



**Vertica for Structured Finance**  
Industry Solution White Paper

## Executive Summary

Organizations that deal with structured products have specialized needs. The requirements to more quickly and accurately model the performance and risk of structured products has risen dramatically over the past few years. This creates new demands for more complete and accurate loan level data, and for faster, more accurate and more advanced analytics to drive investment decisions.

This paper examines the increased business challenges in structured finance, the impact on the underlying enabling technologies, and describes how the use of Vertica's column store, massively parallel processing (MPP) approach to database management and analytic processing can address some of these challenges. This paper also examines how two organizations, Tilden Park Capital Management and Intex Solutions, are using Vertica to successfully address these challenges.

## Industry Challenges

Whether you are an insurance company looking to hedge your existing liabilities, or an investment bank or hedge fund seeking a competitive edge in security selection, the margin for error in structured finance is finer than it has ever been. The industry is more volatile, and the risk of default is higher than ever before. The ability to deeply analyze and understand the underlying loans in structured products, and quickly correlate the various attributes of the individual loan data with the risk and performance of the security under a wide range of market scenarios is critical for success.

These business challenges, and others, are creating new kinds of demands on the technologies that support the business processes and decisions in structured finance.

For example, organizations must now fully analyze many secondary and tertiary fields in the data, such as FICO scores and prepayments. This means that the completeness of the data is critical; requiring that successful organizations integrate data from various sources, and perform advanced, time-series based interpolation techniques to fill in gaps in the data. The accuracy of these secondary and tertiary fields has also become more critical than ever before, necessitating advanced analytics capabilities to identify anomalies and assign confidence levels to the various data values.

The need to manage these large data sets, quickly update the data with new information or with information from new sources, and run sophisticated analytics on time-series data is best achieved by utilizing massively parallel processing (MPP), column store database technologies. The speed, simplicity of the analytics processing, and price point of a traditional row store database, such as MySQL or Oracle, is no match for an MPP, column-store approach.

Organizations that are involved with creating or analyzing structured products are able to achieve unprecedented levels of functionality, performance, and value by using the Vertica Analytic Database to model and store more complete and accurate data on the loans underlying the securities, and perform new, advanced types of analytics on the data to identify information, patterns, and trends to drive their investment decisions.

## Tilden Park Capital Management

Tilden Park Capital Management is a New York-based asset management company that invests in residential and commercial mortgage securities. In order to better service its traders, who utilize large data sets and complex analytic queries to make investment decisions, the company implemented a column store analytic database to support their sophisticated front end applications.

Tilden Park also uses a pre-loaded, third party hosted solution from which they can access and analyze securities data. However, difficulties integrating additional data sources with the hosted solution, and difficulties integrating their front end applications with the hosted solution contributed to the company's decision to evaluate additional approaches.

Tilden Park considered a number of alternatives, and after careful evaluation they implemented the Vertica Analytic Database. The company's implementation of Vertica is providing the following key benefits:

- Superior Price : Performance
- Advanced Analytics for Structured Finance
- Standards Compliance

**Superior Price : Performance.** The Vertica Analytic Database is the only database built from scratch to handle today's heavy business intelligence workloads. In customer benchmarks, Vertica has been shown to manage terabytes of data running on low cost hardware, and answers queries 50 to 200 times faster than competing row oriented databases and specialized analytic hardware.

Vertica is a true column store database. Unlike row store databases where all columns for each row are stored together, the Vertica database only needs to retrieve those columns required for a specific query rather than all columns in the selected rows. This true columnar approach to data management produces dramatic I/O savings for the large majority of decision support queries that only retrieve a subset of columns, making it well suited for delivering the fast, ad hoc, advanced queries required for structured finance applications.

For example, Tilden Park is running Vertica on a cluster of 3 inexpensive Linux servers, and according to the company, the performance for their analytic queries equals or exceeds what was available using million dollar hardware resources in the past.

**Advanced Analytics for Structured Finance.** Vertica's built-in analytics capabilities make it easy to extract information from time-series data associated with structured products. For example, Vertica provides deep analytics capabilities for performing time-series analysis, sliding window analysis, etc. both within and across securities, to identify patterns in the data without the need for specialized products or programming. Using Vertica, Tilden Park is able to easily calculate loan delinquency history and cumulative default behavior using a single query, and compare the results against various time periods for other loans in the security, and for loans in different securities. They are also able to identify patterns associated with a particular underwriter and vintage by analyzing the loans underlying a single security or underlying multiple securities. Previously, it was not possible for the company to perform these types of sliding time-series and cross-security analytics in-house without having to pre-build summary tables. As a result, they can now easily and rapidly model the performance of a security based on a wide variety of different market scenarios. "We're now able to run very involved queries against our in-house data live, rather than having to pre aggregate the data," according to the company. "Vertica eliminates a lot of the work we had to do previously."

In addition, the performance and simplicity of Vertica has made it possible for IT to configure the applications so that their traders can quickly and easily perform many of their own queries. As a result, the company is more flexible and can scale more efficiently. According to Tilden Park, "With Vertica, our platform enables us to dig deep into the data, and get the information in front of our users better than any other system out there."

**Standards Compliance.** Vertica provides a standard SQL interface to users, as well as providing compatibility with existing ETL, reporting, and business intelligence (BI) tools. This makes it easy to migrate existing databases to Vertica and to use other BI technologies with Vertica databases. Vertica is designed to run on inexpensive clusters or “grids” of off-the-shelf Linux servers that use local disk for storage. No expensive SANs or high-end servers are required to run large data warehouses on Vertica (although Vertica performs well using shared SAN storage if that’s a preferred deployment route). Vertica reduces hardware costs (often by up to 90% relative to other data warehouse databases) while improving the ability to answer more queries for more people against more data.

As a result, Tilden Park was able to leverage their existing SQL knowledge with no additional training, and they simply point their existing front end applications at the Vertica database without any additional customization, which was not possible with their hosted solution. They also appreciate the flexibility and options for future changes and growth, and the lack of “vendor lock-in.”

Using Vertica to integrate disparate data from multiple data sources enables Tilden Park to create an enriched, comprehensive set of historical data on the securities and the underlying loans. Although the various data providers all deliver their data in different formats, Vertica provides the platform to house a single, uniform data repository that is used for downstream analytics.

Vertica’s superior price : performance, built-in support for the advanced analytics that are required to deeply analyze structured products, and standards compliance are enabling Tilden Park to speed and simplify its various security selection processes, while dramatically reducing infrastructure costs.

## Intex Solutions

Intex provides deal cashflow models, analytics, and structuring software for RMBS, ABS, CMBS, and CDO securities worldwide. With over 20 years of experience and with more than 20,000 deals modeled, Intex is universally recognized for providing the most accurate, independent, and complete cashflow model library in the industry.

In addition to Intex’s cashflow models, the company offers tools that allow users to perform single security and portfolio analytics on any of the 20,000+ deals modeled by Intex. The tools allow users to control interest rates, prepayments, defaults, triggers, deal specific variables, etc. and solve for various outcomes.

Recent events in the market are driving Intex’s customers to demand even more precise and granular data to enable them to analyze the securities in new ways. Also, many of their customers are asking for new tools that make it easy to perform more powerful and flexible analytics against the data without requiring sophisticated internal IT capabilities. In response, Intex has implemented the Vertica Analytic Database to help deliver these new services.

Vertica is enabling Intex to deliver these new key capabilities to its customers:

- More Powerful and Flexible Research Tools
- More Complete and More Accurate Data
- Integration of Disparate Data Sources

**More Powerful and Flexible Research Tools.** Intex’s customers all want to perform new, sophisticated analytic queries on the data. For example, customers want to analyze loans across

various securities (not just within one security) and across various sliding time windows (6 months, 1 year, etc.) to look for late payments, missed payments, delinquency roll rates, and other risk indicators. Many of the organizations that use Intex's products would rather focus on analyzing cash flows, making successful trades and managing risk than on building and managing an IT infrastructure. Hence, Intex wanted to offer new research tools that would make it easy for customers to perform these types of analyses against their entire library of loans in a simple and cost-effective manner.

Using Vertica, Intex is building its next generation analytics product that provides all of these capabilities - and more - to its customers. Customers can now ask questions of the data that were not previously possible. For example, customers can now run queries to identify loans with similar performance characteristics, to determine how loans from the same underwriter and vintage perform, and to look for relationships between FICO scores and roll rates. Intex relies on Vertica's capabilities to make it possible to identify entirely new scenarios to stress the deals, offering its customers pre-built queries that were not previously available or even possible with its previous tools.

Through the use of Vertica to power its new offerings, Intex is able to offer the state of the art analytics coupled with the best data in the industry, with very high performance and at competitive price points. According to Intex, "Our MySQL database was already slow with just the data within the deal. We were looking for the fastest solution to power these new initiatives. Vertica outperformed all of the other column store databases we considered."

**More Complete and More Accurate Data.** Since many customers are now bucketing off many of the secondary and tertiary data fields in the cash flow models (such as all of the loan modification details, FICO scores, etc.), there is a greater emphasis on the completeness and accuracy of all of the fields in the underlying data.

Before implementing Vertica, Intex managed and delivered the loan data in a set of flat, historical files. For example, a single deal with a five year history might be represented by 60 historical (monthly) files. In total, the details on these 20,000+ deals account for more than 1 terabyte of information. Identifying discrepancies or anomalies in the data was a tedious and cumbersome process.

All of the historical data is now integrated, normalized, and stored in their Vertica database, making it practical for Intex to quickly identify potential inaccuracies in the source data using automated processes and analytics functions. Thousands of monthly details, including payments, triggers, defaults, etc. are now rapidly integrated into the master database and run through their automated quality assurance processes. According to Intex, "it's easy for us to identify potentially bad data, and go back to the servicers and trustees if something doesn't look right." In addition, Vertica's built-in sliding time-series analytics functions enable Intex to easily interpolate the data to fill in gaps, delivering more complete and more accurate data for its customers. Intex now delivers this accurate, integrated, and current data as a single file to its customers, and also uses this data (and the Vertica database) to power its new analytic research tools.

**Integration of Disparate Data Sources.** Using Vertica, Intex is able to easily create mappings to other data sources, with associated confidence levels, addressing customers' requirements. For example, Intex has created mappings with confidence levels to various loan performance and other data.

### Vertica for Structured Finance

This section summarizes key aspects of Vertica's technology that enable such dramatic performance benefits, describes aspects of Vertica that are especially well suited for structured finance, and compares the design of Vertica to other popular relational systems.

## Superior Performance, Scalability, Reliability, and Ease of Use

The key to Vertica's superior performance is three fold:

1. Vertica organizes data on disk as columns of values from the same attribute, as opposed to storing it as rows of tabular records. This organization means that when a query needs to access only a few columns of a particular table, only those columns need to be read from disk. Conversely, in a row-oriented database, all values in a table are typically read from disk, which creates excessive I/O processing.
2. Vertica employs aggressive compression of data on disk, as well as a query execution engine that is able to keep data compressed while it is operated on. Compression in Vertica is particularly effective, as values within a column tend to be quite similar to each other and compress well, often by up to 90%. In a traditional row-oriented database, values within a row of a table are not likely to be very similar, and hence are unlikely to compress well. Columnar compression and direct operation on compressed data shift the bottleneck in query processing from disks (I/O processing is expensive, and is not getting faster) to CPUs (which is less expensive, and is getting faster all the time).
3. Because data is compressed so aggressively, Vertica has sufficient space to store multiple copies of the data to ensure fault tolerance and to improve concurrent and ad hoc query performance. Logical tables are decomposed and physically stored as overlapping groups of columns, called "projections," and each projection is sorted on a different attribute or set of attributes, which optimizes them for answering queries with predicates on its sort attributes. A Vertica database is composed exclusively out of these query-optimized structures on disk, without the overhead of base tables. It's similar in concept to a database made entirely of materialized views with no base tables.

Vertica's compression, column-orientation, and table decomposition are all transparent to the end user.

In addition to delivering superior performance for a wide variety of database workloads, Vertica includes several other features designed to offer performance, scalability, reliability, and ease of use. These include:

1. A shared nothing, grid-based database architecture that allows Vertica to scale effectively on clusters of commodity CPUs.
2. A hybrid data store, where newly inserted records are added to a write optimized portion of the database to allow continuous, high-performance insert operations concurrently with querying to enable real-time analytics. In fact, in December 2008, Syncsort, HP and Vertica teamed up to break the world record for data warehouse data loading. It took 57 minutes 21 seconds to load 5.4TB of TPC-H data into Vertica with Syncsort's data integration software running on an HP BladeSystem C7000, which beat the previous record set by Microsoft (2.36TB in 1 hour).
3. Automated physical database design tools that recommend how data should be organized both locally on each node in a cluster, as well as horizontally partitioned across a cluster. In addition to choosing projections and sort orders, these tools ensure k-safety, meaning that all data is replicated on multiple nodes so that k node failures can be tolerated by the system without interrupting functionality. These tools reduce administrative costs by simplifying physical database design decisions. They also allow Vertica to automatically adapt to on-the-fly the addition or removal of database nodes.

4. High-performance, ACID-compliant database system with a light-weight transaction and concurrency control scheme that is optimized towards loading and querying data. Vertica's failure recovery model is based on replication (k-safety) rather than traditional log-based methods.
5. Flexible deployment options
  - a. Downloaded and installed on Linux servers of your choice
  - b. Pre-configured and shipped on HP BladeSystem hardware
  - c. Licensed and used on an on-demand basis, hosted in the Amazon Elastic Compute Cloud (EC2).
6. Monitoring and administration tools and APIs for controlling performance, backup and disaster recovery, etc.

### Sliding Window Analytics Functions

A key set of features in Vertica that are especially useful for structured finance are the built-in sliding window functions that work with time-series data. These features are native in Vertica, making them easy (no special programming or analytics tools are required) and fast (the computation occurs in the database, close to the data) to perform. These features can be used to:

- Interpolate, or accurately fill in the gaps in missing time-series (loan) data,
- Identify "outliers" in time-series data, which may indicate bad data values that should not be trusted, and
- Calculate delinquency roll rates and identify trends in prepayments, delinquencies, late payments, modifications, and so on.

Vertica's built-in support for time-series analytics capabilities include:

- Moving window aggregate functions
- Time-series and time slice computations
- Event based window definitions, including
  - o `CONDITIONAL_CHANGE_EVENT`, which assigns an event window number to each row when the result of evaluating the argument expression on the current row differs from that on the previous row
  - o `CONDITIONAL_TRUE_EVENT`, which assigns an event window number to each row when the result of the boolean argument expression evaluates true
- Delta functions
- Moving average and exponential moving average
- Gap filling and interpolation

### Conclusion

The Vertica Analytic Database delivers capabilities that enable companies that work with structured products to make the investments decisions they require, with new levels of speed, simplicity, flexibility, and price : performance.

## About Vertica Systems

Vertica Systems provides a true columnar analytic database management system that delivers the fastest, most scalable and easiest way to gain business intelligence from huge volumes of data in near-real-time. Vertica has developed a modern SQL-based analytic database with an MPP architecture that runs on low-cost standard hardware. The technology enables companies to gain a competitive edge by querying enormous databases 50x-1000x faster and for a lower Total Cost of Ownership than other solutions. The Vertica Analytic Database is available as software only, as a hardware-based appliance, as a virtual appliance on VMware or online as a cloud computing solution. The technology's unmatched speed, simplicity and scalability are used by more than 130 customers including Verizon, Guess Inc., Zynga, Capital IQ, Mozilla and Comcast. Vertica is headquartered in Billerica, MA.

For more information or to request a demonstration, contact Vertica at (978) 600-1000 or visit the company's website at <http://www.vertica.com>.