



Data Quality Management Strategy

Based on COVID-19
Data Validation Pipeline

1. Solution for Data Standardization and Quality Visibility

Current Implementation Analysis:

- - Automated validation pipeline (Dagster)
- - Data quality checks (Great Expectations)

Validation dimensions:

- - Completeness, Uniqueness, Standardization, Consistency, Coverage

Proposed Enhancements:

- - Real-time monitoring dashboard
- - Automated alerting system
- - Historical quality metrics tracking
- - Self-service reports

2. Data Quality Processes

1. Data Ingestion:

- - Automated loading & initial format validation
- - Source system verification

2. Quality Validation:

- - Great Expectations checks & custom rules
- - Error logging and reporting

3. Results Processing:

- - Reports generation (Cloud storage integration)

3. Required Technologies

Code editor: VS Code.

Programming language: Python.

Libraries:

- **Pandas and NumPy:** Data manipulation and analysis.
- **Dagster:** Workflow orchestration.
- **Great Expectations (GE):** Data quality validation.
- **Azure Storage Blob:** Cloud storage integration.
- **OpenPyXL and JSON:** Excel and JSON handling.

Data visualization: Power BI.

Cloud service for storing data: Azure Blob Storage.

Version control: GitHub.

3. Roles and Responsibilities

Data Engineers:

- - Pipeline maintenance, infrastructure

Data Quality Analysts:

- - Validation Rules, Rule updates, threshold setting, exception handling

Business Stakeholders:

- - KPI definition, quality specifications, report consumption

4. Stakeholder Engagement Strategy

Communication Plan:

- - Automated reports, weekly updates, trend analysis

Engagement Activities:

- 1. Training programs: Report interpretation
- 2. Regular forums: Review meetings, feedback sessions

Success Metrics:

- - Improved data quality score, reduced validation failures
- - Faster issue detection/resolution, stakeholder satisfaction

Next Steps

Immediate Actions:

- - Form team, set up monitoring, engage stakeholders

Quick Wins:

- - Dashboards, documentation, regular reporting

Long-term Goals:

- - Full automation, predictive measures, company-wide adoption