WEBINAR Q&A:

Where would you start if you were a practitioner?

**Kumar:** It all goes back to the Tug of War slide we started with. Most firms are thinking about establishing a data analytics architecture in-house because they have many outcomes they want to create. New strategies, finding new alpha, retaining customers, new customer acquisition and so on. “What can I combine my proprietary data sets with that is available in the market?” Pick the 4 or 5 things that are the most valuable to you to learn.

Faced with the reality of building a data lake, or centralized data access service, quickly becomes an obstacle. However, access doesn’t need to be a blocking factor if a virtualization approach is in place. You can start with a virtualization approach and swap out for a physical data lake later if that is what you prefer. This will allow you to realize immediate benefits while you build to the longer-term ROI.

**Mike:** In the last year there was a study by an EDM Council member firm that focused on the opportunity to use data science to drive alpha. 1/5th of the global wealth in the marketplace, trillions of dollars, was measured across 15 to 20 different firms. There had been a thinking in place to “let the other company make the mistake” of doing the early investing in data science. Within a year, or year and half, there had started to be a very measurable ROI in generating alpha. It wasn’t linear but more of an exponential type of curve. This sent the message to everyone in the market that if you are waiting for the other guy to figure it out then you are falling behind. The leaders are already starting to see an exponential return on generating alpha through well-managed data science programs. Part of that is taking the hard lessons early on, but continuing that investment and accelerating the journey. This is
what you are talking about when you refer to avoiding landmines and I think your advice will prove very useful.

You mentioned a process that went from 45 minutes to 35 seconds. What created this gain?

Kumar: The use case was about an algorithm model that uses 50-60 factors (building blocks). The performance was impacted by fetching the data from a source and then converting it into a usable format. Whether the data is coming through an API, a file or other source… there is a level of transformation that the data needs to go through. These transformations increase the number of hops, with increases in latency at every step, before you can perform operations. Through data virtualization, with QTPy in our case, we are able to skip those steps in data transformation. Instead, we make the data readily available as python objects so there are no more transformations needed. You simply go to work using your favorite programming language.

Mike: That makes a lot of sense. I’ll share another industry study from CIO Magazine that analyzed the root drivers of successful analytics programs vs. those that may have failed. The number one cause of failure 70% of the time was data supply chain issues (transformations, too many hops, lack of trusted data, bias, etc.). I think you are hitting on the top issue in that if you have good data supply chain and architecture you can really make a difference.

Will data scientists and quants see the use of virtualization as a loss of flexibility in accessing data themselves?

Mark: We have come across this topic a number of times. There is some learning curve when you make the shift because people are used to writing everything from scratch… from source through model. Once they get a taste of what it means to turn the re-usable parts of their models into objects to be called upon time and time again, they find that the coding effort continues to shrink over time. They are able to come up with new ideas, model them and get to the results without having to write nearly as much code as they used to. They also get the benefit of the speed and performance that comes with calling an object rather than source data which requiring transformation. So yes, there is a learning curve and work to be done… but luckily it isn’t a big leap. It is just a tweak as you think about building from blocks rather than from scratch.

Kumar: To extend that Mark, we have noticed that quants having to learn something like SQL or ETL techniques in order to load files, etc. In this case, they have the same flexibility as before without having to know all of these techniques because the data is all exposed through a series of APIs. What they need to come up to speed on is learning the new 15 to 20 APIs. Being such smart people, quants learn these new APIs within a couple of weeks. Once they are up to speed, they are ready to incorporate new data sets without having to worry about where they reside.

How steep is the learning curve for our teams if we want to go down this path?

Mark: We touched on this a bit in my last response. To take it further, I think this does a lot to bring teams together. In some of our customer conversations I hear about processes in which a quant team develops a model and then pass it to a data engineering group for implementation. They will then go
find the data, pull it together, serve it up and then implement, test and deploy the model. Here, with the approach you heard Kumar describe… teams are brought together in a more dynamic way so they can partner on bringing these models forward quicker. If the data engineering folks are working to make sure all of the objects are available, the analytics team can call them in real time. With a product management style approach, both groups can work together to promote newly development models from development, through testing and ultimately into production. This is a far more efficient process for everyone.

I also think there is a benefit (one that Kumar has been working on internally as a thought leadership piece) related to retaining talent. These teams want to work with elegant tech, they want to get to answers, they want to solve problems and they want to minimize busy work. So, we believe there is a real benefit to putting this kind of technology in place not only to receive the business outcomes, but also to retain the teams. Retain the headcount and the leadership that you have developed on your quant teams over the long-haul because they are able to work on the things, they find personally valuable… not busy work.

What steps would you take to get out of the minefield?

Kumar: I hope the presentation has been useful in getting you to think differently. We work with firms that have tremendously talented quant teams who are building these algorithms and they want to show business value add. Soon they realize the concepts of auditability, and other issues, cause them to pull back. In some cases, even having to shut things down and restart from scratch. Before you embark, keep the core concepts we laid out, even the DCAM component 8 perspective, in mind. Remember the 6 operational aspects you need to think through and ask yourself, “have I satisfied this? Have I considered that?” While you may not be able to avoid all the landmines, but we certainly believe you can steer clear of many of them by keeping these ideas in mind. If you were to work with us directly, we are certain you would avoid them altogether.

What are the lessons learned and 2-3 key takeaways for the EDMWebinar participants?

Mark: It is really to take a pause. You heard some new things today and hopefully some of these concepts ring true for you. Maybe because you are living them now or because you have been through some of them in the past. The poll questions were interesting in that 27% of you have some elements of a product management infrastructure. More of you have a smaller number of the elements so you are somewhere in the middle of the journey. I think it is worth taking a pause to look around and see what that allocation of talent is doing. If you find that you are more in the group that spend 80% of their time engineering the data vs. model development… it is worth looking at those subsections of the landmines described to determine if you are over-exposed to any of them. It is hard sometimes to have the courage to pull back and slow the operation down. But, we need to remember that we are talking about some really high cost and sensitive assets within the company in the form of these quant teams. Their time is money. Time spent delivering outcomes is far better than time spent splicing wires. So, hit the pause button, take a look at those landmines and see if you are spending too much time on any of them.
Steve: I just wanted to add that, just as EDM Council has done for our network in financial services and beyond through providing a network in which people can share ideas, Kumar has developed a similar culture at our firm. So, I ask people who may have questions, or may be struggling with some of the concepts we covered today, to reach out so we can set up a strategy session to work through those areas with you.

Kumar: My final thought is to thank you all so much. I know it takes time and effort to participate in webinars like this and look at new things. You wouldn’t make the time unless the subject was near and dear to your hearts. Sincere thanks to you all and to EDM Council for bringing us all together. If at the end of this session we have served as a catalyst for some new ideas about how you can approach alternate ways to accelerate your operation to produce ROI sooner than later… we would consider this session to be a success.