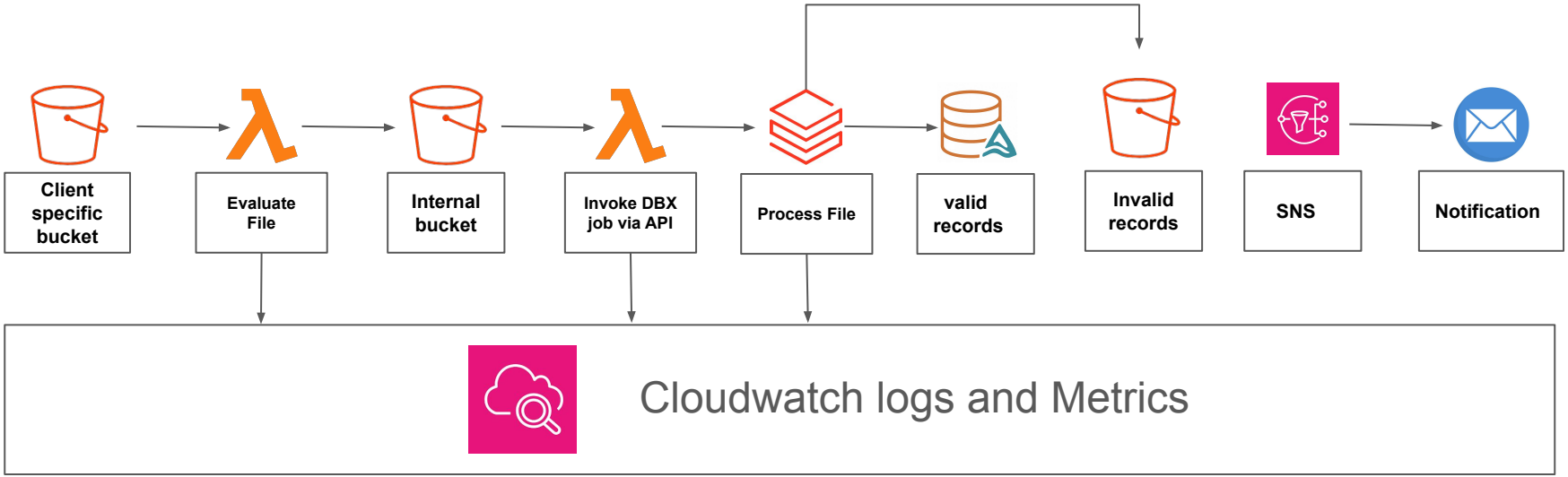
Client specific bucket
Evaluate File
Internal bucket
Invoke DBX job via API
Process File
valid records
Invalid records
SNS
Notification
Cloudwatch logs and Metrics

Radiant Graph Data Ingestion Architecture

- **Client S3 Buckets:** Each client has a dedicated S3 bucket for uploads (created manually or via onboarding).
- **Sanity Check Lambda:** S3 event triggers a Lambda function to validate file metadata and sanity.
- **Input Processing S3 Bucket:** Valid files are moved to a central input bucket (radiant-graph-input).
- **Databricks Trigger Lambda:** S3 event triggers a Lambda that invokes a Databricks notebook job, passing the file path.
- **Databricks Processing:** Notebook reads the file, separates valid/invalid records, applies compliance transformations, writes valid records to a Delta table, and failed records to a failed bucket.
- **CloudWatch Logging:** All processing steps and errors are logged to CloudWatch.
- **SNS Notification:** On error, an SNS topic sends email notifications to support.

Scaling

- **Bucket-per-client:** Each client gets a unique S3 bucket and IAM role, isolating data and access.
- **Event-driven:** S3 event notifications and Lambda functions scale automatically with file volume.
- **Databricks Jobs:** Jobs are parameterized by file path and client, allowing parallel processing.
- **Delta Lake:** Partitioned by client, date, and zip for efficient querying and storage.
- **Monitoring:** CloudWatch and SNS scale with AWS infrastructure, supporting high concurrency.

Error Monitoring

CloudWatch Logs: All errors and metrics are logged.

SNS Email Alerts: Critical failures trigger email notifications.

Ingestion Error Rate: Calculated and monitored via logs and metrics.

Compliance Design Choices

- **Encryption:** All buckets use AES256 encryption.
- **Access Control:** IAM roles restrict access to only necessary users/services.
- **De-identification:** PII is hashed, masked, or redacted before storage.
- **Audit Logging:** All actions are logged for traceability.
- **Tagging:** Resources are tagged for HIPAA/SOC2 compliance.