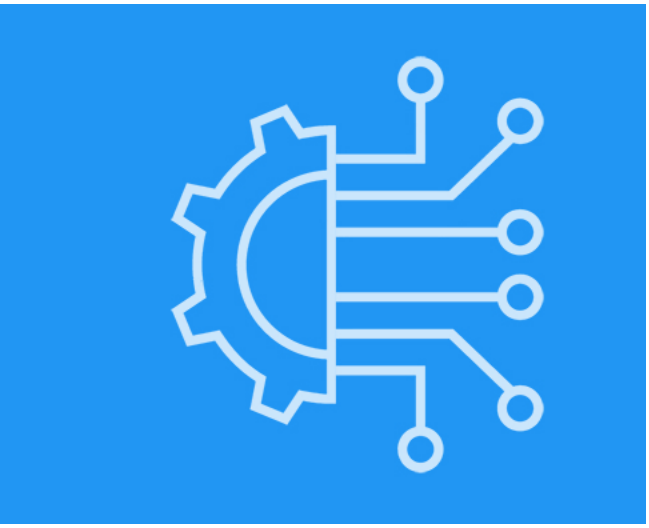
eLuminous™
TECHNOLOGIES PVT LTD

WHITE PAPER



Data Integration in 2020:

Revolutionize Your Data Integration

Table of Contents

Executive Summary

Data is the new oil; goes the conventional saying. But with an explosion of data-generating apps and devices over the past decade, it would be prudent to refer to data as the new crude oil. That is because data in its raw and unprocessed form comprises documents, files, logs, etc. from a variety of applications, devices and systems. Add to that data from non-conventional channels like social media, and it becomes imperative to seam all the data together to make informed business decisions.

Figure 1: Explosion of Data-generation

A vital part of this process is data integration, whereby data derived from disparate streams are combined before being stored and processed to provide important and insightful information. However, traditional data integration methods are not equipped to deal with how data is generated and interpreted in the digital age, and how that is changing the way businesses make decisions. The traditional method usually meant a trade-off between quicker delivery and quality. Selecting the former meant rapid execution of projects, but at the cost of quality. If quality was put at the forefront, timely delivery suffered which led to costs going out of control. In either case, the outcome would be less than ideal.

Next-generation data integration techniques address precisely this trade-off by empowering data management teams to complete projects on-time, within budget and at the expected quality. Such an approach toward data integration harnesses best practices and tools that have gradually gained better capabilities and greater reach to drive a business toward a truly Agile and data-driven strategy.

This white paper delves into the business and IT challenges that are driving the move toward next-generation data integration techniques and defines the central characteristics of a next-generation data integration ecosystem. It also presents a practical guide to aid this transition and provides an overview of how eLuminous' customers harnessed improved quality, speed of integration, and cost efficiency to make business-critical decisions using next-generation data integration techniques.

Next-Generation Data Integration Approach

One of the hallmarks of next-generation data integration is the ability to reuse integration logic, mappings, and components.

[*page no 7*]

Why Traditional Data Integration Approach, Architecture and Methodologies Need An Overhaul?

In the early days, business intelligence systems were built upon data warehouses, which were manually populated with data cleansed and transformed from operational systems.

This clean data formed the basis of business analytics, which provided business users with valuable insights. In today's business environment, however, the focus is more on obtaining near-real-time analytics to automate important business processes and gain vital insights into customer behaviour. A major reason for this change in approach is the massive amounts of data being generated today vis-à-vis an earlier era.

Figure 2: Business Intelligence Systems

Moreover, avenues like machine data and social media are not only contributing to the quantity of data being generated but also its quality. Today, a variety of data is used for analytics including semi-structured and unstructured data, in addition to voice, text, and image data. Together, these two are the driving factors behind the acceptance and rise of machine learning. Compared to traditional human-coded algorithms, machine learning algorithms learn and evolve based on the training data and feedback. Going forward, explicit programming will gradually cede ground to machine learning-based implementations.

However, the effectiveness of machine learning is a function of the training data fed to it. Machine learning models trained with bad data irrevocably lead to bad models. Colloquially referred to as 'dirty data', it remains one of the biggest problems faced by data scientists today and a major impediment to productive machine learning implementations. The only way of overcoming this hurdle lies in harnessing effective data integration, which includes collecting, labelling, cleaning and organizing the raw data.

The traditional data integration approach, however, is ill-equipped to deal with these challenges. Its emphasis on code-primary and manual integration consumes a lot of time, thus hindering data scientists from focusing on more productive and strategic work. In the worst case scenario, the traditional approach can impede a machine learning project altogether, thus laying waste to what is an essential business investment.

> At the heart of any data integration project is the realization that **data volume,variety, and velocity will grow in the future**.
>
> [*page no 8*]

Drawbacks of Traditional Data Integration Approach, Architecture and Methodologies:

Data integration has evolved a fair bit over the years. But that hasn't prevented IT teams from grappling with spiralling costs, expensive delays, and subpar results because of traditional data integration. The blame for this perpetual state of playing catch-up can be laid at the door of data complexity. Not only are new sources and types of data being introduced, but they are done so at never-before-seen volumes and speeds.

These developments have exposed traditional data integration methodologies as being the bottleneck for information-hungry businesses. As seen below, the disadvantages of the traditional approach extend not only to the financial aspect, but also data fidelity, scalability, and customer (businesses) behaviour.

A) Cost too much time and money

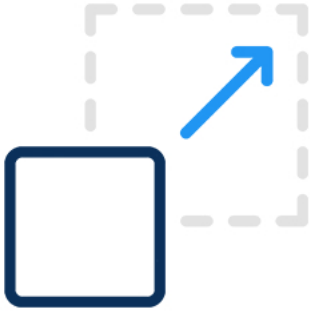
First-generation data management systems are based on inflexible tools and architectures, which means they lack the ability to reuse data integration patterns and other best practices that are essential to processing vast quantities of diverse data. The deficiency of sound metadata further complicates the data management process. Dealing with these complexities while transitioning to newer kinds of data not only takes too much time but also inflates the overhead many times over.

B) Compromise data fidelity

The absence of a holistic approach to data governance and quality combined with the inability of data environments to deal with rapid changes in data sources and types results in data that is difficult to trust. The issue is further compounded by the lack of real-time technologies to update data more frequently, which leads to outdated data. These twin problems are at the heart of why traditional data management is ill-suited to the more fast-paced business world of today.

C) Restrict scalability

Scalability, an essential feature for businesses to grow in the digital age, suffers as traditional data integration approaches are unable to keep up with the massive data volumes, velocity and variety that are characteristic of the new big data era.



Next-Generation Data Integration Approach

The modern-day approach to data integration relies on several technologies to enable reuse of work across the entirety of data integration projects, rather than limiting it to only data warehousing. It also places an Agile approach and modular architecture at the centre stage, which enables businesses to respond more proactively to rapidly-evolving changes and demands.

A) Automation of DI (data integration) processes, development and data validation

In the traditional structure, developers, data professionals and IT were left to fend for themselves when dealing with critical ancillary tasks. But with next-generation data integration, automation is an inherent part of core integration and development work. This ensures development and deployment processes are even more timely and efficient. These benefits extend to even the testing phase of a data integration project, where a minimum of 50 percent time savings through automation is up for the taking. That is a significant achievement as close to 30 percent of a typical DI project's timeframe is taken up by tasks such as testing and validation of data.

B) Application of Agile Methodology

Compared to traditional data warehousing, modern data integration projects are more uncertain and iterative, which makes them well-suited to an Agile approach. The long gestation periods typically associated with previous-generation data warehousing plans are out of place in today's world where business conditions change by the minute.

Business intelligence derived out of such an approach is not only slow but also ineffective. Agile data integration, with its incremental and iterative methodology, helps IT teams respond to new and changing business requirements with greater flexibility, faster turnaround times, and better quality.

C) Reusability of integration logic and universal data access

One of the hallmarks of next-generation data integration is the ability to reuse integration logic, mappings, and components. Such a flexible data integration architecture enables the use of a codeless and graphical mapping environment to build a data movement process, which can then be prototyped and converted into an ETL without recoding the mappings.

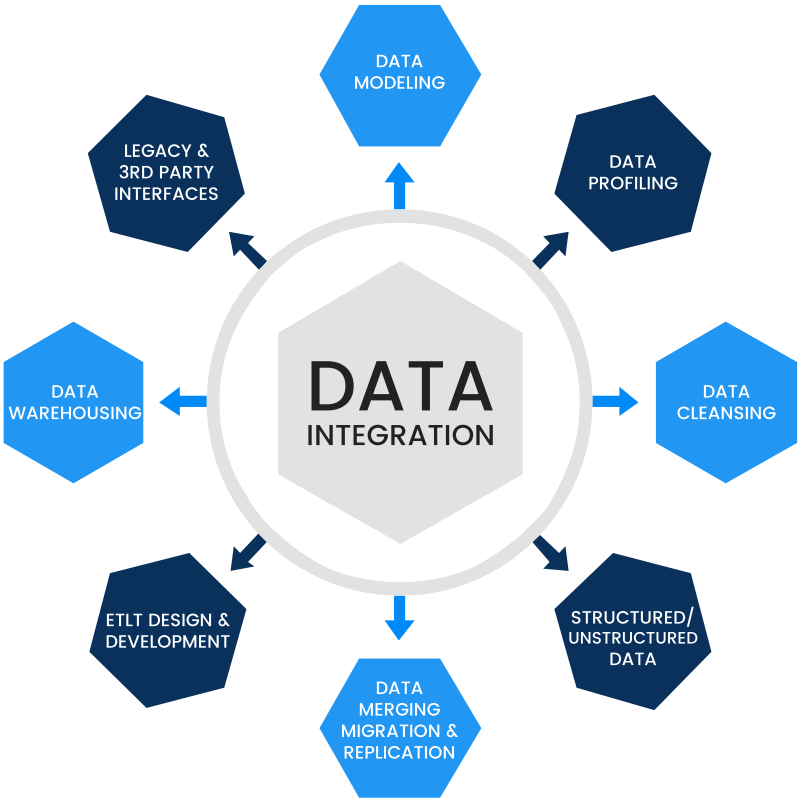
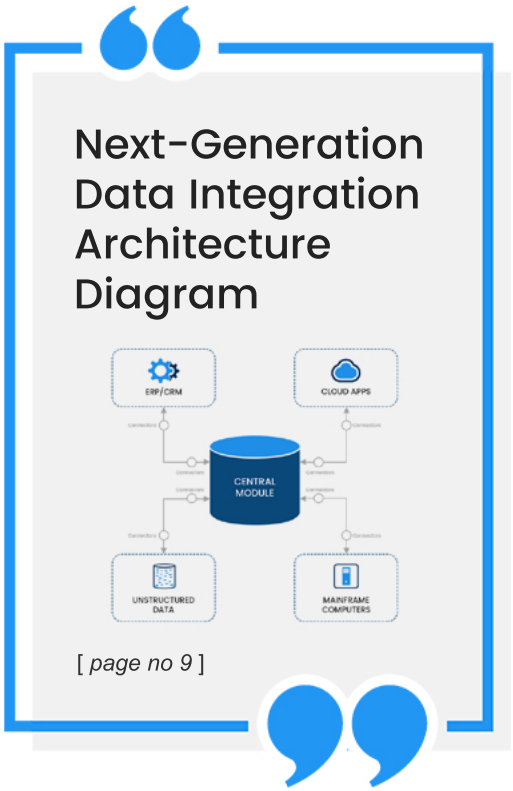


Figure 3: Data Integration

In other words, a traditional data warehouse can be loaded without even moving the data from their sources.

Similarly, while universal data access has been part of previous-generation data integration approaches as well, it assumes greater significance with the advent of newer and usually unstructured data types, platforms, and business use cases. The success of a data integration exercise today is vastly dependent on the ability to harness data from a variety of sources. And with an exponential ri se in the amount of unstructured data, spatial data, and event data that need integration, universal data access helps data integration teams leverage business data from diverse sources in a fast and effective manner.



D) Focus on scalability

At the heart of any data integration project is the realization that data volume, variety, and velocity will grow in the future. Next-generation data integration approach accounts for this growth and makes the transition to a data-driven business easier for a business or enterprise.

In addition to integrating diverse amounts of big data, the next-generation DI approach is also capable of integrating newer technologies with a business's existing systems. In other words, a scalable data integration design not only reduces complexity but saves time and money as well.

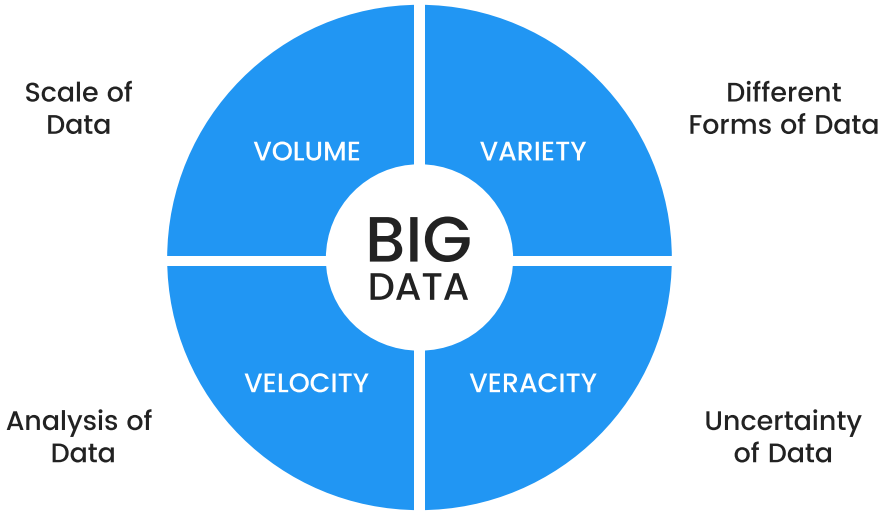


Figure 4: Big Data

What a Next-Generation Data Integration Architecture Looks Like?

A modern-day data integration architecture is more of a process and less of an event. This means it cannot be designed and developed overnight, and instead gradually grows into a productive model as the IT team continuously introduces improvements into the environment. This works to the advantage of business and enterprises looking to adopt a new-generation approach to data integration, but without disrupting their existing business intelligence operations. Instead of replacing the existing data integration architecture altogether, it can be gradually upgraded over a period.

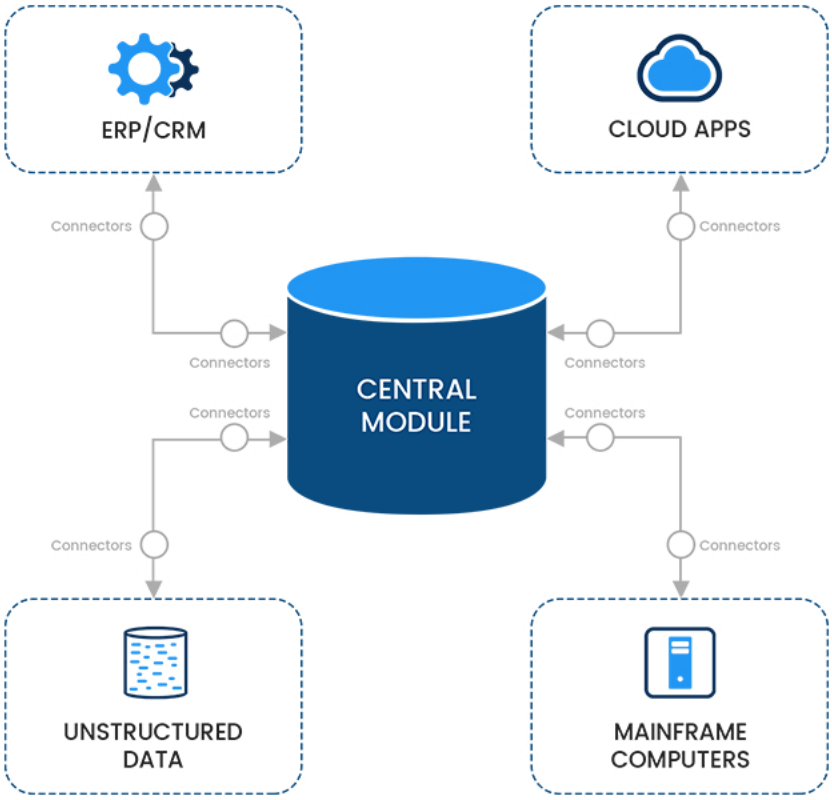


Figure 5: Next-Generation Data Integration Architecture

The figure above outlines the latest generation data integration architecture and its various components.

- The external rectangles represent complex modules that change frequently. Each is representative of a system class that needs to be integrated, like cloud apps, mainframe computers, ERP/CRM, and unstructured data.
- The oval shapes represent connectors, with the one at the external rectangles being source dependent. On the other hand, the oval shape attached to the central module is independent of its respective complex module.
- The central module is designed as per the enterprise standard model and changes less frequently.

Among the highlights of the latest-generation data integration environment is the GUI (graphical user interface)-based drag and drop functionality it offers. As alluded to earlier, this enables DI professionals to build different data moving processes without the need to code. For instance, processes like extraction of raw data from diverse sources, cleansing of that data, and their eventual consolidation into a data warehouse can be accomplished quickly with minimal complexity.

Furthermore, the latest-generation DI environment is built on a VDM (virtual data machine), which ensures DI-related development work remains isolated from the product server. The advantage of a virtual machine is that it allows data engineers to create prototypes without actually moving data from the product server. This further allows them to convert a virtual prototype into a physical ETL, i.e. load a traditional data warehouse, without having to recode the mappings.

A virtual data machine also acts as insura\nce against the fact that data is bound to change and grow, which means different movement technologies might be required for different integration processes. A VDM, as such, can be reused for deploying on physical machines running any kind of technology stack. This ability to reuse data integration logic holds the key to dealing with a fast-changing and unpredictable business environment.

Another cornerstone of latest-generation data integration architecture is DataOps. An automated, process-oriented methodology, DataOps subsumes the roles played by DevOps, Agile and SPC (Statistical Process Control) in modern data integration frameworks.

While DevOps focuses on continuous data integration development, testing, and deployment in an automated environment, Agile concentrates on quicker development in keeping with business goals.

SPC is tasked with uninterrupted monitoring and control of the data pipeline, which ensures DI teams are notified through an automated alert in case of any anomalies in the data. For businesses dependent on data for long-term growth, DataOps has emerged as a critical success factor since it brings software developers, system administrators, and data professionals on the same page.

> Purchasing patterns data on the capacities of machines, models and colours that are preferred in respective regions helped to **improve yield of Digital Marketing campaigns as much by 26%**

Case Study

Problem:

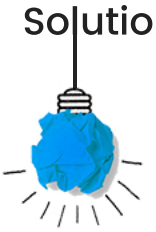
Our client is a USA-based electronic appliances reseller with 900 franchises across 10 countries. While their sales network scaled rapidly and helped to expand market share. The challenge now shifted to perform consistently and sustain profitable growth in the market where competitive forces were putting pressure on the margin.

- Management decided to effectively utilise the sales network by understanding regional consumption pattern & running a selective offer program.
- As an initiative to keep the check on the costs by 5%, they decided to reduce the purchasing decision backed by data.

We started talking with the stakeholders & taking a stock of the Data Practices. We observed that the Purchasing activity was scattered and had various sources like Spreadsheets, Web export in CSV, Excel data & some entries in SAP. Purchase data was not linked with the distribution and computerised reconciliation was next to absent. Less automation also meant data consolidation took too much time and money and had multiple errors. This resulted in a recurring problem of data anomalies.

Due to the lack of quality data, the client was unable to predict demand accurately. This delayed purchase decisions and incentivized gut-based decision-making over data-driven ones.

Solution:



We applied latest-generation data integration techniques to the client's various internal processes. The aim was to create a system that was automated, accurate, and robust. Here is a summary of the solutions.

- We studied the client's business and its pain points by consulting with every stakeholder in the process.
- We designed a data integration system that automated data consolidation and the generation of real-time insights and reports.
- We implemented DataOps to make development, testing, and deployment faster and more accurate. We also applied SPC (Statistical Process Control) to discover data anomalies in real-time through automated alarms.
- We also designed a machine learning-based demand forecasting model to help the client assess product demand accurately.

Results:

Data Integration and Data analysis is the iterative process that progresses with refinement in practices. We encountered procedural and technological hurdles during the process, the client eventually benefitted immensely from the exercise. Here is a summary of the results.

- Our automated alerts system helped the client identify and resolve data anomalies quickly. This further helped them make business-critical decisions with certainty and improved culture & maturity of process.
- Predictive Analysis on purchase data helped to improve Purchasing quantities accurately. This helped to drive country level purchasing contracts and overall helped to reduce the purchasing cost by 3.5% within first 6 months.
- The data modelling on Ordering data & helped us to discover Purchasing patterns on the capacities of machines, models and colours that are preferred in respective regions. This helped the VP Marketing & team to design customised region-wise offers that will gain traction. They cleverly tweaked Advertising campaigns on FB using geo targeting based on Data Analytics inputs. This helped to improve yield of Digital Marketing campaigns as much by 26%

Conclusion

The ever-changing business landscape, along with technological evolutions, globalization, and empowered consumers, is forcing businesses and enterprises to reassess their existing data integration systems and explore new avenues to leverage the business potential of data.

Organizations that have adopted latest-generation data integration principles, technologies, and methodologies are already enjoying rich dividends. Projects are being executed faster, with more quality, and greater cost-efficiency than what was previously possible with traditional data integration practices. As latest-generation data consolidation processes further evolve in terms of principles and technologies, they can help almost any organization achieve its tactical and strategic objectives.

About eLuminous

Your trusted IT Partner for Tech Startups, IT companies & Enterprises. Our expertise is OpenSource, Mobile apps and Data Analytics and BI. With 16+ years 100+ talent pool & CMMI3 practices makes a preferred choice.

We provide solutions that work by accepting challenges & delivering on commitment at a decent cost. Our transparent working models, communication methods, on-time delivery let our clients focus on their competencies by outsourcing to us.

Get In Touch

See what we're all about by visiting our website:

www.eluminoustechnologies.com

sales@eluminoustechnologies.com

+91 8308764279

eLuminous™ TECHNOLOGIES PVT LTD