ONE PAGE SUMMARY

IT & Software › Data Engineering › Data Contracts

Introduction to Data Contracts

Data Contracts: Ensuring Consistency in Data Sharing

DEFINITION

> Data contracts are agreements that define the structure and quality of data shared between systems. They help ensure that data producers and consumers understand and agree on data expectations.

KEY CONCEPTS

Schema Definition

A schema defines the structure of data, including types and relationships, ensuring consistency.

Data Quality

Data contracts specify quality metrics, such as accuracy and completeness, to maintain high standards.

Versioning

Versioning allows for changes in data contracts over time while maintaining backward compatibility.

Contract Validation

Validation checks ensure that data adheres to the defined contract before being processed.

Consumer-Producer Agreement

Data contracts facilitate clear agreements between data producers and consumers regarding expectations.

EXAMPLES

→ A contract specifies that a user ID must be a 10-digit integer.

→ A data contract requires that all dates are in ISO 8601 format.

→ A contract states that a product's price must be a positive decimal number.

MEMORY TIPS

★ Think of contracts as 'data rules' that everyone must follow.

★ Remember 'SQC' for Schema, Quality, and Contract validation.

★ Visualize a handshake to represent agreement between data producers and consumers.

COMMON MISTAKES

✗ Neglecting to update contracts when data changes.

✗ Assuming all stakeholders understand the contract without clear communication.

✗ Overlooking validation checks leading to data quality issues.

QUICK RECAP

Data contracts are essential for defining data structure and quality between systems. They ensure that both producers and consumers have a clear understanding of data expectations, helping maintain data integrity.