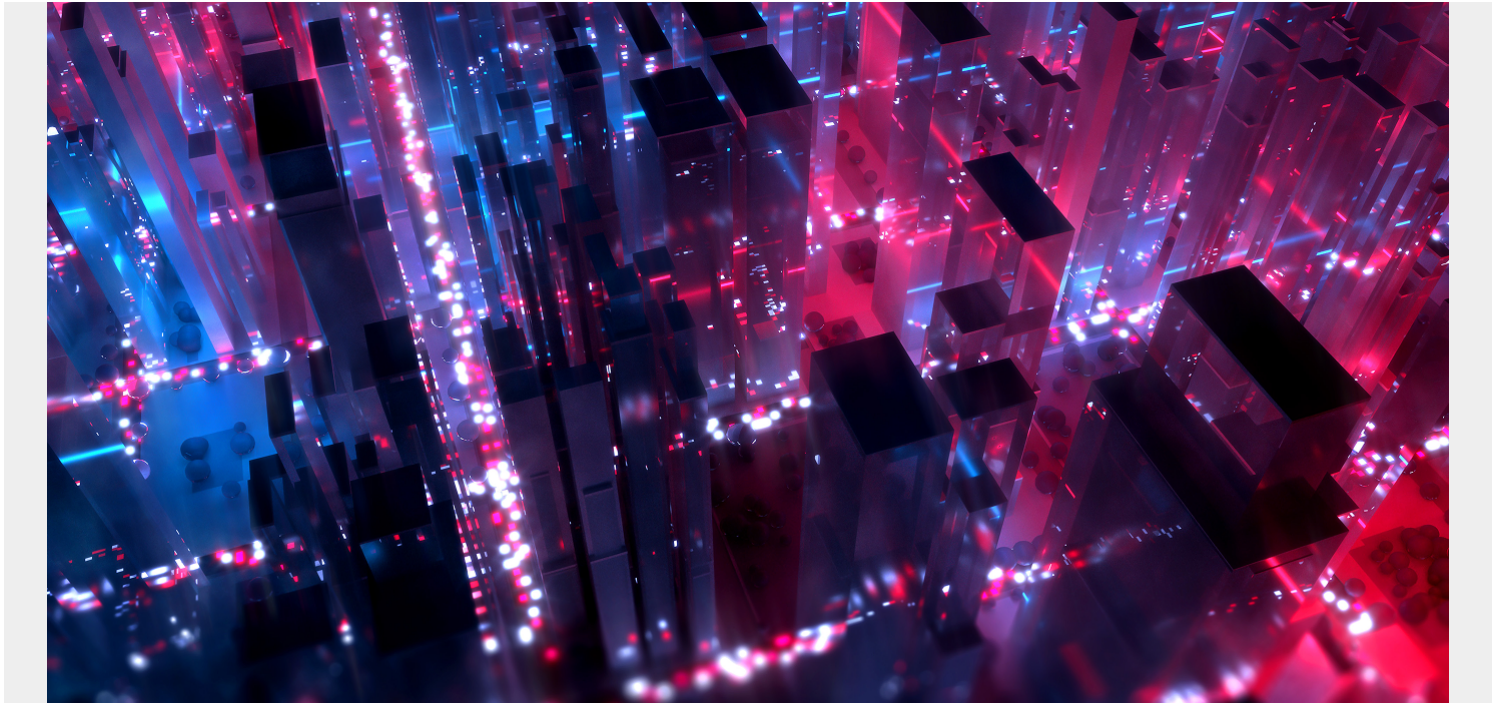# WHAT IS DATA ARCHITECTURE? COMPONENTS, PRINCIPLES & EXAMPLES



Data architecture is a framework for how [IT infrastructure](#) supports your data strategy. The goal of any data architecture is to show the company's infrastructure, including how data is acquired, transported, stored, queried, and secured.

Data architecture is the foundation of any [data strategy](#).

AI technology is radically changing data infrastructures, specifically data architecture and strategies for handling data. Data architecture defines how your organization captures data, how it's stored and managed, and how that data is used. AI applications demand better ways to handle massive volumes of data, as well as increases in computational capacity.

To handle sophisticated AI applications, your data infrastructure must support agility, both for rapidly changing business demands and to handle the fast pace of AI innovation. Your data architecture has to be highly efficient, resilient, and strong, and it must also offer scalability.
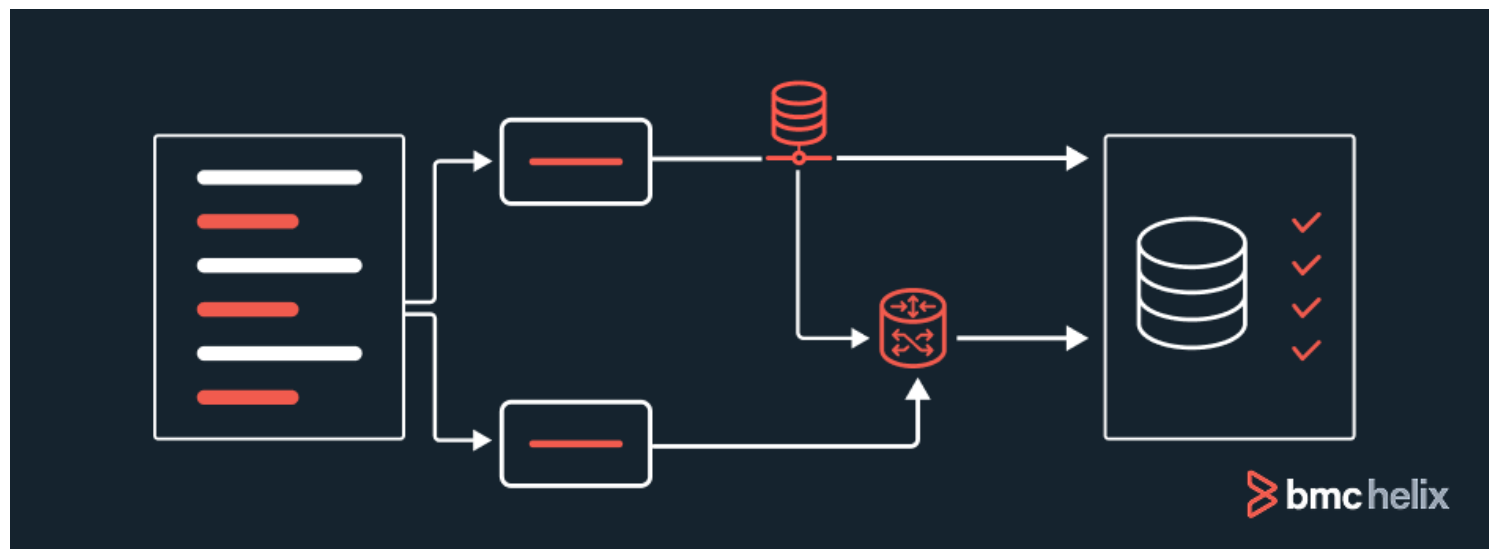
How can you achieve these requirements?

In this article, we'll look at:

- [Data architecture definition](#)
- [Architecture components](#)
- [Common data architecture frameworks](#)
- [Data standards](#)
- [The shift to new architecture](#)

Let's get started.

# What is data architecture?



Data architecture is the structure and organization of how you acquire data, store it, and manage it, and ultimately how your systems access and use it. Data architecture components include data models, rules and policies, data access and security technologies, and analytical processes and outputs.

Data architecture resolves the "how" for implementing your [data strategy](#).

# Data architecture examples

Different data architecture examples include:

- Storing a file as a .csv on a local hard drive and reading the file into [Tableau](#) on a person's computer for analysis.
- Streaming data from a set of point-of-sale registers to accounting.
- Accumulating data in a large-scale data lake and then using big data tools like Spark or Hadoop to process and analyze it.
- Capturing data and placing it where it can be managed by various business units on one platform.
- An enterprise data architecture combines everything from .csv files to data lakes and warehouses to streaming data, using data integration frameworks and business intelligence tools.

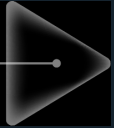# Why is data architecture important?
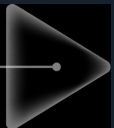
# Data Architecture Explained
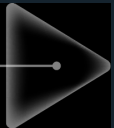
Gives a fuller picture of what is happening in the company.

Creates a better understanding of the company's data.

Grants all teams the ability to make data-driven decisions.

Ensure a system is in place to secure the data.

Outlines how data moves from source to end use.

The data architecture is 100% responsible for increasing a company's freedom to move around the world.

If agility is what is needed to avoid collapse during slow seasons or to capitalize on the spontaneous popularity of a new product, the more advanced the data architecture is, the more capable the company is to take action.

Explicitly, data architecture is important because it:

- Gives a fuller picture of what is happening in the company
- Creates a better understanding of the company's data
- Offers protocols by which data moves from its source to being analyzed and consumed by its destinations
- Ensures a system is in place to secure the data
- Grants all teams the ability to make data-driven decisions

# Key components of data architecture

The architectural components of today's data architecture world are:

- Data pipelines: Refers to the methods used to bring raw data into a data store, typically with some transformation or processing.
- Cloud storage: This model for gathering and keeping data relies on remote devices that you can access via a network.
- Application programming interfaces (APIs): This set of rules provides existing functions for connecting to, communicating with, and sharing among software.
- AI & ML models: These sets of programs find patterns in data to make decisions or predictions to solve tasks.
- Data streaming: Refers to continuously transferring data from its source or sources for use in processing into outputs.
- Kubernetes: This open-source system automates deploying, scaling, and managing applications in containers for efficiency.
- Cloud computing: Involves providing computing services on remote devices that are accessed and managed over the internet.
- Real-time analytics: Uses data, software, and hardware to analyze data as soon as it is generated.

# Data Architecture Components

Real-time analytics

AI & ML models

Cloud computing

APIs

Kubernetes

Cloud storage

Data streaming

Data pipeline
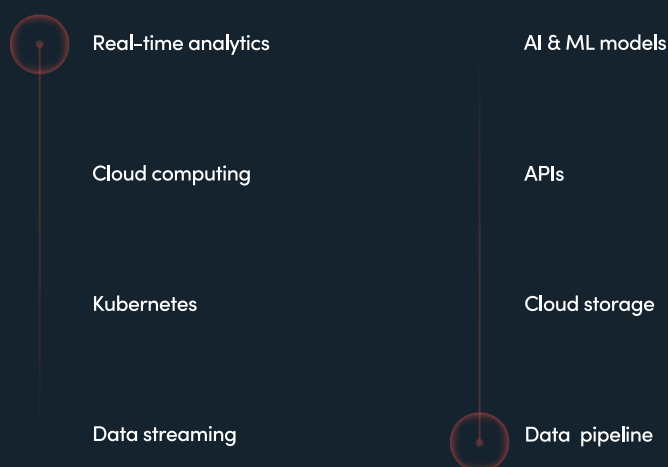
# Common data architecture frameworks

A data architecture framework is a structured approach to defining your data strategy, including how to organize data, process it, analyze it, and document it.

- **The Open Group Architectural Framework (TOGAF)**: A modular approach for creating a hierarchy and content framework that eliminates redundancy and inefficiency while boosting data usability.
- **Data Management-Body of Knowledge (DAMA-DMBOK2)**: Applies best practices for data governance, quality, and security.
- **The Zachman Framework:** Provides a logical matrix structure to support both automated and

manual systems for aligning the IT department with business goals.

# What are data standards?

Data standards are the overarching standards of a data architecture, which you apply to areas such as data schemas and security.

## Data schemas

A data schema defines how you organize data within a database, including specifying its format, relationships, and standards for storage and access. The data schema spells out:

- **Each entity that should be collected.** The Schema for contact info, for example, might include name, phone number, email, and place of work.
- **The type of data each piece should be.** For example, name is text data, phone number is integer data, email is text data, and place of work is text data.
- **The relationship of that entity to others in the database,** such as where it comes from and where it's going.

Most companies update their data schema around changing business needs, applications, and data models. As data becomes increasingly pervasive, companies are shifting away from on-premise databases to scalable cloud-native relational databases.

You can easily add data and combine data from a network of data sources into today's relational (NoSQL) databases without being restricted to a fixed hierarchy. Plus, these relational databases can grow much larger and handle adding data dynamically through integrations with analytics tools that are not possible with traditional SQL databases.

Updating and modifying your data schema, or "versioning" it, is vital. Versioning the data schema helps standardize what to find, where, and the ability to ask *when* a data set was in a location.

*(Explore data storage from database to warehouse to lake and from hot to cold.)*

## Data security

Data standards also help set the security rules for the architecture. These can be visualized in the architecture and schema by showing what data gets passed where, and, when it travels from point A to point B, how the data is secured.

Security protocols can include:

- Encrypting data during travel
- Restricting access to individuals
- Anonymizing data to decrease the value of the information upon receipt by receiving party
- Additional actions

# Shifting to new architecture

AI is driving data architecture trends, reflecting the need for processing data in real time, handling massive volumes of data from diverse sources in a multiplicity of formats, and supporting highly sophisticated queries and analytics. Trends include:

- Decentralizing data management and moving away from centralized data warehouses or even data lakes to domain- or department-specific data collections, all managed on a single platform.
- Unifying data integrations, sometimes called data fabric, using AI and automation to connect data across platforms in hybrid or multi-cloud environments.
- Processing in real-time, or ongoing streaming, to support applications like fraud protection, the function of IoT, and running AI.
- Driving data management decisions with AI at the center to automate the basics of governance, quality checking, and optimization.
- Using distributed databases and multiple models to ensure global scalability with high failover resilience.
- Designing for cybersecurity and compliance with various frameworks and regulations in mind.

When thinking about anything related to data — which is arguably everything — you should always consider the data architecture.

# Related reading

- [BMC Business of IT Blog](#)
- [BMC Machine Learning & Big Data Blog](#)
- [BMC Guides](#), offering multi-part tutorials on a variety of data products
- [Data Ethics for Companies](#)
- [3 Keys to Building Resilient Data Pipelines](#)
- [5 Questions to Ask About Data in Your SaaS Environment](#)