



From Complexity to Efficiency – Modernizing Data Pipelines

SUMMARY

In the fast-paced world of investment asset management, accurate and timely data is crucial for informed decision-making. Data pipelines serve as metaphorical vehicles that facilitate the movement of data from its source to delivery points. Traditionally, investment asset management firms have relied on multiple pipelines for different data sources, resulting in complex and cumbersome processes. This paper explores the concept of modern data pipelines in investment asset management, their advantages over traditional pipelines, and best practices for their implementation.

DATA PIPELINE DEFINITION

A data pipeline in investment asset management refers to the process of moving data from its origin to its intended destination, with intermediate steps of processing and formatting. Data Pipelines often go through three primary stages – Ingest, Process and Store. Modern pipelines may go to five stages, including Ingestion, Processing, Storage, Analysis, and Serving.

- **Ingestion** is the intake of multiple batched or real-time data sources, including both structured and unstructured data into the pipeline and converting it into formats that can be processed and stored.
- **Processing** includes cleaning and validating the data, formatting it into a standardized structure, normalizing values, and enriching it with additional information. Data transformations ensure data consistency and quality before further processing.
- **Storage** takes processed data and stores it in the appropriate data storage systems, including SQL databases, Hadoop, Apache Spark, Good Cloud, Amazon S3, and Snowflake.
- **Analysis** is evaluating stored data using various techniques such as exploratory data analysis, statistical modeling, or AI and machine learning algorithms. The goal is to derive insights, patterns, and trends from the data that can support investment decision-making or further analysis.
- **Serving:** Analyzed data is served to multiple endpoints or users within the investment firm. This can include feeding the data into data marts for specific business units, utilizing business intelligence tools for reporting and visualization, exposing APIs for integration with other systems, or populating data warehouses or lakes for long-term storage and analysis¹.

In investment asset management, companies typically have multiple pipelines tailored to specific data sources, such as ABOR (Accounting Book of Record), IBOR (Investment Book of Record), and vendor feeds. Other data pipelines are tailored to the users or applications such as Equity Analytics, Data Warehouse, quantitative analysts, or trade teams, or for Business Intelligence reporting. These pipelines can be established through point-to-point connections or data-sharing mechanisms.

ISSUES WITH TRADITIONAL DATA PIPELINES

Traditional data pipelines in investment asset management face several challenges. One significant issue is the proliferation of point-to-point connections. This arises from a traditional business model where each department independently funds IT projects to address their specific data needs, resulting in the creation of isolated pipelines that often serve one (or a few) purposes each. Over time, this approach leads to a complex web of connections, making maintenance and troubleshooting difficult.

From Complexity to Efficiency – Modernizing Data Pipelines

ISSUES WITH TRADITIONAL DATA PIPELINES (continued)

Another challenge is the duplication of data quality measures, data checks, and transformations across different pipelines, leading to inefficiencies and increased costs. Costs further add up due to the quality assurance and controls needed to determine why the same measurement may yield a different number in pipelines used for operations vs. client reporting, regulatory reporting, or internal reporting.

Traditional data pipelines also often suffer from limited scalability based on their lack of flexibility and agility in adapting to changing business requirements and emerging data sources. They may not easily manage new data formats, sources, or analysis techniques, limiting a firm's ability to be competitive in the rapidly evolving landscape of the investment industry. For all the reasons above, it is clear why modern data pipelines have emerged.

THE MODERN DATA PIPELINE

In the investment industry, a modern data pipeline represents a streamlined and efficient approach to data management and analysis. It involves the intake of data from various sources, such as market data feeds, trade execution systems, and client portfolios, and the processing of the data to deliver accurate and timely insights for informed decision-making.

A modern data pipeline combines the benefits of data sharing and point-to-point connections. It prioritizes the elimination of silos and encourages the sharing of data across departments and systems, enabling a holistic view of investment assets. This data-sharing approach reduces duplication of efforts and ensures data consistency throughout the organization. At the same time, it recognizes that certain sensitive or high-volume data may still require point-to-point connections for security or performance reasons.

Moreover, a modern data pipeline operates within a governed framework. This framework establishes standards for data quality, transformation, and governance, ensuring data integrity and compliance with regulations. It also emphasizes high data observability, enabling real-time monitoring, and analysis of data flows for pro-active decision-making. By implementing a modern data pipeline, investment asset management and ownership firms can enhance operational efficiency, accelerate data-driven insights, and drive business growth.

BEST PRACTICES FOR MODERN DATA PIPELINES

Implementing modern data pipelines in investment asset management involves following these best practices:

- 1) **Aligning business data delivery methods:** Companies should align data delivery methods with specific business use cases. This entails identifying the most suitable mechanisms for data sharing and point-to-point connections based on factors such as data sensitivity, volume, and latency requirements.
- 2) **Weighing cost of change vs. benefit:** This oft-overlooked best practice is crucial to firms maximizing cost/benefit ratios. When migrating from one-to-one data connections to data-sharing mechanisms, investment asset management firms should carefully evaluate the cost of change against the expected benefits. This assessment should consider factors such as improved efficiency, reduced maintenance efforts, and enhanced data accessibility². Avoid migrating to a modern technology if the cost is higher than the benefit.
- 3) **North Star approach:** The ultimate objective of modern data pipelines is to eliminate point-to-point data movements and promote data sharing as the primary method of data transfer. Establishing a "North Star" vision aligns organizational goals and drives the transformation process toward this objective. North Stars endure despite a journey's unsteady path towards it, allowing for flexibility in reaching goals. Using this approach, the move to modern data pipelines becomes less of a project with a finish, and more of a journey and evolution over time with measured progress and movement forward.
- 4) **Designing a flexible framework:** Investment firms should design a flexible framework that accommodates evolving data needs. This framework should provide guidelines for data ingestion, transformation, governance, and observability, enabling seamless integration of new data sources and technologies without regard to type or complexity. The more a company adopts the framework, the more benefits it can yield.
- 5) **Implement data governance:** In addition to the framework above, establishing data ownership, data lineage, and data access controls provides better benefits to the pipelines. Governance should also address metadata management with practices to capture and store information about data sources, transformation, and lineage. This can aid in understanding data dependencies, impact analysis, and in facilitating data discovery.

(continued)

From Complexity to Efficiency – Modernizing Data Pipelines

BEST PRACTICES FOR MODERN DATA PIPELINES (continued)

6) **Data observability:** Implement robust monitoring and alerting mechanisms to track the health, performance, and availability of the data pipeline. Monitor data latency, job failures, resource utilization, schema drifting, and system metrics. Use centralized logging and dashboarding tools for effective monitoring.

7) **Maintain balance with centralization and flexibility:** Maintain a delicate balance between centralized standards for data observability, security, and compliance, while also delivering self-service data access to developers and analytical teams

CONCLUSION

Modern data pipelines in the investment industry address the limitations of traditional pipelines by promoting data sharing, reducing point-to-point connections, and establishing a governed framework. By following best practices and leveraging expert guidance, organizations can optimize data management processes, enhance decision-making capabilities, and gain a competitive edge in the dynamic investment landscape. In today's competitive landscape, where one in six asset and wealth managers is to consolidate or close by 2027³, it is vital that companies view data as an asset. To stay in the game, firms should establish the most efficient modern practices to maximize their return on investment and gain any edge they can against the competition.

MERADIA'S EXPERTISE IN MODERN DATA PRACTICES

Meradia, a trusted leader in investment asset management data solutions, offers expertise in modern data practices. With a deep understanding of the industry, meradia assists organizations in designing and implementing efficient data-sharing mechanisms, migrating from point-to-point connections, and establishing robust data governance frameworks. To leverage Meradia's expertise, investment asset management firms are encouraged to participate in a data maturity assessment, which provides insights into current data practices and identifies opportunities for improvement.

REFERENCES:

¹From: <https://www.fivetran.com/learn/data-pipeline-architecture> on 7/11/2023

²From Joseph Lisbon (Managing Director, Charles Schwab), panelist at Institutional Investor Data & Analytics Symposium, Building a Modern Pipeline

³From FundFire - [One in Six Asset Managers to Consolidate, Close by 2027: Survey](#) – July 16, 2023



Andrew Jacob, CFA, brings a well-rounded data management perspective to Meradia's global portfolio of clients. Throughout his career, Andrew has supported constituents across the investment management industry, including retail investors, advisors, wholesale teams and institutional investors. He is a seasoned investment professional with diverse experience in business and entrepreneurship, possessing strong acumen of front office operations, data governance and management, and retail client journeys. Andrew's passion for helping people and driving outcomes aligns with Meradia's principles of Perspective, Passion and Impact.