



ENSF 608 – Databases – Study Guide

Dr. Ronnie de Souza Santos, PhD.

Scenario: Vocabularing

System Description: Vocabularing is a daily word challenge. Each calendar day publishes a single secret word. Every participant has up to five guesses to discover that day's word before the daily closing time. Results, progress streaks, and simple stats are shown to the participant and kept for history. People sign up with an email and password. The system keeps a unique internal code for each person, their email, a stored password hash, the account creation time, and an active or blocked status. After sign-up, the person completes a profile with first name, middle name, last name, a short bio, date of birth, a picture link, and location split into city, province or state, and country. The person can choose more than one preferred language for the interface and for messages. The person's age is displayed in the app but is computed from the date of birth. Every date has a single daily challenge with the calendar date, the secret word for that day, a short clue, the release time, and the closing time. The date uniquely identifies that daily challenge. When a participant plays, each guess is recorded with who guessed, which day it belongs to, the guessed word, the order of the attempt from one to five, the time sent, and the outcome such as correct or incorrect. The interface shows how many attempts remain for that day and whether the day was solved. The system also keeps a per-person daily history for streaks. Each streak entry exists only in the context of one person on one date and notes whether that date extended the streak. The running streak length that is displayed is computed from that history rather than fixed. Optional extras for teaching variety: people may store multiple notification channels such as email and mobile push; recovery codes can be issued for account access and expire at a given time.

Requirements Details

1. Access and security

- A person can create an account using email and password so that they can play.
- The system stores a unique internal code for each person, the email, a stored password hash, the account creation time, and an active or blocked status.
- A person can request login recovery codes, and each code expires at a defined time.

2. Profile

- After creating an account, a person can complete a profile with first name, middle name, last name, a short bio, date of birth, a picture link, and location recorded as city, province or state, and country.
- A person can choose more than one preferred language for the interface and messages.
- The app displays the person's age, which is calculated from the date of birth.

3. Daily challenge

- The system creates exactly one challenge per calendar date with the date, the secret word, a clue, a release time, and a closing time.
- The calendar date identifies that day's challenge and is used to tie play sessions to the correct day.

4. Gameplay and guesses

- A person can submit up to five guesses for the current day.
- Each guess records who sent it, which date it belongs to, the guessed word, the attempt number from one to five, the time sent, and the outcome such as correct or incorrect.
- The interface shows the number of remaining attempts for the current day and locks out further attempts after five or after a correct guess.

5. Feedback and streaks

- After each guess, the person receives immediate feedback that indicates whether the guess is correct and simple positional hints as defined by the game rules.
- For each person and date, the system keeps a streak entry that notes whether the streak continues. These entries exist only in the context of the person they belong to.
- The streak length shown on screen is calculated from the person's streak entries rather than stored as a fixed number.

6. Notifications

- A person can register more than one notification channel such as email and mobile push.
- The system may send a reminder shortly after the daily release time if the person has not played yet, according to the person's notification channels.

7. Fair play and administration

- The system prevents playing past the daily closing time and prevents playing future dates.
- An administrator can create, review, and publish upcoming daily challenges with their release and closing times.
- An administrator can deactivate accounts that violate rules and can view high-level play statistics for moderation.

8. Quality expectations

- The daily challenge page loads in under two and a half seconds on a typical 4G connection during peak hours.
- The system supports screen readers and high-contrast modes so that visually impaired players can participate.
- All stored and transmitted personal information is encrypted.

Questions:

1. Based on the Vocabularing system, define two **different types** of entities that belong to the database. For each entity, list its name, attributes, data types, and clearly identify the primary key. Use the format below:

TEAMMEMBER (Member_ID, Name, Role)

- Member_ID → int, primary key
- Name → varchar(50)
- Role → varchar(50)

2. While modeling *Vocabularing*, a data analyst decides to store the player's age directly in the database instead of calculating it from the date of birth. What design issue could this cause?

- A. The value of age would always remain accurate.
- B. It introduces redundancy and may lead to inconsistent data if not updated.
- C. It improves query performance for demographic reports since no calculation is needed.
- D. It eliminates the need to store the date type, reducing the workload.

3. In *Vocabularing*, each Player can submit up to five Guesses per DailyChallenge. Which relationship correctly represents this situation?

- A. One-to-one between Player and Guess, and a method to count five guesses.
- B. Many-to-many between Player and Guess, in case rules change.
- C. One-to-many between DailyChallenge and Player, so DailyChallenge controls the five guesses.
- D. One-to-many between Player and Guess, constrained by business rules to five.

4. A developer forgets to declare the foreign key from *Guess* to *DailyChallenge*. Later, guesses appear for days that never existed. Which integrity problem is occurring?
- A. Insertion anomaly due to missing referential integrity
 - B. Deletion anomaly caused by cascading rules
 - C. Transitive dependency between *Guess* and *Player*
 - D. Multivalued dependency between *Guess* and *DailyChallenge*
5. The *Guess* entity is identified by a combination of *PlayerID*, *ChallengeDate*, and *AttemptNumber*. Why is this considered a composite key?
- A. Because multiple attributes together uniquely identify a record.
 - B. Because it includes a foreign key and a derived attribute.
 - C. Because all attributes are multivalued.
 - D. Because it removes the need for normalization.
6. In *Vocabularing*, the database designer places the **ClueText** attribute of *DailyChallenge* inside the *Guess* table to avoid joins. Which normal-form principle is being violated?
- A. 1NF – atomicity of attributes
 - B. 2NF – elimination of partial dependencies
 - C. 3NF – elimination of transitive dependencies
 - D. BCNF – determinant must be a candidate key
7. Suppose *Vocabularing* introduces new roles such as *Player*, *Admin*, and *ContentReviewer*. Admins can perform more than one role at the same time, for example, they may be both a *Player* and a *ContentReviewer*. Which ER concept best represents this design?
- A. Disjoint specialization, since each user can belong to only one role at the time.
 - B. Generalization, since *Player* and *Admin* merge into one higher-level entity.
 - C. Weak entities, since each role depends entirely on the *User* entity.
 - D. Overlapping specialization, since one user can belong to multiple sub-entities simultaneously.
8. When mapping the *Vocabularing* conceptual model to the logical model, the *Location* attribute in the *Profile* (with *City*, *Province*, and *Country*) is identified as composite. How should this be represented in the logical model?
- A. Keep *Location* as one text column to simplify the design.
 - B. Remove *Location* completely since its parts are self-explanatory.
 - C. Create a new table for the composite attribute, including each sub-part and a foreign key to the *Profile* entity.
 - D. Treat each sub-part as a derived attribute calculated at runtime with a foreign key to the *Profile* entity.
9. In *Vocabularing*, the *Player* entity uses *Player_ID* as its primary key, but *Email* is also unique and could identify each record. What is the correct interpretation of *Email* in this case?
- A. It is a candidate key because it can uniquely identify an entity but is not the primary key.
 - B. It is a foreign key that links to the *Profile* entity and can uniquely identify the entity.
 - C. It is a composite key formed with *Player_ID* to uniquely identify the entity.
 - D. It is a better primary key and should replace *Player_ID* for consistency.
10. When defining the attribute *PreferredLanguage* in the *Profile* table, the database designer specifies it as *VARCHAR(20)* and restricts its values to the set {'English', 'French', 'Spanish'}. What concept is being applied here?
- A. Referential integrity between *Profile* and *Language*.
 - B. Atomicity of the attribute to satisfy 1NF.
 - C. Generalization between languages.
 - D. Definition of the domain and semantics to constrain values.

11. In *Vocabularing*, both Player and DailyChallenge are strong entities. To record which player made which guesses, the database designer decides to connect them through a relationship. How should this relationship appear in the logical model?

- A. It becomes a derived attribute in Player.
- B. It merges Player and DailyChallenge into a single table.
- C. It is represented only as a foreign key in Player.
- D. It becomes a separate table that includes the primary keys of both strong entities as foreign keys.

12. Imagine the *Vocabularing* analytics team wants to improve retention of players. What kind of information (descriptive, predictive, or prescriptive) should they prioritize, and why? Provide one concrete example of a question they could answer using that type of information and what entities and attributes will be involved in the process.