



SELF-SEARCHING STORAGE

A NEW TECHNOLOGY FOR BIG DATA ANALYSIS

BIG DATA ANALYSIS has become so overwhelming it risks passing into the philosophical realm of noisy trees falling in forests.

“If an event exists in your data but you can’t find it, is it really there?”



SUMMARY

NPUsrch™ is a novel, regular expression-based search technology that addresses the escalating mismatch between data accumulation and search capacity. NPUsrch completes rapid, deterministic analysis of petabytes of storage data in minutes without indexing.

THE CHALLENGE OF ANALYZING RAW DATA

Content analysis of raw data is challenging. Perfected search algorithms are too computationally intensive to run quickly. I/O, networking, and compute bottlenecks constrict execution rates. Improved efficiency requires extensive indexing and costly infrastructure. Indexed searches with sophisticated indexing tools and search accelerators constitute today’s usual approach to this extremely important problem.

INHERENT COMPROMISE OF INDEXING

Indexing is fundamentally a compromise and its limitations are brutally exposed by massive data collections. Indices are relatively expensive to build. They may occupy more storage space than the raw data they support. Valuable contextual information is lost. An initial assessment of data utility may limit future value. And catastrophically, indexing does not scale with data volume. Companies with multi-petabytes of storage struggle to afford the computational costs to utilize their own data.

BALANCING PERFORMANCE, COST AND VALUE

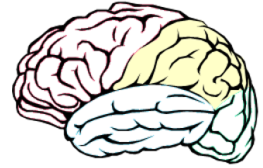
The technical ability to rapidly analyze all applicable content in a data lake for sparse events is fundamental to monetization of big data. Search algorithms run on raw data with traditional methods are computationally intensive and slow. Compute is expensive. Performance varies as a function of both data characteristics and expression complexity, generating long and unpredictable run times.

FULL CONTENT SEARCH REMAINS IDEAL

Petabytes of data accumulate daily and search technology struggles to keep up. AI and machine learning solutions are hampered by limitations of indexing and search. The anticipated benefits of big data will remain elusive as long as content search remains expensive and slow.

NEUROMORPHIC ADVANTAGE

BIG DATA ANALYSIS needs a novel technology capable of rapid, deterministic content search of raw data files without indexing.



SUMMARY

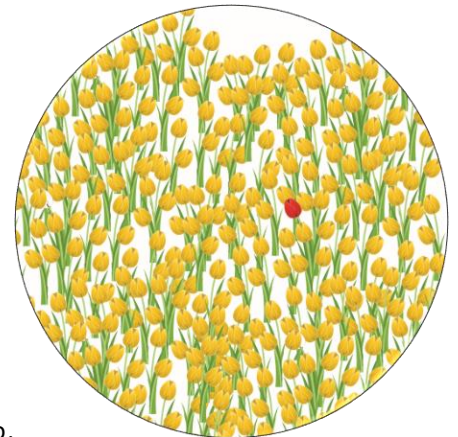
The Neuromorphic Processing Unit (NPU) is a novel hardware technology developed by Lewis Rhodes Labs. Modeled after the human brain, the NPU addresses technical challenges of search from a radically different perspective, providing an entirely new capacity to search petabytes of raw data in minutes.

FEATURES OF SENSORY PROCESSING

Human brains continuously process all sensory input, simultaneously providing real time awareness of sounds, sights, smells, taste and touch. Human reality depends on it. Human senses pass all the information they receive into the brain. Brain regions link together into networks that integrate and analyze input to determine relative usefulness. The increasingly complex hierarchy facilitates efficient, high resolution analysis.

HUMAN CAPACITY FOR PATTERN MATCHING

A human brain recognizes the single red flower in a bouquet of three with the same accuracy and latency as it perceives the one red flower seen in a field of thousands. Brain parallelism and hierarchy together support a constant rate and quality of analysis in the presence of widely varying amount of sensory input.



BRAIN ARCHITECTURE IN THE NPU

Parallelism and hierarchical architecture can empower computer systems too. The extraordinary pattern matching capacity of the NPU allows *NPUsearch* to complete simple or highly complex queries in the same search time.



No amount of thinking overheats your brain.

ENERGY EFFICIENCY

The human brain accomplishes all functions within tight energy and heat requirements. Low heat and low power are benefits of brain architecture.

NPUsearch can be integrated into computational storage because the NPU is an exceptionally low power device that emits very little heat. Even when functioning at full capacity, the NPU is an extremely low power device.

OVERCOMING OBSTACLES TO RAPID, DETERMINISTIC SEARCH

BIG DATA ANALYSIS needs fully scalable, self-searching storage.
 “If you can find it, you can use it!”

SUMMARY

NPUsearch is currently available in a 96 TB self-searching storage appliance behind a regex-accessible, python-based API. Full content search of data requires no indexing and completes in minutes. Multiple appliances running in parallel scan petabytes of data with no increase in search time.

NO I/O AND NETWORK BOTTLENECKS

Content search on unindexed data lakes can take weeks or months with traditional search technology. NPUsearch sits in storage, where the data is. Queries are completed locally. Only files with the requested data get moved across to the CPU, greatly reducing network and compute infrastructure demands. NPUsearch optimizes compute usage by processing only the data of interest.

INCREASE STORAGE, SAME SEARCH TIME

NPUsearch appliances are fully scalable. Additional units may be combined into petabyte racks of self-searching storage with no increase in search time, network demands, or CPU costs. NPUsearch provides a novel solution to an unsolved problem.

FAST AND PREDICTABLE

Content search 96 TB of raw data files in under 25 minutes.
 Simple queries. Under 25 minutes.

Content search 960 TB of raw data files in under 25 minutes.
 Complex queries. Under 25 minutes.

Content search 10 PB of raw data files in under 25 minutes.
 Multiple, massively complex queries. Under 25 minutes.

INVISIBLE TO USER

There is no new interface to learn. A neuromorphic driver and compiler translate the NPUsearch technology to a familiar interface. Data analysts can run standard regular expression searches through a familiar python API such as Jupyter notebook. Neuromorphic technology is invisible to the user.

You could run that query overnight,



or you can NPUsearch it while you get a coffee.



(if it's close by)



SYSTEM SPECIFICATIONS

NPUsearch APPLIANCE

	2U Rack Server (H: 3.42", W: 17.09", D:28.99")
Dimensions	(H: 3.42", W: 17.09", D:28.99")
Central Processor (CPU)	Dual 2 nd gen AMD EPYC™ 7282
Neuromorphic Processor (NPU)	24 x NPUsearch FPGA, 68 GB/s
Search time	25-minute maximum
Memory	DDR4: 64 GB RAM
Storage	96 TB (24 x 4TB NVME SSDs)
Management	1 GbE
Network Connectivity	Dual port 10/25GbE
Power supply	Dual hot-plug 230V, 2400W redundant
OS	Ubuntu 18.04 LTS

NPUsearch APPLIANCE RACK

Multiple appliance system. Racks provide capacity for petabytes of self-searching storage with constant search latency. Contact info@lewis-rhodes.com for more information.



“In our comparisons of the NPUsearch appliance versus our internally-optimized system, a data lake search requiring 10 hours with the current method completed in 5 minutes with NPUsearch.” - DOE customer

