



# Data Governance and Security

ผู้ช่วยศาสตราจารย์ ดร.สมเกียรติ โกศลสมบัติ

สาขาวิชาวิทยาศาสตร์และนวัตกรรมข้อมูล

วิทยาลัยสหวิทยาการ มหาวิทยาลัยธรรมศาสตร์

# Contents

- Data Governance in Cloud Computing
- Data Security in Cloud Computing
- Tools for Data Governance and Security in Cloud
- References

# Data Governance in Cloud Computing





- Data Governance in Cloud Computing refers to the framework, policies, processes, and tools used to manage, protect, and ensure the quality, accessibility, and compliance of data stored and processed in cloud environments.
- It establishes guidelines for how data is collected, stored, accessed, shared, and disposed of, ensuring alignment with organizational objectives, security requirements, and regulatory standards.
- องค์ประกอบของ Data Governance บน Cloud
  - Data Quality Management
  - Data Ownership and Stewardship
  - Data Classification
  - Data Lineage
  - Metadata Management
  - Data Security and Privacy

# Core Components in Cloud Computing

- Data Quality Management
- Data Ownership and Stewardship
- Data Classification
- Data Lineage
- Metadata Management
- Data Security and Privacy



# Data Quality Management

- Data Quality Management (DQM) is the process of ensuring that data is accurate, complete, reliable, and consistent across its entire lifecycle. It involves setting standards, monitoring data quality, and implementing processes to clean, validate, and maintain high-quality data.
    - Accuracy – The data reflects the real-world situation correctly
    - Completeness – No missing values or critical data fields
    - Consistency – Data is the same across different sources or systems
    - Timeliness – Data is up to date and available when needed
    - Validity – Data conforms to the expected format or standard
    - Uniqueness – No duplicate records in the dataset
- 
- 
- 
- 

# Data Ownership and Stewardship

- Data Ownership refers to the individual, team, or role within an organization that has legal and operational authority over a specific dataset.
  - Defining who can access or modify the data
  - Ensuring data quality, security, and compliance
  - Approving how the data is used or shared
- Data Stewardship is the practice of managing and overseeing an organization's data to ensure its quality, consistency, usability, and trustworthiness.
  - Clean and validate data (Fix duplicates, correct errors)
  - Maintain metadata (e.g., data definitions, formats)
  - Monitor data quality over time
  - Coordinate with IT and business teams to resolve data issues

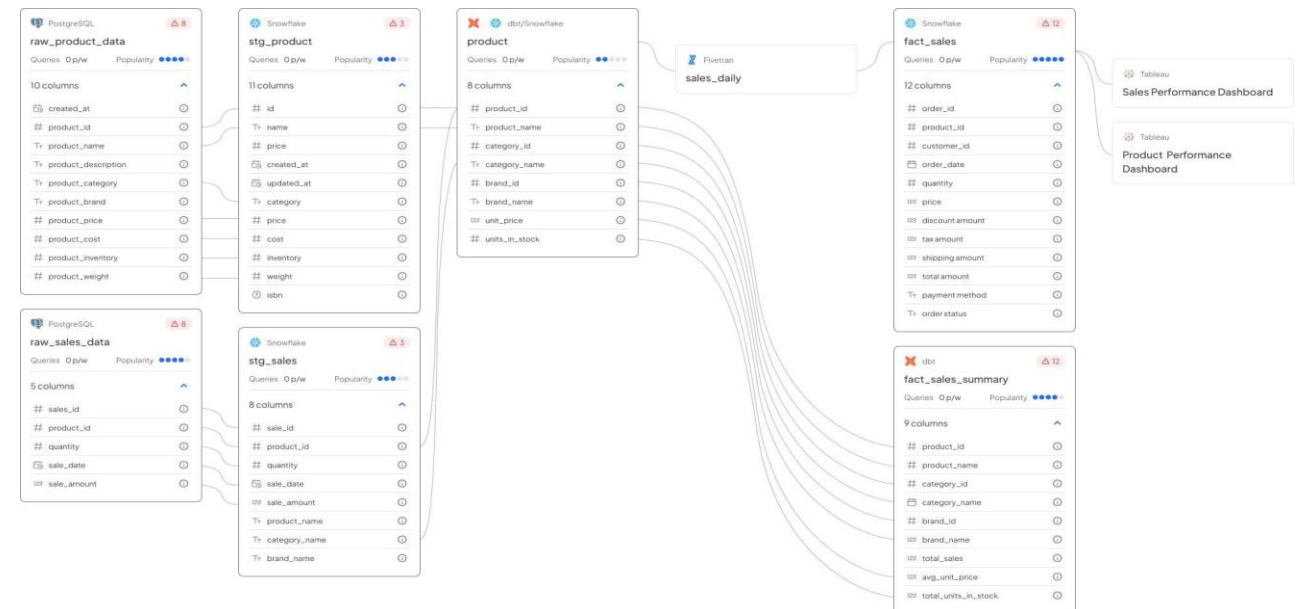
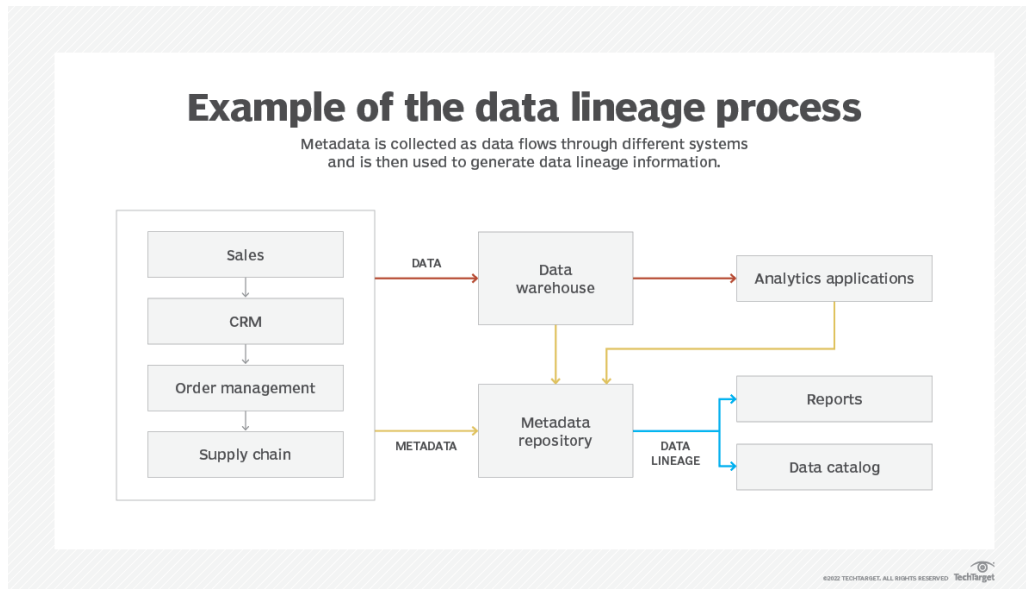
Data Owner (Strategic) - Policy, Access, and Accountability

Data Steward (Operational) – Data quality, usage, and implementation







# Core Components in Cloud Computing

- Data Classification
  - Data Classification is the process of organizing data into categories based on its sensitivity, value, and risk level — so it can be protected, managed, and used appropriately.
- Data Lineage
  - Data Lineage is the process of tracking and visualizing the flow of data through its lifecycle — from the original source to its final destination.





# Metadata Management

- Metadata Management is essential for helping organizations find, understand, and govern their data effectively.
  - It's like a map + dictionary + audit trail for data environment.
  - Technical metadata
    - Data Type, File size, Schema
  - Business metadata
    - Business context, Definition of “Customer ID”
  - Operational metadata
    - Data source, refresh frequency
  - Lineage metadata
    - Source system -> Transformation -> Report
- 
- 
- 
- 



# Data Security and Privacy



# Data Security and Privacy

## Other Information Security Properties



A

ความถูกต้องแท้จริง (Authenticity)

คุณลักษณะเฉพาะเพื่อยืนยันความถูกต้องแท้จริงถึงตัวตนผู้ใช้งาน

A

ความรับผิดชอบ (Accountability)

ความรับผิดชอบที่สามารถตรวจสอบได้

N

การห้ามปฏิเสธความรับผิดชอบ (Non-repudiation)

วิธีการที่ผู้ส่งและผู้รับข้อความ จะไม่สามารถปฏิเสธการส่งหรือการรับข้อความนั้นได้ หากได้ดำเนินการนั้นไปแล้ว

R

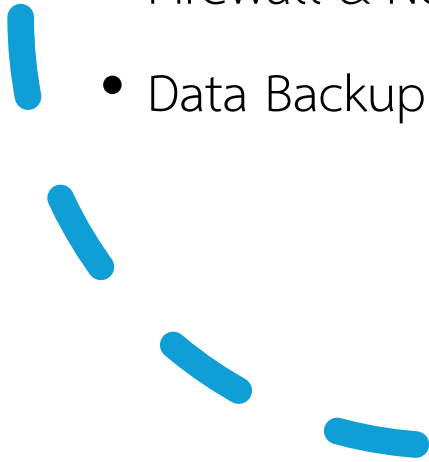
ความน่าเชื่อถือ (Reliability)

ความสามารถในการให้บริการได้ตามที่กำหนดไว้

ภาพประกอบ : Shutterstock.com



# Data Security and Privacy

- Protects data from unauthorized access, corruption, or loss
  - Encryption
  - Access Control
  - Authentication & Authorization
  - Firewall & Network Security
  - Data Backup & Recovery
- 

# Data Security and Privacy

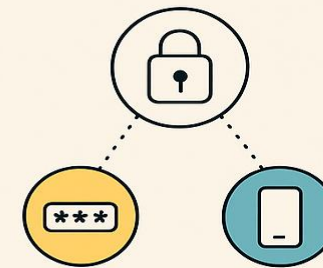
- IAM stands for Identity and Access Management — it's a framework of policies, processes, and technologies that ensures the right people (or systems) have the right access to the right resources at the right time.
- อัตลักษณ์ (Identity) คือ ลักษณะเฉพาะของบุคคลซึ่งสามารถบ่งบอกหรือจำแนกได้โดยคุณลักษณะ หรือชุดของคุณลักษณะที่เกี่ยวข้องกับตัวบุคคลนั้น ตัวอย่างคุณลักษณะ (Attribute) ที่เกี่ยวข้องกับบุคคลธรรมดา เช่น เลขประจำตัว ชื่อ ที่อยู่ วันเดือนปีเกิด เบอร์โทรศัพท์ ภาพใบหน้า อีเมล หรือข้อมูลระบุอุปกรณ์ที่บุคคลใช้งาน เป็นต้น
- การยืนยันตัวตน (Authentication) เป็นกระบวนการตรวจสอบสิ่งที่ใช้ยืนยันตัวตน โดยมีวัตถุประสงค์เพื่อให้มั่นใจว่าบุคคลที่กำลังเข้าใช้บริการครอบครองและควบคุมสิ่งที่ใช้ยืนยันตัวตนนั้นจริง
- การอนุญาตให้เข้าถึงข้อมูล (Authorization) เป็นกระบวนการที่ระบบใช้ในการกำหนดว่า ใครสามารถเข้าถึงอะไรได้บ้าง หรือการอนุญาตให้เข้าถึงข้อมูลหรือทรัพยากรตามสิทธิ์ที่กำหนดไว้

# Authentication

- Authentication Factor
  - Something You Know
    - Password, PIN
  - Something You Have
    - Cryptographic Key, OTP, Out-of-Band Device
  - Something You Are
  - Biometric Data เช่น ใบหน้า, ม่านตา, ลายนิ้วมือ
- Multi Factor Authentication
  - Two Factor เช่น Password, OTP

## TWO-FACTOR vs MULTI-FACTOR AUTHENTICATION

### TWO-FACTOR AUTHENTICATION (2FA)



Uses 2 types from different categories

### MULTI-FACTOR AUTHENTICATION (MFA)



Uses 2 or more types from any category

### CATEGORIES OF AUTHENTICATION FACTORS



SOMETHING  
YOU KNOW



SOMETHING  
YOU HAVE



SOMETHING  
YOU ARE





# Data Privacy

- Data Privacy refers to the right and practice of individuals or organizations to control how their personal or sensitive information is collected, used, shared, and stored.
- It focuses on who has access to data, how it's handled, and whether users have given consent.

ลักษณะ	Data Security	Data Privacy
Focus	How data is protected	Who can use data and why
Goal	Prevent unauthorized access or breaches	Ensure ethical and legal use of data
Tools	Encryption, Masking, Firewall, Access Control	Policies, Consent Forms



# Tools for Data Governance and Security in Cloud

- Cloud Native
    - Azure Purview - Data Catalog, Lineage, Classification
    - Azure AD (Entra ID) – Identity Management and SSO
    - Google Cloud Data Catalog – Metadata and Governance
    - Google Cloud IAM – Identity and Access Control
    - AWS IAM – Identity and Access Management
    - AWS Lake Formation – Data Catalog, Governance
  - Third-Party Data Governance
    - Collibra (Data Ownership and Stewardship), Alation (Data Catalog), Informatica, Apache Atlas, Talend
- 
- 
- 
- 



# References

- Cybersecurity Foundation by THNCA
- <https://www.techtarget.com/searchdatamanagement/feature/15-top-data-governance-tools-to-know-about>
- <https://www.databricks.com/discover/data-governance>