Making big data work for commodities

From automation to AI, and coding with Python to democratised datasets, commodity trading is undergoing a digital transformation.
The Exxon Valdez ran aground on March 24, 1989.
As it did so, 250,000 barrels of crude oil gushed from the damaged ship and impacted the Alaska shoreline.
The accident was an environmental, legal and reputational disaster – but it also set off a chain of events that would upend the global shipping industry forever.

Introduction

Analysis is key to creating insight – and there’s more to analyse than ever

The year after the vessel crashed, politicians in the United States passed laws aiming to prevent similar significant oil spills. Within these laws was the obligation that some tankers should use vessel monitoring and tracking technology. If such ships could be tracked electronically, the argument went, there would be greater transparency about their movements, a lower chance of collisions, and ports could gain a better understanding of a ship’s actions. Similar systems were simultaneously being developed and tested by other authorities around the world.

By 2000, everything had changed. The International Maritime Organisation (IMO) adopted a new requirement for all vessels over a certain size to use automatic identification system transponders, known as AIS. The technology, whose adoption became effective in 2004, works like GPS tracking for ships, and includes a vessel’s identity, location, speed, direction of travel and planned destination.

Now, AIS data powers websites that track ships all around the world, and the information is fed into the commercial systems of multiple industries – including those of traders and commodities traders. The radical improvements in technology have enabled the tracking of ships in real-time to become routine. For commodities traders, the availability of this information opens up the potential to make smarter, data-informed decisions about the movement of goods – from cotton to oil and metal – around the world. The availability of this type of information, among a variety of similar sources, is increasingly becoming the cornerstone of fundamentals trading, which seeks to build a picture of the world using data.

AIS is only one data source – mariners can, of course, turn off their AIS broadcasts to evade tracking, but there are now fewer ways to hide than ever before. The introduction of low cost satellites and high-resolution cameras allows the Earth (and seas) to be photographed at all times. “You can be observed by satellite imagery,” says Leigh Henson, the global head of commodities trading division, at Refinitiv. “And even if there’s cloud cover, they can look at your location using radar”.

Introduction

Analysis is key to creating insight – and there’s more to analyse than ever
Machine learning is even making it possible to combine these data sources and automatically watch the movement of vessels over time. These data science techniques have caught ships illegally fishing, smuggling goods and breaking international laws. Most recently, ships have been found tampering with AIS signals and broadcasting false ones; combining data and machine learning has allowed them to be caught in the act.

The explosion of data is not unique to shipping. Across all commodities there has been an increase in the availability of data about what is happening in the real world. Cheap sensors, ubiquitous internet connections and real-time monitoring make it possible to track and quantify vast swaths of the commodities industry. Oil storage tanks can be monitored. Methane emissions can be tracked. The flows of oil pipelines can be watched. “Increasingly sophisticated types of observations are just sprouting up all the time,” Henson says.

As this picture of the world has grown in scope, so too has fundamentals trading become ever more sophisticated. It is now based on thousands of daily news reports, billions of data points from vast numbers of sensors, almost-instantaneous updates from financial markets and the proprietary data held by commodities firms. Increasingly, that same data is being provided by multiple sources; the measurements of one oil silo can be taken by multiple companies who provide their data to the market.

But simply amassing data is not enough to give traders an advantage over the competition. Traders who have been equipped with the right analytical tools, including the ability to automate research and visualise goods being moved, are able to understand the world in a much more granular way. Being able to effectively utilise this knowledge will become increasingly critical during the coming years. As technology makes trading more automated, every second matters. Companies need to be able to quickly analyse and visualise the data they have, while staff need the skills to operate analytical tools and gain insights. And the whole process starts with data management.

This e-book, which is part of a four-part series, outlines the key data practices and techniques that traders and analysts should be putting in place to gain an edge. It explores how the commodities industry has been revolutionised by data and the steps needed to ingest, standardise, analyse and visualise data. Without these processes in place – either internally or being provided by third-parties – trading companies are likely to fall behind the competition in an increasingly digitised world.
that data. The first are the government agencies that control the flow of information from their localities. Secondly, there are commercial organisations and big data providers that hold certain datasets, such as the weather or news data, in structured ways. Thirdly, there are international monitoring organisations such as the International Maritime Organisation, the body that introduced AIS data to shipping. Finally, there are the exchanges that deal with trading data. Some of these data sources are available for free online, but others require subscriptions and commercial deals.

“In one sense, data is more difficult to come by, since data vendors fully appreciate its value,” Till explains. “However, in another sense, data is easier to come by in that some data sets simply did not exist in the past – at least commercially – that do so now.” Collecting all this data and making it available to people has clear advantages. By providing detailed market information alongside fundamental data, Refinitiv is able to bring greater transparency to the market. It has never been easier for companies to understand where and how commodities are being moved and sold.

Michael Adjemian, an associate professor of agricultural and applied economics at the University of Georgia, says data collection within agriculture is changing both the agricultural industry and the world of trading. “Satellite data on weather and production patterns, as well as data generated by GPS-equipped tractors, drones and soil-sensing applications, are regularly integrated into real-time dashboards to improve the information available to decision makers in the financial economies,” he says. Adjemian adds that this will be crucial given growing worldwide demands for food and the challenge of the climate crisis.

But collecting masses of data causes its own challenges. The volume of data has become incomprehensible for humans, and being able to make the most of it requires being able to effectively analyse it. It’s now crucial that data can work with other data. “The competitive advantage is really evolving from being able to access those additional sources of data, to how well companies can integrate it, commingle it with the data that they already produce themselves, and then apply technology to generate insights,” Refinitiv’s Sanos says.
Commodities trading has been transformed in the last two decades – but there is even greater change coming. Artificial intelligence, sophisticated analytics tools and vastly improved visualisation methods point to a future that includes increased automation – but this doesn’t necessarily mean replacing human traders and traders. Automation can include far more effective data processing and a sophisticated application of available information. Essentially, machines can mine data for intelligence; humans can then act upon it.

But the use of data can’t be supercharged if its fundamental components are missing. Companies and traders need to properly ingest data into their systems, make sure it is standardised, use data science to analyse it, and understand the tools needed to visualise this analysis and make it understandable for people on the ground.

The task of preparing data for the future isn’t an easy one, Sanos says. It is, however, a challenge that Refinitiv has a solution for. Refinitiv’s Data Management Solution (RDMS) is a platform that allows companies to combine Refinitiv’s multiple sources of commodity data into their own processes. Data can be merged and normalised, and it acts as a way for clients to standardise data from Refinitiv alongside third-party data and their own information.

Preparing for success: getting your data ready for business

With the proper tools, data from multiple sources can be standardised and merged, paving the way for automation and artificial intelligence.
The data explosion has created new challenges for companies. “You also need somewhere to put it all,” says Refinitiv’s Henson. Companies need to make the most of the data they are collecting – or accessing from third parties such as Refinitiv – and need to create suitable environments where it can be collected and aggregated. However, Henson explains, increasingly companies are looking to access and manipulate data remotely, adding that traders and analysts want to connect to data feeds – and then ingest it into even bigger databases.

It’s here where the cloud’s quick startup times and seamless ability to expand as required come into play. By taking advantage of cloud hosting and making data available in Application Programming Interfaces (APIs), it is possible for traders to easily access huge data streams and use them in the ways they desire. For instance, Refinitiv’s RDMS allows analysts to create ways to see how much oil is being moved from one location to another. Henson also explains that Refinitiv provides data environments in its terminal product, Eikon, which holds more than 2,000 pricing data sources, with 1,300 providers sending reports. It is also an access point that is light on memory and can be used in virtual offices to aid remote use. The company estimates that Eikon covers more than 22,000 businesses, with historic data stretching back 65 years. This data can then be accessed and shared easily, and subsequently acted upon.

Disparate data united
Combining multiple, continuously updated sources into vast databases can yield exceptional insights – if they can be navigated and parsed.
Making one set of data work with other sets of data isn’t an easy task. As an example, data must first be normalised, and fields in one database need to match those in another database if they are going to be successfully combined. If that doesn’t happen, the result may be an inaccurate picture of the world – and it may subsequently be harder for traders to make accurate decisions as a result.

Sanos says the chief technology officers he speaks to are frustrated about the amount of time their data scientists and staff are spending cleaning up datasets in order to optimise their usefulness. “They’re telling me their analysts spend up to 90 per cent of their time just doing this aggregation and normalisation of data,” he explains. Spending longer making data compatible means less time is spent analysing the data and making decisions based on the intelligence it provides. For many companies, the process of amassing data sources and standardising the information is best left to third parties. However, Refinitiv has dedicated teams – located in Bangalore, Manila, Kenya and Poland – that analyse incoming datasets and ensure they are compatible. They take disparate sources of data and make it possible for them to be easily compared to other data.

For traders, taking in already standardised datasets (through APIs or desktop software) that are ready to be mixed with proprietary or third-party data can vastly cut down on the amount of time needed to make trading decisions. “Previously, our customers may have been happy with just one source of oil flows and cargo tracking,” explains Simon Wilson, head of oil trading at Refinitiv. “Now, they want to take multiple sources.” He adds: “We integrate two or three sources so that they can get a blended view of the cargo tracking.” This approach can help traders make more informed decisions.

Once data has been ingested and normalised, it can then be passed through analytics tools, visualisation processes and – increasingly – artificial intelligence and machine learning systems.
Gone are the days of solely using Microsoft Excel for analysis – the datasets and sources available to traders are now too large for the spreadsheet software, and so newer tools are needed to effectively analyse big data. The coding language Python, which is used across web development and a number of platforms, is increasingly becoming the system of choice for analysts sifting their way through data in the world of financial services. It’s outstripping both Java and C++ as the go-to coding language.

Python’s growth is down to the ease of writing and the huge amount of pre-written code, available through data science libraries, that is accessible online. The language is also flexible enough to enable code to be run in the cloud or put into APIs, and it’s one of the cornerstones of emerging machine learning and artificial intelligence tools. It is increasingly being used to model financial markets and handle the vast amount of data that’s gathered.

The rise of Python is changing the skillsets of trading workforces. Those who can code are in demand and are being hired by financial service companies – numerous large banks are utilising Python-based infrastructure for their work. In fact, some predict that many commodities trading houses and financial services will mostly be hiring only those with Python skills in the coming years.
The value of analysing commodities data comes from the insights it can unlock. Increasingly powerful visualisation tools, along with custom instruction code written in Python, are being used to bring data to life. By making the results of analysis visual, it’s possible for traders to better understand what is happening.

Take, for example, the tracking of ships. Through the use of AIS data, each vessel at sea can be accurately placed on maps of the world. It’s possible to click on a ship and search for its details, see its position, movements and destination. Diving deeper, it is even possible to understand the cargo that it may be carrying and infer what its movements might mean for trading markets.

But this is just the tip of the iceberg. Tools such as Power BI, Tableau and others are allowing standardised data to have Python scripts run against it and produce new results. The tools make it possible to see bigger patterns within the data and put them into a chart for legibility. For instance, a Sankey diagram could show the movement of individual grades of crude oil moving from West Africa to Northern Europe. Data presented in this fashion is often easier to comprehend than numbers in a table, and the process allows current and forecast views of the market to be created. With better forecasts, it’s another tool that makes it easier for traders to make informed decisions.
Trading moves towards a more automated, technological future

Even greater levels of disruption are coming. The last decade has seen the fundamentals commodity market become flooded with these new types of data. But the data acts only as a starting point – what can be built on top of it will provide new opportunities.

Enter machine learning and automation: since 2000, machine learning techniques, which use sophisticated algorithms to examine data and identify patterns within it, have improved exponentially, and they will only continue to grow more powerful and advanced. Much like the influx of data fuelling the commodities industry, this sub-field of artificial intelligence has been democratised. It is possible for individuals or businesses to find all the tools they need to run machine learning applications online and, often, for free.

Throughout the 2020s, the use of machine learning will increase across all industries. For commodities, this could mean more automated processing of data and potentially even changing the way that trades happen. “I think we’ll see more algorithms,” Henson says. “We’ll see more computers trading instead of people. It’s been growing in the equities markets and it’s definitely been growing in the commodities world, too.” AI can step in to perform tasks that are too complex or too time consuming for humans.

The level of disruption that will actually happen is difficult to predict. However, the ultimate goal may be using artificial intelligence to accurately forecast the trading price of individual commodities. It is likely humans won’t be taken entirely out of the loop, but instead receive greater automated insights from machine analysis that they can then apply their expertise to and make better decisions with.

If companies don’t keep up with the current changes – starting with getting the basic data elements in place and fit for use – and prepare for the ones still to come, then they will lose ground that they find becomes impossible to regain. Companies need to be able to properly ingest data that has been standardised and mix it with their own if they are to see the benefits of visualisation and automation.

“ ‘For the majority of all the promising technologies, they need a solid data layer that they can trust,” Sanos says. “This data layer should include proprietary data that companies may currently not be using, as it is not easy to commingle it with other data sources.”

Use of machine learning will increase across all industries. For commodities, this could mean automated processing of data.
Companies need to be able to properly ingest data that has been standardised and mix it with their own if they are to see the benefits of visualisation and automation.

About Refinitiv

Refinitiv, an LSEG (London Stock Exchange Group) business, is one of the world’s largest providers of financial markets data and infrastructure. With $6.25 billion in revenue, over 40,000 customers and 400,000 end users across 190 countries, Refinitiv is powering participants across the global financial marketplace. We provide information, insights, and technology that enable customers to execute critical investing, trading and risk decisions with confidence. By combining a unique open platform with best-in-class data and expertise, we connect people to choice and opportunity – driving performance, innovation and growth for our customers and partners.