File Level Validation Framework

Overview

The File Level Validation Framework is an essential step in our data processing pipeline. It ensures that the files received in the AWS S3 RAW Bucket meet specific criteria before proceeding to the next stages of validation.

Key Components

Checksum Check:

Description: Verifies the integrity of the file using a checksum algorithm.

Implementation: Utilizing the MD5 checksum algorithm to calculate and compare the checksum values.

Business Date Check:

Description: Ensures that the file contains data relevant to the specified business date.

Implementation: Extracting and validating the business date information embedded within the file.

Record Count Check:

Description: Verifies the number of records in the file against an expected count.

Implementation: Comparing the actual record count with a predefined threshold.

File Name Check:

Description: Validates the naming convention of the file.

Implementation: Examining the file name against a predefined pattern to ensure conformity.

Additional Technical File Checks:

Description: Perform one more technical check specific to the file type.

Implementation: Custom check based on file format and additional requirements.

Implementation Workflow

Files are received in the AWS S3 RAW Bucket.

Execute Checksum Check, Business Date Check, Record Count Check, File Name Check, and additional technical file checks.

If all checks pass, proceed to Field Level Validation.

If any check fails, move the failed records to the AWS S3 ERROR Bucket.

Generate a detailed log file for file-level validation, recording passed and failed records.

Store the log files in the AWS S3 LOG Bucket.

Detailed File Checks

Checksum Check

...

Business Date Check

...

Record Count Check

...

File Name Check

...

Additional Technical File Checks

...

Field Level Validation Framework

Overview

The Field Level Validation Framework focuses on detailed validation of individual fields within records to ensure data accuracy and integrity.

Key Components

Schema Check:

Description: Validates that the data adheres to the predefined schema.

Implementation: Comparing each field against the expected data type and format.

Null/Not Null Check:

Description: Verifies if mandatory fields are not null and optional fields are appropriately handled.

Implementation: Checking for null values in mandatory fields and ensuring optional fields are appropriately handled.

Uniqueness/Duplicate Check:

Description: Ensures that key fields have unique values to avoid duplicates.

Implementation: Checking for duplicate values in specified key fields.

Two Additional Technical Field Checks:

Description: Implementing two more technical checks specific to field-level requirements.

Implementation: Custom checks based on data characteristics and business rules.

Implementation Workflow

Execute Schema Check, Null/Not Null Check, Uniqueness/Duplicate Check, and two additional technical field checks on records that passed file-level validation.

If any check fails, move the failed records to the AWS S3 ERROR Bucket.

Generate a detailed log file for field-level validation, recording passed and failed records.

Store the log files in the AWS S3 LOG Bucket.

Detailed Field Checks

Schema Check

...

Null/Not Null Check

...

Uniqueness/Duplicate Check

...

Two Additional Technical Field Checks

...

Overall Process

Report Generation and Email Notification

After successful completion of both file-level and field-level validations, a comprehensive report is generated summarizing:

Number of records passed and failed in file-level validation.

Number of records passed and failed in field-level validation.

This report is then sent via email to the designated recipients.

Implementation Workflow

Perform File Level Validation.

Perform Field Level Validation on trusted records.

Generate an overall report.

Send the report via email.