

Backend Engineer - Product Importer - Advanced

Difficulty level: High

Skills tested: Ability to handle long-running processes, use python, use ORMs

Objective

Acme Inc needs to be able to import products from a CSV file and into their SQL database. Given there are half a million products to be imported into the database, ACME Inc wants you to come up with a scalable solution keeping in mind security, resources, and response times.

Specification

STORY 1:

As a user of the application, I should be able to upload a large CSV file of 500K products ([see here](#)). If there are existing duplicates, it should overwrite the data. Deduplication can be done using the SKU of the product. SKU is case insensitive. Though not in the CSV file, some products should be active and others should be inactive. The SKU is expected to be unique.

STORY 2:

Since the application is processing such a large amount of data the users are also expected to see the progress of the upload. It can be as simple as displaying a progress bar on the UI. Design a backend solution that will address this problem.

Toolkit

Acme Inc. is also opinionated about its tech stack. The tools should be:

1. Web framework: Python-based frameworks - flask, tornado, bottle, django
2. Asynchronous execution: Celery/Dramatiq with RabbitMQ/Redis
3. ORM: SQLAlchemy (if not django).
4. Database: We recommend Postgres (and works with the deployment choice below).

Important Note

The application would be containerized and deployed on stateless servers that have a strict request timeout of 30 seconds. Please keep this in mind while coming up with a solution.

Next Steps

- Prepare a [Loom video](#) describing the architecture of the application.
- Please be very detail-oriented when describing your solutions. The “why(s)” should be answered before the “how(s)”