

Unfolding the winning approach to a modern data management strategy

Abstract

The business requirements of modern enterprises are evolving at a rapid pace, but legacy integration tools can't keep up with the velocity, variety, and volume of data in the cloud-first world. Market-leading organizations need to intelligently manage and modernize their data stacks to advance their cloud analytics and warehousing initiatives.

We had an exclusive chat with Forrester analyst, Noel Yuhanna to unravel an intelligent, future-proof data management strategy for modern enterprises. Here's a walkthrough of business-specific use cases, covering a spectrum of data pipelining solutions, helping you adapt to the changing landscape of modern data stacks.

Question 1: How has the uptake in the adoption of cloud data warehouses like Snowflake, Redshift, and BigQuery impacted traditional ETL frameworks?

Traditionally, ETL frameworks were mostly batch-driven, moving data from transactional and operational systems to data warehouses at a predetermined interval (usually every 24 hours). While this approach worked for decades on-premises, the evolution of the cloud drove the need for quickly ingesting larger volumes of data in an automated way.

Also, organizations wanted to support modern insights that required data from new sources, including cloud, social, and edge, in a self-service manner. In addition, the requirements grew in supporting real-time analytics, extreme performance, and complex data transformation to meet the new demands for businesses.

Unlike traditional on-premises data warehouses, cloud data warehouses and data lakes can scale dynamically, separate compute from storage to scale independently, and process streaming data quickly through parallelism. It offers more optimized data processing and transformation capability, leveraging elastic resources with a high degree of automation to support the most demanding use cases. As organizations move towards supporting more real-time data initiatives, we believe the need to ingest and process streaming data becomes critical to support these data warehouses and data lakes.

Question 2: How can organizations adapt to a changing landscape where Hadoop (ETL) is making way for cloud storage and data warehouses (ELT)?

Today, modern data pipelines can separate ingestion, transformation, and data processing to orchestrate across multiple sources quickly to support large volumes of data. In addition, they focus on the design and run-anywhere approach, allowing them to be repeatable to accelerate other workloads and use cases. Also, data pipeline tools can abstract computing to run across multiple platforms, offering extreme performance and scale for the new generation of operational and analytical use cases.

With a high degree of automation and self-service capabilities built-in into modern data pipelining solutions, designing and deploying data ingestions and processing has become simpler than ever, requiring less effort and technical skills. In some organizations, we find that even business users and analysts leverage modern data pipelines to support their digital transformation initiatives. Today, organizations use modern data pipelines to support many use cases, including data warehouse automation, real-time analytics, data science, operational insights, customer 360, customer intelligence, and data lake transformation.

Question 3: When to consider ETL vs. ELT for your data pipelines?

ETL has been around for decades to support analytical use cases. It offers the ability to move and transform primarily structured data from transactional and operational systems to data warehouses. ETL is mostly associated with batch processing and is mainly used by DBAs, data analysts, and data architects. ETL remains a viable approach when the data volume isn't huge, data sources are static, and transformations aren't too complex. Organizations typically use ETL to support business intelligence (BI), SQL analytics, business reporting, predictive analytics, and integrated analytics.

On the other hand, ELT focuses on loading data quickly into a data warehouse or data lake and then running data transformations based on policies and rules. It is suitable for running more complex workloads where the data volume is huge, running hundreds of terabytes into petabytes, and supports all kinds of data, including semi-structured and unstructured data. We see ELT used by data scientists, data engineers, data analysts, and DBAs. Organizations typically use ELT mainly to support data science, AI/ML, BI, augmented BI, cloud migrations, and complex analytical requirements.

Question 4: What role do data pipelining tools play in establishing an intelligent data management strategy?

We believe data pipelining tools play a crucial role in supporting modern and intelligent data management strategies. Today, most organizations struggle to support self-service, trusted, real-time analytics, and operational use cases. Data pipelining is one of the tools that help overcome these challenges, offering a high degree of automation, self-service, and data intelligence capabilities to accelerate various business use cases. In addition, most data pipelines support multiple public clouds and some hybrid clouds. Data pipelining is used by different personas, including data scientists, data engineers, developers, and business analysts, helping them work collaboratively and more efficiently.

Today, data pipelining is helping organizations modernize their data management architectures to support new and emerging business requirements, including marketing and finance analytics, streaming analytics, data lakes, data science, AI/ML, and data warehouse ingestion. In addition, data pipelines can help organizations support data fabric, data mesh, cloud data warehouses, cloud data lakes, and cloud lakehouses for cloud, distributed, and edge use cases. And with prebuilt connectors, the data pipeline can help access any data stored in apps, databases, events, and logs quickly to support broader granular marketing, finance, and retail analytics.

QUESTION 5: WHAT ARE THE KEY PARAMETERS TO CONSIDER WHILE SELECTING ANY ETL/ ELT TOOL?

There are several key parameters for selecting an ETL/ELT tool. However, a lot depends on the type of use case you are supporting

- Look for tools that offer a single pane of glass to support cloud, on-premises, hybrid, and multi-cloud environments, making it simpler to manage data processing and transformation across various environments.
- Self-service capabilities can help accelerate deployments with minimal user interaction and skills. Look for solutions that have a point-and-click UI.
- Ensure that the tool can deliver the required performance and scale for your use cases, especially when dealing with low-latency requirements. The tool should have the ability to scale linearly and support cloud-native capabilities.
- The tool should be able to reuse existing data pipelines with some customization to support multiple use cases. This allows organizations to scale their requirements quickly without starting from scratch.
- Look for capabilities to support DataOps and MLOps easily without a learning curve and adapting to the business requirements.
- Security should not be overlooked. Ensure that sensitive data is protected using encryption, masking, access controls, and auditing is enabled where needed.
- The tool should have the ability to connect to the most common data sources, including databases, files, data warehouses, data lakes, object stores, and SaaS applications.

DEVELOPMENT, USE, AND CITATION GUIDELINES

The following summary of Forrester's guidelines applies to the use of the above post-engagement Q&A copy:

- Forrester Citation Policy states that all Q&A copy and usage, plus format, is subject to review prior to publication.
- Q&A copy must be company, product, and brand agnostic.
- When published, the Q&A must be client branded. Use of the Forrester logo is subject to additional guidelines and approval.

- Q&A copy contains a minimum of three (3) to a maximum of (5) five questions related to the original engagement content (i.e., webinar/video/speech content), with a two- to four-sentence answer per question.
- Q&A copy must link clearly to the analyst engagement that it goes along with and clearly identify the analyst as a guest speaker at that engagement.
- It must be clear that the questions are coming from the client (or from audience members if a post-event document) and that Forrester is answering them.
- The analyst works with the client to identify relevant questions related to the scope of the engagement, which are then reviewed, edited, and approved by the client.
- The analyst develops responses to the questions based on the engagement, their research coverage, and market knowledge.
- The analyst submits the completed copy for editing, citation review, and approval by Forrester.
- The client offers its audience the final deliverable in web (e.g., blog or landing page) or PDF format (e.g., post-event email offer or download). Alternative formats require citations approval.
- Q&A copy can be cited in other materials specifically related to the engagement. It must not be cited in materials unrelated to the engagement (e.g., quoted in a separate white paper or other presentations).



Scan and
start free 14-day trial



Scan to
schedule a demo

