

Solution Overview Document – Web Scraper Bot

IARE Unofficial Bot

Sathwika Valija

DATE: 05-12-2025

VERSION: 1.0

SOLUTION OVERVIEW DOCUMENT(SOD)

1.0 Project details:

- **Project Title:** Web Scraper Bot for IARE CMS
- **Project Name:** IARE Unofficial BOT
- **Domain:** Educational Automation
- **Prepared By:** V. Sathwika
- **Date:** 05-12-2025
- **Version:** 1.0
- **Objective:** The system scrapes attendance and biometric data from Samvidha CMS, compresses reports, and provides secure, on-demand access without storing student report data. It also features an adjustable UI designed to match student preferences.

1.1 Introduction

- **Purpose of the Document**
 - To define the project needs and requirements for the IARE CMS Web Scraper Bot.
- **Project Background**
 - IARE students face delays and complexity in accessing biometric/lab attendance data. The bot will automate scraping and enhance data visibility.
- **Goals**
 - Improve accessibility, reduce manual effort, and automate report uploads securely.

1.2 Project Objectives

The objective of this project is to develop a Web Scraper Bot that automates the extraction, processing, and reporting of student biometric data, lab attendance, and academic related documents from the IARE Content Management System (CMS). This solution is aimed at addressing the inefficiencies in manual data retrieval by providing a fast, reliable, and user-friendly platform that can be operated by maintainers and admins.

1.3 Problem Statement

A secure and efficient system can enhance functionality by safeguarding user credentials, accurately calculating biometric percentages to help students manage attendance effectively, and automating report downloads and uploads to save time and reduce manual effort.

1.4 Proposed Solution

A Python-based bot using tools like Selenium, Pyrogram, and BeautifulSoup for:

- Biometric/attendance scraping.
- Compression of PDF ($10\text{MB} \leq 1\text{MB}$).

- Telegram bot-based access interface for user interaction.
- PostgreSQL storage for analytics and structured data access.

1.5 Project Background

- IARE students face delays and complexity in accessing biometric/lab attendance data.
The bot will automate scraping and enhance data visibility.

1.6 Project Goals

- Improve accessibility, reduce manual effort, and automate report uploads security.

1.7 Project Objectives

- Streamline attendance and lab upload tasks.
- Enhance visibility into biometric logs.
- Automate PDF compression and file handling.
- Improve performance tracking.

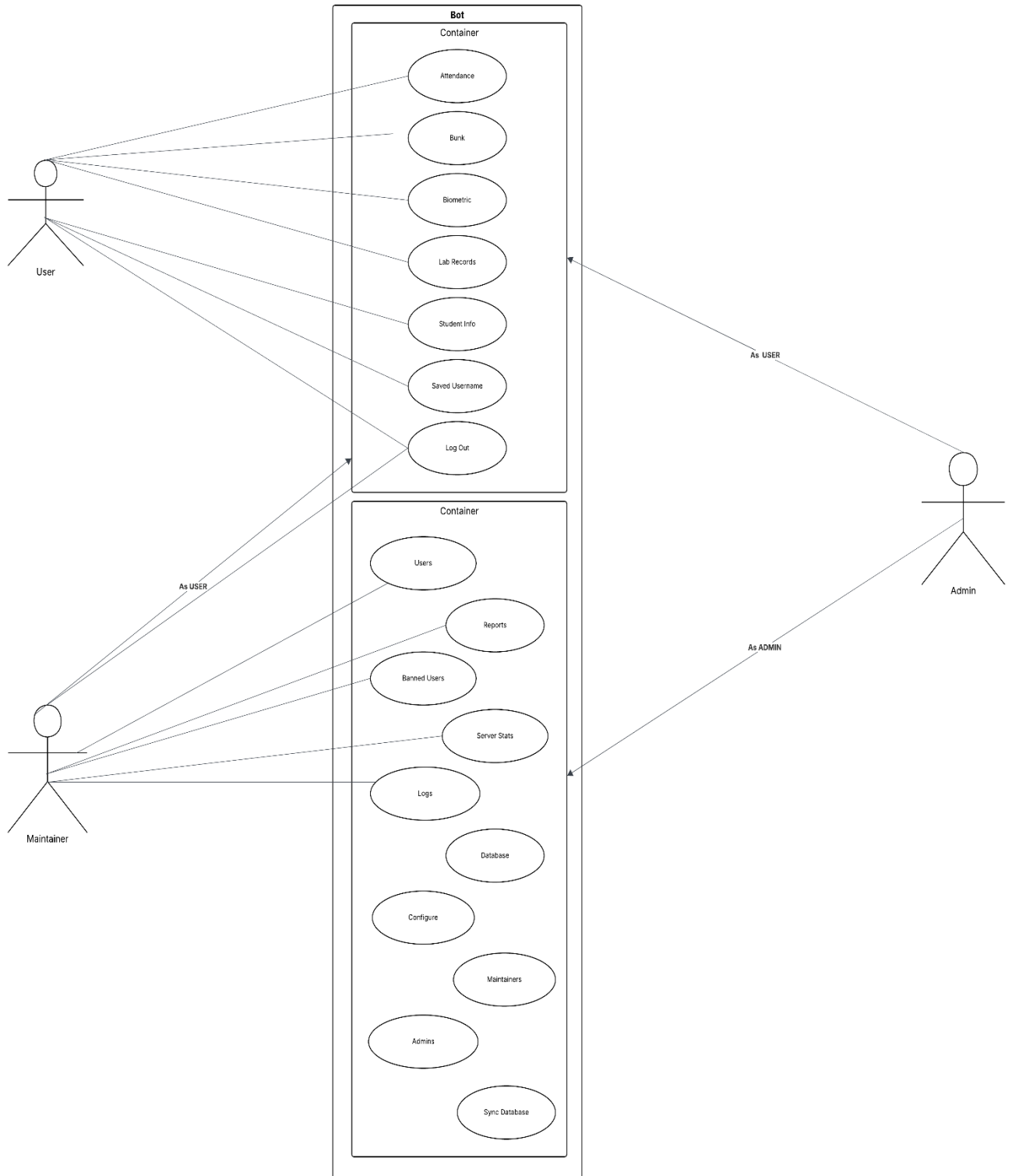
1.8 Key Deliverables

ID	Requirement	Priority	Description
KD-01	Auto Login and Secure Biometric scraping	High	The system must automate login and retrieval of biometric attendance data from IARE CMS.
KD-02	PDF Compression	Medium	The system should support compressing of PDF files to optimize upload speed and the CMS portal only allows uploading PDFs of $\leq 1\text{MB}$.
KD-03	Feedback Extraction	High	The bot must extract student feedback from CMS and generate structured reports.
KD-05	Admin and Maintainer Features	High	Admins and maintainers must have access to extended functionalities like scraping control and bot monitoring.

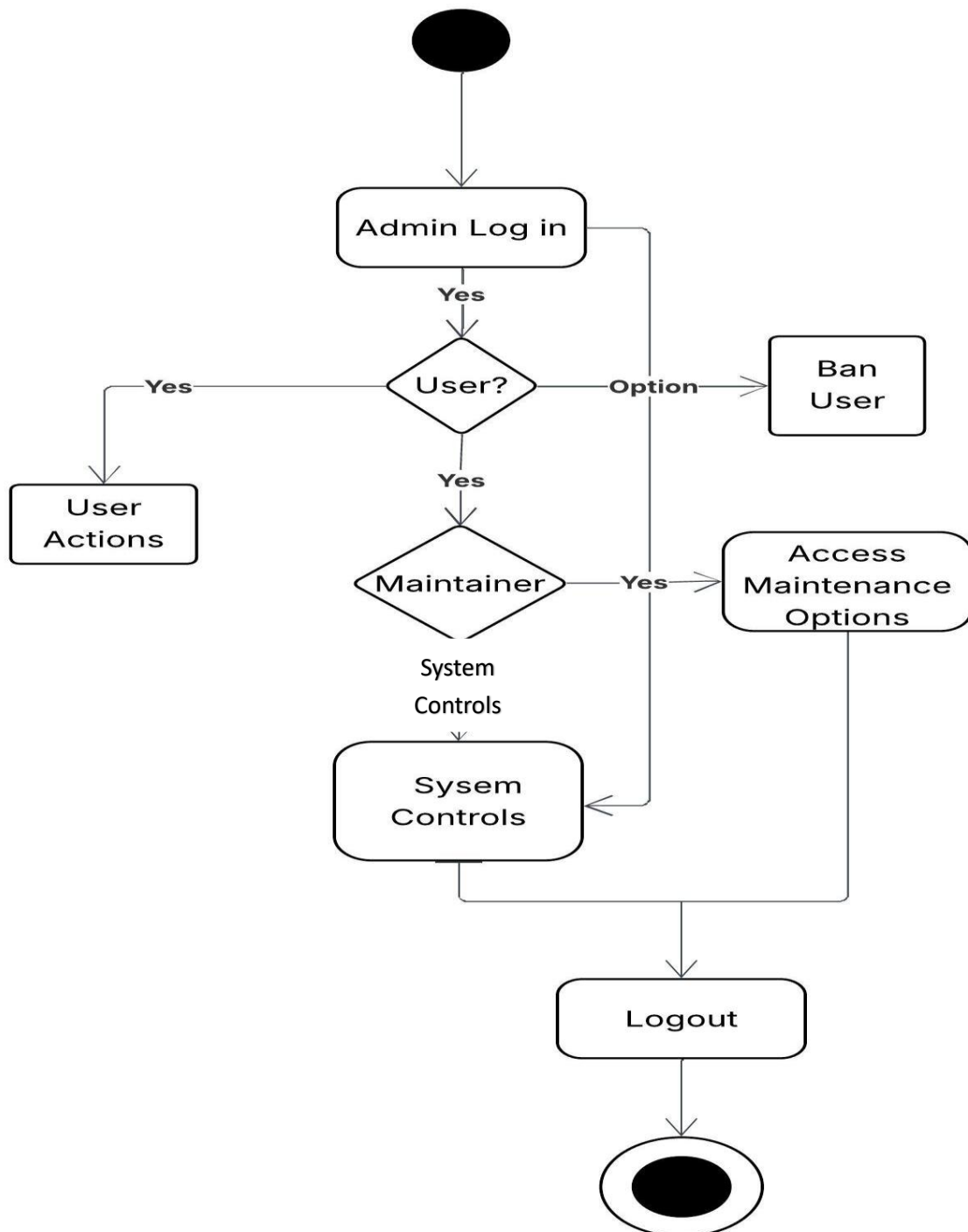
KD-06	Role-based Login	High	The system must implement role-based login and secure session handling for Admins, and Maintainers.
KD-07	Role-based interface	Medium	An interface should display summaries and activity logs tailored to user roles.
KD-08	Fast Execution Time	High	The bot must perform scraping and data generation within 10 seconds on normal load.
KD-09	Data Security Compliance	High	All data must be handled securely and in compliance with institutional data privacy regulations.

1.9 Diagrams

Use Case Diagram



Activity Diagram



1.10 Stakeholders

Stakeholder Type	Names	Role/Responsibility
Project Stakeholders	Mr. K. Keerthi Sathvik (Initial Developer, Admin /Project Manager/Business Analyst) Ms. V. Sathwika (Business Analyst)	Oversees requirements, approves implementation, Manages the project and coordinates tasks and gathers Requirements and oversaw documentation.
Technical Stakeholders	Mr. A. Varun Reddy, Mr. V. Yuvraj (Developers)	Developers Implemented specific scraping and backend functionalities
Functional Stakeholders	End Users (Students)	Interact with the bot to access attendance data

1.11 Assumptions and Constraints

- Assumes Samvidha CMS remains structurally stable.
- Only authorized users have access.
- Bot hosted on Heroku (or similar platform).

1.12 Scope and Limitations

- **In-scope:** Biometric scraping, lab upload, feedback system.
- **Out-of-scope:** Academic evaluation, payment systems, user profile management.

1.13 Conclusion

The IARE CMS Web Scraper Bot is designed to address real-time challenges faced by students in accessing attendance, biometric data, and uploading lab records. By incorporating secure login, data scraping, and responsive features like feedback and compression tools, the bot aligns with institutional needs. The SOD defines a clear roadmap to guide the development process and ensure the solution delivers measurable value.