Automating Data Remediation

Identify, analyze, and fix data issues at scale by automating Lightup's Data Remediation capabilities into existing enterprise workflows.

Datasheet

Introduction

Data remediation, also known as incident remediation, is a systematic process of identifying and addressing data discrepancies in datasets by creating steps to automatically fix data when common issues occur.

Lightup helps automate data remediation processes and efforts by offering an easy way for enterprise data teams to identify, analyze, and fix data issues at scale. These issues may include anything that compromises the integrity and usability of the data, such as:

- **Empty values or nulls**
- **Incorrect formats**
- **Inconsistent or wrong values**

To address these challenges and maintain Data Quality standards, organizations implement a data remediation strategy that outlines an end-to-end process (Figure 1), including eight workflow steps:

1. Data incident identification
2. Severity and impact assessment
3. Root cause analysis (RCA)
4. Remediation plan
5. Remediation implementation
6. Remediation validation
7. Monitoring and adjusting
8. Incident closure

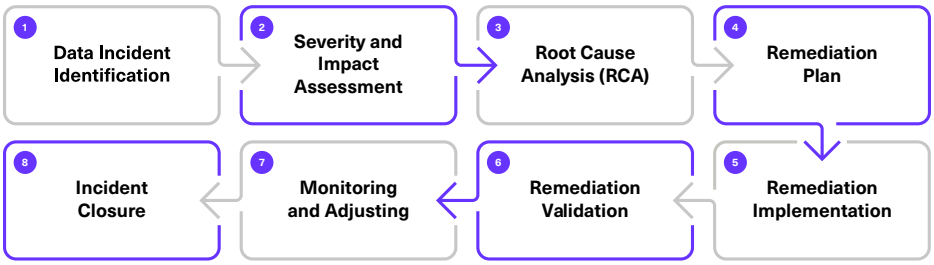

Figure 1: Sample data remediation workflow.

Key Use Cases

Data Quality Management

Address data discrepancies and maintain data integrity across various datasets.

Compliance and Regulation

Ensure compliance with data protection regulations by rectifying data issues promptly.

Operational Efficiency

Streamline data remediation processes to improve operational efficiency and reduce downtime.

Key Features

Automated Metric Collection

Proactively monitor data integrity and detect anomalies with custom metrics.

Integration with Orchestrator Tools

Seamlessly trigger remediation actions based on predefined metrics using tools like Airflow.

Automated Remediation Workflows

Target and execute remediation actions accurately to minimize manual intervention.

Validation and Feedback Loop

Ensure the effectiveness of remediation measures through comprehensive validation mechanisms.

Lightup for Data Remediation

Data remediation is essential for maintaining data integrity and reliability in enterprise data environments. By leveraging Lightup's features and integration capabilities, organizations can streamline the remediation process, mitigate Data Quality issues, and ensure consistent, high-quality data assets.

Lightup enables enterprise organizations to automate data remediation processes within existing workflows.

Detection and Alerting

Lightup performs checks to identify null values or other anomalies within the dataset. Upon detection, Lightup alerts stakeholders about the data incident, signaling the need for investigation and remediation.

Root Cause Analysis (RCA)

Once a data incident is confirmed, organizations initiate a root cause analysis process to understand why the issue occurred. This involves tracing the data flow upstream to pinpoint the origin of the problem, whether it's in data ingestion, processing, or storage.

Remediation Plan

Based on the RCA findings, organizations create a remediation plan divided into preventive measures and corrective actions. For example, to address missing values, organizations may enforce data validation rules or implement default values to prevent future occurrences.

Data Repair

In cases where data needs immediate fixing, Lightup supports data remediation by retrieving information from other sources or executing remediation scripts. This ensures that erroneous data is corrected in place, minimizing downstream impact.

How It Works

In Lightup, the data remediation process begins by measuring the appropriate metrics and monitoring data on an ongoing basis to automatically detect unexpected anomalies, at scale — without setting manual thresholds.

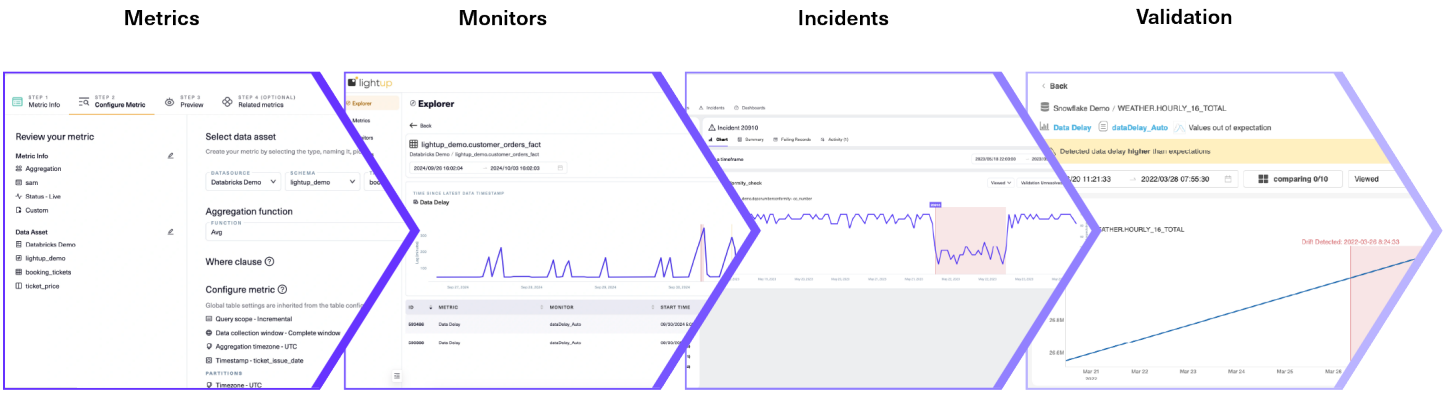

Figure 2: Data remediation in Lightup leverages four components: Metrics, Monitors, Incidents, and Validation.

With Lightup, data remediation becomes a proactive and efficient end-to-end workflow, allowing enterprise organizations to uphold Data Quality standards at scale.

1 Metric Definition and Training

Define Data Quality metrics in Lightup, specifying requirements such as field length or format. Lightup analyzes historically good data to train defined metrics in minutes, identifying patterns and anomalies to refine the remediation process.

2 Incident Detection

Using AI-powered anomaly detection algorithms, Lightup identifies data incidents and triggers alerts for further investigation.

3 Automated Remediation

Automate data fixes based on predefined rules and remediation logic, including bulk operations and simple remediation logic to ensure efficient and accurate data correction.

4 Validation and Monitoring

Validate and verify the effectiveness of remediation actions by monitoring remediated data in real time, adjusting processes as needed to maintain Data Quality standards.

Automating Data Remediation with Lightup

Through pipeline orchestration and programmatic actions, Lightup enables enterprise organizations to automate the entire data remediation process within existing workflows.

Lightup tells you which metric failed and provides all the failed records. Taking the bad records, implement your own remediation in the script. The script will fix the problem and trigger the validate fix in Lightup, which must be successful before moving to the next step.

1. Create a triggered metric in Lightup.
2. Preview and test it.
3. Put triggered metric URL into an orchestrator tool like Airflow.
4. Airflow executes the trigger to Lightup.
5. Lightup runs the metric collection.
6. Lightup runs Anomaly Detection on metric and surfaces incidents.
7. Incidents are created with the Failing Records Query showing bad output.
8. Orchestrator tool receives a failed result that the run was successful but the monitor failed, which triggers a remediation command to fix the bad values in the file using Lightup's preview API.
9. Pass bad rows into a script, which takes that information to apply the fixes. Fix.py will run that setup, run the validate fix, fix the data, and tell Lightup to validate the fix. Airflow will continue after a validated fix.

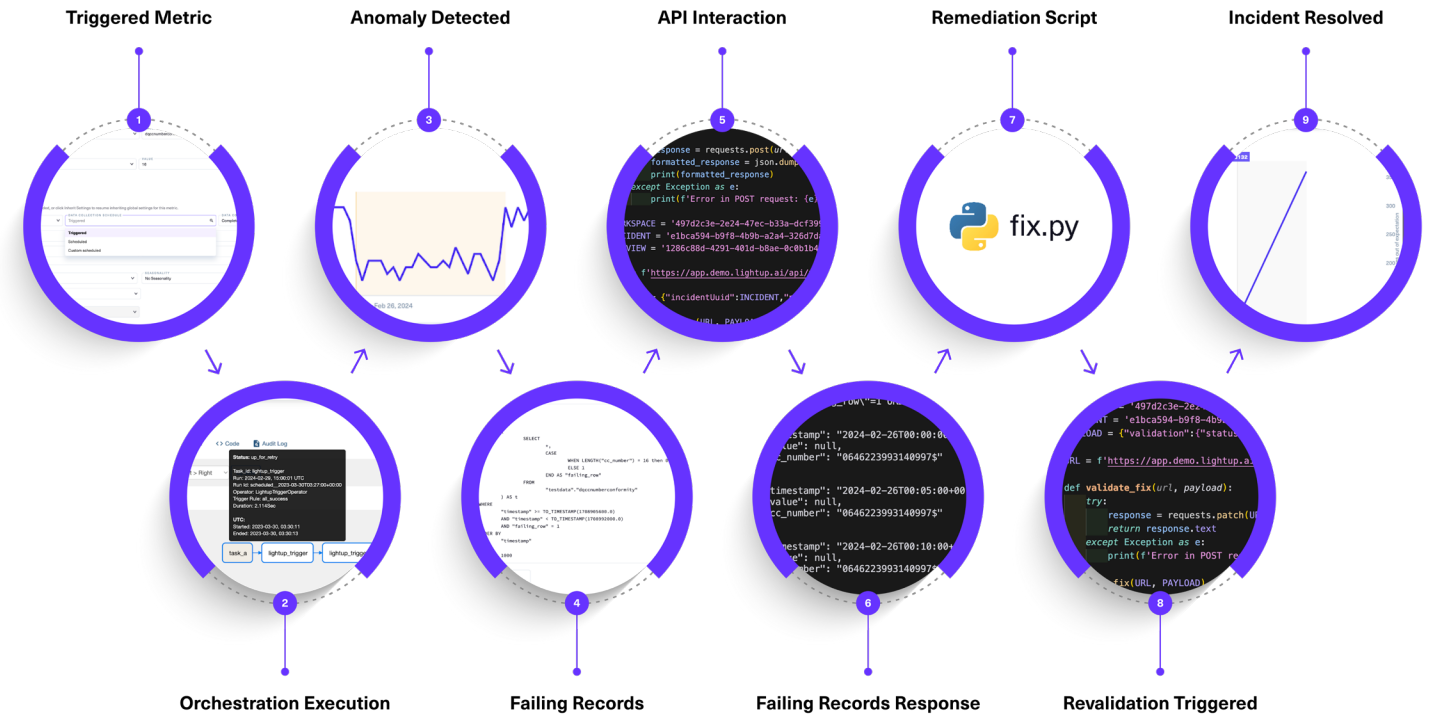

Figure 3: Automating data remediation workflows with Lightup.

Pro Tips

- Automate any possible remediation steps.
- Keep remediation logic simple to avoid unknown, unexpected, and undesirable outcomes.

Try Lightup Free for 30 Days

See how Lightup's Data Remediation works firsthand, start a free 30-day trial today. *No credit card required.*

Start Free Trial

GET IN TOUCH

info@lightup.ai

lightup.ai