Solution Design Document

UiPath

TABLE OF CONTENTS

I. PURPOSE

The Master Project for IMDb Movie Rating Automation and Data Extraction involves developing an automated solution to rate movies and extract detailed information from IMDb. This project will ensure robust performance, scalability, efficiency, and reusability of components while adhering to business restrictions and considerations for future growth.

1. Major Components:

Initialization:

- Description: Sets up the environment and prepares for processing.
- Components:
    - Get Assets: Retrieves paths for input and output Excel files.
    - Initialize Data Tables: Creates data structures for holding movie information and results.
    - Prepare for Web Interaction: Ensures browser is configured and ready for IMDb interactions.

Sign In:

- Description: Handles user authentication on IMDb.
- Components:
    - Get Credentials: Retrieves IMDb user credentials from a secure store.
    - Perform Sign-In: Automates the login process using these credentials.

Movie Rating:

- Description: Searches for movies on IMDb and applies the specified rating.
- Components:
    - Search for Movie: Uses movie name to find the correct IMDb page.
    - Apply Rating: Checks if a rating exists; if so, updates it, otherwise applies a new rating.
- Exception Handling: Manages cases where movies are not found or rating inputs are invalid.

Data Extraction:

- Description: Extracts detailed information about movies and stores it in a structured format.

- Components:
    - Extract Movie Details: Gathers description, directors, writers, and stars.
    - Populate Data Table: Adds extracted data to the structured output table.

Output Handling:

- Description: Manages the results and communicates them.
- Components:
    - Save Output to Excel: Writes the collected movie data to an Excel file.
    - Send Notification: Sends an email with the output file attached.

Close Applications:

- Description: Cleans up resources and ensures no leftover processes.
- Components:
    - Close Browser Tabs: Closes any open IMDb tabs in the browser.
    - Finalize Output: Ensures the Excel file is properly saved and closed.

2. Key Focus Areas:

Robustness:

- Design: Implement error handling and retries to manage transient issues.
- Testing: Extensive testing to handle edge cases and unexpected scenarios.
- Monitoring: Set up logging and alerts to detect and address issues promptly.

Scalability:

- Architecture: Design workflows to handle increased volume, such as batching and parallel processing.
- Configuration: Use parameters and configuration files to adapt to different input sizes and requirements.
- Resource Management: Optimize resource usage to handle peak loads efficiently.

Efficiency:

- Optimization: Streamline workflows to minimize execution time and resource consumption.
- Performance Tuning: Regularly assess and tune the performance of automation components.
- Batch Processing: Use efficient data handling techniques, such as processing movies in batches.

Replicability:

- Modular Design: Create reusable components and workflows that can be easily duplicated for other projects.
- Documentation: Provide detailed documentation on workflow design and configuration for ease of replication.

Reusability of Components:

- Component Library: Develop a library of reusable activities and workflows for common tasks (e.g., login, data extraction).
- Version Control: Use version control to manage changes and updates to reusable components.
- Best Practices: Follow best practices in workflow design to ensure components are easily adaptable.

3. Considerations for Business Restrictions:

- Scheduling: Ensure automation runs during off-peak hours to avoid impacting system performance.
- Peaks: Implement load management strategies to handle spikes in data volume or system load.
- Future Increases in Volume: Design workflows to be flexible and adaptable to future growth, including potential integration with other data sources or systems.

II. AUTOMATED PROCESS DETAILS

Details filled in need to reflect the actual information for the Master Project released for production. The following table will be populated:

Item	Description
Master Project Name	IMDb Movie Rating Automation and Data Extraction
Robot Type	FOR
Orchestrator used?	Yes
Scalable	Yes
UiPath version used	2024.10.1

2 RUNTIME GUIDE

2.1 Architectural structure of the Master Project

Display the interaction between components (package / robots, Orchestrator queues, and running order) in a diagram

2.2 Master Project Runtime Details

Outlines the details of the automated process by filling in the table below.

ITEM NAME	DESCRIPTION
Production environment details	Running on the virtual machine. Scheduled to execute nightly after the input Excel files are updated.
Prerequisites to run	- IMDb credentials must be available in Orchestrator Assets. - Input Excel files must be available in the specified directory. - Web browser (Chrome) must be installed.
Input Data	- 1 Excel file containing a list of movies and their ratings. - Path to this input file is retrieved from Orchestrator Assets.
Expected output	- 1 Excel file containing updated movie data and ratings. - 1 email sent with the output Excel file attached.
How to start the automated process	The process will be started from orchestrator server (demo.uipath.com)
Reporting (queues reporting, Kibana or another platform)	- Orchestrator logs and jobs dashboards will be used for monitoring and reporting the status of the process execution.
How is Orchestrator used?	Orchestrator is used for: - Scheduling the process. - Managing and storing asset passwords. - Handling credentials securely.
Password policies (mention any specific compliance requests)	Credentials for IMDb are stored securely in Orchestrator Assets with compliance to security best practices (e.g., strong passwords, regular updates).
Stored credentials (Never use hardcoded credentials in the workflow!)	Stored in Orchestrator Assets
List of queues names (Naming convention: ProcessName_QueueName)	n/a
Schedule Details	n/a
Multiple Resolutions Supported? (in case of image automation / Citrix and VDI)	n/a
Recommended Resolution	n/a

2.3 IMDb Movie Rating Automation and Data Extraction

ITEM NAME	DESCRIPTION
Environment used for development (name, location, configuration details etc)	*DEV_Env1_EMEA - UiPath development environment located in the EMEA region. This environment is used for designing, testing, and validating the automation workflows.*
Environment prerequisites (OS details, libraries, required apps)	*- Operating System: Windows 10 or Windows Server 2019.* *- UiPath Studio: Latest version with valid license.* *- Microsoft Excel: Installed with support for the required file formats.*
Repository for project (where is the developed project stored)	*https://github.com/beatricecroitoriu/RPA-IMDb-Movie-Rating-Automation-and-Data-Extraction*
Configuration method (assets, excel file, Json file)	*Configuration settings, such as paths to input/output files and IMDb credentials, are managed through UiPath Orchestrator Assets.*
List of reused components	*n/a*
List of new reusable components	*n/a*

2.4 Project(s) workflows

Workflows specific to: IMDb Movie Rating Automation and Data Extraction

Workflow Name	Description
Main	*Orchestrates the execution of all other workflows. Manages the overall process flow and error handling.*
Init_Applications	*Initializes the environment, retrieves asset values, sets up data tables, and prepares for processing.*
Sign_In	*Handles user authentication on IMDb.*
Rating_Movies	*Searches IMDb for each movie, applies ratings, and handles exceptions related to rating.*
Movies_Data_Extraction	*Extracts detailed information about movies from IMDb and updates the output data table.*
Close_Applications	*Closes browser tabs, saves output to an Excel file, and sends an email with the output file attached.*

2.5 Packages

Include the list of packages and high-level description for each of them, to explain their purpose

Package Name	Description
IMDbProject.1.0.0.nupkg	*This package contains the core workflows for automating movie rating and data extraction on IMDb.* *- Init_Applications.xaml: Initializes the environment and prepares data tables.* *- Sign_In.xaml: Manages the sign-in process to IMDb.* *- Rating_Movies.xaml: Applies ratings to movies.* *- Movies_Data_Extraction.xaml: Extracts detailed information about movies.* *- Close_Applications.xaml: Manages closing tabs, saving data, and sending emails.*
UiPath.Excel.Activities.2.10.5.nupkg	*Provides activities for handling Excel files.* *- Read Range: Reads data from an Excel sheet into a DataTable.* *- Write Range: Writes data from a DataTable into an Excel sheet.* *- AutoFit Columns: Adjusts column widths to fit the content automatically.*
UiPath.WebAPI.Activities.1.6.0.nupkg	*Contains activities for making HTTP requests and processing responses.* *- HTTP Request: Used to interact with IMDb's web API for searching and retrieving movie data.*
UiPath.UIAutomation.Activities.21.10.1.nupkg	*Includes activities for automating interactions with user interfaces.* *- Click: Clicks on UI elements like buttons and links.* *- Type Into: Types text into input fields.* *- Element Exists: Checks if a UI element is present.*

UiPath.Mail.Activities.1.13.2.nupkg	*Provides activities for sending and receiving emails.* *- Send Outlook Mail Message: Sends an email with the output Excel file attached.* *- Get Outlook Mail Messages: Retrieves emails if necessary for additional processing.*
UiPath.Core.Activities.21.10.1.nupkg	*Contains core activities used across UiPath workflows.* *- Assign: Assigns values to variables.* *- If: Performs conditional logic.* *- For Each: Iterates over collections.*
UiPath.System.Activities.21.10.1.nupkg	*Provides system-level activities.* *- Kill Process: Terminates processes like web browsers to clean up.* *- Log Message: Logs messages for debugging and tracking purposes.*

2.6 Architectural structure of the Master Project

Display the interaction between components (package / robots, Orchestrator queues, and running order) in a diagram.

3 OTHER DETAILS

Future Improvements

Fill in any improvements that need to be considered for the future:

Example:

- Optimize the processing algorithm
- Implement process error recovery (retry)
- Enable support for multiple template files

Other Remarks

Please mention here any other points that you consider relevant for the automation process.

Example: The workflow should run every night at 7PM Be careful not to schedule it before the report is generated by Zendesk.

The Zendesk generated data is always 1 day old.

4 GLOSSARY

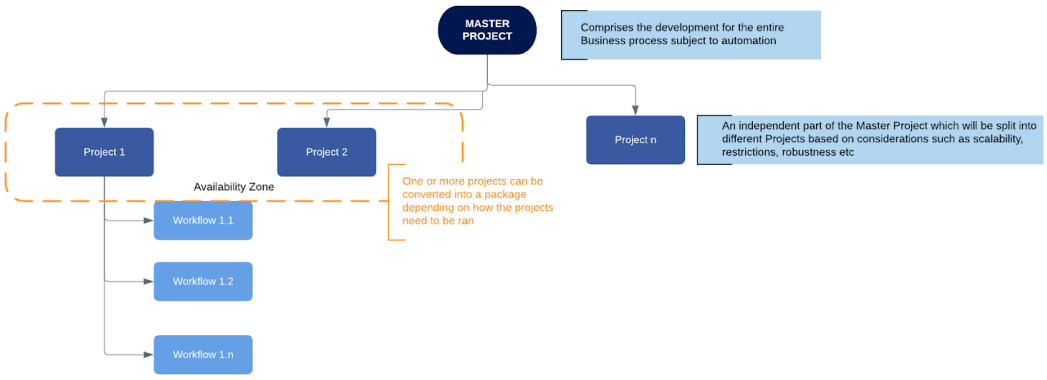
The main terms used in the Solution Architecture Document are defined below:

Master project - the overall output of the development, containing one or multiple projects that together cover the scope of the robotic process automation. There is a 1 to 1 connection between the Master Project and the Process to be automated (As presented in the PDD).

Project - an UiPath Studio project containing one or multiple workflow files. A project can be converted to a package and run independently, covering a particular scope within the master project. Or multiple projects can be converted into one package depending on the aims and restrictions of the automation. The project is used when defining the development and support phase of the automation.

Package - the output of compiling one or multiple projects. A package can be deployed on the robot machine and be executed by the robot service. Only one package can be executed at a given time by a robot. The package is used when defining the running phase of the automation.

Workflow - a component of the package, the workflow encapsulates a part of the project logic. The workflow can be of type: sequence, flowchart or state machine. A workflow is saved as an .xaml file inside the project folder. A workflow file can be invoked from another workflow and by default there is an initial workflow file that will run when executing the package.



Activity - an action that the robot executes.

Sequence - a workflow where activities are executed one after another, in a sequential order

Flowchart - a workflow where activities are connected by arrows and the logic of the workflow can be easily followed in a visual manner. The flowchart can also be exported as an image from UiPath studio.

State machine - a more advanced way of organizing a workflow, similar to a flowchart.

BOR - Back office robot

FOR – Front office robot

Orchestrator – Enterprise architecture server platform supporting: release management, centralized logging, reporting, auditing and monitoring tools, remote control, centralized scheduling, queue/robot workload management, assets management.