

Project Overview: College Attendance Tracking System

Idea Behind the Project

This project aims to automate and digitize the college attendance process by integrating a Java-based real-time API system with a PySpark-based analytics engine. The goal is to ensure accurate, transparent, and time-based attendance tracking for every student and subject.

System Components & Workflow

1. Java API Layer – Real-Time Attendance Logging

- Punch-In Mechanism:

Each classroom is equipped with a punch-in device. When a student enters the class, they swipe their ID card.

- API Trigger:

The moment a student punches in, a Java REST API is triggered.

- Validation Checks:

- Whether the student is registered in the system
- Whether the student is enrolled in the subject scheduled at that time

- Data Recording:

If both conditions are satisfied, the punch timestamp is saved to the attendanceRecord table.

📍 At this point, Java's responsibility ends.

2. PySpark Layer – Duration Check & Monthly Reporting

- Duration Analysis:

PySpark reads the raw punch data and calculates the total time a student remained in each class.

- Attendance Rule:

If the duration is greater than 40 minutes, the student is marked as Present; otherwise, Absent.

- Window Functions:

PySpark uses window functions to find the first and last punch time per student-subject-day and compute the time spent.

- CSV Report Generation:

📁 One CSV per student per subject per month

Report contains:

- Month & Year
- Number of Days Present / Absent
- Attendance %
- Dates Absent

Technology Stack

Component	Technology
API Layer	Java + Spring Boot
Database	PostgreSQL / MySQL
Analytics Engine	PySpark
Reporting Output	Monthly CSV Files

Future Enhancements

- 📲 SMS/Email alerts for low attendance
- 🔑 Integration with RFID/Biometric systems
- 📩 Auto-email reports to faculty/students