



## Tick and Time Series Storage and Analysis December 2008

### Next Generation Market Data Capture and Analysis

When most analysts, quants and traders think of data management, they immediately brace themselves for a time consuming and difficult task. With the majority of software solutions requiring a substantial amount of integration and development just to address common tick data business needs, most users know that managing and analyzing clean data from multiple sources is no easy feat.

That is why One Market Data (OMD) has developed OneTick, the premier enterprise-wide solution for tick data collection, management and research. OneTick differs from the competition by providing users with a platform designed specifically for market data – a distinct advantage over vendors who built generic database products that were retrofitted to support market data.

With extensive backgrounds on Wall Street's business side – using tick data for statistical arbitrage, algo trading, quantitative research, TCA, compliance and surveillance, portfolio trading, program trading and more – OMD's founders have taken special care to design a solution that emphasizes a low total cost of ownership.

With OneTick, users gain access to an array of out of the box adapters that connect to many sources of market data plus a rich API that allows for fast development of new adapters – making integration straightforward and rapid as existing adapters typically just require reconfiguration (not coding) to support new sources.

Through an easy to use GUI for extracting and analyzing data, business users can begin using the product almost immediately without programming support. Thus, OneTick is easy to implement, easy to learn and inexpensive to run.

### Built with Function in Mind

The founders of OneMarketData left Goldman Sachs to start OMD with the express purpose of building the next generation for tick and time series data management. We did not just seek to build another data storage mechanism but rather, a complete tick data and time series management platform that includes a rich business logic and analytical layer so that consumers of tick data and time series could enjoy a far more functional system than those which are available elsewhere in the market. Having built a system which supported the global tick data needs of the entire sell side of Gold Man Sachs, OneMarketData brings a level of domain expertise that is unmatched by our competitors. All key tick data and time series issue have already been encountered and resolved in the past and subsequently addressed in OneTick.

#### *Data Sources*

OneTick places no limitations on sources of data that can be efficiently loaded into it. Data can be loaded both in real-time and in batch modes. Batch mode is useful for loading data sets delivered daily, loading large historical data sets, loading proprietary data, while real-time loaders are used to load data from real-time feeds. We provide out of the box adapters to many providers of real time



# ONE MARKETDATA

market data (Reuters, Wombat, IDC) and providers of data history (TAQ, NYSE Open Book, ARCA, LAVA, OPRA, CME, LSE, Deutsche Bourse, Euronext, etc.). In addition, we provide a generic ASCII adapter that supports loading generic ASCII files by simply writing an appropriate configuration file.

## *Data Products*

OneTick places no limitations on the types of data that can be efficiently stored in it and is agnostic of instrument classes. We support all intraday data types (trades, quotes and order books for all instrument types listed above) as well as lower frequency data (such as closing prices, fundamental data, news, etc), derived and proprietary data (system parameters, calculated statistics), static or slowly changing data (reference data, product data, industry classifications, etc). OneTick is also being used to capture and store options data (OPRA feed), which constitutes the bulk of global tick data.

## *Data Uses*

OMD empowers users to leverage data for market analysis, transactional purposes as well as portfolio and risk analytics:

- **Market Analysis:**
  - OneTick fully supports high frequency asset price and volume data for deeper analyses of market functioning. It also supports data from electronic trading platforms and can provide detailed insight into market behavior
- **Transactional Purposes:**
  - OneTick enables users to manage data for back-testing of strategies, real-time signal generation, collateral pricing in support of repo and securities loan transactions, and decision support for trade execution
  - OneTick also provides built-in portfolio level analytics for a single country or multi-country portfolios and allows for off-the-shelf absolute or relative tracking analysis
- **Portfolio and Risk Analysis:**
  - OneTick fully supports financial market data for portfolio modeling to support asset allocation decisions and for risk analysis

## *Data Volume/Access/Retrieval*

OneTick is a central, single-point storage and analytics solution that supports the querying of real-time data combined with historical data. Providing this as standard functionality, this solution enables the exact same query to run on historical data, real-time data or a combination of both. Users need not specify whether the data is held, in archive or in-memory – OneTick knows where the data resides and will assemble the results across multiple databases if necessary to return a single results set to the client.



# ONEMARKETDATA

## Easy to Deploy, Easy to Integrate

### *Application & Database Software*

A complete out-of-the-box solution, OneTick outpaces the competition by offering an easy-to-use platform that eliminates the need for programming, provides robust data management and analytics capabilities, and connects easily to all leading market data vendors.

In addition, OneTick offers:

- **Deep Granularity of Time Series Information:**
  - OneTick captures and archives every tick. Currently, timestamps are in milliseconds, with microsecond granularity envisioned in a future release if such a need arises
- **Ability to Cross Reference Times Series Data with Other Data Types:**
  - OneTick queries/analytics works seamlessly across multiple databases which can contain data of different types coming from different vendors / sources
- **Ability to Capture Derived Time Series Centrally:**
  - OneTick provides this ability as an out-of-the-box, standard functionality
- **Speed and Consistency in High Volume Input and Query Performance:**
  - OneTick advocates a two-tier architecture for OneTick with streaming data collection performed on one box and the OneTick server residing on a second box responding to queries – ensuring that query execution does not impact real-time data collection
- **Efficient Time Series Query Capabilities and Built-in Data Mining Functions:**
  - Fully multi-threaded, OneTick is designed to allow its server to handle large volumes of concurrent queries from multiple users and applications very efficiently while running on standard hardware. As a benchmark, in three of our largest installations that have deployed OneTick as a central tick data solution, a single server box is used to address the needs of an entire division or an entire firm
- **Capability to Store a Significant Amount of Time Series History Online:**
  - OneTick has no known data limitations or boundaries. We have customers currently storing more than nine terabytes of data within OneTick, customers storing complete U.S. equities history going back to 1993 (TAQ) and years of historical order book data, as well as customers capturing and storing the entire OPRA feed – all that is needed is sufficient disk space to handle the amount of required data
- **Space Compression and Conflation Algorithms Designed for Time Series Data:**
  - OneTick uses a proprietary, in-house developed data compression algorithm for both our archive and in-memory databases. Data compression typically ranges from 3.5 to 5.5 depending on the type of data. By developing our own compression functionality, we have succeeded in creating a specific solution for tick data and time series data that achieves a higher compression ratio than existing compression functionality (e.g., ZIP) which often causes users to lose the data structure after compression, negatively affecting query performance by returning more data than what is actually needed in the query
- **Standards-Based, Out-of-the-Box Interoperability and Compatibility with Oracle:**
  - OneTick has an ODBC interface enabling customers to use SQL for data retrieval. This functionality includes the ability to use SQL to execute pre-saved native OneTick queries. With



# ONE MARKETDATA

one OneTick customer using the OneTick ODBC driver though Oracle's heterogeneous services, OneTick data is transparently available to Oracle users

## *Data Integration*

OneTick enables users to integrate seamlessly with market data vendors whether it includes historical data or real-time market data feeds. Additionally, OneTick can connect easily to complimentary solutions for extensive backtesting, quantitative analysis and trade execution.

Specifically, OneTick delivers:

- **Ability to Integrate with Market Data Systems (e.g., RMDS, WOMBAT):**
  - OneTick includes off-the-shelf adapters to many real-time market data systems including Reuters, Wombat and Interactive Data (IDC). OneTick also provides its users with a complete loader API allowing them to build adapters to any data source. Upon request, OMD engineers can build additional adapters for its customers; development time ranges from one day to three weeks depending on the complexity of the feed.
- **Capability to Load Bulk Data in Batch Mode and Real-Time Data in Real-Time:**
  - OneTick is a twenty-four hour system that collects and loads real-time data into the OneTick in-memory database in real-time, making it immediately available for queries and analytics. Similarly, users can load bulk data in batch mode at any time during the day or night with no impact on real time collection
- **Support for Connectivity Using Native Drivers or ODBC to Major RDBMS Vendors:**
  - OneTick provides a built-in primitive to query the data out of an RDBMS through ODBC. This feature allows users of OneTick to process data from RDBMS and high-frequency data from OneTick databases within a single analytical query
- **Support for Connectivity to Prebuilt In-House Applications with SDK, SOA or SaaS:**
  - OneTick provides an SDK for C++, C#/Net, Java, Perl, Matlab and SAS. An ODBC driver is also provided. Queries built in the OneTick GUI tool are saved as files and can be called from any program written in the above languages

## *Scalability & Performance*

Built by former end-users who understand the importance of data management performance, OneTick provides its users with a highly scalable and efficient platform to meet all data management and analysis needs.

Specifically, OneTick provides:

- **Real-time Infrastructure for Low Latency, High-Volume Data, and Event and Transaction Management:**
  - OneTick collectors and servers scale naturally though scaling is rarely needed. While users can add extra collectors/servers to the system seamlessly, our largest clients' needs currently are met by a single 4-core or 8-core server. Collection of all tick data (excluding OPRA) is typically performed by a single 4-core box. For users collecting OPRA data in real-time, OPRA collection is performed on two 4-core boxes. OneTick's proprietary architecture is fully multi-threaded and allows for great concurrency in query execution. Extra parallelism is achieved by utilizing



# ONE MARKETDATA

multiple cores on the server to execute a single query. In addition, a database can span multiple hosts, allowing simultaneous query processing by multiple hosts

- **Easily Scalable and Expandable Architecture for Additional Data Storage and Complex Queries:**
  - OneTick can easily scale collectors and servers by simply adding more disk
- **Optimized Solution for Real-Time Operations on Time Series Data Sets:**
  - OneTick is optimized to run on large volumes of historical data as well as on the most recently collected data from real-time sources. Utilizing memory management, OneTick analytics are written to operate on real-time data streams, and provide identical behavior for both historical data sets and data that arrives in real time. As such, OneTick functions as both an engine for processing of high-frequency historical data as well as a real-time event-processing engine
- **Low Latency System that Scales to Amount and Use of High-Frequency Tick Data:**
  - OneTick is designed as a high-performance low-latency system which scales naturally with growth in data rates or volumes of queries

## *Security*

Understanding that each institution has its own policies and safeguards when deploying new solutions, OMD has developed a solution that utilizes the latest technologies for authentication security and access control.

Specifically, OneTick provides:

- **Unique User Identifiers:**
  - OneTick supports access control, where users can be assigned privileges to query certain databases and can be restricted from executing certain analytics. Additionally, the OneTick server can act as a repository of per-client licenses to enable access restrictions, such as allowing queries only for specific users and from specific locations
- **Password Standards:**
  - OneTick employs a flexible access control mechanism that can be easily extended to support users' unique password and security requirements
- **Audit Trails for Access Control:**
  - OneTick supports multiple logging levels. Under the highest logging levels, all query information including user, time, object, type, exceptions, etc., gets logged by the server. Data loaders and collectors also log their activity, as well as any warnings and exceptions
- **Source Authentication and Location Verification:**
  - Through OneTick's per-client license repository, managers can restrict access to specific users sending queries from specific hosts or domains. Password-based authentication can be added easily to fulfill specific requirements

## *Reliability and Availability*

With OneTick, users receive a complete sever and data monitoring infrastructure. Also, since the majority of our customers are using OneTick for real-time trading and decision support, OneTick has been designed with high availability and redundancy for collection and server processes.



# ONEMARKETDATA

## Designed by Users, for Users

The founders of OMD understand data management as well as the many challenges associated with tick data capture, storage and analysis. Having previously built a system which supported all of Goldman Sachs' tick data needs globally (i.e., statistical arbitrage, transaction cost analysis, legal/compliance, algorithmic and electronic trading, pre-, during-, post-trade analytics, etc.), they have the extensive experience necessary to understand what users need in their tick data and time series database solutions.

Before the very first line of code was written for OneTick, OMD knew they were building a solution for financial time series data, and not a generic database product. Because all the inherent problems with global market type data were encountered and solved multiple times in the past, we were able to design and architect OneTick to fully account for all key global market data storage issues including corrections / cancellations, corporate actions, calendars, exchange holidays, symbol changes, symbol mapping, futures rolls, options chaining, etc.

OneMarketData's "OneTick" application is a completely proprietary, enterprise-wide data storage solution that has three core architectural components: an archive database (i.e., long-term data storage capability); an in-memory database (for caching data collected in real-time); and lastly, an event processing engine (CEP).

The archive database and in-memory database are designed with the flexibility to store any type of tick data, time series data (economic, security related, etc.,) reference data and any other type of financial data. Architecturally, the OneTick application has no limitation on the amount of data (this is more a function of the amount of available storage both archival and in-memory) and no restrictions on the type of data loaded in OneTick as it supports the storage and retrieval of all asset types, financial data, time series, etc., globally.

In addition to being a rich, flexible enterprise wide market data storage and retrieval platform, OneTick aides in the analysis of data by providing 70+ pre-built parameterized analytical "primitives." Analytical primitives range from low level functions (retrieve data, filter data, merge and play back data) to intermediate functions / aggregations (VWAP, Open, High, Low, Close) to more advanced analytics like order book building and portfolio tracking. The set of primitives can be extended by user-defined ones (written in C++, JAVA, etc) through a specially provided API.

OneTick offers a very rich business logic layer addressing concepts such as corrections/cancellations, corporate actions/symbol continuity/symbol maps, futures rolls, options chaining, etc. Primitives are combined and applied to historical or real-time data to build rich queries that make financial analysis quick and easy to do with OneTick. Lastly, our CEP engine provides the ability to take the exact same queries that typically act in a request-reply mode and register them in the CEP engine to act in listen-react mode. The query is essentially a subscription that listens to streaming data and publishes results when the query conditions are met.

### *Simple Look Up and Visualization*

OneTick provides a very powerful GUI, immediately empowering users to build the same set of queries that are built from C++/JAVA/C#, etc. graphically and without programming. After testing, queries can be saved as files, and later can be called from any of the programming environments



# ONE MARKETDATA

(C++, JAVA, Perl, C#.NET, etc), statistical packages (SAS, Matlab, R, S+, etc.), ODBC, Excel, command line, etc. By eliminating the necessity of programming and providing extensive flexibility with query exporting, OneTick reduces the learning curve while increasing time to production.

Through the OneTick GUI, users can view query results in both table and graphical format as well as embedded and call queries from other applications such as Excel.

Unlike much of the competition, OneTick also has no special language to learn, eliminating the need for expensive intermediate language specialists and consultants. Technology people, quants, analysts, researchers, traders, etc., can all start using OneTick almost immediately with minimal training.

## *Consistency in Query Performance*

From inception, OneTick was designed as a high-performance architecture to address concurrent needs of hundreds of users. It is fully multi-threaded with extra parallelism available so each individual query can specify how many cores it must utilize upon execution. Data is stored as a time-series of sequential ticks. Databases containing billions of ticks per day are currently in use.

OneTick supports the concept of a tick-server, with which clients interface to analyze data. Tick-servers can be deployed in a scalable manner, meaning that as the load increases, new servers can be deployed or cores added for load-balancing purposes. However scaling is only needed when there is a fundamental change in usage (say number of users goes from 100 to 1000). OneTick's special design (including multi-threading, special memory management, etc.), allows over 100 queries to be executed concurrently on a single server without serious degradation of performance.

## *Time Series Support*

OneTick's 70+ analytical primitives make it quite easy to create derived time series data. Once data is transformed or derived using our analytics or customer provided calculations, the results are easily stored in another OneTick db for subsequent retrieval. Derived data can also be generated and inserted into the database in real-time.

OneTick is fully capable of storing different time-series (and non time-series) data types both, in archive and in-memory databases. All of the loaded data is seamlessly available for analytics.

OneTick databases are designed and optimized for very high data volumes, especially market data. OneTick supports tick processing rates over 20mln ticks/sec on low-end 4-core hardware.

OneTick utilizes proprietary compression to reduce disk requirements for tick data storage achieving compression rates ranging from 3.5 to 5.5 times over uncompressed data. Our compression functionality uses statistical properties of a time series and is optimized for market data. The compressing/uncompressing is extremely fast and has no negative effect on the query performance.

## *In-Memory Data Management*

The OneTick query server process maintains in-memory reference data cache to enable efficient queries that require use of both high-frequency tick data and low-frequency reference data. Also,



## ONE MARKETDATA

results of archive queries can be cached in the memory of the server on demand in order to boost performance upon subsequent query.

All analytics are optimized for streaming data processing, and almost all of them use very little memory, reducing the number of page faults and thus improving the performance.

### *Data Migration and Failover*

OneTick supports an extremely flexible schema, where each time series can have its own structure, and that structure can change every day. Therefore, there is no data migration problem and this also uniquely addresses the issues of schema evolution.

Failover is available for collection and server functionality through two redundant collector boxes and server boxes. By configuring collector boxes to receive identical feeds that mirror each other, the processes on each server box access the data from both collectors and can detect outages as well as data transmission gaps. Detected data gaps from the primary feed are immediately filled from the secondary source. OneTick also can stitch data gaps in any symbols using the data from the collector that did not have that gap. If a client process cannot connect to the primary server box, it gets automatically redirected to the backup server box.

### *Uninterrupted Uptime*

With a unique architecture that processes updates automatically, OneTick is able to provide its users with continuous use of all processes without the need for weekly archive updates and that can take users offline for unspecified periods of time. Additionally, since users can deploy OneTick on a number of different operating systems including Linux and Solaris, users are not forced to use Windows operating systems which can require additional system restarts that restrict uptime.

As such, OneTick's tick server and analytics engine are available to its users twenty-four hours a day, seven days a week without exceptions.

## **Leading the Way in Tick Data Management**

By offering an innovative database designed specifically for market data along with an extremely intuitive and user-friendly GUI, OMD has designed a tick data capture, storage and analysis system that streamlines the data management process and reduces the total cost of ownership by eliminating expensive product specialist fees. With OneTick, institutions often can begin querying data immediately – with minimal training and installation time.

Designed by data management experts for deployment across financial institutions that are both large and small, OneTick offers unmatched functionality to users looking to increase efficiency and maximize the value of their market data investments.