OmniFolio

Database & Storage Analysis Report

Generated: February 18, 2026

This report quantifies all database tables, caches, and file storage used by OmniFolio across its data domains: SEC EDGAR, insider sentiment, government spending, LDA lobbying, financial calendars, user portfolio data, community features, and more. Estimates are provided for development baseline, 1,000 users at 12 months, and 10,000 users at 12 months.

1. SEC EDGAR Data

Data sourced from SEC EDGAR (XBRL financials, Form 4 insider transactions, 13F institutional holdings, filing text). Shared across all users — grows only when more symbols are tracked, not with user count.

Table	Avg Row Size	Monthly Growth	Current Est.	12-Month Projection
sec_companies	~300 B	One-time load (10K+ companies)	3–4 MB	4 MB (stable)
sec_filings	~500 B	~200/symbol × 15 symbols	15 MB	50–80 MB
sec_financials	~2 KB	~8 rows/quarter/symbol (70+ NUMERIC cols)	5 MB	20–40 MB
sec_insider_transactions	~400 B	~50/symbol/month	3 MB	15–25 MB
sec_institutional_holdings	~350 B	~500/fund × 3 funds/quarter	2 MB	10–15 MB
sec_filing_sections	5–50 KB	Sporadic (full plain_text + keywords)	20 MB	**100–300 MB ■■**
sec_watchlist	~200 B	Per active user	< 1 MB	< 5 MB
sec_filing_alerts	~300 B	Per filing event/user	< 1 MB	~10 MB
sec_cache_refresh_log	~150 B	1 row/cache key	< 0.1 MB	< 0.1 MB

■■ *sec_filing_sections stores full extracted text (Risk Factors, MD&A) — 15 symbols × 20 filings × 5 sections can reach 750 MB if unconstrained. Recommend capping stored text at 10,000 characters per section.*

SEC subtotal: ~48 MB now → ~200–460 MB at 12 months

2. Proprietary Sentiment & Analysis Caches

Computed from SEC Form 4 and 10-Q/10-K XBRL filings using OmniFolio's proprietary OIC (Insider Confidence) and OES (Earnings Surprise) scoring algorithms. Smart TTL between 2h–72h keeps sizes bounded.

Table	Avg Row Size	Growth Model	12-Month Est.
insider_sentiment_cache	~400 B	1 row/symbol/month × symbols tracked	2–5 MB
insider_sentiment_transactions	~500 B	~50 txns/symbol/month × 15 symbols	5–10 MB

Table	Avg Row Size	Growth Model	12-Month Est.
insider_sentiment_refresh_log	~200 B	Cleaned after 30 days (bounded)	< 1 MB
earnings_surprises_cache	~1.2 KB	4 rows/symbol/year (50+ columns)	3–5 MB
earnings_estimates_history	~200 B	Snapshot per analyst revision	1–2 MB
earnings_surprises_refresh_log	~200 B	Cleaned periodically	< 1 MB

Sentiment subtotal: ~12 MB now → ~12–24 MB at 12 months

3. Government & Public Data Caches (Senate LDA, USAspending.gov)

100% public government data. Senate Lobbying Disclosure Act (LDA) filings via lda.senate.gov, federal contract awards via usaspending.gov. Both caches use weekly TTL (168h) and 30-day log cleanup.

Table	Avg Row Size	Growth Model	12-Month Est.
senate_lobbying_cache	~800 B	~20 filings/symbol × 15 symbols (JSONB arrays)	5–10 MB
senate_lobbying_refresh_log	~200 B	Cleaned after 30 days	< 1 MB
usa_spending_cache	~600 B	~50 awards/symbol × 15 symbols	5–10 MB
usa_spending_refresh_log	~200 B	Cleaned after 30 days	< 1 MB

Government data subtotal: ~12 MB now → ~12–22 MB at 12 months

4. Calendar & News Caches

IPO and earnings calendars sourced from SEC EDGAR (S-1/10-Q filings). Economic calendar uses public macroeconomic event feeds. Twitter/X feed cache rolls over every 15 days.

Table	Avg Row Size	Growth Model	12-Month Est.
ipo_calendar_cache	~500 B + JSONB	~200 IPOs/year	2–5 MB
earnings_calendar_cache	~400 B + JSONB	~5,000 earnings reports/quarter	10–20 MB
economic_calendar_cache	~300 B	~50 events/week × 52 weeks	1–2 MB
twitter_feed_cache	~1 KB	Rolling 15-day window	5–10 MB
crypto_fear_and_greed	~100 B	1 row/day (daily update)	< 0.1 MB
ipo_calendar_meta	~100 B	Few config rows	< 0.1 MB
earnings_calendar_meta	~100 B	Few config rows	< 0.1 MB
economic_calendar_meta	~100 B	Few config rows	< 0.1 MB
cache_metadata	~100 B	1 row per cache name (static)	< 0.1 MB

Calendar/News subtotal: ~18 MB now → ~18–37 MB at 12 months

5. User Portfolio Data

All user-owned financial data: holdings, transactions, snapshots. Scales linearly with user count. price_snapshots is bounded by a 48-hour auto-cleanup function.

Table	Avg Row Size	Growth per User	Per 1K Users (12 mo)
encrypted_state_snapshots	2–50 KB	1 row/user (upserted, E2E encrypted)	2–50 MB
portfolio_snapshots	~500 B	1 snapshot/day × 365	180 MB
price_snapshots	~100 B	Cleaned every 48 hours	5–10 MB (rolling)
users / profiles	~500 B	1 row/user	0.5 MB
user_subscriptions	~300 B	1 row/user	0.3 MB
user_usage	~300 B	1 row/user/day × 365	110 MB
cash_accounts	~200 B	Max 10–50 entries	10 MB
savings_accounts	~200 B	Max 10–50 entries	10 MB
crypto_holdings	~300 B	Max 10–50 entries	15 MB
stock_holdings	~300 B	Max 10–50 entries	15 MB
crypto_transactions	~300 B	Moderate	30 MB
stock_transactions	~300 B	Moderate	30 MB
trading_accounts	~200 B	Max 10–50 entries	10 MB
real_estate	~400 B	Few per user	5 MB

Table	Avg Row Size	Growth per User	Per 1K Users (12 mo)
valuable_items	~300 B	Few per user	5 MB
expense_categories	~200 B	10–20 per user	4 MB
income_sources	~200 B	Few per user	2 MB
tax_profiles	~300 B	1 per user	0.3 MB
exchange_rates_history	~100 B	30 currencies × 365 days (shared)	5 MB
user_currency_preferences	~100 B	1 row/user	0.1 MB

User data per 1,000 users: ~430 MB → ~4.3 GB for 10,000 users

6. Community Data

Social features: posts, comments, likes, follows, hashtags. Comments can grow 5× faster than posts if engagement is high.

Table	Avg Row Size	Growth Model	Per 1K Users (12 mo)
posts	~500 B	~10 posts/user/month × 12 months	60 MB
comments	~300 B	~5 comments/post	180 MB
post_likes	~50 B	Variable engagement	10 MB
follows	~50 B	Variable	5 MB
hashtags	~100 B	Unique hashtag registry	1 MB
post_hashtags	~60 B	~3 tags/post	4 MB

Community subtotal per 1,000 users: ~260 MB → ~2.6 GB for 10,000 users

7. File Storage (Supabase Storage Buckets)

Object storage for binary files. post-images is the dominant growth driver and should have upload size limits enforced (recommend max 2 MB per image, compressed to 800px width server-side).

Bucket	Avg File Size	Growth	Per 1K Users (12 mo)
avatars	100–500 KB	1 avatar/user	100–500 MB
post-images	200 KB–2 MB	~5 images/user/month × 12 months	**1–10 GB ■■**

■■ *post-images is the largest wildcard. Without compression, 10K active users posting images could generate 10–100 GB of file storage.*

8. Grand Summary

Database totals include ~20% overhead for indexes. File storage totals are separate from database storage.

Category	Dev / Baseline	1K Users @ 12 Mo	10K Users @ 12 Mo
SEC EDGAR (DB)	~48 MB	200–460 MB	200–460 MB (shared)
Sentiment & Analysis (DB)	~12 MB	12–24 MB	12–24 MB (shared)
Gov Data / LDA (DB)	~12 MB	12–22 MB	12–22 MB (shared)
Calendar & News (DB)	~18 MB	18–37 MB	18–37 MB (shared)
User Portfolio Data (DB)	< 1 MB	~430 MB	~4.3 GB
Community (DB)	< 1 MB	~260 MB	~2.6 GB
Reference & Config (DB)	~10 MB	~25 MB	~200 MB
Index overhead (~20%)	~20 MB	~190 MB	~1.5 GB
DB Total	**~121 MB**	**~1.2–1.5 GB**	**~8–10 GB**
File Storage (Buckets)	~10 MB	1–10 GB	10–100 GB ■■
Grand Total	**~131 MB**	**~2–12 GB**	**~18–110 GB**

9. Key Insights & Recommendations

■■ **sec_filing_sections**	Full extracted text (Risk Factors, MD&A;) can balloon to 750 MB+ unconstrained. Cap stored text to 10,000 characters per section or store AI-generated summaries instead.
■■ **post-images bucket**	Without limits this dominates file storage. Enforce max 2 MB upload size and resize/compress server-side on ingest (you already handle avatar uploads via API — apply the same pattern).
■ **price_snapshots is well-managed**	The 48-hour cleanup function keeps this table at a constant ~5–10 MB regardless of user count.
■ **Shared data doesn't scale with users**	SEC, government, and calendar caches (~300–550 MB) are shared across all users. This is a significant architectural advantage.
■ **Smart TTL system**	The dynamic TTL logic (OIC/OES caches, insider sentiment) adapts to market hours, filing activity, and weekends — preventing unbounded growth.
■ **portfolio_snapshots growth**	At 1 snapshot/day/user, this table reaches 180 MB per 1K users/year. Consider weekly snapshots for FREE/BASIC users and daily only for PRO+.
■ **user_usage table**	365 rows/user/year at 300 B = 110 MB per 1K users. Aggregate older records monthly (sum counts) after 90 days to reduce row count.
■ **Supabase plan recommendation**	Supabase Free tier provides 500 MB DB + 1 GB file storage — sufficient only for early dev. You will need the Pro plan ($25/mo → 8 GB DB + 100 GB storage) before reaching ~500 active users. Plan for Team tier ($599/mo → 100 GB DB) at ~5K users.

All figures are estimates based on schema analysis and typical usage patterns. Actual sizes depend on content length (especially JSONB fields and free-text columns), user activity level, and cleanup job execution frequency. Re-run this analysis after each major schema change.