

Enhancing asset management insights with alternative data ingestion

Benefits of leveraging structured + unstructured data

Whitepaper | May 2023



Analytical contacts

Mohit Modi

Global Head of Data & Analytics

CRISIL GR&RS

mohit.modi@crsil.com

Nick Dalbis

US Data & Analytics Leader

CRISIL GR&RS

nicholas.dalbis@crsil.com

Contents

Leveraging the power of alternative data5

Baking in AI, inventing intelligent pipelines6

Evaluating third-party and open-source tools.....7

Conclusion8

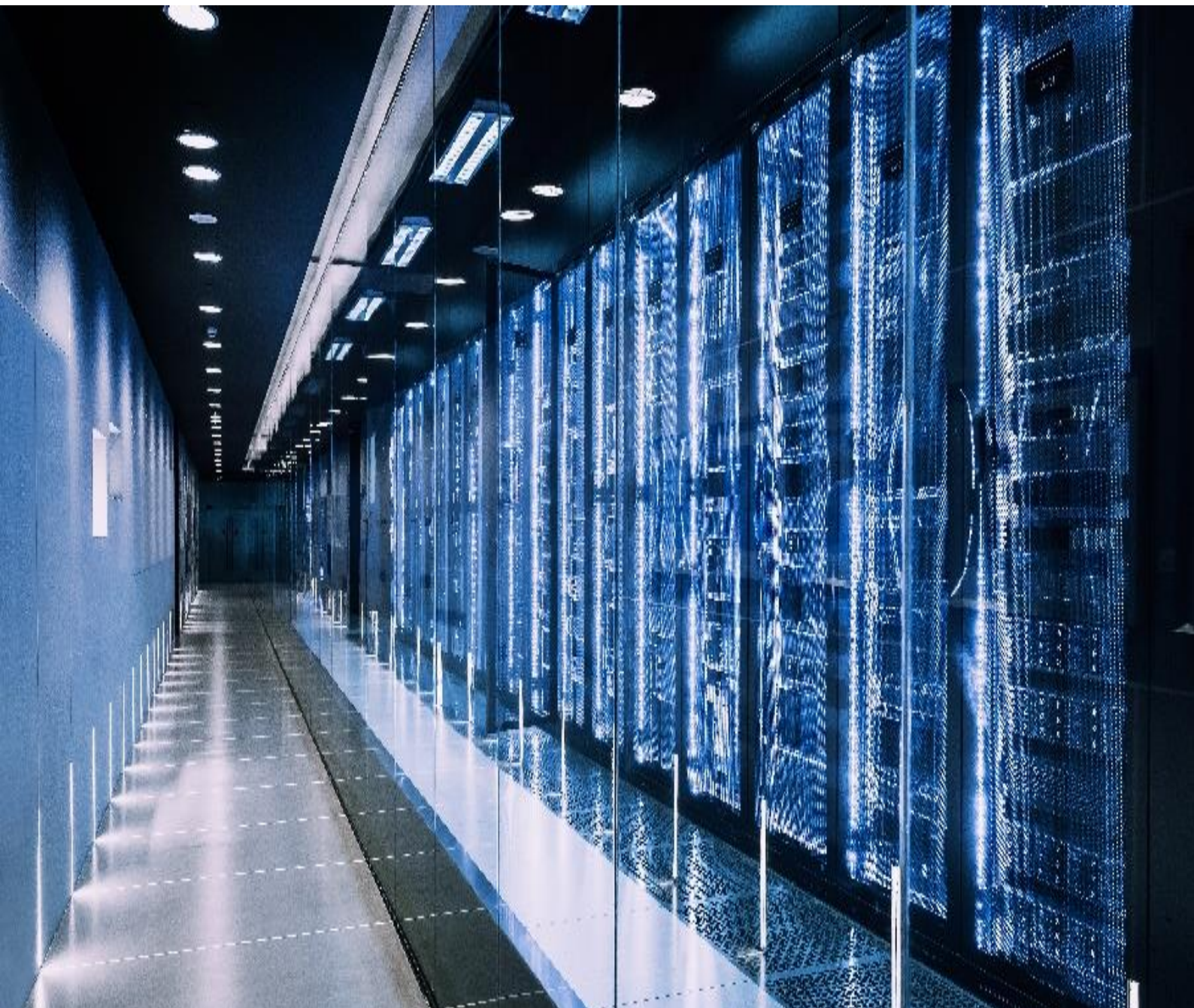
This whitepaper underscores the possibilities and benefits of alternative data ingestion for asset managers — from basic to complex solutions. It also discusses why they should consider proactively embracing ‘intelligent’ data pipelines if they want to be future-ready.

Data pipelines — a set of processes that ingest data from various sources, transform, optimize, and move it to a destination for storage and analysis — are an evolving area with burgeoning applications ranging from field it research to businesses and financial services.

Many firms offer end-to-end solutions to set such pipelines up, and most of these have easy connectors to a multitude of data sources, from csv files to simple storage service (S3).

But practitioners feel the real killer app here is the one that can ingest both structured and unstructured data effectively to offer rich insights and dashboards. It also needs curating capabilities to learn on its own.

Not having such a solution is a limitation, including for asset managers with fiduciary remit over investor money in a high information-velocity world.



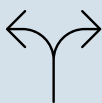
Leveraging the power of alternative data

Social media has evolved from being the personalized newspaper of the common man to the sentiment maker of the masses within no time. The recent banking crisis in the US, which saw large-scale deposit withdrawals, illustrates the frenzy created by, and the power of, alternative data sources in the investing world.

But with uber-connectivity and explosion of information come great opportunities, too. For asset managers, specifically, some of the key benefits include:



1. **Alpha generation:** Asset managers can get information faster than competitors from blended datasets delivered in real time. They can also gain new insights into market trends, consumer behavior and economic indicators that were previously unavailable. This could help them make more informed investment decisions, increase process efficiencies to reduce costs, identify potential risks and opportunities, and increase revenue potential. Financial services firms are using web scraping techniques to extract information from various data sources to gain a more holistic view of the market and identify stronger patterns.

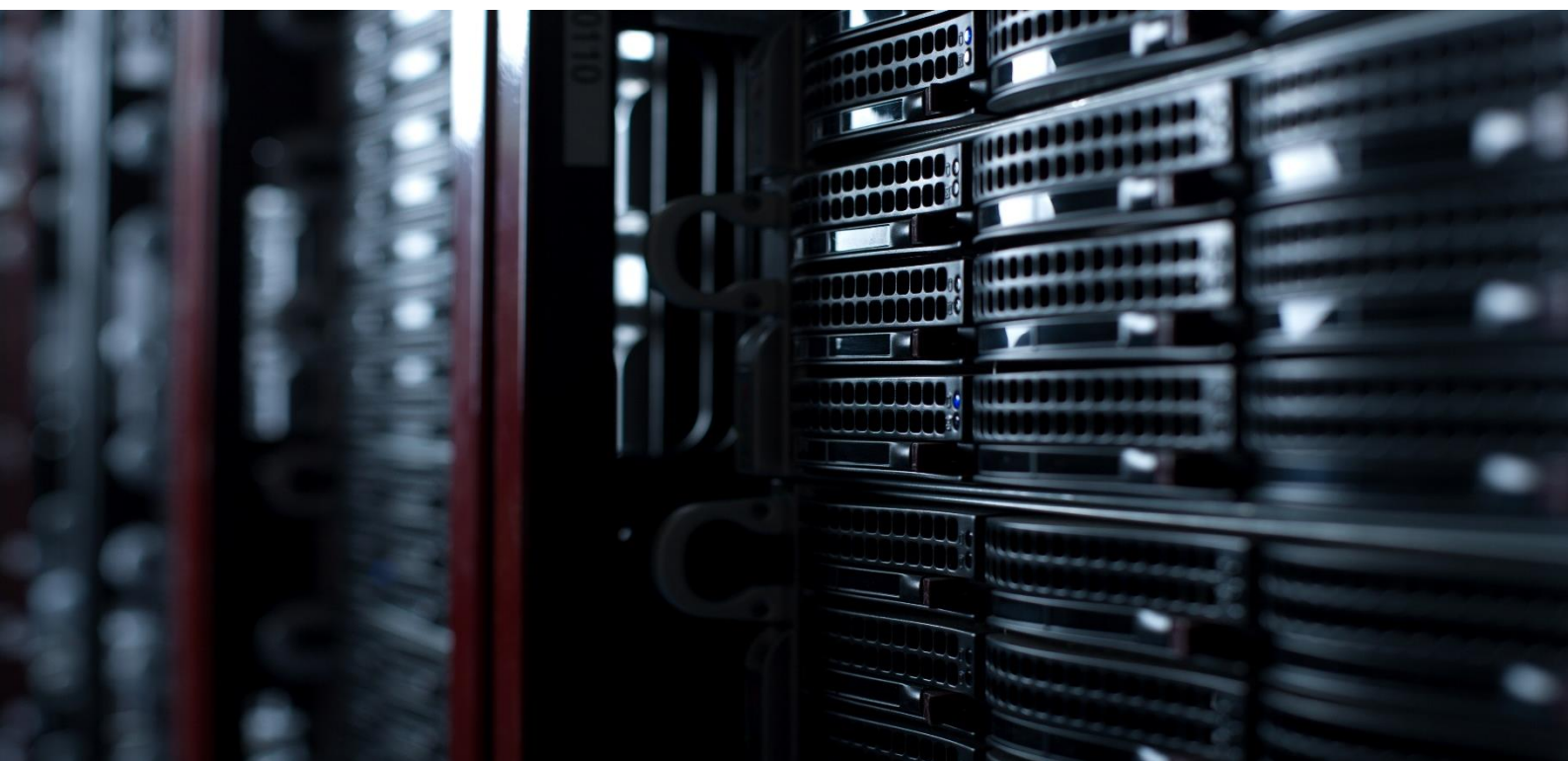


2. **Improved decision-making:** Innovative alternative datasets mixed with traditional asset management data provide a more holistic view for projections, help optimize decision-making capabilities, and enhance the quality of risk assessment. Asset managers are using a multitude of data sources to improve ETL processes that, in turn, streamline visualization and reporting output to simplify business decisions.



3. **Greater efficiency:** Automated data preparation and augmented analytics help data engineers reduce data wrangling time. This translates into efficiency for analysts, by feeding cleaned-up data straight into their models for faster and timely insights. Companies are ingesting unstructured financial statement data through data extraction tools to quickly standardize data for faster turnaround time for data processing and insights generation.

Overall, the use of alternative data has the potential to revolutionize the way asset managers operate, providing them with a powerful tool to gain new insights and drive investment returns.



Baking in AI, inventing intelligent pipelines

When the use of alternative data is combined with artificial intelligence (AI) and machine learning (ML), data pipelines could drive richer, more advanced data transformations.

As Gartner® states: “The AI community has traditionally focused on improving outcomes from AI solutions by tweaking the AI models themselves, but data-centric AI shifts the focus toward enhancing and enriching the data used to train the algorithms. In addressing AI-specific data considerations, data-centric AI disrupts traditional data management, but organizations that invest in AI at scale will evolve to preserve evergreen classical data-management ideas and extend them to AI in two ways: add capabilities necessary for convenient AI development by an AI-focused audience that is not familiar with data management, and use AI to improve and augment evergreen classics of data governance, persistence, integration and data quality.”¹

This calls upon firms to rapidly adapt and integrate generative AI into their technology stacks.

Indeed, the asset management industry, like many others, is at the cusp of an AI-led revolution. Externally, the Fourth Industrial Revolution is bringing on newer challenges and opportunities for the industry.

The rapid rise of ChatGPT is a prime example. Market leaders who can produce commercially viable solutions from ChatGPT will gain significant first-mover benefits. These solutions could be efficiency oriented, e.g., on automation of client communication, or could be revenue accretive, e.g., on leveraging alternative data for accelerating deal flows, or on other intuitive applications to drive client and internal benefits.

Here are four distinct ways asset managers will increasingly benefit from advances in this field:



1. **Use of generative AI:** Generative AI has emerged as a major trend in recent years, driven by advancements in deep learning and neural network technologies. This type of AI can create new data and content, rather than just analyzing and processing existing information. Its use was initially limited to more mundane tasks such as customer query responses, but we expect it to gain acceptance in high value-added tasks such as automated investment note preparation.



2. **Data science-assisted automated data preparation and augmented analytics:** Machine learning (ML) algorithms automate the process of data cleaning and preparation, allowing businesses to access insights more quickly and easily from alternative data sources. ML is also used in identifying patterns and insights into large datasets and deriving inferences for faster decision making.



3. **Edge computing for real-time data processing:** Processing data closer to the source rather than sending it to a central location can enable real-time data processing and analysis, making it easier for businesses to respond quickly to changing market conditions. Technology firms are creating distributed processing systems to increase bandwidth and reduce response times.



4. **Cloud-based data ingestion and analysis:** Cloud provides businesses with access to large-scale computing resources and flexible data storage solutions, making it easier to manage and analyze large datasets, while also reducing infrastructure costs.

Evaluating third-party and open-source tools

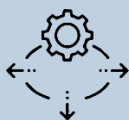
Many solution providers have developed pluggable components based on third-party as well as open-source tools. These components provide highly effective and specific, configurable solutions for the industry. Assessment of these tools requires close attention to critical success factors and a deep domain expertise-led approach that pure IT firms cannot provide. A blended technology and financial services firm can offer the astute understanding to both sides of the house to achieve the desired outcomes. While evaluating any such tool or component for use, an organization needs to check and validate a few parameters:



1. **Ability to quickly onboard new data sources:** To remain flexible and efficient, the key is in ingesting data from new sources and with unforeseen variables, without customization or code development.



2. **Comprehensive data extraction:** All data points within various file formats must be extractable. Reliance on a hard-coded scheme for each source could cause problems as ingestion of alternative data scales.



3. **Scalability:** Ingestion engines must be capable of avoiding bottlenecks and downtime when ingesting and processing multiple data sources at the same time.



4. **Security and compliance:** Infosec is a key consideration across the value chain. It becomes all the more relevant when ingesting data from diverse external sources. Another important aspect is compliance with laws, especially the ones covering copyright and personal data, while scraping public data.



Case study

CRISIL's automated data ingestion solution for a top financial services firm

**50% reduction in effort hours,
25-50% better data latency,
improved time to market**

A large financial services firm was looking for a scalable solution that:

- Ingests large volumes of data, including unstructured data, across 25 asset classes, 30,000 entities, and 200,000 documents
- Is agile, with low dependency on the central technology team or coding resources
- Meets high data quality expectations and still provides possibility of reactive data quality measures
- Works on multiple destination systems with varied consumption formats

To solve these challenges, we implemented our proprietary tool that manages the complete data operations lifecycle. This included a taxonomy-based approach for the architecture, a highly configurable platform enabling end-users to define the data to be collected and disseminated, automated quality checks on the data, and an ML-based model to automate extraction of financial data from scanned documents.

The solution generated tangible business benefits across the data lifecycle, including:

- ~50% reduction in effort hours due to automated extraction of data
- Improved time to market, with 25-50% improvement in data latency
- Configurability enabling 90% reduction in lead time for making changes
- 4,000+ automated checks on the data, thereby enhancing the overall quality
- High uptime and enhanced user experience
- Fast and easy onboarding of new asset classes

Not all solutions need to be complex. Asset managers can use a multitude of low-code/no-code platforms to address their specific data ingestion and workflow challenges. These solutions can be low-cost as well as functionally pliant, but may not have high-end capabilities such as AI-enabled optical character recognition engines for financial statements.

About CRISIL Limited

CRISIL is a leading, agile and innovative global analytics company driven by its mission of making markets function better.

It is India's foremost provider of ratings, data, research, analytics and solutions with a strong track record of growth, culture of innovation, and global footprint.

It has delivered independent opinions, actionable insights, and efficient solutions to over 100,000 customers through businesses that operate from India, the US, the UK, Argentina, Poland, China, Hong Kong and Singapore.

It is majority owned by S&P Global Inc, a leading provider of transparent and independent ratings, benchmarks, analytics and data to the capital and commodity markets worldwide.

About Global Research & Risk Solutions

Global Research & Risk Solutions is the world's largest and top-ranked provider of high-end research and analytics services. We are the world's largest provider of equity and fixed income research support to banks, and the foremost provider of end-to-end risk and analytics services to trading and risk management functions at world's leading financial institutions. We provide corporate research and analytics solutions to operations, strategy, and sales and marketing teams of corporations globally. Coalition provides analytics and business intelligence to 14 leading global investment banks. We operate from 8 research centers in Argentina, China, India and Poland, working with clients across time zones and languages. Being part of CRISIL enables us to attract and retain top quality talent. We have over 2,300 employees, 75% of whom hold advanced degrees in finance, accounting and management. We employ the largest number of CFAs and CFA aspirants in India. We have won top honours at the World HR Congress on Talent Management and HR Project for the year 2015. We have also won the NASSCOM Exemplary Talent Practices Award (NExT Practices) for skill development for two years in succession in 2011 and 2012. The award recognizes us as a firm that has the vision to proactively invest in its people and get them future-ready.

We are committed to delivering cutting-edge analysis, opinions, and solutions. This underscores our proposition of being the best people to work with.

CRISIL Privacy Notice

CRISIL respects your privacy. We may use your contact information, such as your name, address, and email id to fulfil your request and service your account and to provide you with additional information from CRISIL. For further information on CRISIL's privacy policy please visit www.crisil.com.