This document is designed to meet the requirement of the decomposition project. The top portion in this page describes how should the document provides provide documents to S3 bucket.

AWS Dynamo or SimpleDB

AWS Cloud Front

Document Consumers searched by Metadata from Dynamo or SimpleDB, then subsequently retrieves the data from CloudFront.

EMR Cluster(s) calls the ETL applications to Process metadata while file content still stores at S3. Then the processed metadata will be store in Dynamo or SimpleDB for searching purpose, obviously each metadata has reference to its data in S3

If not

Provided Presigned URL

Document Providers have access to AWS S3 Bucket

Multipart Upload

This page explains the operation of the development team and provides a way for them to quickly and accurately deploy new codes using AWS Code Deploy feature. At the end, the EMR clusters will be used to process metadata as depicted in the above page.

AWS EC2/EMR which can be used to processed old/new documents

AWS Code Deploy pushes revision deployment

Extraction / Enrichment Team