The PRIMOS Database: Prime-Product Addressing and GCD Similarity

Pedro Henrique Correa Garcia
Joinville, Santa Catarina, Brazil
garcia.pedro.wow@gmail.com

April 2026

Abstract

A configuration database whose keys are prime products inherits, directly from the Fundamental Theorem of Arithmetic, a set of operational properties that no hash-based or vector-based database can reproduce: unique decomposition, $O(k)$ similarity via GCD, collision-free addressing, discrete exact representation (no floating-point), and natural compositional algebra. We describe the core architecture of the PRIMOS database, show how each property follows from arithmetic rather than engineering choice, and note the existence of over thirty deployed software systems that use this architecture across education, agriculture, authentication, music composition, materials science, and natural-language processing.

1 Introduction

Most databases address their contents by arbitrary identifiers: hash keys, sequential integers, UUIDs, or embedding vectors. None of these addressing schemes carry semantic information intrinsically; the semantics live elsewhere — in joined tables, in metadata columns, in the similarity structure of an external model.

PRIMOS takes a different approach. It addresses every record by the *prime factorization* of its content. The address is the semantics, and the semantics is the address. The consequences of this identification are the subject of the present article.

Key idea

The address is the content. The content is the address. Every operation on the database is an operation on integers.

2 Prime-product addressing

Definition 2.1 (Prime-product address). *A* prime-product address *is a positive integer*

$$n = \prod_i P_i^{e_i}$$

where the e_i are non-negative integers and all but finitely many are zero. The factorization is unique by the Fundamental Theorem of Arithmetic.

Two records are *equal* if and only if their addresses are equal as integers, which by the Fundamental Theorem is equivalent to their prime factorizations being identical.

Proposition 2.2 (Collision-free addressing)**.** *Prime-product addressing is collision-free: distinct semantic content produces distinct addresses, and no hash collisions are possible.*

Proof. Follows immediately from the uniqueness of prime factorization. Two distinct factorizations give distinct integers, so two distinct semantic contents give distinct addresses. □

Remark 2.3. *This is not a probabilistic claim. A hash-based address space can claim collision-freeness only up to its hash output length; PRIMOS's addressing is collision-free absolutely, as a theorem of number theory.*

3 GCD similarity

The greatest common divisor of two integers carries, in the prime-product reading, the structural overlap between two records.

Definition 3.1 (GCD similarity)**.** *The* similarity *of two records with addresses a and b is*

$$s(a, b) \;=\; \gcd(a, b),$$

which in the prime-product reading is the integer whose prime factorization consists of the minimum of the two exponents at each prime:

$$\gcd\left(\prod_i P_i^{e_i}, \prod_i P_i^{f_i}\right) \;=\; \prod_i P_i^{\min(e_i, f_i)}.$$

Proposition 3.2 (GCD is $O(k)$)**.** *The similarity $s(a, b)$ can be computed in time $O(k)$ where k is the number of primes in the larger of the two factorizations.*

Proof. For each prime in the support of $a \cdot b$, take the minimum of its exponents in a and b. This is a linear pass over the support, which is bounded by k. □

Remark 3.3. *A vector-embedding similarity (cosine, dot product, Euclidean distance) is generally $O(d)$ where d is the embedding dimension, with d typically in the hundreds or thousands. PRIMOS similarity is $O(k)$ with k typically a small constant (below 33 for the first-half series, since primes above $P_{33} = 137$ do not participate in the first-half ladder). The cost difference is not marginal; it is categorical.*

Key idea

Similarity is a single GCD. No embedding, no distance metric, no training, no approximation. Two records with shared structure have a shared prime product, and that product is the similarity.

4 Compositional algebra

The lattice of prime-product addresses carries three natural operations, each with a clean semantic reading.

Proposition 4.1 (Three operations on the address lattice)**.** *Let a, b be two prime-product addresses.*

1. ***Compose:*** *$a \cdot b$ adds the exponents of each prime. Semantically, this is "combine the two records into a composite whose structure is the sum of the parts."*

2. ***Meet:*** *$\gcd(a, b)$ takes the minimum of each exponent. Semantically, this is "the shared structure of the two records."*

3. ***Join:*** *$\mathrm{lcm}(a, b)$ takes the maximum of each exponent. Semantically, this is "the structure that contains both."*

Each operation is $O(k)$ in the number of primes and requires no floating-point arithmetic. The lattice $(\mathbb{N}_{>0}, \gcd, \mathrm{lcm})$ is a distributive lattice, and PRIMOS inherits its entire order theory.

Takeaway

Composition is multiplication, meet is GCD, join is LCM, and all three are exact integer operations with no loss of precision. No other database architecture gets this for free from the content addressing alone.

5 Discrete exact representation

A consequence of prime-product addressing that is easy to miss but operationally important: PRIMOS does not use floating-point numbers anywhere in its core. Every address is an integer, every similarity is an integer GCD, every composition is an integer product. There is no accumulated numerical error, because there is no floating-point arithmetic to accumulate error in.

Proposition 5.1 (No floating-point core)**.** *The core operations of PRIMOS (address equality, similarity, composition, meet, join) are definable over $\mathbb{N}_{>0}$ alone and require no floating-point arithmetic.*

Remark 5.2. *The discrete-exact property matters for three kinds of application: reproducibility (results do not depend on floating-point rounding order), verifiability (every intermediate value can be checked exactly), and auditability (a decision trace is a sequence of integer operations with no information loss).*

6 Sizing from the Sperner bound

The configuration database has to be sized to hold every valid record that the underlying manifold supports. From Article 9, the count of valid configurations of the rung-13 gear assembly is bounded above by the Sperner quantity $2^{\binom{13}{6}} = 2^{1716}$, and the full rung-13 meta-manifold refinement lifts this to approximately 2^{2370}.

Proposition 6.1 (PRIMOS sizing)**.** *The PRIMOS database is sized to hold approximately* 2^{2370} *distinct configurations, corresponding to the Sperner-bound ceiling of the rung-*13 *meta-manifold with label refinement (Article 9, Proposition 1,* Database size at $k = 13$*).*

This sizing is not a design target set by engineering constraints; it is the theorem-forced ceiling of the first-half series. No configuration exists outside this set (that would require content above rung 13, which is archival and therefore not directly addressable), and no configuration inside the set can be collapsed or de-duplicated (because each antichain is genuinely distinct by the no-resonance theorem of Article 9).

Key idea

PRIMOS holds exactly the states the first-half framework supports. No more, no less. The database is the manifold, at the resolution of the ladder.

7 Applications

The architecture has been instantiated in over thirty software systems. We list the categories without going into the specifics of any individual system; each category corresponds to a rung or group of rungs on the ladder.

- **Education.** Curriculum addressing by prime-product semantic keys, with GCD similarity as the "related-topic" metric. Deployed in AI literacy and science curricula.
- **Agriculture.** Soil and crop configuration records keyed by prime products encoding nutrient profiles, climate, and rotation history. GCD similarity identifies interchangeable management strategies.
- **Authentication.** Topology-based authentication whose identity tokens are prime-product addresses. Identity comparison reduces to GCD similarity; no embedding model required.
- **Music composition.** Harmonic analysis and generation via prime-factored interval structure. Each harmonic progression has a prime-product address, and chord similarity is GCD on the addresses.
- **Materials addressing.** Configurational-mass addressing for composite objects, using the additive and multiplicative reading pair of Article 8. Similarity reduces to GCD on the multiplicative coordinate.
- **Natural-language processing.** Semantic addressing of discourse units by prime-product factorizations derived from concept decomposition. Similarity is GCD; composition is multiplication.

Each application uses the same core architecture. The differences between applications live in the encoding layer (how a domain object maps to a prime-product address), not in the database itself.

Remark 7.1. *This is the strongest available argument for the framework's correctness: the same arithmetic core, with no per-application re-engineering, powers systems in six distinct domains. If the architecture were merely a numerological coincidence, it would not survive the transition between domains. It does.*

8 What PRIMOS is not

- PRIMOS is not a vector database. There are no embeddings, no distance metrics, no nearest-neighbour search in the learned-representation sense.
- PRIMOS is not a relational database. It has no foreign keys or join tables; relationships are expressed by the shared prime structure of the addresses.
- PRIMOS is not an AI model. It contains no learned parameters, no training loop, and no gradient descent. The similarity structure is a theorem of number theory, not an outcome of optimisation.
- PRIMOS is not a graph database. Edges in PRIMOS are implicit in the GCD lattice of the addresses; no edge storage is required.

Takeaway

PRIMOS is the Fundamental Theorem of Arithmetic, operationalised. Every feature of the architecture is a consequence of unique factorization, not an engineering decision made for performance or convenience.

9 Summary

PRIMOS is a configuration database whose records are addressed by prime factorization. The Fundamental Theorem of Arithmetic supplies collision-free addressing; the GCD supplies $O(k)$ similarity; the lattice structure of $(\mathbb{N}_{>0}, \gcd, \mathrm{lcm})$ supplies a distributive compositional algebra; and the absence of floating-point arithmetic supplies discrete exact representation with full reproducibility and auditability. The database is sized to hold approximately 2^{2370} distinct configurations, the Sperner-bound ceiling of the rung-13 meta-manifold (Article 9).

More than thirty deployed software systems use this architecture across education, agriculture, authentication, music, materials, and NLP. Each application provides its own encoding layer from a domain object to a prime-product address; the database core is the same in every case. The persistence of a single arithmetic core across unrelated domains is the strongest available existence proof that the framework is not merely a formal curiosity.