**Core Database Tables**

**1. users Table**

This table will store essential information about all registered users of the application.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints / Description** |
| user\_id | UUID (Primary Key) | A unique identifier for each user. Using UUIDs helps prevent ID collisions in distributed systems. |
| Email | VARCHAR(100) | The user's email address, must be unique. Also used for login and communication. |
| Display\_name | VARCHAR(256) | User’s Display name. |
| password\_hash | VARCHAR(255) | Stores the securely hashed version of the user's password. **Never store plain text passwords.** |
| role | VARCHAR(20) | Defines the user's access level (Only ‘user’ and ‘admin’ allowed for the time being). |
| credits | INT(50) | Number of allowed uses/operations remaining |
| created\_at | TIMESTAMP | Automatically records the timestamp when the user account was created. |
| updated\_at | TIMESTAMP | Last profile update |
| last\_login | TIMESTAMP | Last login time |

Notes:

* For registered, credits can be initialized higher or managed by our business logic.

**2. translation\_projects Table**

This table allows users to organize their subtitle files into logical projects. It enhances user experience, especially for users with many files or recurring translation needs.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints / Description** |
| project\_id | UUID (Primary Key) | A unique identifier for each translation project. |
| user\_id | UUID (Foreign Key) | Links to the users.user\_id. Owner of the project. |
| project\_name | VARCHAR(100) | A user-defined name for the project (e.g., "Lecture Series," "Movie Subtitles"). |
| description | TEXT | An optional longer description of the project's purpose. |
| created\_at | TIMESTAMP | Records when the project was created. |
| updated\_at | TIMESTAMP | Automatically updates whenever the project details are modified. |

**3. subtitle\_files Table**

This is a central table for storing metadata about all subtitle files in the system, including both original uploads and their translated versions. The actual file content will be stored in Azure Blob Storage / S3 / MinIO with a reference here.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints / Description** |
| file\_id | UUID (Primary Key) | A unique identifier for each individual subtitle file record. |
| project\_id | UUID (Foreign Key) | Links to translation\_projects.project\_id. Nullable if the file is not associated with a specific project. |
| user\_id | UUID (Foreign Key) | Links to users.user\_id. Identifies the user who uploaded or is associated with this file. |
| original\_filename | VARCHAR(255) | The name of the file as it was originally uploaded by the user (e.g., MyMovie.srt). |
| storage\_path | VARCHAR(512) | The URL or unique path within Azure Blob Storage where the actual file content is stored. |
| file\_format | VARCHAR(10) | The format of the subtitle file (e.g., 'SRT', 'VTT', 'ASS', 'DFXP'). |
| file\_size\_bytes | BIGINT | The size of the file in bytes. |
| is\_original | BOOLEAN | TRUE if this record represents the original uploaded file; FALSE if it's a translated version. |
| is\_public | BOOLEAN | TRUE if public, False is private |
| has\_profanity | BOOLEAN | True if it has, False if doesn’t |
| source\_language | VARCHAR(10) | The BCP-47 language tag of the subtitle content (e.g., 'en' for English, 'es' for Spanish). |
| created\_at | TIMESTAMP | Records when this file record was created in the database. |

**4. translations Table**

This table logs each instance of a translation operation, linking the original file to its translated output and recording details about the translation process.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints / Description** |
| translation\_id | UUID (Primary Key) | A unique identifier for each specific translation event. |
| file\_id | UUID (Foreign Key) | Links to subtitle\_files.file\_id where is\_original is TRUE. Identifies the source file. |
| translated\_file\_id | UUID (Foreign Key) | Links to subtitle\_files.file\_id where is\_original is FALSE. Identifies the translated output file. (Nullable if translation is ongoing or failed). |
| source\_language | VARCHAR(20) | The BCP-47 language tag of the language the file was originally. |
| target\_language | VARCHAR(20) | The BCP-47 language tag of the language the file was translated *into* (e.g., 'fr', 'de'). |
| translation\_status | VARCHAR(20) | The current status of the translation process (e.g., 'pending', 'completed', 'failed', 'editing', 'queued'). |
| translation\_service | VARCHAR(50) | The specific translation service used (e.g., 'Azure AI Translator', 'OpenAI LLM', 'Hybrid'). |
| requested\_at | TIMESTAMP | The timestamp when the translation request was initiated. |
| completed\_at | TIMESTAMP | The timestamp when the translation process was completed (nullable if pending). |
| has\_profanity | BOOLEAN | True if it has, False if it doesn’t |
| translation\_cost | DECIMAL(10,4) | Optional: Stores the estimated cost incurred for this specific translation API call. |
| manual\_edits\_count | INTEGER | Records how many times this specific translated version has been manually edited. |
| last\_edited\_by\_user\_id | UUID (Foreign Key) | Links to users.user\_id. Records the ID of the user who performed the last manual edit. (Nullable if no edits). |
| last\_edited\_at | TIMESTAMP | The timestamp of the last manual edit performed on this translation. |

**5. live\_sessions Table (For Real-Time Translation Logging)**

This table will store metadata about each real-time translation session (e.g., a live webinar). The actual full transcript content from these sessions would likely be stored in Azure Blob Storage, with a reference here.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Constraints / Description** |
| session\_id | UUID (Primary Key) | A unique identifier for each live translation session. |
| user\_id | UUID (Foreign Key) | Links to users.user\_id. The user who initiated or was associated with this live session. |
| session\_title | VARCHAR(100) | An optional name for the live session (e.g., "Monday Stand-up," "Product Demo Webinar"). |
| source\_platform | VARCHAR(50) | The platform from which captions were sourced (e.g., 'Microsoft Teams', 'Zoom', 'Custom RTMP'). |
| source\_language | VARCHAR(10) | The original spoken language of the live captions during the session. |
| target\_language | VARCHAR(10) | The language into which the captions were translated during the session. |
| start\_time | TIMESTAMP | The timestamp when the live session translation began. |
| end\_time | TIMESTAMP | The timestamp when the live session translation concluded. |
| full\_transcript\_path | VARCHAR(512) | Path to the complete (original or translated) transcript of the session stored in Azure Blob Storage for post-event access. |
| translation\_log\_path | VARCHAR(512) | Optional: Path to more detailed, granular translation event logs for auditing/debugging. |

**Relationships**

* **One-to-Many:**
  + users to translation\_projects (one user can have many projects)
  + users to subtitle\_files (one user can upload many files)
  + users to translations (one user can initiate many translations)
  + users to live\_sessions (one user can have many live sessions)
  + translation\_projects to subtitle\_files (one project can contain many subtitle files)
* **One-to-One / One-to-Many (depending on implementation):**
  + subtitle\_files to translations (an original file can have many translated versions, and each translated version links back to one original).