

## TRADE TRENDS

# Big Data: Manage it, don't drown in it

BY MARCUS KWAN

**We are swimming in an ever-expanding ocean of data. Its exponential growth is challenging our ability to process that data effectively to exploit what it contains.**

Wikipedia describes Big Data as “data sets that grow so large and complex that they become awkward to work with using on-hand database management tools. Difficulties include capture, storage, search, sharing, analytics and visualizing.”

Making decisions based on too much information that is not properly managed and classified can be just as dangerous as making decisions based on too little. This is the challenge of Big Data.

Direct connection vs. aggregator  
Collection & storage  
Buy vs. build

Big Data has become a common buzz phrase in the trading industry and the implications for market professionals, from the individual trader to large firms and trading desks, are real and significant. Big Data confronts us with several questions: How do we tackle collection and distribution? How do we take data soup and manage it in such a way that it is informative and offers cues to act? When profitable opportunities rely on market inefficiencies, and when

## BYTE ME

We all have heard of kilobytes, megabytes and gigabytes, but the exponential growth of Big Data has forced data processors to look at larger units.

1 Bit = Binary Digit
8 Bits = 1 Byte
1024 Bytes = 1 Kilobyte
1024 Kilobytes = 1 Megabyte
1024 Megabytes = 1 Gigabyte
1024 Gigabytes = 1 Terabyte
1024 Terabytes = 1 Petabyte
1024 Petabytes = 1 Exabyte
1024 Exabytes = 1 Zettabyte
1024 Zettabytes = 1 Yottabyte
1024 Yottabytes = 1 Brontobyte
1024 Brontobytes = 1 Geopbyte

Source: [whatsabyte.com/](http://whatsabyte.com/)IBM Dictionary of Computing

these inefficiencies may appear only for moments, how do we equip ourselves to manage these processes fast enough to seize them? Big Data no longer rings hollow as a buzz phrase but as a solid challenge that needs to be addressed. Before

we dig into ways to navigate this new world, let's look at the evolution of data.

- **Data past:** In recent history, traders managed market data with pen and graph paper. Data granularity was confined to limits of open, high, low and close prices, and data management was restricted by physical capacity. When parsing price data to an analyzable and actionable format was a manual process, the scale of manageable data was limited. Data delivery was slower, which constrained the amount received in a given period of time — over the past 35 years we have moved from data delivery via floppy disk, satellite and low-bandwidth internet connectivity. The seemingly finite nature of data in the recent past makes the current information explosion appear overwhelming.

- **Data present:** The futures industry now finds itself in a situation of ballooning data volume and low-latency, or even ultra-low-latency, access. As with nearly all things technological, we have progressed rapidly in an impressively short time period. Trading decisions now hinge on a

huge quantity of market information. Electronic exchanges offer quick access to highly granular data. Tick data is updated in microsecond time frames. Real-time price action for similar or related instruments listed on different exchanges is visible on a single trading screen. In addition to the speed and complexity of structured data, factor in unstructured data such as real-time news from traditional sources and non-traditional sources like Twitter, which increasingly provides news before it is news. It is possible to add sentiment gauging on social media platforms to the trader's or firm's ordinary workflow. Now consider the number of symbols in that workflow and multiply accordingly. Thus our future: Big Data.

- **Data future:** These factors distill to two core Big Data challenges: 1) Effectively and efficiently collecting and distributing data and 2) reliably and intelligently analyzing the data and executing on it.

### Collection and distribution

Data access and distribution are the most straightforward of the challenges, but intelligent strategies can be rewarded with cost savings and optimized performance that translates to competitive advantages.

Data access can be via direct connection or through an aggregator maintaining its own direct connections (see "Buy vs. build," page 34). For algorithmic traders, latency introduced by anything less than a direct connection may be a deal-breaker. On the other hand, direct connectivity is not easily scalable and its speed comes at a price. Each new connection requires parsing a new data protocol, so development and maintenance costs quickly can become prohibitive for companies requiring multiple connections.

Conversely, aggregated connections come with the advantage of a single parser obtaining data from multiple sources, packaged and ready to plug into a firm's workflow. There may be some reduction in speed, though often in microsecond time frames. Additionally, for full value, real-time data requires historical context, and this need adds up-front purchase costs and data storage burdens. Aggregated connectivity may offer a way



to skirt this mechanical archiving challenge and a significant data purchase cost. For a market data provider, customer requirements may compel a middle ground where direct connectivity to key markets is essential and ancillary data sourced from third-party aggregators is perfectly acceptable.

Regardless of direct or aggregated connectivity, if an individual or organization must redistribute data, this poses additional hurdles. Quick and reliable data redistribution to a user base demands server farm performance. Hardware costs and scalability issues can be significant. Part of the solution is straightforward: Farms simply must have the capacity to serve clients effectively. The other part — hardware optimization — is where dealing with Big Data intelligently can yield a competitive edge. With flat file storage, SSD, multithreading and other methods, it is possible to make redistribution more sophisticated and less a brute-force issue of server farm expansion.

### Analysis and execution

For market professionals, collecting, storing and managing Big Data are increasingly necessary challenges, but these are only requisite steps to addressing the heart of the issue. Truly valuable Big Data solutions must allow for robust visualization of data that enables constructive analysis, whether it is in the analysis that powers trading decisions or the tools that provide a clear picture of risk. With this sort of solution in hand, it is possible to execute decisions with bottom-line impact more reliably. This is the greater challenge, requiring the synthesis of detailed analysis with big-picture thinking.

The complexity associated with the scale of market data in the futures industry can mean that signals are buried in noise or visible only when analysis incorporates an adequately large or diverse data set, so a pressing question arises: Do you have the in-house tools and talent to make sense of Big Data? For the individual trader, this means continuing

to cultivate intimate market knowledge, but perhaps incorporating new sources of information and new data visualization tools into methodology as well.

To implement productive and reliable analytical systems, firms need savvy personnel who understand data nuances and can formulate the right questions (i.e., it is critical to have a sense of what you are looking for before plowing head-long into data). These may be individuals who are deeply rooted in mathematics and quantitative analysis, but it is important that the end result is a system that provides decision makers and business drivers, who may or may not be so fundamentally versed, with a broad, accessible and informative picture of complex data. Here, decision makers who necessarily are not versed highly in quantitative analysis avoid situations where they are presented solely with the end results of calculations and left to make choices based largely on trust. They are, rather, able to bring their own dynamic analysis and decision-making to bear.

Increasingly, advanced visualization tools allow traders to expand their market view simultaneously while speeding and streamlining workflow. Consider this challenge in the market data world: A traditional decision-making process often takes a single symbol or a small set of symbols, each with very granular real-time and historical detail, and then adds varying degrees of complex analytics. It is increasingly necessary to expand the scope of this process to include many



For more technology articles, go to [futuresmag.com/Technology](http://futuresmag.com/Technology)

symbols, often in many markets, while maintaining robust analytics and deeper correlations. This becomes computationally expensive very quickly, so traders and firms are confronted with a balancing act in their workflow: Processing the number of markets and conditions that satisfy strategy requirements on one hand against a burden on speed on the other. Solutions to deal with this sort of Big Data hurdle already exist, for example, in the form of programs that condense data that previously may have occupied hundreds of charts and multiple moni-

## BUY VS. BUILD

You can access data from many different sources and there are pros and cons on how you manage that access.

### Direct connectivity

#### PROS

**Speed** – direct connections introduce the least amount of latency

**Stability** – if one exchange connection is lost, other direct connections may remain intact

#### CONS

**Cost** – expenses include those related to infrastructure, expertise/personnel and maintenance

**Scalability** – a new parser is typically required for each new connection

### Aggregator connectivity

#### PROS

**Breadth** – coverage of multiple data sources can be obtained through a single connection

**Consistency** – data received from multiple sources is likely available in a standardized format

**Cost** – infrastructure, expertise/personnel and maintenance costs can be significantly lower than those associated with direct connectivity

#### CONS

**Stability** – a single point of failure exists if all data is received via a single connection

Source: CQG

tors into an accessible, single-screen view. Data that otherwise may have been very difficult or impossible to manage can now be parsed by the human eye.

### Unstructured data

For unstructured data, big questions remain. As an example in the social media realm, a tweet aggregator-of-choice may have a dedicated display in a trader's array of screens, but how does this trader sort through 340 million tweets per day to find relevant information? An apt historical analogy may be that of market news where previously unstructured news content now often is available in elementized feeds. The companies developing sentiment-gauging tools to elementize Twitter and other social media sources in a sort of next generation way face a far greater challenge not only because of information volumes, but also because so many more sources contribute to the stream. Thus, what we may see are tools that allow firms and individuals to configure sentiment rules based on multiple sources.

Although the challenges presented above are significant and the solutions

are evolving, business needs and strategy should inform any approach to Big Data. The first step is to articulate the strategy. It is possible to match strategy with the appropriate technology after questions and objectives are defined in both the areas of data collection and distribution as well as analysis and execution. With forethought, often it is possible to reduce costs associated with data acquisition, storage and dissemination without sacrificing performance. In the more challenging space of analysis and execution, though much of making sense of Big Data may boil down to quantitative analysis, new data visualization solutions should create transparency and accessibility. With the right combination of strategic thinking, technology and effective new tools, futures industry decision makers, from the individual trader to key players in large institutions, will have control over Big Data and be able to see the big picture. Traders and managers will need to, lest they risk drowning in the ever-expanding ocean of Big Data. **F**

Marcus Kwan is vice president of product strategy and design at CQG.

Copyright of Futures: News, Analysis & Strategies for Futures, Options & Derivatives Traders is the property of Summit Business Media and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.