Analyst Paper

Revolutionizing AI Infrastructure: An In-Depth Exploration of the VAST DataBase and VAST Data Platform

Sumit Pal

Table of Contents

Introduction **3**

Why Should Organizations Look Beyond Traditional Data Systems in the AI Era? **3**

The VAST DataBase **4**

- Architecture 4
- Access 5
- Performance 6
- Deployment 6
- Integrations 7
- The VAST Catalog 8
- Transactional Support 8
- Filtering 8
- Inserts 9
- Updates / Deletes / Schema Changes 9

Analysis **9**

Introduction

The AI flood has begun, ushering in a fresh imperative for underlying data storage and compute foundations to evolve in tandem. To date, enterprises focused on harnessing the power of AI to stay competitive have patchworked storage and processing technologies developed in the pre-AI era. To prepare organizations for the coming AI renaissance, VAST Data has introduced the VAST Data Platform, a unique offering that stores and manages both structured (tables) and unstructured (files and objects) data in a single all-flash system based on VAST's DASE scale-out architecture. The system is built for large-scale data processing and low latency data access across diverse workloads.

This report explores the VAST DataBase, a next-generation database within the VAST Data Platform optimized for transactional, analytical, and real-time data workloads in the deep learning era. Having spearheaded innovation and revolutionized datacenter storage over the past three-plus years with its cost-effective, scalable, and flash-driven unstructured data platform, VAST now aims to empower enterprises and data teams to pioneer advancements with the VAST DataBase, while capitalizing on other novel features within its next-generation data infrastructure solution, the VAST Data Platform, all of which this report will comprehensively explore.

Why Should Organizations Look Beyond Traditional Data Systems in the AI Era?

The current generation of analytics platforms can be described as inflexible and unable to meet the demands of the AI era. Modern data-intensive applications are forcing organizations to re-evaluate their data management and compute engines. Heterogeneous data is continuously generated from edge to cloud, stored and used across data center boundaries. Maintaining this data across distributed storage and security boundaries is challenging, complex, convoluted, and cost-prohibitive. Moreover, organizations are losing innovative and competitive edges in this process.

Traditional data warehouses and MPP databases hit networking and compute bottlenecks as they scale and cannot operate at modern scale factors. Data lakes scale well but are inflexible and can't be transacted without expensive workarounds and rewrites. Data lakehouses bring flexibility to data lakes with transactions and time-travel but incur write amplification, adding fragility. The more the data lake concept is extended, the more it appears a complex patchwork attempting to conceal fundamental flaws.

VAST Data solves these AI-driven challenges with a globally distributed consistent system with high-performance and low-cost access to files, objects, and tables supported with a data catalog and a semantic layer. It is a system that seamlessly helps build data workflows across data center boundaries and heterogeneous workloads with low latency and pluggable compute architectures. VAST's goal is to manage data and computational gravity efficiently with a system that intelligently manages the oblique requirements of distributed systems across storage, compute, and I/O.

The VAST DataBase

The VAST DataBase is the first all-flash transactional and analytical system for AI. The goal of the VAST DataBase is to simplify data engineering by eliminating tiers of data management and combining transactional properties of a relational database with the schema of a data warehouse.

The DataBase is a database management service that writes tables into the system and enables real-time, fine-grained queries over tabular data that is cataloged. It breaks tradeoffs between transactional row-based OLTP databases with support for columnar-based analytical queries - akin to a flash-based data warehouse - with the scale and affordability of a data lake and the capabilities of a data lakehouse.

The VAST DataBase is built with open standards for seamless migration of enterprise workloads so that legacy systems can continue using standard file and object protocols while at the same time having access to AI-era data infrastructure and tools. It is built with core enterprise security and data management features that allows it to be the trusted system of records.

The VAST DataBase is one component of the VAST Data Platform. Customers who buy the platform get access not only to the DataBase but to the other components within:

- the VAST DataStore, a scalable and performant enterprise NAS platform with a distributed data processing engine and a novel disaggregated and shared everything (DASE) architecture over NVMEs
- the VAST DataSpace, a geo-distributed interface that enables a unified computing fabric and storage namespace to data stored across data centers spanning on-premises, cloud and edge.
- the VAST DataEngine, slated to be released in 2024, is a declarative compute execution module that has a topographically-aware global data positioning map of data locality that intelligently combines processor locality and availability to route and allocate functions and jobs across the available resources.

All of the platform components work together. The VAST DataBase, for example, leverages the same low-level storage elements of the VAST DataStore across a variety of storage protocols to benefit from functionalities like snapshots and global data compression.

Architecture

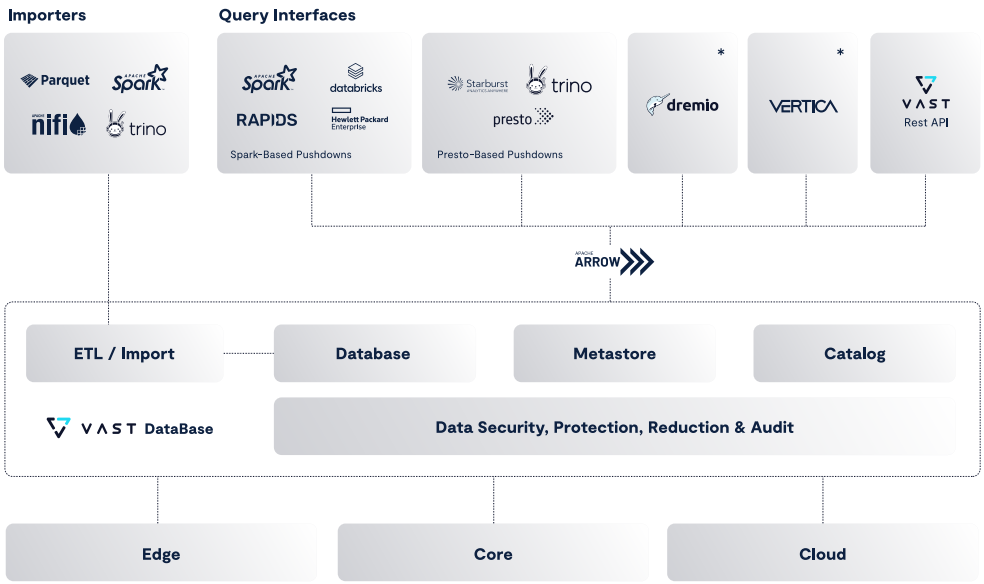
VAST's DASE architecture enables every core in the engine to have a global view to both short-term and long-term memory spaces. This concept avoids east-west cross traffic and provides scale with transactional consistency. DASE clusters are embarrassingly parallel and can scale linearly as the number of CPUs grows, eliminating the intercommunication between processors and state management that often slows down distributed systems at scale.

VAST clusters are built by decoupling system processors from the system's state and connecting these computational elements with a shared, low-latency data fabric. This allows the compute and storage layers to scale independently, eliminating the need to maintain a fixed capacity vs performance ratio.

Because the platform is architected to be stateless, customers can elastically vary the number of compute nodes in the pool based on their workload requirements. This is often a simple UI operation or an API call to automate the process. This happens at the stateless server layer, which necessitates no data movement.

The core data block of size 32KB in VAST is called an Element, which is an abstraction of a file, object, or table. The VAST Element Store is a high-performance key-value store with distributed sorting and projection filtering support. With this abstraction, the system can serve files as objects over the network and query against unstructured data like text files, audio, video, etc. This structure provides computational scale by allowing parallel execution of system functions capable of providing linear speedup with additional processing units or SSDs.

The VAST DataStore uses a persistent, low-latency write buffer for data to be streamed, captured, processed, and manipulated before it's stored in long-term low-cost flash storage. This gives an opportunity to order writes before they are persisted in long-term storage to optimize for low-latency retrieval. Data is reduced using VAST's unique similarity-based algorithms, helping users quickly discover correlations across large datasets.



Access

The VAST DataBase provides the engine for data analytics and AI ecosystems. It supports REST-based API for CRUD and querying with ACID semantics. It has a direct, ultra-fast interface into the VAST DataStore and integrates with industry-established analytic tools that leverage its scalability, performance, data reduction, and snapshot capabilities. It also incorporates a decentralized lock manager and a global read/write buffer. This facilitates VAST clusters to read,

process, and write in the global namespace by breaking tradeoffs between local performance and global consistency.

Performance

The VAST DataBase is targeted at analytical or high-volume workloads that benefit from high-performance and scalability. The best fits are queries on structured PB-sized datasets that need to avoid expensive re-sharding as volumes grow. Another good use case is any large analytical workload served by data on HDFS or S3 with containers managing different data and formatted files that traditionally introduce inefficiencies and slow query performance.

The VAST DataBase leverages Apache Arrow as the data-interchange format for in-memory representation and vector processing. This opens integration with modern analytics tools, minimizing the serialization/deserialization process during query executions across different query engines. With the combination of random-access columnar data structure with an embarrassingly parallel system, the VAST DataBase delivers immense query speedup compared to established players in the market with a much lower TCO.

The VAST DataBase writes tables transactionally and organizes table data into columns as they move from the buffer into low-cost long-term flash. VAST has pioneered a new columnar data structure for flash with 10^3 times smaller footprint than the standard Parquet row group. By avoiding disk-centric storage formats, VAST can achieve very high speedups, with two orders of magnitude acceleration over conventional data lake technologies such as Delta Lake and Apache Iceberg.

The VAST DataBase compute nodes perform efficient low-level operations for query execution through push-down predicates. It does not have a query engine or a query optimizer or a SQL parser, query planner. It rests on the shoulder of giants and relies on the compute nodes for fast, efficient low-level compute. Newer versions of the VAST DataBase will leverage distributed sorting and semi-sorted projections to enable even better performance.

Once the VAST DataEngine is available, the VAST DataBase will allow compute functions to freely flow from one data center to another essentially mapping the data topology and the processor topology to move compute to the data or vice-versa, whichever is most optimal. This allows customers to store and process data on-prem, at edge and in cloud.

Deployment

Data-driven applications deployed with Kubernetes can leverage VAST to be the persistent store, either as an NFS file server or with the VAST CSI plugin. CSI containers request dynamically provisioned storage from VAST with a VAST CSI driver that supports different Kubernetes storage classes. This can support multiple storage paths with access policies, path protection, and replication policies.

VAST DataBase functions are encapsulated in containers as stateless services for scale and resiliency and are deployed across edge, core, or cloud. VAST's infrastructure and resource management monitors deployed services across different cloud resources, optimizing for availability, data sovereignty, and cost. It allows tenants to set resource quotas and the system automatically adjusts workloads to match resource demands of data pipelines.

The containers persist state while the functions are stateless and instantiated only when an event

is triggered in the system. These functions can be developed as Python, SQL and Java code or with RESTful event API.

Integrations

The VAST DataBase has connectors for popular SQL query engines like Trino, Apache Spark and Lakehouses like Dremio that use push-down or offload work to VAST, where database operations can be executed faster with the most optimal resources and reduced data transfer between the query engine and the storage. This provides higher performance with less hardware compared to the Lakehouse approach.

VAST implements pushdown plugins to work with leading data science tools such as Apache Spark, the RAPIDS software libraries for data science on Nvidia GPUs, and the open-source Trino query engine in partnership with Starburst. VAST is open-sourcing pushdown plugins to provide a path for developers to integrate query engines of their choice.

The VAST DataBase supports four types of push-down operations:

- Row-pruning predicate pushdowns, where the query expression evaluates to TRUE, FALSE or UNKNOWN.
- Column-pruning projection pushdowns, where only the requested columns in the query are returned to the execution engine - as opposed to the entire row.
- Column expression projection where a computation of a column can be returned as TRUE/FALSE/UNKNOWN
- Count aggregations like “SELECT COUNT (*) can be pushed-down from the query engine where the DataBase will return only the result.

VAST’s S3 compatible object interface allows integration with analytics toolsets such as MapReduce, HIVE and Spark. It supports integration with distributed query engines (i.e., Dremio, Trino, Presto) and GPU-accelerated data warehouses like SQream.

Because VAST’s namespace is NFSv3 and NFSv4 compatible, the system also supports traditional analytics applications and databases such as SAS Grid, Postgres and KDB. Teradata, Splunk, ElasticSearch, and Vertica can integrate using either S3 or NFS. Support for the Windows SMB file sharing protocol allows easy access to reports and outputs from local file systems. This enables the VAST DataBase to integrate with:

- Legacy data warehouse-style analytics applications
- The intermediate layer of data processing within a data lake
- ETL & ELT
- Ad hoc analytics using tools
- Emerging AI/ML tools

The VAST Catalog

Scaling storage and data beyond the petabyte zone introduces extreme challenges in managing metadata. VAST addresses these challenges with a built-in metadata index called the VAST Catalog, the foundation for a semantic layer that adds structure to unstructured data.

The VAST Catalog indexes audit logs, access policies, replication, encryption, multi-tenancy, QOS metrics, and snapshots for time travel. Every file and object is cataloged. The Catalog quickens findability and searchability, major painpoints for most organizations with data scattered across clouds and on-premises.

The VAST Catalog treats the filesystem as a database that can be queried with SQL. The metadata data structure on which the Catalog is based is extensible and can be enriched with functions executed by the VAST DataEngine, keeping context and content synchronized. The VAST Catalog allows atomic operations with the file/object namespace and provides access through SQL, CLI, and a GUI.

Transactional Support

The VAST DataBase is ACID compliant. All operations are transactional and the boundaries of these transactions are adaptable to cover complex use cases. The nature of how data is managed in the metastore goes a long way to provide low latency for transactional and rollback semantics.

VAST has innovated with a new transactional data structure that exploits the NVMe and low-latency commodity fabrics to ensure data consistency and integrity with concurrent access. All basic CRUD-based database operations are supported at a granular level. Updates across rows or columns are simple metadata operations that make deletions and updates very fast.

Managing updates and rewrites in traditional data lakes is convoluted and can result in rewriting entire Parquet files. With the VAST DataBase this is a simple atomic object that can be easily deleted and updated without the need for any database cleanup or vacuuming.

The VAST DataBase supports transactions across multiple statements and tables, unlike data lakehouses.

Filtering

The VAST DataBase supports filtering with predicates that can be used with APIs or used by analytics engines to facilitate “push down” operations. For filtering and search, the VAST DataBase performs a distributed tree scan over blocks in the Element Store. It compares metadata in bitmap blocks, eliminating sections/blocks to search based on min/max value and optimizing data scan.

Inserts

Row inserts into the VAST DataBase occur in storage class memory (SCM) while bulk inserts result in a data chunk written to SCM. Insertion obeys isolation guardrails; inserted data is exposed to clients only after transaction completion. Transactions across different clients are grouped asynchronously and discrete objects in SCM are written to persistent read-intensive

NVMe along with updates to the associated metadata structures. Incoming data is hashed to cluster by similarity and compressed together and stored to flash memory (QLC).

As data is written into SCM buffers, an automatic migration process lays efficient stripes in lower cost, scalable flash. The layout chosen optimizes data querying with compiled statistics and metadata collocated with the data and results in significant reduction in I/O and processing to retrieve result sets. This process of handling ingestion into SCM has been successfully used for file and object applications on VAST for years, which facilitates millions of operations per second for both random and sequential workload profiles.

Updates / Deletes / Schema Changes

Row-level appends or updates write data to SCM, where table related metadata objects point to newly written data. Row-level deletes work like row updates or upserts. Table metadata is updated, logically removing the row by tombstoning it.

DDL operations, such as adding a column, are quick transactional metadata operations that don't result in data updates. As a columnar store, the VAST DataBase ensures no impact on subsequent inserts or updates. Column removals are transactional, with tombstones and immediate inaccessibility.

Analysis

Today the entire data ecosystem is crowded with shiny objects and dazzling buzzwords all fighting for investment. Data ecosystems are becoming data jungles, with data teams struggling to control entropy of complex modern data stacks.

How does one create an end-to-end data-driven platform in a heterogeneous ecosystem with modern tools and frameworks? The disaggregation and unbundling of the data stack and the proliferation of tools and vendors across the spectrum are causing organizations to "duct tape" products and frameworks to build their end-to-end solutions, products, and delivery processes. On the other end of the spectrum are all-in-one systems that are hard to change and adapt and are cost prohibitive.

VAST Data gives a peek into the future data platform and paves the way for an era where data becomes the API to the real world with a unified system for managing, processing, and accessing structured and unstructured data. Today VAST is flourishing in markets where performance is key, spanning several vertical and customer types with divergent deployments.

Most importantly, organizations looking to build a modern AI-capable data platform don't need a parallel file system. The VAST dream team has built a software integration layer to their DataBase that can leverage CPUs and GPUs with vertically-integrated storage to be the enterprise generative AI workhorse.

Contrary to what is happening in the data ecosystem with the disaggregation of the data stack, VAST Data, with its all-in-one approach, reduces and eliminates the need for expensive data integration tools. Integration and migration processes do not involve actual data movement but rather pointer maneuvering within the VAST DataStore and a corresponding update to the VAST Catalog. The VAST DataBase supports a Parquet file importer, which is essentially a pointer to

the location on the VAST DataStore cluster. The well-integrated VAST Catalog also eliminates the need for having an external or third-party data catalog and metadata services.

While Databricks and Snowflake are morphing from data lakes and cloud data warehouses to data lakehouses and data platforms, VAST goes a few steps further with its value proposition, vision, and goals. VAST can effortlessly work with extreme-scale and complex datasets like genomics data and unstructured data. The VAST DataBase can handle transactional queries across multiple tables and does not need tuning of its I/O subsystems, which already relies on the VAST DataStore's unmatched storage capabilities. VAST's platform and architecture can easily integrate and leverage different types of compute units (CPU, GPU), based on workload profile with its containerized services.

VAST's innovative approach to solving data problems for enterprises is exciting, forward thinking, and valuable. However, data systems don't exist in isolation, and VAST's major challenge will be to strategically integrate with other tools across the ecosystem. The data industry has historically seen multiple all-in-one systems and such systems have been successful in certain areas and not-so successful across others. With such systems, customers always get wary of vendor lock-in to proprietary systems.

VAST is countering the above challenges by entering into alliances to drive its total addressable market (TAM) and increase its footprint. Two of VAST's key technology partnerships are with HPE, where VAST will be the brains for HPE's GreenLake-based file storage solution, and with NVIDIA, which has certified VAST as the first NAS solution for its DGX SuperPOD.

VAST Data serves as the data foundation for organizations like Pixar to rapidly develop films, Lawrence Livermore National Labs to address energy research challenges, Agoda to optimally drive user booking experiences, to name a few customers. 100% of customers surveyed from Gartner Peer Insights highly recommend VAST to their peers.

In a rapidly changing technology world with increasingly complex problems to address, VAST Data is continually innovating, with disruptive changes for a better future. The VAST DataBase is a refreshing change, building atop VAST's powerful, successful flash-based storage layer to deliver next-generation DBMS services and integrations with established vendors and frameworks for data processing and data integration.