

# Actian Vector Hybrid Data Warehouse

## High performance data warehouse for operational workloads

### 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

100x SQL Performance Increase over traditional tools

Execute updates without any performance penalty

Get consistent query results even if the data changes

Exploit dedicated CPU core and caches running 100x faster than RAM

Scan data faster using self-indexed blocks

Native Spark powered direct queries

### Flexible Deployment

Linux 

Windows 

Hadoop 

Google 

AWS 

Azure 

Data sheet

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

With Actian Vector, organizations can make decisions based on fresh data and the flexibility to explore their data beyond the beaten path.

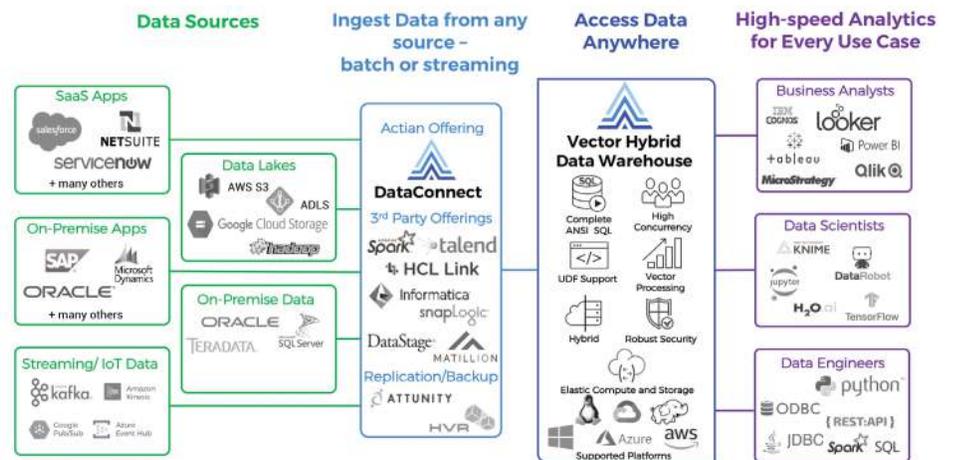


Figure 1. Actian Vector works in any enterprise, on any data

### Deliver fast analytics for your Operational Data Warehouse

- **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 delivers blazing fast response on ad-hoc BI queries using the freshest data. Unlike other data warehouses, Vector enables real-time updates to data in the data warehouse without adding latency. You can be assured of accessing the latest data to power your analytics without paying a performance penalty. Vector's advanced columnar implementation enables the least I/O performed while retrieving data from disk. It's vectorized compute leverages CPU Single Instruction, Multiple Data (SIMD) and processes data in the L1/L2 CPU cache instead of RAM, leading to significantly faster performance.

## Features

- Vectorized query execution
- Highly scalable MPP architecture
- Full ACID Compliance
- Zero-Penalty Real-Time Updates
- CPU Cache Optimization
- Column-based Storage
- Fully SQL-2016 compliant
- User Defined Function support for Python, SQL and JavaScript
- Support for diverse data types, including JSON
- Open, industry standard data access, including Spark, ODBC, JDBC, .NET
- Accelerated ML Execution Support
- Deploy on-premises and in the cloud

## Hybrid, On-Premise, Private and Managed Cloud for Maximum Flexibility

Action Vector can be delivered as an on-premise solution on Windows and Linux or as a private or managed cloud offering on Google Cloud, AWS, or Azure. This enables organizations to realize the true potential of hybrid by bringing the compute power of Vector to the place where their data resides. With Vector, you can leverage the same database engine, the same physical data model, the same ETL/ELT tools, and the same BI tools of your choosing.

Vector takes advantage of patented features and industry best practices such as the following:

**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. Additionally, Vector running on Hadoop enables Scalar UDF creation.

**Faster Machine Learning Execution Capability:** With Vector, 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.

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

**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.

#### What this means for you:

- **Lower cost and improved performance:** Processing and querying data where applications are producing it, whether on-premises or the cloud, often produces the best results in terms of cost and performance since data movement from the cloud is expensive and slow
- **Simplified, futureproof architecture:** Since Vector relies on the same patented vectorized database engine both in the cloud and on-premises, you will work with a single data model, consistent ETL integration, and one technology to learn
- **Stronger compliance and security:** You have the option to retain complete control over sensitive datasets with encryption at rest and in transit and through columnar-level masking. Highly secure workloads can remain in the data center if the CISO demands it, while the cloud is used for other workloads.
- **Amortized on-premises investments:** Workloads that are optimal for the cloud can move immediately to the cloud while those that can be handled on-premises can run on infrastructure that has already been paid for and may be cheaper
- **Flexible CapEx/OpEx pricing:** Share your expenses between CapEx and OpEx as your needs dictate
- **Phased, non-disruptive migrations:** Workloads can stay on-premises until it is ready to move to the cloud

#### Built for Real-Time Performance

With Vector, you can achieve blazing fast response on ad-hoc BI queries using the freshest data. Unlike other cloud data warehouses, Vector enables real-time updates to data in the data warehouse without adding latency. You can be assured of accessing the latest data to power your analytics without paying a performance penalty.

Its advanced columnar implementation enables the least I/O performed while retrieving data from disk. Avalanche's vectorized compute leverages CPU Single Instruction, Multiple Data (SIMD) and processes data in the L1/L2 CPU cache instead of RAM, leading to significantly faster performance.

#### Rapidly Deploy with Native Integration for Enterprise BI & Analytics

Vector has out-of-the-box integration with multiple Data Integration tools, including Actian DataConnect which includes over 200 connectors and templates for easily sourcing and moving data from SaaS applications to Vector at scale—no special ETL is required. You can load structured and semi-structured JSON data, including event-based and streaming, without coding.

Through ODBC, JDBC, .Net and Python, Vector enables business analysts, data scientists and data engineers to use their favorite tools for visualization, reporting, and advanced analytics. Examples include Looker, Power BI, Tableau, and MicroStrategy.



www.actian.com | Toll Free +1 888 446 4737 | Tel +1 650 587 5500  
2300 Geng Rd., Suite 150, Palo Alto, CA 94303



© 2021 Actian Corporation. Actian is a trademark of Actian Corporation and its subsidiaries. All other trademarks, trade names, service marks, and logos referenced herein belong to their respective companies. (DS03-0321)