

Data architecture definition

Data architecture describes the structure of an organization's logical and physical data assets and data management resources, according to [The Open Group Architecture Framework \(TOGAF\)](#). It is an offshoot of enterprise architecture that comprises the models, policies, rules, and standards that govern the collection, storage, arrangement, integration, and use of data in organizations. An organization's data architecture is the purview of [data architects](#).

Data architecture goals

The goal of data architecture is to translate business needs into data and system requirements and to manage data and its flow through the enterprise.

Data architecture principles

[According to Joshua Klahr](#), vice president of product management, core products, at Splunk, and formerly vice president of product management at AtScale, six principles form the foundation of modern data architecture:

1. **Data is a shared asset.** A modern data architecture needs to eliminate departmental data silos and give all stakeholders a complete view of the company.
2. **Users require adequate access to data.** Beyond breaking down silos, modern data architectures need to provide interfaces that make it easy for users to consume data using tools fit for their jobs.
3. **Security is essential.** Modern data architectures must be designed for security and they must support data policies and access controls directly on the raw data.
4. **Common vocabularies ensure common understanding.** Shared data assets, such as product catalogs, fiscal calendar dimensions, and KPI definitions, require a common vocabulary to help avoid disputes during analysis.
5. **Data should be curated.** Invest in core functions that perform data curation (modeling important relationships, cleansing raw data, and curating key dimensions and measures).
6. **Data flows should be optimized for agility.** Reduce the number of times data must be moved to reduce cost, increase data freshness, and optimize enterprise agility.

Data architecture components

[Dataversity](#) says data architecture can be synthesized into three overall components:

- **Data architecture outcomes.** These are the models, definitions, and data flows often referred to as data architecture artifacts.
- **Data architecture activities.** These are the forms, deploys, and fulfillments of data architecture intentions.
- **Data architecture behaviors.** These are the collaborations, mindsets, and skills of the various roles that affect an enterprise's data architecture.

Data architecture vs. data modeling

According to [Data Management Book of Knowledge \(DMBOK 2\)](#), data architecture defines the blueprint for managing data assets by aligning with organizational strategy to establish strategic data requirements and designs to meet those requirements. On the other hand, DMBOK 2 defines data modeling as, "the process of discovering, analyzing, representing, and communicating data requirements in a precise form called the data model."

While both data architecture and data modeling seek to bridge the gap between business goals and technology, data architecture is about the macro view that seeks to understand and support the relationships between an organization's functions, technology, and data types. Data modeling takes a more focused view of specific systems or business cases.

Data architecture frameworks

There are several enterprise architecture frameworks that commonly serve as the foundation for building an organization's data architecture framework.

- [DAMA-DMBOK 2](#). DAMA International's Data Management Body of Knowledge is a framework specifically for data management. It provides standard definitions for data management functions, deliverables, roles, and other terminology, and presents guiding principles for data management.
- [Zachman Framework for Enterprise Architecture](#). The [Zachman Framework](#) is an enterprise ontology created by John Zachman at IBM in the 1980s. The "data" column of the Zachman Framework comprises multiple layers, including architectural standards important to the business, a semantic model or conceptual/enterprise data model, an enterprise/logical data model, a physical data model, and actual databases.
- [The Open Group Architecture Framework \(TOGAF\)](#). TOGAF is an enterprise architecture methodology that offers a [high-level framework](#) for enterprise software development.

Phase C of TOGAF covers developing a data architecture and building a data architecture roadmap.

Characteristics of modern data architecture

Modern data architectures must be designed to take advantage of emerging technologies such as artificial intelligence (AI), automation, internet of things (IoT), and blockchain. Dan Sutherland, distinguished engineer and CTO, data platforms, at IBM, says [modern data architectures should hold the following characteristics in](#) common:

- **Cloud-native.** Modern data architectures are designed to support elastic scaling, high availability, end-to-end security for data in motion and data at rest, and cost and performance scalability.
- **Scalable data pipelines.** To take advantage of emerging technologies, data architectures support real-time data streaming and micro-batch data bursts.
- **Seamless data integration.** Data architectures integrate with legacy applications using standard API interfaces. They are optimized for sharing data across systems, geographies, and organizations.
- **Real-time data enablement.** Modern data architectures support the ability to deploy automated and active data validation, classification, management, and governance.
- **Decoupled and extensible.** Modern data architectures are designed to be loosely coupled, enabling services to perform minimal tasks independent of other services.

Data architect role

Data architects are senior visionaries who translate business requirements into technology requirements and define data standards and principles. The data architect is responsible for visualizing and designing an organization's enterprise data management framework. This framework describes the processes used to plan, specify, enable, create, acquire, maintain, use, archive, retrieve, control, and purge data. The data architect also "provides a standard common business vocabulary, expresses strategic requirements, outlines high-level integrated designs to meet those requirements, and aligns with enterprise strategy and related business architecture," according to [DAMA International's Data Management Body of Knowledge](#).

Data architect responsibilities

According to [Panoply](#), typical data architect responsibilities include:

- Translating business requirements into technical specifications, including data streams, integrations, transformations, databases, and data warehouses
- Defining the data architecture framework, standards and principles, including modeling, metadata, security, reference data such as product codes and client categories, and master data such as clients, vendors, materials, and employees
- Defining reference architecture, which is a pattern that others can follow to create and improve data systems
- Defining data flows, i.e., which parts of the organization generate data, which require data to function, how data flows are managed, and how data changes in transition
- Collaborating and coordinating with multiple departments, stakeholders, partners, and external vendors

Data architect vs. data engineer

The data architect and [data engineer](#) roles are closely related. In some ways, the data architect is an advanced data engineer. Data architects and data engineers work together to visualize and build the enterprise data management framework. The data architect is responsible for visualizing the "blueprint" of the complete framework that data engineers then build. According to [Dataversity](#), data architects visualize, design, and prepare data in a framework that can be used by data scientists, data engineers, or data analysts. Data engineers assist data architects in building the working framework for data search and retrieval.

Data architects have the ability to:

- Design models of data processing that implement the intended business model
- Develop diagrams representing key data entities and their relationships
- Generate a list of components needed to build the designed system
- Communicate clearly, simply, and effectively

Data architect skills

According to [Bob Lambert](#), analytics delivery lead at Anthem and former director of CapTech Consulting, important data architect skills include:

- **A foundation in systems development.** Data architects must understand the system development life cycle, project management approaches, and requirements, design, and test techniques, Lambert says.
- **Data modeling and design.** This is the core skill of the data architect and the most requested skill in data architect job descriptions, according to Lambert, who notes that this often includes SQL development and database administration.
- **Established and emerging data technologies.** Data architects need to understand established data management and reporting technologies, and have some knowledge of columnar and NoSQL databases, predictive analytics, data visualization, and unstructured data.
- **Communication and political savvy.** Data architects need people skills. They must be articulate, persuasive, and good salespeople, Lambert says, and they must conceive and portray the big data picture to others.