MOLECULA

DATASHEET—

FeatureBase

A feature-oriented database platform that powers real-time analytics and machine learning applications by simultaneously executing low-latency, high-throughput, and highly concurrent workloads.

FeatureBase: The Real-Time Database for Any Scale

Our Novel Data Format

Traditional OLAP systems (column-oriented) improved processing of analytical workloads, but they were not built to handle real-time, streaming data flexibly and affordably. At Molecula, we've created a feature-oriented data format that minimizes I/O on queries by allowing the database engine to read and write exactly the data it needs and intelligently compress that data in memory.

This approach makes it possible for engineers and developers to meet real-time analytics requirements at a fraction of the cost and complexity.

Ideal Environment

FeatureBase is beneficial for organizations that need to access large quantities of real-time events each day, joined with many fragmented, terabyte-scale data sources. FeatureBase excels at complex analytical workloads where source data is fragmented across silos and where a user or machine wants to apply a number of filters or criteria to a query.

Implementation

FeatureBase ingests from various streaming data sources, data warehouses, and change data capture (CDC) plugins. Our experienced customer engineering managers are trained to bring these integrations, customized for your complex data environment, to production during the implementation period.

Product Capabilities

- Handle massive scale datasets
- Make instant decisions
- Make decisions w/the freshest data
- Stop throttling your resources
- Include multiple values for a single field
- Do more with fewer resources

How FeatureBase Works

High-Throughput Ingest: FeatureBase has the unique ability to scale out elastic ingest servers independently from your actual database servers. This means you can ingest as much streaming and historical data as you need when you need it without impacting your database hardware. When data is ingested into FeatureBase, it is automatically transformed into our highly performant feature-oriented format.

Fresh Data: FeatureBase allows you to eliminate costly and time-consuming preaggregation steps meaning you are continuously operating on your freshest data. Avoid being dependent on rigid scripts tied to rollup tables, OLAP cubes, data marts, or summarized views, and instead view and compute data in its most granular, up-to-date form.

Low-Latency Queries: Run even your most complex queries and workloads and expect millisecond results, regardless of data cardinality, size, or scale. Our novel data format structures data to allow for minimum I/O when processing analytical queries, meaning users avoid compute-intensive scans and reduce their cost per query.

High-Concurrency Querying: FeatureBase employs a technique called Multiversion Concurrency Control (MVCC) meaning that any time it writes, whatever is being written gets written entirely new and any ongoing reads or reads that start while that write is happening will not be impacted by the write. This ensures that reads are never blocked by other reads or writes.

FeatureBase in Action:

A company that provides video advertising solutions for brand advertisers to engage consumers across multiple internet-connected devices struggled to achieve the throughput and latency they needed. Previous solutions they considered were offline tools with thousands of boxes (e.g., Hadoop) and some more advanced servers (e.g., Vertica). Still, these solutions forced anywhere between 24 to 72 hours before they could target an audience.

With FeatureBase and the feature-oriented format:

- Able to capture more market share by creating something that is not yet seen in the market today: a real-time processing engine.
- Workloads continuously ingest anywhere from 500K-1M records per second while simultaneously running thousands of concurrent queries.
- Maintain sub-second query latency regardless of ingest volume

FeatureBase

FeatureBase is priced based on reserved memory instances, which are sized in alignment with your data footprint, and immediately available CPU time. Contact us for more information.