Summary of Data Processing Workflow

# Chunk 1: Data Fetching and Logging

In this chunk, the system fetches data from various API endpoints provided by ASA. The list of endpoints includes information about players, games, teams, and more. The key fields are extracted from the response data and stored in DataFrames for further processing. Additionally, a log of each successful data fetch is written into the 'ChangeLogs' table in SQL Server. This log helps track all the data fetching operations for the relevant datasets.

# Chunk 2: Data Fetching with Required Columns

This chunk focuses on fetching data from the ASA API with specific required columns for each dataset. For each endpoint, a request is made, and the relevant columns are extracted and stored in separate DataFrames. If the data fetched is not in the expected format, an error message is logged to the 'ChangeLogs' table. This ensures that only valid data is processed and stored for further operations.

# Chunk 3: Data Cleaning and Validation

The third chunk of the workflow involves cleaning and validating the data retrieved from the API. It handles missing values, duplicate rows, and unwanted columns. Specific cleaning operations are applied based on the dataset, such as filling missing values in 'birth\_date' or 'position' columns, and removing rows with unhashable types. Additionally, certain columns are validated, and their data types are adjusted to ensure compatibility with SQL Server. Foreign key validations are also performed to ensure the integrity of relationships between datasets.

# Chunk 4: Data Upload and Indexing to SQL Server

The final chunk is responsible for uploading the cleaned and validated data into SQL Server. It dynamically generates SQL queries to create the necessary tables and populate them with the data from the DataFrames. Indexing is applied to the tables to enhance query performance, with clustered and non-clustered indexes created for relevant columns. After the data is uploaded and indexed, a final log of the successful upload operation is recorded in the 'ChangeLogs' table.