Assignment:

Create a program using Langchain or LLaMAindex that can read data from PDFs, store them with the date and other custom metadata like version tags, security and access tags, and retrieve specific document chunks based on various criteria.

Example:

Suppose you have a collection of PDFs containing mathematical reasoning, each with different version tags. The program should be able to:

1. Extract the reasoning from each PDF and store it along with the date, version tag, and other relevant metadata.
2. Given a specific date, the program should retrieve the reasoning from the PDF that was created closest to that date.
3. If multiple PDFs have the same date, the program should retrieve the reasoning from the PDF with the latest version tag.
4. Additionally, the program should allow users to specify custom metadata like security and access tags when storing the PDFs, and use these tags to control who can access and retrieve specific document chunks.

This program can be used for various purposes, such as:

- Versioning and tracking changes in documents over time
- Retrieving specific information from large collections of PDFs
- Controlling access to sensitive information based on security and access tags

Example:

Consider two PDF documents, PDF1 and PDF2, each containing two versions of the same mathematical reasoning:

PDF1:

- Version 1:
    - Mathematical reasoning: "If a triangle has two equal sides, then it is an isosceles triangle."
    - Date: `2023-01-01`
    - Version tag: v1.0
    - Security tag: Public
- Version 2:
    - Mathematical reasoning: "If a triangle has two equal sides and two equal angles, then it is an equilateral triangle."
    - Date: `2023-03-15`
    - Version tag: v2.0
    - Security tag: Confidential

PDF2:

- Version 1:
    - Mathematical reasoning: "If a triangle has three equal sides, then it is an equilateral triangle."
    - Date: `2023-02-01`
    - Version tag: v1.1
    - Security tag: Restricted
- Version 2:
    - Mathematical reasoning: "If a triangle has three equal sides and three equal angles, then it is a regular triangle."
    - Date: `2023-04-01`
    - Version tag: v2.1
    - Security tag: Top Secret

Test Case:

- Given the date `2023-03-08`, the program should retrieve the mathematical reasoning from PDF1, version 2, because it is the closest version to the given date.
- Given the date `2023-02-15`, the program should retrieve the mathematical reasoning from PDF2, version 1, because it is the latest version with that date.
- Given the security tag "Confidential" and the date `2023-03-01`, the program should retrieve the mathematical reasoning from PDF1, version 2, because it is the only version with that security tag and a date close to `2023-03-01`.

These test cases will help ensure that the program can accurately extract, store, and retrieve mathematical reasoning from PDFs based on various criteria, including date, version tag, and security tag.