Yurii Ovsiuchenko
Candidate of Economic Sciences,
Associate Professor at the Department of Economic Cybernetics
and Management of Economic Security
Kharkiv National University of Radio Electronics

Овсюченко Ю.В.
кандидат економічних наук, доцент,
доцент кафедри економічної кібернетики
та управління економічною безпекою
Харківського національного університету радіоелектроніки

VERIFICATION ORGANIZATION QUESTIONS OF BIG DATA STORES AT THE ENTERPRISE

ПИТАННЯ ОРГАНІЗАЦІЇ ВЕРИФІКАЦІЇ СХОВИЩ ВЕЛИКИХ ДАНИХ НА ПІДПРИЄМСТВІ

In the rapidly evolving landscape of modern enterprises, the effective management and verification of big data stores have become paramount. As organizations accumulate vast amounts of data from diverse sources, ensuring the integrity, security, and quality of this data is crucial for making informed business decisions and maintaining regulatory compliance. This introduction sets the stage to explore the organization of verification processes within enterprise big data environments.

Big data refers to datasets that are too large or complex for traditional data-processing applications to handle [1]. Enterprises harness big data for various purposes, including customer insights, operational analytics, and predictive modeling. However, the inherent challenges of big data, such as volume, velocity, variety, and veracity, necessitate robust verification mechanisms to maintain data accuracy and reliability.

Verification of big data stores involves the systematic validation and quality assurance of data throughout its lifecycle. This process encompasses data acquisition, storage, processing, and analysis stages [2]. It ensures that data is accurate, consistent, secure, and compliant with organizational standards and regulatory requirements.

Organizations must adopt structured approaches to organize and streamline the verification of their big data stores. This involves establishing clear methodologies, deploying appropriate technologies, and implementing governance frameworks to support data quality initiatives [3]. Furthermore, collaboration between data engineers, data scientists, business analysts, and compliance officers is essential to align verification efforts with business objectives and regulatory guidelines.

Let's explore key components of organizing verification within enterprise big data environments.

Data Governance Framework: Developing policies, procedures, and controls to ensure data quality, privacy, and security.

Data Quality Management: Implementing tools and processes to monitor, cleanse, and enhance data quality throughout its lifecycle.

Metadata Management: Maintaining comprehensive metadata to track data lineage, provenance, and usage.

Compliance and Security Measures: Adhering to regulatory standards (e.g., GDPR, HIPAA) and implementing robust security protocols to protect sensitive data.

Technology Infrastructure: Leveraging advanced technologies such as data validation tools, automated testing frameworks, and machine learning algorithms to improve verification efficiency and accuracy.

Cross-Functional Collaboration: Fostering collaboration between IT, data management teams, and business stakeholders to ensure alignment with organizational goals and priorities.

This researching provides a foundational understanding of the importance of organizing verification processes within enterprise big data environments. Subsequent discussions will delve deeper into specific strategies, challenges, and best practices for effectively verifying and managing big data stores to derive actionable insights and drive business value.

Analyzing typical existing problems in enterprises related to organizing verification of big data stores reveals several challenges that organizations commonly encounter in managing and ensuring the quality of their large-scale data repositories. These challenges stem from the complexities inherent in big data environments, including data volume, variety, velocity, and veracity. Let's explore some of the typical problems faced by enterprises in organizing verification of big data stores.

Data Quality Issues: one of the primary challenges is maintaining data quality across diverse and rapidly expanding data sets. Enterprises often struggle with inconsistencies, inaccuracies, duplicates, and missing values within their big data stores. Poor data quality undermines decision-making, impedes analytics, and erodes trust in enterprise data assets.

Data Integration Complexity: enterprises frequently encounter challenges in integrating data from disparate sources, including structured, semi-structured, and unstructured data. Data integration complexities arise due to differences in data formats, schemas, semantics, and governance policies. Managing data integration across hybrid cloud and on-premises environments further exacerbates these challenges.

Lack of Data Governance: inadequate data governance practices pose significant obstacles to organizing verification efforts. Enterprises often lack comprehensive data governance frameworks encompassing data quality management, metadata management, access controls, and compliance enforcement. Without robust governance, it's challenging to establish accountability and traceability in data verification processes.

Limited Metadata Management: iInsufficient metadata management contributes to ambiguity and inefficiency in verifying big data stores. Enterprises struggle to maintain accurate metadata catalogs that provide insights into data lineage, provenance, and usage. Inadequate metadata hampers data discovery, impacts data validation, and complicates compliance efforts.

Scalability and Performance Issues: verifying big data stores at scale presents scalability and performance challenges. Enterprises must ensure that verification processes can handle massive volumes of data efficiently without compromising performance. Inefficient data processing, resource contention, and bottlenecks in data pipelines can lead to delays in verification and hinder timely insights.

Data Security and Privacy Concerns: ensuring data security and privacy during verification is a critical challenge for enterprises. Big data stores often contain sensitive and personally identifiable information (PII) that must be protected against unauthorized access, breaches, and data leaks. Compliance with regulations such as GDPR, HIPAA, and CCPA adds complexity to data security practices.

Complexity of Analytical Workloads: verification of big data stores involves complex analytical workloads, including data validation, anomaly detection, and pattern recognition. Enterprises struggle with selecting and deploying appropriate analytical tools and algorithms that can effectively process and analyze large volumes of data in real-time or near-real-time.

Limited Collaboration Across Teams: siloed organizational structures hinder collaboration among teams responsible for data verification. Lack of alignment between data engineers, data scientists, business analysts, and compliance officers leads to disjointed verification processes and suboptimal outcomes. Effective cross-functional collaboration is essential for addressing verification challenges comprehensively.

Technology Fragmentation and Tool Proliferation: enterprises often face challenges related to technology fragmentation and tool proliferation in their big data environments. Multiple data platforms, databases, analytics tools, and validation frameworks may coexist within the enterprise, resulting in interoperability issues, licensing complexities, and increased operational overhead for verification.

Continuous Monitoring and Adaptation: lastly, enterprises struggle to establish robust mechanisms for continuous monitoring and adaptation of verification processes. Big data environments are dynamic and evolve rapidly, requiring proactive monitoring, feedback loops, and adaptive strategies to detect and respond to emerging verification issues in real-time.

In summary, typical existing problems in enterprises related to organizing verification of big data stores revolve around data quality, integration complexity, governance gaps, scalability challenges, security concerns, analytical workloads, collaboration barriers, technology fragmentation, and the need for continuous monitoring and adaptation. Addressing these challenges requires a holistic approach encompassing technology investments, organizational alignment, process improvements, and a data-driven culture focused on ensuring the reliability and utility of enterprise big data assets.

References:
1. Yashar Mansouri. Big Data Management: Data Repository Strategies and Data Warehouses. Available at: https://towardsdatascience.com/big-data-management-data-repository-strategies-and-data-warehouses-bfa9b45c95de
2. Leandro DalleMule, Thomas H. Davenport. What's Your Data Strategy? Available at: https://analyticsconsultores.com.mx/wp-content/uploads/2019/03/Whats-Your-Data-Strategy-L.-DalleMule-T.H.-Davenport-HBR-2017.pdf
3. Emmanuel Ohaba. The Impact Of Big Data Analytics On Business Decision-Making. Available at: https://elearningindustry.com/the-impact-of-big-data-analytics-on-business-decision-making