

# Advance Image Downloader/Extractor

## Architecture Design Document

Written By	Shreyas Parab, Harshad Kadam
Revision Number	<b>1.0</b>
Last date of revision	<b>31-08-2021</b>

## Document Version Control

Date Issued	Version	Description	Author
31-08-2021	1.0	Initial Architecture Design Document – V1.0	Shreyas Parab

## Contents

Document Version Control .....	2
1 Introduction .....	4
1.1 Why this Architecture Design Document? .....	4
2 Architecture .....	5
2.1 Architecture Description .....	5
2.1.1 User Input .....	5
2.1.2 Validating Input .....	5
2.1.3 Job Scheduling .....	5
2.1.4 Web Scrapping .....	6
2.1.5 Database Insertion .....	6
2.1.6 Email Notification .....	6
2.1.7 Files Creation .....	6
2.1.8 Download Link .....	6
2.1.9 Deleting Records .....	6

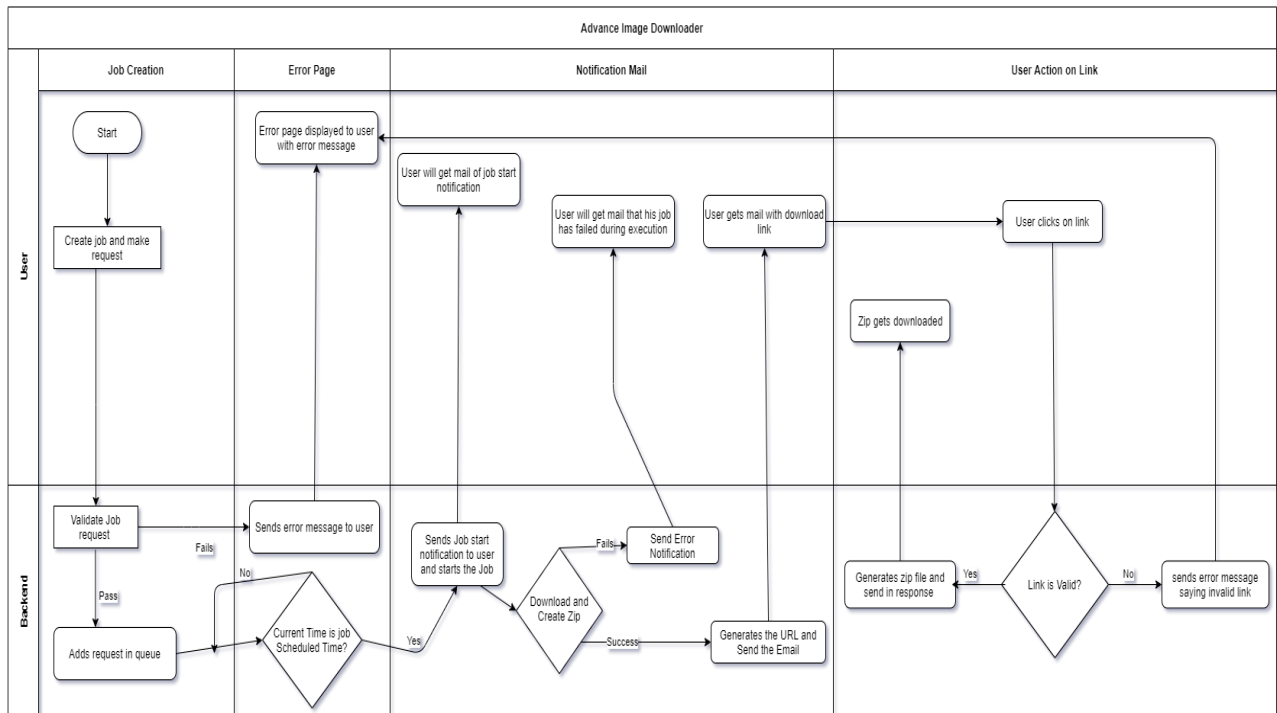
# 1. Introduction

## 1.1 Why this Architecture Design Document?

The purpose of this document is to provide a detailed architecture design of the Advance Image Downloader/Extractor by focusing on four key quality attributes: usability, availability, maintainability, testability.

This document will address the background for this project, and the architecturally significant function requirements. The intension of this document is to help the development team to determine how the system will be structured at the highest level. Finally, the project coach can use this document to validate that the development team is meeting the agreed-upon requirements during the evaluation of the team's efforts.

## 2. Architecture



### 2.1 Architecture Description

#### 2.1.1 User Input

In the user input process, user will be treated with the html page for inserting its query, which will contain inputs like Search Query term, Date, Time and Number of images which the user wants to download. Once this is filled in, user can click on submit button to submit the query.

#### 2.1.2 Validating Input

Validation is done both at the front end and backend level. Email address and Date validation is carried out at the frontend. For handling the scenarios where the user inserted the past date and time, we are treating such situations at the backend site by validating the date-time provided by the user with the current date-time.

#### 2.1.3 Job Scheduling

If the requested job is validated for both frontend and backend validation, then the job is queued to be scheduled. Once the current date-time is equal to the posted job date-time then the process will start its execution.

#### **2.1.4 Web Scrapping**

For getting the images data, web scraping is carried out using selenium. Chrome drivers are responsible for getting the data from Google's search engine and then storing that image data into the databases.

#### **2.1.5 Database Insertion**

- a. Database creation and connection – Create the database with the keyspace name passed (If it's not already present). Connect with the database. Create the table in that keyspace (If it's not already present).
- b. Database Insertion – Selenium will fetch the data from the internet and the links of the images will then be stored into the Database for further processing.

#### **2.1.6 Email Notification**

Users can get an email notification a maximum of three times

- a. Job Start – Once the current time is equal to the scheduled time, the user will get a notification about the job starts.
- b. Job End – Once the file gets downloaded from the internet user will get the download link for downloading the images.
- c. Job Error – If some error occurs at the backend side, the user will get notified by this by sending him an email notification about the job failure.

#### **2.1.7 Files Creation**

Images files will get downloaded and store on the server-side. Since to reduce the size of the images, zip functionality has been added.

#### **2.1.8 Download Link**

Once the zip file is created, then a download link will be sent to the user. This link will be active for a particular interval of time (30 minutes). Users can download within this time interval. After crossing the time limit, the user will get a 'link has expired' error from the frontend.

#### **2.1.9 Deleting Records**

Once the time interval ends, the files which are created at the backend and the Database records will get deleted from the system upon which if the user tries to click on the same link. users will get 'link has expired' error.