Collibra

Must know terms for data quality and observability

Key concepts for ensuring reliable data

DATA QUALITY DIMENSIONS

Data accuracy:

The degree to which data is factually correct and precise

Data completeness:

Whether all data fields required are available

Data consistency:

The degree to which data does not conflict across sources

Data validity:

The degree to which data conforms to the expected format, type and range

Data uniqueness:

Number of distinct values and duplicate data records

Data freshness:

Whether the data is up-to-date and reflects the current state of affairs

DATA QUALITY MONITORING AND VALIDATION

Data quality rule:

The technical code used to measure and validate data against metrics and thresholds

Data profiling:

The process of analyzing data to gather information about its structure, attributes and quality

Anomaly detection:

The process of identifying patterns in data that are outside the normal ranges

Schema drift:

Columns and data types that have been added, removed or changed in data sources

Duplicate detection:

The process of identifying entities in multiple records and linking them together

Data reconciliation:

The process of validating source and target data match after data movement

Data quality and observability:

Data quality and observability is the process of monitoring and validating data quality across applications, data pipelines, and data stores, and notifying stakeholders when data does not meet quality standards.