

# Course: Software Technologies and Enterprise Architecture



## **WEEK 11: Data Architecture**

**Lemlem Kassa(Dr.)**

**Addis Ababa Science and Technology University (AASTU)**

**June , 2026**

# WEEK 11: Data Architecture

---

## Contents

- Introduction to Data Architecture
- Benefits of Data Architecture
- Key Data Architecture Patterns
- Characteristics of Modern Data Architecture
- Types of Data Architecture
- Key Frameworks in Data Architecture
- Challenges in Data Architecture and the Solutions
- Top Best Practices for Effective Data Architecture



# WEEK 11: Learning Objectives



**By the end of this unit, the student should be able to**

- Define Data Architecture and explain its importance.
- Describe the benefits of Data Architecture in organizations.
- Identify and compare key Data Architecture patterns
- Explain the characteristics of modern data architectures.
- Differentiate between various types of data architectures.
- Understand major frameworks used in data architecture.
- Analyze common challenges in data architecture and propose solutions.
- Apply best practices for designing effective and scalable data architectures.

# Introduction to Data Architecture

---

- Data architecture is like the blueprint for how data flows and is organized in a system.
- Data architecture is the plan that shows how data is collected, stored, accessed, and used in a company or system.
- It defines the rules and standards for handling data. This includes:
  - Where the data will be stored (databases or cloud storage),
  - How different pieces of data are connected (like relationships between tables),
  - How the data will move between different parts of the system (data pipelines).
  - Good data architecture ensures that data is easy to find, secure, and reliable, which is important for making decisions and solving problems.

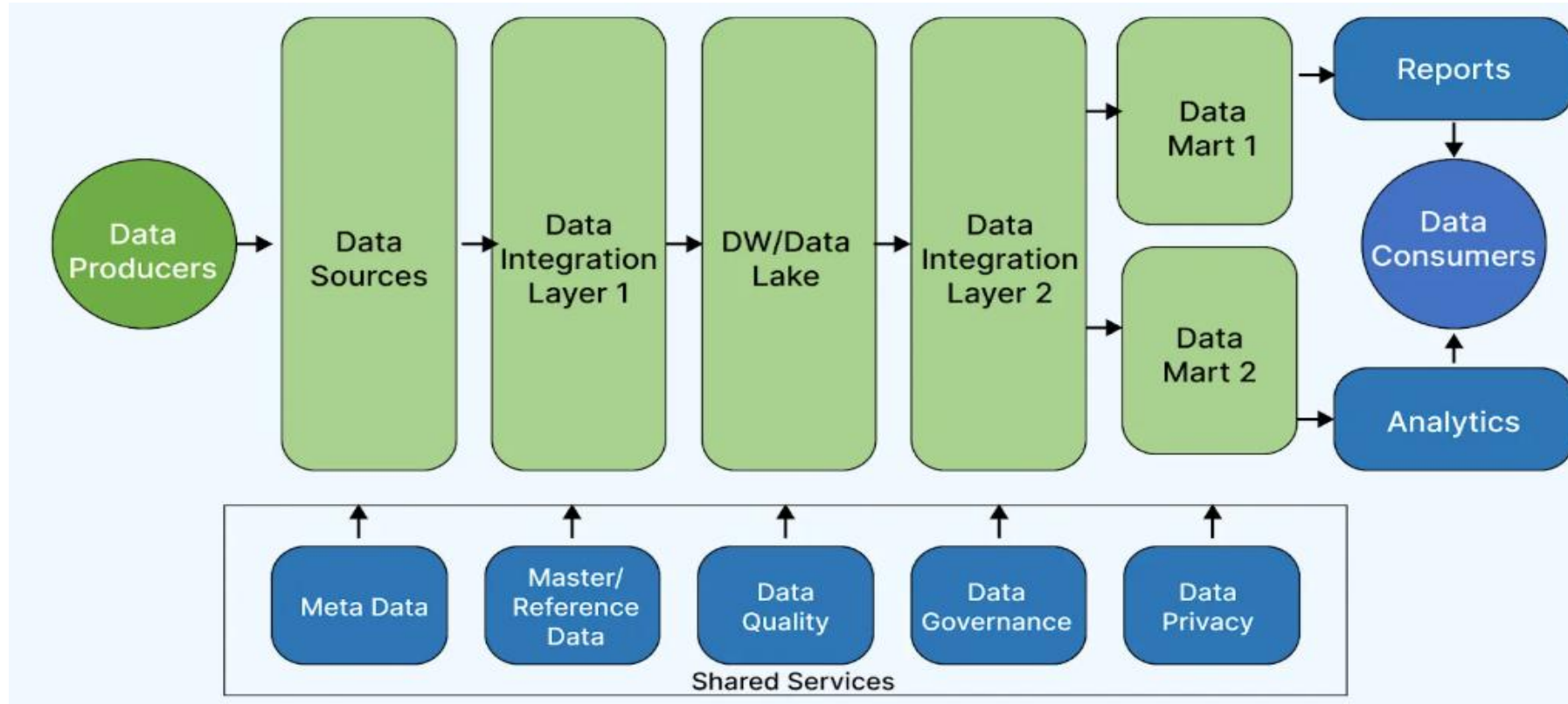
# Cont. ...Introduction to Data Architecture

---

- Info-Tech's five-tier data architecture model summarizes an organization's data environment at a logical level.
- Data flows from left to right but can also flow from the presentation layer (how the data is presented in the real world, i.e., reports, dashboards, etc.) back to the warehousing layer for repatriation of data.
  - *Sources* - Where all the data enters the organization.
  - *Integration and Translation* – Where integration, transformation and aggregation occur.
  - *Data Warehouse* – Where the data rests in long-term storage.
  - *Analytics* – Where data is used for a purpose.
  - *Presentation* – Where data is presented in a knowledge form.

# Cont. ...Introduction to Data Architecture

- Generic data Architecture Model



[1]. What Is Data Architecture?, Rohit Sharma, UpGrad, Jan. 2025.  
<https://www.upgrad.com/blog/what-is-data-architecture/>

# Cont. ...Introduction to Data Architecture

---

## Why is data architecture important?

- A modern data architecture can help unify and standardize enterprise data, enabling seamless [data sharing](#) across business domains.
- It also provides a scalable foundation for advanced use cases like real-time [data analytics](#) and [generative AI](#), helping teams extract value from data faster and more reliably.
- As technologies like the [Internet of Things](#) (IoT) generate new data sources, a well-designed architecture ensures that data remains manageable, integrated and useful throughout its lifecycle. It can reduce [redundancy](#), improve [data quality](#) and help eliminate [silos](#) by connecting systems across the enterprise.

# Benefits of Data Architecture

---

## 1. Scalability for Growing Data Needs

- Data volume has been growing. This growth increases the need for scalable data architectures to manage storage and processing effectively.
- Advanced data architectures ensure scalability to manage increasing data volumes effectively.

## 2. Support for Decentralized Frameworks

- Modern architectures like data mesh promote domain-oriented ownership of data.
- This supports distributed teams and ensures better adaptability for large enterprises.

# Cont. ...Benefits of Data Architecture

---

## 3. Enhanced Decision-Making

- AI-powered architectures integrate real-time analytics. Real-time analytics helps businesses process large datasets faster, enabling quicker and more accurate decisions.

## 4. Improved Data Quality and Governance

- Data architectures implement policies and standards for consistent, high-quality data. High-quality data ensures regulatory compliance and improves operational efficiency across organizations.

## 5. Increased Flexibility with Cloud Integration

- Sovereign cloud initiatives in India reduce reliance on global providers and ensure localized data control, enhancing compliance and operational efficiency.

# Cont. ...Benefits of Data Architecture

---

## 6. Increased Flexibility with Cloud Integration

- Sovereign cloud initiatives in India reduce reliance on global providers.
- Localized systems improve data control and enhance [digital security](#) for businesses.
- Flexible architectures enable localized data management and therefore enhanced data control.

## 7. Cost Optimization

- Advanced architectures streamline workflows and eliminate redundancy.
- Efficient workflows in modern data architectures reduce storage costs and optimize expenses for scaling operations.

# Cont. ...Benefits of Data Architecture

---

## 8. Facilitation of Real-Time Collaboration

- Modern systems provide decentralized data access. Decentralized access enables real-time collaboration and efficient data sharing."

## 9. Future-Proof Infrastructure

- It is expected that by 2027, more than 50% of organizations will adopt cloud data warehousing for scalability.
- Cloud [data warehousing](#) with modern data architecture ensures infrastructures remain adaptable to emerging technologies and needs.

# Cont. ...Benefits of Data Architecture

---

## 10. Global Competitiveness

- Projects like OpenAI's \$500 billion Stargate initiative highlight global advancements.
- Modern data architectures equip businesses to stay competitive in the digital economy.

## 11. Better Resource Utilization

- Decentralized frameworks reduce bottlenecks and improve efficiency.
- Domain ownership improves resource allocation for better operational efficiency.

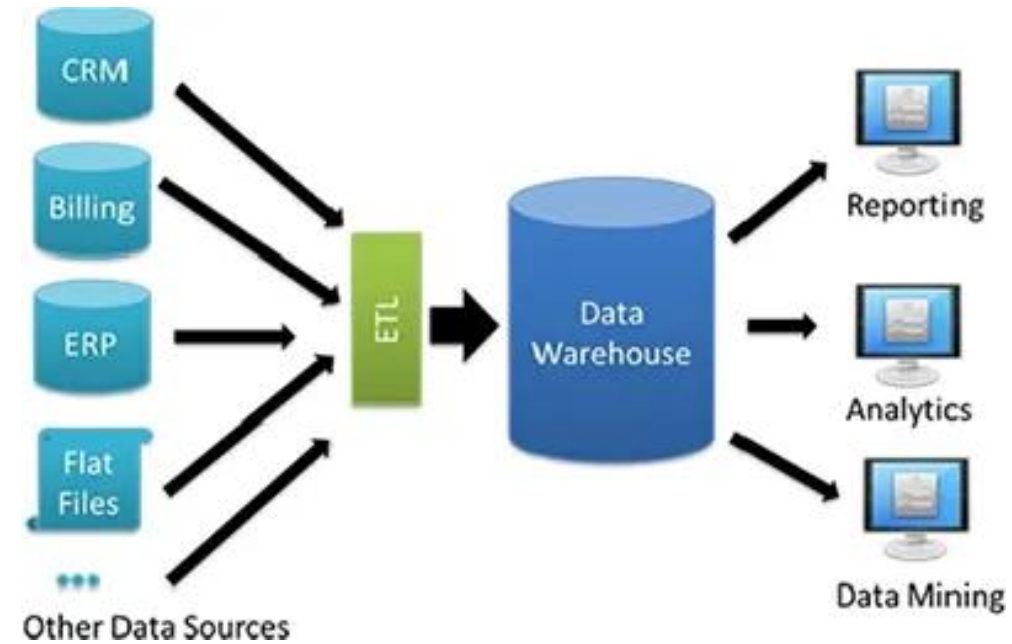
[1]. What Is Data Architecture?, Rohit Sharma\_UpGrad, Jan. 2025.  
<https://www.upgrad.com/blog/what-is-data-architecture/>

# Key Data Architecture Patterns

- Several key patterns have emerged as essential for managing and leveraging data effectively.
- Each pattern has its strengths, weaknesses, and ideal use cases.

## 1. Data Warehouses

- Data warehouses are centralized repositories designed to store structured data for reporting and analysis.
- They are optimized for query performance and data aggregation, making them ideal for business intelligence applications.



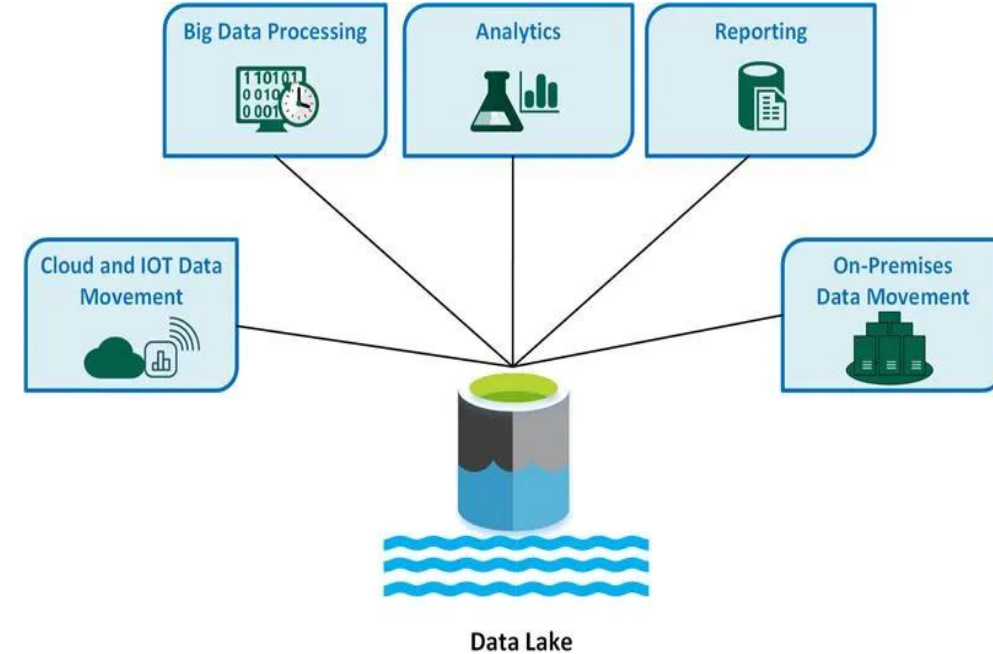
# Cont. ...Key Data Architecture Patterns

## 2. Data Lakes

- Data lakes store vast amounts of raw, unstructured, and semi-structured data.
- Unlike data warehouses, data lakes do not impose a strict schema on data, allowing for more flexibility in data ingestion and storage.

### Characteristics:

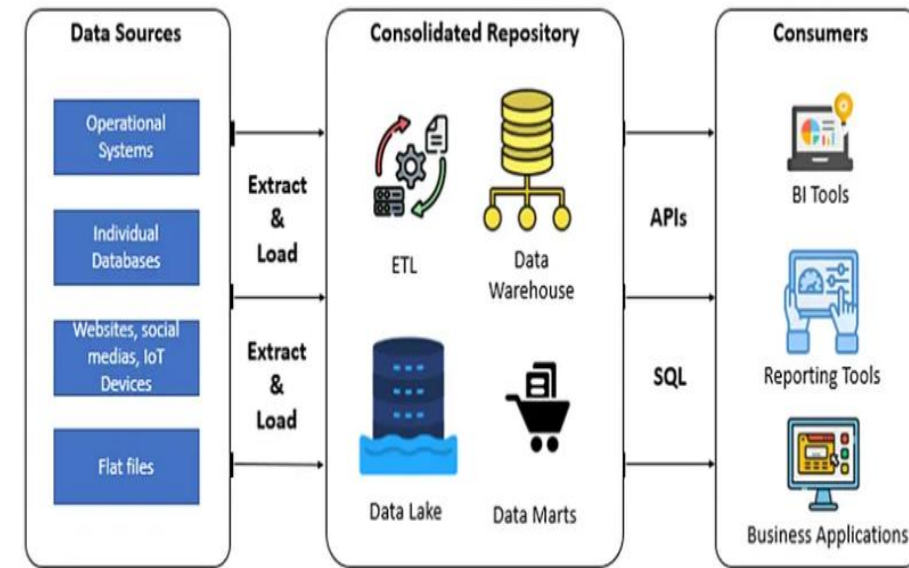
- Raw Data Storage: Stores data in its native format.
- Scalability: Can handle large volumes of data.
- Schema-on-Read: Schema is applied when data is read, not when it is stored.



# Cont. ...Key Data Architecture Patterns

## 3. Data Fabric

- An architectural approach that creates a unified data environment by connecting disparate data sources and applications across on-premises and cloud environments.
- It aims to provide seamless access and data integration, ensuring consistent data management and governance.



**Characteristics:** Provides a single point of access to all data sources. Connects diverse data environments, enabling data movement and integration.

- Utilizes AI and machine learning to automate data management tasks. Ensures data quality, security, and compliance across the data landscape.

# Characteristics of Modern Data Architecture

---

- Modern data architecture leverages AI-driven systems, shared data platforms, and elastic resources to address the challenges of growing data complexity.
- Modern data architecture frameworks emphasize user-driven designs and AI integration, ensuring systems adapt to evolving data needs.

## a) User-Driven

- Modern data architecture empowers users by enabling self-service analytics. Tools like Power BI provide drag-and-drop interfaces for creating reports without coding.
- This reduces reliance on IT teams, speeding up insights and decision-making processes.

# Cont. ...Characteristics of Modern Data Architecture

---

## b. Built on Shared Data

- Shared data architecture eliminates silos, enabling seamless access for all stakeholders. Shared datasets also improve collaboration across departments, aligning everyone with a unified source of truth.
- **Example:-** Amazon Redshift uses centralized storage, ensuring data consistency and avoiding duplication

## c. Automated

- Automation in data architecture reduces manual interventions by automating processes like data ingestion, cleaning, and transformation.
- **Example:-** Apache NiFi automates data pipelines by setting triggers and workflows, ensuring tasks are completed consistently and without delays

# Cont. ...Characteristics of Modern Data Architecture

---

## d. AI-Driven

- AI integration enables intelligent decision-making and real-time processing. Systems like Snowflake use AI to detect anomalies in data, optimize query performance, and suggest data trends.
- This ensures businesses extract meaningful insights quickly and stay competitive.

## e. Elastic

- Elastic architectures dynamically scale resources based on workloads, ensuring uninterrupted operations during high demand.

**Example:-** Google BigQuery scales up automatically, while scaling down during low usage periods reduces operational costs.

- This elasticity also helps control costs, as unused resources are scaled down during low usage periods.

# Cont. ...Characteristics of Modern Data Architecture

---

## f. Simple

- Simplicity is achieved through intuitive designs and minimal configurations. Databricks offers a unified platform for data engineering and [machine learning](#), simplifying complex workflows into user-friendly dashboards.
- This reduces training time and boosts adoption.

## g. Secure

- Security is reinforced through encryption, access controls, and compliance standards.
- For instance, Azure Data Lake encrypts data both at rest and in transit, while role-based access ensures only authorized users handle sensitive information.
- Built-in compliance with standards like GDPR further safeguards organizations from legal risks.

# Types of Data Architecture

- Data architecture evolves to meet diverse business demands, emphasizing scalability and security. Frameworks like Data Mesh and Cloud-Native Architecture enable adaptability.
- Breakdown of key types, their features, and examples integrated into descriptions:

Type	Key Features	Use Cases
Traditional Data Architecture	<ul style="list-style-type: none"><li>• Centralized and rigid structure. Relies on on-premise databases with limited scalability and flexibility.</li><li>• For example, older ERP systems like SAP ECC used traditional architectures to manage fixed workloads and static data.</li></ul>	Ideal for legacy systems in manufacturing and small-scale businesses with predictable data requirements.

# Cont. ... Types of Data Architecture

Type	Key Features	Use Cases
Modern Data Architecture	<ul style="list-style-type: none"><li>• Highly flexible, scalable, and AI-ready. Integrates cloud and real-time analytics.</li><li>• Platforms like Amazon and Flipkart use modern architectures to offer real-time inventory tracking and personalized recommendations.</li></ul>	<p>Suitable for dynamic industries like e-commerce, healthcare, and logistics requiring agile data systems.</p>
Data Mesh	<ul style="list-style-type: none"><li>• Decentralized and domain-oriented. Assigns data ownership to individual teams for efficiency.</li><li>• Netflix uses data mesh to manage microservices, ensuring scalability while empowering teams with independent data ownership.</li></ul>	<p>Ideal for large organizations with distributed teams, enabling better collaboration and autonomy.</p>

# Cont. ... Types of Data Architecture

Type	Key Features	Use Cases
Hybrid Data Architecture	<ul style="list-style-type: none"><li>• Combines on-premise and cloud systems. Ensures compliance with local regulations while leveraging cloud benefits.</li><li>• Indian banks use hybrid setups during cloud migrations to balance security and regulatory requirements.</li></ul>	Suitable for industries transitioning from legacy systems while adhering to strict compliance norms.
Data Warehouses	<ul style="list-style-type: none"><li>• Optimized for structured data and analytics. Requires data transformation during ingestion.</li><li>• Banks use data warehouses to generate quick financial reports, such as dashboards tracking account balances.</li></ul>	Best for structured, query-intensive operations in finance, retail, and supply chain management.

# Cont. ... Types of Data Architecture

Type	Key Features	Use Cases
Cloud-Native Data Architecture	<ul style="list-style-type: none"><li>• Fully cloud-based and scalable. Easily integrates with AI/ML tools for automation and advanced analytics.</li><li>• Snowflake powers AI-driven analytics for companies, offering dynamic scaling based on workloads.</li></ul>	Ideal for startups and enterprises focused on scalability, automation, and diverse data processing needs.

# Cont. ... Types of Data Architecture

---

## Core Components of Data Architecture

- Efficient data architecture relies on core components that streamline data collection, processing, storage, and analysis for seamless operations.
- Key components, such as data pipelines and APIs, are integral to understanding what data architecture is in practice. Below is a detailed breakdown of these components:

### 1. Data Pipelines

- Data pipelines manage the flow of data from collection to refinement and storage. They automate tasks such as cleaning, transforming, and aggregating data.
- **Example:** Companies like Uber use [Apache Kafka](#)-based pipelines to process millions of real-time ride and driver updates.
- DataOps platforms, such as Informatica's Intelligent Data Pipeline, now integrate AI for predictive pipeline monitoring and optimization.

# Cont. ... Types of Data Architecture

---

## 2. Cloud Storage

- Cloud storage provides scalable and secure repositories for data. It allows access to data from anywhere with an internet connection.
- **Example:** Google Cloud Storage is widely used for storing large datasets for analytics and machine learning projects.
- The rise of sovereign clouds in India, such as RBI's proposed cloud services, ensures compliance with local data protection regulations.

## 3. APIs (Application Programming Interfaces)

- APIs enable seamless communication between systems, allowing applications to share data efficiently.
- **Example:** Stripe's API allows real-time payment processing for e-commerce platforms.
- OpenAPI 3.1 introduced improved JSON Schema support, simplifying API integration and enhancing developer productivity.

# Cont. ... Types of Data Architecture

---

## 5. Data Streaming

- Data streaming involves processing continuous flows of data in real-time for instant analysis and decision-making.
  - Example: Netflix uses Apache Flink for real-time streaming to improve user experience by minimizing playback delays.
- Confluent's streaming platform now supports advanced schema registry features, reducing errors during data serialization.

[1]. What Is Data Architecture?, Rohit Sharma\_UpGrad, Jan. 2025.  
<https://www.upgrad.com/blog/what-is-data-architecture/>

# Cont. ...Types of Data Architecture

---

## 6. Kubernetes

- Kubernetes manages the deployment and scaling of containerized applications, including data processing workloads.
  - Example: Shopify uses Kubernetes to handle traffic spikes during seasonal sales by scaling its applications dynamically.
- Kubernetes 1.28 introduced security enhancements, such as network policy support for sandboxed containers, boosting data safety.

# Cont. ... Types of Data Architecture

---

## 7. Cloud Computing

- Cloud computing powers the analysis, storage, and management of data without on-premise infrastructure. Multi-cloud setups enhance flexibility and reduce vendor lock-in.
- **Example:** AWS Lambda provides serverless computing to execute code in response to events, reducing operational overhead.
- Multi-cloud solutions like Anthos by Google allow businesses to operate seamlessly across multiple cloud platforms.

## 8. Real-Time Analytics

- Real-time analytics processes and analyzes data as it is generated, enabling quick decision-making.
- **Example:** Stock trading platforms use real-time analytics to process and analyze market data instantly for better trades.
- Google BigQuery launched real-time anomaly detection for identifying irregular patterns in streaming data.

# Cont. ..Key Frameworks in Data Architecture

- Data architecture frameworks define clear methods for building systems that ensure data consistency, support compliance, and address evolving business needs effectively.
- Data architecture frameworks like TOGAF and Zachman simplify creating scalable and compliant systems. They ensure consistency, scalability, and alignment with business goals.

Framework	Role in Data Architecture	Use Cases and Updates
DAMA-DMBOK 2	A comprehensive framework by DAMA International for data management. It standardizes terminologies and guides policies for consistent governance.	Used in large organizations to establish uniform data standards across teams. Updated definitions reflect emerging needs like data ethics and AI governance.
Zachman Framework	Developed by John Zachman at IBM, it includes multiple layers for creating data models, databases, and physical structures.	Employed in enterprise-wide projects where alignment of data layers is critical. Widely used in finance for regulatory compliance models.

# Cont. ..Key Frameworks in Data Architecture

TOGAF (The Open Group Architecture Framework)	Focuses on enterprise architecture development, with Phase C dedicated to designing data architecture and roadmaps.	Popular in software enterprises for building scalable solutions. Recent updates include cloud-native strategies for faster deployment.
Data Vault 2.0	A methodology for scalable data warehouses. Focuses on historical tracking and provides flexibility for changing business needs.	Ideal for real-time analytics and data integration in industries like healthcare. Incorporates AI for automated ETL processes.
Cloud Adoption Framework (CAF)	A framework designed by cloud providers like AWS and Microsoft Azure to simplify cloud migration and architecture.	Adopted by Indian banks for secure cloud migrations while meeting RBI's compliance standards. Updated for multi-cloud strategies.

# Challenges in Data Architecture and the Solutions

---

- Integrating data from multiple sources often leads to compatibility issues and delays. Organizations adopting hybrid architectures face challenges syncing on-premise systems with cloud environments.
- Addressing these obstacles effectively ensures scalability, security, and efficient data operations.

## a) Complexity in Integration

- Integrating data from multiple sources often leads to compatibility issues and delays. Organizations adopting hybrid architectures face challenges syncing on-premise systems with cloud environments.
- **Solution:** Use integration platforms like MuleSoft or Informatica to streamline API connections. Netflix, for instance, uses APIs to ensure seamless data transfer between microservices.

# Cont. ..Challenges in Data Architecture and the Solutions

---

## b) Data Security Concerns

- Data breaches and compliance violations remain top concerns, especially with sensitive information. Migrating to cloud systems raises risks without proper security measures.
- **Solution:** Implement end-to-end encryption, identity access management, and real-time threat detection tools. Azure's advanced security features ensure compliance with GDPR and RBI standards.
- Kubernetes 1.28 introduced network policy upgrades, improving containerized application security for data-centric organizations.

# Cont. ..Challenges in Data Architecture and the Solutions

---

## c)Data Quality and Governance Issues

- Poor-quality data impacts decision-making and leads to inefficiencies. Inconsistent governance across teams results in fragmented systems and errors.
- **Solution:** Implement frameworks like DAMA-DMBOK 2 for governance and automated tools for quality checks. Snowflake integrates governance features like automated data lineage and version control.
- AI-powered governance tools, such as Informatica CLAIRE, ensure data quality and consistency in real-time.

# Cont. ..Challenges in Data Architecture and the Solutions

---

## d) Cost and Resource Management

Scaling data systems without overspending is challenging. Cloud services often lead to spiraling costs without careful monitoring.

- **Solution:** Use cloud cost management platforms like AWS Cost Explorer to track and optimize spending.
- Real-time scaling tools like BigQuery help minimize resource wastage.
- Microsoft Azure introduced pricing calculators for multi-cloud strategies, aiding cost-efficient deployments.

# Cont. ..Challenges in Data Architecture and the Solutions

---

## e) Rapid Technological Changes

New technologies disrupt existing architectures, leading to compatibility and scalability challenges. Adopting AI and ML requires significant infrastructure adjustments.

- **Solution:** Invest in adaptable frameworks like Data Vault 2.0 and adopt microservices-based architectures.
- OpenAI's ongoing integration with cloud platforms demonstrates how scalable architectures are vital for supporting large-scale AI models.

# Top Best Practices for Effective Data Architecture

---

## a) Promote Collaboration Across Teams

Collaboration between business stakeholders and IT teams is critical.

- Decision-makers identify essential data, while architects design systems to meet those needs.
  - **Example:** In e-commerce, marketing teams collaborate with IT to ensure customer data is accessible for targeted campaigns.
  - Platforms like Jira and Confluence now offer features tailored for cross-department collaboration in real-time data projects.

[1]. What Is Data Architecture?, Rohit Sharma, UpGrad, Jan. 2025.  
<https://www.upgrad.com/blog/what-is-data-architecture/>

# Cont. ..Top Best Practices for Effective Data Architecture

---

## b) Prioritize Data Governance

High-quality, well-governed data is crucial for effective decision-making.

- Implementing governance frameworks ensures compliance, security, and consistency.
  - **Example:** Financial institutions use DAMA-DMBOK 2 to manage sensitive transaction data. Data stewards maintain quality and enforce governance policies.
  - Informatica's CLAIRE now integrates AI to automate data governance tasks like lineage tracking and anomaly detection.

# Cont. ..Top Best Practices for Effective Data Architecture

---

## c) Ensure Scalability and Agility

Data architecture must adapt to new technologies and changing data types.

- Avoid rigid systems tied to specific platforms or formats.
  - **Example:** Netflix's transition to data mesh demonstrates how agile architectures enable rapid scalability for distributed teams.
  - **Update:** Google's BigQuery Omni now supports multi-cloud scaling, ensuring agility across cloud platforms like AWS and Azure.

# Cont. ..Top Best Practices for Effective Data Architecture

---

## d) Embed Security at Every Level

Security protects sensitive data across industries, from healthcare to e-commerce and government.

- Use encryption, access controls, and compliance frameworks.
  - **Example:** E-commerce platforms secure customer payment data using PCI DSS standards, while government systems enforce encryption for classified information. Azure offers advanced security layers for compliance and protection
  - Kubernetes 1.28 enhances container security, making it easier to safeguard data in distributed environments.

# Cont. ..Top Best Practices for Effective Data Architecture

---

## e) Leverage Real-Time Analytics

Organizations need insights from real-time data to stay competitive.

- Architectures should support streaming data for instant analysis.
  - **Example:** Stock trading platforms use Apache Kafka for streaming analytics to make instant trade decisions.
  - Real-time analytics tools, like Snowflake's Streaming Data Service, simplify implementing streaming pipelines in complex architectures.

# Cont. ..Top Best Practices for Effective Data Architecture

---

## f) Design for Multi-Cloud and Hybrid Environments

Businesses increasingly adopt multi-cloud strategies for flexibility and risk mitigation.

- Architectures must support seamless integration across platforms.
  - **Example:** Indian banks use hybrid cloud setups to comply with RBI regulations while leveraging cloud scalability.
  - Microsoft's Cloud Adoption Framework now includes guidance for multi-cloud governance and optimization.

# Cont. ..Top Best Practices for Effective Data Architecture

---

## g) Focus on Documentation and Standards

Clear documentation ensures teams understand the data architecture and its workflows.

- Adopting widely used standards, like ISO/IEC 27001, promotes consistency and security.
- Following industry standards promotes consistency and simplifies onboarding.
  - **Example:** Retailers document supply chain and customer data flows, ensuring alignment between logistics and analytics teams for better decision-making.
  - OpenAI's standardized API documentation approach has become a benchmark for clarity in complex data systems.

# Summary

---

- Data architecture is the blueprint that defines how data is collected, stored, integrated, and used within an organization. It establishes standards and processes to ensure data is accessible, reliable, secure, and useful for business operations and decision-making.
- A well-designed data architecture improves data quality, governance, scalability, and operational efficiency. It enables organizations to make faster, data-driven decisions while reducing redundancy and supporting future growth.
- Data architecture patterns such as Data Warehouses, Data Lakes, and Data Fabrics provide different approaches to storing and managing data. Each pattern is designed to meet specific business needs related to analytics, flexibility, integration, and scalability.
- Modern data architectures are user-driven, automated, AI-powered, scalable, secure, and built on shared data platforms. These characteristics help organizations efficiently manage growing data volumes and extract insights in real time.

# Cont. ..Summary

---

- Different types of data architecture are designed to address various business requirements, including cloud-native, hybrid, and decentralized environments. They provide flexibility, adaptability, and efficient data management across organizations.
- Key components such as data pipelines, cloud storage, APIs, data streaming, Kubernetes, cloud computing, and real-time analytics work together to support data processing and delivery. These components ensure seamless data flow and accessibility.
- Organizations often face challenges related to data integration, security, governance, cost management, and rapidly changing technologies. Addressing these challenges is essential for maintaining reliable and efficient data systems.
- Effective data architecture requires strong governance, collaboration, security, scalability, documentation, and support for real-time analytics. Following these best practices helps organizations maximize the value of their data assets.

# References

---

1. What Is Data Architecture?, Rohit Sharma, UpGrad, Jan. 2025. <https://www.upgrad.com/blog/what-is-data-architecture/>
2. “But We Have the Data” – The Importance of Data Architecture, Rick Bawcum, Cimatri, 2020. <https://cimatri.com/but-we-have-the-data-the-importance-of-data-architecture/>
3. What Is a Data Architecture?, Tom Krantz and Alexandra Jonker, IBM Think (IBM), 2025. <https://www.ibm.com/think/topics/data-architecture>.
4. Data Architecture: A Brief Overview, Jayant Nehra, Towards Data Engineering (Medium), 2024. <https://medium.com/towards-data-engineering/data-architecture-a-brief-overview-a93286f3e1f7>.

