The Search for Execution Quality Part Two: Challenges to Implementation

Shyamkiran Rao
SVP Quantitative Strategies
Jefferies LLC

Regulatory change, competition, market fragmentation, and other factors are driving buy-side and sell-side firms to implement analytics systems to help them improve trading quality and reduce the cost of compliance. This two-part white paper explores best practices for approaching and planning such implementations.

Introduction

In part one of this paper, we made the case for optimizing trade execution quality, describing some of the trade execution quality challenges that firms face and articulating the need for a comprehensive firm-wide analytics approach for understanding and improving trade execution. In part two, we describe some of the challenges that firms face as they seek to implement trade analysis solutions and make several key recommendations to help the people responsible for implementing and operating analytics systems make good decisions.

Implementation Challenges

The most significant challenge to implementing an effective trade analytics system is typically the poor quality of the data itself. The effort to produce good, clean data can easily account for 30 to 40 percent of the time and investment required to implement the system.

Essential tasks include verifying the accuracy of incoming data, distinguishing historical data from real time data, parsing and collecting views of data sources, and maintaining metadata dictionaries. Complicating factors include corporate actions and the need to make sense of symbolical systems that vary between countries and exchanges.

Achieving the flexibility necessary to adapt to the ever-expanding volume of data also poses a significant challenge. The amount of data that requires analyzing increases constantly – and the act of analyzing data itself creates new source data. Most firms will handle a range of 2.5 to 6 billion ticks per day in equities data alone. U.S. markets alone trade roughly 8.5 billion shares every day (average for May 2019). Over the last decade, the volume of data that should be analyzed has grown eightfold to tenfold – some firms are analyzing more than a petabyte (a million gigabytes) of data. And storing, managing, and analyzing all this data can be expensive.

Finally, a third challenge is how to respond to the changing market structure. Regulations such as the European Union's Markets in Financial Instruments Directive II (MiFID II) require collecting enormous amounts of data on orders, executions, and venues. Under MiFID II, firms need to know, for example, why a broker chose a specific venue.

Artificial Intelligence and the Investment Process

Artificial intelligence (AI) has roots in statistical mathematics going back approximately 150 years. Likewise, machine learning (ML), which is also closely related to statistics, arose from theoretical developments in computer science. The diagram below from the SAS Institute illustrates the relationship between Statistics, AI, ML, data mining, and pattern recognition.

ML is highly automated. In simplified terms, a programmer inputs a small set of assumptions and the system uses various algorithms to learn from the data. The system then generates extended data sets based on the output of those algorithms. AI goes further and makes inferences based on the output of ML algorithms.

Analyzing trading using ML and AI involves four important steps:

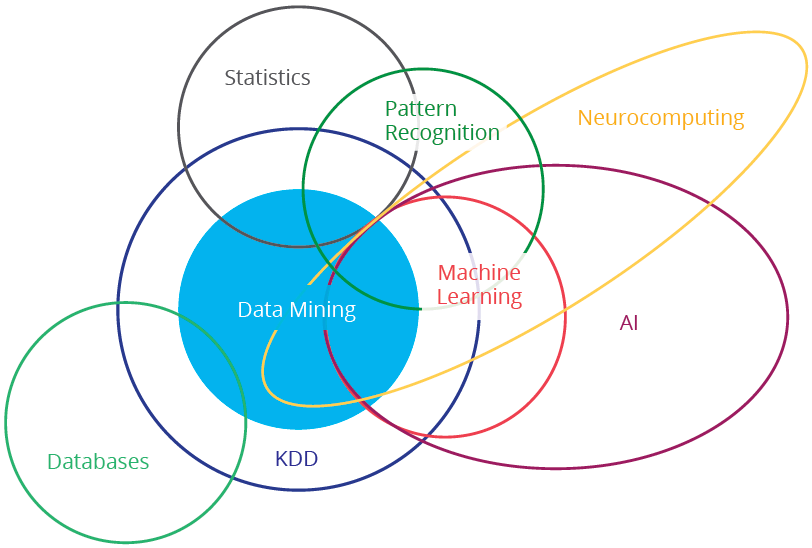
Identify the goal of the exercise. Why am I doing this? What do I plan to gain from this exercise? What is the expected end result?

Determine what data resources are required.

Mine the data using econometric models. Build neural nets or other ML/AI algorithms to extract even more information from the data.

Develop an implementation strategy.

Despite a popular fear that robots using AI could dominate and subjugate humans, AI in reality has limited capacity to master the most complex of human tasks. It cannot replicate the intuitive decision-making capabilities of people in every situation, including the fast-changing environment of electronic trading. As financial professionals well know, live trading doesn't always follow historical patterns, and so-called "black swan" events happen with great regularity.



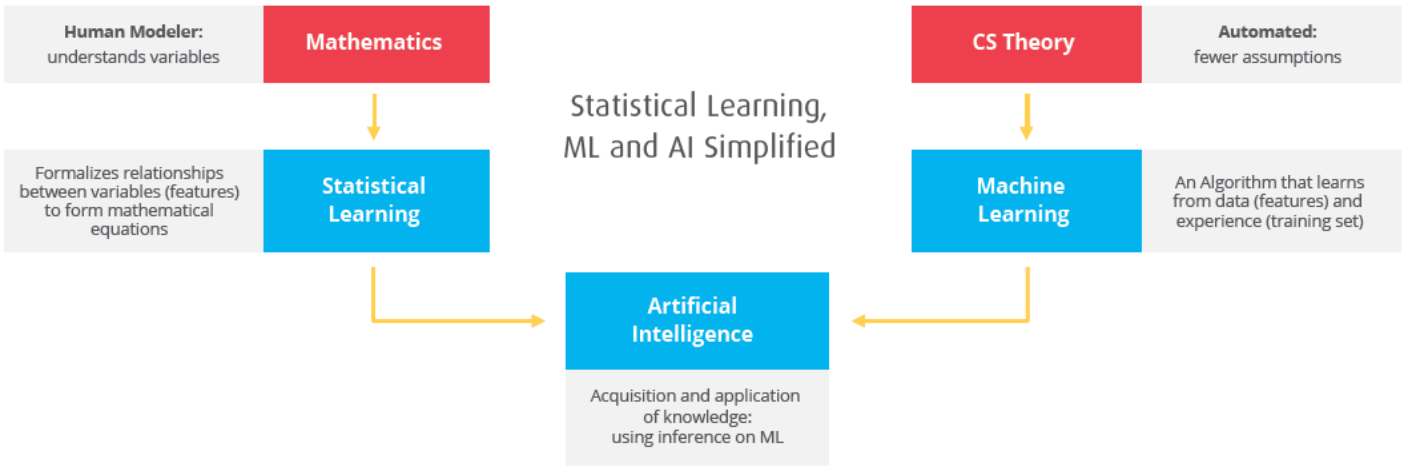
Source: The SAS institute, KDD: Knowledge discovery in DBs

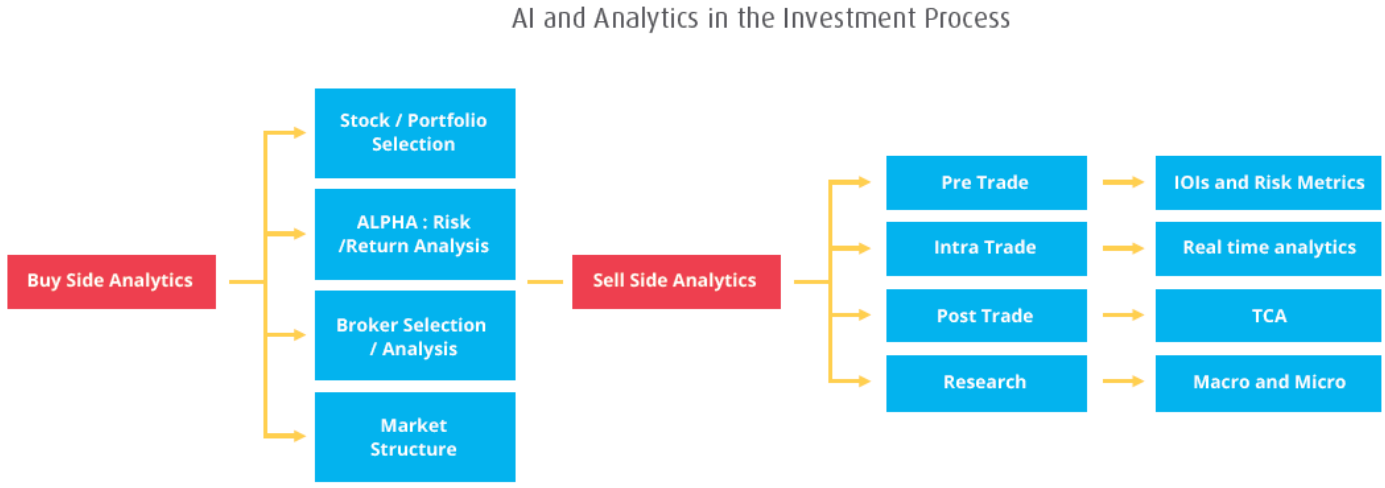
A good way to think of AI is as an intelligent intern. For example, use AI tools to review execution data going back several years. The AI algorithms do not need to understand the market in order to look for correlations that help identify anomalous behavior. Use this way, AI tools can highlight problematic behavior quickly and give the trading desks new insights they can then use to make informed execution decisions going forward.

A better way to think of AI, especially in the trading world, is as an intelligent intern. For example, imagine a firm asks an intern – a graduate student in statistics, for example – to review its execution data going back several years. The intern, who knows little about the market, looks for correlations to identify anomalous behavior. Artificial intelligence tools used in this way can highlight problematic behavior quickly and give the trading desks new insights they can use to advise customers and make informed execution decisions.

Another myth around AI, as well as "Big Data" projects that involve AI, is that they require expensive outlays and the work of hundreds of data scientists, quants, and programmers. But in reality, a small, nimble team of as few as twenty-five

people can create, implement, and manage world-class AI tools for use by quants, including tools that support entire order flow from origination to post-trade analysis.





To achieve maximum execution quality, firms should use AI and analytics throughout the entire investment process, and not just apply these tools to post-trade analysis.

Buy-side firms can use AI and analytics to calculate alpha and determine the risk and return for each available instrument. On the buy side, risk/return analytics is particularly helpful in portfolio selection. With these tools, firms can answer such questions as: Should we “layer the book” or “sweep available liquidity” to achieve the desired price and size? Firms can test broker performance by, for example, systematically sending orders to multiple brokers to measure execution quality.

Sell-side firms can use AI to conduct pre-trade analysis to estimate the market impact of the portfolio selected by the buy side and provide cost estimates for planned trades. Post-trade or TCA-execution-quality measurement relies on processing and analyzing enormous amounts of data in real time, and the system must be flexible enough to allow frequent and fast customization.

Real-time analytics tools such as complex-event processing (CEP) and data visualization can generate signals to feed AI and trading algorithms and to develop bots that can respond to intra-trade queries from clients.

A good visual analytics platform can work nicely as a rapid application development tool in this context. Direct access to such tools can help quants use model behavior to make hold/sell decisions – ideally, they should be able to modify them

on demand without requiring help from an IT department. Clients appreciate receiving highly detailed TCA reports, and these can be a useful tool for client retention.

Avoid Infrastructure Traps

Firms should avoid getting bogged down in the hardware and software infrastructures behind these tools. Three core technology areas can present potential infrastructure traps. The first is data management. Gaining a precise view of markets requires gathering historical content as well as streaming real-time sources, which help establish benchmarks for use in pre-trade, intraday, or post- trade performance analysis.

The data must first be complete, precise, and accurate, and then it must be consolidated, coalesced, and time-aligned.

Time alignment is an especially critical element of data management because of the fragmented nature of equities markets and the number of different liquidity providers. For example, accurately determining the liquidity that's available at any given time – essentially, the best price – is difficult. Not every region has a mandated National Best Bid and Offer (NBBO) system as does the United States. It is not available at all in European or Asian markets, but the concept is valuable enough that smart firms will create their own using their own analytical tools and the right selection of data resources.

Analyzing trade executions, is necessarily complex. Statistical models can compute benchmarks and make comparisons between the firm's orders and those benchmarks. They can help traders conduct market impact analysis and assess implementation shortfalls or slippage against known benchmarks.

How will users consume all the information the analytical models produce? Overwhelming key decision makers with information is a real risk. Visualization, the second core technology area, is at the "top of the stack" – it enables firms to transform the mass of data including benchmark prices and performance metrics into something clear, understandable, and actionable. Visualization makes it easy to identify outliers, anomalies, trends, and clusters – both from real-time streaming data from trading systems and the market as well as from historical data.

The visual analytics layer, however, can only be as good as the data itself and the infrastructure it is built on. A layered approach is best: stream processing (or event processing) for live data, in-memory analytics for today's data and this week's data, and long term (slower) storage for the full history, with aggregations and calculations from history stored in memory. This approach increases the responsiveness of the visual layer. Traders don't want to click a button and wait for half an hour – they need answers immediately, while that information can still make difference.

The third core technology infrastructure is a tick database. This fast data store works like an analytical engine for data enrichment, aggregation, and calculation. Projects that make "Big Data" tools like Hadoop, NewSQL, or NoSQL central to their analytics infrastructures will struggle. These systems may enable an IT team to get an environment up and running quickly, but someone still must write all the required analytical functions, which can be time consuming, especially when nanosecond-level time series data is involved. Tick databases provide a wealth of built-in functionality that makes implementing analytical functions for electronic trading data easy, including such functions as stitching two time series together or recalculating an order book.

Recommendations

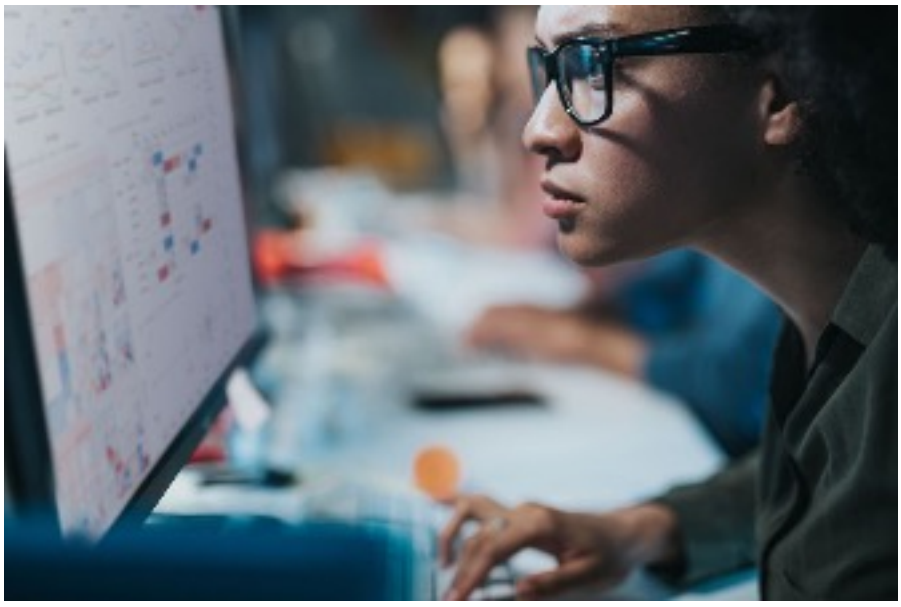
Responsibility for the design of trading analytics systems must be driven by the people working on the trading desks. With proper access to the right tools, they can optimize trading processing on an ongoing basis by monitoring activity, investigating anomalies, back-testing new approaches, and implementing successfully tested approaches as part of their day-to-day operations.

Firms that have approached the development of an analytics infrastructure as an IT project, with software engineers or data science people in charge, have tended to fall behind in terms of how useful their systems are in improving execution quality

Keep in mind that, the most successful trading analytics projects are generally up and running quickly – usually in weeks instead of months.

Consider these questions when developing your trading analytics implementation plan:

- Where will you get the data?
- How will you clean that data and ensure that it's accurate and correctly time aligned?
- Will your existing order-management system provide the time-stamp precision required?
- How will you give your quants and traders the tools they need to build their own analytical interfaces so they don't have to wait for an IT development team?
- How will you ensure that your system is truly responsive to the needs of your trading desks?
- How much capital and human resources will it take to build all this, as well as handle the daily maintenance and administration of the system?
- If you choose open-source technologies for key components, are you being realistic about how mature the software is and what kind of hardware will be required to make it work in your particular use cases?
- Will your quants be able to back test using huge data sets without diminishing the performance of your live trade monitoring systems?
- Will someone new be able to learn quickly and easily how everything works, or will it look like an impenetrable mess of spaghetti in terms of data pipelines, analytical queries, data source subscriptions, and user interfaces?



Think real time. The databases at the heart of their trade analytics system, their event processing engines, and their visualization systems must all be capable of handling true real-time streaming data. All data used in intraday analysis should be in real-time systems.

Make your build-versus-buy decisions based on your answers to the above questions while keeping a close eye on how those decisions affect your schedule.

Because getting clean data is crucial, question the quality of the data you plan to use, obtaining sufficient sample data from proposed sources and testing it thoroughly before choosing what to buy or whether to build. In addition, they should determine where AI will give them the biggest advantage – it won't replace traders, but it can give them tools that will help them make smarter and better-informed decisions.

And, finally, think real time. The databases at the heart of the trade analytics system, the stream processing (event processing) engines, and the visualization systems must all be capable of handling true real-time streaming data. Storing data for longer than a few days in a standard row-oriented database is fine as a way to keep costs down, but all data used in intraday analysis should be in a real-time system.

Implementing effective solutions to help optimize execution quality can be challenging. But the benefits are significant: accessible, visible, actionable, real-time information that traders can use to improve the overall quality of their trades and return greater value to their clients.

Lead Author

Shyamkiran Rao
SVP Quantitative Strategies
Jefferies LLC

Mr. Rao is responsible for driving the business and technological strategies for Equities Analytics and TCA at Jefferies. He has extensive experience in implementing platforms and trading analytic solutions at Jefferies Equities Quantitative Strategies, using One Market Data and most recently with Altair Panopticon.

Prior to joining Jefferies, Shyam was a VP, Architect and developer of the Trading Automation Systems at Goldman Sachs & Co. Shyam has over 20 years of experience at implementing and managing projects for Equities, FX and Fixed Income trading.

Earlier in is his career he was a Technology solutions lead of various products at SunGard Trading Systems, Brass (Now FIS) and a Software Engineer at C-DAC India, building application software and parallel computing solutions.

Mr. Rao was assisted by Louis Lovas and Peter Simpson of One Market Data in the development of this white paper.

Learn more at altair.com/panopticon