Quantifying Intuition

MAPPING THE DATA SCIENCE LANDSCAPE
IN THE HEDGE FUND INDUSTRY

Jefferies

Quantifying Intuition

MAPPING THE DATA SCIENCE LANDSCAPE IN THE HEDGE FUND INDUSTRY

Data Science in the Hedge Fund Industry

Machine learning. Big data. Artificial intelligence. Neural networks. The data science tide that has swept across many industries is increasingly shaping the way hedge funds invest, execute, and run their businesses. Watershed moments seem obvious in hindsight, but we rarely understand the significance of all the small changes that precede a considerable shift until later.

2016 seems to have marked a tipping point for the hedge fund industry's mainstream embrace of data science. A number of firms who would not have cited quant skills or technology as an historic core competency began aggressively exploring and leveraging data science in their businesses, and we saw 'quantamental' become an increasingly popular way to describe firms incorporating these solutions.

Data science, the umbrella term for the ever-expanding innovations that identify, gather, aggregate and digest enormous amounts of discrete data points, has permeated the hedge fund industry. Regulators use it to identify bad actors and market irregularities, as well as gather and aggregate industry information. Investors use it to conduct ongoing operational due diligence on hedge fund managers, or build internal dashboards to more holistically understand their cross-portfolio risk profile. Counterparties, exchanges and clearinghouses are using data science to develop new business lines, selling real time data to clients. And hedge fund managers leverage different solutions to improve their investment approaches, streamline execution and run more efficient and effective businesses. Even the CFA Institute is close to requiring CFA hopefuls to answer questions on artificial intelligence, big data and robo-advice in their notoriously rigorous exams.

> **Quantifying Intuition: Mapping the Data Science Landscape in the Hedge Fund Industry** argues that we have reached a tipping point - data science is now, and will increasingly be, of strategic importance to *all* hedge funds in the coming years. This mainstreaming of data science solutions means managers, investors and counterparties should be informed about and conversant in the opportunities and challenges offered by these technologies.

This is the first in a series of publications focused specifically on data science at hedge funds. In this framing piece, we focus on the *practical implementation* of data science at investment managers. We have also hosted four Artificial Intelligence Summits across the globe in the last 9 months, and co-sponsored the Learn 2Quant Conference to help discretionary investors better understand this fast-moving space.

We hope this piece helps explain how investment managers are putting data science theory into practice, and contributes to the growing body of knowledge about the practical implications of rapidly emerging innovations. We look forward to partnering with you as you explore the way forward with data science in your organizations, and are happy to answer any questions you may have.

CONTACTS

Shannon Murphy
+1 (212) 336-1139
shannon.murphy@jefferies.com

Dan Furstenberg
+1 (212) 323-3336
dfurstenberg@jefferies.com

Richard Carbone
+1 (212) 336-7472
rcarbone1@jefferies.com

Jefferies

EXECUTIVE SUMMARY

Hedge funds have always differentiated and succeeded by finding an 'edge' – leveraging information asymmetries via new sources of information, or new ways of understanding pre-existing information. We moved from better analysts to channel checks to expert networks and beyond. Data science is just another evolution - there is more data available now than at any point in history. How firms effectively leverage this explosion of information is now a strategic business issue that will shape long term success and sustainability.

Some firms have successfully leveraged data science for more than thirty years; their efforts are not our focus here. Nor does this piece exhaustively catalogue current available data science solutions. Rather, we explore how hedge funds have more recently employed and capitalized on the rapidly growing data science industry. We hope to offer practical information to help inform principals, CIOs and other decision makers to better understand emerging industry trends and how peers may be embedding these innovations strategically and structurally at their organizations.

There are five core themes of this paper:

1. **Data science is an exciting, innovative and critical business strategy issue – but it also remains a completely amorphous one.** This is not limited to hedge funds. As a May 26th *Wall Street Journal* article reported, CEOs across all industries are feverishly working to understand the impact of disruptive technologies on their businesses.[1] This is a tough balancing act – these technologies are increasingly mission critical, but the landscape changes so rapidly, leaders may have a hard time understanding the landscape, and effectively implementing and adapting over time.

2. **Hedge funds are in different stages of embedding data science across their organizations – but funds of *all* shapes, sizes and strategies are increasingly dedicating headcount and resources to the space.** Jefferies estimates more than 20% of the current Billion Dollar Club has someone who spends at least half their time focused on the issue, or a *dedicated* employee managing data science efforts across the firm. For firms over $8 billion in AuM, it's more like 40%. Anecdotal information indicates considerable appetite to dedicate budget to these solutions in the coming years. And a small number of managers report launching vehicles specifically dedicated to investing in these disruptive technologies as well.

3. **"Excel is not a strategy." PMs, analysts and traders want to understand possible solutions, ranging from unstructured big data feeds to machine learning tools or technology consulting.** Even firms that may not have a long history of quantitative experience are elevating this issue to understand what *potential* exists. They explore using data science solutions as a complement or overlay to longstanding investment processes, better understanding risk profiles and unintended bets, or improving execution or business processes.

4. **Successfully *implementing* data science across hedge funds can be tricky, and cultural challenges may create unintended inefficiencies.** It's not enough to say: "Get me a quant, any quant!" Hedge funds are idiosyncratic organizations, and they need to think strategically about the purpose, *implementation* and execution of these solutions for their specific firm. Particularly for those focused on leveraging data science on the investment or alpha-capture side, implementation and cultural embrace of these solutions becomes paramount to avoid a clash between organizational efficiency and internal competition.

5. **Many questions remain about data science – ranging from legal and regulatory to existential.** Will machines eventually out-invest humans? What is the future for crowd-sourced algo funds? And from a regulatory perspective – what legal issues arise from these solutions and is there any cross jurisdictional consistency? We're still in the nascent stages of these innovations, and their long term implications are still taking shape.
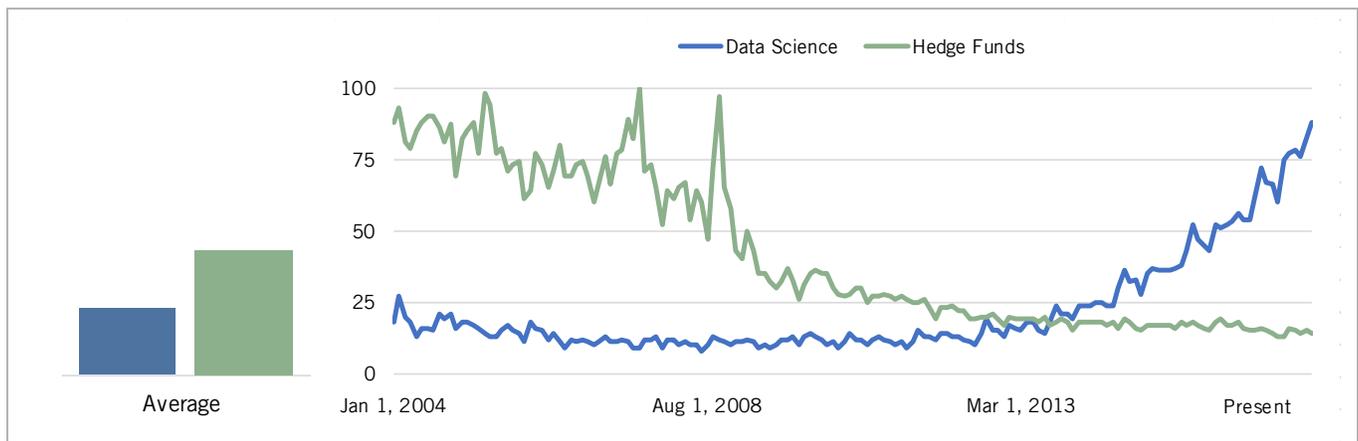
---

[1] "Wall Street to CEOs: Disrupt Your Industry, Or Else" The Wall Street Journal, May 26th, 2017

Jefferies

This paper is broken into four parts. **First**, because of the diverse and often colloquial nature with which people discuss data science, **we will define** the most frequently used terms. **We then map the current landscape** of data science in the hedge fund community, from early adopters, to early mainstreamers, to mainstream, to nascent adopters, and briefly discuss each. **The third section broadens the discussion** to understand how other actors in the hedge fund industry are leveraging these technologies, and we explore industry wide trends that are impacting the space. **Finally, we offer a vision of what may lie ahead** for data science at hedge funds – exploring opportunities, challenges, and what questions remain unanswered.

A quick exercise in data visualization – the below represents the relative interest in "hedge funds" vs. "data science" as mapped by Google Trends (which uses machine learning). Since 2004, worldwide interest in hedge funds vastly outstripped that of data science (peaking in 2008, for obvious reasons), but after inverting in 2014, the two have continued to diverge.

CHART 1

Interest over Time of "Hedge Funds" and "Data Science" from January 1, 2004 to present

Source: Google Trends

Jefferies

I. DATA, DATA EVERYWHERE, AND WAY TOO MUCH TO DRINK

Data, of course, has a nearly limitless history. The earliest recorded exhibit of "data" is thought to be a 20,000-year-old bone (the Ishango Bone), discovered in Uganda that seems to show Paleolithic tribespeople notching trading activity and supplies.[2] While we have come a long way from notches and bones – IBM estimates we create 2.5 quintillion bytes of data *per day* now – basic issues like how to identify, source, aggregate and understand this information endure. [3]

Simply: how do we "get it and digest it?"

Because the answer to that is *not* simple, and the language around it can often be confusing and pedantic, it's worth taking a moment to lay out some definitions.

Data science: an umbrella term that captures all methods, processes and systems to extract knowledge or insights from data in its various forms.

"Big" data: Most of the time this refers to any large set of data or information that can be computationally leveraged to reveal patterns, signals or trends. Big data has long existed, but the methods for gathering and extracting information from these sets are new; for example, credit cards have existed for more than 50 years, but investors' ability to parse credit card transaction data is more recent. This data may be unstructured or structured, with firms seeking to mine data sets themselves *or* have both the data sets and potential signals delivered in a more digestible manner. IBM breaks big data into four dimensions: volume, variety, velocity and veracity.

Alternative data: very simply, any data arising from sources not considered traditional financial data but that may offer market insights. Data doesn't need to be "big" to be alternative, but both alternative and big data are the fuel for data science. Drones that count cars at Wal-Mart or satellites that track ships are two examples of alternative data sources.

Data visualization: the ability to see and understand patterns or insights from data sets. Tableau, Spotfire and Qlikview are some commonly used data visualization solutions.

Artificial intelligence (AI): the ability of a computer to do things that used to require "human" intelligence – like speech recognition, decision making, linguistic translation or vision (image recognition). Two examples: i) Tesla's continuous updating of data gathered from its fleet daily to inform its autopilot feature, and ii) Pandora's system for recommending music.

Machine learning: a way that computers learn from themselves to discover and display patterns in data, adapting to new information on their own, the way the human brain does. Machine learning takes three forms: i) supervised (more traditional algos and stats like determining sentiment from having been given positive, negative or neutral tags), ii) unsupervised (interpreting patterns in a data set automatically – i.e. via "clustering") or iii) deep (which tries to mimic the work of human neural networks in the brain to iterate or "learn" when given new information). Siri (Apple) uses machine learning technology to better predict and understand our natural language questions.

Neural networks: how the human brain makes sense of the world via trial-and-error; retaining information and iterating or adapting what it "knows" when presented with new data points. Neural networks are the underpinning frameworks for how some artificial intelligence systems are constructed – namely; how can we make a computer "know" things, then adapt or update this knowledge when presented with new data.

Natural language processing (NLP): statistical techniques for learning language (i.e. – translation tools). One realm that's still just mostly for human intelligence? Sarcasm.

The data market is slated to nearly double by 2020, to upwards of $200 billion.[4] What does all this mean for hedge funds? Nothing short of a potential sea change in how information is sourced, digested and acted upon.

Markets continuously evolve, but data science and technology have prompted a shift that many feel will transform the entire finance industry. Traditional data sources have been largely commoditized; data science attempts to solve the new informational asymmetry question. Those who don't understand or can't leverage these opportunities across their businesses may be left behind.

Let's turn to the hedge fund landscape to understand where managers currently fall across the spectrum of embedding data science across their organizations.

---

[2]   A Brief History of Big Data Everyone Should Read. The World Economic Forum. 25 February, 2015

[3]   https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

[4]   Worldwide Semiannual Big Data and Analytics Spending Guide, International Data Corporation (IDC), October 2015
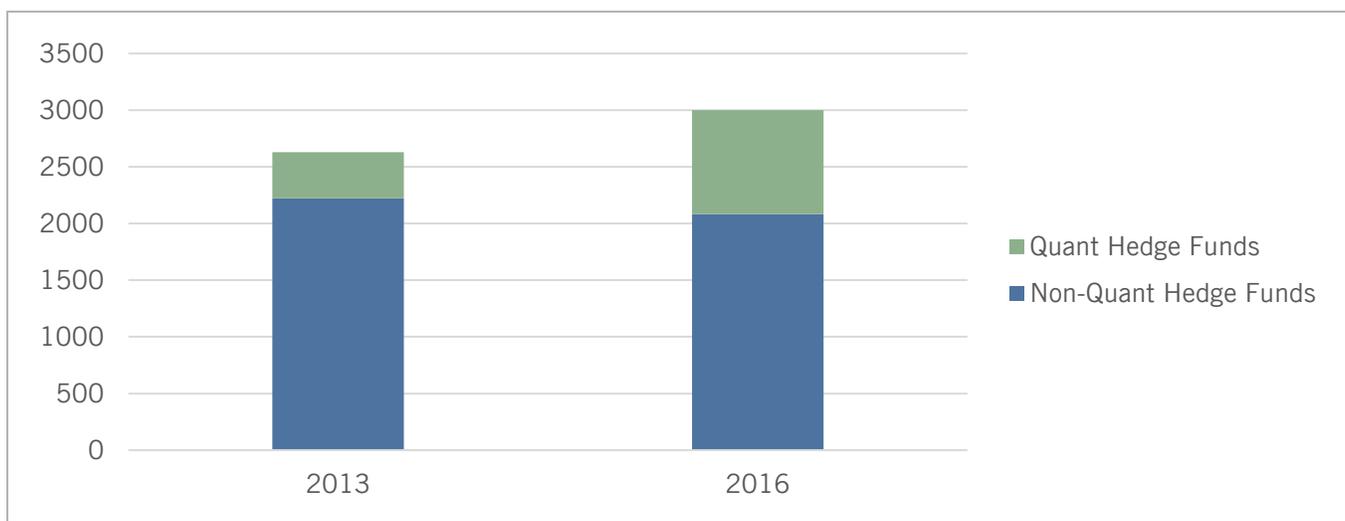
Jefferies

II. EVERYONE'S A "QUANT" THESE DAYS

Maybe for good reason. Quantitatively focused funds have doubled their assets since 2009 to nearly $1 trillion, and are one of the fastest growing areas of interest (and allocation) for investors.[5] But you don't have to be a quant fund to increasingly want to leverage quantitative resources and personnel in your organization. Firms with various core competencies are actively exploring how to improve investment theses, risk management or business strategy with the help of different data science solutions.

In the words of one data scientist, "Everyone seems to want to be able to quantify their intuition these days." An arms race is underway, with some sourcing various big data solutions and machine learning processes to create better predictive models. It remains to be seen whether Google searches predict sales of BMWs better than meeting a management team can, or if both are needed?

CHART 2

Hedge Fund Assets Broken Out by Quant and Non-Quant Strategies

Source: Hedge Fund Research

Both established managers and new launches are working to figure this out. Jefferies estimates that of the current Billion Dollar Club – 434 firms at last count of diverse focus, background and location – more than 20% currently have designated headcount focused on data science, or someone who spends more than half their time on the issue. What those individuals focus on, the types of teams they lead and their autonomy over budget and spend varies. And of course, these firms have reached a certain point of maturity to be managing at least $1 billion of AuM.

But new launches are also increasingly hiring data science talent *on day one*. While it may be too early to call it a trend, we are increasingly seeing managers launch at inception (or shortly thereafter) with one fundamentally focused analyst and one quantitative researcher/analyst. What is clear is that a diverse array of hedge funds is increasingly finding it important to incorporate data science or data *scientists* into their businesses, at all stages of the hedge fund lifecycle.

A number of managers of all sizes also report they're starting to field questions from allocators querying: "How do you approach data?" While open ended in nature and not specifically focused on big data or machine learning specifically, the question clearly indicates investors understand the changing nature of the information and data landscape, and managers increasingly need to be prepared to respond. One large investor notes, "We now ask the question, but without presuming an answer. We don't want people trying to do things 'just to do them' – but we think people should be exploring what's out there."

---

[5]   Hedge Fund Research, Q1 2017 Industry Report

Jefferies

A review of the hedge fund industry reveals firms currently fall into one of four buckets:

1. Early adopters   2. Early mainstream   3. Mainstream   4. Nascent adopters

CHART 3
Spectrum of Data Landscape Across Hedge Funds

	Type of firm	Headcount	Types of data science used	Use in firm	Budget	Overall
EARLY ADOPTERS	Quant specialty; Systematic	Dedicated data scientists, more decentralized model	Multiple big data sources, machine learning, AI	Source ideas or signals, risk management, business processes	Dedicated recurring OpEx with appetite for growth	Data science embedded in firm's DNA
EARLY MAINSTREAM	"Quantamental" across strategies	At least one data scientist or analyst full time, more centralized model	Big data most common, start to leverage unstructured data sets; may use machine learning & AI	Overlay for investment processes and some business	Dedicated budget; may fall under IT or charged to investment teams	"incorporated" investment and risk processes
MAINSTREAM	All strategies	Has at least one employee looking into data science solutions	Big data most common; typically structured data sets	Usually complement to existing investment processes	Sometimes dedicated; may be part of broader investment budget	Data science seen as "complement" or overlay to existing processes
NASCENT ADOPTERS	All strategies	None	Has not explored yet	Has not explored yet	None	In nascent stages of leveraging data science

Source: Jefferies Prime Services

Hedge funds fall along a spectrum of **embedding big data, data aggregation platforms, machine learning solutions, data visualization tools or technology consulting** into their firms.

Early Adopters – there are a number of mature, quant focused firms (typically systematic strategies) for whom data science has been embedded in their DNA since day one. These shops leverage massive data feeds via API and may have computing power that rivals the entire U.S. research library system *combined*. Early adopters have leveraged all that data science has to offer – artificial intelligence, machine learning, deep learning, big data from countless sources. Their goal is to embed data science ubiquitously across their firms. They leverage multiple data sets to systematically triangulate new signals, and view the convergence of these various sets as a useful source of signals, themes and ideas.

They have Chief Data Scientists and teams of individuals dedicated to the space, and use data solutions on the business side, from HR to compliance or operations. Early adopters are most likely to have started bringing the mountain to Mohammad, by co-founding a university research center on quantitative finance or data science, or opening outposts in Silicon Valley and staffing them with engineers from tech firms. These firms are also most likely to have hired high profile employees from places like Google, Netflix or Microsoft, and constantly hire talent at across all levels to focus on the issue.

From a structural perspective, early adopters are more likely to have discrete verticals of data scientists across the organization, including, but not limited to: data acquisition, sanitization, digestion and finally application. They are more likely to "pull" data, testing countless data sets on an endless hunt for obscure or less crowded and more accurate signals – equally focused on what data sets they're missing as what they have. For Early Adopters, data is as critical as the PCs on their desktop; as such they have considerable, dedicated budget for these resources, with appetite to grow.

Early Mainstream – these firms have been working to incorporate data science into their investment process (and sometimes more broadly across the firm) for the last few years, with the goal of scaling appropriately across the organization. These firms likely have dedicated or working groups managing data science implementation, and incorporate data scientists into investment idea generation and research processes. Early Mainstream firms are more likely to have centralized data science groups, who use trial and error to decide which data sets are most effective, but also may have quant analysts in discrete investment pods.

They may use unstructured data, but more typically leverage multiple structured sets. Early Mainstream firms sometimes overlay multiple data sets to identify signals, but typically view data through the lens of a pre-existing portfolio theme, as a discretionary complement "overlay" or for risk purposes. They dedicate budget to data science and understand this allocation is likely to grow in the coming years.

Jefferies

While Early Mainstream firms have a process for using data science on the investment side, it is less commonly used on the business side. And they most frequently face a cultural dilemma of balancing their discretionary backgrounds with embedding these technologies across the firm. There seems to be a trend among some Early Adopters and Early Mainstream in not only sourcing alternative data, but in allocating venture capital focused on investing in these innovations as well.

Mainstream – firms who have been adding (or looking into adding) data science solutions as a complement or overlay to pre-existing investment processes. These firms may have thought about adding quantitative personnel on the investment side, or have started to source structured data to further build out or audit their investment theses. These firms are in the early stages of understanding how data science may fit into their organizations; mainly seeing data science as a "supplement" to other efforts or approaches rather than something to be embedded in all aspects of the business. They typically leverage stand-alone data sets, and their spend is situational, based on particular data sets or technologies.

These firms are most likely to have identified data science as something that can help bolster performance or improve productivity, but remain in early innings of fully understanding how to effectively incorporate some solutions at their firms.

Nascent Adopters – Nascent adopters may not yet have begun exploring how data science could impact their organizations. They do not have any headcount focused on data science, nor are there formalized resources dedicated to exploring the space yet, or solutions deployed across the platform. Many nascent adopters either believe data science is just another fad, or that these resources won't serve as an effective business complement.

Even those who aren't embracing data science yet in their own firms are studying how data science is transforming the industries they invest in. Energy analysts, for example, may want to understand how Shell leverages big data in its exploration process to enable its geologists to make more accurate recommendations about where to drill. Big Oil, Big Pharma, tech firms (Google among those shifting from "mobile first" to "AI first"), retail, and other many industry groups all utilize big data and various data science solutions now, and analysts increasingly need to incorporate this information in their models.

CHART 4

Data Science Usage Across Various Hedge Fund Verticals

Idea Generation	Research Process & Back testing	Risk Management	Execution	Business Processes
Identify & Collect	**Analyze**	**Assess**	**Execute**	**Streamline**
Big Data • Structured or unstructured sets • Social Media, web search trends, consumer transactions, satellite or drone data, governmental or regulatory data Data Aggregation • Transforms raw data to more structured, digestible data sets Technology Consulting • Assistance in understanding, shaping and embedding data science solutions	Big Data Data Aggregation Machine Learning • Supervised, unsupervised or reinforced learning • Processes reams of data to identify trends, signals or outliers Data Visualization • Understanding unintended bets across the portfolio • Isolating particular risks • Technology Consulting	Big Data Data Aggregation Machine Learning Data Visualization Technology Consulting	Big Data • Sourced from trading venues, exchanges, and/or clearinghouses to optimize execution and settlement	Big Data Data Visualization Data Aggregation Technology Consulting

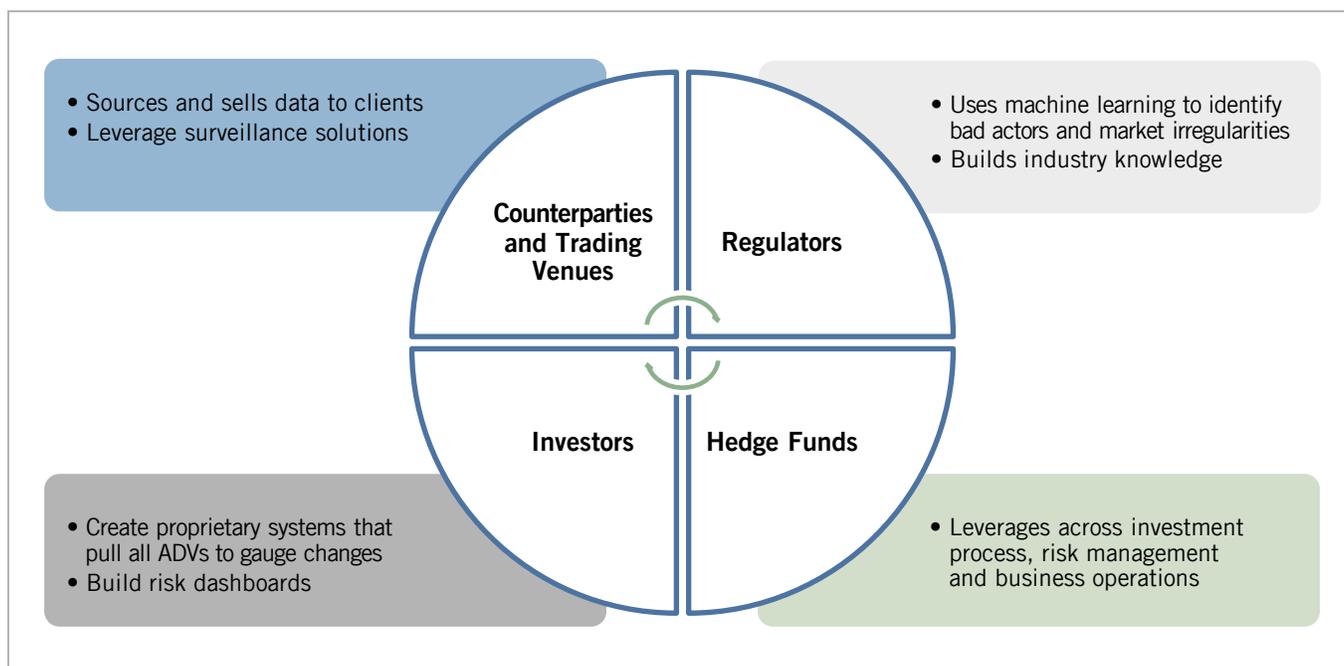Source: Jefferies Prime Services

Jefferies

There is considerable variety in how hedge funds have employed new technologies and solutions at their firms. Firms that try to incorporate data science without clear purpose or having thought through what's involved in bringing new innovations on risk inefficiency, waste and distraction. But when approached thoughtfully and incrementally, data science *may* improve productivity, enhance decision making and deliver new sets of actionable data not previously available.

III. DATA SCIENCE IN THE BROADER HEDGE FUND ECOSYSTEM

Counterparties like exchanges and other trading venues are increasingly selling data feeds to clients as new product lines. Regulators use data science to better understand market activity, and ferret out bad actors or market irregularities. Investors use data science to build internal risk dashboards to understand comprehensive portfolio risk, as well as conduct ongoing due diligence on managers. We explore each in turn.

CHART 5

Data Science in the Broader Hedge Fund Ecosystem

Source: Jefferies Prime Services

Counterparties or exchanges

In part because of shrinking margins, the rising cost of doing business, and lower volumes, counterparties and exchanges are increasingly turning to selling data or data analysis solutions to augment revenue. The Tabb Group estimates that revenues from exchange data and technology are up 62% over the last five years, with an overall revenue growth of 16% during the same time.[6]

The London Stock Exchange Group recently "raised its earning margin growth forecasts for coming years to future underline its reorientation from equity trading to high value market data and clearing."[7] A number of exchanges, including Intercontinental Exchanges (ICE) and the Chicago Mercantile Exchange (CME), estimate they can grow their data businesses at a rate of about 5-6% annually for the next few years.[8] And a number of exchanges have made a considerable acquisitive push of data companies in recent years, converging execution and content – Liquidnet's May 2017 acquisition of OTAS, as one recent example.

Even large asset managers report continuing to pivot their business models to focus on technology and data. As Blackrock noted in its 1Q2017 earnings call: "Technology is changing how the world invests and how we interact with clients…[we] intend to

---

[6]   The Market Data Deathmatch: The Increasingly Brutal Fight Over Equity Market Data Costs, The Tabb Group, January 2016
[7]   "LSE Lifts Guidance for Earnings Margin Growth," The Financial Times, June 12, 2017
[8]   Surinder Thind, Jefferies Equity Analyst, Asset Managers, Brokers & Exchanges

Jefferies

sharpen the focus on our technology and risk management businesses."[9] If the world's largest asset manager believes it's possible for considerable forward revenue growth to come from data, software and data analytics – what are the implications for the rest of financial services?

Allocators & Investors

Allocators and investors are increasingly interested in understanding how their managers are approaching the explosive growth of available data. They are starting to ask the question of existing and emerging managers, and while firms don't necessarily need to have onboarded data science solutions, at a minimum they are starting to have to give thought to the issue.

From a due diligence perspective, data science remains an emerging area – but sophisticated investors are increasingly using solutions to conduct ongoing due diligence on their managers – in some cases vetting thousands of ADVs to map changes, or scanning all investor letters for certain keywords or phrases. They've adopted these enhanced techniques as part of their overall risk management processes. A small number of investors are working to build better internal dashboards to understand their cross-portfolio risk profile and exposures to specific factors by leveraging big data and/or data visualization, often in conjunction with third party consultants.

Investors have also identified these solutions as worthy investments as an end in themselves. There is at least one fund-of-funds dedicated to investments in Artificial Intelligence funds, with more than $50 mm invested, and a number of investors report launching disruptive technology vehicles, with big data, machine learning and AI solutions among their holdings.

Regulators

Regulators leverage data science to: i) better map and understand market activity and ii) identify bad actors or market irregularities. Scott Bauguess, acting director of the SEC's Division of Economic Risk and Analysis, remarked in an October 2016 speech, that the agency was using machine learning algorithms to map topic and tonality signals to measure risk::

> "Some of these [new] analytical methods are allowing analyses of previously impenetrable information sets –
> for example, those without structure, such as freeform text. This has been of particular interest to the SEC,
> where registrant filings are often in the form of a narrative disclosure. So, as a result, we have begun a host of
> new initiatives that leverage the machine learning approach to behavioral predictions, particularly in the area
> of market risk assessment, which includes the identification of potential fraud and misconduct."[10]

The SEC's MIDAS (the Market Information Data Analytics System) initiative seeks to improve transparency in and understanding of the market, and the early stage Consolidated Audit Trail (CAT) is working to build a global database of market activity. Both leverage big data and machine learning technologies. FINRA and the CFTC have also engaged with vendors to develop machine learning programs for transaction surveillance. The CFTC has launched CFTC 2.0, a part of its LabCFTC initiative, to engage with new technologies like machine learning, artificial intelligence and other advanced analytics to "improve the effectiveness and efficiency of the agency in carrying out its day-to-day activities."[11]

These initiatives have helped regulators identify market irregularities or questionable trading patterns by triangulating trading patterns to uncover insider trading, as in some recent high profile cases. And regulators, exchanges, banks and funds alike leverage the expertise of firms like Digital Reasoning, whose machine learning technologies can conduct natural language processing for surveillance of potential internal threats.

Hedge Funds

The various ways hedge funds are embedding data science across their organizations is covered in section II.

---

[9]   Blackrock 1Q2017 Earnings Call Transcript, April 19th, 2017

[10]   Scott Bauguess, U.S. Securities & Exchange Commission "Has Big Data Made Us Lazy," October 21, 2016 speech before the Midwest Region Meeting of the American Accounting Association

[11]   CFTC website: CFTC 2.0, accessed at http://www.cftc.gov/LabCFTC/CFTC2_0/index.htm on June 12, 2017
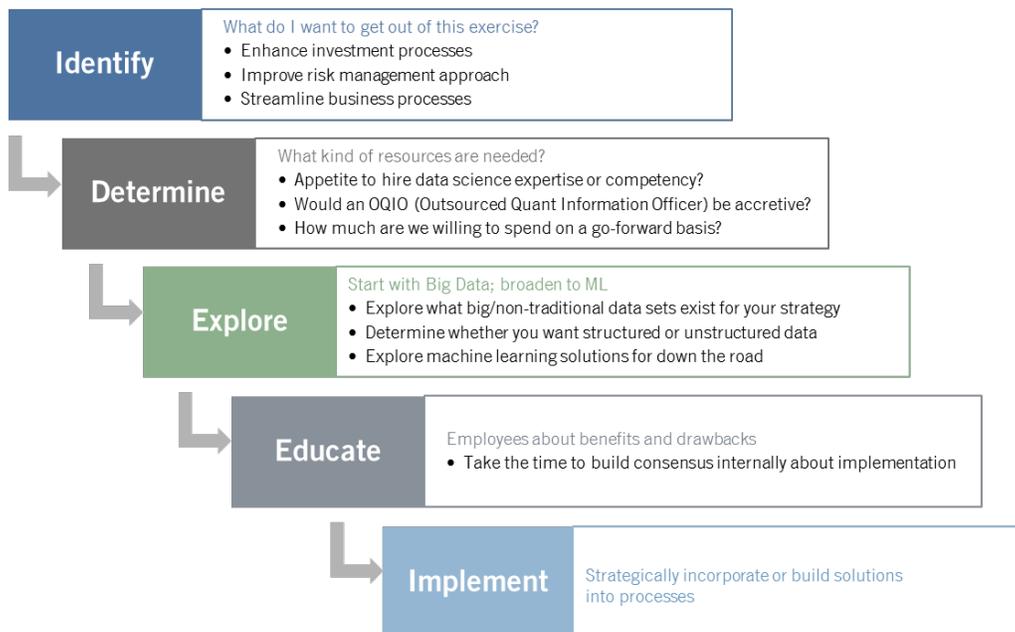
Jefferies

IV. WHAT LIES AHEAD

Available data continues its exponential growth march – with estimates of 44 zettabytes (or 44 trillion gigabytes) of data points generated annually by 2020.[12]

The most common question from hedge funds that have not begun exploring data science is: "Where do I even start?" We offer the following as a basic five-step framework for starting to understand and leverage data science at your organization.

Basic Initial Framework to Understand and Incorporate Data Science into Hedge Funds

Identify — What do I want to get out of this exercise?
- Enhance investment processes
- Improve risk management approach
- Streamline business processes

Determine — What kind of resources are needed?
- Appetite to hire data science expertise or competency?
- Would an OQIO (Outsourced Quant Information Officer) be accretive?
- How much are we willing to spend on a go-forward basis?

Explore — Start with Big Data; broaden to ML
- Explore what big/non-traditional data sets exist for your strategy
- Determine whether you want structured or unstructured data
- Explore machine learning solutions for down the road

Educate — Employees about benefits and drawbacks
- Take the time to build consensus internally about implementation

Implement — Strategically incorporate or build solutions into processes

Source: Jefferies Prime Services

Identify

What do I hope to get out of data science? Why am I even exploring different solutions for my firm?

Improving investment theses. Expanding inputs or ideas. Enhancing risk management. Identifying unintended bets. Running a more efficient organization. Staying current with industry themes

Is it possible effective implementation of data science could attract incremental investors, new talent, or new enthusiasm about the research and development (R&D) happening at our firm?

Determine

What expertise can I leverage in better understanding the data science landscape? What resources am I willing to dedicate to these efforts?

Not everyone needs to hire a dedicated data scientist to work in house. Expertise comes in many forms and may be sourced from current employees with an interest in the topic (typically a CTO or head of risk), third party consultants, or even experts at counterparties. Surveying others can help determine the most efficient and effective data science solution for a particular firm

In some cases, an "OQIO" (Outsourced Quant Information Officers) might be of help. Just as Outsourced Chief Investment Officers (OCIOs) have become a solution for some asset owners to manage their investment portfolios, it's possible we may increasingly hear of OQIOs serving as consultants to assist fundamental managers in understanding, selecting and managing the data solutions that make the most sense for their organizations

Should I attend any events hosted by data science providers or my counterparties that focus on these issues?

---

[12] "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things," EMC, April 2014

Jefferies

What is my anticipated spend for product, system and implementation? Is it a recurring or one-time cost – and what is my appetite for that spend growing on a go-forward basis?

Data science solution costs cross a broad spectrum; beyond recurring subscription costs, one time implementation or project costs can add up. It's critical to budget base and bull cases for bringing these innovations on board, but costs can scale quite linearly

Where will these allocations live in our budget? Under IT? Under research costs? Do we want to syndicate the costs across investment teams (if organizationally appropriate)?

Explore

What types of big data solutions exist? How should I be thinking about machine learning down the line – if at all?

There are already an endless – and growing – number of big data brokers. It's important to understand which will best serve the goals of your firm and which are least likely to reflect data decay in the short term (i.e. – that the sets get so commoditized, their signaling benefits are minimized)

Identify the data sets, sectors or themes that are most efficient and effective for your processes

Start to learn about machine learning and other methods of digesting these big data sets for a basic understanding of these capabilities, even if you don't incorporate them in the short term

Educate

There is often a cultural component to trying to incorporate data science into our organization

John Maynard Keynes said, "Difficulty lies not so much in developing new ideas as in escaping from old ones." Principals, CTOs and risk managers from across the organization need to buy in and fully understand both the opportunities and pitfalls for bringing new solutions on board. As with most cultural issues, data science needs to be embraced at the top. No group can advance to efficiently and effectively onboard these solutions without buy in from the principal. Implementation may be vertical (the principal buys in, and pushes the issue down across the organization), or horizontally shared across the C-suite, potentially including the CTO, COO, Head of Risk or Trading or senior technology analysts

Sometimes a few training or information sessions are required to fully educate employees and minimize disruption or conflict as new solutions are brought on

Implement

Roll out solutions on a strategic timeline, allowing sufficient space for incorporating and iterating as needed

Education of these solutions is now an ongoing effort; PMs, analysts or risk managers should work to stay abreast of emerging innovations

Other Considerations

Do I really need to spend time on this, or is it just another fad?

The exponential growth in data sets and processing solutions has meant that many people don't even know what they don't know. The universe is sufficiently vast and novel that at a minimum, it's worth exploring what might be of use to enhance pre-existing approaches or improve productivity

Could new data science solutions become part of your marketing or investor relations approach?

We are aware of a number of managers who have started incorporating more quantitative analyses in their pitches, tear sheets and marketing materials to show how they're leveraging new data sets or machine learning solutions. They're expanding the number of firm employees who meet with potential allocators to include data scientists or quantitative analysts, where appropriate. This is currently a differentiated approach; but it's possible 24 months from now, it will be common. There is considerable opportunity in turning the black box into a grey box and helping inform investors how new quantamental approaches are supplementing more traditional processes

What are the legal implications of incorporating data science into my firm?

There are numerous legal and regulatory questions that arise for these innovations, including: prohibitions around data scraping (is your data "clean?"), privacy laws, trespassing, misappropriation, and confidentiality. At this point, there is little clarity among many of these issues, and little standardization across jurisdictions

Jefferies

How Jefferies Can Help

The scope of the data science universe is enormous, and growing – further complicating things is that **there simply is not a one size fits all approach to adopting these solutions effectively** for all firms. The idiosyncratic nature of hedge funds and their investment processes means that principals, PMs, traders, CTOs and risk managers need to carefully consider the right approach for their organizations.

Jefferies has centralized its data science expertise, resources and solutions to better help clients approach this issue strategically, to maximize the return on your invested time covering the issue. Please do not hesitate to let us know how we can be of assistance as you look to learn more about the ever-changing landscape of data science and its possibilities for your firms, or if we may follow up on anything discussed herein.

IMPORTANT DISCLAIMER

Jefferies