Introduction to Data Frameworks

Data frameworks are the foundation for effectively managing and deriving insights from data. They provide a structured approach to data ingestion, transformation, storage, and analysis - enabling organizations to make data-driven decisions and drive innovation.

by Prashant Shelke

The Importance of Data Frameworks

1 **Enhanced Data Governance**

Robust data frameworks ensure integrity, security, and compliance of critical data assets.

2 **Improved Decision Making**

Reliable, accessible, and well-structured data empowers data-driven decision making.

3 **Scalable Data Infrastructure**

Flexible data frameworks can accommodate growing data volumes and evolving business needs.

Common Data Framework Architectures

Lambda Architecture

Combines batch processing and real-time stream processing for robust data handling.

Kappa Architecture

Focuses solely on real-time stream processing, simplifying the overall architecture.

Tiered Architecture

Separates storage and processing layers for improved scalability and performance.

Defining Data Requirements

1

Business Objectives

Clearly define the strategic goals and data-driven use cases.

2

Data Inventory

Catalog existing data sources and assess their quality and relevance.

3

Data Modeling

Determine the necessary data structures and relationships to support the use cases.

Data Ingestion Strategies

Batch Ingestion

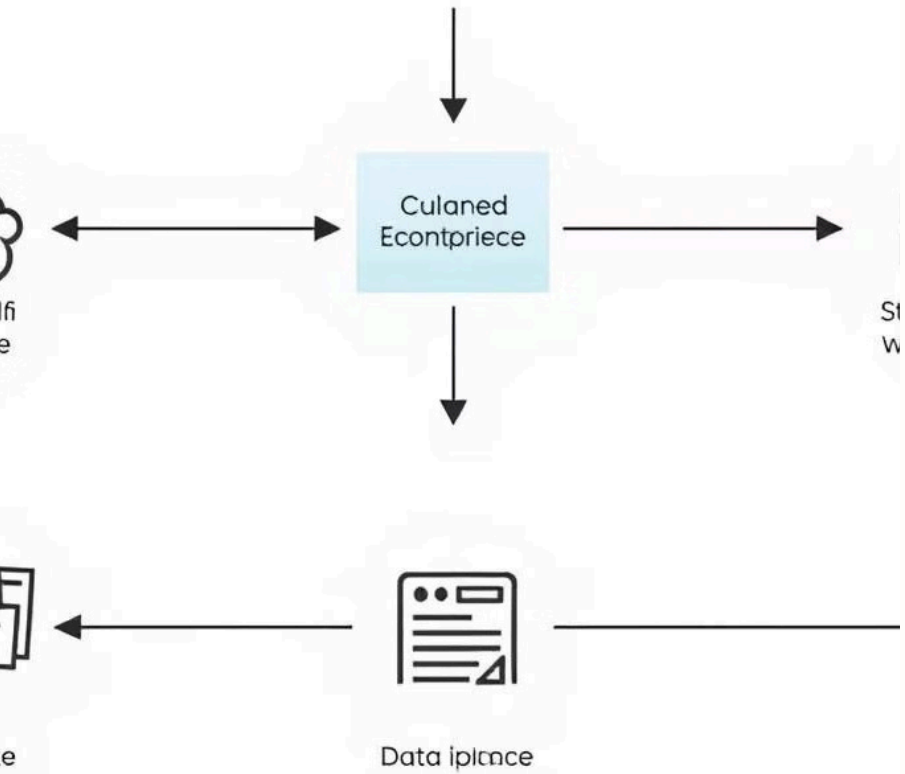
Periodic data transfers from source systems to the data framework.

Real-Time Ingestion

Continuous data streams flowing into the framework as events occur.

Hybrid Ingestion

A combination of batch and real-time ingestion to cater to diverse data needs.



Data Transformation and Normalization

1

Extract

Retrieve data from various sources.

2

Transform

Apply data cleansing, formatting, and enrichment.

3

Load

Store the processed data in the target data stores.

Data Storage and Governance

Data Warehouse	Centralized storage for structured, historical data.
Data Lake	Repository for raw, unstructured data from diverse sources.
Metadata Management	Catalog and manage data assets, policies, and lineage.

Analytical Modeling and Visualization

Reporting

Dashboards and reports to monitor key performance indicators.

Predictive Analytics

Machine learning models to forecast trends and identify patterns.

Descriptive Analytics

Insights into current state and historical performance.

Continuous Improvement and Scalability

Iterative Refinement

Regularly review and optimize the data framework based on evolving requirements.

Cloud-Native Scaling

Leverage cloud computing resources to dynamically scale storage and processing capabilities.

Automation and DevOps

Implement DevOps practices to automate deployment, monitoring, and maintenance.

Thank You...