



TickSmith

BAD DATA HURTS

May 2019

Why Does Bad Data Hurt?

"Banks have access to less than 10% of their data and use less than 1% in any form of analytics"¹

Financial institutions create and consume enormous amounts of data, about 50-100 petabytes of data per year. Market data vendors alone sold over \$27 billion of feeds and prices to the industry.² Yet for every dollar spent on data, a financial institution will spend four times that amount on administration, storage, retrieval, data quality checks, billing, distribution, and formatting. Imagine the thousands of people doing nothing more than data validation and reconciliation. Imagine further that the amount of data is doubling every three years.

Unfortunately bad data quality is not solely a cost problem but it also affects the ability for firms to make money for their investors and their customers. Critical data is typically scattered across siloed databases; spreadsheets are lost in a numbers graveyard known as a data warehouse. Not helping are the endless number of data consolidation projects that go on to create archaeological layers of data, derived data, fragments, and just plain detritus.

The broken record: "Mapping to mapping to mapping..." it's small wonder then why banks have access to less than 10% of the data they have and use less than 1% of their data in any form of analytics.



Confirmation Bias

Even when data is available, other human factors such as echo chambers conspire to prevent an optimal decision for the firm. We see this in news and journalism of course. People seek out those news sources that agree with their existing philosophy, preventing wider consideration of dialogue with different views. The same thing happens in corporations where individual teams may focus exclusively on their tasks and not the bigger picture. Becoming a master of enterprise data is not only about tools and governance, but also about culture.

Consider Goldman Sachs in the months leading-up to the 2007 Credit Crisis. The cross firm risk committee saw mark-to-market risk with AIG so there was a challenge from fixed income, which led to the purchasing of credit default swaps.

If this didn't save the franchise, it was close. The key was a culture that asked and encouraged people to connect the dots. Challenge and dissent were allowed when talking to someone from another department.



Data Strategy Still Under-utilized

Digital transformation does not mean simply automating manual tasks. It means a dramatic rethink of how to play the game. Think of, MoneyBall, Michael Lewis' story of a professional sports team that transformed decision making from folklore and habit ("he doesn't have a baseball body") toward a data-driven approach using the advanced analytics called SaberMetrics.



Effective data curators know that data is dynamic and contextual. Historical data (e.g. backward looking data) needs a golden source or a single version of the truth. Firms require some consistency in order to make informed decisions about the past. In contrast, forward-looking data is about approximations or ranges of outcomes and is used by risk managers and portfolio managers to be smarter about the future. Data curation is about understanding the context, source, and uses of data so we don't mix up past and future views.

For example, hedge fund investors want to know internal rates of return and net asset values that are a point-in-time estimate. They also understand the probabilities and ranges of outcomes about risk and liquidity. There is also a continuous democratisation of data, which includes active users. Consider the "Wikipedia Model," which enables users to contribute, alter, and edit data. At the core, there is still an editorial override and final decision, but there is still an open opportunity for participation, which all great data-driven firms are striving for.

Unlock More Value from Existing Data Assets

TickSmith allows financial institutions to focus on higher order value-creation work. They provide a platform that institutions can install in-house or in the cloud that serves as the main foundation for distribution, aggregation, and advanced analytics. Not only does this remove the physical presence of data storage, but more importantly, also creates a data lake where there is no need to worry about formatting, linkage, access, analytic block, and other challenges.

Using technology to facilitate collaboration, dialogue and insight is at the forefront of every firm's mission. Technologists would describe this as taking data and quickly allowing it to be stored and retrieved in a highly flexible data management platform.



The CME Group, one of the largest exchange groups in the world, adopted the TickSmith technology as a solution to provide faster distribution to clients as well as allowing storage and advanced analytics.



The CME Group Selects TickSmith

The CME Group enlisted TickSmith to power a big data platform for managing their immense quantities of data for monetization. At the time, they did not have a standard system for handling and processing data and as a result, their data was dispersed over multiple departments. It seemed an impossible task to normalize, centralize, and extract the vast amount of historical data and make it easily accessible online. The CME financial market company is distinct because they are also a mercantile exchange. Therefore, the requirements included normalization of both traditional and alternative stock data.

Until 2011, the CME was using hard disk drives to ship this data to their clients, who would then download the data and return the disk within a month. With historical market data going back to the 1970s and beyond, the CME Group was in need of a strong data management solution in order to migrate into the digital world.

TickSmith was an easy choice, being experts in implementing big data technology for capital markets. "The CME Group has always offered customers a robust set of historical data, and working with TickSmith allows us to deliver it efficiently and cost-effectively," says Craig Mohan, the CME Group Managing Director of Market Technology and Data Services.

"The CME has always offered customers a robust set of historical data, and working with TickSmith allows us to deliver it efficiently and cost-effectively"

"We don't need to touch the actual data all that much, we mostly touch everything around the data," says Nicolas Doyen, Product Owner responsible for the Data Monetization Platform at TickSmith.

"Clients can enhance their data monetization processes, and they can choose specific pieces to offer on the system. Our value prop is that we enable clients to offer a full range of products, our technology scales infinitely, and we are constantly building new features to improve it. Clients are able to build a platform that fits their needs, exactly as they need it."

The End Result: The CME DataMine

The CME DataMine webstore is the official source of the most comprehensive price information available for the CME Group markets, with more than 750 terabytes of historical market data. It provides a broad array of data types including Market Depth, End-of-Day and Block Trades among others, which can help customers discover insights to capture market opportunity. The CME Group has also integrated third party datasets that provide data around the instruments traded on their exchange, such as crop health analytics and oil tanker fill estimates. A new use case for data monetization came along the way for the CME CF Bitcoin Reference Rate (BRR) and the CME CF Bitcoin Real-Time Index (BRTI), a standardized reference rate and spot price index for online bitcoin trading.



One of the main challenges of this project was the general management of big data distribution— a huge pain point for any company looking to monetize massive amounts of data. This required a revolutionary solution to the traditional SFTP (Secure File Transfer Protocol) model. Traditional SFTP approaches would require the Data Distributor to set up client-specific SFTP servers and continuously copy data into those servers.

“It makes it that your big data problem becomes a very, very small data problem for the end consumer. For instance, some financial institutions don’t have big data technology expertise, so dealing with a 40 terabyte historical dataset is hard for them,” says Doyen. “So if they’re able to get smaller slices of data, they can more easily leverage them. They can actually load these files on their computers, which they couldn’t do before.”

The CME Group jumped on the big data bandwagon pretty early compared to other financial institutions and has since reaped major benefits. DataMine is essentially an online web store or data catalog, powered by TickSmith’s Data Monetization Platform, which customers can subscribe to and access data instantaneously via web interfaces and APIs. “Our clients now have a simple way to access the data they need,” says Mohan.

TickSmith on AWS

The CME Group is among the many financial institutions choosing Amazon Web Services (AWS) for their cloud journey. AWS provides the Financial Services industry with the world's most comprehensive and broadly adopted cloud platform to power infrastructure, strengthen security postures, become more agile, and lower costs. With AWS, customers own and control access to their own data— much of which is highly sensitive – in an industry that is closely regulated.

"The sophistication of its cloud offering, specifically regarding the range of services available and the flexibility of these services, makes AWS easily the best choice."

By migrating to AWS, the CME achieved a significant level of privacy assurance at a much lower cost (in both effort and dollars invested).

TickSmith is an Advanced Technology Partner in the AWS Partner Network and has achieved Financial Services Competency. This means AWS recognizes TickSmith for its industry expertise, certified staff, and well-architected solution. As a result, the platform's cloud infrastructure follows an accredited and secure workflow from ingestion to delivery, with entitlement and monitoring capabilities throughout the process.

"The sophistication of its cloud offering, specifically regarding the range of services available and the flexibility of these services, makes AWS easily the best choice," adds Doyen. "For actual clients, there's proven success with AWS regarding security. In terms of capabilities alone in enabling us to accomplish our goals, you couldn't have a better cloud. They have everything we need."



Final Thoughts

"The curation of data requires 80% of expensive and scarce data scientists' time to clear and format data³"

Consider that today the curation of data (an advanced form of information management) requires 80% of expensive and scarce data scientists' time to clear and format data rather than working on higher order tasks such as predictive modelling and insights on customers' value or risk management.

Information-based management companies understand data and analytics. They are focused on scaling the franchise. They use AI and machine learning for streamlined data quality efforts. They form a culture that seeks to minimize time spent on the plumbing of data using capabilities such as TickSmith's technology in order to focus on higher pursuits. In sum, you want to allow easy componentized API access so the real value of insight and analytics can begin. The endlessly-increasing amount of data generated by business processes serves as a reminder that financial institutions need to prioritize data as part of the core strategy at the big table of executive management.

They can rely on TickSmith's platforms to deliver high-quality data and address time-consuming pains related to financial data management at scale.





References:

[1] “A Smarter Way to Jump into Data Lakes.” *McKinsey & Company*, 2017, www.mckinsey.com/business-functions/digital-mckinsey/our-insights/a-smarter-way-to-jump-into-data-lakes.

[2] MarketsandMarkets. *IoT Middleware Market by Platform Type (Device, Application, and Connectivity Management), Organization Size (Large Enterprises and SMES), Vertical (Manufacturing, Government & Defense, Automotive & Transportation), and Region - Global Forecast to 2023*. MarketsandMarkets, OAD, pp. 1–118.

[3] CrowdFlower. *Data Science Report*. CrowdFlower, 2016, pp. 1–10.

Author Contact Details

Mark Rodrigues joined the TickSmith board in August 2017. Rodrigues further develops TickSmith’s overall strategy , customer service, and global expansion.

Mark Rodrigues
TickSmith Board Member
mark.rodrigues@ticksmith.com