CONTENTSTACK

Guide

How to migrate content from AEM to Contentstack

Table of contents

Introduction	3
Terminologies	3
Discovery process	4
Planning	5
Types of migrations	5
Standard Processes	6
Pre-requisites	6
Methods of exporting content from AEM	6
Framework configuration and execution	8
Frequently encountered issues and resolutions	9

How to migrate content from AEM to Contentstack

Introduction

This document provides a detailed guide to migrating from a legacy CMS to Contentstack. It lists the end-to-end process required to meet your business objectives by migrating to Contentstack.

Terminologies

1. **Project team:** Whether from the customer or a solution integrator, the team is responsible for the migration.
2. **Business users:** Individuals or teams who are subject matter experts for the respective business.
3. **Third-party applications:** Business applications are integrated with the website.
4. **Technical Solutions Organisation (TSO):** The technical solutions division of Contentstack is focused on customer enablement.
5. **CMS:** All legacy Content Management Systems, collectively.
6. **Migration framework:** A set of TSO-supported and documented tools, scripts, and migration procedures. Several legacy CMS platforms are currently supported. In addition, content transformation and content import from JSON file formats into the Contentstack DXP system are available. The framework provides more expansive guidelines that can be customized according to the customer's dataset.

Discovery Process

The project team must determine the following as part of the discovery process.

1. **The goal:**
    a. **Redesign or replatform:** Determine if the site is getting redesigned.
        I. A redesign will typically take longer than a platform due to additional processes in implementing the new designs and approaches.
    b. **Current infrastructure:** Determine if the current CMS is an On-premise or Cloud version. Refer to the current architecture diagram and verify accuracy.
        I. Identify the gaps in the current architecture diagram that will be fixed in the future state.
    c. **Source of truth:** Is the current CMS the source of truth for the data? If not, seek an architecture diagram. In such cases, data will always be converted using middleware connected to the current CMS. Refer to the current architecture diagram and verify accuracy.
        I. Identify the gaps in the current architecture diagram
        II. Fix any gaps if the same architecture will be retained in the future
2. **Content Audit:** Legacy content has to be audited to ensure only relevant content is migrated to Contentstack. The following is the outcome of the exercise:
    a. The list of Pages v/s the Sitemap must be checked, and any standalone pages must be reviewed for migration
    b. Identify content that is irrelevant and needs to be removed from the migration
    c. Identify content that needs to be changed or updated. Content authors should curate the content before migrations
    d. Identify content to be migrated as is
3. **Technology stack:** The project team must select the appropriate Tech stack based on which the site will be built.
4. **Deadlines:** Determine crucial dates, milestones, and deliverables for the project.

Planning

The following tasks are essential during the planning stage:

1. **Project plan:** The project team must prepare a detailed project plan. Below is a sample plan that lists high-level tasks that can be further normalized. The absence of a project plan is a critical risk to the project.
2. **Team composition:** Identify the development team and upskill them regarding the tech stack being used. They should also start their Contentstack Academy courses, which are designed to equip them with best practices.

	Task Name	Duration	Dependencies	Resources Involved
2	Discovery			
3	Identify project trigger/purpose/objectives			
8	Audit functional and technical documentation on the current state			
11	Identify infrastructure and technology on the future state			
18	Review/scope known security & compliance requirements			
26	Define/document project requirements			
29	Deadlines			
32	Budget Earmark			
35	Planning			
45	Design			
58	Content Architecture & Modeling Phase			
76	Migrations (Optional)			
98	Development			
106	Testing & Documentation			
112	Deployment & Training Phase			
118	Maintenance & Support			

Types of Migration

Based on the discovery process, migrations can be classified into two broad categories:

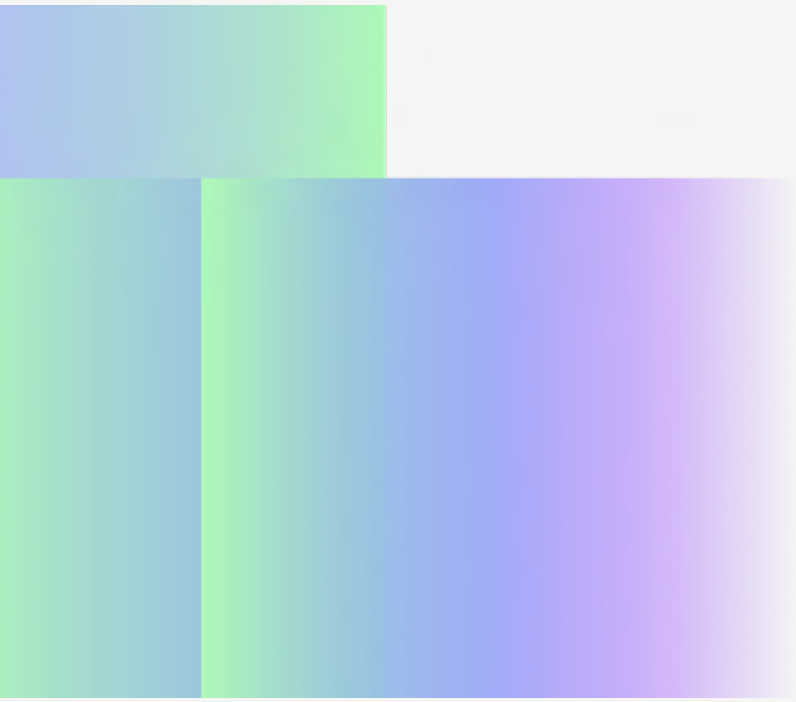
1. **Site redesign:** The following activities must be completed while planning for migrations where the site will be redesigned.
    a. A content audit should be completed
    b. Designs should be approved
    c. Content modeling must be completed and approved
    d. Once content modeling is completed, data must be exported in Contentstack's JSON-based format.
2. **Site replatform:** The following is the list of activities that must be completed while planning for migrations where the site is being replatformed.
    a. A content audit should be completed
    b. Data has to be exported in Contentstack's JSON-based format
    c. Content modeling is to be done as per best practices recommended by Contentstack

Standard Processes

1. Freeze content modeling
2. Migrations always have to be done in an internal sandbox or non-production environment
3. Sample migrated content has to be reviewed before the complete migration
4. Integrate with the revised code base and test it thoroughly, along with all integrations
5. This testing must be done in a non-production environment to avoid any customer impact
6. Defect fixes
7. Create backups of the source data
8. Go live for activities to be performed

Prerequisites

1. Contentstack account as per respective Cloud infrastructure. AWS, Google Cloud, or Microsoft Azure
2. Understanding the Command Line Interface (CLI)
3. Node version 16+



Methods of exporting content from AEM

Padding of infinity.json on HTML pages to export the AEM data into a JSON format

Access your AEM account. Users should have admin access to export content.

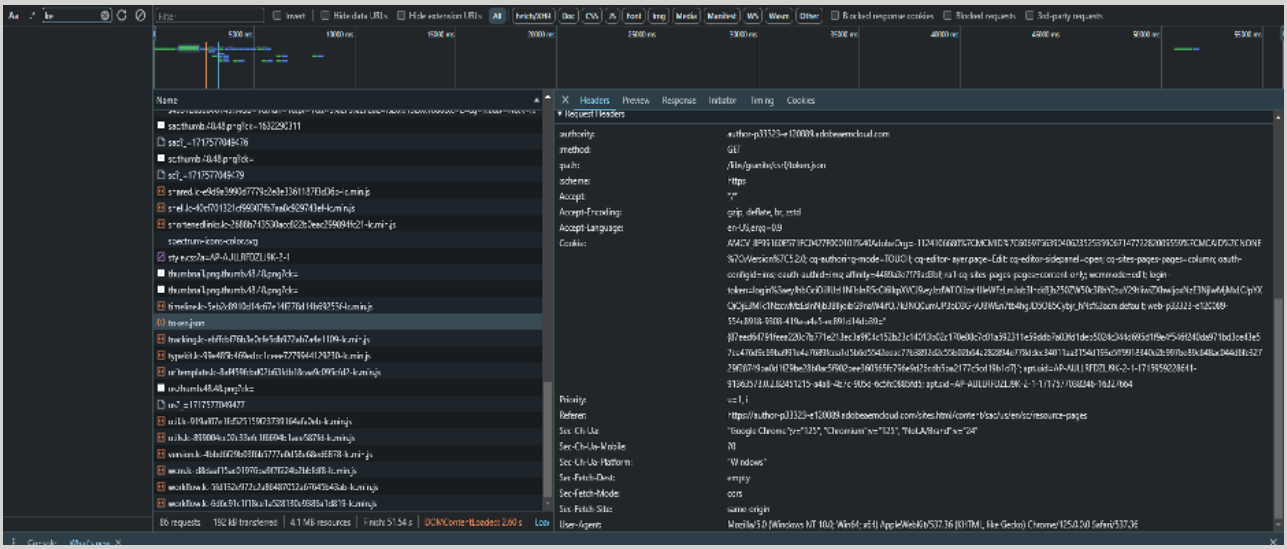
The exported data should be in the same folder format as it is shown on the AEM site.

If the user does not have access to export the data from the AEM, they can use this GitHub link to export the data from the AEM.

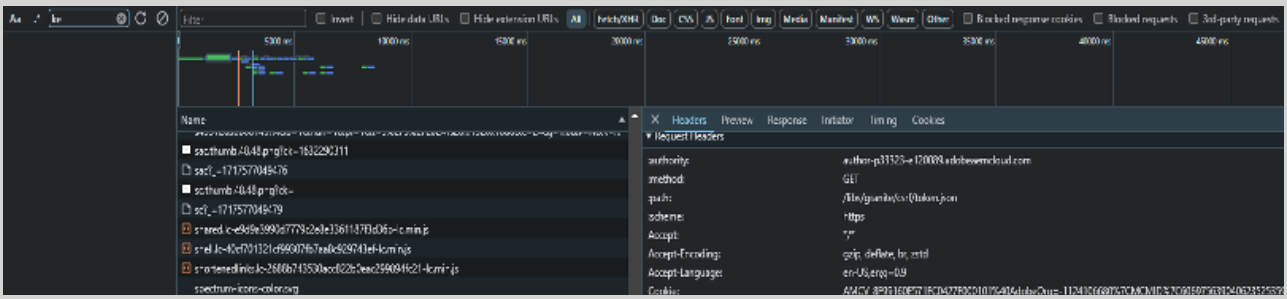
https://github.com/contentstack-expert-services/aem-extraction/tree/feature/shreyas

run npm i.

Create the env file. It needs authority and cookies. This can be found in the network tab. Open any request. Go to the headers tab. Copy authority and header value.

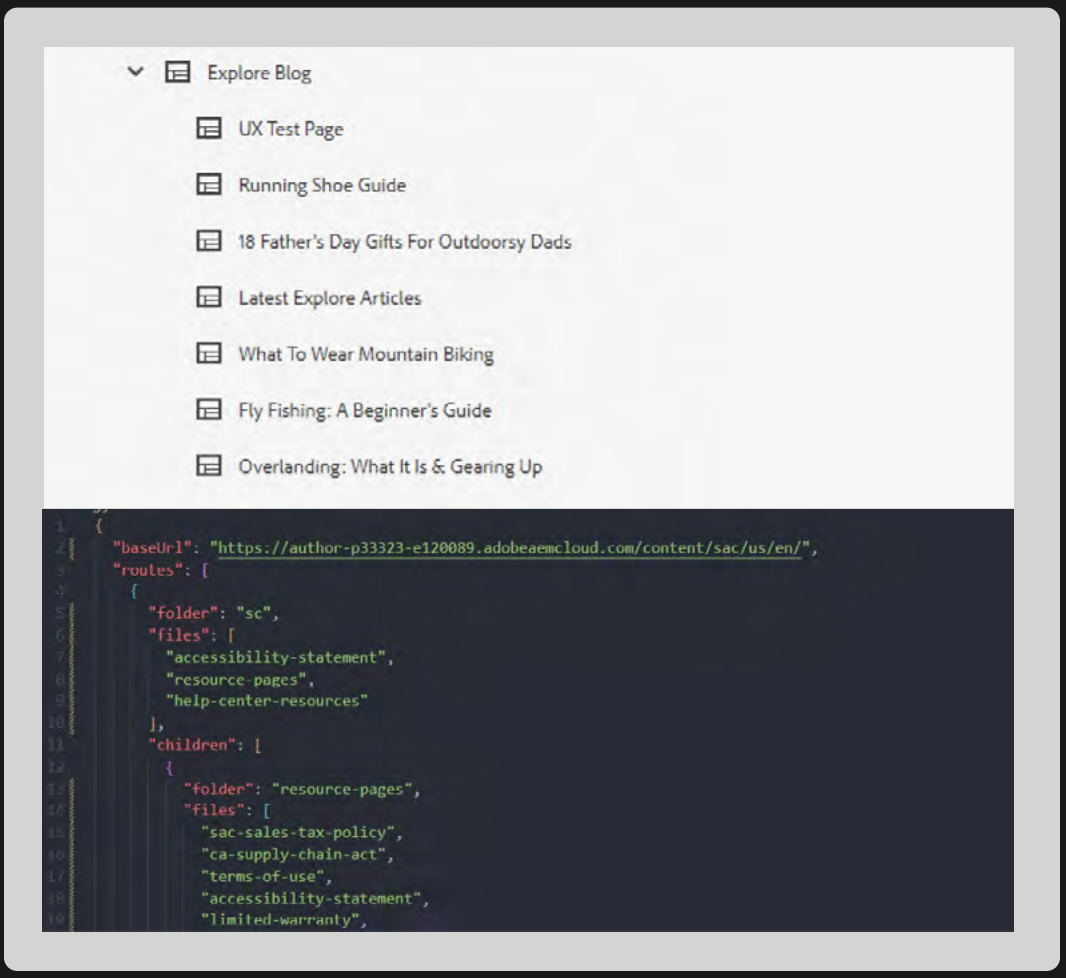


.env file looks like this:



CONTENTSTACK

5. Create the configuration file. It contains the folder structure and base URL.



1. Run node.index.js
2. The user should mention the "**.infinity.json**" format to export the JSON from the AEM. This code supports only the "**.infinity.json**" format.
3. More users can follow the readme file to export the data from AEM.
4. After running the above code, the user has to save the downloaded file in the same way as shown on the AEM site.
5. Users can refer to this Migration-AEM code to convert their data into Contentstack support format.
6. Git repo link to convert data from the AEM site to CS support format: https://github.com/contentstack-expert-services/migration-aem
7. After cloning the repo, the user has to do the following steps to execute:
8. Run npm install to install the required node modules.
9. Open the assets.js file from the lib folder and add the prefix to download all the assets.
10. This will download all the assets from the AEM folder and save them in the **csMigrationData** folder, where the user can see the Assets folder. In the Assets folder, all the downloaded assets will be visible.

Framework configuration and execution

A migration framework for a specific CMS can be downloaded as an NPM package on your local machine or server instance.

Global setup:

You can follow this link for setup or follow the below steps to install Contentstack CLI on your server or system and perform the migration.

Copy **npm install -g @contentstack/cli**

Open a terminal and enter the following line:

```
npm install -g @contentstack/cli
```

Region selection: Select the region and cloud infrastructure on Contentstack where the migration has to be performed. The user running the migration should be provisioned for this region.

```
csdx config:set:region: <set region>
```

a. NA
b. EU
c. AZURE-NA
d. AZURE-EU
e. GCP-NA

Log in with your Contentstack credentials.

```
csdx auth:login -u username@email.com -p ******
```

Now, we have to import the data converted from AEM to CS. So, we have to run the CLI import command to import the data.

```
csdx cm:stacks:import -k <stack_API_Key> -d <Folder_Path>

Eg:- csdx cm:stacks:import -k blt0123456789 -d "/user/admin/aemMigrationData"
```

Frequently encountered issues and resolutions

Observation 1:
If the user is using an existing stack

Solution 1:
Stacks master-locale should be ‘en-us’

Observation 2:
To download assets directly from the AEM DAM

Solution 2:
One can easily download assets from DAM using Axios if one has proper credentials. In the Axios parameter, the user has to send an auth, which will contain the username and password along with the URL to fetch
the original assets, as shown in the picture.

The URLs provided in the API JSON and AEM DAM are different. The API contains the URLs of mobile and desktop versions, while the URL contained in the AEM DAM is the original version.

```
const response = await axios({
  method: 'get',
  url: url,
  responseType: 'stream',
  auth: {
    username: username,
    password: password,
  },
});
```

CONTENTSTACK

Contentstack is in the business of real-time digital experiences, equipping the world's most iconic brands with the technology to connect authentically with their customers like never before. Companies including Walmart, Mattel and MongoDB rely on our platform to cut through the noise across any channel with personalized, data-driven experiences that drive the bottom line.

Request a demo