

# Demystifying Data Marts, Lakes, Hubs, and Warehouses & Their Role in Privacy-Aware Modern Architectures

Technology Operations Track



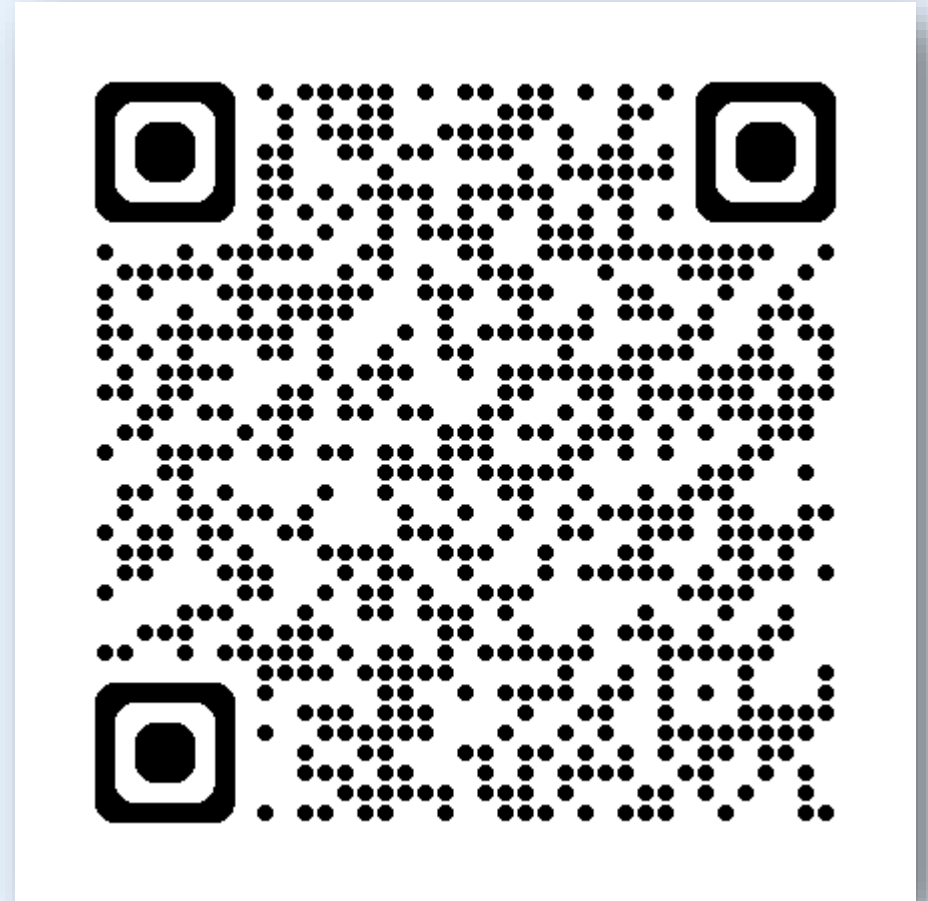
# Presenter: Mike Butler

- Senior Director of Data Delivery at OneAmerica Financial
- Creator of VERY long session titles
- 20+ Year Career as a Technologist and Data Practitioner
- Prior Director of BI, CTO, CTIO, CDO, Consultant, Adjunct Professor
- BSU Alumni

## Connect with me!

 LinkedIn -> [www.linkedin.com/in/mikebutlerin/](https://www.linkedin.com/in/mikebutlerin/)

- Blog -> [www.thedatahitchhiker.com](https://www.thedatahitchhiker.com)
  - Slides for this session will be made available



# Key Questions Today



Why are building data platform and programs so challenging?



What are some of key concepts and technologies in modern architectures?



How does privacy concern impact design and leadership decisions for modern architecture?



What should I be doing now to prepare for the AI onslaught?

# Why is building good data platforms so challenging?



**Governance**  
(Regulation, Privacy, Ethics)



**Technology complexity**



**Infinite Game**



**Talent**



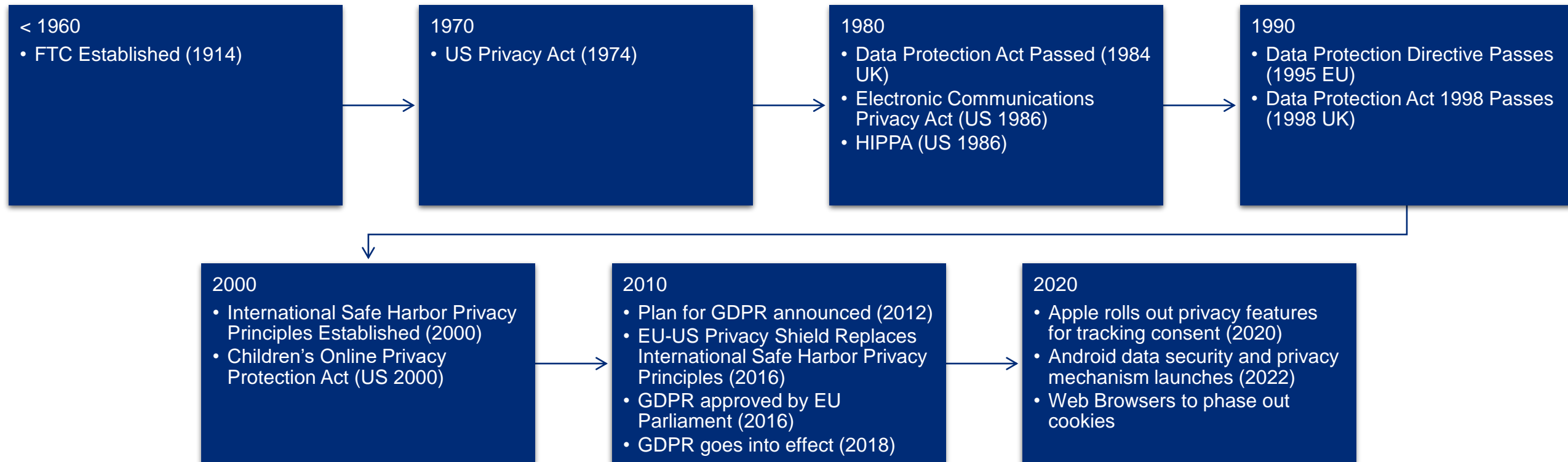
**Monetization of Data**



**3 V's**  
Volume, Variety, and Velocity

# Data Privacy

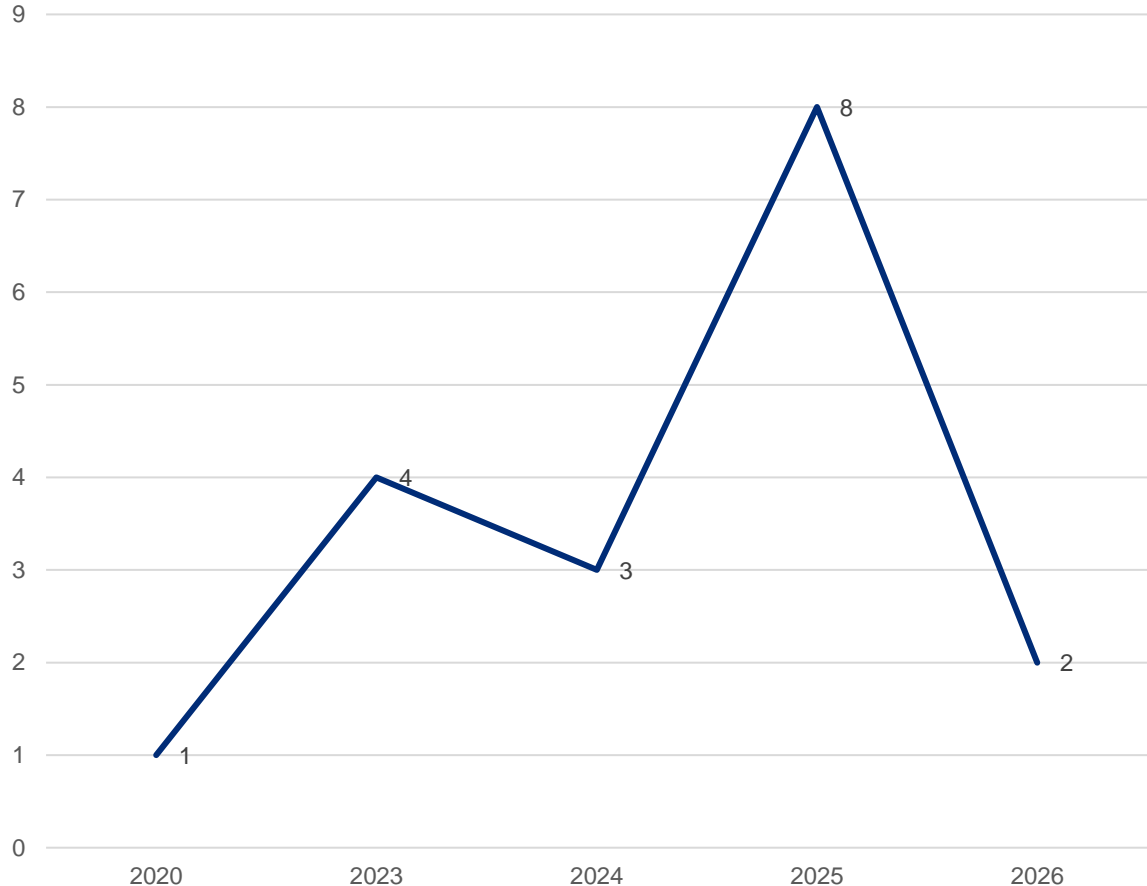
# A Brief History of Data Privacy



# Data Privacy Landscape

STATE	LAW SIGNED	EFFECTIVE FROM
CALIFORNIA	<a href="#">CALIFORNIA CONSUMER PRIVACY ACT</a>	JAN. 1, 2020
COLORADO	<a href="#">COLORADO PRIVACY ACT</a>	JUL. 1, 2023
CONNECTICUT	<a href="#">CONNECTICUT DATA PRIVACY ACT</a>	JUL. 1, 2023
DELAWARE	DELAWARE PERSONAL DATA PRIVACY ACT	JAN. 1, 2025
INDIANA	INDIANA CONSUMER DATA PROTECTION ACT	JAN. 1, 2026
IOWA	IOWA CONSUMER DATA PROTECTION ACT	JAN. 1, 2025
KENTUCKY	KENTUCKY CONSUMER DATA PROTECTION ACT	JAN. 1, 2026
MARYLAND	MARYLAND ONLINE DATA PRIVACY ACT	OCT. 1, 2025
MINNESOTA	MINNESOTA CONSUMER DATA PRIVACY ACT	JUL. 31, 2025
MONTANA	MONTANA CONSUMER DATA PRIVACY ACT	OCT. 1, 2024
NEBRASKA	NEBRASKA DATA PRIVACY ACT	JAN. 1, 2025
NEW HAMPSHIRE	NEW HAMPSHIRE PRIVACY ACT	JAN. 1, 2025
NEW JERSEY	NEW JERSEY DATA PRIVACY ACT	JAN. 15, 2025
OREGON	OREGON CONSUMER PRIVACY ACT	JUL. 1, 2024
TENNESSEE	TENNESSEE INFORMATION PROTECTION ACT	JUL. 1, 2025
TEXAS	TEXAS DATA PRIVACY & SECURITY ACT	JUL. 1, 2024
UTAH	<a href="#">UTAH CONSUMER PRIVACY ACT</a>	DEC. 31, 2023
VIRGINIA	<a href="#">VIRGINIA CONSUMER DATA PROTECTION ACT</a>	JAN. 1, 2023

Number of State Privacy Acts Going into Effect



# Data Privacy Basic Principles

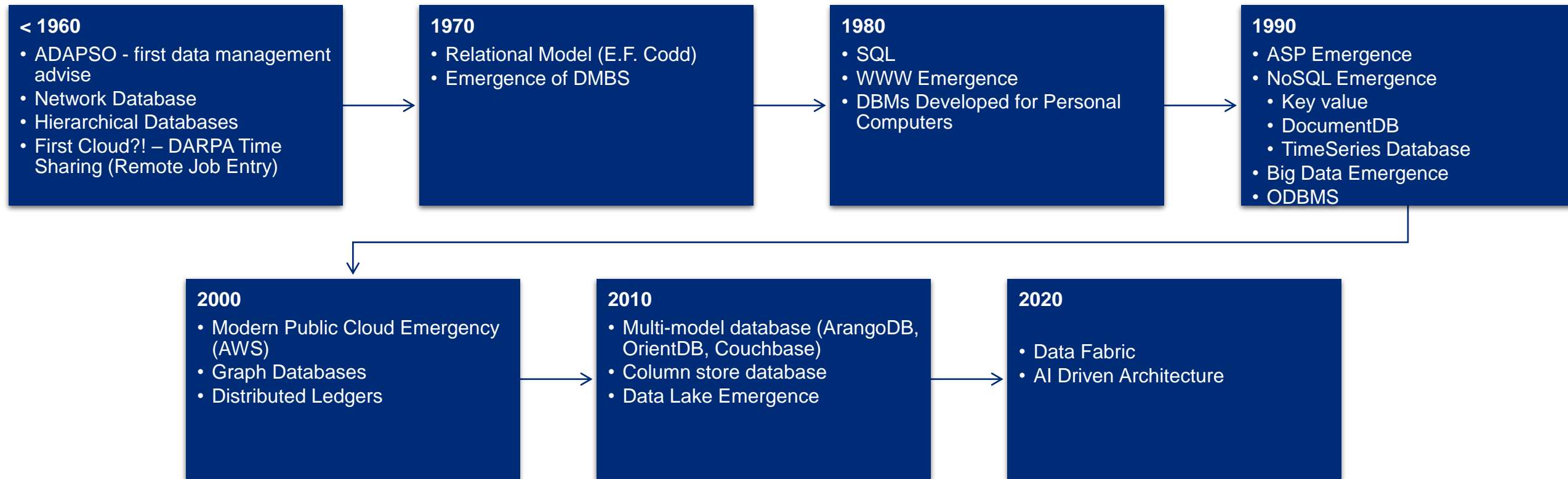
- 1. Right to Access:** Consumers can view the data a business collects about them and see which third parties it is shared with.
- 2. Right to Rectification:** Consumers can ask for corrections to any inaccurate or outdated personal data.
- 3. Right to Erasure:** Consumers can request that their personal data be deleted.
- 4. Right to Restrict Processing:** Consumers can limit how businesses process their data.
- 5. Right to Data Portability:** Consumers can request their data in a commonly used format.
- 6. Right to Opt-Out:** Consumers can choose to prevent their data from being sold to third parties.



# Data Management



# A Brief History of Data Management



# Gartner: The Hype Cycle for Data Management 2023



As of July 2023

Plateau will be reached: ○ <2 yrs. ● 2-5 yrs. ● 5-10 yrs. ▲ >10 yrs. ⊗ Obsolete before plateau

# Data Landscape (Sample)

Operational Data Store

Data Fabric

Cloud Data Warehouse

Data Governance

Columnar Database

Data Warehouse

Data Mesh

Distributed Database

Data Lake

Data Observability

Data Mart

ML/AI Classifiers

Graph Database

Data Inventory

AI-Driven Architecture

Data Offense / Data Defense

Data Lakehouse

Kappa Architecture

Vector Database

NoSQL Database

Lambda Architecture

Data Pipeline

Data Hub

Data Products

ETL

DataOps

Marketplace

ELT

Data Integration

Data Monetization

Active Metadata

Hadoop

GenAI  
(for Data Management)

Metadata

Data Sharing

Data Democratization

# Data Landscape: Organizing the Chaos

## Concepts

Data Fabric	AI-Driven Architecture	Data Integration
Data Mesh	DataOps	ETL
Data Hub	Data Observability	ELT
Data Governance	Data Offense / Data Defense	Lambda Architecture
Data Products	Data Monetization	Kappa Architecture
Data Marketplace	Data Pipeline	Active Metadata
Data Sharing	Data Democratization	

## Tools

Operational Data Store	Vector Database	Data Inventory
Data Warehouse	NoSQL Database	ML/AI Classifier
Data Lake	Graph Database	Columnar Database
Data Lakehouse	Distributed Database	Cloud Data Warehouse
Data Mart	GenAI (for Data Management)	Metadata
Knowledge Graph	Hadoop	

# Exploring a Key Concepts for 2025

## ETL vs ELT

ETL: Data is extracted from sources, transformed, and then loaded into a data warehouse. (Traditional)

ELT: Data is extracted and loaded first, then transformed as needed. (Modern)

## Data Fabric

An architecture that provides a unified data management framework, integrating various data sources, usually virtually, for seamless access and processing.

## Data Observability

Data observability refers to the practice of monitoring, managing and maintaining data in a way that ensures its quality, availability and reliability across various processes, systems and pipelines within an organization.

## Data Marketplace

A platform where data providers and consumers can rent, checkout, buy, sell, or exchange data assets, often with tools for data discovery and governance.

## AI-Driven Architecture

An architectural pattern that enables data for efficient usage in AI workloads. This may consist of pushing data from a warehouse into a lake (reverse pattern from today).

## Data Mesh

A data mesh is a decentralized data architecture where domain-specific teams own and manage their data as products, using shared infrastructure and federated governance principles.

## Active Metadata

Active metadata is metadata that is continuously collected, processed, and used to automate data management tasks and improve data quality through intelligent, action-oriented systems.

## Data Offense / Data Defense

Data Offense is an approach to strategy to use your data for custom focused uses (revenues).

Data Defense is an approach to data strategy that focuses on legal, financial, compliance, and IT concerns.

# Exploring Key Tools for 2025

## Data Mart

A subset of a data warehouse focused on a specific business line or team.

*Teradata, DB2, SQL Server, Oracle, Tableau, PowerBI*

## Data Warehouse

A system used for reporting and data analysis, storing structured data from multiple sources.

*Teradata, DB2, SQL Server, Oracle*

## Data Lake

A centralized repository that stores raw data in its native format until needed.

*Azure ADLS, Amazon S3, Google Cloud, Databricks*

## Data Lakehouse

A hybrid architecture that combines the features of data lakes and data warehouses.

*Databricks, Snowflake*

## Cloud Data Warehouse

A scalable, managed service hosted in the cloud for storing and analyzing data with flexible and scalable compute.

*Databricks, Snowflake, Azure Synapse, Amazon Redshift, Google Bigquery*

## Vector Databases

A database optimized for storing and querying high-dimensional vector data, often used in AI.

*Dedicated: Pinecone, chroma, Milvus, Weaviate, LanceDB*

*Supported Search: PostgreSQL, redis, elasticsearch.*

## ML/AI Classifiers

An algorithm that categorizes data into predefined classes based on input features.

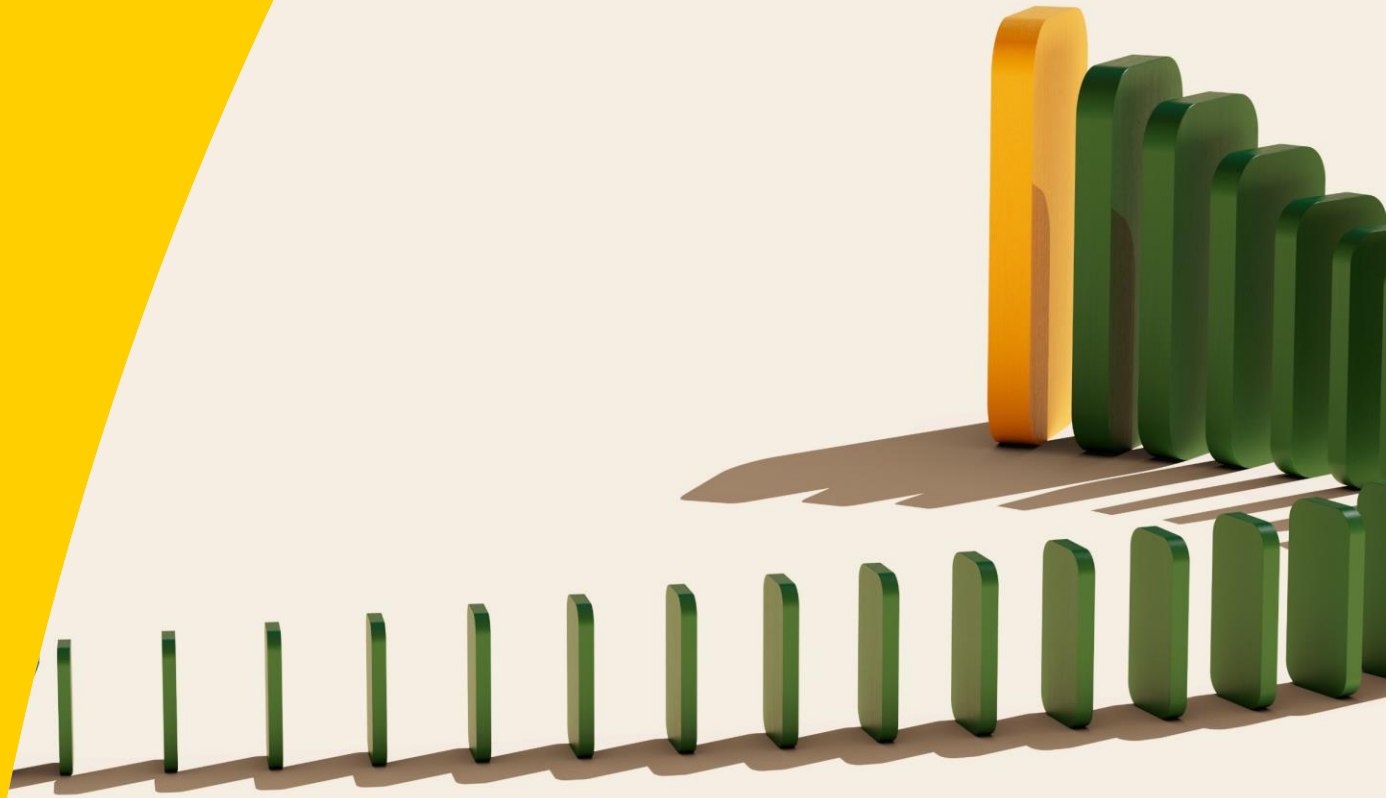
*Microsoft Purview, Informatica, Apache Atlas, Python/R*

## GenAI/LLMs

Short for Generative AI, it refers to AI models that can generate new content, such as text, images, or music, based on learned patterns.

*GPT-4, Snowflake Artic, Hugging Face, MLFlow*

# Aligning Your Strategy





# 4 Key Considerations for Your Data Future

1

Look to implement data observability and modern ELT for transparency and accuracy.

2

Leverage hybrid-architectures of Data Lakes, Warehouses, Lake houses, and Marts with ELT for modern pipelines for offence and defense.

3

Loosely couple your data architecture to prepare for constantly changing privacy.

4

Consider your future use of AI and design your architecture to support the workloads with centralized controls for privacy.

# One America<sup>SM</sup> Financial

**Life Insurance**

**Retirement**

**Employee Benefits**

**Long-Term Care**

[OneAmerica.com](https://www.OneAmerica.com)