

ETL & BPMN Workflow Report

© 2023 PUCIT - All Rights Reserved

PUCIT Centralized ETL Workflow

BPMN + SSIS Style Pipeline Documentation

BS Data Science - Fall 2023

Lab 09 - 27 November 2025

Instructor: Dr. Khurram Shahzad

Teacher Assistants: Ali Hasan, Shahzeb Ali, Amna Yaseen

ETL & BPMN Workflow Report

© 2024 Project Alpha, Inc. All rights reserved.

TABLE OF CONTENTS

1. Problem Overview	3
2. ETL Requirements	4
3. BPMN Workflow Model	5
4. Final Data Warehouse Targets	6
5. Executive Summary	7

ETL & BPMN Workflow Report

© PUCIT 2023. All rights reserved.

1. Problem Overview

PUCIT maintains student, faculty, and attendance data across two independent campus servers (Old Campus and New Campus), each storing similar data but in different file formats. The administration requires a centralized Data Warehouse to support unified reporting, analytics, and campus-level comparisons.

Data Sources

- Server A (Old Campus): Students.xlsx, Faculty.db, Attendance.csv
- Server B (New Campus): Students.csv, Faculty.json, Attendance.xlsx

2. ETL Requirements

The ETL pipeline must extract data from both servers, standardize formats, clean inconsistent values, merge student and faculty entities, unify attendance logs, and load everything into the PUCIT Centralized Data Warehouse.

Primary Goals

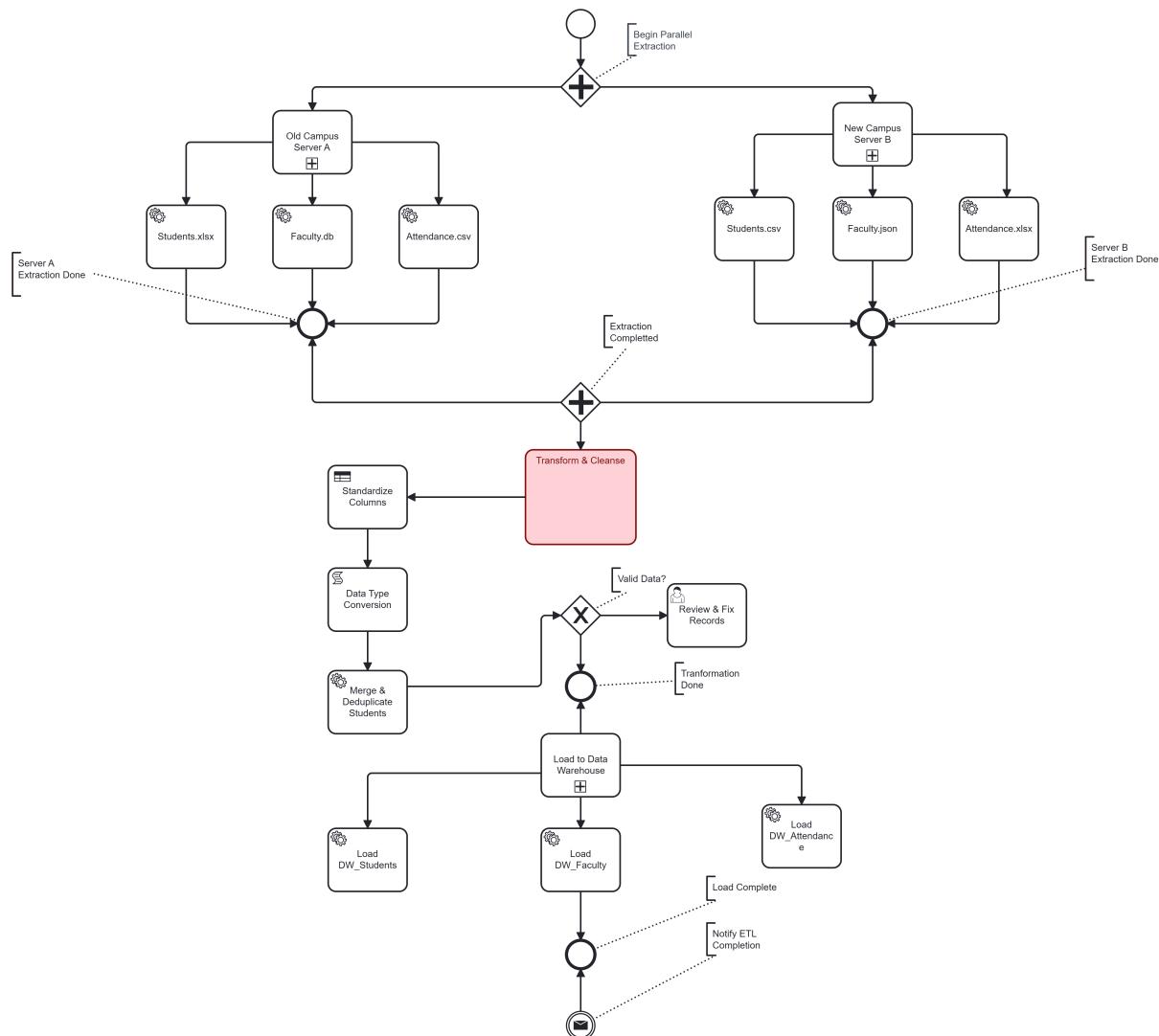
- Combine student records from Excel + CSV
- Merge faculty records from SQLite + JSON
- Unify attendance logs from CSV + Excel
- Apply transformations and standardization
- Load DW_Students, DW_Faculty, DW_Attendance

3. BPMN Workflow Model

The diagram below represents the complete ETL workflow modeled using BPMN.

Figure 1: Complete BPMN ETL Workflow

ETL & BPMN Workflow Report



4. Final Data Warehouse Targets

- DW_Students (StudentID, Name, Campus, Degree, Standardized Info)
- DW_Faculty (FacultyID, Name, Campus, Department, Standardized Info)
- DW_Attendance (StudentID, Date, Status, Campus, Cleaned Fields)

ETL & BPMN Workflow Report

© PUCIT Academic Data Pipeline 2025

5. Executive Summary

This ETL pipeline successfully consolidates all PUCIT academic data into a unified Data Warehouse. Using BPMN modeling principles, the workflow clearly defines extraction, cleaning, merging, and loading steps across two campuses. The resulting warehouse supports reporting for total students, faculty details, attendance summaries, and cross-campus comparisons.