



# The Data Management Revolution

The Drive for Real-Time Transparency

CONCLUSIONS PAPER

Insights from a roundtable discussion sponsored by  
*Wall Street & Technology* and SAS

Featuring:

Greg MacSweeney, Editorial Director, *Wall Street & Technology*,  
and *InformationWeek* Financial Services

Larry Tabb, founder and Chief Executive Officer, TABB Group

David M. Wallace, Global Financial Services Marketing Manager, SAS

## Table of Contents

<b>Table of Contents</b> .....	<b>ii</b>
<b>Introduction</b> .....	<b>1</b>
The Growing Data Imperative .....	2
<b>Rethinking Data Management</b> .....	<b>3</b>
<b>The Technologies for the Data Management Revolution</b> .....	<b>4</b>
Redefining the Possible with High-Performance Analytics .....	5
<b>Closing Thoughts</b> .....	<b>6</b>
For More Information .....	7

## Introduction

***“Working as we do in an industry focused on the intangible, we quantify our goals, progress and risks with data – in fact, lots of it, surrounding ourselves at times with an overwhelming amount of data about our investments, positions, risks and opportunities. We also interrogate that data: How much can we make? What’s our downside? How much risk are we exposed to? What happens if we’re right? And, more important, what if we’re wrong?”***

***Because our world revolves around data, you would think that our data infrastructures are robust. But unfortunately, they’re not nearly as robust as you’d think, or for that matter, as regulators require.”***

That’s the troubling assessment from Larry Tabb, founder and CEO of TABB Group, a research and strategic advisory firm focused exclusively on capital markets (TABBforum, October 22, 2012). Few firms would disagree, as they grapple with the data challenges associated with new regulatory requirements and big data – as well as the imperative to provide more clarity into their operations to internal and external audiences.

“Going back to 2010 and looking forward to 2013 and 2014 as well, the capital markets’ focus is about regulation,” said Greg MacSweeney, Editorial Director of *Wall Street & Technology*. In a roundtable discussion co-hosted by his organization and SAS, MacSweeney described an environment rife with new Wall Street reforms and consumer protections that profoundly change data, analysis and reporting requirements.

“Most of the regulation is focused on transparency and risk management, trying to get better insight into what firms are doing, how they are doing, who they’re trading with and what the risks may be,” said MacSweeney. “The second primary focus is about [internal] transparency and getting a better grasp on the financials inside of a firm.”

The Dodd-Frank Act, the Consumer Protection Act, Basel III, the Volcker Rule on proprietary trading, regulations governing trading of over-the-counter derivatives, adviser regulation related to hedge funds and private equity, Consumer Financial Protection whistleblower statutes... the effects of these changes are acutely felt and likely to become even more influential in coming years.

“Everybody in this room is having to deal with the ramifications of this,” said Tabb. “You get an upset person who leaves, and then all of a sudden they’re out crying to the SEC [Securities and Exchange Commission] that ‘My company did something bad.’ So the people developing systems and data need to really understand who has access to what, what they’re doing and whether we are doing things legitimately. How do you scan all the information that’s going out? How do we monitor practices of many thousands of people who work for some of the large organizations?”

“Then you’ve got the issue around Basel III [requirements for financial institutions to have a larger capital cushion]. You’re trying to manage capital, understand how to aggregate it, what’s included and not included, and how we risk-weight assets, understand where our capital is coming from, and how to reduce our reliance on overnight funding with its associated penalties. And then you have SEC issues just outside of Dodd-Frank, such as HFT [high-frequency trading] and the proper structure for the equities market.”

## The Growing Data Imperative

The common denominator to all these questions is ratcheted demands for data – exploding data volumes, new data sources, greater problem complexity, mounting performance expectations and shrinking decision-making windows.

“New data feeds and new types of data are coming in – such as mobile and Web – and many managers want to include different types of information and data in their calculations, whether for financial calculations, risk modeling or financial modeling,” said MacSweeney. “They’re sometimes bringing these ideas to the data management technology teams, and saying, ‘Hey, look at this feed we can get. Can we include this in our data management strategy somehow?’”

“The data management environment is becoming increasingly complex, because much of this data isn’t normalized; it’s not in a format that is common across different groups. And this data is coming very fast at a high volume, it’s coming from all different parts of the business, and the managers want it quickly. They want to be able to analyze this data, churn through it and come up with some meaningful information.

“Instant data analysis isn’t just for the trading desk anymore. It’s for risk managers, it’s for portfolio managers, it’s for across the business and C-level executives as well. Enterprise data integration – getting that data from different parts of the business – becomes very important.”

Optimization is no longer a quarterly or monthly reporting cycle; it is an activity of responding quickly to market, capital and risk factor changes as they happen.

“Increasingly we hear talk that folks are trying to get more risk metrics in more real time, released multiple times during the day,” said Tabb, “not just an end-of-day or weekend exercise.”

To further complicate matters:

- **Regulatory compliance is a moving target.** “Before you can adapt to one change in regulation, there’s another new one,” noted another participant. “You’re always chasing a new target. You may have a system that meets Basel 2.5, but does it meet Basel III? I’m sure before we can even touch Basel III, there’s going to be a 3.5.”
- **Outsourced data is costly.** “We’ll buy a lot of the data, but it gets very expensive to try to do any kind of historical analysis and things like that,” said a roundtable participant. “The more historical data you want to buy, the more expensive it is to support your models.”

**“Instant data analysis isn’t just for the trading desk anymore. It’s for risk managers, it’s for portfolio managers, it’s for across the business and C-level executives as well. And enterprise data integration, getting that data from different parts of the business, is very important.”**

**Greg MacSweeney**  
*Editorial Director, Wall Street & Technology*

- **The data flow is a flood.** For example, high-frequency trading generates an estimated 6.5 million messages per second on US exchange products. One roundtable participant reported that his firm has 20 petabytes of data. To put it in perspective, a petabyte is a quadrillion bytes; Google processes about 24 petabytes of data per day. How do you manage all of this real-time data? Should you capture it all in massive databases or look to third parties such as Reuters or Bloomberg to get what you need when you need it?
- **Much of the data is unstructured.** For instance, Twitter users generate millions of tweets, potentially yielding clues to whether you should be buying X or selling Y. However, most firms aren't ready to capitalize on that opportunity. "There's a large amount of data that we need to analyze to find that proverbial needle in a haystack," said a participant. "But for us, the focus is on internal data as opposed to mining social media feeds and things like that." For now, the opportunity sits largely untapped.
- **Internal or external, data quality is suspect.** "We compare a lot of the reference data, the pricing we get, against multiple vendors, and we consistently get errors," said a roundtable participant. "I'm sure every other firm is getting the same thing. So we could report it back, and it does get better, but then new issues arise." Participants agreed that data quality is a perennial issue.
- **Data management is redundant.** "There isn't one kind of consolidation further upstream somewhere that does the comparison for you and sends it down," said a participant. "It's up to each individual group – and of course, it's repeated within the organization many times for every person who gets the data feed. In our organization, we're trying to move it further upstream to clean it up, but there are still issues, not just for our organization but industrywide."

Satisfying all these requirements is a tall order, especially in the face of fast-growing data volumes and processing requirements – *big data*. Everything that was important before is more important now. Everything that was problematic before will just be bigger now.

## Rethinking Data Management

"Traditional data management strategies may not be up to par or able to handle a lot of this new data that's coming in. It requires a new way of thinking and new technologies," said MacSweeney. "Moving from a data management strategy to data analytics to instant analysis is very important given the need for transparency, looking at financial information in real time depending on market events."

**“Even though your risk calculations and GLs meet certain requirements today, you can't know if it will satisfy new regulations and local regulations for new markets. You have to make sure the data structure – your metadata that fits in today – doesn't lose its applicability tomorrow because you didn't have the foresight to make it that granular.”**

Roundtable participant

Banks, capital markets firms and asset managers have relied on traditional transactional and reporting technology since the early mainframe systems were introduced in the 1970s. In the 1980s and 1990s, quantitative measures and calculations extended the analytic framework beyond mere transaction reporting, but the vision of a predictive analytics framework remained just that, a vision. A huge obstacle was the apparent need to combine as much transaction data as possible into one central repository to be able to cross-correlate market, portfolio and event data. Results took days to calculate and were then further obscured by being summarized into dimensional analysis cubes that required a team of specialists to interpret.

The pressure for more immediate answers requires a different kind of technology framework, one that can recalculate exposures based on emerging, dynamic risk factors – and distribute the results in real time. The right technology foundation would deliver:

- An integrated view of data from multiple, disparate systems, with assured data quality and consistency, and without the requirement to create one enormous database or to forklift the data around for analysis.
- Faster processing speeds – by several orders of magnitude – so you get key risk results within the time windows that critical decisions demand.
- Considerably greater precision in extremely complex portfolio valuations, versus overly simplistic approximations.
- Always up-to-date portfolio views of aggregated risk by accommodating incremental arrival times for data relating to portions of the portfolio.
- The right level of detail with granular access control – from executives needing high-level summary views, to analysts needing to drill into the most specific details in response to regulatory inquiries.

**Financial firms need the computing power, algorithms for finding approximate solutions to these computationally difficult problems at the speed of the market, and a management philosophy that disciplines the firm to anticipate changes in the opportunity set when making optimization decisions.**

## The Technologies for the Data Management Revolution

Which data architecture is the right one to deliver on these promises: centralized or distributed, in-house or outsourced? These are persistent questions for an industry in flux.

Centralized or decentralized data repositories? “I’ve seen [the data platform] go from big iron [mainframes] 30 years ago to distributed and back to centralizing it again – and now throw the word ‘virtualization’ on it,” said a participant. “It’s a pendulum that’s swinging back and forth. But I don’t know what’s going to happen with the high-frequency trading systems. We’re just basically waiting to see what’s going to come out. We don’t want to build anything and find out that we’re in the wrong direction.”

In-house or purchased data? “It depends,” said a participant. “In the majority of the cases, we store the historical data for everything, and we load that data into the cache for, say, next week or 10 days or a month – whatever the requirement is – and we use it for calculating risk and management reporting. But if we are doing real-time trading, then we have no choice but to go to Bloomberg or other [purchased] feeds and make decisions based on that.”

Own or outsource the data platform and processes? Should you build, own and manage your data infrastructure in your data center, or just ship your algorithms to a provider, let them back-test and move proven models into production? “It depends on the cost,” said a participant. “If it’s cheaper doing it ourselves, we will take that path. We definitely do the cost analysis to determine which way is better: outsourcing or doing it ourselves, even though we have fewer resources.”

## Redefining the Possible with High-Performance Analytics

“In the past, the challenge was identifying the data, pulling it to one place, getting it standardized and prepared for analysis,” said David Wallace, Industry Marketing Manager for Financial Services at SAS. That part of the equation is now easy to conquer. “A part of our stack of solutions includes a full suite of information management solutions – from federated and enterprise data governance all the way down to data integration and data quality solutions for extremely large data sets.” If firms still have issues with data quality and integration, it’s not that the solution isn’t available, just that they haven’t implemented it yet.

“We are also revamping all of our analytical products and solutions to operate on big data,” said Wallace. This accelerated processing with huge data sets is made possible by four primary technologies:

- With **grid computing**, the processing workload is distributed among different hardware components in a centrally managed grid infrastructure. Multiple applications and users share a grid environment for efficient use of hardware capacity and faster performance. “Since many firms – all of the firms represented here today – have already had enterprise grids for a long time, we have grid-enabled our software solutions,” said Wallace.
- **In-database analytics** moves relevant data management, analytics and reporting tasks to where the data resides for faster results, less data movement and better data governance. “Rather than move the data around to get analyzed by SAS®, we’re moving the SAS analytical routines into the enterprise data warehouse appliances,” said Wallace. In-database analytics reduces the time needed to prepare data, and then build, deploy and update analytical models.
- An all-new **high-performance analytics** initiative moves SAS analytical solutions directly into memory for exponentially faster speed to answer. “Wherever the data lives – be it in an enterprise data warehouse appliance or Hadoop; structured data, semistructured data or unstructured data – we can move the routines and process and analyze the data directly,” said Wallace. In-memory risk analytics reduces the time required to revalue large, complex portfolios from hours to minutes, and from minutes to seconds.
- “In addition to an expanded focus on interacting with Hadoop, we’re applying our in-memory analytical expertise to **real-time data visualization**. We now have solutions where a billion or more rows of data can be pulled directly into memory. It can be interrogated, correlated, graphed and charted by not just data scientists, but by business analysts and business professionals,” explained Wallace.

**“SAS took the old ways of doing things and came up with a more advanced way to both process and calculate the risk vectors and to be able to manipulate them in memory to provide an instantaneous drill-down response.”**

David Wallace  
Industry Marketing Manager for Financial  
Services at SAS

“Business users have a much greater need to interrogate the data directly and see it on screen,” said Wallace. “The insight that business professionals get may then result in models being developed by data scientists and modelers and put into production. But we need to get the data exposed to the business professionals first so they can see if there is a new market or business opportunity.”

Since most firms have complex event processing software in place, generally on the trading floor, SAS also offers an event stream processing solution that connects directly to SAS analytics products, said Wallace. “You could move from end-of-day risk analysis to more of an intraday or on-demand analysis. The event stream processing solution could receive and process your internal and external data and pulse that data directly into an in-memory analytical version of our risk engine. The limit checks that you do on the floor today can still be done by the event stream processing engine, but whenever you want to do an on-demand valuation of your portfolio, you can signal that revaluation to be done directly – it’s a full revaluation – and then you can push the approximations back out to the traders and to the desks.

“With all of the additional regulatory requirements, we’re also seeing quite a bit of interest in using that risk engine to aggregate all the data that comes out of risk systems. Normally that data is coming out at the end of the day on a portfolio or asset class basis. The runs never seem to finish at the same time. When they do finish, the checks of the data file that comes out are semiautomated at best. The data movement always seems to be a challenge, and then when you finish all that work, often the aggregation is still being done in Microsoft Excel spreadsheets.

“We have been working on a capability that will allow that portfolio of valuation data to be pulsed into an in-memory aggregation engine. You calculate risk the same way, but then you bring it in to one big cube that’s in memory. A user with the appropriate security clearance can drill into that data to whatever level of detail desired, whether it’s by desk, by business unit, by portfolio or product – all the way down to the instrument level – right before your eyes.”

## Closing Thoughts

If there’s any silver lining to the financial crisis, it is that it put a spotlight on the need to get rapid insights from big data, which in turn will accelerate much-needed improvements in firms’ information management practices. “The data guys are getting their say at the corner office and getting the budget to drive a lot of these regulatory initiatives,” said Tabb.

The technologies implemented for compliance then drive business improvements, such as:

- Accelerated analysis of portfolio-level market, credit and liquidity risk.
- Up-to-the-minute assessments of risk exposures for large, complex portfolios of financial instruments.

**“SAS has been investing in these kinds of new technologies because in the age of big data, it’s important to be able to analyze the data in streams and still call upon deep analytics to provide real answers to business questions and predict what’s going to happen.”**

David Wallace  
*Industry Marketing Manager for Financial Services at SAS*

**“Companies that are under a good deal of regulatory scrutiny are definitely weaving the budgets for their data governance and data management programs into the overall budgeting process more than before.”**

Roundtable participant



- Rapid analysis of near-real-time/intraday incremental value at risk (VaR), counterparty exposures and liquidity measures coupled with full risk aggregation.
- Risk analysis results provided as in-memory cubes for aggregation, exploration and drilling to the level of business unit, desk, portfolio, instrument or horizon.
- Dynamic, interactive stress testing that risk managers can use to assess the impact of extreme events on portfolio values.

***Under pressure for greater transparency and risk awareness than ever, firms stand to gain a lot from the data management revolution. With high-performance analytics and visualization, you can continually analyze events and update intelligence as new events occur. You can capture opportunities and value that would otherwise be lost because the information came too late or didn't account for late-breaking market conditions. When the next financial crisis looms, the industry should be able to see the icebergs ahead – in time to change course.***

### For More Information

Download the SAS white paper *Evolving from Quantitative Risk Management to a High-Performance Risk Management Analytic Framework: Insights on a New Direction for Risk Management* at: [sas.com/reg/gen/corp/1052188](https://sas.com/reg/gen/corp/1052188)

For more about SAS® High-Performance Risk, visit: [sas.com/highperformancerisk](https://sas.com/highperformancerisk)

### Innovative Strategies for Big Data Analytics

- A flexible enterprise architecture that supports many data types and usage patterns.
- In-memory SAS analytics for enterprise data warehouses and Hadoop.
- Real-time visualization and advanced analytics to accelerate understanding and action.
- A real-time event streaming engine for low-latency, high-throughput needs.
- In-memory risk analytics and aggregation engine for on-demand valuations.
- A common analytical framework across the enterprise.

## About SAS

SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 60,000 sites improve performance and deliver value by making better decisions faster. Since 1976, SAS has been giving customers around the world THE POWER TO KNOW®. For more information on SAS® Business Analytics software and services, visit [sas.com](http://sas.com).



**SAS Institute Inc. World Headquarters +1 919 677 8000**

To contact your local SAS office, please visit: [sas.com/offices](http://sas.com/offices)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2013, SAS Institute Inc. All rights reserved. 106157\_S107569\_0413