

I. What is Metadata?

- **Metadata:** Data that describes, identifies, and contextualizes other data to determine its baseline properties and potential utility.
- **Generic Metadata:** Descriptions of baseline, inherent properties of a dataset (e.g., format, encoding, row/column count) to ensure usability.
- **Custom Metadata:** Evaluations that build on generic properties using organizational context to determine strategic value, utility, and blending potential.

II. The 5 Core Metadata Characteristics

Characteristic	Strict Technical Definition
Structure	The format, schema, and encoding layout of records and fields, ranging on a spectrum from rectangular to jagged or heterogeneous.
Granularity	The level of depth, specificity, or scale represented by a single record, defining the discrete entity or transaction captured.
Accuracy	The correctness, quality, consistency, and truthfulness of values populating the record fields of a dataset.
Temporality	The representation of time within a dataset, recognizing that records represent snapshots of entities at a specific point in time.
Scope	The breadth and coverage of a dataset, evaluated across attribute width (field count) and population coverage.

III. Refined Stage Remediation Definitions

- **Structural Remediation:** Standardizing file formatting and schema layout variations into uniform rectangular tables, and expanding categorical variables into binary indicator flags.
- **Granularity Adjustment:** Preserving the finest analytical record-level resolution in the canonical dataset while systematically creating coarser pre-aggregates to speed up reporting.
- **Accuracy Corrections:** Rectifying field anomalies by removing corrupt records, flagging inaccurate values, or performing statistical imputation.
- **Temporality Reconciliation:** Resolving snapshot conflicts and staleness by mapping valid date ranges, version control attributes, or tracking systemic changes.
- **Scope Harmonization:** Blending and merging disparate tables using aligned schemas (via joins or unions) and calibrating for systematic population bias.

Implementation: Viewing Metadata in Pandas

```
import pandas as pd
# 1. THE RAW DATA
data = {
    'customer_id': [101, 102, 103],
    'age': [22, 28, 25],
    'signup_date': pd.to_datetime(['2025-01-15', '2025-02-20', '2025-03-10'])
}
df = pd.DataFrame(data)
# 2. ADDING DESCRIPTIVE & ADMINISTRATIVE METADATA
df.attrs['dataset_name'] = "Customer_Signup_Q1"
df.attrs['author'] = "Data Science Team"
df.attrs['last_updated'] = "2026-08-11"
df.attrs['description'] = "Ages and signup dates for early 2025 cohort."

# 3. EXTRACTING AND VIEWING THE METADATA
print("--- 1. STRUCTURAL METADATA ---")
print(df.dtypes)
print("\nDataFrame Shape:", df.shape)

print("\n--- 2. DESCRIPTIVE METADATA ---")
for key, value in df.attrs.items():
    print(f"{key.upper()}: {value}")
```