

## AUTHORS

Nicko Gorhau, Charlie Meyer, Tyler Grover,  
Cole Zucosky, Arielle Madeam,  
Brandon Wang

## AFFILIATIONS

Advisor: Filippos Vokolos  
Stakeholder: Courtney Wilgucki

# PROJECT ROLLCALL



## PROBLEM

- **Manual Inefficiencies:** Managing headcount data across the ES&F department currently relies on fragmented data sources (Workday, external contractor systems, and localized Excel files), leading to slow, error-prone manual tracking.
  - If there is no issues in the data, the process can be ran in around 30 minutes
  - If there are issues/errors, it can take as much as 2-3 days to resolve
- **Data Integrity Mismatches:** The current process cannot adapt to changes within the data, usually produced by changes in the Workday data the process pulls from.

## COMPONENTS

- **Ingestion (AWS SES & S3):** Receives Workday and contractor data streams securely via targeted cloud email storage.
- **Queueing (AWS SNS & SQS):** Coordinates ingestion triggers and throttles high-volume message passing to prevent resource exhaustion.
- **Real-time Monitoring & Alerts (Cloudwatch):** Alarms are actively mapped to both Lambda1\_Errors and Lambda2\_Errors to instantly flag execution failures for system administrators.
- **Troubleshooting & Debugging (DLQ):** Isolates and preserves data payloads from any failed events across the services for postmortem analysis.
- **Processing Core (AWS Lambda):** Executes containerized Python scripts to clean missing items, standardize department names, and format records.
- **Storage (AWS RDS MySQL):** A hardened relational database serving as the immutable core storage layer.

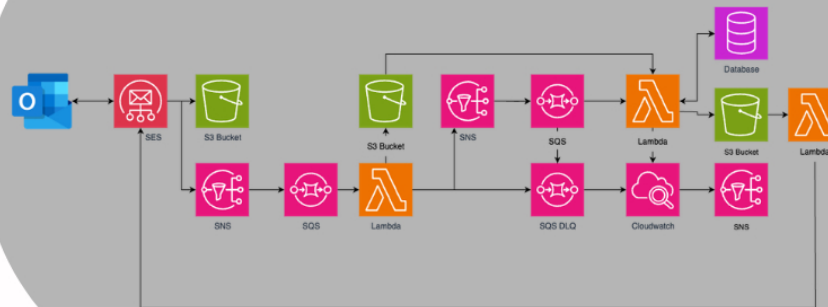
## KEY FEATURES

- **Automated Weekly ETL Pipeline:** Python scripts seamlessly extract, transform, and load workforce data on a strict weekly loop.
- **Centralized Data Repository:** An AWS-hosted database unifies both full-time employees (FTE) and contractor details into a standardized schema.

## SOLUTION

An automated and centralized weekly data-processing pipeline, reducing runtime to ~5 minutes, eliminating manual Excel/macros, and delivering a complete dataset for seamless Tableau dashboard reporting.

## PROCESS DIAGRAM



## TECH STACK

