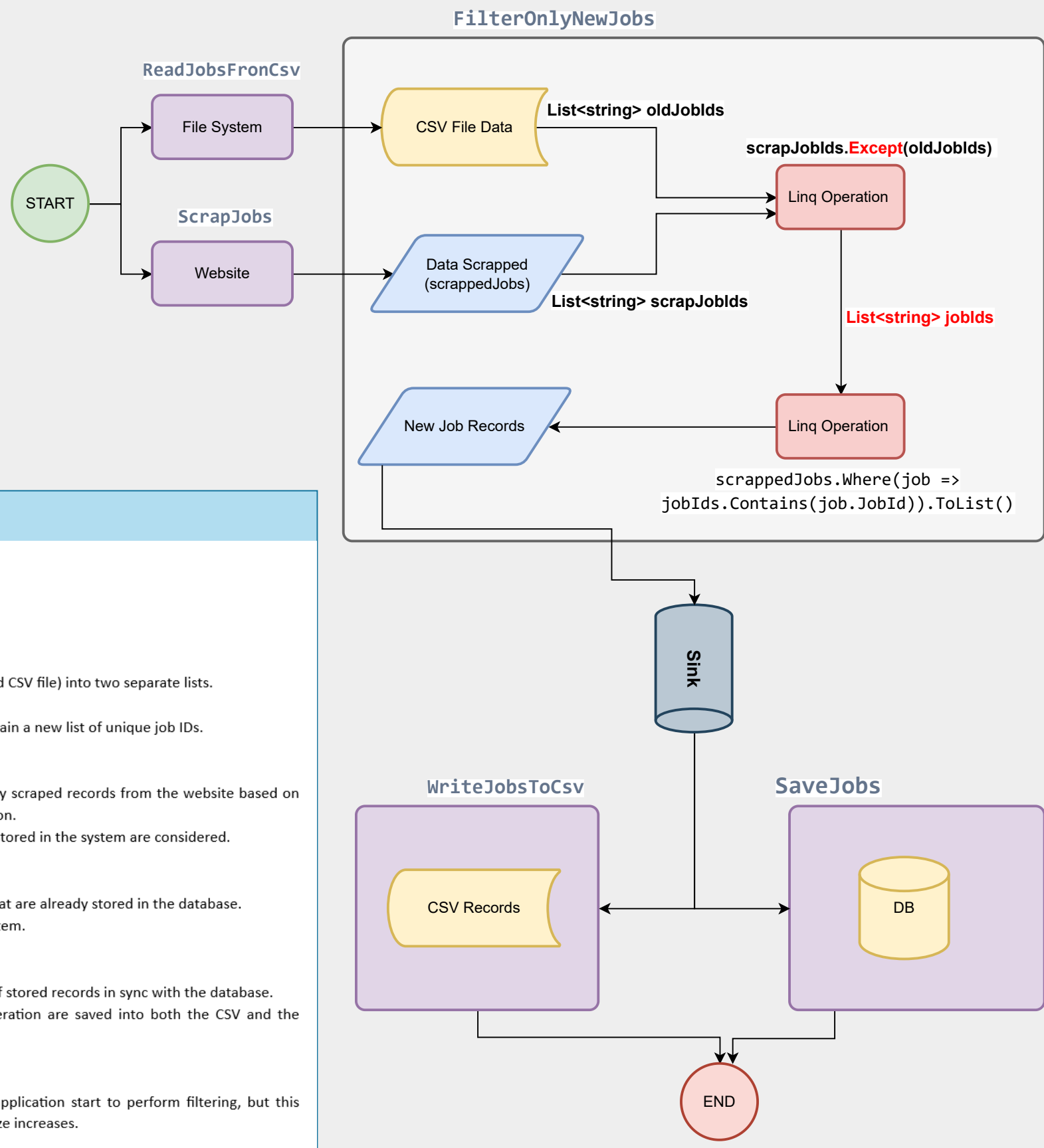


GhanaJob



How It Works

Initial Data Extraction:

- Scrape data from the website.
- Load CSV data from the file system.

Job ID Extraction:

- Extract job IDs from both sources (scraped data and CSV file) into two separate lists.
- First LINQ Operation:
- Use a LINQ operation to process these lists and obtain a new list of unique job IDs.

Retrieval of New Records:

- Perform a second LINQ operation to retrieve newly scraped records from the website based on the list of job IDs obtained in the first LINQ operation.
- This step ensures that only records not previously stored in the system are considered.

Filtering for Redundancy:

- Use another LINQ operation to filter out records that are already stored in the database.
- This step helps prevent data redundancy in the system.

CSV and Database Synchronization:

- The CSV file serves as a mechanism to keep track of stored records in sync with the database.
- New records filtered out in the second LINQ operation are saved into both the CSV and the database to maintain consistency.

Alternative Approach (Database Loading):

- Some may choose to load database records on application start to perform filtering, but this approach might impact performance as the data size increases.