



MAKERERE UNIVERSITY BUSINESS SCHOOL

Business Intelligence and Data Warehousing

TOPIC 6

DATA QUALITY AND GOVERNANCE

BUC 2225



Topic Outline

2

- Evaluate the significance of data quality within Business Intelligence (BI) and data warehousing.
- Comprehend the practices of data governance and the roles of data stewardship.
- Apply data quality frameworks to assess and enhance data accuracy and consistency.



Introduction to Data Quality and Governance

3

- In today's data-driven world, organizations face critical concerns regarding the quality and governance of data.
- **Data Quality:** Refers to the accuracy, completeness, consistency, timeliness, and reliability of data, crucial for effective decision-making.
- Data quality is paramount for organizations relying on data analytics, business intelligence (BI), and data warehousing systems to derive actionable insights and support strategic decision-making.
- High-quality data ensures that organizations can trust their data-driven analyses, enabling them to identify trends, patterns, and opportunities accurately.
- Data quality issues can manifest in various forms, including inaccuracies, incompleteness, inconsistencies, and outdatedness, posing significant challenges to organizational effectiveness and competitiveness.
- Poor data quality can lead to erroneous conclusions, misinformed decisions, and operational inefficiencies.

Introduction to Data Quality and Governance



4

- **Data Governance:** Involves establishing policies, processes, roles, and responsibilities to ensure proper data management, protection, and utilization.
- Data governance provides the framework for managing and ensuring the quality, integrity, security, and compliance of organizational data throughout its lifecycle.
- It involves defining policies, standards, procedures, and accountability structures to govern data-related processes effectively.
- Data governance establishes clear roles and responsibilities for data management including data stewards who act as custodians responsible for overseeing specific data.
- It aligns data management practices with organizational goals and regulatory requirements, promoting data-driven decision-making while mitigating risks associated with data misuse, security breaches, and non-compliance.



Practical Applications of DQ & G

5

- Financial Institutions:
 - Data Quality: Accurate customer account information is vital in banking for transactions, loan approvals, and risk assessment. Errors like incorrect balances or outdated contact details can lead to financial statement inaccuracies and customer dissatisfaction.
 - Data Governance: Financial institutions enforce rigorous data governance practices to meet regulatory standards (e.g., GDPR, Basel III). They assign data stewardship roles for ensuring accuracy and security, conduct data quality checks in transaction systems, and utilize governance tools for monitoring and auditing.

- Healthcare Sector:
 - Data Quality: Patient records in healthcare must be complete, accurate, and current to support diagnosis, treatment, and research. Inaccurate or incomplete data can result in medical errors, compromised patient safety.
 - Data Governance: Healthcare organizations implement governance frameworks to uphold patient data integrity and confidentiality. They set data entry standards, enforce access controls for sensitive information, and conduct regular audits to comply with healthcare regulations like HIPAA.



Practical Applications of DQ & G

6

- E-commerce Platforms:
 - Data Quality: E-commerce firms depend on precise product details, pricing, and customer profiles to enhance shopping experiences and boost sales. Errors like inaccurate descriptions or pricing can harm customer satisfaction and revenue.
 - Data Governance: E-commerce platforms enforce governance measures to maintain consistent and reliable product data across various channels. They set quality metrics, employ validation methods during data handling, and utilize governance tools to uphold data integrity.

- Government Agencies:
 - Data Quality: Government bodies rely on precise demographic data, census figures, and socioeconomic indicators to guide policy decisions and resource allocation. Mistakes like incorrect population estimates can lead to misallocated funds and ineffective policies.
 - Data Governance: Government agencies implement governance practices to ensure public data integrity. They establish quality standards for data collection, employ validation checks, and promote transparency through open data initiatives to enhance accountability and accessibility.



Data Quality Challenges

7

- Data Entry Errors: Mistakes made during manual data entry processes, such as typos, incorrect values, or transposition errors, can lead to inaccuracies in the dataset.
- Incomplete Data: Missing information in datasets, whether due to oversight, non-mandatory fields not being filled, or partial data capture, can reduce the completeness and reliability of the data.
- Inconsistencies in Data Formats: Variation in the format of data entries, such as different date formats, units of measurement, or naming conventions, can hinder data integration and analysis.
- Outdated Data: Data that is not regularly updated can become stale and no longer reflect the current state of affairs, leading to inaccurate analyses and decisions.
- Duplicate Records: Duplicate entries of the same data can skew statistical analyses, inflate counts, and misrepresent the true size of datasets.



Data Quality Challenges

8

- Bias and Inaccuracies: Biases introduced during data collection, such as sampling bias or selection bias, can distort the representation of the population and lead to erroneous conclusions.
- Lack of Data Standardization: Absence of standardized data formats, definitions, or quality control processes can hinder interoperability and complicate data management and analysis efforts.
- Data Silos: Data silos, where information is isolated within specific departments or systems, can inhibit data sharing, collaboration, and integration efforts, leading to fragmented and incomplete datasets.
- Human Error: Human error in data handling processes, such as misinterpretation of data, improper data manipulation, or data misclassification, can introduce inaccuracies and undermine data quality.
- External Factors: External factors such as changes in regulations, market conditions, or technological advancements can impact data quality by introducing new data sources, requirements, or challenges that need to be addressed.

Impact of Poor Data Quality on BI & DW



9

- ❑ Inaccurate Insights: Erroneous data leads to flawed analyses, compromising decision-making in BI initiatives.
- ❑ Operational Inefficiency: Reliance on unreliable data disrupts business operations, causing delays and errors.
- ❑ Reduced Trust: Persistent data quality issues erode confidence in organizational data, hindering stakeholder trust.
- ❑ Financial Losses: Incorrect financial data leads to regulatory penalties and revenue leakage.
- ❑ Missed Opportunities: Flawed insights result in missed revenue opportunities and innovation.

Impact of Poor Data Quality on BI & DW



10

- Reputational Damage: Poor data tarnishes the organization's reputation and brand image.
- Legal and Compliance Risks: Non-compliance with regulations due to inaccurate data.
- Resource Wastage: Time and resources spent on data cleansing and validation.
- Customer Dissatisfaction: Incorrect information leads to poor customer experiences and churn.
- Decision-Making Paralysis: Lack of confidence in data-driven insights leads to decision-making delays.



Establishing Data Quality

11

- Data Cleansing: Implementing data cleansing processes to rectify inaccuracies, remove duplicates, and standardize data formats.
- Data Validation: Applying data validation checks to ensure data accuracy, completeness, and consistency during data entry and processing.
- Data Governance: Establishing robust data governance practices, including policies, standards, and procedures for data management, stewardship, and compliance.
- Continuous Monitoring: Implementing tools and processes for continuous data monitoring, auditing, and validation to maintain data quality over time.
- Training and Education: Providing training and education to employees on data quality best practices, tools, and techniques to ensure awareness and adherence to data quality standards.
- Benchmarking: Establishing benchmarks and targets for data quality metrics based on industry standards, organizational goals, and stakeholder expectations.



Data Quality Frameworks – Six Sigma

12

- ❑ Six Sigma is a data-driven framework aimed at reducing defects and variability in processes, including data processes.
- ❑ It emphasizes rigorous measurement, analysis, and improvement to achieve near-perfect performance and minimize errors.
- ❑ The focus is on identifying root causes of defects, optimizing processes, and ensuring consistency in data quality.
- ❑ Six Sigma principles are applied to analyze data processes, identify areas for improvement, and implement strategies to enhance data accuracy and consistency.



Data Quality Frameworks - TQM

13

- TQM (Total Quality Management) is centered around continuous improvement and customer satisfaction.
- It emphasizes the involvement of employees at all levels, process optimization, and the pursuit of excellence in all aspects of operations.
- TQM aims to enhance data quality by fostering a culture of quality, where every employee is responsible for maintaining and improving data integrity.
- TQM principles are applied to establish quality standards, promote data-driven decision-making, and continuously monitor and improve data processes to meet customer expectations.



Data Governance Practices

14

- Establishing Data Governance Frameworks: Develop structured frameworks, policies, and procedures to govern data management practices within the organization.
- Defining Data Ownership: Clearly designate ownership and accountability for different data assets, ensuring responsibility for data quality, security, and compliance.
- Data Stewardship: Appoint data stewards responsible for overseeing specific data domains, ensuring data quality, integrity, and accessibility.
- Data Classification: Classify data based on its sensitivity, value, and regulatory requirements to determine appropriate handling and protection measures.
- Data Standards and Policies: Define and enforce data standards, naming conventions, and policies to ensure consistency and compliance across the organization.
- Data Quality Management: Implement processes and controls for monitoring, measuring, and improving data quality throughout its lifecycle.
- Data Lifecycle Management: Establish procedures for managing data from creation to archival or deletion, ensuring compliance with retention policies and regulatory requirements.

Data Governance Practices



15

- Data Access Controls: Implement access controls and authentication mechanisms to ensure appropriate access to data based on user roles and permissions.
- Data Security: Implement security measures, such as encryption, access controls, and data masking, to protect sensitive data from unauthorized access or disclosure.
- Data Privacy Compliance: Ensure compliance with data privacy regulations, such as GDPR or HIPAA, by implementing policies and controls for data collection, storage, and processing.
- Data Governance Committees: Establish governance committees or councils to provide oversight, guidance, and decision-making authority for data governance initiatives.
- Data Training and Awareness: Provide training and awareness programs to educate employees on data governance policies, procedures, and best practices.
- Data Audit and Monitoring: Conduct regular audits and monitoring activities to assess compliance with data governance policies, identify issues, and track remediation efforts.
- Continuous Improvement: Foster a culture of continuous improvement by soliciting feedback, measuring performance against established metrics, and implementing corrective actions to enhance data governance practices over time.



Data Stewardship

16

- ❑ Data stewards are individuals tasked with managing specific data domains or sets within an organization. They are charged with the following roles and responsibilities.
- ❑ Custodianship of Data: Data stewards are responsible for the overall management and protection of specific data domains or datasets assigned to them. They ensure the accuracy, completeness, and reliability of the data under their stewardship, implementing measures to maintain data quality and integrity.
- ❑ Data Quality Management: Data stewards play a crucial role in ensuring data quality within their assigned domains. They define and enforce data quality standards, conduct data profiling and analysis to identify quality issues, and implement corrective actions to improve data accuracy and consistency.
- ❑ Data Access and Security: Data stewards are responsible for controlling access to data and maintaining its security. They establish access controls and permissions, ensuring that only authorized individuals have appropriate access to the data. They also implement security measures to protect data from unauthorized access, breaches, or misuse.



Data Stewardship

17

- **Data Documentation:** Data stewards maintain documentation related to their assigned data domains. They document data definitions, lineage, and usage guidelines. This documentation facilitates data discovery, understanding, and usage across the organization.
- **Data Governance Compliance:** Data stewards ensure compliance with data governance policies, regulations, and standards. They monitor data usage and handling practices to ensure adherence to governance policies and regulatory requirements.
- **Collaboration:** Data stewards collaborate with various stakeholders effectively regarding data-related matters. They work closely with business units, IT teams, data users, and other stakeholders to understand data requirements, address concerns, and align data management practices with organizational goals.
- **Communication** They also serve as subject matter experts on data-related issues and provide guidance and support to data users as needed. Effectively communicating data-related matters, policies, and procedures to relevant stakeholders across the organization.



Data Governance Frameworks

18

- ❑ **Data Management Body of Knowledge (DMBOK):** Developed by the Data Management Association (DAMA), DMBOK provides a comprehensive framework outlining the principles, practices, and processes for effective data management, including data governance.
- ❑ **Control Objectives for Information and Related Technologies (COBIT):** COBIT is a framework developed by ISACA for governance and management of enterprise IT. It includes principles and practices for data governance within the broader context of IT governance.
- ❑ **The Data Governance Institute (DGI) Framework:** DGI offers a framework comprising principles, processes, and best practices for establishing and managing data governance programs within organizations.
- ❑ **The Open Group Architecture Framework (TOGAF):** TOGAF is an enterprise architecture framework that includes guidelines and methods for designing, planning, implementing, and governing enterprise architectures, including data governance aspects.



Data Governance Frameworks

19

- ❑ **Data Governance Maturity Models:** Various maturity models, such as the Data Governance Maturity Model by Gartner or the Data Management Maturity (DMM) Model by CMMI Institute, provide frameworks for assessing the maturity of an organization's data governance practices and guiding its improvement efforts.
- ❑ **Data Governance Framework by Oracle:** Oracle offers a framework for data governance that covers various aspects, including data quality management, metadata management, data security, and compliance.
- ❑ **Data Governance Framework by Microsoft:** Microsoft provides guidance and tools for implementing data governance within its data platform, including Azure Data Governance capabilities and best practices for governing data in Microsoft environments.
- ❑ **ISO/IEC 38500:** This international standard provides principles and guidance for corporate governance of information technology, which includes aspects of data governance.

Applying Data Governance Frameworks



20

- Assessment and Planning: Evaluate current practices, define objectives, and assess readiness.
- Establish Governance Structures: Form committees and define roles for oversight and accountability.
- Develop Policies and Standards: Create comprehensive guidelines aligned with best practices and regulations.
- Technology Enablement: Implement tools for automation and support of governance processes.
- Data Quality Management: Implement processes for monitoring and improving data quality.
- Change Management and Communication: Communicate the importance of governance and facilitate adoption.
- Implementation and Execution: Roll out initiatives gradually, monitor progress, and address challenges.
- Continuous Improvement: Monitor, evaluate, and refine governance practices over time.

Significance of Data Quality and Governance



21

- Informed Decision Making: Reliable data ensures accurate decisions, improving strategic outcomes and operational efficiency.
- Regulatory Compliance: Data governance ensures adherence to regulations, safeguarding data privacy and integrity.
- Customer Satisfaction: High-quality data enables personalized experiences, enhancing customer loyalty.
- Operational Efficiency: Well-governed data processes reduce errors, driving cost savings and productivity.
- Risk Mitigation: Effective governance identifies and mitigates data-related risks, protecting organizational reputation.
- Innovation: Reliable data fuels analytics and innovation, unlocking new opportunities for growth.
- Trust and Credibility: Consistent data builds stakeholder trust, enhancing organizational credibility.
- Collaboration: Governance fosters cross-functional collaboration, aligning data practices with organizational goals.
- Continuous Improvement: Monitoring and refining governance practices ensures ongoing data quality enhancement and business optimization.



Assignment

Applying Data Governance Frameworks



23

You've been provided with the dataset below. Use it to answer the questions that follow:

Customer Orders Data for March, 2024					
Order ID	Customer Name	Email	Phone Number	Order Date	Items Ordered
1001	John Smith	johnsmith@gmail.com	555-123-4567	2023-12-01	Shirt, Hat
1002	Jane Doe	j.doe123	(888) 555-1212	2024-03-20	
1003	John Smith	johnsmith@gmail.com		2023-11-15	
1004	Michael Lee			03/22/2024	Shoes
1005		peter.look@g.c			Coffee Maker, Blender
1006	Sarah Jones	[email address removed]		2024-03-18	
1007	David Williams	david.williams@email.com	123-456	2024-03-23	Jacket
1008	Emily Brown	[email address removed]	(415) 333-7890	2024-03-15	
1009	Jane Doe	j.doe123	(888) 555-1212		
1010	Charles Miller	[email address removed]	555-555-0101		

Applying Data Governance Frameworks



24

□ **Tasks**

i) **Data Quality Assessment:**

- Analyze a provided dataset to identify potential data quality issues.

i) **Data Governance Implementation:**

- Propose and implement data governance practices for the dataset.
- Define roles and responsibilities for data stewards in addressing data quality concerns.

i) **Continuous Improvement Plan:**

- Develop a plan for continuous data quality improvement.
- Establish metrics and benchmarks to monitor the effectiveness of governance practices.



Thank You