

BeaverDB: The Complete Guide

SYALIA S.R.L.

Table of contents

1	Introduction	3
1.1	What is BeaverDB?	3
1.2	The BeaverDB Philosophy	3
1.3	Ideal Use Cases	5
1.4	How This Guide is Structured	5
2	Quickstart Guide	6
2.1	Installation	6
2.2	Your First Example in 10 Lines	7
I	The User Guide	10
3	Key-Value and Blob Storage	11
4	Lists and Queues	12
5	The Document Collection (<code>db.collection</code>)	13
6	Real-Time Data	14
7	Concurrency	15
8	Deployment & Access	16
II	The Developer Guide	17
9	Core Architecture & Design	18
10	Concurrency Model	19
11	Search Architecture	20
12	Future Roadmap & Contributing	21

1 Introduction

Welcome to BeaverDB!

If you’ve ever found yourself building a Python application that needs to save some data, but setting up a full-blown database server felt like massive overkill, you’re in the right place. BeaverDB is designed to be the “Swiss Army knife” for embedded Python data. It’s built for those exact “just right” scenarios: more powerful than a simple pickle file, but far less complex than a networked server like PostgreSQL or MySQL.

1.1 What is BeaverDB?

At its heart, BeaverDB is a multi-modal database in a single SQLite file.

It’s a Python library that lets you manage modern, complex data types without ever leaving the comfort of a local file. The name itself tells the story:

B.E.A.V.E.R. stands for **B**ackend for mbedded, **A**ll-in-one **V**ector, **E**ntity, and **R**elationship storage.

This means that inside that one `.db` file, you can seamlessly store and query:

- Key-Value Pairs (like a dictionary for your app’s configuration)
- Lists (like a persistent to-do list)
- Vector Embeddings (for AI and semantic search)
- Documents & Text (for full-text search)
- Graph Relationships (to connect your data together)
- ...and much more.

All this power comes from building on top of the world’s most deployed database engine: SQLite.

1.2 The BeaverDB Philosophy

BeaverDB is built on a few core principles. Understanding these will help you know when and why to choose it for your project.

Robust, Safe, and Durable

Your data should be safe, period. BeaverDB is built to be resilient. Thanks to SQLite’s atomic transactions and Write-Ahead Logging (WAL) mode, your database is crash-safe.

If your program crashes mid-operation, your data is never lost or corrupted; the database simply rolls back the incomplete transaction.

Furthermore, it's designed for concurrency. It's both thread-safe (different threads can share one BeaverDB object) and process-safe (multiple, independent Python scripts can read from and write to the same database file at the same time). For tasks that require true coordination, it even provides a simple, built-in distributed lock.

Performant by Default

BeaverDB is fast. It's not just “fast for a small database”—it's genuinely fast for the vast majority of medium-sized projects. Because it's an embedded library, there is zero network latency for any query.

Let's be clear: if you're building the next X (formerly Twitter) and need to handle millions of documents and thousands of queries per second, you'll need a distributed, networked database. But for almost everything else? BeaverDB is more than fast enough. If your project is in the thousand to tens-of-thousands of documents range, you'll find it's incredibly responsive.

Local-First & Embedded

The default, primary way to use BeaverDB is as a single file right next to your code. This means your entire database—users, vectors, chat logs, and all—is contained in one portable file (e.g., `my_app.db`). You can copy it, email it, or back it up. This “local-first” approach is what makes it so fast and simple to deploy.

Minimal & Optional Dependencies

The core BeaverDB library has zero external dependencies. You can get started with key-value stores, lists, and queues right away.

Want to add vector search? Great! Install the `[vector]` extra, and BeaverDB will activate its faiss integration. Need a web server? Install the `[server]` extra, and it unlocks a fastapi-based REST API. This “pay-as-you-go” approach keeps your project lightweight.

Pythonic API

BeaverDB is designed to feel like you're just using standard Python data structures. You shouldn't have to write complex SQL queries just to save a Python dict or list. The goal is to make the database feel like a natural extension of your code.

Standard SQLite Compatibility

This is the “no-magic” rule. The `my_app.db` file that BeaverDB creates is a 100% standard, valid SQLite file. You can open it with any database tool (like DB Browser for SQLite) and see your data in regular tables. This ensures your data is never locked into a proprietary format.

Synchronous Core with Async Potential

The core library is synchronous, which makes it simple and robust for multi-threaded applications. However, BeaverDB is fully aware of the modern async world. For every data

structure, you can call `.as_async()` to get a fully awaitable version that runs its blocking operations in a background thread, keeping your asyncio event loop from getting blocked.

1.3 Ideal Use Cases

BeaverDB shines in scenarios where simplicity, robustness, and local performance are more important than massive, web-scale concurrency.

- **Local AI & RAG:** Perfect for building Retrieval-Augmented Generation (RAG) applications that run on your local machine. You can store your vector embeddings and their corresponding text right next to each other.
- **Desktop Utilities & CLI Tools:** The ideal companion for a custom tool that needs to remember user preferences, manage a history, or cache results.
- **Chatbots:** A persistent list is a perfect, simple way to store a chatbot’s conversation history for a user.
- **Rapid Prototyping:** Get your idea up and running in minutes. Start with a local `.db` file, and if your project grows, you can deploy it as a REST API without changing your application logic.

1.4 How This Guide is Structured

We’ve designed this documentation to get you the information you need, whether you’re building your first script or contributing to the core.

This guide is split into two main parts:

- **Part 1: The User Guide**

This is your starting point. After the Quickstart, this is where you’ll find an in-depth guide that walks you through how to use every single feature of BeaverDB. We’ll explore each “modality” one by one with practical examples.

- **Part 2: The Developer Guide**

This part is for power users and contributors. We’ll go under the hood to look at the why behind the design. This is where we do deep dives into the core architecture, the concurrency model (threading and locking), and the internals of how features like vector search are implemented.

2 Quickstart Guide

This is where the fun begins. Let's get BeaverDB installed and run your first multi-modal script. You'll be up and running in about 30 seconds.

2.1 Installation

BeaverDB is a Python library, so you can install it right from your terminal using pip.

The Core Install

If you just want the core features—like key-value dictionaries, lists, and queues—you can install the zero-dependency package.

```
# This has NO external dependencies
pip install beaver-db
```

This gives you all the core data structures and is perfect for many simple applications.

Installing Optional Features

BeaverDB keeps its core light by making advanced features optional. You can install them as “extras” as needed.

- `beaver-db[vector]`: Adds AI-powered vector search (using faiss).
- `beaver-db[server,cli]`: Adds the fastapi-based REST server and the beaver command-line tool.

For this guide, we recommend installing the `beaver-db[full]` package, which includes everything, so you can follow along with all the examples.

```
# To install all features, including vector search and the server
pip install "beaver-db[full]"
```

With that, you're ready to write some code.

2.2 Your First Example in 10 Lines

Let's create a single Python script that shows off BeaverDB's "multi-modal" power. We'll use three different data types—a dictionary, a list, and a document collection—all in the same database file.

Create a new file named `quickstart.py` and add the following:

```
from beaver import BeaverDB, Document

# 1. Initialize the database
# This creates a single file "my_data.db" if it doesn't exist
# and sets it up for safe, concurrent access.
db = BeaverDB("my_data.db")

# 2. Use a namespaced dictionary (like a Python dict)
# This is perfect for storing app configuration or user settings.
config = db.dict("app_config")
config["theme"] = "dark"
config["user_id"] = 123

# You can read the value back just as easily:
print(f"App theme is: {config['theme']}")

# 3. Use a persistent list (like a Python list)
# This is great for a to-do list, a job queue, or a chat history.
tasks = db.list("daily_tasks")
tasks.push({"id": "task-001", "desc": "Write project report"})
tasks.push({"id": "task-002", "desc": "Deploy new feature"})

# You can access items by index, just like a normal list:
print(f"First task is: {tasks[0]['desc']}")

# 4. Use a collection for rich documents and search
# This is the most powerful feature, combining data and search.
articles = db.collection("articles")

# Create a Document to store.
# We give it a unique ID and some text content.
doc = Document(
    id="sqlite-001",
    content="SQLite is a powerful embedded database ideal for local apps."
)

# 5. Index the document
# This not only saves the document but also automatically
```

```

# makes its text content searchable via a Full-Text Search (FTS) index
# with optional fuzzy matching.
articles.index(doc, fts=True, fuzzy=True)

# 6. Perform a full-text search
# This isn't a simple string find; it's a real search engine with fuzzy matching!
results = articles.match(query="database", fuzziness=1)

# The result is a list of tuples: (document, score)
top_doc, rank = results[0]
print(f"Search found: '{top_doc.content}' (Score: {rank:.2f})")

```

Here's a line-by-line explanation of what you just did:

- **from beaver import BeaverDB, Document** `BeaverDB` is the main class, your entry point to the database. A `Document` is a special data object used when you're working with `db.collection()`.
- **db = BeaverDB("my_data.db")** This is the most important line. It finds `my_data.db` or creates it if it's not there. It also automatically enables all the high-performance and safety features (like Write-Ahead Logging) so it's ready for use.
- **config = db.dict("app_config")** Here, you're asking `BeaverDB` for a dictionary. `"app_config"` is the "namespace." This means you can have *many* different dictionaries (`app_config`, `user_prefs`, `cache`, etc.) that won't interfere with each other. The `config` object you get back behaves just like a standard Python `dict`. When you do `config["theme"] = "dark"`, that change is instantly saved to the `my_data.db` file.
- **tasks = db.list("daily_tasks")** Same idea, but for a list. You get back a `tasks` object that acts like a Python `list`. You can **push** (append) items, get items by index (`tasks[i]`), or **pop** them. You can also insert and remove items at an arbitrary index, and it all works instantaneously (in CS terms, it's $O(1)$ for all operations).
- **articles = db.collection("articles")** This gets you a "collection," which is the most powerful data structure. A collection is designed to hold rich data, like articles, user profiles, or AI embeddings.
- **doc = Document(...)** To put something in a collection, you wrap it in a `Document`. Here, we give it a unique `id` and some `content`. You can add any other fields you want just by passing them as keyword arguments.
- **articles.index(doc, ...)** This is where the magic happens. When you call `.index()`, `BeaverDB` saves your document. But it *also* reads the `content` field and automatically puts all the words into a Full-Text Search (FTS) index and a clever fuzzy index, which are optional.
- **results = articles.match(query="database")** This line runs a search. Because `index()` already did the work, this query is fast. It searches the FTS index for the word "database" and finds your document.

When you run the script, you've created a single `my_data.db` file that now contains your config, your task list, and your searchable articles.

Part I

The User Guide

3 Key-Value and Blob Storage

Chapter Outline:

- **3.1. Dictionaries & Caching (db.dict)**
 - A Python-like dictionary interface: `config["api_key"] = ...`
 - Standard methods: `.get()`, `del`, `len()`, iterating with `.items()`.
 - **Use Case: Caching with TTL:** Using `.set(key, value, ttl_seconds=3600)`.
- **3.2. Blob Storage (db.blobs)**
 - Storing binary data (images, PDFs, files) with metadata.
 - API: `.put(key, data, metadata)`, `.get(key)`.
 - The Blob object: Accessing `.data` and `.metadata`.

4 Lists and Queues

Chapter Outline:

- **4.1. Persistent Lists (`db.list`)**
 - A full-featured, persistent Python list.
 - Full support for: `push`, `pop`, `prepend`, `deque`, slicing `my_list[1:5]`, and in-place updates `my_list[0] =`
- **4.2. Priority Queues (`db.queue`)**
 - Creating a persistent, multi-process task queue.
 - Adding tasks: `.put(data, priority=N)` (lower number is higher priority).
 - Consuming tasks: The blocking `.get(timeout=N)` method.
 - **Use Case:** A multi-process producer/consumer pattern.

5 The Document Collection (`db.collection`)

Chapter Outline:

- **5.1. Documents & Indexing**
 - The `Document` class: `id`, `embedding`, and `metadata`.
 - Indexing and Upserting: `.index(doc)` performs an atomic insert-or-replace.
 - Removing data: `.drop(doc)`.
- **5.2. Vector Search (ANN)**
 - Adding vectors via the `Document(embedding=...)` field.
 - Querying: `.search(vector, top_k=N)`.
 - **Use Case:** Building a RAG system by combining text and vector search.
 - **Helper:** The `rerank()` function for hybrid search results.
- **5.3. Full-Text & Fuzzy Search**
 - Full-Text Search (FTS): `.match(query, on=["field.path"])`.
 - Fuzzy Search: `.match(query, fuzziness=2)` for typo-tolerance.
- **5.4. Knowledge Graph**
 - Creating relationships: `.connect(source, target, label, metadata)`.
 - Single-hop traversal: `.neighbors(doc, label)`.
 - Multi-hop (BFS) traversal: `.walk(source, labels, depth, direction)`.

6 Real-Time Data

Chapter Outline:

- **6.1. Publish/Subscribe (`db.channel`)**
 - High-efficiency, multi-process messaging.
 - Publishing: `channel.publish(payload)`.
 - Subscribing: `with channel.subscribe() as listener: for msg in listener.listen(): ...`
- **6.2. Time-Indexed Logs (`db.log`)**
 - Creating structured, time-series logs: `logs.log(data)`.
 - Querying history: `.range(start_time, end_time)`.
 - **Feature:** Creating a live dashboard with `.live(window, period, aggregator)`.

7 Concurrency

Chapter Outline:

- **7.1. Inter-Process Locks (`db.lock`)**
 - Creating a critical section: `with db.lock("my_task", timeout=10): ...`
 - Guarantees: Fair (FIFO) and Deadlock-Proof (via TTL).
- **7.2. Atomic Operations on Data Structures**
 - Locking a specific manager: `with db.dict("config") as config: ...`
 - **Use Case:** Atomically getting and processing a *batch* of items from a queue.

8 Deployment & Access

Chapter Outline:

- **8.1. The REST API Server (`beaver serve`)**
 - Exposing your database as a FastAPI application.
 - Command: `beaver serve --database my.db --port 8000`
 - Accessing the interactive OpenAPI docs at `/docs`.
- **8.2. The Command-Line Client (`beaver client`)**
 - Interacting with your database from the terminal for admin and debugging.
 - Example: `beaver client --database my.db dict config get theme`
- **8.3. Docker Deployment**
 - Running the server in a container for stable deployment.
 - `docker run -p 8000:8000 -v $(pwd)/data:/app apiad/beaverdb`

Part II

The Developer Guide

9 Core Architecture & Design

Chapter Outline:

- **9.1. Guiding Principles (Developer Focus)**
 - A deeper dive into the “Why” from `design.md`.
 - Standard SQLite Compatibility as a “no-magic” rule.
 - Convention over Configuration.
- **9.2. The Manager Delegation Pattern**
 - How `BeaverDB` acts as a factory.
 - How managers (e.g., `DictManager`) are initialized with a reference to the core `BeaverDB` connection pool.
 - How all tables are prefixed with `beaver_` to avoid user-space conflicts.
- **9.3. Type-Safe Models (`beaver.Model`)**
 - Using the `model=...` parameter for automatic serialization and deserialization.
 - Inheriting from `beaver.Model` for a lightweight, Pydantic-compatible solution.

10 Concurrency Model

Chapter Outline:

- **10.1. Thread Safety (`threading.local`)**
 - How BeaverDB provides a unique `sqlite3.Connection` for *every thread*.
 - Why this is the key to preventing thread-related errors.
 - Enabling WAL (Write-Ahead Logging) for concurrent reads.
- **10.2. Inter-Process Locking (The Implementation)**
 - How `db.lock()` works under the hood.
 - The `beaver_lock_waiters` table as a fair (FIFO) queue.
 - The `expires_at` column as a deadlock-prevention (TTL) mechanism.
- **10.3. The Asynchronous `.as_async()` Pattern**
 - How the `Async...Manager` wrappers are implemented.
 - Using `asyncio.to_thread` to run blocking I/O without blocking the event loop.

11 Search Architecture

Chapter Outline:

- **11.1. Vector Search (ANN) Internals**
 - The “Hybrid Index System”: Base Index and Delta Index.
 - **Crash-Safe Logging:** How additions and deletions are written to SQLite logs (`_beaver_ann_...` tables).
 - **Background Compaction:** The `compact()` process.
- **11.2. Text Search (FTS & Fuzzy) Internals**
 - **FTS:** How `beaver_fts_index` is a `fts5` virtual table.
 - **Fuzzy Search:** How BeaverDB builds a custom trigram index (`beaver_trigrams` table) and uses Levenshtein distance.

12 Future Roadmap & Contributing

Chapter Outline:

- **12.1. The Future of BeaverDB**
 - The `BeaverClient` as a drop-in network client.
 - Replacing `faiss` with a simpler, pure-numpy linear search.
- **12.2. How to Contribute**
 - (Standard contribution guidelines, linking to Makefile, etc.)