Slashing Market Data Costs by Employing a New Paradigm for Market Data Development

TABLE OF CONTENTS

1 INTRODUCTION

The future of Market Data Development belongs to a new generation of tools that will enable organizations to dramatically reduce costs, improve application quality, and realize huge efficiency gains. Consider a market data paradigm that would:

- Eliminate the majority of real-time exchange costs.
- Provide market data 24/7 to QA, UAT and Development groups, even when exchanges are closed.
- Allow the market data stream from notable trading days to be saved and replayed at any time.
- Provide off-shore resources with market data anywhere, anytime in a simple and low-cost manner.
- Completely isolate the production environment while still allowing Development, QA, and UAT access to production-like data.
- Reduce the number of circuit connections and feed handlers required.
- Make it possible to create specific market scenarios against which to test market data–dependent logic.
- Certify applications perform correctly under failure conditions such as stale data and feed down.
- Simplify capacity planning by simulating increased update rates.
- Allow QA and development to build automated market data regression test suites.
- Make it possible to carry out a true production-like DR test with a full-blown data set and not just a small subset.
- Provide low-cost latency monitoring all the way to an application, thus allowing administrators to pinpoint exactly when and where latencies occurs.
- Provide a single rationalized, low-cost solution for capturing tick-by-tick and bar data for all projects across an organization.

This whitepaper describes a new approach to market data development that achieves all of the above, resulting in huge reductions in overall market data costs. Through this approach, a sample development community with 100 active developers and associated operator support would realize a combined hard and soft dollar saving of between $1 - 3 million per year. (See Section 4 for an itemization of these savings.)

Requiring minimal upfront investment, proportional savings are available to firms of all sizes. Indeed, as market data processing software becomes ever more central to the trading of all asset classes, firms that adopt this new paradigm will obtain a substantial competitive advantage over their competitors who delay in making the transition.

2   MARKET DATA DEVELOPMENT TODAY

The vast majority of today's market data development relies on real-time production data flowing over production networks. At some organizations, separate development networks distribute real-time data or attempt to cache and delay real-time updates.

These approaches lead to a number of critical limitations. Perhaps most consequential are the constraints placed on development by the inability to vary the content and update frequency of live testing data. Likewise, without a facility to repeat or "replay" market scenarios, development and testing are largely dependent on the operating schedules of individual exchanges. In the U.S. equity markets, for instance, the conventional approach limits development and testing to the hours between 9:30 AM and 4:00 PM Eastern Time.

Moreover, delivering real-time market data to users is an extremely expensive undertaking. In the past, firms have managed these costs by negotiating non-professional fees with receptive exchanges and by custom building delayed environments. These solutions are limited because:

- They require extensive contact with potentially numerous exchanges.

- Delayed data cannot be available on a 24/7 basis.

- The frequency and content of data cannot be varied.

3   GAINING CONTROL OF COSTS

The world of market data development is going through rapid and fundamental change. Among the key trends driving up costs and complexity:

- The trend toward electronic trading means market data development has moved beyond simple applications for displaying market updates.

- The trend toward off-shore development means market data must be available to development teams in all parts of the globe and at all hours.

- Trends in global volume and volatility have led to soaring update rates, which taxes applications as well as infrastructure, and makes data latency an ever more central business issue.
- Exchanges have become increasingly aggressive about charging for market data that is delayed and/or intended for non-professional use.

In this challenging environment, market data and development managers are turning to recorded data, coupled with sophisticated development tools, to service their firms' market data development needs. In addition to dramatically improving developer and QA efficiency, these innovations substantially reduce the hard costs associated with building and testing any applications that consumes real-time market data.

CodeStreet's ReplayService is an example of a new generation of tools that leverage the full power of recorded data to create dedicated development environments for market data. The myriad advantages to this new approach are detailed below.

3.1   Dramatically Enhanced QA Capabilities and Efficiency

In the traditional development paradigm, test data is limited to the real-time stream of tick updates flowing on a development network at any given time. This approach allows developers to test their market data subscriptions, and little else—it is wholly inadequate for testing the logic of their applications.

Indeed, what is critically absent from the conventional model is the ability to create specific market data scenarios, with the appropriate content and update rates, that can be applied to applications in an automated and repeatable manner. Without this capability, development groups must resort to "testing in production"—a decidedly haphazard and unrigorous QA method.

In summary, because they are limited to using live data, developers of market data applications face these common obstacles:

a) Incorrect or inappropriate data on the network.
b) No way to test for error conditions on the network.
c) No means of varying data rates when testing applications.
d) No means of testing particular scenarios of interest.
e) No access to market data when working away from the network.

The new development model, as epitomized by ReplayService, addresses each of these obstacles head-on. At its core is the philosophy that modern software testing

methodology must be brought to the world of market data: In short, developers and testers must be able to create specific market scenarios on demand that can be persisted for use in automated testing.

Under the new model, improvements in the rigor and completeness of testing are matched by increases in the overall efficiency of development. Equipped with what is in effect his or her own personal market data platform, each market data developer in an organization becomes significantly more productive, requiring far less "handholding" by market data administrators. In fact, a dedicated development environment and tool suite for market data is estimated to improve developer productivity between 5% and 25% on average.

3.2   Market Data Available 24/7 in Any Location

As development goes global, the challenge of providing market data to distributed software teams becomes ever more pressing. With typical development networks, once a market of interest is closed, data will stop flowing on the network—time zone differences thus mean market data is often unavailable during developer working hours. For U.S. Equities markets, trading stops as early as 4 PM Eastern Time, cutting off valuable development time for groups across the country or overseas.

As a result, banks and other financial institutions are exploring models that make market data available at all times, in all locations, but without the live feeds, complex infrastructure, and administrative burden. Used with the appropriate tools, recorded data allows any organization to establish a market data presence in the remotest of locations, making it possible for far-flung development and test groups to work even when the relevant markets are closed. Simulated real-time market data streams are available at all times—no live feed, distribution infrastructure, or administrator support is needed. By leveraging modern development platforms built around recorded data, larger financial institutions can potentially save hundreds of thousands of dollars a year in off-shore development costs.

3.3   Low-cost Dev, QA, and UAT Environments

Providing market data to developers around the globe can exact a daunting burden on infrastructure. Licensing costs, hardware costs, and most crucially, circuit costs are all difficult to absorb in the home market, and can well be prohibitive for smaller distributed organizations abroad.

Contrast that with the new recorded-data model, which allows Dev, QA and UAT environments to be established quickly in any location, with little operational support

required. The cost savings for organizations that have taken this approach have been hundreds of thousands of dollars annually.

Indeed, by giving each developer control over a simulated market data platform, the new model eliminates the need to have a Thomson Reuters Enterprise Platform (TREP) or competing platform deployed in multiple locations. Users are able work from remote locations—including international offices and their own homes—without having to connect to a live network. As a typical development network with associated feeds and maintenance costs at least $50,000 per year per location, the infrastructure savings from using recorded data can be quite significant.

3.4   Reduced Real-time Exchange Costs

In recent years, fees paid to exchanges have become an increasingly substantial burden on market data development. These fees are not standardized and will vary widely from firm to firm depending on their leverage in negotiations with each exchange. Exchanges now typically charge for any real-time access to market data, regardless of whether it is actually traded on, or just used for development purposes. Even when a particular exchange does not mandate a fee for non-business use of their data, consolidators that deliver this data often will.

Real-time exchange fees for developers vary widely, and are particularly expensive at European and Asian exchanges. At the time of this writing, fees were as high as 168 GBP/month for Level 2 data from the London Stock Exchange; even the average fees for smaller markets were in the neighborhood of the London Metal Exchange's 48 GBP/month. At the time of this writing, the LSE does not provide a delayed data fee, thus requiring all developers to pay the full price for live feeds. Given the sheer number of developers at even medium-sized organizations, such monthly fees can quickly constitute a major ongoing business expense.

The recorded-data model described here allows organizations to drastically reduce or even eliminate altogether the amount spent each month on real-time exchange fees. This represents hard-dollar savings that can be realized immediately. Depending on the size of the organization, these fees can run into the millions of dollars annually. Consider, for instance, a community of 100 developers with access to data from a small sample of exchanges such as the CBOT or CME ($90.95/month) and Euronext Cash Level 2 ($92.02). Moving to recorded date would immediately result in a monthly hard-dollar savings of over $10,000. A recent ReplayService customer was able to save $38,000/month in real-time exchange fees for the equity derivatives desk alone, resulting in an ROI of less than 6 months.

3.5   A Rationalized Approach for Tick Capture

Beyond the cost savings related to live market-data feeds, an intelligently designed recorded-data platform can also deliver a number of ancillary improvements in a wide variety of business areas.

Prominent among these is the case of tick capture—storing market data in a relational database is a key part of countless projects throughout the financial industry. As it stands today, if a costly enterprise-class solution for tick capture is not appropriate, such needs are answered through custom development. In an organization of any size, this soon results in numerous custom solutions, each of which requires significant upfront and ongoing expenditures to properly maintain.

An integrated environment that provides low-cost tick capture facilities holds the promise of making this costly, time-consuming inefficiency a thing of the past. Indeed, by providing an out-of-the-box tool to serve as the starting point for any project requiring tick capture, however simple or complex, individual groups need never again resort to custom development for market data capture. The potential cost savings from this rationalization of a standard activity are substantial, both in terms of development expenses and, especially, long-term maintenance costs.

3.6   Low-cost, Lightweight Latency Monitoring

Leveraging recorded-data also provides an elegant solution for monitoring data latency, which in today's electronic trading world has become increasingly crucial to success. Currently, most stand-alone latency-monitoring solutions operate using specialized hardware that is expensive to implement and complicated to administer. Besides their high cost and complexity, these hardware-based solutions are unable to accurately reflect the true latency experienced by applications.

By storing and comparing the timestamps appended to updates at various locations on a network, a recorded-data platform makes it possible to integrate latency monitoring into a standardized market data development environment. This lightweight, software-based approach is achieved at a fraction of the cost of hardware solutions, and allows organizations to monitor the full trajectory of latency, up to and including the endpoint of a vendor API. With the ability to track market-data latency both from an originating exchange and from particular data feeds (i.e., across TREP or other market data delivery system), it becomes possible to instantly chart various latencies against time, as well as to set up software alerts for when thresholds are breached.

3.7   An Integrated, Transparent Toolbox for Market Data

The inefficiency associated with market data development can be traced, perhaps above all else, to the opacity of existing market data platforms, which make it discouragingly difficult to access information on the data flowing over these networks. The productivity of developers and operators alike can be dramatically enhanced by a solution that integrates a wide array of easy-to-use tools into a single standardized development environment for market data.

Indeed, the day-to-day activities of today's market data users demand, among others, tools that unpack options chains, tools that present data in grid form, tools for isolating update-rate peaks, and tools for tracking data-availability issues and determining their causes. It is presently not uncommon for simple data-content questions to entail day-long turnarounds; an integral element of the new recorded-data model is the elimination of these hidden costs through the widespread use of thoughtfully designed, wizard-driven tools.

3.8   Reduced Support Needs

The conventional model, whereby development communities depend directly on live distribution networks such as TREP, necessitates a large team of market data administrators to operate and support the platform. These administrators must continuously manage feeds, monitor infrastructure components, and train developers—a costly and time-consuming proposition.

The new paradigm of recorded data and software-simulated distribution platforms would make developers entirely self-sufficient with regards to their market data needs. The widespread deployment of a development environment and associated tools based on these principles is estimated to cut the support time required of market data administrators by at least 25%.

3.9   Automated Testing

In the vast majority of financial institutions, market data testing is still performed manually; due to the difficulty of sourcing and repeating particular market scenarios, such methods can be dangerously imprecise. Testing, after all, typically takes place when only random market data is available either in live or delayed form, leading to haphazard guessing and estimation around application behavior.

In contrast, the recorded-data model allows for precise, repeatable market data scenarios to be applied to any application under test. This makes it possible to include market data–consuming applications in daily build and regression-test exercises, thus finally bringing market data development into the world of modern programming. The result is dramatically improved reliability and speedier development.

3.10   DR Testing and Capacity Planning with Realistic Market Conditions

With few exceptions, disaster recovery and capacity-planning testing must occur after hours, when exchanges are closed. Thus, they take place in an environment unrepresentative of typical real-world market conditions. The modern market data environment envisioned under the recorded-data model would allow live feeds to be simulated using actual production data, thereby enabling tests with realistic DR failure scenarios and high–update rate scenarios that can be repeated on demand.

3.11   Streamlined Issue Resolution

As a byproduct of recording production data, organizations are able to isolate and analyze data-quality problems at an unprecedented level of detail. With an easily accessible audit trail of all market data flowing on a distribution network, administrators are able to isolate precisely the data responsible for any issue under investigation. Problems in data content or distribution can thus be identified and resolved quickly, before they become ongoing detriments to developer and production productivity.

Beyond application and infrastructure testing, recorded data can also be used for a variety of compliance, risk, and reference-data purposes. In short, the consistency and reliability of having a single source of recorded data dramatically consolidates market data activity across an enterprise, realizing enormous cost savings in the process.

4   EXAMPLE COST REDUCTIONS

The following exercise demonstrates the savings that can be achieved by deploying a modern market data development environment that leverages recorded data in the manner described above. The result is a remarkable return on investment, with leading firms seeing potential savings in the millions of dollars.

The table below itemizes the estimated savings associated with each of the areas detailed above. As shown, it is possible for a community of 100 developers to reduce market data expenditures, in combined hard dollar savings and efficiency gains, by up to $3 million per annum. This is achieved at a cost of roughly 10–20% of the savings. Larger communities should expect to see correspondingly larger cost reductions. The savings are broken down into $870,000 in hard dollar savings and the remainder arises from efficiency savings. Therefore, realized savings will be in the $1-3Million range depending on how effective the organization is in translating efficiency improvements into genuine cost reduction.

Though rough approximations, the figures below should make obvious the profound inefficiencies of market data development under the prevailing model that relies solely on live data flowing over a development network. They illustrate a stark reality: As the market data community moves toward sophisticated tools for delivering, replaying, and modifying recorded data, organizations that fail to adopt the modern paradigm will find themselves at a cost and competitive disadvantage that becomes increasingly acute with each passing year.

Savings include $870,000 in hard dollar savings and up to $3M dollars in combined hard and soft annual savings for 100 developers using real-time market data

|	Assumption	Yearly Savings
Exchange Fees	$100/monthly access/user	*$120,000[*1]*
Development Networks	No need for dev network in remote locations	*$50,000*
Market Data Administrator Support	25% efficiency gain	$40,000
24x7 development	Provides 5% efficiency gain	$500,000
Off-shore work	No need to route market data to off-shore locations	*$500,000*
Reduced custom coding	All market data capture is done with common low-cost tool	$250,000
Automated testing	Decrease QA expense 25%	$1,000,000[*2]
Enhanced developer debugging and general efficiency	5% efficiency gain	$500,000
Faster issue resolution	5 issues at $10K per issue	$50,000
Reduced tick capture costs	Avoid buying expensive third party tools	*$100,000*
Reduced latency monitoring costs	Avoid buying expensive third party tools	*$100,000*
TOTAL YEARLY SAVINGS		**$3,210,000**
*1 These fees will vary by organization, exchanges of interest, and usage. *2 This number is highly variable based on the nature of the work. Automated trading groups are those most likely to benefit from automated regression testing. Items in italics reflect hard dollar savings		

5 WHAT IS REPLAYSERVICE?

ReplayService is the first modern development environment built around the paradigm-shifting principles described in this document. It was designed for any and all users of real-time market data, including developers, QA groups, market data operators and administrators, quantitative traders, business analysts, and compliance and risk officers. Through an elegant graphical interface, ReplayService allows users to contrive market data scenarios of interest and play them back into their applications by simulating a Thomson Reuters Market Data System (TREP) or a Reuters Data Feed (RDF).

Every ReplayService user thus controls what appears to be a complete market data platform, one in which they can select data content, data update rates, and specific time periods or market activity of interest. Augmenting this core functionality is a complete array of market data–related tools that enhance productivity and make it possible for even novices to work effectively with market data in electronic form.

ReplayService consists of a client application that runs on each user's local machine and a shared Recording Backbone that provides the recorded data for use by the client applications. A centralized facility shared between all ReplayService users, the Recording Backbone allows data to be captured once and then utilized for a wide variety of purposes. ReplayService is designed to scale elegantly from the simplest deployment of a single client on a simple laptop machine, to an enterprise-wide deployment supporting hundreds of users.

The diagram below shows the relationship between the market data platform, the Recording Backbone, and the ReplayService client applications.
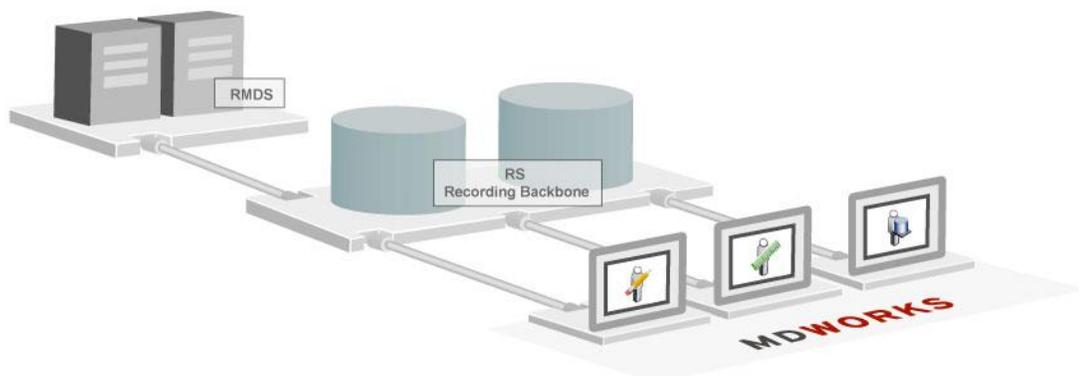
6   REPLAYDATA.COM - DATA SAMPLES FOR USE WITH REPLAYSERVICE

To provide the data necessary for working with ReplayService, Thomson Reuters and Codestreet have combined to offer a web based service called replaydata.com. This web site provides samples of Thomson Reuters tick history product for use with ReplayService. Users of ReplayService can register on this public web site and download samples of tick history data, import them into ReplayService using the tick history import wizard and subsequently replay the imported data.

Replaydata.com includes data samples from each of the 10 tick history asset class sections. It can also be used as a repository to share proprietary ReplayService recordings within a given firm by uploading recordings to a firm specific area within the web site.

7   SUMMARY

The market data development world has up to now been without comprehensive tools suitable for building modern trading applications. Development methodologies continue to rely on production or delayed data flowing on distribution platforms. This decades old paradigm is proving to be highly costly and inefficient and produces poor quality applications.

In the new paradigm, a common set of recorded data is leveraged in all development activities allowing for automated testing of the most complex solutions using repeatable scenarios. This ability to leverage modern programming methodologies that focus on test driven development is an essential capability for those firms wishing to remain competitive.

In an environment where financial services firms are looking for material cost savings, market data is often the first place to look. While reducing business use of the data is generally not an option, reducing overall market data spend by addressing the developer community is a simple and painless way to cut costs while also improving the organization's capabilities.

The following acronyms are frequently used in this document:

Acronym	Meaning
API	Applications Programmer Interface
TREP	Thomson Reuters Enterprise Platform Market Data System
RS	ReplayService CodeStreet's market data development environment

8 LICENSE AGREEMENT

© Copyright CodeStreet 2008. All rights reserved. The information contained in the documentation is subject to U.S. and International copyright laws and treaties. Any copying, modification, re-publication or re-distribution of CodeStreet content or of any content used in this whitepaper, is expressly prohibited without prior written consent of CodeStreet.

All brand and product names are registered trademarks or trademarks of their respective holders.

This publication is provided "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

This publication could include technical inaccuracies or typographical errors. Changes are periodically added to the information herein; these changes will be incorporated in new editions of this publication. CodeStreet reserves the right to make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time.

Please contact us if you have any comments or concerns regarding CodeStreet ReplayService Market Data Works™ and the accompanying documentation.

codestreet

Address	CodeStreet LLC,
	11 West 42nd Street, 9th Floor
	New York, NY 10036-8002
Telephone	+1 (646) 442-2800
Facsimile	+1 (212) 620-CODE
Technical Support Email	support@codestreet.com
URL	www.codestreet.com