Data Governance, Cataloging, and Lineage

by (Bangun Sasongko)

Trainer Profile

Bangun W. Sasongko

Data Engineer at BRI
Ex. Devops Engineer
Bachelor of Physics, Electrical and Instrumentations
@sasongkobgn
sasongkobgn@gmail.com

Table of Content

Content

About this Course

Dive into the world of views! This chapter will equip you with the skills to create and query them. You'll not only learn the basics but also explore advanced management techniques. Finally, we'll demystify the distinction between materialized and non-materialized views.

The Objectives

By the end of this course, you will be able to:

Data Fellowship
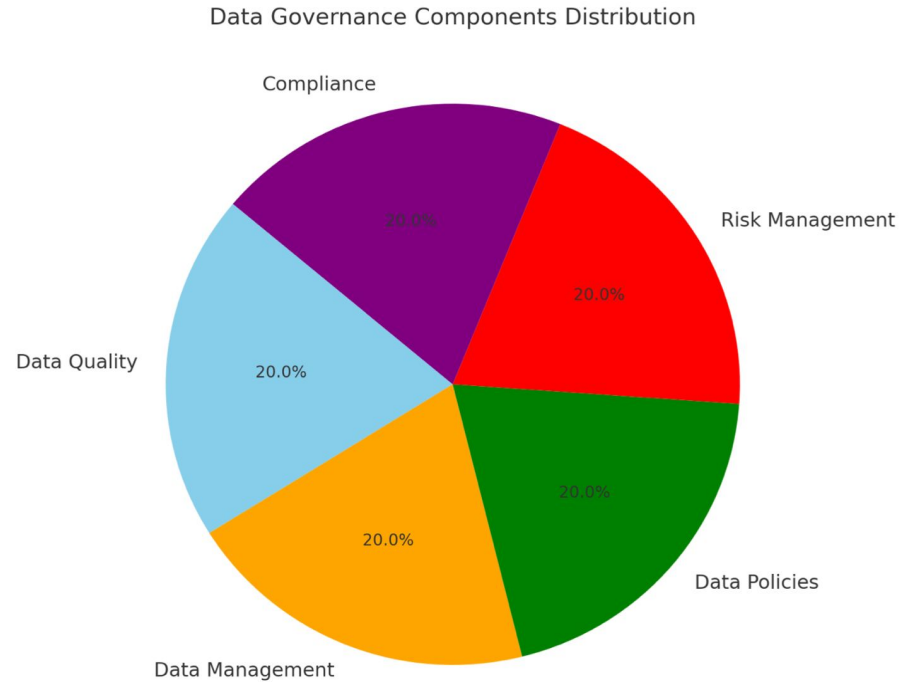
What is Data Governance?

What is Data Governance?

Definitions of Data Governance are numerous. However, they agree on a key principle - that it's dedicated to the organisation of people, processes, and technology to enable effective data management.

Data Governance includes:

- Understanding the **roles and responsibilities** of those who use data.
- The **technologies, tools and systems** to support capabilities in the areas of data management, quality, security, integration, discovery, accessibility, and availability.
- **Processes and controls** implemented to support the consistency, integrity, usability and privacy of data.
- **Policies, procedures and standards** to support regulatory compliance, ethics, and provide clarity over responsibilities.
- **Data quality** - including metadata and information security - to ensure reliability of data.

What is Data Governance?

Data Governance Components Distribution

key objectives of implementing data governance

- **Data Quality**: Demonstrates the emphasis on ensuring the accuracy and reliability of data.

- **Compliance**: Highlights the importance of adhering to legal and regulatory requirements, particularly pertinent in the context of the GDPR in Europe.

- **Data Security**: Stresses the need for strong security measures to protect data.

- **Decision-Making**: Indicates the role of data governance in enhancing informed decision-making processes.

- **Data Management**: Points to the importance of efficient and effective management of data within organizations.

Benefits of Implementing Data Governance

- Enhanced Operational Efficiency Effective

- Augmented Data Usability and Accessibility

- Fortified Competitive Advantage

- Fortified Competitive Advantage

- Enhanced Risk Management

- Cultural Transformation

Use Case 1

Data Quality

The Mars Orbiter was supposed to be the first weather observer on another planet. However, as it approached Mars on 23 September 1999, it vanished in space.

Why did it happen?

The software controlling the orbiter's thruster calculated the force the thrusters needed to exert in pounds. However, a separate piece of software took in the data assuming it was in the metric unit: newtons.

What were the consequences?

Significant loss of money. According to NASA, the cost of the mission was $327.6 million total for the orbiter and lander, comprising $193.1 million for spacecraft development, $91.7 million for launching it, and $42.8 million for mission operations.

How could better Data Governance Practices have helped avoid this?

- Data quality - consistency across fields, definitions, and units of measurement.
- People and communication - managing change and communication safeguards against costly and dangerous mistakes.

Use Case 2

Improper Disposal of Sensitive Data

Used body cameras are commonly for sale online,

and several other hackers followed suit by purchasing decommissioned police cameras from eBay to see what they could find.

Why did it happen?

- Data was not destroyed properly before retiring and on-selling equipment.
- There was no encryption on the devices.

What were the consequences?

- Reputation damage.
- Leaked footage of police at work.
- Sensitive military information/secrets could possibly be obtained from this.

How could better Data Governance Practices have helped avoid this?

Better data literacy and understanding of the information lifecycle.

Information Management Lifecycle

Right:
Information Management Lifecycle
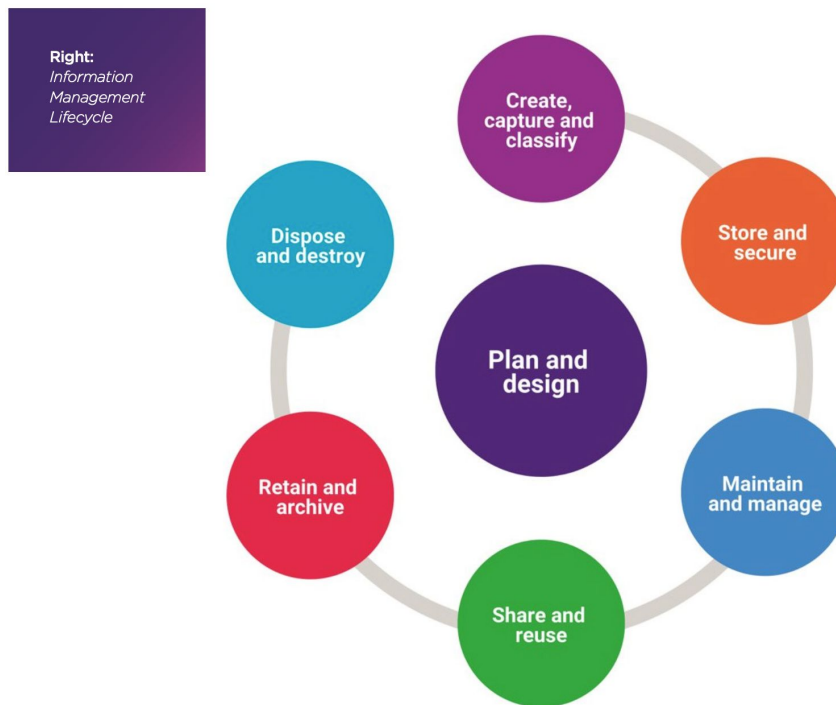
Data Catalog

Data Lineage

Thank you!

nothin' is impossible until its done