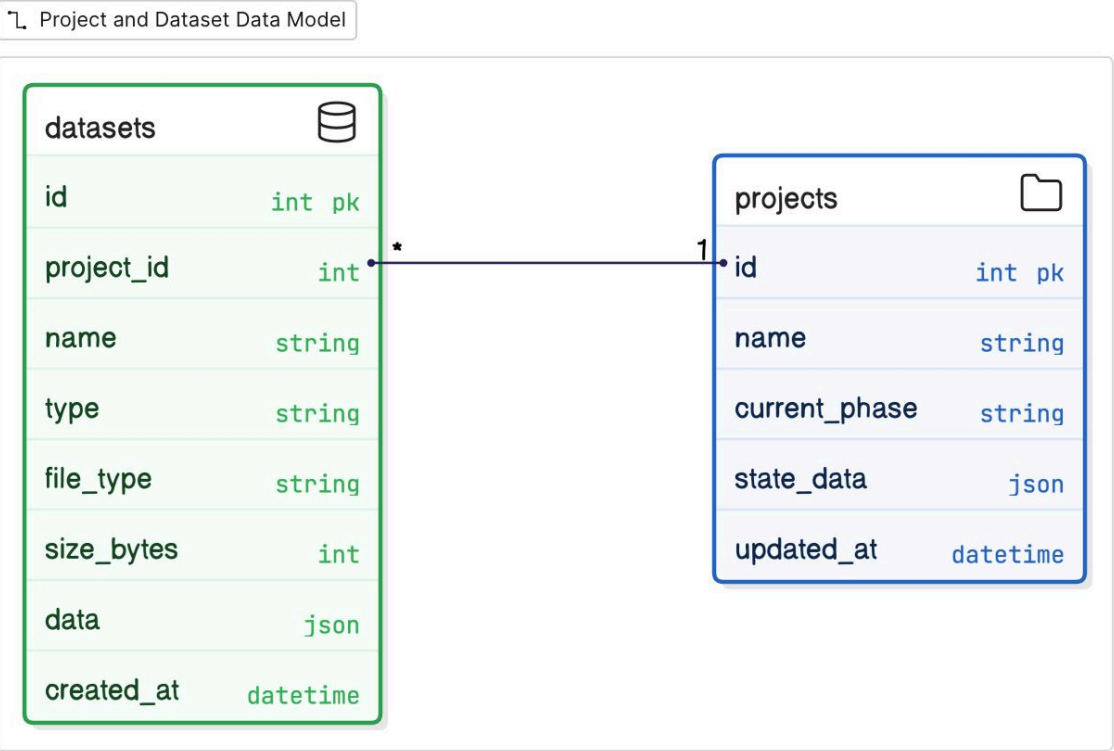


Easy_ACC: Visual document representing the database structure and data storage process

Entity Relationship Diagram



Workflow Logic

Project Initialization: A project must exist first to generate a `project_id`.

Dataset Linking: Every dataset saved via `save_dataset` is physically "stamped" with that `project_id`.

Scoped Selection: When you view a project, the system only fetches datasets where `Dataset.project_id == active_project_id`.

Phase 1: The Gatekeeper (Validation)

Before the database is even touched, the `validate_file_type` function checks the file extension.

Allowed: `.csv`, `.txt`, `.xlsx`

Rejected: Any other extension triggers an `UnsupportedFileTypeError`.

Phase 2: Atomic Transaction (The Save)

When `save_dataset` is called, the system performs a session transaction:

It checks for the existence of the parent `project_id`.

It calculates the `size_bytes` of the JSON payload.

It updates/inserts the dataset.

It "touches" the parent project's `updated_at` timestamp so the UI shows the project was recently modified.

If any step fails, the `session.rollback()` ensures no partial data is saved.

Phase 3: Cascade Cleanup

Because of the Foreign Key constraint with `onDelete="CASCADE"`, deleting a project automatically wipes all its datasets from the disk space.

Operational Documentation for Developers

Storage MonitoringUse `get_storage_usage()` to retrieve the total byte count of all datasets.

Use Case: Triggering UI alerts when the database nears a size limit (e.g., 2GB).

Data IsolationTo retrieve data for a specific project, always use the `project_id` filter:

```
[stmt = select(Dataset).where(Dataset.project_id == active_project_id)]
```

Warning Deleting a row from the Projects table will permanently and immediately delete all associated rows in the Datasets table.

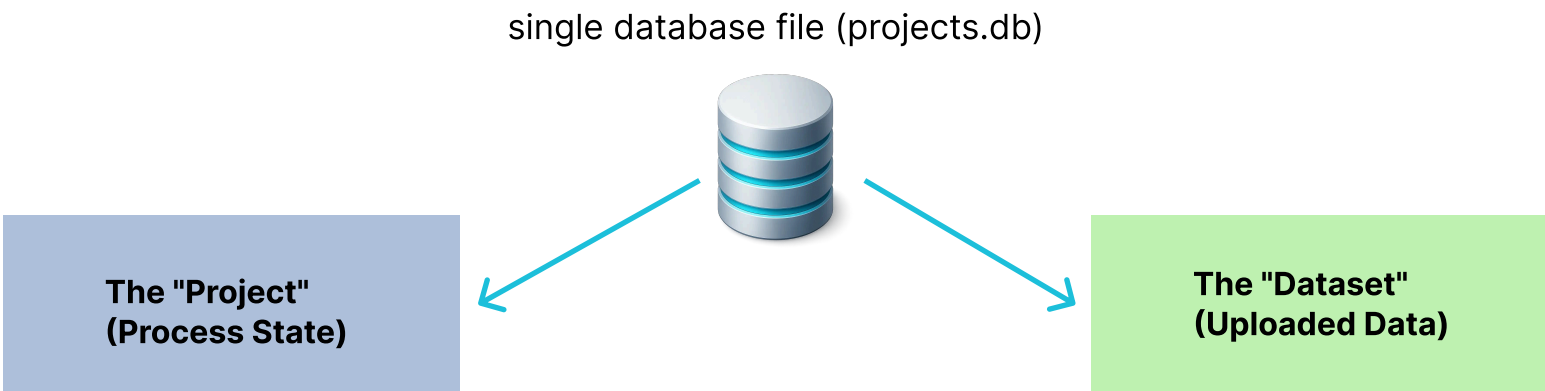


Table: projects

Purpose: Primary container for research sessions and workflow state.

Column	Type	Constraints	Description
id	Integer	PK	Unique identifier.
name	String	Unique, Not Null	Human-readable project name.
current_phase	String	-	Workflow status (e.g., EDA, Modeling).
state_data	JSON	-	Serialized UI/Process variables.
updated_at	DateTime	Auto-update	Timestamp of the last user activity.

Table: datasets

Purpose: Stores the actual data content, strictly linked to a parent project.

Column	Type	Constraints	Description
id	Integer	PK	Unique identifier.
project_id	Integer	FK	Reference to <code>projects.id</code> (Cascade on delete).
name	String	Not Null	Filename or internal label.
type	String	-	Category (e.g., "raw", "clean").
file_type	String	-	Restricted to: <code>csv</code> , <code>txt</code> , <code>xlsx</code> .
size_bytes	Integer	-	Data size for storage monitoring.
data	JSON	Not Null	The serialized content of the file.
created_at	DateTime	Default: Now	Timestamp of initial upload.