(1) Who are your team members?

Mirra Klimov & Spencer Ye

(2) Briefly describe your target domain

Our database revolves around track and field results from the Centennial Conference (CC).

(3) Give a reasonably comprehensive and representative list of the kinds of English questions you would like your system to be able to answer (minimum 15).

1. What is the personal-best (PB) time for Mirra Klimov in the 100m Dash?
2. Which athletes from Ursinus College competed in the 2025 Indoor Season?
3. What is the average winning time for the 5000m Run at all meets held in New York last year?
4. Who are the top 3 athletes in the Javelin Throw for the 2025 Outdoor Season from each university in the Centennial Conference?
5. How many athletes from the Centennial Conference competed in the Decathlon at the 2024 National Championships?
6. What is the personal best for every current athlete from a school in Baltimore in the Long Jump?
7. Which meets were held on Saturday, April 12, 2025 that CC athletes attended?
8. Which CC races had the winning time that within 0.5 seconds of the meet record?
9. Which CC athletes competed in both the 400m Dash and the 4x400m Relay at the Penn Relays in 2024?
10. What are the team scores for Johns Hopkins University at the 2024 Outdoor Conference Championship for Men?
11. Which cities hosted a track meet with more than 500 CC competitors?
12. What are the results from all races for Spencer Ye in 2024?
13. What is the season record in the 200m Dash for Alex Colletti in each season?
14. Who won the 800m dash in the 2024 Centenial Conference Outdoor Championships?
15. Which athlete has competed in the most relays?
16. How many seniors does each school have?

(4) Design and show a relational data model that you plan to use for your system, with a preliminary implementation in standard SQL data-definition-language syntax. This specification should include appropriate primary key, foreign key and domain specifications for each relation/attribute, as well as the not null constraint when appropriate. You may also find it useful, but not required, to create a few insert-into statements that populate your schema designs with representative values (both to document your choices and to exercise them. You are welcome to change and augment your design and its specification by Phase II, but any time investment now will reduce effort later.

SEE .SQL FILE IN REPO

(5) Submit a set of SQL statements that will implement a representative sample of your target queries, including some of the more interesting or challenging cases. This is primarily to get you to think about your design and how it will be exercised as well as any limitations, so focus on queries that would be useful for doing so, rather than creating trivial or non-insightful queries just to fill space.

Which CC athletes competed in both the 400m Dash and the 4x400m Relay at the Penn Relays in 2024?

```
SELECT A.athletename
FROM AthleteSeason AS AtS
JOIN Athlete AS A ON A.athleteid = AtS.athleteid
WHERE AtS.seasonyear = 2024
  AND EXISTS (
    SELECT 1
    FROM DidEvent AS DE
    JOIN MeetEvent AS ME ON ME.meeteventid = DE.meeteventid
    JOIN Event AS E ON E.eventid = ME.eventid
    JOIN TrackMeet AS M ON M.meetid = ME.meetid
    WHERE DE.athleteseasonid = AtS.athleteseasonid
      AND M.meetname LIKE ‘%Penn Relays%’
      AND E.eventname = ‘400m’
  )
  AND EXISTS (
    SELECT 1
    FROM DidEvent AS DE
    JOIN MeetEvent AS ME ON ME.meeteventid = DE.meeteventid
    JOIN Event AS E ON E.eventid = ME.eventid
    JOIN TrackMeet AS M ON M.meetid = ME.meetid
    JOIN RelayTeam AS RT ON RT.relayteamid = DE.relayteamid
    JOIN RelayTeamMembers AS RTM ON RTM.relayteamid = RT.relayteamid
    WHERE RTM.athleteseasonid = AtS.athleteteamid
      AND M.meetname LIKE ‘%Penn Relays%’
      AND E.eventname = ‘4x400m’
  )
```

Who won the 800m dash in the 2024 Centennial Conference Outdoor Championships?

```
SELECT A.athletename
FROM DidEvent AS DE
JOIN MeetEvent AS ME ON ME.meeteventid = DE.meeteventid
JOIN TrackMeet AS M ON M.meetid = ME.meetid
JOIN Event AS E ON E.eventid = ME.eventid
JOIN AthleteSeason AS AtS ON AtS.athleteseasonid = DE.athleteseasonid
JOIN Athlete AS A ON A.athleteid = AtS.athleteid
WHERE E.eventname = ‘800m Dash’
  AND M.meetname = ‘Centential Conference Outdoor Championships 2024’
  AND DE.place = 1
```

Calculate the total points Johns Hopkins University would have earned at the 2024 Outdoor Conference Championship, using a standard 10-8-6-5-4-3-2-1 scoring system for top 8 places.

```
SELECT
    S.SchoolName,
    SUM(
        CASE
            WHEN DE.Place = 1 THEN 10
            WHEN DE.Place = 2 THEN 8
            WHEN DE.Place = 3 THEN 6
            WHEN DE.Place = 4 THEN 5
            WHEN DE.Place = 5 THEN 4
            WHEN DE.Place = 6 THEN 3
            WHEN DE.Place = 7 THEN 2
            WHEN DE.Place = 8 THEN 1
            ELSE 0
        END
    ) AS TotalPoints
FROM DidEvent AS DE
JOIN MeetEvent AS ME ON ME.MeetEventID = DE.MeetEventID
JOIN TrackMeet AS TM ON TM.MeetID = ME.MeetID
LEFT JOIN AthleteSeason AS AtS ON AtS.AthleteSeasonID = DE.AthleteSeasonID
LEFT JOIN RelayTeam AS RT ON RT.RelayTeamID = DE.RelayTeamID
JOIN School AS S ON S.SchoolID = COALESCE(AtS.SchoolID, RT.SchoolID)
WHERE
    TM.MeetName = 'Centennial Conference Outdoor Championships 2024'
    AND S.SchoolName = 'Johns Hopkins University'
GROUP BY S.SchoolName;
```

(6) Provide a plan for how you will load the database with values.
– If you plan to extract/import data from on-line sources, briefly describe what are the sources (e.g. personal data, or provide URL's) and what are any format conversion issues you expect to encounter.
– If you plan to imput your data primarily through a web or form-based interface, briefly describe this interface and the issues involved.

We will download performance lists for each season, by each school to get all performances by that school (ex. link). This will take about 20 clicks per year. We will create a simple python script to scrape all the necessary data.
We will add the meet location by hand. There are approximately 50 meets per year.
We also plan on collecting weather data from a public api after we know all big locations.

(7) Very briefly describe the form/type of output or result you plan to generate or any special user interface issues (e.g. views) that you plan to implement.

- Users cannot add data to our database, so there are no input forms.
- A user will be able to look up specific athletes, schools, meets, and events. Upon inputting what they're looking for they will see entries that include athlete name, event, meet, year, season, and performance.

- There will be a special sheet that predicts the results at the Conference Championship meet (indoor or outdoor).

(8) What are the specialized/advanced topics you plan to focus on in your database design?

Students in 601.415/615 should do major work in at least one of these areas and minor work in at least one other. Students in 601.315 should do at least minor work in one of these additional areas of specialization. More details will be discussed in class.

We definitely plan on needing to do complex data extraction from online sources, as we need to parse pdfs in order to get the data we need. Another places we can explore include minor work in particularly advanced GUI report generation. We may also do natural language interfaces and some machine learning predictions.

Notes On Completion:

Try different prediction models, attempt to categorize improvement over season, visualization (export data from db) for trends (over time, etc.),

Utilizing Neon instead of MySQL, Next.js

Refresh data button on a update info page (and add number of updated results)

Prediction Page
- Predictions based on PRS
- Predictions based on improving vs burnt out
- Prediction based on average performance thoughout
- Predictio of just the teams or events also
- Essentially maybe some AI/ML to do said prediction

Tffrs Page
- Lookup results
    - By athlete
        - Personal best
        - Season bests
        - Historical trends (graph)
        - Any predictive models (improving, burnt out, stagnant, etc)
        - Rankings
    - By school
        - Records
        - Historical trends
        - Seasons bests
        - Athlete Rankings (school, conference)
    - By meet
        - Rankings
        - Historical trends
    - Overall conference results

THINGS TO NOTE ABT HOW ITS SET Up

- Results like `DNF`, `DNS`, `DQ`, `FOUL` are skipped entirely — they won't appear in the database.
- If a meet is scraped multiple times with different dates (e.g., day 1 and day 2), the `StartDate` and `EndDate` automatically expand: