**Q1. What is Web Scraping? Why is it Used? Give three areas where Web Scraping is used to get data.**

**Web Scraping** is the process of extracting data from websites in an automated manner. It involves fetching web pages and parsing their content to retrieve specific information.

**Why is it used?**

* To collect large volumes of data efficiently from online sources.
* To gather data that is not readily available in a structured format, enabling data-driven decision-making.
* To monitor, analyze, and utilize web-based information dynamically.

**Three areas where Web Scraping is used:**

1. **E-commerce:** To compare prices, monitor competitor trends, and analyze customer reviews.
2. **Real Estate:** To gather property listings, pricing, and availability from various platforms.
3. **Research & Analytics:** To collect data for academic or business research, including market trends, social media analytics, and news aggregation.

**Q2. What are the different methods used for Web Scraping?**

1. **Manual Scraping:** Copy-pasting data manually (not scalable).
2. **Using Libraries/Frameworks:** Libraries like Beautiful Soup, Scrapy, and Selenium are commonly used.
3. **APIs:** If a website provides an API, you can fetch data in a structured format like JSON.
4. **Headless Browsers:** Tools like Puppeteer and Selenium simulate user interactions with websites to scrape dynamic content.
5. **Data Extraction Tools:** Tools like Octoparse, ParseHub, or WebHarvy provide visual interfaces for scraping.

**Q3. What is Beautiful Soup? Why is it used?**

**Beautiful Soup** is a Python library used for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data.

**Why is it used?**

* To navigate and search the HTML structure of a webpage easily.
* To extract specific elements such as titles, links, or tables.
* To clean and preprocess web data before further analysis.

**Q4. Why is Flask used in this Web Scraping project?**

**Flask** is a lightweight web framework in Python used to create web applications.

**Reasons for using Flask in Web Scraping projects:**

1. To serve the scraped data dynamically via a web interface.
2. To provide an API endpoint for others to access the scraped data.
3. To integrate web scraping functionality with a user-friendly front-end interface.
4. To handle requests and responses for scraping tasks triggered by users.

**Q5. Write the names of AWS services used in this project. Also, explain the use of each service.**

1. **Amazon EC2 (Elastic Compute Cloud):**  
   Used to host and run the web scraping scripts and web application (Flask) in a virtual server environment.
2. **Amazon S3 (Simple Storage Service):**  
   Used for storing scraped data in a scalable and cost-effective manner, such as saving large datasets or logs.
3. **AWS Lambda:**  
   Used to execute web scraping scripts on demand in a serverless environment, reducing the need for managing infrastructure.
4. **AWS RDS (Relational Database Service):**  
   Used for storing and managing structured data collected from web scraping tasks.
5. **AWS CloudWatch:**  
   Used for monitoring the performance and logging the activities of the web scraping scripts or web application.

These AWS services collectively help in building, deploying, and managing scalable and efficient web scraping projects.