Big Data

Big Data and Smart Trading

How a Real-Time Data Platform Maximizes Trading Opportunities

© 2013 SAP AG or an SAP affiliate company. All rights reserved.

The Best-Run Businesses Run SAP™

Table of Contents

3 **Big Data in Trading**
Big Data, But Big Information?
From HFT to Smart Trading
Big Data: Extracting Big Information
Enriched Data: Rich Trading Opportunities
Integrating the Smart Trading Assembly Line
Real-Time Big Data Processing
Managing Big History
Pulling It All Together
Learn More

© 2013 SAP AG or an SAP affiliate company. All rights reserved.

Big Data in Trading

The convergence of "Big Data" and "smart trading" is leading to the rapid development of new, profitable, and often fascinating trading strategies. These strategies extract and harness information from a myriad of conventional and not-so-conventional data sources, often on a massive scale. Trading organizations employing new Big Data trading strategies are searching for sustained profitability over extended time periods.

In a few short years, Big Data has evolved from a hot technology trend to a significant business reality. Huge volumes of data are now an accepted part of the landscape for financial industry organizations. Some of this data is generated by traditional sources, such as transaction systems, while other data can be traced to sensors, social media, and call center records.

From the early discussions in leading technology journals to the breathless warnings of business publications, Big Data has generated significant attention and focus. Yet since its appearance as a Wikipedia entry in 2010, it's clear the Big Data train has left the Silicon Valley station. The phrase has now entered mainstream lexicon, as evidenced by several ardently discussed panels at the recent South by Southwest Conference. In aggregate, experts seem to agree that harnessing and leveraging Big Data is becoming a competitive game-changer, with some rather unique implications.

For those of us who work in capital markets, dealing with Big Data is hardly news – it's not even new. Daily, we contend with the ultimate reality of Big Data: market data. Many financial organizations are already seeing terabytes of data accumulate each year, and the annual volume is likely to double soon. That's really Big Data.

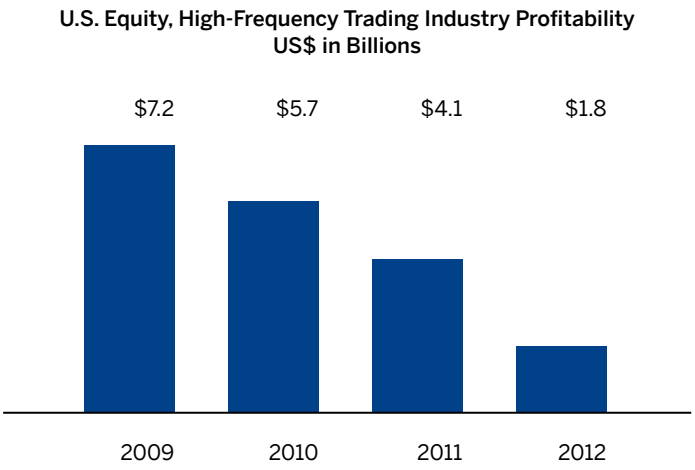
BIG DATA, BUT BIG INFORMATION?

Market data is Big Data. But does it readily translate into the kind of information we can use? In many cases, information density can be rather low. Anecdotally, we have seen a recent ratio of quotes to trades of only about 4,000 to 1. Further, in level 1 large capitalization U.S. equities, fewer than roughly 15% of quotes set a new midpoint price. And level 2 market data has an even lower information density. To be useful and actionable, raw market data must be enriched and refined through a systematic process.

FROM HFT TO SMART TRADING

The rise of Big Data technologies and techniques happens to coincide with changes in electronic trading – namely, high-frequency trading (HFT). Thanks to increasing regulatory issues and higher costs, HFT volumes are waning. As recently as 2009, HFT profitability in U.S. equities was US$7.2 billion. Only three years later, profits had fallen to $1.8 billion. (See Figure 1.)

Figure 1: High-Frequency Trading's Decline in Profitability



Source: Tabb Group

© 2013 SAP AG or an SAP affiliate company. All rights reserved.

Organizations are evolving toward "smart trading" – a practice that entails extracting and analyzing information from a variety of big and complex sources and data types to create a **holistic view of market state.**

There is no doubt that HFT has left a clear and irrevocable mark on the electronic securities markets and technology. And historically, the more successful trading strategies routinely rely on low-latency market access. However, low-latency trading is heading toward commoditization, and therefore there are diminishing returns using low latency as a sole strategic differentiator. Assuredly, some low-latency traders will continue to push the state of the art and prosper, assuming public backlash and looming regulations don't shut down their boxes. However, the majority of trading organizations will need to augment a low-latency approach for sustained profitability.

In their search for new profitable trading strategies, organizations are evolving toward smart trading. Smart trading entails extracting and analyzing information from a variety of big and complex sources and data types to create a holistic view of market state. Smart trading's goal is to extract and exploit unique trading signals, particularly those persistent in time. And while smart trading exploits low-latency market access, the persistence of signals is a key attribute that HFT lacks.

BIG DATA: EXTRACTING BIG INFORMATION

Effectively extracting information from Big Data and integrating this information with other sources lies at the core of smart trading. Consequently, the volume and message rate of structured and unstructured Big Data render it necessary to use a well-defined data enrichment process. The process is key. Data enrichment is primarily a data management and organization activity. Questions that must be answered include the following: what type of strategy are we constructing; what signals are we looking for; what data do we need; which data sources can we use; how do we organize, integrate, and aggregate the data; and which analytics tools are available? Fortunately, rich platform technology is readily available to support a well-thought-through data enrichment process.

Data enrichment processes involve normalization, cleaning, summarization, and aggregation of structured and unstructured data sets. Besides market data, other generally structured data includes reference data, fundamental data, and econometric information. Nonconventional data types that are often unstructured include electronic news and sentiment, weather information, and traffic patterns. The enrichment process transforms this raw data from high-message rate, high-volume, low-information density content to a low-message, low-volume, high-information density content. Enrichment is typically performed on both real-time and historical data. The enriched data is then applied to the core of the business, where a strategy jumps to life – research, backtesting, validation, and then real-time trading.

ENRICHED DATA: RICH TRADING OPPORTUNITIES

A complex enrichment process involves cleaning, normalization, integration, and aggregation of market data from multiple markets and geographies as well as a variety of asset classes. This data is integrated with other structured and unstructured sources to create enriched streams. These streams provide a comprehensive and integrated view of the market state and are used for a variety of activities such as signal generation, trade execution, real-time P&L reporting, and real-time risk measurement.

As an example, let's consider how the enriched data for a global equity index arbitrage strategy comes together. Both structured and unstructured data must be enriched and provided to a signal generation process. Market data for the equities comprising the index – along with exchange-traded funds (ETFs), futures, and options tracking the index and index components – are integrated with other structured data, such as macroeconomic data, fundamental data, and commodity spot prices. Reference data maintains the relationships between both like and underlying securities. Using real-time currency, prices are converted to base the currency. Along with voluminous structured data, massive unstructured data sets are typically required to extract relevant information on consumer sentiment and to locate and analyze relevant news and sentiment.

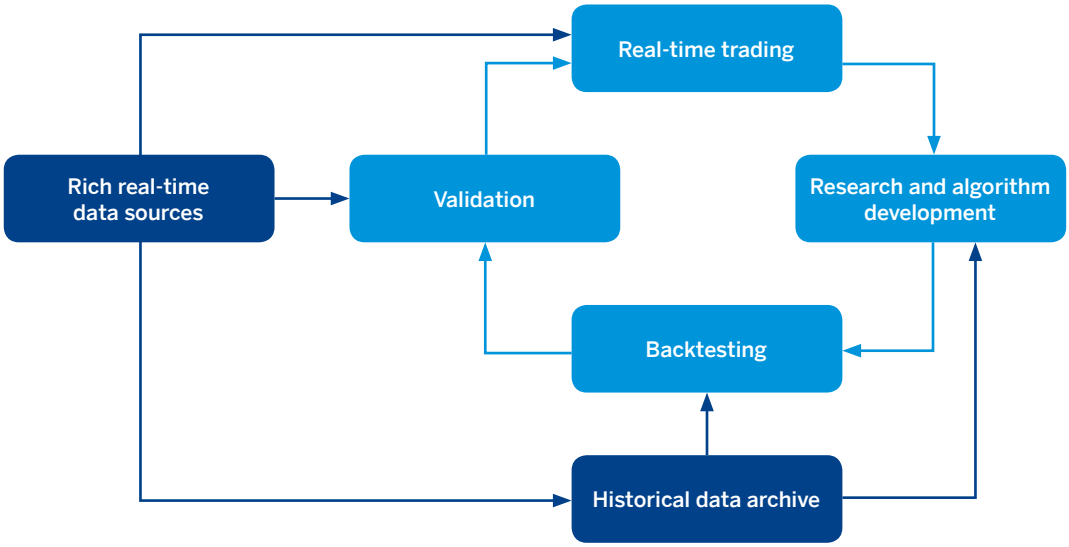
The enrichment process significantly reduces data volume, converting the raw structured and unstructured data into a compact form suitable for signal extraction. The massive volumes of unstructured data are reduced to a small number of relevance scores. Enriched market data includes a rich set of summary statistics, updated in real time. These measures can include price momentum, spreads, trading volume, volatility, and liquidity information.

© 2013 SAP AG or an SAP affiliate company. All rights reserved.

INTEGRATING THE SMART TRADING ASSEMBLY LINE

Smart trading's virtuous circle starts with research to develop analysis methods and algorithms. It then proceeds through rigorous backtesting to a period of simulated real-time trading or validation, and it concludes with live production with the new processes. The same cycle applies to most smart trading processes, including signal generation methods, trading algorithms, and market risk models. This typical workflow repeats continuously, often in cycles overlapping in time. (See Figure 2.)

Figure 2: The Integrated Cycle of Smart Trading



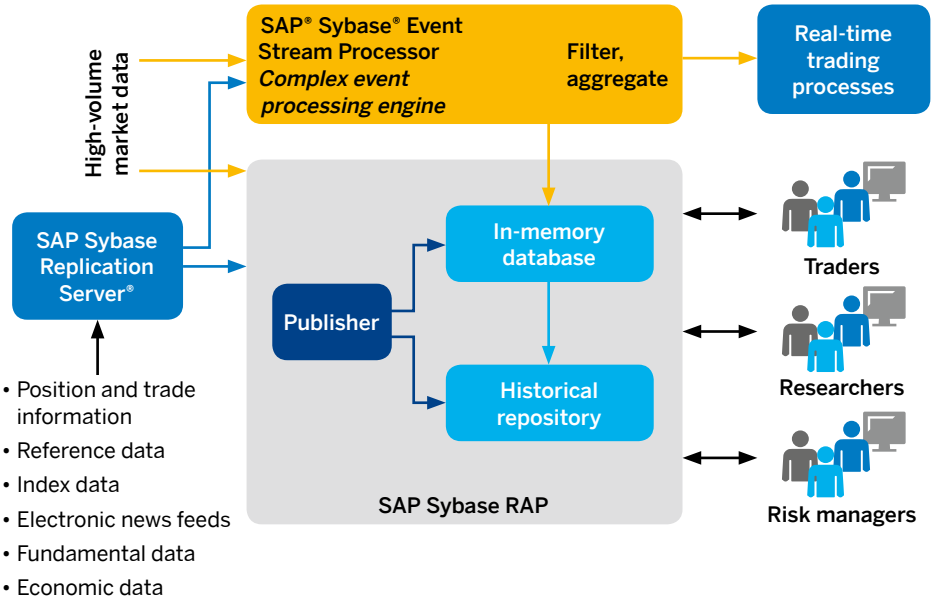
© 2013 SAP AG or an SAP affiliate company. All rights reserved.

An integrated platform manages and analyzes both real-time and historical Big Data that is structured and unstructured, and it quickly and efficiently seizes the best trading opportunities using a unified, holistic view of market state. Complex event processing (CEP) technology for real-time processing, integrated with a column-oriented database for historical data storage, management, and replay, creates this optimum environment. The platform facilitates rapidly creating, testing, and taking sophisticated smart trading processes into production. (See Figure 3.)

Integration with research tools, such as the powerful open source programming language R, greatly extends the analytical capabilities of the column-oriented database and CEP platforms. The hundreds of analytics packages available for the R language can be harnessed to perform research. Sophisticated time series analysis, text analysis, risk modeling, and performance attribution are all found in R packages. In addition to R, researchers commonly use MATLAB and similar tools.

Data set sizes in R and similar tools are limited by physical memory. So the use of radically reduced volumes of highly enriched market data provides researchers with comprehensive market state information without overwhelming the capacity of these commonly used tools.

Figure 3: Integrated Environment for Smart Trading



© 2013 SAP AG or an SAP affiliate company. All rights reserved.

Big Data fuels smart trading, and CEP technology is ideally suited to the **real-time enrichment of data.**

REAL-TIME BIG DATA PROCESSING

Big Data fuels smart trading, and CEP technology is ideally suited to the real-time enrichment of data. The CEP platform provides an environment for rapid development, testing, and, most important, production of enrichment processes. With CEP technology, large volumes of data from multiple sources and of multiple types can be enriched and integrated, and subsequently, information can be extracted. These enriched information streams are fed to the real-time smart trading processes on the same platform.

The CEP platform is used not only to enrich the real-time market data but also to integrate other structured and unstructured data. These data types can range from the fairly static, such as fundamental data and macroeconomic data, to the nearly real time, such as sentiment scores.

With real-time market tick data, the CEP platform performs vital enrichment and integration tasks in milliseconds or less. As examples, prices can be adjusted using the latest FX data for global trading and order book summaries, while real-time index levels can be computed along with a multitude of real-time analytics and risk measures. Real-time filtering, integration, and aggregation within the CEP platform provide streams of highly enriched data to downstream trading processes.

MANAGING BIG HISTORY

Management of Big Data for trading involves storage, retrieval, and playback. Data enrichment is integrated with the playback process. Both structured and unstructured data is queried and replayed in time order. Data enrichment reduces the volume that must be processed with each query.

Column-oriented databases provide a powerful platform for the management of Big Data. Most structured Big Data, including market data, is inherently organized into time series that makes the column-oriented architecture natural and highly efficient. A column-oriented, "share-everything," massively parallel processing architecture database allows multiple queries to be executed simultaneously on archives of both structured and unstructured Big Data. Further, by placing the enrichment processes close to the data archive, queries return smaller volumes of enriched data rather than high-volume raw data.

An effective database for the management of big market tick data must have the flexibility and tools to store and integrate multiple data types. Different types of data, with different frequencies, can be integrated and enriched with online analytical processing (OLAP) tools.

Historical market data queries for backtesting must return time series with continuous values, adjusted for corporate actions. Symbol mapping and rollovers must be applied over time. Finally, financial calendar filters are required to normalize time stamps and filter time series for exchange hours, holidays, half days, and trading halts.

PULLING IT ALL TOGETHER

The intersection of Big Data technology and smart trading processes has opened up exciting new directions for profitable trading. In smart trading, rich data sets generate profitable trading signals that persist in time. The integrated environment of CEP and column-oriented database technology provides an ideal platform for the efficient development and deployment of new smart trading processes. Exploiting one of these powerful platforms stands in contrast to building and maintaining a purpose-built trading platform. The latter is typically constructed over a series of one-off and ad hoc projects. Using the integrated platform technology instead allows people in trading organizations to focus on smarter trading rather than infrastructure construction and maintenance. The integrated environment improves opportunities to trade smarter and more profitably while shrinking lifecycle costs.

LEARN MORE

To find out more about how your organization can effectively manage Big Data, please contact your SAP representative or visit us online at www54.sap.com/industries/banking/solutions/sub-ind/capital-markets.html.

© 2013 SAP AG or an SAP affiliate company. All rights reserved.

www.sap.com/contactsap

CMP22602 (13/01)
© 2013 SAP AG or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries.

Please see http://www.sap.com/corporate-en/legal/copyright/index.epx#trademark for additional trademark information and notices.

The Best-Run Businesses Run SAP™