7. DATABASE

7.1 Database Overview

The Study in Woods platform uses PostgreSQL 15.x as its primary relational database management system, storing all persistent application data across 30+ tables. The database schema is designed following normalization principles (3NF) to minimize data redundancy while maintaining referential integrity through foreign key constraints. GORM (Go Object Relational Mapping) library manages database operations, automatic migrations, and relationship handling.

7.2 Core Tables

7.2.1 User Management Tables

Table: users

Stores user account information including authentication credentials and profile data.

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Auto-incrementing user identifier
created_at	TIMESTAMP	NOT NULL	Account creation timestamp
updated_at	TIMESTAMP	NOT NULL	Last modification timestamp
deleted_at	TIMESTAMP	NULL,	Soft delete timestamp

Column	Type	Constraints	Description
		INDEX	(NULL = active)
email	VARCHAR(255)	UNIQUE, NOT NULL	User email address (login identifier)
password_hash	VARCHAR(255)	NOT NULL	bcrypt hashed password (cost factor 12)
password_salt	BYTEA	NOT NULL	Random salt for key derivation
name	VARCHAR(255)	NOT NULL	Full name of user
role	VARCHAR(20)	DEFAULT 'student'	User role: 'student' or 'admin'
semester	INTEGER	DEFAULT 1	Current semester number (for students)
token_version	INTEGER	DEFAULT 0	Incremented to invalidate all user tokens

Indexes: email (unique), deleted_at (partial), role

Relationships: Has many ChatSessions, ChatMessages, UserCourses, AdminAuditLogs

7.2.2 Academic Hierarchy Tables

Table: universities

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	University identifier
name	VARCHAR(255)	UNIQUE, NOT NULL	Full university name
code	VARCHAR(50)	UNIQUE, NOT NULL	University code (e.g., "AKTU", "DU")
location	VARCHAR(255)	NULL	City/State location
website	VARCHAR(255)	NULL	Official website URL
is_active	BOOLEAN	DEFAULT TRUE	Active status flag
created_at	TIMESTAMP	NOT NULL	Creation timestamp
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

Table: courses

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Course identifier
university_id	INTEGER	FOREIGN KEY, INDEX	References universities(id) ON

			DELETE CASCADE
name	VARCHAR(255)	NOT NULL	Course name (e.g., "MCA", "BCA")
code	VARCHAR(50)	UNIQUE, NOT NULL	Course code
description	TEXT	NULL	Course description
duration	INTEGER	DEFAULT 4	Duration in semesters
created_at	TIMESTAMP	NOT NULL	Creation timestamp
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

Table: semesters

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Semester identifier
course_id	INTEGER	FOREIGN KEY, INDEX	References courses(id) ON DELETE CASCADE
number	INTEGER	NOT NULL	Semester number (1, 2, 3...)
name	VARCHAR(50)	NULL	Display name (e.g., "Semester 1")
created_a	TIMESTAMP	NOT NULL	Creation timestamp

Column	Type	Constraints	Description
t			
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

Table: subjects

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Subject identifier
semester_id	INTEGER	FOREIGN KEY, INDEX	References semesters(id) ON DELETE CASCADE
name	VARCHAR(255)	NOT NULL	Subject name
code	VARCHAR(50)	NOT NULL	Subject code
credits	INTEGER	DEFAULT 0	Credit hours
description	TEXT	NULL	Subject description
knowledge_base_uuid	VARCHAR(100)	NULL	DigitalOcean Knowledge Base UUID
agent_uuid	VARCHAR(100)	NULL	DigitalOcean Agent UUID
created_at	TIMESTAMP	NOT NULL	Creation timestamp

Column	Type	Constraints	Description
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

7.2.3 Document Management Tables

Table: documents

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Document identifier
subject_id	INTEGER	FOREIGN KEY, INDEX	References subjects(id) ON DELETE CASCADE
type	VARCHAR(20)	NOT NULL	Type: 'syllabus', 'pyq', 'book', 'reference', 'notes'
filename	VARCHAR(255)	NOT NULL	Original filename
original_url	TEXT	NULL	Original source URL (if crawled)
spaces_url	TEXT	NOT NULL	DigitalOcean Spaces full URL
spaces_key	VARCHAR(25	NOT NULL	S3 object key in

Column	Type	Constraints	Description
	5)		Spaces
data_source_id	VARCHAR(100)	NULL	Knowledge Base data source ID
indexing_job_id	VARCHAR(100)	NULL	Knowledge Base indexing job ID
indexing_status	VARCHAR(20)	DEFAULT 'pending'	'pending', 'in_progress', 'completed', 'failed'
indexing_error	TEXT	NULL	Error message if indexing failed
file_size	BIGINT	DEFAULT 0	File size in bytes
page_count	INTEGER	DEFAULT 0	Number of pages (PDF)
uploaded_by_user_id	INTEGER	FOREIGN KEY	References users(id) ON DELETE SET NULL
created_at	TIMESTAMP	NOT NULL	Upload timestamp
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

7.2.4 Syllabus Tables

Table: syllabuses

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Syllabus identifier
subject_id	INTEGER	FOREIGN KEY, INDEX	References subjects(id) ON DELETE CASCADE
document_id	INTEGER	FOREIGN KEY	References documents(id) ON DELETE SET NULL
subject_name	VARCHAR(255)	NULL	Extracted subject name
subject_code	VARCHAR(50)	NULL	Extracted subject code
total_credits	INTEGER	DEFAULT 0	Total credit hours
extraction_status	VARCHAR(20)	DEFAULT 'pending'	'pending', 'processing', 'completed', 'failed'
extraction_error	TEXT	NULL	Error message if extraction failed
raw_extraction	TEXT	NULL	Raw LLM output for debugging
created_at	TIMESTAMP	NOT NULL	Creation timestamp
updated_at	TIMESTAMP	NOT NULL	Last update

Column	Type	Constraints	Description
			timestamp

Table: syllabus_units

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Unit identifier
syllabus_id	INTEGER	FOREIGN KEY, INDEX	References syllabuses(id) ON DELETE CASCADE
unit_number	INTEGER	NOT NULL	Unit number (1, 2, 3...)
title	VARCHAR(1000)	NOT NULL	Unit title
description	TEXT	NULL	Unit description
raw_text	TEXT	NULL	Verbatim text from syllabus
hours	INTEGER	DEFAULT 0	Teaching hours allocated
created_at	TIMESTAMP	NOT NULL	Creation timestamp
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

Table: syllabus_topics

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Topic identifier
unit_id	INTEGER	FOREIGN KEY, INDEX	References syllabus_units(id) ON DELETE CASCADE
topic_number	INTEGER	NOT NULL	Topic order within unit
title	VARCHAR(500)	NOT NULL	Topic title
description	TEXT	NULL	Topic description
keywords	TEXT	NULL	Comma-separated keywords
created_at	TIMESTAMP	NOT NULL	Creation timestamp
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

7.2.5 Chat System Tables

Table: chat_sessions

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Session identifier
user_id	INTEGER	FOREIGN	References users(id)

Column	Type	Constraints	Description
		KEY, INDEX	ON DELETE CASCADE
subject_id	INTEGER	FOREIGN KEY, INDEX	References subjects(id) ON DELETE CASCADE
title	VARCHAR(255)	NULL	Session title (auto-generated or user-set)
created_at	TIMESTAMP	NOT NULL	Session creation timestamp
updated_at	TIMESTAMP	NOT NULL	Last message timestamp

Table: chat_messages

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Message identifier
session_id	INTEGER	FOREIGN KEY, INDEX	References chat_sessions(id) ON DELETE CASCADE
subject_id	INTEGER	FOREIGN KEY, INDEX	References subjects(id) ON DELETE CASCADE
user_id	INTEGER	FOREIGN KEY, INDEX	References users(id) ON DELETE CASCADE
role	VARCHAR(20	NOT NULL	'user', 'assistant',

Column	Type	Constraints	Description
)			'system'
content	TEXT	NOT NULL	Message content
citations	JSONB	NULL	Array of citation objects from KB
tokens_used	INTEGER	DEFAULT 0	Token count for this message
model_used	VARCHAR(100)	NULL	AI model identifier (e.g., 'llama-3.3-70b')
response_time	INTEGER	DEFAULT 0	Response time in milliseconds
is_streamed	BOOLEAN	DEFAULT FALSE	Whether response was streamed via SSE
metadata	JSONB	NULL	Additional metadata
created_at	TIMESTAMP	NOT NULL	Message timestamp
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

7.2.6 System & Audit Tables

Table: external_api_keys (DEPRECATED)

Note: This table is deprecated. API keys now stored client-side.

Table: api_key_usage_logs

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Log identifier
user_id	INTEGER	FOREIGN KEY	References users(id) ON DELETE CASCADE
service	VARCHAR(50)	NOT NULL	Service name (e.g., 'firecrawl', 'tavily')
endpoint	VARCHAR(255)	NULL	API endpoint called
status_code	INTEGER	NULL	HTTP response status code
created_at	TIMESTAMP	NOT NULL	Log timestamp

Table: user_activities

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Activity identifier
user_id	INTEGER	FOREIGN KEY, INDEX	References users(id) ON DELETE CASCADE
action	VARCHAR(100)	NOT NULL	Action type (e.g., 'login', 'upload_document')

Column	Type	Constraints	Description
resource_type	VARCHAR(50)	NULL	Resource type affected
resource_id	INTEGER	NULL	Resource ID affected
details	JSONB	NULL	Additional activity details
ip_address	VARCHAR(45)	NULL	User IP address (IPv4/IPv6)
created_at	TIMESTAMP	NOT NULL	Activity timestamp

Table: admin_audit_logs

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Log identifier
admin_id	INTEGER	FOREIGN KEY	References users(id) ON DELETE CASCADE
action	VARCHAR(100)	NOT NULL	Admin action performed
target_type	VARCHAR(50)	NULL	Target resource type
target_id	INTEGER	NULL	Target resource ID
changes	JSONB	NULL	JSON of changes made

Column	Type	Constraints	Description
created_at	TIMESTAMP	NOT NULL	Log timestamp

Table: app_settings

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Setting identifier
key	VARCHAR(100)	UNIQUE, NOT NULL	Setting key
value	TEXT	NULL	Setting value
description	TEXT	NULL	Setting description
created_at	TIMESTAMP	NOT NULL	Creation timestamp
updated_at	TIMESTAMP	NOT NULL	Last update timestamp

Table: jwt_token_blacklist

Column	Type	Constraints	Description
id	SERIAL	PRIMARY KEY	Entry identifier
user_id	INTEGER	FOREIGN	References users(id)

Column	Type	Constraints	Description
		KEY	ON DELETE CASCADE
token_hash	VARCHAR(255)	UNIQUE, NOT NULL	SHA-256 hash of invalidated token
expires_at	TIMESTAMP	NOT NULL	Token expiration (for auto-cleanup)
created_at	TIMESTAMP	NOT NULL	Blacklist timestamp

7.3 Indexes and Performance

The database employs strategic indexing to optimize query performance. Primary indexes include B-tree indexes on all foreign keys (user_id, subject_id, session_id, course_id, semester_id), unique indexes on email addresses and codes, composite indexes on (user_id, created_at) for user activity queries, and partial indexes on indexing_status='pending' for background job processing. GIN indexes on JSONB columns (citations, metadata) enable fast containment queries using @> operator.

Connection pooling through pgx driver maintains 25-100 concurrent connections with maximum connection lifetime of 1 hour. Query optimization includes automatic prepared statement caching, N+1 query prevention through GORM preloading, and query result caching in Redis with 5-minute TTL for frequently accessed data.

7.4 Additional Tables

Beyond the core tables documented above, the database includes additional specialized tables for extended functionality:

Chat Memory Management: chat_memories (stores conversation context), chat_memory_batches (batch processing for memory optimization), and chat_compacted_contexts (compressed long-term memory storage for efficient retrieval).

PYQ (Previous Year Questions) System: pyq_papers (examination paper metadata), pyq_questions (individual questions extracted from papers), pyq_question_choices (multiple choice options for MCQ questions), pyq_crawler_sources (configured crawling sources), and pyq_crawled_papers (tracking of crawled papers with status).

Syllabus Management: The syllabi table (also referenced as syllabuses in the schema) works with syllabus_units and syllabus_topics to provide a hierarchical structure for curriculum management.

Background Processing: indexing_jobs (document indexing job queue) and indexing_job_items (individual items within indexing jobs) manage asynchronous document processing workflows.

User Engagement: user_notifications (system notifications and alerts for users) and user_courses (enrollment tracking between users and courses).

API Management: api_keys (user API key storage with encrypted values) and api_key_usage_logs (detailed tracking of API consumption for rate limiting and analytics).