67-262 Database Design and Development, Fall 2025
Project Phase-2 Report

Project Title: *[Duolingo]*

Team Number: *[T1-08]*

Team Members: *[Jullie Cach (jcach), Lok Hariharan (lharihar), Parishi Jain (parishij)]*

1. Updated List of User Stories

Use the table below to list your updated user stories after incorporating feedback from phase-1. ***Briefly explain how you addressed your TA mentor's feedback as well as any additional changes made since phase-1.*** Ensure that all user stories meet the requirements in Step 3 of phase-1 instructions.

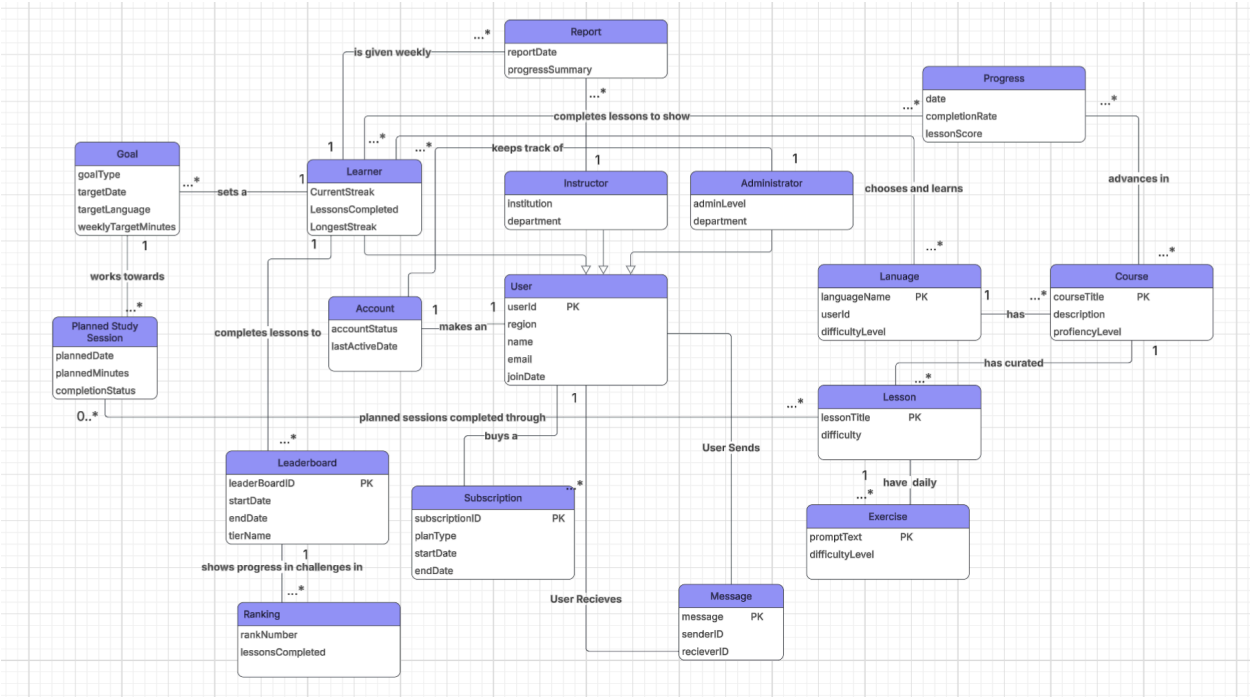
ID	Verb+Noun	As a <role>	I want <goal>	So that <reason>	Simple / Complex	Operational / Analytical
US1	Track progress	Learner	I want to view my daily lesson completion and streak status	So that I can see my progress and stay motivated to practice every day	Simple	Operational
US2	Analyze progress	Learner	I want to see how my lesson completion rate and lesson scores change over time across different lessons	So that I can identify which topics I am improving in and which ones I need to review	Complex	Analytical
US3	View leaderboard	Learner	I want to see my ranking among friends and other learners	So that I can stay motivated through friendly competition	Complex	Operational
US4 **NEW**	Evaluate study plan adherence	Learner	I want to compare the study sessions I planned with the lessons I actually	I can understand how well I am following my study plan and adjust	Complex	Analytical

			completed	my goals if needed		
US5	Assign lessons	Instructor	I want to review each learner's progress across the lessons in a course, including their completion rates and lesson scores	So that I can identify which learners may need additional support and which lessons are causing difficulty	Complex	Operational
US6	Generate report	Instructor	I want to generate a report showing students' accuracy improvement over time	So that I can measure which students are making progress	Complex	Analytical
US7	Message students	Instructor	I want to send feedback to students within the platform	So that I can communicate without external apps or emails	Simple	Operational
US8	Analyze regional learner engagement	Adminstrator	I want to analyze learner engagement across regions by comparing the number of active accounts and subscription types	So that I can identify which regions show strong platform usage and prioritize resources or marketing accordingly	Complex	Analytical
US9	View subscription growth	Adminstrator	I want to see monthly subscription growth by region	So that I can evaluate marketing performance and regional trends	Complex	Analytical
US10 (Trigger function)	Deactivate accounts	Adminstrator	I want to remove inactive or spam accounts from the system	So that I can maintain database quality and user security	Simple	Operational

2. Updated Conceptual Model

[Insert the image of your updated UML conceptual model diagram here (exported from your diagramming tool). Ensure the image is high resolution so all texts and diagram elements are legible.]

Briefly explain how you addressed your TA mentor's feedback as well as any additional changes made since phase-1.



Assumptions

List any key assumptions made when modeling entities, attributes, and relationships.

3. Relational Model

List the relations converted from your updated conceptual model, following the steps we discussed in class. Clearly label all PKs and FKs. Ensure the chosen PKs do not lead to inadvertent business consequences, and state any assumptions you made when selecting the PKs.

Final Combined Relational Model

Users(userID, region, name, email, joinDate)

Learner(userID, currentStreak, lessonsCompleted, longestStreak)

Instructor(userID, institution, department)

Administrator(userID, adminLevel, department)

Account(userID, accountStatus, lastActiveDate)

Language(languageName, difficultyLevel, userID)

Goal(goalType, targetDate, targetLanguage, weeklyTargetMinutes, userID)

Course(courseTitle, description, proficiencyLevel)

Lesson(lessonTitle, difficulty, courseTitle)

Exercise(promptText, difficultyLevel, lessonTitle)

Progress(userID, date, lessonTitle, completionRate, lessonScore)

PlannedStudySession(plannedDate, userID, lessonTitle, plannedMinutes, completionStatus)

Leaderboard(leaderboardID, startDate, endDate, tierName)

Ranking(userID, leaderboardID, rankNumber, lessonsCompleted)

Subscription(subscriptionID, planType, startDate, endDate, userID)

Report(reportDate, progressSummary, LearnerID, InstructorID)

Message(message, senderID, recieverID)

Combined Assumptions

1. Each user is uniquely identified by `userID` across all roles (User, Learner, Instructor, Administrator).
2. Lessons are uniquely identified by `lessonTitle`. Courses are uniquely identified by `courseTitle`.
3. A learner may complete a lesson many times across different dates; the Progress table enforces one record per (userID, lessonTitle, date).
4. A planned study session is uniquely identified by (plannedDate, userID, lessonTitle).
5. PlannedStudySession requires an existing learner and an existing lesson.
6. Progress captures actual learning behavior and may differ from planned sessions. This supports adherence evaluation.
7. Message content is assumed to be unique, so `message` serves as the primary key.
8. Reports are uniquely identified by reportDate and link exactly one learner and instructor.
9. Exercises belong to lessons; lessons belong to courses; courses belong to languages.

4. Functional Dependencies and Normalization

For EACH relation obtained from the section above:

- Identify ALL **functional dependencies** in that relation. Clearly state any assumptions upon which the FDs are based. Your assumptions must be reasonable in the business context.
- For **each FD**, workout the **closure** of its LHS, and determine if the FD is **good or bad**. For any bad FD, identify what type of bad FD it is (**partial** dependency or **transitive** dependency).
- Based on the above, indicate the highest **normal form** (1NF, 2NF, 3NF, or BCNF) the relation satisfies.
- If the relation is already in BCNF, no further action is needed. Else if the relation is **not in BCNF**, perform and show the **BCNF decomposition process** (as discussed in class). For each relation in the final decomposition results, list its relevant FDs, verify that it is indeed in BCNF (i.e., contains no bad FDs), and clearly label all PKs and FKs.

Users

- FD: userID → region, name, email, joinDate
- BCNF: userID is the primary key and the only determinant → BCNF.

Instructor:

- FD: userID → currentStreak, lessonsCompleted, longestStreak
- BCNF: userID is the key and determines all attributes → BCNF.

Instructor:

- FD: userID → institution, department
- BCNF: Key is userID → BCNF.

Administrator:

- FD:userID → adminLevel, department
- BCNF: Determinant is the primary key, userID → BCNF

Account:

- FD: userID → accountStatus, lastActiveDate
- BCNF:userID is the only determinant → BCNF.

Language:

- FD: languageName → difficultyLevel

- BCNF: languageName is the primary key → BCNF.

Goal:

- FD:(goalType, targetDate, targetLanguage) → weeklyTargetMinutes, userID
- BCNF:Composite key is only determinant → BCNF.

Course:

- FD: courseTitle → description, proficiencyLevel
- BCNF: Primary key courseTitle determines everything → BCNF.

Lesson:

- FD: lessonTitle → difficulty, courseTitle
- BCNF:lessonTitle is the only determinant → BCNF.

Exercise:

- FD: promptText → difficultyLevel, lessonTitle
- BCNF: Primary key promptText determines all attributes → BCNF.

Progress:

- FD: (userID, date, lessonTitle) → completionRate, lessonScore
- BCNF:Only the full composite key determines other attributes → BCNF.

PlannedStudySession:

- FD: (plannedDate, userID, lessonTitle) → plannedMinutes, completionStatus
- BCNF:Full composite key is the only determinant → BCNF.

Leaderboard:

- FD:leaderboardID → startDate, endDate, tierName
- BCNF:Primary key is only determinant → BCNF.

Ranking:

- FD: (userID, leaderboardID) → rankNumber, lessonsCompleted

- BCNF:Composite key determines all attributes → BCNF.

Subscription

- FD:subscriptionID → planType, startDate, endDate, userID
- BCNF: subscriptionID is the only determinant → BCNF.

Report

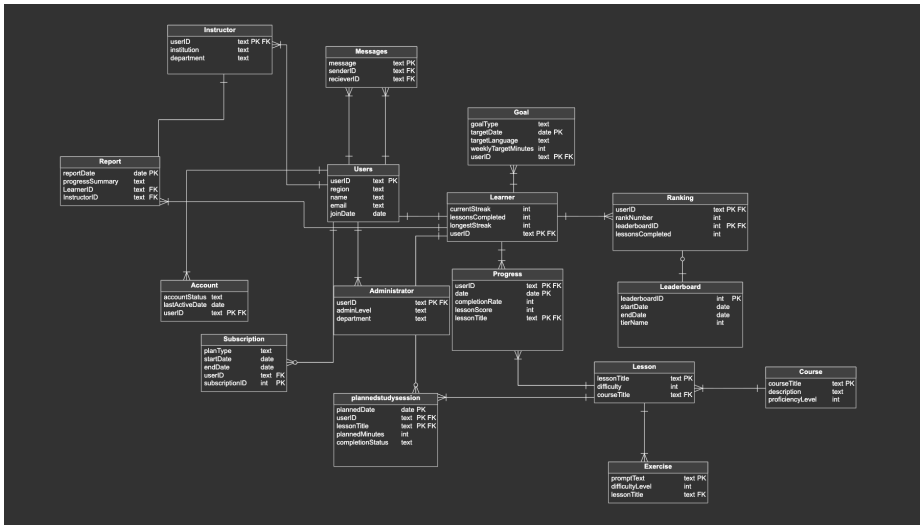
- FD: reportDate → progressSummary, LearnerID, InstructorID
- BCNF:reportDate is the primary key → BCNF.

Message

- FD: message → senderID, receiverID
- BCNF message is unique content and the primary key → BCNF

5. Physical Model

[Insert the image of your physical model created in Vertabelo. Ensure the image is high resolution so all texts and diagram elements are legible.]



6. References (if applicable)

Provide references to any web sources, datasets, or materials consulted. Use a consistent citation format.