

Python Program Documentation:

Step-by-Step Process:

1. Connect to AWS S3:

- a. Establish a connection with AWS S3 using the provided credentials.

2. Download FHIR Files from S3:

- a. Retrieve the most recently created folder from the specified S3 bucket.
- b. Download all JSON files within the latest folder containing FHIR data.

3. Normalize FHIR Data:

- a. Flatten the JSON structure of each FHIR file to create a tabular format.
- b. Normalize the FHIR data into DataFrames, separating each resource type.

4. Deidentify Data:

- a. Anonymize or deidentify sensitive information in the data, such as personally identifiable information (PII), using a specified set of columns.

5. Ingest Data into Elasticsearch:

- a. Establish a connection with Elasticsearch.
- b. Prepare the data for ingestion by adding metadata and creating unique identifiers.
- c. Create or update Elasticsearch indices as necessary.
- d. Bulk index the deidentified data into Elasticsearch for each resource type.

6. Logging and Error Handling:

- a. Log information and errors using the configured logging system.
- b. Handle exceptions gracefully, logging any errors encountered during the process.