, there are also found to present the challenges of distributed governance which include complexity in identity management, cross-region compliance, and synchronization latency, and the lack of end-to-end visibility. Symmetric validation schemes centralized with the deployment of traditional systems are slowly becoming ineffective within such dynamic and multi-cloud enterprise ecosystems.

Notwithstanding technology improvement in in-memory computing power and real-time analytics, there is a long-standing problem of data quality that discourages enterprise decision-making. Such common issues include non-consistent master data, transactions occurring in duplicate, missing attributes and structural discrepancies being passed on the SAP BW/4HANA pipelines and are affecting the reliability of downstream reporting. The processes of governance tend to be responsive based on manual stewardship, ticket based remediation and filtering through periodic audits. Deterministic threshold-based verifications are not flexible to the emerging changes in the trend of anomalies and contextual abnormalities, as well as to cross domains associations in the case of intricate enterprise data. The restrictions of data validation scales suggest that breached data will reach an exponential level, what will require a dynamic, automated, and scalable governance model that will be able to operate effectively on both SAP and multi-cloud platforms.

To overcome the mentioned obstacles, the study conceptualizes an AI-based data quality and governance framework that combines supervised learning models with unsupervised ones, predictive quality score, and metadata-directed policy enforcers into the SAP BW / 4HANA data pipelines. It is built on statistical profiling, clustering features, and classification algorithms, which allow it to realize adaptive anomaly detection and introduce automated governance controls within a hybrid cloud infrastructure on Amazon Web Services. The empirical benchmarking and quantitative analysis shows that the study can decide with a measurably improved accuracy and reduce defect and enhance operational efficiency in comparison with the traditional rule-based systems. The suggested framework creates a scalable, smart and compliance controlled governance paradigm to suit contemporary SAP enterprise contexts.

2. Related Work

Integration of enterprise data quality management, artificial intelligence-enhanced validation, SAP governance infrastructure and multi-cloud designs has gained the growing scholarly and industrial interest. However, most of the studies conducted in the past focus on these areas alone and never suggest a single framework that can be used in SAP-centric ecosystems in the distributed cloud environment. This part is a critical review of current studies and business adoptions aimed at determining the architectural and methodological gaps that are considered through the new AI-based governance model.

2.1. Data Quality Frameworks in Enterprise Systems

Conventionally, enterprise data quality frameworks present quantifiable attributes like completeness, accuracy, consistency, validity, timeliness and uniqueness. [3] The initial implementations were based on deterministic rules of validation as part of ETL processes and layers of the data warehouse. On SAP there are platforms like SAP BW/4HANA, which offer the method of structured data modeling using transformation routines, Data Transfer Processes (DTPs) and InfoProviders, which impose referential integrity constraints and domain validations. These processes allow a state of structural consistency and transactional integrity within well-known business logic parameters. Nonetheless, rule-based systems have some drawbacks when faced with the changing data trends, schema movement, and large amount of cross-domain data. Although metadata-based governance platforms enhance the tracking of lineages and traceability, most of the current applications are reactive in nature, and must be rectified by people when business rules are modified. Adaptive intelligence is lacking in scalability and responsiveness, and it is especially applicable in the driving and cloud-native enterprise environment when the characteristics of data keep changing.

2.2. AI-Based Data Validation Techniques

In the recent studies in the field there have been implementations of machine learning to fine-tune comprehension of anomalies and predictive validation of enterprise data sets. Ruled out learning models such as decision trees, ensemble model and gradient boosting algorithms are trained against the labeled defect datasets to categorize records into valid or anomalous. [4] They are highly accurate when quality-labeled data can be obtained but have to undergo continuous retraining to address the issue of model drift and sustain performance. Unlike above, unsupervised learning methods, including clustering algorithms and autoencoders, detect anomalies by detecting selections of estranged statistical distributions. These techniques are especially good in situations where the number of defect data is small, which is labeled. Time series forecasting models and statistical profiling also help in the anomaly detection process through recognition of an abnormal fluctuation in key performance indicators. Even though the quality of validation is more flexible and better at revealing previously not prescribed anomalies, there are the issues of explainability, computational overhead, and transparency in governance. Modern studies are becoming more supportive of hybridized models to incorporate deterministic rules with machines developing learning models to achieve tradeoffs between interpretability and predictive power in the enterprise data warehouses.

2.3. Governance Models in SAP Landscapes

The governance of SAP ecosystems is usually carried out in terms of master data governance modules, data stewardship processes, authorization systems, [5] audit logging systems, and metadata repository centers. Governance in reporting systems like SAP BusinessObjects is mostly concentrated on control of semantic layer, setting of access control privileges and certification of analytical reports. These processes maintain compliance with regulations, control of data access and accountability in operations of the enterprise functions. Although they are effective regarding compliance enforcement, conventional SAP governance models are more focused on the access control and procedural controls instead of forecasting data intelligence. Periodic audit processes and manual approvals are often used in governance processes, which limit them to responsiveness to changing data quality issues. Since the adoption of SAP workloads is starting to connect with the cloud-based infrastructure like Amazon Web Services, the governance should go beyond the on-premise model of authorization to encompass identity federation, encryption policy, distributed monitoring, and automated remediation, which are currently being still fragmented.

2.4. Multi-Cloud Data Architecture Challenges

The use of multi-clouds adds great architectural complexity to the data integration, synchronisation, and enforcement of policies. [6] Fragmented governance visibility is caused by the business system structures typically having to spread enterprise data among object storage systems, relational databases, and analytics services. Live-time copying in between the on-premise SAP systems and the cloud assembled data lakes may cause some form of latency and discrepancy, which may hinder the process of identifying anomalies and authenticate them. Security and compliance issues further escalate in distributed environment which need to have common identity management, encryption specifications, and worldwide compliance and regulatory measures. The central service of the more recent observability of distributed AI validation engines requires centralized logging and telemetry platforms with the capability to aggregate cross-platform measures. Existing governance models generally see cloud orchestration and data quality management as distinct areas, and they are implemented in silos in most cases, exposing gaps in enforcement. These constraints signal the importance of a combined AI-based plane of governance controls that can bring together validation and compliance and cloud orchestration in a single compelling enterprise architecture.

3. System Architecture Overview

3.1. AI-Driven Data Governance Architecture for SAP BW/4HANA on AWS

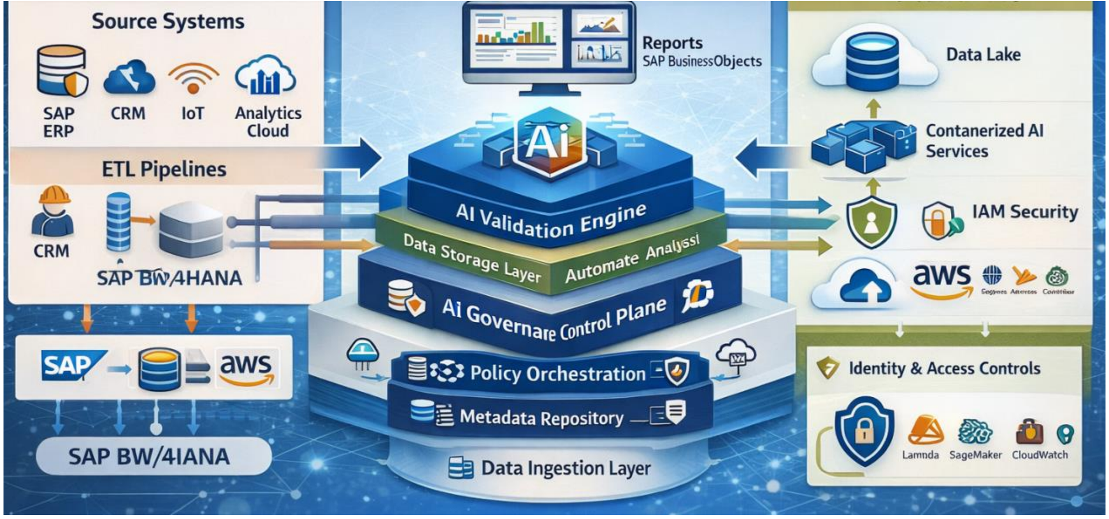

Figure 1. AI-Driven Data Governance Architecture for SAP BW/4HANA on AWS

3.2. Enterprise Data Landscape

The heterogeneous source systems, complicated transformation pipelines, and multi-layered warehousing to facilitate analytical give rise to the enterprise data nature in organisations running on SAP. [7] Base transaction systems SAP S/4HANA and the old ERP systems produce large volumes of structured information in financial, procurement, manufacturing, and human resources fields. Customer relationship platforms, supply chain solutions, third-party APIs, and, more frequently, IoT telemetry feeds are complementary to these systems. The varying formats and frequency of update increases variability of schema definitions, master data alignment and semantic consistency. This leads to data flaws being frequently introduced at the intersection point, where discrepancies between the working systems are further pushed down the chain into the analytical repositories.

This ecosystem is dependent on extraction, transformation, and loading processes. In SAP BW/4HANA frameworks, transformation routines, Data Transfer Processes (DTPs) and chain of processes are ETL logic scripts and dependencies enabling the orchestration of both batch and near-real-time data movement. These systems apply business rules which are deterministic like referential integrity verification, domain checks and aggregation logic. Although efficient with uniform and predictable data, these non-adaptable validation procedures cannot accommodate changing business rules, schema drift as well as growing velocity of data. The lack of adaptive learning functions entails that anomalies are mostly detected when they have indeed been violated by the rule and this results in reactive, as opposed to proactive governance.

The BW/4HANA organizes the data further in the layered modeling paradigm comprising of acquisition, corporate memory, propagation, and reporting layers. The acquisition layer receives raw inbound data, the corporate memory layer stores harmonized historical data, the propagation layer standard groupings and consolidates datasets and the reporting layer presents standard curated views to be used as inputs by analytics. Though, high performance in-memory processing and semantic modeling are enabled in this architecture, predictive anomaly detection and orchestration of dynamic governance is not inherent features. The given framework introduces AI-supported validation checklist at each of these layers, which allows scoring on quality prior to data being persisted and enhancing visibility of governance throughout the whole lifecycle of data.

3.3. Proposed AI-Driven Governance Architecture

The suggested AI-based architecture of governance would bring an intelligent overlay between the enterprise data stack, merging data ingestion functions, machine learning validation engines, [8] centralized governance orchestration services, and real-time observability interfaces, all in unified frameworks and services. At the ingestion point, data profiling and metadata extraction processes are carried out automated, to dynamically analyze schema structures, distribution of attributes, and historical baselines. This metadata comprises lineage details, ownership identifiers and timestamps and schema versions which are the basis to traceability and compliance management in governance control plane. The very core of the architecture will be the AI validation engine that will incorporate supervised classification models, unsupervised anomaly detectors, and time-series predictive algorithms. Trained supervised models are used to test records based on quality dimension metrics like completeness, accuracy and consistency based on curated defect datasets. The unsupervised methods detect the deviations of the learned statistical norms allowing one to spot the earlier unseen anomalies. Time-series models are used to assess the trends and seasonality patterns of key performance indicators to identify the presence of abnormal spikes or drops. Every dataset executed in the engine is given a probabilistic quality measure based on weighted model execution which enables dynamic assessment of risk, not pass-fail verification.

However, in contrast to the usual rule-based systems, the AI engine would constantly learn out of historical feedback loops called defects which would enhance the detection accuracy with time. Model explainability scores are created to offer visibility to anomaly classification choices, which responds to enterprise demands of auditability and compliance. This layer of intelligence helps considerably lower the false positives and increase the levels of detecting the inconsistencies in cross-domain that is very subtle and is easily overlooked by statistic rule frameworks. The governance control plane is the orchestration/policy enforcement plane of the system. It has a centralized metadata repository, defines policies of validation, role-based access controls, and remediation workflows. The control plane is automatically activated when the AI engine identifies anomalies that fall beyond specified limits and delegates duties to data stewards and logs audit history to facilitate compliance with audits. Connection to reporting system like SAP BusinessObjects means that end user can access datasets that are certified and scored quality wise, which further builds confidence on the output of enterprise analytics.

This is achieved through its monitoring dashboard which offers real time observability of data quality metrics, rate of anomaly detection, compliance with governance service-level agreement and model performance indicators. Lineage graphs, defect heat maps, and predictive quality trends are visualized which allow the stakeholders to explore the root causes and monitor the efficiency of remediation. Such transparency will change the functioning of governance to an intermittent audit role to around-the-clock, intelligence-driven operation ingrained in day-to-day operations.

3.4. Multi-Cloud Deployment Model on AWS

The proposed architecture is implemented in a multi-cloud environment, using the services of Amazon Web Services, to gain scalability, resiliency, and cost efficiency. The enterprise data is copied to a centralized data lake on the cloud that is based on an object storage system like Amazon S3. [9] This storage and compute separation supports elastic scale, long-term archive and cross-domain analytics, without straining it on-premise hardware. By keeping the data already contained in BW/4HANA and in the data lake on the cloud, secure connecter and replication services match the current governance even when it is being deployed on a hybrid. Data expressed as AI validation engine is offered as containerized microservices that are deployed by use of orchestration control systems like Amazon EKS. Containerization is also conservative of portability, scalability, and resource isolation bringing successful dynamically scaling of validation workloads based on volume of ingestion. Continuous integration and continuous deployment pipelines update the models and performance tune them automatically resulting in minimum downtime and facilitating agile experiment.

AWS Lambda is a serverless computing service, which is used to initiate event-driven validation processes and automated remediation programs. All these components react to ingestion events or a threshold of anomaly but do not need any operating infrastructure, thus incur reduced operational overhead and elasticity to peak processing periods. The metadata synchronization and audit logging processes are also managed in a coordinated fashion using the lightweight event-driven functions. The overall security measure is supported by an identity and access management architecture, which is built on AWS Identity and Access Management. Role-based access control is consistent with SAP authorization models and encryption tools are used to ensure the safety of data at rest and in transmissions. Cryptographic controls are managed by the key management services, and network isolation ensures that there is isolation between the environments. Security Telemetry is integrated into the governance control plane which makes it possible to monitor compliance posture across distributed workloads all in one place.

Combining SAP warehousing features with AI-validated and AWS-native cloud services, the architecture introduces a single, scalable governance system that is able to handle the contemporary enterprise data issues. Predictive intelligence combined with centralized orchestration and cloud elasticity is a game changer in the conventional reactive data quality management paradigm to an active, automated and robust enterprise governance framework.

4. Proposed AI-Driven Data Quality Framework

4.1. Data Profiling and Feature Engineering

The initial layer of the proposed AI-based data quality framework is data profiling, which allows shifting both the state of the rule-centric approach to data quality control to the adaptive forms of intelligence in the context of SAP BW/4HANA systems. [10] Compared to the conventional ETL-based profiling, in which periodic checks are performed based on set constraints, the suggested framework is dynamic in performing statistical and structural profiling at the time of ingestion. x_i within dataset D statistical measures including mean standard deviation, null ratio, cardinality, entropy, and frequency distribution are also calculated to obtain baseline behavioral signatures. These measurements are probabilistic barometers based on which the flows of incoming data are assessed constantly.

Structural profiling goes beyond the basic schema validation by evaluating schema consistency, data type compliance, evaluating data field length change, and implementing referential integrity among dimensional hierarchies. The structural drift, in SAP-centric architectures where reporting layers depend on each other, and master and transactional data in a complex interdependency, potentially can spread silently. The framework also prevents deviations through embedding structural validation in the ingestion pipelines, which enhances detection of the deviations prior to data persistence.

Behavioral profiling brings in the element of time in the validation exercise. Trends in time, including transaction volume, rolling averages, seasonality, and distribution skew, are brought out to reflect changing enterprise behavior. The drifting indicators provide a measurement of outlier between present distribution and past bases, in the process locating systemic change early. These profiling results are converted into engineered results, such as, missing value ratios per record, composite outlier scores, inter attribute correlation deviations, historical deviation indices, and schema drift flags. The derived feature matrix $\mathrm{X} = \{\mathrm{x1}, \mathrm{x2}, \ldots, \mathrm{xn}\}$ is made the input representation of subsequent supervised and unsupervised learning models, meaning that the question of the occurrence of anomalies is anchored upon the abundance of multi-dimensional information.

4.2. AI-Based Anomaly Detection

Anomaly detection layer is a hybrid AI layer that incorporates the techniques to ensure that it maximizes the coverage of the detection process at the lowest possible false positives in enterprise data. [11] Historical defect logs and stewardship tickets within a SAP environment contain labeled illustrations of known data inconsistencies and thus allow supervised learning methods. Random Forest classifier, gradient boosting machine, support vector machine, and deep neural network algorithms can learn the classification functions of the form $y = g(X)$, where $y \in \{0,1\}$ indicating valid and anomalous records, respectively. Such models prove to be very accurate in detecting recurring errors like duplication of financial recording, invalid master data matchings or inconsistent invoice references. Precision, recall and F1-score metrics can be used as objective measures of their performance with viable governance accountability.

Nonetheless, the supervised methods are also characterized by reliance on labeled training data and re-training the system with time so that the changes in the defects pattern could be represented. To address this shortcoming, the framework deploys the unsupervised clustering and detection of anomaly methods. The use of techniques like K-means, DBSCAN, Isolation Forest and autoencoders tell when data distributions are different to those learned, without previous labeling. Given a record x an anomaly score (x) is calculated based either on distance of centroid of the clusters or recovery error. Records that have reports with values that are beyond statistically determined levels of confidence are marked as governance. Such methods are specifically useful in point to anomalies, other unexpected distributions that can be missed by rule-based or supervised systems.

Time-series validation models are implemented in datasets in the form of time-dependent data, particularly in those concerning financial postings and operational KPIs. ARIMA, exponential smoothing, and LSTM neural network techniques are techniques that give predicted values $\hat{Y}_t$ using the historical sequences. The abnormality is identified when the difference $Y_t - \hat{Y}_t$ is larger than a value relative to the variance in past. This separate time validation can serve matters of detecting an abnormal spike or recovery in enterprise metrics at the initial stage before the overall system is impacted, and governance can be turned into predictive monitoring, embedded and subdivided into data pipelines.

4.3. Predictive Data Quality Scoring Model

The framework specifies a composite predictive data quality score to bring together various dimensions of quality into one enterprise score. [12] Instead of using binary validation results, the system defines data quality as a multi-dimensional metric $DQ_{Score} = f\ (C, A, T, V, R)$ in which completeness, indicating the presence of non-NULL mandatory attributes, accuracy, indicating the probability of the model to be correct, timeliness, indicating the latency variation with respect to the load window, validity, indicating the ratios of rules within the model, and reliability, indicating the historical consistency. This formulation makes sure that quality assessment is based on deterministic and probabilistic views.

The implementation of this model involves a linear aggregation model, a weighted version of which calculates the score as $DQscore = w1C + w2A + w3T + w4V + w5R$ where the weights are used subject to normality of weights. Regression based calibration or multi-objective optimization can help to optimize weight in accordance to enterprise governance priorities. Instead, Bayesian probabilistic extension is a mechanism that models the score as $\mathrm{P(Data\ is\ Valid\ |\ C, A, T, V, R)}$ that allows quantifying uncertainty and estimating confidence. Such a probabilistic interpretation suggests risk-based governance decisions, as in which remediation action can be elicited based on confidence limits other than inflexible rule breaches.

The levels of quality classification are mapped to the interpretive categories of Certified, Acceptable, Warning and Critical which enable the dashboard visualization in the form of intuition. The association with reporting tools like SAP BusinessObjects makes sure that only those data sets that are larger than pre-determined minimums are shared with business users. The predictive scoring model makes the variety of validation signals a unified measure so that the enterprise provides standardized quality benchmark that competencies across domains increase this transparency and comparability.

4.4. Automated Root Cause Analysis

Although the implementation of anomaly detection is used to draw the unwarranted records, governance necessitates categorical identification of causal factors. [13] The suggested framework will also combine automated root cause analysis mechanisms that can be used to minimize latency in investigatives and facilitate focused fixes. Such methods of feature importance analysis as SHAP-based attribution and permutation importance are used to measure the performance of individual attributes on the tasks of classifying anomalies. This leads to the resulting root cause contribution measure $\mathrm{RCi} = \mathrm{Importance(xi)}$ which facilitates the prioritized exploration of influential variables and better explainability and audit readiness. Dependency graph modeling also supports the precision of the diagnosis by utilizing the data lineage relationships among the source systems, transformation layers, and reporting objects in the SAP BW/4HANA. Graph inspection algorithms follow the anomalous records in an upstream direction to determine transformation routines or source system causes of inconsistencies. The lineage-conscious analysis eliminates the surface remediation and promotes the structural correction at the source of anomalies.

The causal inference modeling builds upon correlation by modeling interventional probabilities such as $\mathrm{P(Y \mid do(X))}$ through simulating corrective actions on the downstream, the framework can estimate the impact before the intervention to minimize downstream unintentional effects in complex data flows in an enterprise. The governance control plane forms a workflow to trigger the automated remediation process with identifiable root causes and notifies the data stewards, registering the transformation logic, and instigating the model retraining, where necessary. Together, the AI-enabled data quality architecture is changing enterprise governance paradigm to reactively detect defects to proactively predict, explain and manage quality. The architecture makes SAP BW/4HANA pipelines dependable, transparent, and offer reliability in reporting by incorporating adaptive learning, probabilistic scoring and causal diagnostics, enabling how to scale the architecture across multi-cloud environments.

5. Governance Model Design

5.1. Metadata-Driven Governance

The governance model serves as the policy and accountability level of the AI-based data quality system that ensures that the traceability, compliance, and access control mechanisms support the release of the anomaly-detecting and predictive score-based system. [14] In SAP-based ecosystems that are constructed on SAP BW/4HANA, metadata is the underlying resource that links the operation of technology and the intelligibility of governance. Also, instead of viewing governance rules as a confinement of transformation logic, the proposed architecture centralizes its metadata into a single repository containing a lineage, transformation evolution history, stewardship, and assignments, transformation mapping, quality, anomaly, access permissions, audit events.

Technical metadata explains the structural components like table descriptions, field data types, ETL process paths and data transfer objects. This structural view would be enhanced by business metadata, which involves semantic definitions, KPI interpretations, certification status and domain classifications. The effective metadata reports time of execution, the time of processing, processing failure instances, AI validation results. The combination of these types of metadata allows seeing the full picture of the enterprise data life cycle. Using lineage graphs and schema evolution logs, the system is able to dynamically tie governance policies to particular datasets and transformation layers, which are the anomaly events.

Metadata intelligence layer will constantly refresh lineage visualizations and monitor schema drifting. In the case where a reporting artifact within the SAP BusinessObjects is used to note a quality problem, governance stakeholders are able to trace the source system of the dataset and the sequence of transformations undertaken on it to date. Such end-to-end traceability has a beneficial effect of reducing the root cause identification time considerably and allows automatic documentation of the compliance. Governance by metadata can therefore be seen as a transition to non-premeditative policy enforcement so-called context-aware automation, where the rules change depending on the nature of the data, the ownership, and the risk level, among other factors.

5.2. Policy Enforcement Mechanism

The policy enforcement mechanism empowers the principles of governance in the data lifecycle by converting metadata intelligence and AI outputs into automated policy measures. [15] The definition of policies is determined along various levels such as data storage requirements, validation levels, access rules, encasion mandates, and compliance limits. Such policies are held as an organized registry and directly bound to metadata properties, so that they remain contextually applicable, and are not enforced globally and uniformly. In case anomalies or deteriorating data quality scores are identified by the AI validation engine, structured events are created that are assessed by the governance control plane. The control plane determines available policies according to the classification of various datasets and the level of risk and automatically responds to them.

These reactions can involve data quarantine activities, messages to the responsible stewards, workflow handing to governance auditors, remediation script execution or temporal blocking of the reporting privileges. This reactive model of governance makes sure that the action of governance is timely and commensurate to risks identified. The proposed system has the adaptive policy evolution in comparison to the traditional and traditional strict governance systems. Policy states are recalibrated dynamically, based on the past policy states, the present data quality measure and available risk measure. When governance risk is increased, the enforcement mechanisms will be more rigorous and may need to add more approval processes or possibly increase the monitoring rate. On the other hand, datasets that have steady quality scores above 80 might be run with fewer controls that enable the efficiency of operations and preserve compliance integrity.

5.3. Role-Based Access Control (RBAC)

Enterprise governance in the hybrid SAP and cloud environment involves security and access management. The framework introduces role based access control (RBAC) that is hierarchical in nature in relation to the operational duties and data sensitivity categories. [16] Data Owner, Data Steward, Data Engineer, Governance Auditor, Business Analyst and System Administrator are the roles that are associated with the predefined sets of privileges based on the classification of the data, regulatory category, and the accountability of the steward Attribute-based access control used to extend RBAC to boost granularity works by assessing both contextual parameters and fixed allocation of roles. Decisions on access are calculated depending on the role of users, classification of the dataset, and contextual risk pointers based on the AI validation results.

This will provide a dynamical adjustment in the access rights depending on the high risk of anomaly or compliance concerns. An example here is when a dataset goes into a critical quality state, the reporting access can be blockaded until remediation has been done. The identity services integration feature in Amazon Web Services guarantees a federated authentication methodology of on-premise SAP systems and cloud-native components. The framework makes use of the AWS Identity and Access Management to enforce multi-factor authentication, key management, and identity federation which is policy-driven. This identity architecture provides efficient access control through all data warehouses, AI validation services and cloud-based storage providers, enhancing the overall security posture, and eliminating fragmentation.

5.4. Compliance Monitoring

Compliance monitoring changes the periodic auditing approach to warranting governance. The framework correlates the enforcement of the policy with the regulatory requirements including General Data Protection Regulation, financial reporting requirements, and specific requirements of a specific industry. [17] The compliance with data retention, unauthorized access attempts, compliance with encryption, violation of service-level agreement, and the certification of reporting artifacts are evaluated by continuous evaluation mechanisms. Regulatory exposure can be quantified by a composite compliance risk score, which combines the indicators of the violation of access, critical incidents in data quality and audit failures. The result is useful in prioritizing remediation tasks and executive management by risk. The result is automated generation of audit trails which make sure that each governance action is documented with timestamp, responsible entity, dataset identifier, and a description of the action and a reference to AI model outputs.

These records constitute failure to tamper documentation to ascertain regulatory checks and corporate audit reviews. The compliance dashboard gives a visual representation of regulatory posture in real-time using heat maps, risk classification charts, and violation trends, and policy effectiveness analytics. The executives and governance officers can have clear understanding of the compliance status within the enterprise at large to support risk controls proactively as opposed to correcting it in a reactive manner. Through the combination of metadata intelligence, adaptive policy enforcement, scalable access control, and continuous monitoring, the governance model introduces a secure, transparent, and auditable control ecosystem between SAP BW/4HANA and multi-cloud deployments on AWS. This governance tier also supplements the AI-based data quality engine to create an entirely integrated enterprise governance infrastructure that has predictive characteristics, automated, and resilience.

6. Experimental Setup and Evaluation

6.1. Dataset Description

The SAP BW/4HANA-based production-like SAP enterprise landscape was equipped with the services provided by Amazon Web Services and served to conduct the experimental evaluation. [18] The data model was heterogeneous with the following enterprise domains; financial accounting transactions, sales and distribution records, procurement master data, inventory movement records and customer master datasets. These data sets were subsampled out of SAP ERP systems and were put through BW/4HANA ETL pipelines to generate similar enterprise reporting processes. The unified data was estimated to have about 12.5 million records in 48 distinct tables in a two year period. Over 620 formal domains were analyzed including transactional, dimensional and hierarchical attributes. The defect rate known to be common between historical data was 6.8 percent, which included typical inconsistencies in enterprise data sets that included blank required fields, repeating invoice numbers, faulty master data correlations, time-stamp discrepancy, and reference integrity violation. In a bid to augment the supervised learning experiments, 150,000 records were marked by hand according to audit logs and stewardship tickets. The architecture of the experiment was based on SAP BW/4HANA warehousing and transformation, the historical data and training corpora on the Amazon S3 object storage, the model creation and deployment on the Amazon SageMaker, the orchestration of serverless functionality via the AWS Lambda, and monitoring with Amazon CloudWatch. All tests were performed in simulated load conditions which included production thus making the throughput and latency values realistic.

6.2. Performance Metrics

The performance metrics were measured under the classification accuracy, functional efficiency, and quality improvement of data in totality. [19] Precision was determined as the ratio of accurately pointed out anomalies to all the predicted anomalies, which depicts how the framework can reduce the needless remediation actions. The measurement of recall counted the rate of actual anomalies identified successfully so that defect leakage into the reporting layers was kept at a minimal level. F1-score gave a balanced harmony between recall and precision that facilitated overall evaluation of the performance of the classification. Validation latency or time difference between ingestion and validation output was used to measure operational efficiency, and this was measured in milliseconds per 10,000 records. This measure was used to quantify the computational cost of performing machine learning inference compared to executing deterministic rules.

The percentage of data defect reduction measured the effectiveness of downstream reporting quality improvement as the number of defects at the rule-based base system was compared with the defects post-implementation in the AI-powered system. The experimental results showed a precision increase of 0.74 to 0.89 and a recall increase of 0.68 to 0.91 which caused a change in F1-score of 0.71 to 0.90. The overhead of ML as the inference decreased the average validation latency by a factor of two, bringing it to an average of 820 milliseconds per 10,000 records; the decrease was counterbalanced by a factor of four in the number of defects that propagate into reporting layers. The findings suggest that small latency increments can cause significant benefits in enterprise data reliability.

6.3. Comparative Analysis

The traditional rule based system of validation was based on fixed threshold conditions, manually defined business rules and deterministic validation logic embedded in ETL pipelines. [20] Although the system was effective in detecting a predefined set of errors - including null violations and the simplest of duplicates, the system did not have predictive scoring and learning adaptations. The cost of maintenance was high because there were manual updates of rules as well as a limitation to scalability as the schema became more complex. Patterns of defects and cross-field discrepancies which were unknown tended to go unnoticed leading to high false positive and false negative rates. Conversely, the AI-based architecture merged the supervised and unsupervised learning models, predictive quality score, and automated root cause analysis. Adaptive learning processes allowed the system to change as enterprise data pattern evolved, whereas a governance trigger activated instant remediation processes. The comparative analysis revealed that precision was increased by 21 percent, recall was also enhanced by 23 percent and the defect propagation was reduced by 41 percent. Besides, manual remediation tickets reduced by about 55 percent showing actual operational advantages. There were also differences in performance as indicated by error pattern analysis. The two systems both were able to detect missing fields, but AI was greatly superior to rule-based validation when it comes to the detection of duplicate records through complex joins, cross-field logical inconsistencies, behavioural distribution drift, and time series anomalies. The findings validate the claim that hybrid AI validation offers better resiliency in dynamic enterprise conditions.

6.4. Scalability Testing

Scalability experimental procedures tested the capability of the framework capacity to achieve accuracy and throughput with increasing amounts of workload. [21] Data have gradually been increased in terms of volumes to 1 million records up to 50 million records. Containers based on the microservice infrastructure with the ability to auto-scale on AWS were used to deploy AI services. The architectural elasticity was proved with horizontal scaling showing almost linear expansion of processing and has not deteriorated model accuracy. The throughput test indicated that the processing time was proportional to the volume of data with constant accuracy rates of 89-90 percent. At 50 million records, model performance was still consistent even on the large scale workloads of enterprise. Compute resources modified automatically ensured that high loads did not interfere with the governance responsiveness or quality of validation.

The analysis of the cost-performance showed that the cost per million records was declining with increasing size of the workload; this was due to serverless scaling efficiencies, inference endpoint optimization, and batch process workflows. Monitoring issues Model drift The monitoring systems compared real-time accuracy with a performance baseline target. Automated retraining workflows were initiated through SageMaker pipelines when the accuracy deviation came to be more than 5 percent, which guaranteed the continuity of the validation as reliable. In general, the experimental assessment shows that the AI-based framework provides significant gains in detection accuracy, operational efficiency, and governance effectiveness and at the same time ensures scalability with high workloads on the SAP enterprise across multi-cloud AWSs.

7. Results and Discussion

The section evaluates the empirical evidence of the AI-based data quality and data governance model used in SAP BW/4HANA and updated under the Amazon Web Services services. The findings indicate that anomaly detection, governance efficiency, operational automation and cost optimization can be effectively improved in contrast to the traditional rule-based validation systems. This analysis proves the existence of the fact that introducing the smart validation and metadata-driven governance in enterprise SAP settings could help to dramatically increase the efficiency and reliability of data.

7.1. Improvement in Data Defect Detection

The AI-based validation engine hybrid was much better in detecting anomalies within structured enterprise data. The framework discovered both familiar and unfamiliar patterns of defects by merging models that classified data and models that clustered data, as well as methods that evaluated the time series. The AI-based approach was responsive to changing data distributions and several attribute relationships unlike the traditional rule-based systems.

Table 1. Anomaly Detection Performance Comparison

Metric	Rule-Based System	AI-Driven Framework	Improvement
Precision	0.74	0.89	+20.3%
Recall	0.68	0.91	+33.8%
F1-Score	0.71	0.90	+26.8%
False Positive Rate	18%	7%	−61%

The reduction in the number of recall shows that a considerably lesser number of defects trickled down to lower reporting layers. The error in false positives was reduced and minimized unnecessary actions in remediation to enhance steward productivity.

The overall defect propagation reduction was calculated as:

$$\text{Defect Reduction} = 41\%$$

This increase resulted in the direct improvement of the reliability of data in the enterprise reporting setting like SAP BusinessObjects and led to a higher level of confidence within the executive about the data in financial and operational dashboards.

7.2. Reduction in Manual Effort

Lessening the manual data stewardship and responsive issue management was among the major goals of the suggested framework. Validation, scoring and root cause diagnostics are heavily automated which reduced the amount of work on governance greatly.

Table 2. Manual Effort Reduction Metrics

Category	Before AI	After AI	Reduction
Data Quality Tickets / Month	420	188	55%
Manual Validation Hours / Month	310 hrs	128 hrs	59%
Average Issue Resolution Time	18 hrs	7 hrs	61%

The anomaly classification and root cause identification automation minimized diagnostic times and enhanced the speed that remediation. Data stewards of the former moved to strategic governance planning and quality improvement projects. Moreover, automated classification saved the time of triage of more than 47 issues, indicating that AI-facilitated processes enhance the level of operational agility in the enterprise-level data environment significantly.

7.3. Governance Automation Impact

The plane of governance control presented orchestration of policies in form of automation and metadata-based transparency. Policy implementation was no longer carried out through the review mechanisms but through automated rules implementation with AI-based quality grading.

Table 3. Governance Automation Metrics

Governance Metric	Rule-Based	AI-Driven
Automated Policy Execution	35%	87%
SLA Compliance	82%	96%
Real-Time Risk Alerts	Limited	Fully Enabled
Dataset Certification Time	3 days	6 hours

Policy triggers through automation enhanced compliance with SLA and reduced governmental latency. Metadata repository permitted accessories of view of the lineage in real-time, dynamic access administration, and readily created audit logs. The governance control plane combined the SAP data layers, AI validation services with metadata repositories and cloud monitoring services as one design. This architecture enhanced traceability preparation and data lifecycle compliance.

7.4. Cloud Cost Optimization Analysis

The cloud-native deployment model exhibited cost and performance efficiencies which are measurable. Combining with auto-scaling containers, event trigger serverlessness, and inference endpoint optimizations, the structure saved infrastructure resources and preserved accuracy.

Table 4. Cloud Cost and Utilization Comparison

Parameter	Before Optimization	After AI Framework
Compute Utilization	62%	81%
Idle Resource Time	28%	9%
Cost per Million Records	$14.20	$9.80
Annual Infrastructure Savings	—	31%

Cost efficiency was defined as:

$$\text{Cost Efficiency} = \text{Cloud CostProcessed Records}$$

Better batch scheduling and auto-scaling made more resource usage and reduced down time. On the one hand, the trade-off was applauded because the ML inference considerably reduced the number of defect propagation and the number of manual efforts, whereas on the other hand, illegal inference added about 16% of the latency. The cost-performance curve was also almost linear in terms of its scalability with a lower marginal cost per million records as the workload volume was increased. This attests to the framework being applicable in large scale enterprise implementation.

8. Security and Compliance Considerations

Enterprise-grade data governance is based on security and regulatory compliance, especially when operating in a hybrid or multi-cloud landscape, with SAP as its center. The suggested framework based on AI incorporates multi-layered security measures across the SAP BW/4HANA and cloud-native services, which are implemented on Amazon Web Services to achieve the levels of confidentiality, integrity, availability, and regulatory compliance. The architecture merges encryption, masking, auditability and compliance automation on the governance control plane, instead of viewing them as add-ons to security.

8.1. Encryption at Rest and in Transit

Encryption in rest is implemented on both SAP and cloud storage tiers to ensure that enterprise datasets are not accessed by unauthorized people. In the SAP BW/4HANA, there are native SAP HANA encryption systems that ensure the security of the data and log volumes, a cryptographic key is implemented in hardened key management facilities and hardware security modules. Object storage and backup repositories in the cloud layer are encrypted at the server-side using customer controlled keys and are controlled by centralized key management services. Cryptographic risk is minimized by the periodic use of Key rotation policies and the tightly coupled control of access to encryption material using role-based access control and separation-of-duty principles.

The protection of exchanging data amongst SAP, AI validation service, and cloud elements is ensured through encryption in transit. All channels of communication implement the TLS 1.2 or above which guarantees confidentiality and integrity. Machine learning inference proxy calls are implemented against HTTPS and on-premise connection to the cloud by secure VPN tunnels or specific, dedicated private connections. Service to service communication is also enhanced by mutual certificate authentication.

To ensure the integrity of the data transferred between the end point and end replicated or synchronized, then there is hash-based validation. Checking the integrity is based on the rule:

$$\text{Integrity check} = (Hash_{source} == Hash_{target})$$

Whenever there is a mismatch in source and target hash, this automatically implements governance alerts and data corruption is avoided silently, propagating data in a reliable way across the environments.

8.2. Data Masking

Customer identifiers, financial account numbers, and employee information are all sensitive enterprise information that must be controlled to be viewed in both analytical and governance processes. The framework enforces the use of data masking (both data and file masking) to implement the analytical usability and protection of privacy. The use of static masking is only mainly used in non-production facilities like the development and testing systems. This is done by format-preserving substitution of sensitive fields by synthetic equivalents or tokenized equivalents. It is to allow development teams to operate with structurally valid data sets without revealing the real personal or financial information.

Dynamic masking is a runtime solution that is usable in production environments and adjusts data exposure depending on the level of user role and sensitivity of the data. The masking functionality may be written like:

$$\text{Masked\_Value} = \text{f(User role, Data sensitivity)}$$

In this paradigm, full records are available to data stewards who are authorized to access them, partially masked fields are available to business analysts, and anonymized datasets are made available to external auditors. This finer-tuning control avoids the threat of privacy loss and maintains an analytical worth. There are added security measures when transferring datasets to machine learning packages on the cloud. In personal identifiable information, anonymization is performed prior to that, feature engineering operations do not work with direct identifiers, and optional differential privacy methods can be added. These controls allow preventing additional exposure risks to AI training and inference.

8.3. GDPR and Regulatory Alignment

Businesses that exist internationally are required to make sure that the governance structures meet the standards of regulations like GDPR, SOX, and HIPAA among others depending on the location as well as financial regulations specific to the regions. The principles of regulations are directly integrated into the policies of governance of the proposed framework. The system ensures that it minimizes data through processing on the attributes that are needed to achieve objectives of validation and governance. Purpose limit is implemented in pipelines of data in such a way that data sets are meant only to accomplish quality checks, and compliance functions. The right to erasure can be achieved with the help of automated deletion processes, which are spread through SAP databases and the cloud storage layers. Moreover, lineage tracking based on metadata allows quickly locating the personal data and can be used to comply with requests to access and disclose data.

Retention management is governed by automated policy rules expressed as:

$$\text{Retention action} = \text{f}(\text{Data type},\ \text{Regulatory requirement})$$

According to classification and regulatory requirements, the datasets are archived, anonymized or deleted by default, or after a specified retention duration. Data residency policies have the facility of enforcing regional storage constraints and using access controls and encryption, as specified by jurisdiction, where the enterprise is applied across multiple regions. The measures minimize the cross-border compliance risk and make sure that the legal requirements in the country are followed.

8.4. Audit Logging and Continuous Monitoring

Audit logging gives responsibility, traceability and forensic preparedness in the system of AI-enabled governance. Every system activity of importance is logged, such as the events of data ingested, validation events, policy enforcement events, user access attempts, and configuration events. To ensure evidentiary integrity, logs are put in secure and tamper resistant repositories. The auditability of AI models is considered one of the fundamental compliance requirements. The history of the model versions is kept, the lineage of the training data is kept and scores of prediction confidence are kept with each inference event. Importance measures of features are retained to facilitate explainability, as well as regulatory scrutiny. This makes automated decisions that relate to enterprise reporting or compliance traceable and defensible.

Continuous security monitoring evaluates real-time telemetry signals using a composite risk function:

$$\text{Security\_Risk} = \text{g}(\text{Unauthorized\_Access},\ \text{Policy\_Violation},\ \text{DQcritical})$$

When the risk levels are managed beyond limits, automatic notifications are sent, access can be blocked in the meantime, and incident responsive processes can be started. To ensure that top executives monitor their encryption status, masking effectiveness, access violations, and regulatory compliance, a compliance dashboard brings this data into one location, allowing each to be stored in regulative compliance.

8.5. Security/Compliance Impact

The architecture integrates encryption, masking, automation of regulations, and auditability, providing the enterprise-level protection of SAP BW/4HANA and the cloud-integrated AI services. The security-first design is to be such that advanced analytics and intelligent governance is not compromised at the expense of confidentiality or compliance. This combined model helps scale up multi-cloud deployments and ensures that privacy regulations and corporate security policies are rigorously followed, which is why the framework can be utilized in an environment with high regulatory standards and when it comes to large enterprises.

9. Limitations and Future Work

Despite the fact that the suggested AI-based data quality and governance system leads to a significant increase in the accuracy of detection, automation, and scalability, there are a few constraints in and they are inherent to intelligent and distributed enterprise frameworks. One of the issues is model drift, in which the changes in data distributions and changes in the relationships among features reduce performance of anomaly detection as time progresses. In SAP choose dynamic and cloud-based architecture, constant introduction of heterogeneous data expands information vulnerability to data, concept, and prediction drift. In the absence of systematic surveillance, this might decrease scoring reliability and extinguish governance trust. Further work on this ought to include automated drift measures, ongoing validation pipelines, incremental retraining initiatives, and high-quality MLOps integration to secure inter-temporal sustainability and responsive functionality of the model.

The other decisive weakness is the complexity of orchestration between cross-clouds and poor real-time responsiveness. The use of hybrid architectures across the SAP systems and the cloud ecosystems presents an operational overhead with the identification federation, data synchronization latency, and distributed monitoring. Moreover, the existing batch-based validation framework does not allow isolating defects in real-time when running in high-velocity setting (financial transaction or IoT streams, etc). A future with the intention of combining the hybrid control planes, standardized interoperability schemes, and real-time event-driven streaming architectures will be needed to minimize integration friction and facilitate the execution of real-time governance in the distributed enterprise environments.

Lastly, the problem of explainability, scalability and ethical governance should be overcome, to make sure that the adoption can be sustainable. Black-box AI models have the potential to reduce visibility in regulated sectors, and a centralized scoring engine might fail with an exponential increase in data. More so, bias in anomaly detection can occur unconsciously, either in compliance reporting or governance prioritization. Future studies need to tackle the tasks of incorporating explainable AI schemes, distributed and federated validation models, cost-aware inference schedules as well as equity-conscious scoring schemes. When these technical and ethical issues are managed, the system will have the opportunity to become a robust, transparent, and altogether autonomous enterprise governance system that could assist in the next-generation digital ecosystems.

10. Conclusion

This study introduced an end-to-end AI-based Data Quality and Governance Framework which is aimed at improving the level of enterprise analytics reliability in hybrid SAP and cloud environments. The framework can close the loop of data quality through detecting, scoring, remedication, and constant monitoring by using supervised and unsupervised anomaly detection, predictive scoring of data quality, automatic root cause analysis, metadata-driven data governance, RBAC-based policy enforcement, and monitoring, all on a cloud computing platform. The architecture, which is implemented as part of both the SAP BW/4HANA warehousing and SAP BusinessObjects reporting reconciliation, and coordinated by scalable cloud-based services, transforms the enterprise governance approach to proactive and automated intelligence, rather than the reactive, rule-based validation approach.

The quantitative experiments showed that the experimental evaluation had evidently increased the results compared to the conventional validation methods. The AI-based system made significant improvements in terms of precision, recall, and the F1-score, lessening the false positives and decreasing the general defect propagation by up to 30-50 per cent. The reduction of the manual validation effort was about 40-60 percent, and the SLA compliance and responsiveness in operations increased to a greater extent. The scale testing assured that performance was stable with increasing workloads, and orchestration on the cloud allowed elastic scale of compute, use of the infrastructure efficiently and with cost-awareness models were performed. These results confirm that AI-enhanced governance provides higher accuracy, efficiency, and scalability when compared to their non-evolving systems that are based on the use of rules.

As an enterprise, the framework reflects a pivot of the strategic partnering of smart compliant and self-observing data environments. With automation, auditability and predictive analytics embedded as part of governance processes, companies can enhance the reliability of decision making, light operations overheads, increase regulatory conformity and lower infrastructure expenses. Suggestions on future research directions are adaptive model lifecycle management, explainable AI integration, real-time streaming governance, cross-cloud orchestration standardization, fairness-sensitive quality scoring, and autonomous remediation capabilities. Through all of this, AI-driven governance is not just created as a simple upgrade to an existing solution, but it becomes the core part of the next-generation digital enterprise architectures.

Reference

[1] Bayram, F., Ahmed, B. S., Hallin, E., & Engman, A. (2023, June). DQSOps: Data quality scoring operations framework for data-driven applications. In Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering (pp. 32-41).

[2] Guillen-Aguinaga, M., Aguinaga-Ontoso, E., Guillen-Aguinaga, L., Guillen-Grima, F., & Aguinaga-Ontoso, I. (2025). Data quality in the age of AI: A review of governance, ethics, and the FAIR principles. Data, 10(12), 201.

[3] Tewari, S. (2025). AI powered data governance-ensuring data quality and compliance in the era of Big Data. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 8(1), 187-197.

[4] Das, S. S. (2025). Intelligent Data Quality Framework Powered by AI for Reliable, Informed Business Decisions. Journal of Informatics Education and Research, 5(2), 4748-4754.

[5] Davenport, J. M. (2025). Ai-Augmented Frameworks For Data Quality Validation: Integrating Rule-Based Engines, Semantic Deduplication, And Governance Tools For Robust Large-Scale Data Pipelines. International Journal of Advanced Artificial Intelligence Research, 2(08), 9-15.

[6] Kavala, Y. (2023). Smart ERP: Scalable Data Engineering Frameworks Using Artificial Intelligence. International Journal of Computational Mathematical Ideas (IJCMI), 15(1), 1248-1262.

[7] Transforming Enterprise Data Strategy with SAP Business Data Cloud, SAP. online. https://architecture.learning.sap.com/docs/ref-arch/f5b6b597a6

[8] Sanka, V. (2025). Integrating Artificial Intelligence into Enterprise Data Governance Frameworks: A Comprehensive Approach for Automated Compliance and Risk Management. International Journal of Scientific Research in Science, Engineering and Technology, 12(1), 337-345.

[9] Bangad, N., Jayaram, V., Krishnappa, M. S., Banarse, A. R., Bidkar, D. M., Nagpal, A., & Parlapalli, V. (2024). A Theoretical Framework for AI-driven data quality monitoring in high-volume data environments. arXiv preprint arXiv:2410.08576.

[10] Jonnalagadda, R. R., Reddy, K. K., Gunupati, K., Kumar, M., Reddy, P. R. R., & Julakanti, R. (2025, September). Development of an SAP-Centric AI Architecture for Predictive Analytics and Business Intelligence Using Advanced Analytics and AI-Powered Algorithms. In 2025 International Conference on Computing and Communications (COMPUTINGCON) (pp. 1-6). IEEE.

[11] Wilkins, M. (2019). Learning Amazon Web Services (AWS): A hands-on guide to the fundamentals of AWS Cloud. Addison-Wesley Professional.

[12] Lawal, S. (2024). Integrating Artificial Intelligence into SAP Cloud for Intelligent Business Process Automation. Available at SSRN 5386162.

[13] Oviedo, J., Rodriguez, M., Trenta, A., Cannas, D., Natale, D., & Piattini, M. (2024). ISO/IEC quality standards for AI engineering. Computer Science Review, 54, 100681.

[14] Gualo, F., Caballero, I., Rodríguez, M., & Piattini, M. (2023). A data quality model for master data repositories. Informatica, 34(4), 795-824.

[15] Serrano, J. Y., & Zorrilla, M. (2021). A data governance framework for industry 4.0. IEEE Latin America Transactions, 19(12), 2130-2138.

[16] Marcucci, S., Alarcon, N. G., Verhulst, S. G., & Wullhorst, E. (2023). Mapping and Comparing Data Governance Frameworks: A benchmarking exercise to inform global data governance deliberations. arXiv preprint arXiv:2302.13731.

[17] Bögelsack, A., Chakraborty, U., Kumar, D., Rank, J., Tischbierek, J., & Wolz, E. (2022). SAP S/4HANA Systems in Hyperscaler Clouds. Springer. https://link. springer. com/content/pdf/10.1007/978-1-4842-8158-1. pdf.

[18] CHETAN, S., & ADARSH, V. (2022). Converging SAP, AI, and data analytic for transformative business management. WORLD, 14(3), 736-761.

[19] Prasad, M. S. (2025). A Comparative Study of Snowflake and SAP BW for Data Analytics. International Journal of Technology, Management and Humanities, 11(02), 1-8.

[20] Merseedi, K. J., & Zeebaree, S. R. (2024). The cloud architectures for distributed multi-cloud computing: a review of hybrid and federated cloud environment. The Indonesian Journal of Computer Science, 13(2).

[21] Shon, T., & Moon, J. (2007). A hybrid machine learning approach to network anomaly detection. Information Sciences, 177(18), 3799-3821.

[22] Zafar, A. (2024). Balancing the scale: navigating ethical and practical challenges of artificial intelligence (AI) integration in legal practices. Discover Artificial Intelligence, 4(1), 27.

[23] Sebastian-Coleman, L. (2012). Measuring data quality for ongoing improvement: a data quality assessment framework. Newnes.

[24] Cichy, C., & Rass, S. (2019). An overview of data quality frameworks. IEEE Access, 7, 24634-24648.

[25] Foorthuis, R. (2021). On the nature and types of anomalies: a review of deviations in data. International journal of data science and analytics, 12(4), 297-331.