Data Operations System Project

# Purpose Statement

The purpose of this application is to streamline the management and execution of data analyst workflows by enabling efficient creation of work queues, ensuring seamless navigation and task execution for analysts, and providing actionable insights through dashboard-level reporting. This application will support a variety of data projects such as de-duplication, data completeness, record creation, and record updates. By optimizing workflow management, the application will improve efficiency, enhance accountability, and provide clear visibility into key metrics.

# Value Proposition Statement

This application will immediately address critical inefficiencies in the data analyst workflow by delivering a functional prototype that reduces operational costs, improves throughput, and enhances data accuracy. By centralizing task management, providing real-time visibility into performance, and automating the tracking of key metrics like backlog and throughput, the prototype will enable better decision-making and resource allocation. This foundational phase will establish a scalable framework to drive measurable impact and support future growth initiatives.

# Business Requirements

1. Work Queue Creation
   * Allow the creation of new work queues by uploading spreadsheets in one of five predefined standard formats.
   * Enable future flexibility to support different formats for various data projects.
2. Queue Management
   * Display work queues with clear navigation for analysts.
   * Support various project types, including:
   * De-duplication (identify duplicate records and mark the survivor).
   * Data completeness (update or add missing data).
   * Record creation (new entries based on provided data).
   * Record updating (edit existing records).
3. Task Assignment and Locking
   * Ensure records can only be accessed by one analyst at a time.
   * Automatically unlock records after:
     + Research is completed.
     + 1 hour of inactivity.
     + Enable administrative users to manually unlock records if needed.
4. Task Execution
   * Enable analysts to:
     + Select a record from the queue.
     + Edit or update record details.
   * Take one of the following actions:
     + Update: Confirm the record is updated.
     + Not Found: Mark as unresolvable.
     + Skip: Return to the queue for reassignment.
   * For duplicate records: Identify the duplicate and designate the survivor.
5. Dashboard Reporting
   * Provide actionable insights through a dashboard that tracks:
     + Backlog size.
     + Records added in the past 24 hours.
     + Records completed in the past 24 hours.
     + Average time spent per record.
     + Completion rate by analyst.
     + Average throughput (records/hour) by analyst.
6. User Roles and Permissions
   * Implement user authentication with role-based access:
   * Data Analysts: Access and complete tasks within queues.
   * Administrators: Manage user accounts, assign roles, create work queues, and unlock records.
7. Scalability and Flexibility
   * Support predefined standard formats for spreadsheet uploads initially, with plans for greater flexibility in the future.
   * Allow additional metrics and reporting capabilities to be added as needed.

## Functional Requirements

1. Work Queue Creation
   1. Allow administrators to upload spreadsheets in one of five standard formats to create new work queues.
   2. Validate spreadsheet format upon upload:
      1. Reject invalid formats.
      2. Display an error message specifying the issue.
   3. Optional: Provide a data preview of valid uploads before queue creation.
   4. Enable administrators to append additional data to an existing queue while maintaining data integrity.
2. Navigation and Search
   1. Provide a navigation bar listing all available work queues.
   2. Include search functionality within queues, allowing analysts to search by:
      1. Record ID.
      2. Company name.
   3. Implement a “Get Next” function to retrieve the next available record for the analyst to work.
3. Record Display and Editing
   1. Display all available data for each record to the analyst while they work.
   2. Highlight any missing fields that need to be populated.
   3. Allow analysts to edit or update record details as required.
   4. Enable analysts to take the following actions:
      1. Update: Mark the record as complete with updated data.
      2. Not Found: Mark as unresolvable.
      3. Skip: Return the record to the queue for reassignment.
      4. For duplicate records - Identify which record is the duplicate and which is the survivor.
4. Record Locking and Completion
   1. Lock a record when accessed by an analyst to prevent others from working on it simultaneously.
   2. Automatically unlock records after:
      1. Completion.
      2. 1 hour of inactivity.
   3. Allow administrators to manually unlock records.
   4. Upon completion of a record:
      1. Move it to a “Completed” status.
      2. Remove it from the active queue.
      3. Write it to a downloadable “Completed Records” CSV file within the queue. - Include the date and time the record was completed.
5. Administrative Capabilities
   1. Enable administrators to:
      1. Create new queues.
      2. Edit queue descriptions.
      3. Delete queues.
      4. Assign permissions for queues to specific users.
      5. View dashboards and reports.
   2. Manage user accounts, including:
      1. Creating new users.
      2. Updating user passwords.
   3. Append additional data to an existing queue.
6. Dashboards and Reporting
   1. Provide dashboards with the following metrics:
      1. Backlog size.
      2. Records added in the past 24 hours.
      3. Records completed in the past 24 hours.
      4. Average time spent per record.
      5. Completion rate by analyst.
      6. Average throughput (records/hour) by analyst.
   2. Ensure dashboards and metrics are accessible to administrators.

## Work Queue Templates

Initially we will need to develop 2 work queue templates. When creating a work queue from uploaded data, the user will need to select the appropriate template. The system will also need to ensure the uploaded data can be used with the selected template; if not, an error should be shown.

* Missing Data Template – The purpose of this queue is for the data researcher to view available information and then research to fill in missing data elements. For the purpose of this project, the missing data element will be included as an empty column.
* Duplicates Template – The purpose of this queue is for the data researcher to compare 2 records side by side and determine:
  + Are they duplicates?
  + Which record should remain and which should be removed/dissolved

Additionally, to help you get started, I have also included a very old image I build (literally years ago) of what the table of records in queue should look like. This is what shows up after the user selects the queue they intend to work (in this case the Missing DUNS queue, which would use the Missing Data Template). It doesn’t need to exactly like this; I wanted to include a representative example to help your thought process.

Feel free to bug me with questions anytime.