LTIMindtree

WHITEPAPER

EDW Modernization - A Practical Approach to Implementation

In today's dynamic and data-driven world, staying ahead demands a strategic overhaul of Enterprise Data Warehouses. This paper explores actionable insights and best practices, equipping organizations to navigate the challenges and modernize their data infrastructure. Embrace the future of data with confidence as we delve into the core of EDW transformation.

Introduction

Enterprise Data Warehouse (EDW) modernization breathes new life into business data infrastructure. It does so by unlocking the untapped potential of digitization to empower data-related processes with unparalleled speed and advanced analytics. While data warehouses have long fueled mission-critical decisions, relying on weekly or monthly reports is no longer sufficient for thriving businesses. The imperative now is real-time, actionable insights, which legacy data warehouses struggle to provide, given their inherent lack of agility.

The meteoric growth in data volume, variety, velocity, and veracity makes EDW modernization a business imperative to remain competitive in today's dynamic markets. Embracing this transformation equips organizations with the tools to conquer the challenges of tomorrow, turning data into a strategic advantage for sustainable success.

That said, navigating the ever-expanding data landscape poses challenges that demand innovative solutions for maximizing the potential of data-driven insights. **Key challenges in using traditional data warehouses include:**

Efficient data structure management

As data volume within a legacy EDW expands, structuring it becomes complex. This results in an inefficient extract, transform, and load (ETL) process, making it challenging for system managers to qualify data for advanced analytics. A well-designed and configured warehouse is essential to derive true value from data, streamlining the organization of details for diverse processes and functions.

Meeting user expectations

Business users expect refined results from data analysis. However, the bulk of information in a traditional EDW makes it difficult for the management system to delve into, resulting in slow and potentially inaccurate outcomes. The solution lies in modernized, self-managed data warehouses that empower technical and non-technical users to master data analysis. Cloud-based EDWs, in particular, offer point-and-click solutions, enabling seamless data warehouse operation without disrupting workflows, thus meeting user expectations effectively.

Enhancing data quality

In traditional EDW, manual administration leads to human errors and missed updates, resulting in corrupt or outdated information. Consequently, data processing becomes unreliable, leading to suboptimal business decisions. Moreover, duplicate data copies for different use cases, aggregated analyses, and individual process dashboards further complicate matters, creating governance, security, and compliance concerns within complex legacy architectures.

Addressing high costs

Businesses need agile and adaptable IT environments to keep up with their digital transformation strategies. Legacy on-premise data warehouses lack such attributes. Their rigid architecture makes introducing even minor changes in data models for new reporting requirements a highly complicated process. The technical challenges and limited scalability of such EDWs drive up costs and timelines and lead to project failures. In contrast, a cloud-based modernized data warehouse is a game-changer, significantly reducing the total cost of ownership (TCO). Providing a more accommodating architecture eliminates the expenses associated with hardware maintenance and ongoing systems engineering. This modernization ensures that EDWs can seamlessly scale storage, possess stateless and resilient computational power, enable scale-invariant data sharing, and uphold compliant storage operations, effectively meeting diverse business requirements.

Modernizing the data landscape to stay competitive with data-driven business decisions

The significance of data in shaping business decisions cannot be overstated. Through data collected via sensors, surveys, and social media, businesses gain actionable insights and the capacity to optimize performance across multiple dimensions. By leveraging this information, they can test the success rate of their strategies and drive sustainable growth. Real-time updates to data warehouses enable organizations to support new initiatives and adapt to evolving business needs. Furthermore, such updated data allows them to effectively demonstrate the outcomes of their decision-making to stakeholders, reinforcing transparency and informed decision-making across the organization.

Data-based decisions empower organizations to gain valuable insights across various functions and departments by cultivating a data-driven culture and fostering an ecosystem where employees continually learn and innovate for enhanced productivity. By sharing data insights, business units can effectively communicate measurable results, facilitating seamless collaboration and optimizing revenue-generation processes intelligently.

Moreover, in a constantly evolving digital world, leveraging data is crucial to making informed choices based on industry trends, ensuring sustained competitiveness and relevance for the organization. With data as the cornerstone of strategic planning and operational efficiency, organizations can confidently navigate the dynamic landscape, achieving success in a data-rich era.

Modernizing EDW to achieve scalability, faster insights, and cost optimization

EDW modernization is the first step for a business that aims for the benefits of data-driven decisions. Aging legacy data warehouses cannot support the rising business requirements, making introducing new KPIs for reports difficult. They are incompatible with the emerging self-service business intelligence (BI) and analytics tools. Their ETL pipelines running on a fixed schedule do not deliver real-time insights.

Migrating traditional data warehouses to the cloud is a compelling solution, effectively addressing these issues and unlocking several benefits. Cloud-based modernization accelerates speed-to-value, empowering businesses with serverless analytics and fortified data security, instilling more trust in managing operations. Moreover, the cloud offers unprecedented scalability, allowing organizations to tailor their data infrastructure to match evolving business needs. The rapid expansion of the data warehouse is met with high network speeds, ensuring efficient data processing and enabling the application of advanced analytics, yielding actionable insights.

With cloud-based EDW modernization, organizations can scale their data infrastructure based on evolving business needs. Even as the data warehouse expands, high network speeds allow organizations to continue processing data faster and apply advanced analytics for actionable insights.

Cost optimization is another widely proven benefit of transitioning to the cloud. EDW hosted on the cloud helps to move from capital budgeting to need-based forecasting, delivering business results with increased savings.

Trends show that businesses are already understanding the benefits of modernizing EDWs. According to Gartner, 35% of data center infrastructure will be managed from a cloud-based control panel by 2027, a significant surge from the less than 10% observed in 2022.

Why adopt GCP for data warehouse modernization

Google Cloud Platform (GCP) is a leading solution for EDW modernization. As a feature-rich cloud data platform, it enables scalability, enhanced security, and cost optimization, cultivating innovation and agility in data-driven decisions for operational excellence. It expedites an organization's digital transformation with data democratization, app and infrastructure revitalization, strong connections between people, and secure transactions.

When a company chooses GCP for its EDW modernization, it can expect the following benefits:

Intelligent decisions

Google Cloud offers built-in AI capabilities, enabling speed, scale, and security. It breaks down data and process silos, facilitating engaging user experiences, deriving global-scale insights, and transforming data into real-time decisions, bolstering business growth.

Better team collaboration

GCP integrates collaboration channels like email, chat, video conferencing, and document sharing into a unified platform. The Google Workspace segment employs a zero-trust approach for enterprise-grade access management, data encryption, and endpoint protection.

Building and running apps anywhere

GCP empowers businesses to avoid vendor lock-in, accelerate digitization, and embrace a multi-cloud strategy with an evolved open-source cloud infrastructure. Users can run workloads in the most suitable locations, leveraging a global network that provides exceptional throughput at lower administration costs than other cloud services.

Advanced security systems

Through the same zero-trust model, GCP also secures business data and apps against online risks and fraudulent access to the data warehouse. Instead of burdening customers with cybersecurity responsibilities, the cloud platform adopts secure-by-default configurations, safe blueprints, policy hierarchies, and consistently available advanced security features, offering high assurance of controls.

Seamless scaling

GCP can be seamlessly scaled up or down to accommodate changing business requirements. This ability also makes it cost-effective as customers have a pay-as-you-go model for their cloud platform subscription.

Pitfalls and challenges in EDW modernization

As businesses increasingly recognize the benefits of EDW modernization, they must approach migration with a comprehensive plan, as transitioning from a traditional to a cloud platform can be complex. **The common challenges faced in executing the move include:**

Poor governance

An effective data warehouse requires comprehensive data coverage from all endpoints and functions to eliminate system vulnerabilities. Without a robust governance model or proper adherence to compliance guidelines, the migration of the EDW can be riddled with issues and potential pitfalls.

Access control

A well-defined data strategy framework for access control is imperative during data warehouse migration. Organizations often face challenges in correctly identifying personnel or departments requiring access to specific information and applications in a modernized EDW. Ineffectively managed resources and inaccurate permissions can place an unnecessarily heavy load on the new system.

Weak data strategy

Insufficient focus on data strategy can impede an organization's ability to fully realize and harness the advantages of a modernized data warehouse. Engaging key stakeholders in the planning process is crucial to effectively strategize the migration of relevant data from one platform to another. This can help ensure a well-thought-out and cohesive data strategy, maximizing the value of their modernized data warehouse.

Business disruption

The lift-and-shift process of moving a data warehouse comes with inherent risks of disrupting day-to-day operations, causing downtime, or data loss. These concerns arise due to inadequate documentation, a lack of in-house expertise, highly technical and complex systems, and an incomplete understanding of existing code logic.

Spiraling costs

While modernizing their data warehouse, enterprises are also concerned about preserving their investments and the years of effort put into writing business logic. Failure to reuse strategic legacy workloads, including analytics, DML, ETL, orchestrator, and reporting scripts, will only make the migration process more expensive. Additionally, design and code defects stacked in legacy systems may snowball into technical debt that causes several operational issues. To mitigate these risks and avoid technical debt, modularizing the architecture becomes essential.

The right approach and successive stages for completing EDW modernization

The seemingly complicated process of data warehouse modernization can be successfully handled with careful planning and execution. **The steps to update the existing infrastructure and migrate to a more scalable and agile environment are:**

Step 1: Assessment

The initial step in the EDW modernization journey involves analyzing the existing EDW environment and identifying potential areas for improvement. This comprehensive assessment includes the architecture, data structure, custom software, and integration processes. Employing data profiling, lineage, and mapping techniques provides a detailed view of the data warehouse assets and facilitates a thorough understanding of their dependencies. This crucial step lays the foundation for informed decision-making throughout the modernization process.

Step 2: Defining goals

After identifying the assets earmarked for migration, it becomes essential to pinpoint their specific pain points that require addressing. Areas such as data quality and governance issues, slow query performance, weak analytics capabilities, application security concerns, and challenges in integrating new data sources, should be given due attention. By defining these goals, the organization can lay a clear roadmap for tackling critical issues and ensuring a successful EDW modernization journey.

Step 3: Migration planning

This step helps prioritize use cases and estimate a realistic migration timeline within the budget. The EDW migration team must diligently select appropriate technologies for ETL, data storage, integration, modeling, and analytics within the chosen cloud platform that will host the data. Thoughtful planning at this stage ensures a smooth and efficient migration process, laying the groundwork for a successful transition to the modernized EDW.

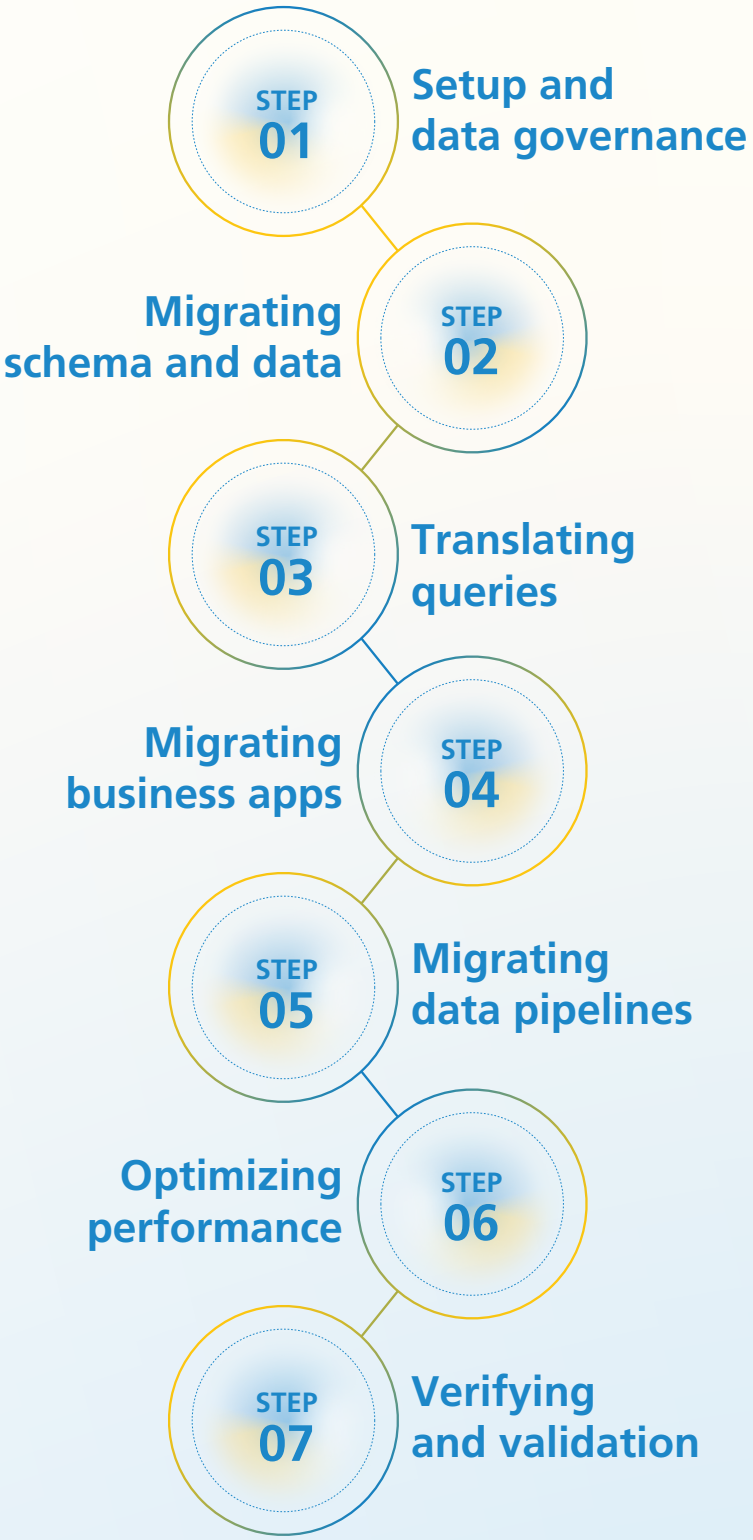
It is crucial to get stakeholders' buy-in and keep them involved through all these stages of transitioning to a new EDW. A CTO or another business leader is critical in approving the migration plan and allocating necessary resources. A project lead takes charge of guiding execution and monitoring milestones while managing the project scope. Also, there should be technical leads to look into specific technical areas for database migration. The stakeholders would also include subject matter experts from different business departments whose functions would be impacted by the modernization. Ensuring strong engagement and collaboration among these stakeholders fosters successful EDW modernization, aligning business objectives and technical requirements for a seamless transition.

Secure EDW migration: Establishing the cloud foundation

The eventual process of shifting data from the legacy data warehouse to a new environment involves updating the schema (that defines how data is structured), translating queries to the new data query language, verifying data quality, and migrating data pipelines and applications. Collaborating with a proven and experienced data migration partner is highly beneficial, as they can efficiently allocate workloads and contribute to improved outcomes during the data warehouse migration process.

Opting for GCP as the platform for EDW modernization introduces businesses to the advanced capabilities of BigQuery, a sophisticated SaaS technology that facilitates serverless data warehouse operations. This powerful tool offers seamless access to structured data storage, processing, and analytics, boasting scalability, flexibility, and cost-effectiveness. These essential features ensure that storage and processing resources are readily available as needed, allowing organizations to extract maximum value from their data. By harnessing the potential of BigQuery, businesses can optimize their data operations, making data-driven decisions with efficiency and precision.

BigQuery also helps to apply machine learning (ML) algorithms in understanding emerging data patterns and to test new hypotheses. As a result, the business gets timely insights into its performance and modifies processes for better results. **The key steps for migration to GCP include:**



Step 1: Setup and data governance

Setup is the foundational groundwork that enables the smooth execution of use cases on GCP. It includes essential configurations such as projects, network setup, and virtual private cloud establishment. Similarly, data governance plays a pivotal role in outlining comprehensive policies, procedures, responsibilities, and controls governing data management throughout its lifecycle, from acquisition to utilization and eventual disposal.

Step 2: Migrating schema and data

This crucial phase focuses on data structuring and establishing relationships between data entities. The schema forms the core of data design, significantly influencing various upstream and downstream processes.

Step 3: Translating queries

During migration, the organization can opt for batch SQL translation to migrate SQL code in bulk or leverage interactive SQL translation to handle ad hoc queries efficiently.

Step 4: Migrating business apps

The migration to GCP encompasses a wide range of business applications, including dashboards, custom apps, operational data pipelines that facilitate feedback loops, and transactional systems.

Step 5: Migrating data pipelines

Efficiently moving legacy data pipelines to GCP becomes a streamlined process when relying on comprehensive data pipeline documentation. This documentation specifies the necessary procedures, available options, and technologies for executing large-scale data warehouse migrations.

Step 6: Optimizing performance

BigQuery offers unparalleled processing capabilities, seamlessly accommodating datasets of all sizes. With its robust support for data analytics tasks, organizations can achieve optimal performance in their modernized, cloud-based EDW without requiring extensive modifications.

Step 7: Verifying and validation

Throughout each iteration, the success of use-case migration is crucially validated by thoroughly verifying the following key aspects:

Complete migration of schema and data

Proper functioning of data governance

Seamless functionality of migrated data pipelines

Accurate translation of queries

Establishment of automated maintenance and monitoring procedures

Successful access to migrated data and queries by business applications

Addressing data governance and security

In a cloud-based EDW, data governance ensures that information is collected, maintained, used, and disseminated in alignment with the organization's data integrity and security requirements. **To build a robust data governance practice within GCP, organizations can focus on three essential practices:**

01 A framework enabling stakeholders to define, agree to, and enforce data policies

02 Processes for control, oversight, and stewardship over all data asset

03 Tools and technologies to oversee and manage data policy compliance

Google further reinforces data governance and security with a framework built on the principles of zero-trust security. **This dynamic model enforces strict access controls, requiring users or devices attempting to access resources to:**

01 Authenticate itself with multi-factor authentication

02 Be authorized to access the said resource

03 Communicate using encryption

Ensuring observability and monitoring

Effective monitoring of the data warehouse environment is vital for high-performing business teams' success. **As per definitions given by DevOps Research and Assessment (DORA):**

> *"Monitoring is tooling or a technical solution that allows teams to watch and understand the state of their systems."*
>
> *"Observability is tooling or a technical solution that allows teams to actively debug their system."*

The best practices for effective monitoring and observability include:

- Reporting on overall system health, ensuring proper functioning and availability of resources
- Reporting on system state in line with customer experience
- Tracking key business and systems metrics
- Tooling to comprehend and debug systems in production
- Tooling to identify the previously unknown attributes of the data warehouse
- Leveraging tools and data to diagnose infrastructure bottlenecks in the production environment, including interactions between services

Conclusion

This paper underscores the limitations of traditional EDWs in meeting the demands of digitally transforming businesses. Their inability to manage data structuring and quality hampers crucial business functions, while the lack of support from legacy technologies prevents the automation of repetitive processes. Moreover, these outdated systems don't offer the robust security to combat modern cyber threats.

For enterprises aiming to consolidate siloed data, enable custom reporting, conduct ad-hoc analysis, and drive strategic business growth, the time to innovate their EDW is now. Embracing a modernized data warehouse empowers organizations to thrive in the dynamic landscape of data-driven opportunities.

About the authors

Harshavardhan Jb

Associate Director - Program & Project Management | Data & Analytics

Has over 17 years of experience in Data and Analytics Consulting, crafting solutions to enable Self Service Analytics and data solutions. Agile evangelist proficient in transforming business from traditional approach to Agile way of working. Has performed various roles both in technology and program management. He is specialized in EDW modernizations enabling customers move from on-premise data ecosystem to cloud native systems. He is currently working as a GCP Solution Architect and Technology consultant for LTIMindtree.

Rajat Goswami

Senior Specialist – Architecture | Data & Analytics

With over 12 years of experience in the IT industry across product management & consulting in critical business applications to Data Analytics and AI/ML, Rajat has a unique blend of experience that lets him align the right technology solutions to the business problems of today.

A technology enthusiast at heart, and an alumnus of IIM Calcutta, he has conceptualized and developed AI enabled business products, and lead complex Data Analytics and AI/ML client engagements from initial stages to delivering value for business.

Currently, he is working as a Data & Analytics consultant at LTIMindtree, driving GTM strategies and solutions for the D&A practice.

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700 clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by 83,000+ talented and entrepreneurial professionals across more than 30 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen and Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit **www.ltimindtree.com.**

LTIMindtree Limited is a subsidiary of Larsen & Toubro Limited