Alphabets, Grammars, Calculators, and the End of Hand-Crafted Systems

Stratos Idreos
Harvard University
stratos@seas.harvard.edu

ABSTRACT

The AI revolution is transforming every scientific field and business sector, driving an unprecedented demand for data-centric computation. As new data types, hardware platforms, and workloads appear faster than ever before, the backbone systems that power this revolution must evolve just as quickly. Yet a single system architecture, whether tuned for computing analytics, generative AI, or machine learning, faces a design space larger than 10^{100} alternatives, and we still cling to a handful of "good" templates that each require years of manual design and implementation tuning. It is time to abandon this artisanal practice and embrace self-designing systems: systems that can reason about and refactor their own architecture. We show that by modeling the design space of systems as an alphabet of low-level design primitives and whole architectures as sentences in a grammar over that alphabet, "systems calculators" can now synthesize fresh systems blueprints on demand. The Data Calculator explores trillions of previously unknown data-structure variants to pick an optimal layout; Cosine and Limousine generate novel NoSQL stores that run up to three orders of magnitude faster than today's best deployments; the Image Calculator co-designs entirely new storage formats and neural networks to speed vision pipelines by 10; and LegoAI and TorchTitan invent novel distributed training algorithms for large AI models that extract every flop and byte from modern accelerators. These results signal a future in which systems research increasingly focuses on crafting richer alphabets and grammars while machines write the sentences, freeing designers and researchers to pursue more profound questions and enabling practitioners to dial in cost, latency, and accuracy with surgical precision.

PVLDB Reference Format:
Stratos Idreos, Harvard University. Alphabets, Grammars, Calculators, and the End of Hand-Crafted Systems. PVLDB, 18(12): 5537 - 5537, 2025.
doi:10.14778/3750601.3760522

BIOGRAPHY

Stratos Idreos is the Gordon McKay Professor of Computer Science at Harvard's John A. Paulson School of Engineering and Applied Sciences and serves as Faculty Co-Director of the Harvard Data Science Initiative. Stratos leads DASlab, the Harvard Data Systems Laboratory. His research pursues a "grammar of data systems," enabling machines—not humans—to design and tune systems architectures, resulting in systems that are tailored to their context, faster and more scalable. Stratos's work has been recognized by the community with honors such as the ACM SIGMOD Jim Gray Dissertation and ERCIM Cor Baayen awards (2011), IEEE TCDE Rising Star (2015), NSF CAREER and DOE Early Career awards, the ACM SIGMOD Contributions Award (2020) and Test-of-Time Award (2022), as well as a Sloan Research Fellowship and Harvard's McDonald Mentoring Award (2023). He has co-chaired ACM SIGMOD 2021 and IEEE ICDE 2022, co-founded the ACM/IMS Journal of Data Science, and currently serves as the chair of the ACM SoCC Steering Committee.

Proceedings of the VLDB Endowment, Vol. 18, No. 12 ISSN 2150-8097.
doi:10.14778/3750601.3760522