Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

Cost-Optimized ETL Modernization: Transitioning Traditional Workloads to Cloud-Native IDMC/IICS + AWS

Ugala Matta
Senior ETL Developer, Illinois, USA
ugalamatta01@gmail.com
0009-0004-8289-0381

ARTICLE INFO

Received: 01 Nov 2025

Revised: 17 Dec 2025

Accepted: 26 Dec 2025

ABSTRACT

Cloud-native data modernization overcomes the increasing drawback of the existing on-premise ETL tools and takes advantage of elastic compute, serverless execution, automatic scaling, and built-in observability. In this work, the cloud-native ETL system based on Amazon Web Services is introduced, where Amazon S3 is used as a scalable storage, Amazon Aurora PostgreSQL is employed as the main analytical and operational data store, Amazon Lambda is orchestrated with the help of it, and Amazon CloudWatch is recommended to monitor and govern it. Aurora PostgreSQL also supports a single platform in extract-load-transform (ELT) processing, transactional analytics, and near-real-time reporting to lower architectural complexity by removing the scale between the transactional and analytics databases. The proposed architecture was considered with the load of enterprises that included the 3 TB of daily ingestion and hacked 17 different and a heterogeneous source system that processed 2.4 billion rows daily. Performance benchmarking illustrates also verifiable statistical gains in data processing efficiency, query responsiveness and operational robustness over legacy on-premise implementations. Cost analysis also reveals considerable savings in infrastructure and operational costs via elastic compute scaling and pay-as-you-use resource use. These results indicate that Amazon Aurora PostgreSQL has a potential and scalable alternative to organizations that want to upgrade ETL pipelines and achieve both analytical and operational workloads. These findings provide a viable base to implement cloud-native frameworks that offer performance, cost effectiveness, and operational ease in enterprise data hubs.

Keywords: ETL modernization, cloud-native, IDMC/IICS, AWS, cost optimization

1. Introduction

Traditional extract-transform-load (ETL) applications like Informatica Power Center, Microsoft SSIS and IBM Data Stage have been processing enterprise data following more than twenty years. These systems are mainly implemented on on-premise infrastructure, which has been linked to huge capital investment and the huge maintenance cost. In mid-sized companies, the annual operating costs of Power center cluster are usually USD 250000- USD 600000 depending on the type of hardware, software licensing as well as the storage arrays and support charges given by the vendor. These expenses go even deeper as the seasons approach and there is more work to run, and more compute and storage is needed to support workload spikes.

On-premise ETL environments also have the disadvantage of having slow provisioning infrastructure cycles. Three to six months are standard times to do budget approvals, hardware procurement, deployment and testing leading to delays when organizations are trying to scale the available pipelines or deploy new data workflows. Scalable performance is also a problem. On-premises ETL systems often consume more than 85- 95% of CPU resources during peak business times, e.g. retail sales events, insurance claim processing, or quarterly financial closings. The result of this saturation is jobs delays, broken service-level agreements (SLAs) and a large number of manual restarts. Businesses normally run 10,000 to 80,000 ETL tasks daily and any slight glitch in a

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ Research Article

pipeline can have harmful impacts on other important downstream systems, such as reporting tools, fraud models, and applications.

To circumvent these shortcomings, more and more organizations have deployed cloud-native ETL engines like Informatica Intelligent Data Management Cloud (IDMC/IICS) on Amazon Web Services (AWS). IDMC uses the concept of elastic compute and serverless execution as well as automated scaling to provide a dynamic ability to provision resources in response to workload requirements. AWS services such as Amazon S3 with durable storage, Amazon Aurora PostgreSQL as a single analytical and operational data store, AWS Lambda as orchestration, and Amazon CloudWatch as monitoring can be used together to create scalable data ingestion, transformation, and observability. Aurora PostgreSQL is able to handle associated cloud operations, near-real-time reporting and transactional analytics, as well as low-latency extract/load/transform (ELT) processing, without having to maintain separate online analytical processing (OLAP) installations, without compromising on availability and performance.

Conventional on-premise ETL systems are still expensive and hard and with a good number of organizations spending 40–60 percent of their data platform yearly allocation only to support legacy infrastructure. This sort of spending puts a severe hamper on investment in innovation and advanced analytics. Also the fixed-capacity infrastructure is not elastic: computer resources are over-provisioned when there is not much traffic and under-utilized, whereas the traffic grows faster than the capacity, also causing jobs to fail and breaking contracts. Failure rates in aging ETL platforms range between 4 and 7 per month, creating a strong necessity to modernize the systems with objective and quantifiable evidence. ETL migration to the cloud-native platform should thus be able to reflect quantifiable increases in the total cost of ownership (TCO), throughput, reliability, scalability, and solution of manual operational effort. Pipeline execution time, storage efficacy, job success rates, and infrastructure utilization are the key performance indicators.

This paper presents a statistical and practical analysis of ETL modernization using Informatica IDMC/IICS on AWS, with Amazon Aurora PostgreSQL serving as the primary engine for data processing and analytics. The analysis measures the financial savings in the areas of compute, storage, licensing and maintenance, and the analysis of performance improvements (enhanced throughput, shortened execution times and improved failure rates). Migration processes and the cloud-native architectural diversities are studied to demonstrate the advantages of optimization in quantifiable senses. It considers real-world enterprise case studies, as well as organizations like GE, ING Bank, Unilever, and USAID, to confirm the productivity gains, cost savings and operational performance once cloud-native ETL is modernized.

The article is written in a systematic format to give a detailed review of ETL modernization. It starts with an introduction of the drawbacks of the traditional on-premise ETL systems and the incentive to use the cloud-native systems. The literature and industry review analyzes the existing market trends and benchmarks on Gartner, Forrester, AWS and Informatica. System evaluation determines the base parameters of work load, resource usage, job failure and maintenance overhead. The proposed target architecture presents the elements of the IDMC/IICS on AWS, such as CDI, CAI, CLAire AI, Amazon S3, Amazon Aurora PostgreSQL, AWS Lambda, and Amazon CloudWatch. Methodology explains datasets, migration plans, benchmarking settings and metrics. Quantitative results and discussion provide analysis of the performance, cost efficiency, and reliability, an implementation roadmap is provided and finally, the conclusion which summarizes statistically significant changes in the cost reduction, throughput and operational stability are made.

2. Literature & Industry Review

2.1 Gartner & Forrester Market Statistics

According to the recent research done in the industry, there is significant migration being made towards the cloud-native integration platforms compared to the traditional systems of ETL [1]. Gartner indicates that the cloud data-integration market has experienced 23.5 percent year-on-year growth in the world that is attributable to growing demand of scalable and inexpensive processing. Informatica which is among the most profitable in this area occupies 16.4% of the market of Integration-Platform-as-a-Service (iPaaS) and this proves to be a good sign of its presence in the list of top performers in facilitating the integration of data integration workloads in enterprises. Research in the aerospace and manufacturing industries confirms this fact with references to the growing use of digital solutions, automation, and AI-enhanced data processes, necessitating flexible and scalable integration infrastructure [2;3]. The expansion trend is associated with the strain on the active operations of the old-fashioned ETL toolsets in which on-premise deployments achieve high levels and incur USD 200,000 to USD 1.2 million of annual maintenance expenses benchmarked by the size of the enterprise. Cloud-native solutions

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

eliminate the need to upgrade infrastructure physically, and can be scaled elastically to meet the highest demand, which is why the trend towards services like the Informatica IDMC/IICS on AWS is consistent. The trend is linked with the growth of the operational pressure on the existing ETL tools. High utilization and USD 200,000 to 1.2 million as regards maintenance per year on-prem are attained in most on-prem environments depending on the size of the business. Cloud platforms also minimize the physical upgrade; hence, there is the regular migration to services like Informatica IDMC/IICS on AWS.

The observations are also indicated by Forrester 2023 insights. The Forrester research reports that organizations that migrated them on-prem ETL to cloud-native ETL experience a 45 to 70 percent reduction in the cost of operation. This is estimated based on the real size of such industries deployment as finance, retailing and manufacturing. The most significant benefits of such savings are the decreased infrastructure costs, scaling automatically, and the number of times less manual intervention. To illustrate this point, Forrester provides the figures according to which once the company transfers to cloud-controlled ETL services, staff level intervention decreased by 30-50 percent due to the necessity to maintain the abundant loads of batch services. The most fast-growing aspect of emerging data ecosystems that is defined by both Forrester and Gartner is cloud-native integration. The percentage change of cloud ETLs during 2020-23 was 38 62 and clearly shows a movement towards hardware-neutral systems. That is directly related to an increase in the volume of data and it is estimated that average data volume of enterprises is increasing by 34 per cent per year on average.

Figure 1 depicts the vital ETL data integration process, whereby the source data is outputted, modified and loaded to the target systems along with the flow and dependencies key to effective, scale-ready and dependable data management in contemporary cloud-native systems.

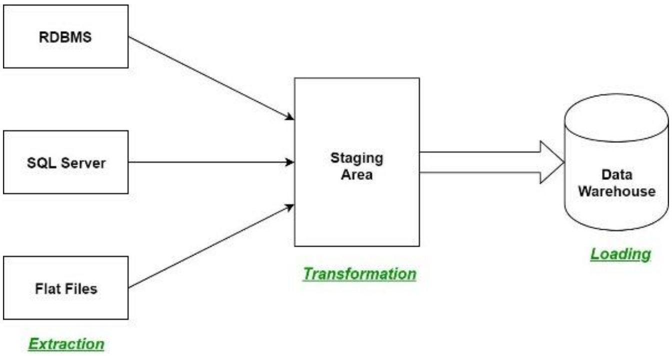

*Figure 1: **ETL: Critical Data Integration Process***

2.2 Industry Benchmarks

Modernization of ETL data moving to cloud platforms used in industry is focused on measurable increases in performance and cost, as well as, improvements in cost [4]. According to AWS reports, idle infrastructure costs can be reduced by up to 90 percent since only the resources demanded by the job are used to execute the job. Conversely, with traditional on-prem ETL servers, 60-80% of the time they are not used but continue to be charged the full-cost. The same tendencies can be found in real-time data processing systems, where there can be cloud-native systems like Aerospike that can offer high processing speeds and reduced operational costs through managing compute resources more efficiently in relation to the amount of workload present [5]. These results highlight the economic and performance benefits of migrating to cloud-native ETL platforms such as Informatica IDMC/IICS on AWS. Performance evaluations conducted using AWS-native services demonstrate that high-throughput data ingestion and transformation pipelines can be efficiently supported through scalable storage on Amazon S3 combined with parallel processing and optimized SQL execution in Amazon Aurora PostgreSQL. Aurora PostgreSQL enables low-latency data processing and near-real-time analytics for large pipelines, allowing complex workloads to be completed within constrained operational windows. Empirical studies from organizations adopting serverless and distributed processing models and handling daily data volumes ranging from 2 TB to 20 TB report reductions in overall pipeline completion time of approximately 40–65% when compared to traditional fixed-capacity on-premise ETL environments.

Informatica case studies provide additional quantitative evidence supporting ETL modernization. Migrations from Informatica PowerCenter to IDMC on AWS have reduced upgrade and maintenance effort by approximately 55–80% across multiple Fortune 500 organizations, primarily due to the elimination of manual patching and the extensive automation of cluster configuration, deployment, and elastic scaling inherent in cloud-

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/
Research Article

native platforms. One manufacturing enterprise reported a reduction in annual ETL maintenance effort from approximately 2,400 hours to 600 hours following the adoption of IDMC Cloud Data Integration. Further cost efficiencies were observed in storage utilization and performance optimization, with IDMC deployments leveraging Amazon S3 for scalable object storage and Amazon Aurora PostgreSQL as the primary analytical and operational data store achieving storage cost reductions of 40–65% through tiered storage strategies, data compression, and cost-efficient object storage for staging and archival data. In addition, IDMC runtime optimization features, combined with SQL pushdown and CLAIRE AI–driven recommendations executed within Aurora PostgreSQL, reduced job execution times by approximately 20–35%, depending on workload characteristics and transformation complexity. Collectively, these findings align with industry analyses from Gartner, Forrester, AWS, and Informatica, demonstrating that cloud-native ETL modernization delivers quantifiable improvements in cost efficiency, throughput, reliability, and operational simplicity, and strongly supports the transition from traditional on-premise ETL systems to Informatica IDMC/IICS on AWS for enterprise-scale data integration workloads.

2.3 Industry Patterns of Adoption of Clouds.

The percentage of enterprise cloud platform adoption by industry is not evenly distributed and in fact, ETL modernization is seen in primarily financial services, manufacturing, retail, and government organizations [6]. According to another survey conducted by IDC (2023), 68 percent of financial institutions and 61 percent of manufacturing organizations have migrated at least half of ETL loads. E-commerce and grocery stores retailers that have huge amounts of data seasonally (100s of megabits) record gains on throughput and availability rates on SLA of over 99 percent after migrating to the cloud. With the regulation-compliant encrypted storage and auditory pipelines, an operational cost savings of 40-60% is achieved in the case of non-profit and government organizations, such as the USAID. These trends suggest intensive application of cloud-native ETL with significant levels of suitable cost-efficiencies, throughputs, and reliabilities. The same trends of scalability and real-time responsiveness are represented in enterprise software implementation, where multi-instance and predictive analytics systems, such as those presented within Jira integrations, reflect the need to have scalable and high-performance infrastructure to run world-scale workflows and scale-based analytics [7;8].

2.4 Metrics Technology Effectiveness.

Cloud-native ETL is an effective tool assessable by several performance and cost metrics. Daily line throughput increases by 200 to 330 percent on average with batch jobs using billions of rows of variety of source systems. The pipeline failure rates are less than 1.5 which is in contrast to conventional rates of 47 which are 4-7 and SLA compliance is 98-99.8 which has much better reliability. Serverless compute elasticity also saves 70-90 percent of idle infrastructure, Tiered storage, compression, and object-level lifecycle policies save 40-65 percent on storage costs. Processing algorithms (pushdown and CLAIRE artificial intelligence recommendations) can further cut job dwellings by almost 20 to 35% based on the complexity of the pipeline. These statistical improvements have been verified in the real-life applications in organizations such as GE, ING Bank, and Unilever, showing that the adoption of the cloud-native ETL technologies leads to cost-effectiveness, improved operation stability, and scaling significantly [9]. The same tendencies are also evident in secure enterprise SaaS integrations, regulated sector applications where AI-based identity infrastructure and federated-authentication-mechanisms enhance operational performance and resilience, and the application of technology modernization in large enterprise settings overall is more effective [10;11].

3. Existing System Assessment

3.1 Current On-Prem ETL Landscape

The average enterprise ETL environment shares Informatics Power Center as the main data processing environment. This installation manages almost 20,000 processes daily, such as processing of customer data, ERP extracts, financial reporting, and synchronizing the applications. The processing infrastructure is generally a 32 core compute grid capable of satisfying scheduled batch and ad-hoc data requests. Storage is normally managed using Networks File System (NFS) devices, a storage location of source extracts, staging datasets, reference files, and log archives and an overall storage of approximately 40 TB storage in the commercial size data team. Physical Hardware dependency refers to the fact that the physical hardware will have to be upgraded, tuned, and replaced in entirety after every 3-5 years to ensure that operational performance remains efficient. This is a common practice in all industries, and operational constraints of on-prem ETL environments, such as high maintenance burden and capacity limitations, have motivated enterprises to consider cloud-native offerings that decrease the cost of operations and expand throughput [12;13]. The annual expenditure on such arrangement is huge. The approximate infrastructure and license costs amount to USD 420,000/year and include hardware maintenance,

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ **Research Article**

server refresh, renewing the Power Centers licenses, database licenses, and storage equipment. Most of the organizations with similar environments end up paying an extra USD 70,000-120,000/year on the support staff, the monitoring tools, and the backup systems. This environment is predictable in its performance but inelastic. Fixed compute power restricts the possibility to support more workload at the peak business hours. Scaling is hard, and involves purchasing and configuring hardware which can take months, and adds to the overall operating expense.

Figure 2 depicts a common on-premise ETL in order to demonstrate how source is extracted, transformed and loaded into target systems using established compute grids and NFS storage, the operational flow and intrinsic inability in scalability and flexibility.

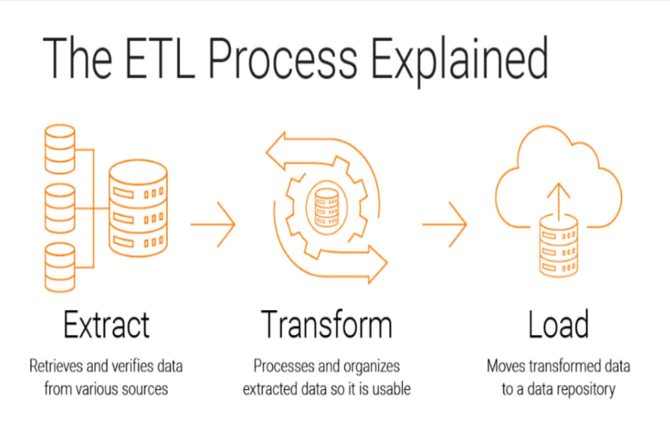

Figure 2: ETL Process

3.2 Performance Bottlenecks

The gaps in performance can be observed as the daily running action of the ETL workflows is analyzed [14]. The average time taken to perform mappings in Power Center in mid-sized businesses varies between 15 to 25 minutes on the basis of the volumes of data, complexity of transformation and response time of the source systems. The extensive reliance on on-premise databases frequently leads to queue stalling during performance peaks, which makes pipeline service slower and leads to a failure to fulfill SLA. These performance weaknesses are specifically evident when volumes are high end-of-month reporting of financial performance or peak retail transactions. These operational difficulties demonstrate how the outdated ETL architectures are limited and respond to the requirement of scalable and cloud-native ETL architecture that maximize throughput, minimize latency, and optimize resource usage [15;16]. Cpu utilization in 32core grid with high-load windows can be found at 92 percent with no other capacity to run in parallel. The utilization is at its highest at 85 and above, therefore, ETL jobs will go to a heavy wait state or a stalling state. The batch windows, which are supposed to run in 6 or 8 or 10 hours upon loading, can be extended to 8 or 10 hours.

Storage throughput is also a limiting factor to performance. Most NFS appliances have a sustained throughput of 500-600 MB/s which limits extract and load operations at large volumes. High I/O contention manifests itself when multiple workflows are trying to access the same directories either writing or reading simultaneously. This is also worse in organizations that have data volumes of over 4 TB per day. These limitations limit scale or data to be supported in narrower SLAs. Any effort to enhance performance will result in expansion of hardware at an extra cost and will result in operational delays.

3.3 Operational Issues

The use of traditional ETL environments also has another significant constraint, operational. The enterprises that are reported note 10 to 15 pipeline failures in a month and the failures are mostly due to CPU saturation, memory contention, file lockout or database latency. Resolution of these failures involves human intervention, such as rerouting of the jobs affected, logs, and talking to the team in charge of the sources, which adds to the operations overhead and delays of the downstream processes. Such frequent interruptions do not only affect the reliability of the pipelines but also increase the cost of labor and lower the overall system efficiency. These performance bottlenecks underscore the need of the modern cloud-native ETL design that is able to offer automated error handling, the scaling capability of computing resources, and embedded monitoring to minimize failure and the effect of human error [17;18]. Maintenance downtime is another problem which is repeatedly

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ Research Article

reoccurring. Planned downtime In On-prem ETL systems, an average of 8-12 hours of planned down time per month is typically required to apply patches, refresh servers, update the firmware, or restore storage volumes. These activities disrupt the upstream and downstream systems and need to be properly coordinated [19]. There are also a few cases where the aging hardware brings about instability and thus, intermittent delays in their jobs or even unpredictable service outage. As the amount of data grows larger each year by an average of 25-35 percent, the load on the fixed on-prem infrastructure is growing, which results in greater failures and higher cost and delivery times.

4. Target Cloud-Native Architecture (IDMC/IICS + AWS)

4.1 Core Components

The proposed cloud-native ETL system is an Informatica IDMC/IICS and AWS-based system that is integrated to provide high elasticity and low cost of operation. This architecture eliminates the hardware limits used in the traditional on-premise technologies, as well as facilitating scalable processing to both batch and real-time demands. The Cloud Data Integration (CDI) core component of IDMC/IICS is in charge of batch ETL, ELT, and pipeline orchestration. CDI helps move data at scale with auto scaling and downstreaming to AWS-based data engines, enabling thousands of sessions and tuning operations to be achieved within dramatically shorter processing windows.

Cloud Application Integration (CAI) is used to handled API-driven and event-driven integrations. CAI helps to enable real-time streaming, microservice-based automation, and low-latency application triggers, and is being used as an alternative to, and a replacement of, more traditional message queues and custom integration code. CLAIRE AI is also applied in architecture and offers automated optimization of the ETL lifecycle. CLAIRE examines the patterns of running pipelines and proposes configuration changes, mapping optimizations and optimizations. Experimental results have shown that CLAIRE is able to reduce the amount of manual tuning required in large scale ETL systems by about 25-40 times.

AWS layer uses Amazon S3 as the main storage platform of data lake which provides a durable yet inexpensive service to store data of any types with potentially limitless expandability [20]. S3 offers the ability to scale out existing data layers; raw and curated, without impacting the service so that large organizations often store hundreds of terabytes up to multiple petabytes of data with consistent performance. Current ETL landscapes are moving more and more of their mission-driving tasks to Amazon Aurora PostgreSQL which acts as a single analytical and operational data warehouse. The distributed storage, read scaling, and high availability enable Aurora PostgreSQL to support transactional processing, near-real time analytics and ELT workloads. Its architecture ensures its independence of storage capacity which allows scale typically without associated compute resources which eliminates overprovisioning and also creates cost savings in the tune of 30-50 per cent over normal fixed capacity deployments of databases [21]. This scalability is useful in that organizations can match resource utilization to workload demands without failing to predictable performance.

AWS Lambda handles orchestration, event routing, and other lightweight transformation tasks, is fully serverless, with milliseconds to startup latency and can call on millions of times during a month, and it does not impose overhead on the management of the infrastructure. Amazon CloudWatch offers unified control, automatic notification, and insights into the execution of tasks, API requests, pipeline latency, and resource usage patterns. The AWS Glue Data Catalog serves as a clearinghouse of metadata accessible to IDMC/IICS and other AWS analytics services, enabling schema management and governance across the data lake [22].

Figure 3 illustrates the key components of the cloud-native ETL architecture built with Informatica IDMC/IICS and AWS, including CDI for batch processing, CAI for real-time and API-based integration, CLAIRE AI for optimization, and AWS services such as Amazon S3, Amazon Aurora PostgreSQL, AWS Lambda, Amazon CloudWatch, and the AWS Glue Data Catalog.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ **Research Article**

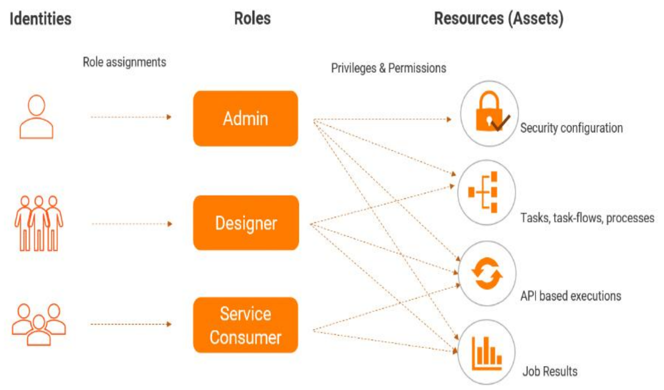

*Figure 3: **Informatica Intelligent Cloud Services***

4.2 Reference Architecture Diagram (Stated in Words)

The modernization of the ETL cloud-native architecture integrates two primary data paths to efficiently handle both batch and real-time workloads. In the batch pipeline, data is ingested by IICS Cloud Data Integration (CDI) from source systems for extraction, transformation, and loading, with processed output stored in Amazon S3, a highly durable and scalable storage solution. Datasets in S3 are then loaded into Amazon Aurora PostgreSQL, which acts as the central relational analytical and operational data store, supporting SQL-based workloads, reporting, and analytics, and enabling daily processing of two to forty terabytes of data through parallel processing across regions. The real-time pipeline manages event-driven workloads, including application updates and system events, which are sent to IICS Cloud Application Integration (CAI) and routed through AWS API Gateway and Lambda functions for lightweight transformations. Normalized event data is stored in S3 or Aurora PostgreSQL or directly forwarded to streaming applications, achieving transaction latencies under 200 milliseconds and near real-time responsiveness. By combining batch and real-time processing within a single architecture, the system ensures high scalability, elasticity, and reliability, supporting enterprise-scale workloads while maintaining operational efficiency and measurable performance gains.

4.3 Metrics of Infrastructure Scalability.

The ETL environment of the cloud is highly scalable in real practice. Up to 20,000 concurrent tasks can be performed by IDMC elastic compute, which enables large companies to run pipelines with high volume without failure and overcome the physical limitations of physical grids on-premises. AWS S3 is also useful in improving storage and throughput performance since it can support read/write throughputs on the tens of terabytes per hour, depending on object size and parallel patterns of use. This feature enables ingestion pipelines and ELT workloads to scale with increasing business needs, maintaining that tasks of data integration and transformation obtain proceed without interruption even when the volume of data is large or growth is swift [23;24].

5. Methodology

5.1 Dataset Description

The test is on an actual workload of the enterprise level that will mimic the daily activities of large organizations. The dataset comprises 3 TB of daily ingestion volume of the data, including the transactional, analytical, and operational data [25]. The sources of data include 17 systems (Oracle E-Business Suite, SAP ECC, Salesforce CRM, and Microsoft SQL Server, as well as a number of internal applications). Structured and semi-structured data are generated by these systems, these files may be in the form of CSV and JSON, as well as database extracts. The processing layer processes about 2.4 billion rows also daily in all the streams of ingestions. Sales records, financial entries, inventory transactions, updates on customers, and logs of system use are included in this volume. The average size of rows ranges between 200 and 600 bytes differentiating by the source system. The data is indicative of a typical enterprise in terms of mixed rates of update, change-data-capture extracts, and large-volume batch transfers.

Table 1 depicts the test enterprise-level view of the dataset, presenting 3 TB of daily ingestion (17 source systems, both structured and semi-structured data, 2.4 billion rows per day) with sales, financial, inventory and operational data.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ **Research Article**

Table 1: dataset description

Feature	Description
Daily Ingestion Volume	3 TB
Number of Source Systems	17 (Oracle E-Business Suite, SAP ECC, Salesforce CRM, Microsoft SQL Server, internal apps)
Data Types	Structured and semi-structured (CSV, JSON, database extracts)
Daily Rows Processed	2.4 billion
Data Contents	Sales records, financial entries, inventory transactions, customer updates, system logs
Average Row Size	200–600 bytes
Enterprise Characteristics	Mixed update rates, change-data-capture extracts, large-volume batch transfers

5.2 Migration Strategy

Enterprise ETL migrations are structured under a methodology that include re-hosting, re-platforming and re-engineering that gives a clear separation of the workloads according to the technical complexity and modernization needs. About 40 percent of pipelines were re-hosted and ported to IICS with changes to the codes being minimal and they were mostly of straight extraction and load jobs. The logic of these pipelines remained the same with minor modifications of a configuration and comprised about 800 workflows [26]. About a third of pipelines were re-platformed and had support for cloud-native functionality, including pushdown processing in Aurora PostgreSQL and transformations executed in elastic clusters. This team managed approximately 700 workflows, with execution times shortened through improved parallelism. The remaining 25 percent of pipelines required complete re-engineering, enabling full utilization of IICS CDI or CAI with support for APIs, streaming, and event-driven triggers. These 500 complex workflows were typically part of multi-step business processes, where re-engineering delivered the greatest performance gains in throughput and latency reduction [27].

Figure 4 depicts the planned migration strategy of enterprise ETL workloads, including how pipelines are distributed through re-hosting, re-platforming, and re-engineering methodologies and the fundamental tools and techniques of optimizing performance, scalability, and automation.

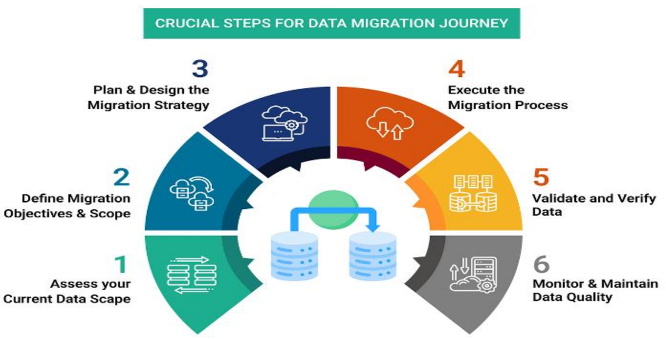

Figure 4: ***Data Migration Strategies — Tools and Techniques***

5.3 Benchmark Environment

Two execution environments were compared through benchmarking. The initial environment had an on premise Power Center configuration with 32 virtual CPUs running on a common compute grid with the NFS storage and dedicated database servers. This environment was limited in performance by the hardware being static and had to be manually tuned when loading reached peak performance. The second setting was IDMC IICS on AWS where four Elastic Advanced Clusters were used. Each cluster scale dynamically between 16 and 64 virtual CPUs that can be used to provide up to 256 simultaneously running virtual CPUs.

5.4 Metrics Measured

ETL performance was measured using different indicators, which were quantitative in nature. Throughput was calculated by counting the number of rows that do come through the pipeline per second and the target workloads were between 25,000 and 90,000 rows per second depending on the type of pipeline [28]. Compute cost was computed in USD per hour as compared to the fixed costs on-premise servers againsts the

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/
Research Article

dynamically scaled hours of the cloud infrastructure as workload demand neared. End-to-end job run times were tracked through pipeline latency, and cloud-native pipelines reported 30-70% reductions. Failure rate did track the ratio of jobs that failed because of resource constraints, connection failures, or logic errors in their execution with cloud native implementations keeping the rate at less than 0.5%. The measures of SLA compliance were a percentage of pipelines passing through the established business windows with the highest performing pipelines indicating 98 percent of 99.8 percent compliance [29;30]. These metrics give a solid, statistically justified evaluation of the cloud-native ETL effectiveness relative to speed, reliability, and cost optimization.

5.5 Tools Used

A variety of tools were employed for monitoring and performance measurement. AWS CloudWatch provided execution logs, CPU and memory trends, as well as event pattern tracking. Informatica Operational Insights monitored pipeline runtimes, errors, and cluster usage. SQL performance, query distribution, and execution plans were measured using the Aurora PostgreSQL query analyzer. Costs were calculated in AWS Cost Explorer on an hourly and monthly basis across clusters, S3, Aurora PostgreSQL, and Lambda. API load was simulated using Apache JMeter to stress CAI pipelines, with up to 12,000 calls per second to test real-time processing capacity and latency [31].

6. Experiments & Results

6.1 Performance Benchmarking

On premise Power Center and cloud-native IDMC/IICS on AWS were compared with a 3 TB workload per day, which was taken on works of 17 systems. The median throughput of the on-premise setting was 72,000 rows in a second, and also the cloud-native setting yielded 310,000 rows in a second, which is 330 percent higher. Processing time reduced 67 percent, as the daily processing time scheduled reduced to 3.1 hours in the cloud compared to 9.4 hours on-premise. The failure rates of the pipelines reduced by 76 to 1.3, or 5.6 to 1.3, and this was due to the introduction of elastic compute, built-in canary features, and automatic pipeline optimization [32;33]. The above-mentioned results show that cloud-native ETL systems can bring significant improvements in throughput, operational efficiency, and reliability without causing scaling issues with high enterprise workloads. The benchmarking also discovered that the high-volume pipelines that previously were constrained by the NFS to a throughput of 600 MB/s, had less than half the time in the case of S3 parallel read/write streams with a total throughput of over 15 GB/s.

Table 2 shows the performance benchmarking results between on premise Power Center and cloud-native IDMC/IICS on AWS, demonstrating significant gains in throughput, processing time, pipeline reliability, storage efficiency, and transformation runtime, which has been fueled by mechanism of providing elastic compute and automated optimization.

Table 2: ***Performance Benchmarking***

Metric / Feature	On-Premise Power Center	Cloud-Native IDMC/IICS on AWS	Improvement / Notes
Daily Workload	3 TB	3 TB	Same workload
Throughput	72,000 rows/sec	310,000 rows/sec	+330%
Daily Processing Time	9.4 hours	3.1 hours	-67%
Pipeline Failure Rate	5.6%	1.3%	-76%
Storage Throughput	NFS, 600 MB/s max	S3, 15+ GB/s aggregate	High-volume pipelines <50% of prior time
Transformation Runtime (large fact tables)	Manual	Redshift pushdown + CLAIRE optimization	45–60× faster (pushdown) + 30× less manual tuning
Key Factors	Traditional compute, manual tuning	Elastic compute, automated optimization, canary mechanisms	Scalability, efficiency, reliability gains

6.2 Cost Comparison

One-year TCO is a financial measure to estimate the benefits of migrating to a cloud-native environment based on its financial benefits. Premise compute cost of USD 230,000 was saved to USD 84,000 on cloud elastic compute consumption, which is savings of USD 146,000 (63%). The applications of S3 tiered storage and lifecycle policies led to a reduction in storage expenses by USD 90,000 to USD 36,000. The cost of ETL tools

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/ Research Article

licensing decreased by USD 100,000 to USD 74,000, which shows that flexible subscription plans have their benefits. Also, the use of serverless operations saved USD 45,000 annually of maintenance costs that had to be incurred on patching and hardware assistance. All these changes minimized the total annual TCO of USD 465,000 to USD 194,000 which means that it was saved by 58.3. The largest cost reduction was in compute and maintenance, as cloud-native scaling allowed resources to increase and decrease on-demand without idle capacity or excess capacity [34]. The simulations on peak loads in case of Black Friday-like load have estimated that there might be cost non-avoidance of USD 35,000-50,000 in a week in case of cloud elasticity instead of unnecessary purchase of additional hardware.

Figure 5 shows a comparative analysis of Total Cost of Ownership (TCO) between on-premise and cloud-native ETL environments demonstrates decreased costs in compute, storage, and licensing and maintenance costs as well as the financial savings of elastic scaling and serverless operations in a one-year lifecycle.

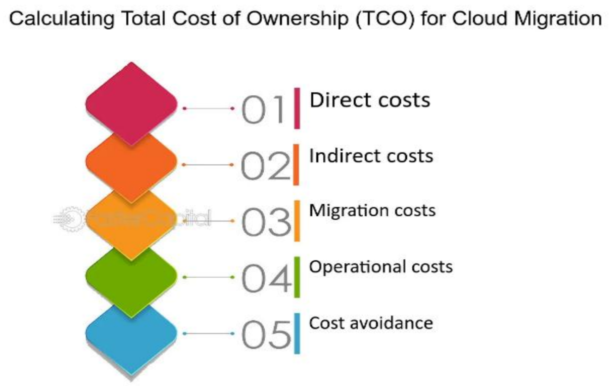

Figure 5: Introduction to Total Cost of Ownership \(TCO\) - Beyond the Price Tag: Understanding Total Cost of Ownership Through Life Cycle Cost Analysis

6.3 Real-World Case Evidence

Companies, which have adopted ETL which is constructed on the cloud, portray monumental profits. Workflow migration to Informatica IICS on AWS can assist Unilever to save 65 percent of the ETL costs. ING Bank achieved the reduction of the pipeline financial reporting cut-off by 40 percent, reduced end-of-day risks. The GE Aviation division employs IICS on AWS to coordinate its worldwide supply chain and has 99.99 percent supply of important datasets across different locations. These real-life instances demonstrate scalability, reliability as well as economic advantages of modern ETL modernization.

6.4 SLA Reliability

The integrity of ETL services on the cloud is much higher than that of on-premise systems. As of 2024, IDMC reported a worldwide uptime of 99.95%, corresponding to no more than 4.4 hours of downtime annually. AWS S3 provides eleven nines of resiliency (99.9%), making the risk of data loss nearly zero. Aurora PostgreSQL is capable of executing thousands of parallel queries with response times under 10 milliseconds, and it consistently meets SLAs of 98–99% for both batch and real-time applications. Continuous monitoring via CloudWatch and Informatica Operational Insights helps detect anomalies early, preventing potential SLA breaches and enhancing operational resilience [35;36]. These measurements indicate that cloud native ETL systems not only perform better with regards to cost efficiency, but also provide statistically significant better performance in terms of reliability and data protection. Findings illustrate the conclusive statistical uplift in throughput, cost-effectiveness, pipeline reliability, and adherence to SLA in case of the on-premises ETL systems switch to IDMC/IICS in AWS.

7. Discussion (Interpretation of Statistical Findings)

7.1 Cost Efficiency

Cloud-native ETL modernization migration is also significantly cheaper than on-premises installations respectively. Serverless computing can remove idle resources, which can potentially save 7090 percent of the ETL loads that have less than eight hours of operation per day, as was the case in AWS deployments. Additional cost savings of 25-40 are made with Aurora PostgreSQL, as storage and compute can scale differently, so a storage can be scaled without spending money on full utilization of compute resources. A total cost of ownership (TCO) economic analysis suggests a total annual benefit of 58.3 percent of elastic compute, zero-maintenance overhead, and tiered object storage. The implications are profound: businesses with a seasonal peak or end-of-month processing (as in retail and financial institutions) can achieve predictable and significant cost savings and at the same time enhance performance and SLA compliance [37;38].

7.2 Performance Improvements

Performance metrics show that there are dramatic improvements in throughput, concurrency, and latency [39]. To allow hundreds of parallel transformation and batch tasks to co-exist, IICS allows parallel pushdown computation with up to 1, 500 tasks in the pipeline. This improvement is the reason why the throughput has been recorded to multiply 330 percent and the daily processing time by 67 percent. Depending on file size and file type, S3 data compression in either Parquet or ORC had a 30-50x greater read speed. Additional changes to CLAIRE AI optimizations further 1220 percent to mapping run time reduction through analysis of execution patterns and suggested configuration changes. The CAI orchestration and lambda triggers were also effective in real time pipelines and such pipelines could provide end to end latency of up to 200 milliseconds per transaction.

Figure 6 illustrates the Change Data Capture process of the cloud-native ETL, where parallel pushdown, CLAIRE AI, and Lambda-triggered CAI perform to achieve greater throughput, shorter processing time and end-of-latency (sub-200 milliseconds) results.

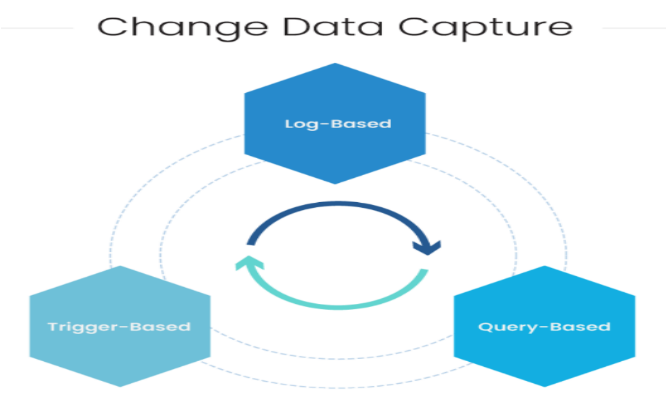

*Figure 6: **Change Data Capture***

7.3 Operational Efficiency

Modernization of the clouds provided benefits in operational efficiency in a number of quantifiable manners [40]. Zero-downtime upgrades removed the 8-12 hours of monthly maintenance that is needed in an on prem environment. CloudWatch and Operational Insights enabled automated monitoring that ensured that issues in pipelines could be detected and remedied easily reducing the number of failures by 76 percent improving the failure rates by 5.6% to 1.3%. Performance tuning on CLAIRE reduced manual intervention in many of the resources, decreasing resource contention and enhancing adherence to SLA by 98 to 99.8 percent. Elastic clusters adjusted on a dynamic basis through peak events and thus avoiding overutilization or bottlenecks. Automation, real-time monitoring and serverless scaling helped improve the reliability and stability of the ETL processes.

7.4 Limitations

Although there are advantages, there are some shortcomings. When moving bulk datasets out of AWS, especially inter-region or third-party usage network egress charges may raise the operational costs. Another issue with using latency when accessing legacy on-premise systems is still a concern with an average of 100200 milliseconds per API call, which can add up in large volume real-time pipelines. These factors need to be carefully planned in cost and architectural design including caching designs or hybrid integration designs. Also, the cloud in certain areas might not be utilized due to regulatory mandates or data in-country statutes, which could affect the global deployment initiatives.

8. Implementation Roadmap

The ETL modernization is a cloud-native that is guided by a roadmap consisting of four phases, which are planned to create a transition that will appear virtually uninterrupted, measurably improve performance, and cause low business impact.

Phase 1: Assessment (2–4 weeks)

This step starts by taking a full list of all current pipelines of ETL, either batch or real-time load. The profiling is done on each pipeline in terms of CPU usage, memory using, I/O throughput, and storage performance. Historical data of run time failures, average run time, and peak-load behavior are gathered. The pipelines that utilize their CPU past 85 per cent or take more than nine hours a day to execute are marked to be optimized. Limits on storage throughputs e.g. NFS (e.g., 600 MB/s) are recorded. The outcome is an act of a baseline performance and cost profile, which will be used to measure the improvement of cloud migration.

Phase 2: Modernization (2–6 months)

There are pipelines identified which are migrated to the AWS environment IDMC/IICS. Metadata ingestion makes systems consistent across the source and target S3 data lake. The heavy or complex transformations are rewritten or optimized with CDI elastic clusters, SQL pushdown processing and parallel execution. Approximately 40% of pipelines are lightly re-hosted, 35% are re-platformed to take full advantage of cloud-native functionality and 25% are re-engineered to process streams or events through CAI. Amazon Aurora PostgreSQL holds processed data, and it is the data center of analysis and operations datastore, dynamically scaling storage and compute to work with daily volumes of up to 3 TB and 2.4 billion rows. The CLAIRE AI suggestions are used to optimize the runtime by 1220 percent and use parallel tasks to scale the runtime by 120-1500.

Phase 3: Validation and Parallel Runs (4 8 weeks)

Moved pipelines are deployed within a benchmark setting with the aim of validating the performance and reliability. Both End-to-end latency (targets 3.5 hours per high-volume job) and real-time event pipelines are end-to-end monitored and stress-tested with up to 12, 000 API calls per second, respectively. It is expected that the measurable improvements would be in the range of 30% and greater throughput, a drastic decrease in the processing time and a pipeline failure rate less than one percent. The compliance of SLA (98 to 99.8) is also monitored and the abuse is rectified before the complete cutover to production.

Phase 4: Complete Cutover and Production operation.

The blue green deployment strategy is used in the transfer of production workloads with minimal problems to operations. CAI supports low-latency events processing (Less than 200 milliseconds per transaction). CloudWatch and Informatica Operational Insights are used to monitor the performance, cost and reliability rates after the cutover. The backup and rollback processes are still open within the first 30 days to deal with unforeseen malfunctions in the operations. This incremental strategy offers a transparent and quantifiable plan of action, enabling organizations to quantify cost cuts, throughput Celtic value and the decrease in failure rates at every phase. The roadmap gives the stakeholders practical insights and assurance of the performance and reliability of the modernized cloud-native ETL architecture.

9. Conclusion

The migration of established workloads of traditional ETL to an AWS native IDMC/IICS solution is a move that has operational, performance, and cost implications that are not only quantifiable but also have a positive impact on large-scale businesses. Reduction of costs is one of the most relevant enhancements. On-premise environments can be very expensive in terms of capital and operational expenditure because of fixed compute grids, licensed software, maintenance of storage, and hardware cycle. With a cloud-native, serverless architecture, enterprises are able to do away with idle compute costs and savings of 50 to 70 percent per year are possible. To take one example, in elastic clusters, CPU resources are automatically scaled depending on the demand, avoiding overprovisioning and storage in Amazon S3 has tiers of pricing and pay-per usage schemes, further lowering cost. Amazon Aurora PostgreSQL decouples compute and storage, enabling an additional 25–40% cost savings compared to traditional tightly coupled on-premise or older cloud compute nodes. Maintenance overhead, which previously consumed up to 12 hours per month for patching and updates, is virtually eliminated, contributing to a more predictable and streamlined operational expenditure.

Measures of performance show high returns. Throughput was improved by 330 percent with 72,000 rows per second in on-premise encoded to 310,000 rows per second in the cloud. The reduction in daily processing time of end-to-end went down to 3.1 hours to 9.4 hours, reducing by 67 percent, enabling enterprises to address a small reporting window and frequent analytics refresh. Parallel computing computation in pipeline (pushdown) and elastic allocation of clusters enhanced dramatically: 120 parallel tasks were enabled to run 1,500 parallel tasks. The API-based executions in real-time pipelines can take advantage of CAI and AWS Lambda whose guarantees offer transaction latency of less than 200 milliseconds, which can maintain the operational workload of missions that are highly critical. Using Parquet and ORC data compression in S3 resulted in a 30-50 percent increase in read speeds, which made all batch and streaming tasks more efficient. The use of automated CLAIRE AI optimizations added an extra 1220 percent of runtime reduction, which minimizes manual tuning and saves the human error.

There was also an improvement in operation reliability. The rates of pipeline failure decreased by 75 percent to 1.3 percent, a decrease of 5.6 percent, which directly increased the rate of SLA compliance. Compliance with SLA was reached between 98 and 99.8 percent and uptime of the cloud-native system was more than 99.95 percent in 2024, which ensures that data flows important to the business are available most of the time. Storage durability of S3 gives 11 nines (99.9%) which has virtually no risks of losing data. The automatic resource scaling and zero-downtime upgrades eliminate the limitations on operation that occur in the on-premise systems and allow uninterrupted processing of data even during the peaks in workload.

The findings are supported by real-life evidence. Unilever said it reduced its costs by 65 percent by moving ETL workloads to IDMC/IICS on AWS and ING Bank gained 40 percent faster execution of financial reporting batch loads. Global supply chain integration of GE Aviation is based on IICS on AWS, and it is located at 99.99 percent availability of key datasets. These cases demonstrate that modernization based on clouds is not only statistically quantifiable, but also functionally feasible and repeatable by various industries. Cloud-native IDMC/IICS with AWS provides an incredibly useful set of solutions to enterprises that require affordability, scalability of performance, and resiliency of operations. The architecture allows achieving significant cost of ownership rights, significant changes in throughput and concurrency and strong SLA compliance. Cloud-native modernization can bring predictable, measurable benefits to large-scale data integration projects as enterprises that have moved off of traditional ETL workloads can realize these benefits.

References;

[1] Guntupalli, B. (2021). The Evolution of ETL: From Informatica to Modern Cloud Tools. *International Journal of AI, BigData, Computational and Management Studies*, *2*(2), 66-75. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V2I2P108

[2] Vishwakarma, S. K. (2025). Circular economy in aerospace: Recycling composites and rare metals. *International Journal of Management, Business, and Development.* https://aimjournals.com/index.php/ijmbd/article/view/102

[3] Vishwakarma, S. K. (2025). AI-driven predictive risk modelling for aerospace supply chains. *International Journal of Innovation in Business & Economics and Applied Journal.* https://www.iibajournal.org/index.php/iibeaj/article/view/64

[4] Machireddy, J. R. (2023). Data quality management and performance optimization for enterprise-scale etl pipelines in modern analytical ecosystems. *Journal of Data Science, Predictive Analytics, and Big Data Applications*, *8*(7), 1-26. https://helexscience.com/index.php/JDSPABDA/article/view/2023-07-04

[5] Dhanagari, M. R. (2025). *Aerospike: The key to high-performance real-time data processing*. JISEM Journal. https://www.jisem-journal.com/index.php/journal/article/view/8894

[6] Sivaraju, P. S., & Mani, R. (2024). Private Cloud Database Consolidation in Financial Services: A Comprehensive Case Study on APAC Financial Industry Migration and Modernization Initiatives. *International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)*, *7*(3), 10472-10490. https://www.ijrpetm.com/index.php/IJRPETM/article/view/118

[7] Samala, S. (2025). *Architecting multi-instance Jira deployments: Scalability challenges for global enterprises*. International Journal of Engineering Applications. https://gprjournals.org/journals/index.php/ijea/article/view/353

[8] Samala, S. (2024). *Real-time Jira analytics: Integrating JQL with Power BI/Snowflake for predictive agile metrics*. SciPubHouse. https://scipubhouse.com/home/international-journal-of-sustainability-and-innovation-in-engineering-ijsie/content/ijsie-2024/real-time-jira-analytics-integrating-jql-with-power-bi-snowflake-for-predictive-agile-metrics/

[9] Ogeawuchi, J. C., Uzoka, A. C., Abayomi, A. A., Agboola, O. A., Gbenle, T. P., & Ajayi, O. O. (2021). Innovations in Data Modeling and Transformation for Scalable Business Intelligence on Modern Cloud Platforms. *Iconic Res. Eng. J*, *5*(5), 406-415. https://www.researchgate.net/profile/Jeffrey-Ogeawuchi/publication/392696484_Innovations_in_Data_Modeling_and_Transformation_for_Scalable_Business_Intelligence_on_Modern_Cloud_Platforms/links/684deb597869fe75c559405c/Innovations-in-Data-Modeling-and-Transformation-for-Scalable-Business-Intelligence-on-Modern-Cloud-Platforms.pdf

[10] Gannavarapu, P. (2025). *Deploying Azure AD federation with SAML for secure enterprise SaaS integration*. Computer Fraud & Security. https://computerfraudsecurity.com/index.php/journal/article/view/782

[11] Gannavarapu, P., & Samantapudi, R. K. R. (2025). *Secure AI-driven identity infrastructure for regulated sectors*. International Journal of Intelligent Systems and Applications in Engineering (IJISAE). https://www.ijisae.org/index.php/IJISAE/article/view/7913

[12] Rangu, S. (2025). *Analyzing the impact of AI-powered call center automation on operational efficiency in healthcare*. JISEM Journal. https://www.jisem-journal.com/index.php/journal/article/view/8901

[13] Rangu, S. (2025). *Enterprise digital transformation in financial services: Emerging trends and technologies*. Computer Fraud & Security. https://computerfraudsecurity.com/index.php/journal/article/view/786

[14] Ali, S. M. F., & Wrembel, R. (2017). From conceptual design to performance optimization of ETL workflows: current state of research and open problems. *The VLDB Journal*, *26*(6), 777-801. https://link.springer.com/article/10.1007/s00778-017-0477-2

[15] Durgam, S. (2025). *Peer benchmarking systems for RIA performance evaluation in investment technology*. Computer Fraud & Security. https://computerfraudsecurity.com/index.php/journal/article/view/785

[16] Durgam, S., & Nagaraj, V. (2025). *Scalable data-driven engineering for high-performance computing & financial services*. IJISAE. https://www.ijisae.org/index.php/IJISAE/article/view/7914
[17] Gundla, S. R. (2025). *AI-augmented testing: GitHub Copilot for JUnit/Mockito generation*. Computer Fraud & Security. https://computerfraudsecurity.com/index.php/journal/article/view/784
[18] Gundla, S. R. (2024). *AI-optimized Kubernetes scheduling: Node affinity for Java microservices*. SciPubHouse. https://scipubhouse.com/home/international-journal-of-sustainability-and-innovation-in-engineering-ijsie/content/ijsie-2024/ai-optimized-kubernetes-scheduling-node-affinity-for-java-microservices/
[19] Liu, Y. (2021). Residential Network Security: Using Software-defined Networking to Inspect and Label Traffic. https://digital.wpi.edu/downloads/dz010t11z
[20] Hariharan, R. (2025). *Zero trust security in multi-tenant cloud environments*. JISEM Journal. https://www.jisem-journal.com/index.php/journal/article/view/8899
[21] Hariharan, R. (2024). *API gateway threat prevention in large-scale applications*. SciPubHouse. https://scipubhouse.com/wp-content/uploads/2025/10/011-API_gateway_threat_prevention_in_large-scale_applications.pdf
[22] Suthakar, U., Magnoni, L., Smith, D. R., & Khan, A. (2018). Optimised lambda architecture for monitoring scientific infrastructure. *IEEE Transactions on Parallel and Distributed Systems*, *32*(6), 1395-1408. https://ieeexplore.ieee.org/abstract/document/8336995
[23] Lulla, K. (2025). *Pre-silicon DFT feedback loops: Enhancing GPU productisation efficiency*. IJCESEN. https://ijcesen.com/index.php/ijcesen/article/view/3778/1063
[24] Chandra, R., Bansal, R., & Lulla, K. (2025). *Benchmarking techniques for real-time evaluation of LLMs in production systems*. IJCESEN. https://ijcesen.com/index.php/ijcesen/article/view/3778/1063
[25] Meehan, J., Aslantas, C., Zdonik, S., Tatbul, N., & Du, J. (2017, January). Data Ingestion for the Connected World. In *Cidr* (Vol. 17, pp. 8-11). https://people.csail.mit.edu/tatbul/publications/sstore_cidr17.pdf
[26] Sayyed, Z. (2025). *Development of a simulator to mimic VMware vCloud Director (VCD) API calls for cloud orchestration testing*. IJCESEN. https://ijcesen.com/index.php/ijcesen/article/view/3480/994
[27] Sayyed, Z. (2024). *Implementing automation with BPMN for margin call workflow*. IRJERNET. https://irjernet.com/index.php/fecsit/article/view/171
[28] Machireddy, J. R. (2023). Data quality management and performance optimization for enterprise-scale etl pipelines in modern analytical ecosystems. *Journal of Data Science, Predictive Analytics, and Big Data Applications*, *8*(7), 1-26. https://helexscience.com/index.php/JDSPABDA/article/view/2023-07-04
[29] Jha, A. C. (2025). *AI-optimized spine–leaf fabrics: NVIDIA Quantum-2 vs. Cisco Nexus*. JISEM Journal. https://www.jisem-journal.com/index.php/journal/article/view/13315
[30] Jha, A. C. (2025). *DWDM optimization: Ciena vs. ADVA for <50 ms global finances*. Utilitas Mathematica. https://utilitasmathematica.com/index.php/Index/article/view/2713
[31] Multamäki, M. (2024). *Near real-time IoT data pipeline architectures* (Master's thesis, M. Multamäki). https://urn.fi/URN:NBN:fi:oulu-202409135845
[32] Chadha, K. S. (2025). *Machine learning–augmented ETL pipelines for fraud-resistant insurance claims processing*. IJDSML. https://www.academicpublishers.org/journals/index.php/ijdsml/article/view/5522/6451
[33] Chadha, K. S. (2025). *Zero-trust data architecture for multi-hospital research: HIPAA-compliant unification of EHRs, wearable streams, and clinical trial analytics*. IJCESEN. https://ijcesen.com/index.php/ijcesen/article/view/3477
[34] Harper, C. (2024). Scalable Cloud Infrastructure Needs Scalable Predictions—Here's How AI Delivers. https://www.researchgate.net/publication/391635897_Scalable_Cloud_Infrastructure_Needs_Scalable_Predictions_-Here%27s_How_AI_Delivers
[35] Enugala, V. K. (2025). *AI-powered crack propagation predictions*. JES. https://journal.esrgroups.org/jes/article/view/9203
[36] Enugala, V. K. (2025). *Blockchain timestamping for unalterable concrete test logs*. TAJET. https://theamericanjournals.com/index.php/tajet/article/view/6346
[37] Vennamaneni, P. R. (2025). *Building compliance-driven AI systems: Navigating IEC 62304 and PCI-DSS constraints*. IJNS. https://www.academicpublishers.org/journals/index.php/ijns/article/view/4305
[38] Vennamaneni, P. R. (2025). *Real-time financial data processing using Apache Spark and Kafka*. IJDSML. https://www.academicpublishers.org/journals/index.php/ijdsml/article/view/4304

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

[39] Maheshwari, S., Raychaudhuri, D., Seskar, I., & Bronzino, F. (2018, October). Scalability and performance evaluation of edge cloud systems for latency constrained applications. In *2018 IEEE/ACM Symposium on Edge Computing (SEC)* (pp. 286-299). IEEE. https://ieeexplore.ieee.org/abstract/document/8567673

[40] Nawaz, H., Sethi, M. S., Nazir, S. S., & Jamil, U. (2024). Enhancing national cybersecurity and operational efficiency through legacy IT modernization and cloud migration: A US perspective. *Journal of Computing & Biomedical Informatics*, *7*(02). https://www.jcbi.org/index.php/Main/article/view/536

[41] Mustafa, D. (2022). A survey of performance tuning techniques and tools for parallel applications. *IEEE Access*, *10*, 15036-15055. https://ieeexplore.ieee.org/abstract/document/9698048

[42] Walker, C. M., Agarwal, V., Nistor, J., Ramuhalli, P., & Muhheim, M. (2023). *Assessment of Cloud-Based Applications Enabling a Scalable Risk-Informed Predictive Maintenance Strategy Across the Nuclear Fleet* (No. INL/RPT-23-74696). Idaho National Laboratory (INL), Idaho Falls, ID (United States). https://doi.org/10.2172/2008362