

# [CS209A-25Fall] Final Project Grading Scheme (100 points)

## Data Collection (10 points)

- **Collect at least 1,000 Java-tagged threads** from Stack Overflow using the official REST API.  
(5 points)
- **Persist data locally** in a structured database (e.g., PostgreSQL, MySQL) or well-organized local files like json. (5 points)

## Data Analysis (60 points)

### 1. Topic Trends (15 points)

- Topic identification: **5**
- Time dimension: **5**
- Meaningful measurement of "activity": **5**

### 2. Topic Co-occurrence (15 points)

- Topic identification: **5**
- Correct co-occurrence logic: **10**
- **Deductions:**
  - N too small or too large: **-2**

### 3. Common Pitfalls in Multithreading (15 points)

- Identifying pitfalls for multithreading: **10**
- Pitfall aggregation (find the common ones): **5**
- **Deductions:**
  - Only analyze tag, without thread body content: **0**
  - No text analysis (e.g., use AI generated common pitfalls): **0**
  - Limited or wrong approach for identifying pitfalls

### 4. Solvable vs. Hard-to-Solve Questions (15 points)

- Identify solvable questions: **3**
- Identify hard-to-solve questions: **3**
- Identify at least 3 meaningful contributing factors: **9**
- **Deductions:**
  - Shallow or wrong analysis
  - < 3 factors (-3 each)

## RESTful APIs (10 points)

- Available REST API: **5**
- Returns valid JSON with correct result: **5**

- **Deductions:**

- Returns non-JSON (e.g., HTML, plain text): **-5**
- Poor API design (e.g., confusing resources and parameters): **-3**
- < 2 REST API: **-5 each**

## Visualization & Insights (20 points)

For each of the 4 question:

- Visualization: 3 points (good, fair, poor). A good visualization must be **accurate, clear, aesthetically appropriate, and user-friendly** to answer the question.
- Insights: 2 points (good, poor). You should discuss the insights during presentation.

## Framework & Compliance (Deduction-only)

- Not using Spring Boot: **-40**
- Precomputed/static results: **-20**
- Use other language (e.g., Python) for data analysis: **-20**. Yet, it's okay to use Python **only for** data collection with Stack Overflow REST API.
- AI-generated analysis: **0 points for the affected question(s)**