Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Architecting a GxP-Compliant Data Lakehouse for Pharma: Leveraging Azure Databricks for Regulated Analytics

Vinod Balasaheb Parhad
Independent Researcher, USA

ARTICLE INFO

Received: 12 July 2025

Revised: 24 Aug 2025

Accepted: 04 Sept 2025

ABSTRACT

This article presents a comprehensive framework for implementing GxP-compliant data lakehouse architectures using Azure Databricks within pharmaceutical environments. The article addresses the fundamental tension between regulatory requirements and analytical innovation, proposing an architectural approach that satisfies stringent compliance standards while enabling advanced analytics capabilities. The article establishes design patterns for maintaining data integrity, traceability, and auditability across the pharmaceutical data lifecycle. The article demonstrates integration approaches for critical systems, including SAP S/4HANA, Laboratory Information Management Systems, and Manufacturing Execution Systems, with particular attention to data lineage implementation and validation strategies. The article reveals practical insights from implementation within a mid-size pharmaceutical organization, highlighting both challenges and successful outcomes. Discussion of best practices encompasses regulatory compliance patterns, performance optimization considerations, cost management strategies, and governance recommendations. While acknowledging current limitations in validation efficiency and specialized expertise requirements, the article establishes a foundation for GxP-compliant cloud data platforms that can transform pharmaceutical operations while maintaining the highest standards of data integrity and regulatory compliance.

Keywords: Pharmaceutical Data Lakehouse GxP Compliance Azure Databricks Validation Regulatory Data Lineage Delta Lake Audit Trails

Introduction

The pharmaceutical industry is experiencing an unprecedented digital transformation, accelerated by the convergence of advanced analytics, artificial intelligence, and cloud computing capabilities. Organizations increasingly recognize the strategic value of their data assets, yet face unique challenges in managing regulated environments where data integrity, traceability, and compliance are non-negotiable requirements. Traditional data architectures—characterized by siloed systems, fragmented data governance, and limited scalability—have proven inadequate for meeting modern analytical demands while maintaining regulatory compliance [1].

The emergence of the lakehouse paradigm represents a significant architectural evolution, combining the flexibility and scalability of data lakes with the reliability and performance of traditional data warehouses. This hybrid approach offers pharmaceutical organizations a compelling solution for consolidating diverse data types—from structured clinical trial results to unstructured research documents and real-time manufacturing data—within a unified, governed framework. However, implementing such architectures in Good Practice (GxP) regulated environments introduces complex considerations around validation, audit trails, and data lineage that require careful examination.

This research explores the design and implementation of a GxP-compliant data lakehouse using Azure Databricks, addressing the specific requirements of pharmaceutical organizations managing regulated processes. The proposed architecture integrates critical enterprise systems, including SAP S/4HANA and Laboratory Information Management Systems (LIMS), while supporting real-time operational data streams from Manufacturing Execution Systems (MES). The paper provides a comprehensive framework for maintaining regulatory compliance while enabling advanced analytics capabilities that can accelerate drug development, optimize manufacturing processes, and enhance patient outcomes.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

By examining validation strategies, lineage mechanisms, and governance approaches, this research aims to establish practical guidelines for pharmaceutical organizations navigating the complex intersection of innovative data technologies and stringent regulatory requirements.

II. Regulatory Framework for Pharmaceutical Data

Overview of GxP requirements for data management

Pharmaceutical data management operates within a complex regulatory framework where Good Practice (GxP) principles establish fundamental requirements for maintaining data integrity throughout the product lifecycle. These principles encompass Good Laboratory Practice (GLP), Good Clinical Practice (GCP), and Good Manufacturing Practice (GMP), collectively defining standards for data collection, processing, storage, and retrieval [2]. Central to GxP compliance is the ALCOA+ framework (Attributable, Legible, Contemporaneous, Original, Accurate, plus Complete, Consistent, Enduring, and Available), which provides practical criteria for ensuring data quality in regulated environments.

FDA 21 CFR Part 11 compliance considerations

The FDA's 21 CFR Part 11 regulation establishes requirements for electronic records and electronic signatures, mandating controls that ensure trustworthiness and reliability equivalent to paper records. Key provisions include system validation, audit trails, record retention, and electronic signature requirements. For cloud-based platforms like Azure Databricks, this necessitates robust validation protocols, documented system controls, and mechanisms to maintain the integrity of audit trails throughout data transformations.

GDPR and other relevant regulatory frameworks

Beyond pharmaceutical-specific regulations, data platforms must comply with broader data protection frameworks such as the General Data Protection Regulation (GDPR) in Europe, which imposes strict requirements on personal data processing. Additional considerations include regional regulations like China's NMPA guidelines and Japan's PMDA requirements, creating a complex global compliance landscape that pharmaceutical data architectures must navigate.

Key compliance challenges in modern data environments

Modern data environments present unique compliance challenges, including maintaining data lineage across diverse processing stages, implementing appropriate controls for cloud-based systems, and ensuring consistent governance across hybrid infrastructures. The integration of real-time data streams and advanced analytics further complicates compliance efforts, requiring thoughtful architectural approaches to balance innovation with regulatory requirements.

III. Azure Databricks Lakehouse Architecture

Foundational components of the lakehouse paradigm

The lakehouse paradigm combines data lake storage capabilities with data warehouse functionality, enabling organizations to process structured, semi-structured, and unstructured data within a unified architecture. This model eliminates traditional data silos while providing transactional integrity, schema enforcement, and governance capabilities previously associated only with data warehouses.

Azure Databricks capabilities for regulated environments

Azure Databricks provides several capabilities specifically relevant for regulated pharmaceutical environments, including workspace isolation, private link implementation for secure connectivity, and integration with Azure security services. The platform's notebook-based development environment supports validation through version control, reproducible execution, and comprehensive documentation, addressing key GxP requirements for traceability and accountability.

Delta Lake for ACID transactions and data quality

Delta Lake, a core component of the Databricks lakehouse architecture, provides ACID (Atomicity, Consistency, Isolation, Durability) transaction support on cloud object storage, enabling reliable data

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

operations at scale. Its time travel capabilities maintain historical versions, support audit requirements, and facilitate investigations of data anomalies. Schema enforcement and constraints help maintain data quality through automated validation at ingestion time.

Unity Catalog and governance capabilities

Databricks Unity Catalog delivers centralized governance across workspaces, providing fine-grained access controls and comprehensive audit logging. This capability addresses critical regulatory requirements for data access management and usage tracking. The centralized metadata repository simplifies lineage tracking and provides visibility into data transformations, supporting both compliance documentation and impact analysis for change management [3].

IV. Proposed GxP-Compliant Lakehouse Reference Architecture

Architectural layers and components

The proposed GxP-compliant lakehouse architecture consists of four distinct layers: ingestion, storage, processing, and consumption. The ingestion layer incorporates validated connectors and landing zones with automated data quality checks. The storage layer implements a multi-tier approach with bronze (raw), silver (validated), and gold (analytics-ready) zones, each with appropriate controls for data integrity and access management. The processing layer leverages Databricks compute clusters with separate development, validation, and production environments. Finally, the consumption layer provides validated reporting tools, analytical applications, and secure APIs for downstream systems [4].

Data ingestion patterns for regulated sources

Regulated data sources require specialized ingestion patterns that preserve data integrity and provide comprehensive audit capabilities. For each source system, the architecture implements source-specific adapters that capture metadata alongside content, establishing clear provenance. Change data capture (CDC) mechanisms detect and propagate modifications while maintaining historical records. Each ingestion pipeline incorporates automated validation rules that verify data completeness, format compliance, and logical consistency before acceptance into the storage layer.

Integration strategy for SAP S/4HANA and LIMS systems

Integration with SAP S/4HANA leverages Azure Data Factory with SAP-certified connectors to extract master data, transactional records, and quality parameters. The architecture implements both batch extraction for historical data and near-real-time integration through OData services for critical operational data. For Laboratory Information Management Systems (LIMS), the approach combines scheduled exports of structured results with event-driven triggers for time-sensitive data, such as stability testing and release testing results. Both integration paths maintain comprehensive technical context metadata to support regulatory compliance.

Real-time streaming architecture for operational systems (MES)

Manufacturing Execution Systems (MES) generate continuous data streams that require real-time processing for timely insights and process control. The architecture implements Azure Event Hubs as the entry point for these streams, with Databricks Structured Streaming providing stateful processing capabilities. Stream processing jobs validate incoming data against predefined quality thresholds while maintaining complete auditability. Delta Lake's transaction log ensures that streaming writes maintain ACID properties even during processing node failures, addressing a critical compliance requirement for manufacturing data integrity [5].

Security and access control implementation

Security controls span multiple layers, beginning with network isolation through Azure Private Link and service endpoints. Azure Active Directory integration provides identity management with multi-factor authentication, while role-based access control (RBAC) restricts data access based on job function and regulatory requirements. Column-level security and dynamic data masking protect sensitive information, while comprehensive audit logging captures all access attempts. The

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

architecture implements encryption both in transit and at rest, with key management through Azure Key Vault to facilitate regulatory compliance.

V. Data Lineage and Traceability Implementation

End-to-end lineage across the data lifecycle

The architecture implements comprehensive data lineage tracking from source systems through transformation pipelines to analytical outputs. Each dataset carries metadata identifying its origin, processing history, and quality assessment results. Lineage information captures both technical dependencies (which datasets and transformations contributed to a result) and business context (which processes and decisions were supported). This approach enables backward traceability to verify data provenance and forward impact analysis to assess the consequences of source changes or corrections.

Leveraging Unity Catalog for automated audit trails

Unity Catalog serves as the central governance mechanism, providing automated audit trails that document all data access and manipulation events. The implementation captures user identity, timestamp, operation type, and affected objects for each interaction with the data platform. These audit records are immutable and preserved according to record retention policies, supporting both routine compliance verification and targeted investigations. Programmatic access to audit logs enables automated compliance reporting and anomaly detection to identify potential security or quality concerns.

Delta logs for transparent data history

Delta Lake's transaction log architecture provides a foundation for transparent data history throughout the data lifecycle. The implementation preserves log entries indefinitely for regulated datasets, enabling point-in-time recovery and forensic analysis as required by GxP standards. Each transaction includes metadata capturing the business context, the associated process step, and the responsible user. Time travel capabilities allow authorized users to reconstruct the precise state of data at any historical point, supporting both audit activities and scientific investigations of analytical results [6].

Metadata management strategies

The architecture implements a layered metadata management strategy spanning technical, operational, and business domains. Technical metadata describes data structures and transformations, while operational metadata captures processing statistics and quality metrics. Business metadata provides context through controlled vocabularies and taxonomies aligned with industry standards. The implementation federates metadata from multiple sources, including automated extraction from source systems, transformation pipeline annotations, and manual curation of business context. This comprehensive approach ensures that data remains interpretable and traceable throughout its lifecycle.

VI. Validation Approach for Databricks Environments

Computer system validation methodology

Computer System Validation (CSV) for Databricks environments adapts traditional validation methodologies to cloud-native architectures while maintaining GxP compliance. The approach implements a streamlined V-model that defines distinct validation phases: user requirements specification, functional specification, configuration specification, installation qualification, operational qualification, and performance qualification. Each phase produces documented deliverables that establish an audit trail from requirements through testing. The validation package incorporates platform-specific considerations, including workspace configuration, cluster settings, and integration points with other Azure services, ensuring complete coverage of the regulated environment [7].

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Component	Traditional Approach	Azure Databricks Lakehouse Approach	Regulatory Benefit
Data Storage	Siloed systems with fragmented governance	Multi-tier architecture (Bronze/Silver/Gold zones) with unified storage	Enhanced traceability across the data lifecycle
Audit Capability	Manual audit logs with limited retention	Automated audit trails through Unity Catalog with immutable records	Comprehensive documentation of all data access and changes
Validation Process	Document-heavy V-model with extensive manual testing	Risk-based qualification with automated test patterns	Reduced validation overhead while maintaining compliance
Data Lineage	Limited or manual traceability between systems	End-to-end automated lineage with technical and business context	Complete provenance tracking for regulatory inquiries
Security Model	System-level controls with limited granularity	Multi-layered security with RBAC and column-level protections	Fine-grained access management aligned with regulatory roles

Table 1: Comparison of Key Components in GxP-Compliant Lakehouse Architecture [7, 10]

Risk-based qualification framework

The risk-based qualification framework categorizes system components according to their impact on product quality, patient safety, and data integrity. Critical components undergo comprehensive validation, while lower-risk elements receive targeted testing proportionate to their potential impact. Risk assessment considers multiple factors, including data criticality, processing complexity, and regulatory exposure. This framework applies GAMP 5 principles to cloud infrastructure, implementing streamlined validation for configurable components while reserving detailed scrutiny for custom code and integrations with GxP systems. The approach significantly reduces validation overhead while maintaining regulatory compliance.

Notebook validation workflow and documentation

Notebook validation follows a structured workflow that treats each notebook as a discrete computational unit requiring verification. The process begins with requirements mapping, identifying GxP-relevant functions within each notebook. Static code analysis tools verify compliance with coding standards and identify potential vulnerabilities. Execution validation then confirms deterministic behavior under controlled conditions, with results captured in validation records. Version control integration ensures that validated notebooks remain immutable in production, with formal change control procedures governing modifications. The documentation approach embeds validation evidence within notebook metadata, creating self-documenting artifacts that simplify regulatory inspections.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Test automation patterns for continuous compliance

Test automation implements continuous compliance verification through multiple patterns: regression testing confirms that validated functionality remains intact after platform updates; data quality testing validates transformation logic using predefined test datasets; performance testing ensures that processing times meet operational requirements; and security testing verifies access control mechanisms. Automated test suites execute on schedule and trigger alerts when deviations occur. This infrastructure-as-code approach to testing produces consistent, reproducible evidence of system fitness while reducing manual effort. The implementation captures test results in immutable records that satisfy regulatory documentation requirements [8].

VII. Case Study: Implementation in a Mid-Size Pharmaceutical Organization

Business requirements and legacy environment

A mid-size pharmaceutical organization specializing in oncology therapeutics faced significant challenges with its fragmented data landscape. The legacy environment consisted of disconnected systems: a validated on-premises data warehouse for regulatory reporting, departmental data silos with inconsistent governance, and manual processes for cross-functional analytics. Business requirements centered on accelerating research insights, improving manufacturing efficiency, and enhancing regulatory reporting capabilities while maintaining strict compliance with FDA and EMA requirements. Key drivers included reducing time-to-insight for clinical data, enabling real-time manufacturing analytics, and streamlining regulatory submissions through improved data integration.

Migration approach and implementation phases

The implementation followed a phased approach to minimize business disruption while establishing a validated foundation. Phase one established the core infrastructure with initial validation of the Azure Databricks environment and implementation of governance frameworks. Phase two migrated critical GxP data from the legacy warehouse, focusing on manufacturing and quality data with established regulatory importance. Phase three integrates real-time streams from manufacturing systems and laboratory equipment. The final phase implemented advanced analytics capabilities, including predictive maintenance for manufacturing equipment and pattern detection for quality deviations. Each phase included comprehensive validation activities with regulatory documentation.

Challenges encountered and mitigation strategies

Several significant challenges emerged during implementation. Integration with legacy validated systems required custom connector development with extensive validation testing to ensure data integrity. Initial performance issues with large-scale transformations necessitated architecture refinements, including optimized cluster configurations and improved partitioning strategies. Resistance from quality assurance teams accustomed to traditional validation approaches required education on risk-based validation and demonstration of improved compliance capabilities. The organization also encountered challenges with data ownership and governance across departmental boundaries, which were addressed through a cross-functional data governance committee with clear decision-making authority [9].

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

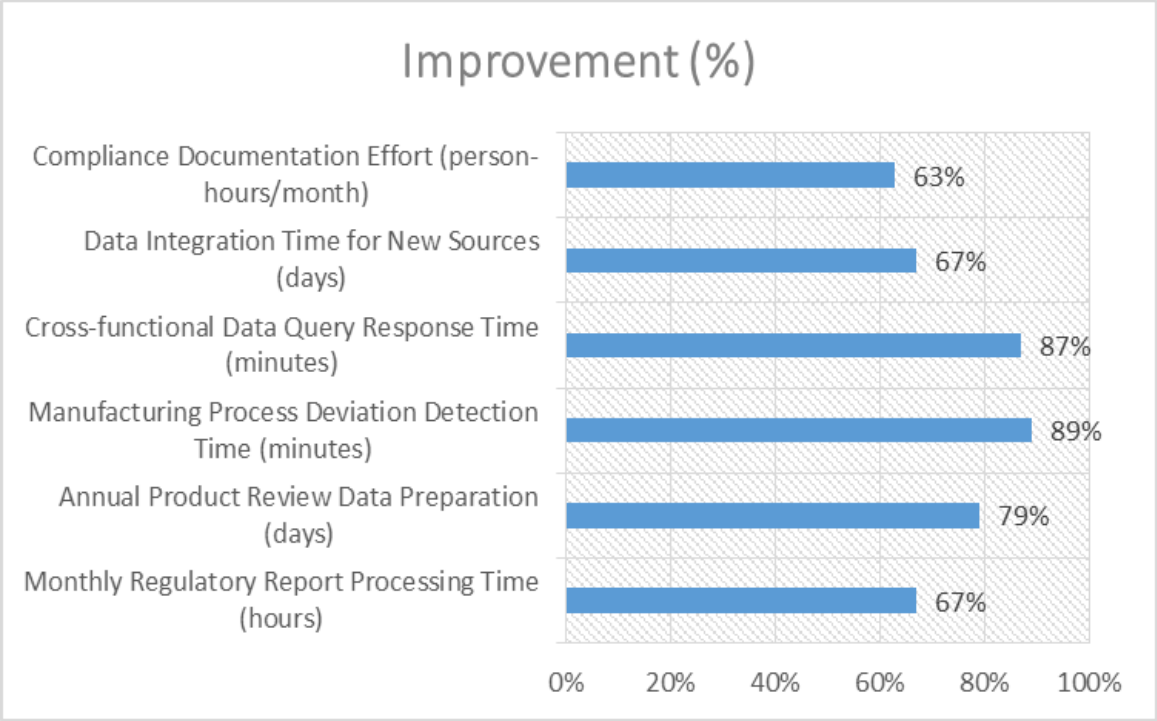

Fig 1: Performance Improvement Metrics After GxP-Compliant Lakehouse Implementation [9]

Performance and compliance outcomes

The implemented architecture delivered substantial improvements in both analytical capabilities and compliance posture. Processing time for monthly regulatory reports decreased by 67%, while data preparation for annual product reviews was reduced from weeks to days. Real-time manufacturing analytics enabled prompt intervention for process deviations, improving product quality and reducing waste. From a compliance perspective, the organization successfully passed FDA inspection with the new system, with inspectors specifically noting improved data traceability and audit capabilities. The centralized governance model reduced compliance overhead while improving data quality metrics across functional areas, demonstrating that advanced analytics and regulatory compliance can be simultaneously achieved through appropriate architecture and validation approaches.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Phase	Focus Area	Key Activities	Validation Considerations
1. Foundation	Core Infrastructure	Establish Azure Databricks environment, implement governance frameworks, and Configure security controls	Platform qualification, Documentation of configuration settings, Security validation
2. Data Migration	Critical GxP Data	Migrate manufacturing and quality data, implement SAP S/4HANA and LIMS connectors, and establish data quality rules	Data reconciliation testing, Connector validation, Audit trail verification
3. Real-time Integration	Operational Systems	Implement MES data streams, develop real-time processing pipelines, and Configure streaming checkpoints	Stream processing validation, Latency testing, Failure recovery verification
4. Advanced Analytics	Insights Generation	Implement predictive maintenance models, develop quality deviation detection, and build regulatory reporting automation	Algorithm validation, Report output verification, Performance qualification

Table 2: Implementation Phases for GxP-Compliant Lakehouse Migration [8, 9]

VIII. Discussion and Best Practices

Design patterns for regulatory compliance

Several design patterns have emerged as particularly effective for maintaining regulatory compliance in pharmaceutical data lakehouses. The immutable data pattern preserves original records while implementing compliant update mechanisms through append-only operations with explicit versioning. The segregation of duties pattern enforces separation between development, validation, and production environments with controlled promotion workflows. The documentation-as-code pattern embeds compliance evidence within the infrastructure definition, creating self-documenting systems that simplify audit preparation. Finally, the automated reconciliation pattern implements continuous verification of data consistency across system boundaries, detecting potential integrity issues before they impact regulated processes.

Performance optimization considerations

Performance optimization in GxP-compliant environments requires a careful balance between processing efficiency and validation integrity. Cluster sizing should implement auto-scaling capabilities while maintaining documented configurations for validated workloads. Data partitioning strategies should align with query patterns while preserving complete audit trails across partition boundaries. Query optimization techniques should leverage Delta Lake's data skipping and Z-ordering

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

capabilities to accelerate performance without compromising data integrity. For streaming workloads, checkpointing mechanisms must maintain exactly-once processing guarantees while allowing for recovery from infrastructure failures without data loss or duplication.

Cost management strategies

Cost management for regulated data platforms requires governance mechanisms that balance fiscal responsibility with compliance requirements. Resource tagging should identify GxP versus non-GxP workloads, enabling differentiated policies for retention and performance. Automated shutdown of development and validation environments during inactive periods can significantly reduce compute costs without impacting production systems. Storage tiering should implement lifecycle policies that transition historical data to lower-cost storage while maintaining accessibility for regulatory inquiries. Reserved capacity commitments can reduce costs for predictable workloads, particularly for production validation environments with consistent utilization patterns.

Change management and governance recommendations

Effective change management in GxP-compliant data environments requires structured processes that maintain validation status while enabling continuous improvement. Changes should be categorized by regulatory impact, with streamlined approval for low-risk modifications and comprehensive assessment for critical components. Impact analysis should leverage automated lineage capabilities to identify all potentially affected downstream systems and data products. Governance committees should include representation from quality, compliance, and technical teams to ensure balanced decision-making. Finally, ongoing monitoring should verify that implemented changes achieve their intended outcomes without introducing compliance risks [10].

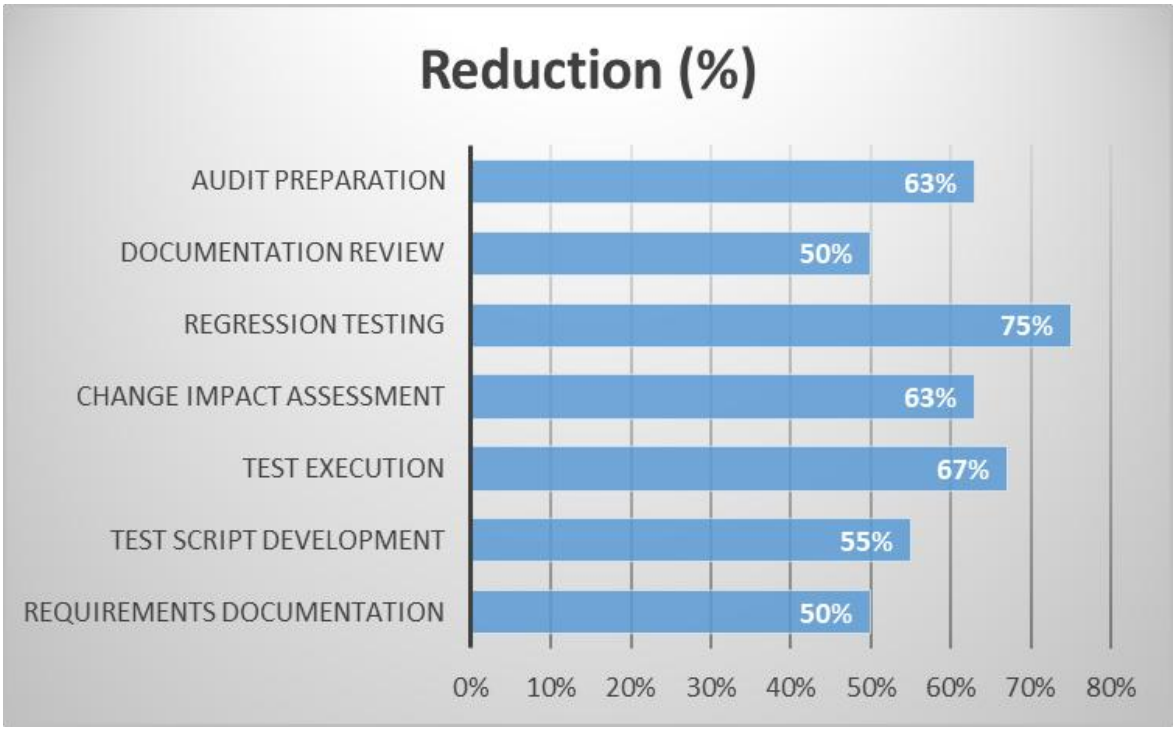

Fig 2: Resource Utilization Before and After Risk-Based Validation Implementation [7, 8]

IX. Future Research Directions

Summary of findings

This research has demonstrated that Azure Databricks lakehouses can successfully implement GxP-compliant data environments that support advanced analytics while maintaining regulatory compliance. Key findings include: (1) the lakehouse architecture effectively addresses the integration

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

challenges of diverse pharmaceutical data sources; (2) Delta Lake transaction logs provide comprehensive audit capabilities that satisfy regulatory requirements; (3) risk-based validation approaches can be successfully applied to cloud-native platforms; and (4) centralized governance through Unity Catalog significantly improves compliance posture while reducing administrative overhead. These findings suggest that modern data architectures can simultaneously enhance analytical capabilities and strengthen regulatory compliance when properly implemented.

Limitations of the current approach

Despite its strengths, the current approach has several limitations. First, validation methodologies remain relatively labor-intensive, requiring significant documentation effort despite automation improvements. Second, real-time processing of manufacturing data introduces latency challenges that may impact time-sensitive quality decisions. Third, integration with legacy validated systems often requires custom connectors that create potential compliance gaps during system updates. Fourth, the rapidly evolving cloud platform landscape creates validation challenges as new features are continuously introduced. Finally, the approach requires specialized expertise in both pharmaceutical compliance and modern data architectures, creating potential resource constraints for smaller organizations.

Areas for future investigation and enhancement

Future research should address several promising areas for enhancement. Automated validation frameworks that leverage AI for test generation and execution could significantly reduce validation overhead while improving coverage. Advanced lineage visualization techniques could enhance understanding of complex data relationships for both technical and non-technical stakeholders. Integration of regulatory intelligence systems could automate compliance checks against evolving regulatory requirements. Real-time compliance monitoring could detect potential issues before they impact product quality or patient safety. Finally, industry standardization of validation approaches for cloud-native data platforms would reduce implementation costs and accelerate adoption across the pharmaceutical sector.

Conclusion

The integration of GxP compliance principles with modern lakehouse architectures represents a significant advancement for pharmaceutical data management, enabling organizations to harness the analytical power of cloud platforms while maintaining regulatory integrity. This article has demonstrated that Azure Databricks, when properly configured with appropriate governance mechanisms and validation approaches, provides a viable foundation for regulated analytics across the pharmaceutical value chain. The article architecture addresses critical compliance requirements through comprehensive lineage tracking, immutable audit trails, and risk-based validation approaches, while enabling the performance and flexibility needed for contemporary analytics workloads. As pharmaceutical organizations continue their digital transformation journeys, the patterns and frameworks presented in this paper offer practical guidance for balancing innovation with compliance obligations. While challenges remain in areas such as automated validation, real-time processing, and specialized expertise requirements, the path toward GxP-compliant cloud data platforms is now clearer and more accessible. The evolution of these architectures will likely accelerate as regulatory frameworks adapt to technological change and as pharmaceutical organizations increasingly recognize data as a strategic asset requiring both protection and activation.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

References

[1] Judith Nwoke, "Regulatory Compliance and Risk Management in Pharmaceuticals and Healthcare". International Journal of Health Sciences, 7(6), 60–88. 2024-09-08. https://carijournals.org/journals/index.php/IJHS/article/view/2223
[2] KPMG, "GxP compliance in cloud infrastructure", Aug 2022. https://assets.kpmg.com/content/dam/kpmg/sg/pdf/2022/09/gxp-compliance-in-cloud-it-infrastructure.pdf
[3] Michael Amburst, et al., "Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics", CIDR '21, Jan. 2021, Online, https://www.cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf
[4] Nailya Uzhakova (née Sabirzyanova), Stefan Fischer, "Data-Driven Enterprise Architecture for Pharmaceutical R&D". Digital 2024, 4, 333-371, 22 April 2024. https://www.mdpi.com/2673-6470/4/2/17
[5] Chris Schwartz, "GxP Validation: Overview and Best Practice Guide", April 9, 2024. https://www.leapwork.com/blog/how-to-secure-gxp-compliance-in-pharma-with-test-automation
[6] Sudarshan Singh, "What Is Data Lineage and Why Is It Required in Today's Complex Data Environment?", September 19, 2024. https://www.acceldata.io/blog/what-is-data-lineage-and-why-is-it-required-in-todays-complex-data-environment
[7] Pravin Ullagaddi, "A Framework for Cloud Validation in Pharma". Journal of Computer and Communications, 12, 103-118. doi: 10.4236/jcc.2024.129006, September 2024, https://www.scirp.org/journal/paperinformation?paperid=136101
[8] USDM Life Sciences, "Continuous Compliance & Validation". https://usdm.com/capabilities/trust/continuous-compliance-validation
[9] KPMG, "Digitalization in life sciences: Integrating the patient pathway into the technology ecosystem", https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2018/01/digitalization-in-life-sciences.pdf
[10] Shivaram P R, "Data Governance Model: How Leading Companies Ensure Compliance and Security". acceldata, March 25, 2025, https://www.acceldata.io/blog/data-governance-model-how-leading-companies-ensure-compliance-and-security