

Actian Vector for Hadoop Analytic Database

The industry's fastest SQL database for Hadoop



Key Benefits

- Run complex, ad hoc queries against billions of records in seconds
- Process hundreds of records in a single CPU instruction cycle with vector processing
- Execute updates without any performance penalty
- Get consistent query results even if the data changes
- Utilize dedicated CPU core and caches running 100x faster than RAM
- Scan data faster using self-indexed blocks
- Leverage your existing data lake for new analytic workloads
- 100x SQL Performance Increase over traditional tools
- Comprehensive Hadoop Security Integration
- Native Spark powered direct query access across Hadoop data formats
- Combine SQL and NoSQL in a single database table

Actian Vector delivers on the promise of in-the-moment analytics with the industry's fastest analytics database. Actian Vector makes analytics more accessible to business users, freeing them from the common limitations of traditional data warehouses. Achieve extreme performance on commodity hardware or cloud-based platforms, with little or no database tuning.

Organizations have made significant investments in their Hadoop data lakes and continue to look for ways to leverage that expanse of data to gain more value and faster insights. Hadoop was designed for scale, and Vector for Hadoop was designed for speed. Together they are enabling a new class of real-time and performance-oriented analytics with existing data lake investments.

Financial services, retail, telecommunications, media and many other enterprise users need to make decisions based on fresh data and the flexibility to explore their data beyond the beaten path.

Deliver fast analytics for your Hadoop Data Lake

Execute queries in seconds not hours.

Analyze larger datasets faster whether operational or streamed

Apply updates from operational systems with no impact to query performance

Support more concurrent users to increase the return from your data investment

Iterate more quickly – more responsive ad hoc queries without tuning

Record-breaking analytic performance

Actian Vector leverages vector processing technology to unlock hidden performance features in your existing hardware. The results below compare the sums of query execution times for all 22 TPC Benchmark^{TMH} (TPC-H) queries using popular SQL on Hadoop solutions. This workload benchmarks decision support systems that examine large volumes of data, execute queries with a high degree of complexity, and give answers to critical business questions.



Key features

Vectorized query execution: that exploits Single Instruction, Multiple Data (SIMD) capabilities in commodity Intel x86 architecture CPUs, enabling processing of hundreds or thousands of data values using a single instruction.

MPP architecture provides exceptional scalability on Hadoop clusters which scale-out to thousands of users, hundreds of nodes, and petabytes of data, with built-in data redundancy and system-wide data protection.

Full ACID compliance: performs data updates with multi-version read consistency, maintaining transaction integrity.

Zero-penalty real-time data updates: enable in the moment computing using patented Positional Delta Trees (PDTs) for incremental small inserts, updates and deletes without impacting query performance.

CPU cache optimization: using dedicated CPU cores and caches as execution memory to run queries 100x faster than from RAM, delivering significantly greater throughput than conventional in-memory approaches.

CPU optimizations: include hardware-accelerated string-based operations, for accelerating selections on strings using wildcard matching, aggregations on string-based values, and joins or sorts using string keys.

Column-based storage: reduce I/O to relevant columns and provide the opportunity for greater data compression and enable storage indexes to maximize efficiency.

Data compression: provides multiple options to maximize compression, from 4-10x for Hadoop storage.

Storage indexes: provide automatic min-max indices to enable fast block skipping on reads and eliminate the need for an explicit data partitioning strategy.

Parallel execution: use adaptive algorithms to maximize concurrency while enabling load prioritization.

Spark powered direct query access: that provides direct access to Hadoop data files stored in Parquet, ORC, or other standard formats allowing users to realize significant performance benefits without converting to the Vector file format first.

User Defined Function Support: User-defined functions (UDFs) let the user extend the database to perform operations that are not available through the built-in, system-defined functions provided by Vector. Vector for Hadoop 6 provides the capability to create Scalar UDFs.

Faster Machine Learning Execution Capability: With Vector for Hadoop it is now possible for the deployment of machine learning (ML) models that run alongside the database leveraging the new extended UDF capabilities. By deploying machine learning models alongside the Vector database, data movement is reduced, thus allowing for faster scoring of data.

Combine SQL and NoSQL in a Single Database: JSON functions in Vector enable you to combine NoSQL and relational concepts in the same database. Now you can combine classic relational columns with columns that contain documents formatted as JSON text in the same table and parse and import JSON documents in relational structures. Bridging semi-structured data with relational databases creates a solution that is more flexible and can handle additional use cases where the underlying data structures change more rapidly.

Extensive SQL support with standard ANSI SQL and advanced analytics, including cubing, grouping, and window functions.

Deploy on-premises and in the cloud: on single servers or in Hadoop clusters scaling to hundreds of nodes. Cloud deployment supports perpetual and subscription licensing, pay as you go (hourly and annual plans) or as a service, managed by Actian, on your choice of cloud service provider.

