

The background of the slide is a photograph of a business meeting. A man in a blue suit and glasses stands in the center, holding a tablet and looking at it. He is surrounded by three other people: a man in a light blue shirt sitting to his left, a woman in a black top sitting to his right, and another man in a light blue shirt sitting in the foreground with his back to the camera. They are all smiling and appear to be in a collaborative discussion. The setting is a modern office with large windows in the background.

Key considerations for ETL modernization

Efficiency and agility in business processes drive the growth of any successful organization. ETL (Extract, Transform and Load) tools play a critical role in delivering the speed an organization requires to access its data efficiently. In the digital age, ETL modernization is becoming a must to keep pace with the growing data and data sources.

The need for modernization of ETL platforms

As the data landscape of enterprises increases exponentially, traditional ETL tools do not work best to handle the complexity of the data from various sources: today, data could be stored on cloud or on premise; it could be static or stream data, and may be stored in repositories sitting in different countries with different data protection laws. Traditional tools were created at a time when the requirement was to manage lower volumes of data and processes. These tools may not be able to scale up for the modern data landscape requirements.

Traditional ETL product licences cost millions to organizations and hence they seek to employ open source frameworks to perform ETL operations, which enable similar and better functionalities as traditional ETL products.

Legacy ETL tools also face issues while handling real-time data processing from various social media channels. Scalable, faster and flexible environments for new age digital applications

ensure data processing is done in real-time. Existing ETL data pipelines need to be modernized to support real-time data in addition to transactional and analytical data workloads conversion.

Existing ETL tools have challenges in providing efficient and flexible metadata management and lineage across systems for robust regulatory and governance needs.

Embarking on the modernization journey

Many large enterprises have been exploring ways to transform traditional ETL platform data pipelines and workflows leveraging open source processing frameworks.

Traditional ETL data processing pipelines, predominantly meant and built over decades for batch processing, are under pressure while open source processing frameworks are catching up with the momentum. Also, these frameworks are well aligned with the Big Data applications for processing and managing huge amounts of structured/semi-structured/unstructured data being generated from several new age and existing enterprise systems.

Open source processing frameworks predominantly supported by Scala/Java/Python programming languages and equipped with out-of-box libraries and utilities, and scalable for huge volumes of data processing on cloud, on-premise and hybrid environments, are taking the centre stage for catering to the digital and innovative re-imagined business scenarios.



The right approach to ETL modernization

There is no easy path or way to convert and migrate thousands of legacy ETL jobs developed over decades in the organization's landscape to modern approaches such as Spark and microservices based processing frameworks.

Some of the key considerations while embarking on the ETL modernization journey are:

- Involve right stakeholders from IT and business across business lines by justifying the appropriate use cases and return on investment
- Assess and prioritize subject areas and associated business process-related data pipelines and dependent workloads from other process areas in the current data estate for conversion to define overall roadmap
- Do deep analysis of the legacy data pipelines and segment them into one-to-one conversion to open source ETL or re-architecture/re-factor scenarios. Categorize pipelines under re-architecture where one-to-one conversion of batch data processing workflow may not be relevant and need to re-engineer for the new age and digital use cases
- Select the right approach, solution, platforms and frameworks to achieve conversion automation, semi-automation and required manual custom coding
- Package, orchestrate, containerize and deploy modernized pipelines for creating scalable run time environments and clusters on cloud, on-premise and hybrid frameworks

- Define and establish well governed processes for operationalization to monitor and sustain the environment
- Induct right talent as well as upskill traditional ETL pipelines development and architect staff to work on open source framework components to ensure optimal investments and acceleration of the modernization journey

To conclude

ETL modernization with its cost-saving approaches to transactional and analytical data processing is becoming a key strategy for organizations' IT estate rationalization. ETL modernization help businesses reimagine their business processes and integrate their enterprise applications data with external systems such as merchants and channels in real-time in a more flexible and scalable manner.

Open Source ETL frameworks offer numerous business use cases such as developing customer knowledge graphs for better understanding of the prospects/customers, payment processing of billions of transactions, real-time fraud analytics, compliance and regulatory related data processing pipelines, among others. A right ETL modernization strategy aligned to the enterprise's digital strategy and the IT estate rationalization roadmap, and an implementation framework and execution approach with the aptly-skilled resources are the key success factors of the ETL modernization journey.

About the author

Mohan Mahankali

Practice Leader and Principal Architect - Information Management, Data Analytics & Artificial Intelligence, Wipro Ltd.

Mohan has 20+ years of business and IT experience in the areas of information management and analytics solutions for global organizations. In his current role, he is responsible for practice vision and strategy,

solution definition, customer advisory, consulting, competency development, and nurturing of emerging trends and partner ecosystem in the areas of data and information management. Mohan is the co-owner of a patent in data management and governance awarded by USPTO (United States Patent and Trademark Office).



Wipro Limited

Doddakannelli, Sarjapur Road,
Bangalore-560 035, India

Tel: +91 (80) 2844 0011

Fax: +91 (80) 2844 0256

wipro.com

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading global information technology, consulting and business process services company. We harness the power of cognitive computing, hyper-automation, robotics, cloud, analytics and emerging technologies to help our clients adapt to the digital world and make them successful. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, we have over 175,000 dedicated employees serving clients across six continents. Together, we discover ideas and connect the dots to build a better and a bold new future.

For more information,
please write to us at
info@wipro.com

