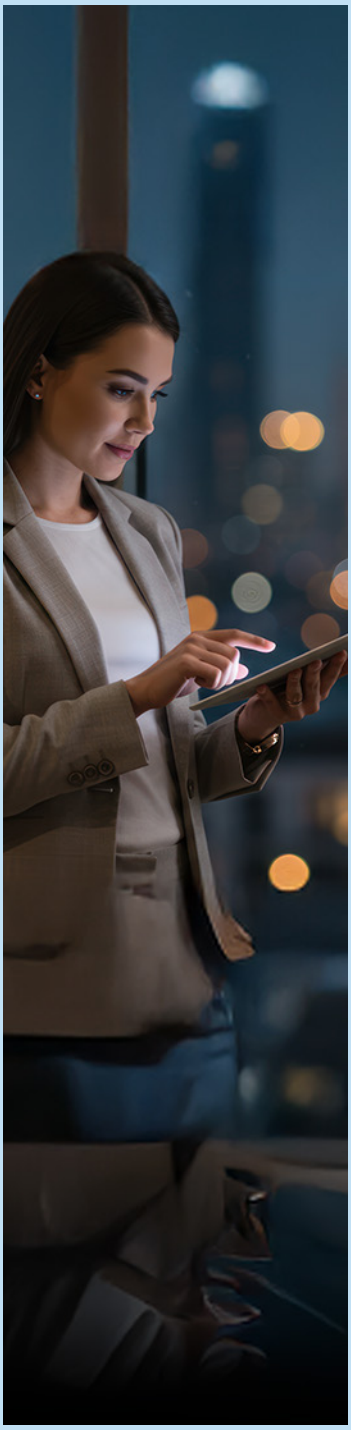
Infosys topaz

VIEW POINT



UNLOCK NEXT-GEN DATA SCIENCE & ENGINEERING

THE STRATEGIC SHIFT TO SNOWFLAKE SNOWPARK

The complexity of managing traditional Apache Spark environments presents organizations with operational challenges that extend far beyond visible infrastructure costs. Teams are frequently burdened with time-intensive tasks such as cluster provisioning, scaling, dependency resolution, and environment-specific configurations—particularly in multi-cloud or hybrid architectures. These complexities are compounded in large-scale deployments, where separate clusters must be maintained for development, testing, and production, each requiring dedicated monitoring, DevOps expertise, and continuous attention to security, compliance, and disaster recovery.

snowflake®

Infosys®
Navigate your next

Infosys, as a trusted Snowflake Elite Services Partner, recognizes the deep impact of this legacy architecture: Data Scientists wait days instead of minutes, and Data Engineers spend more time managing infrastructure than delivering pipeline logic.

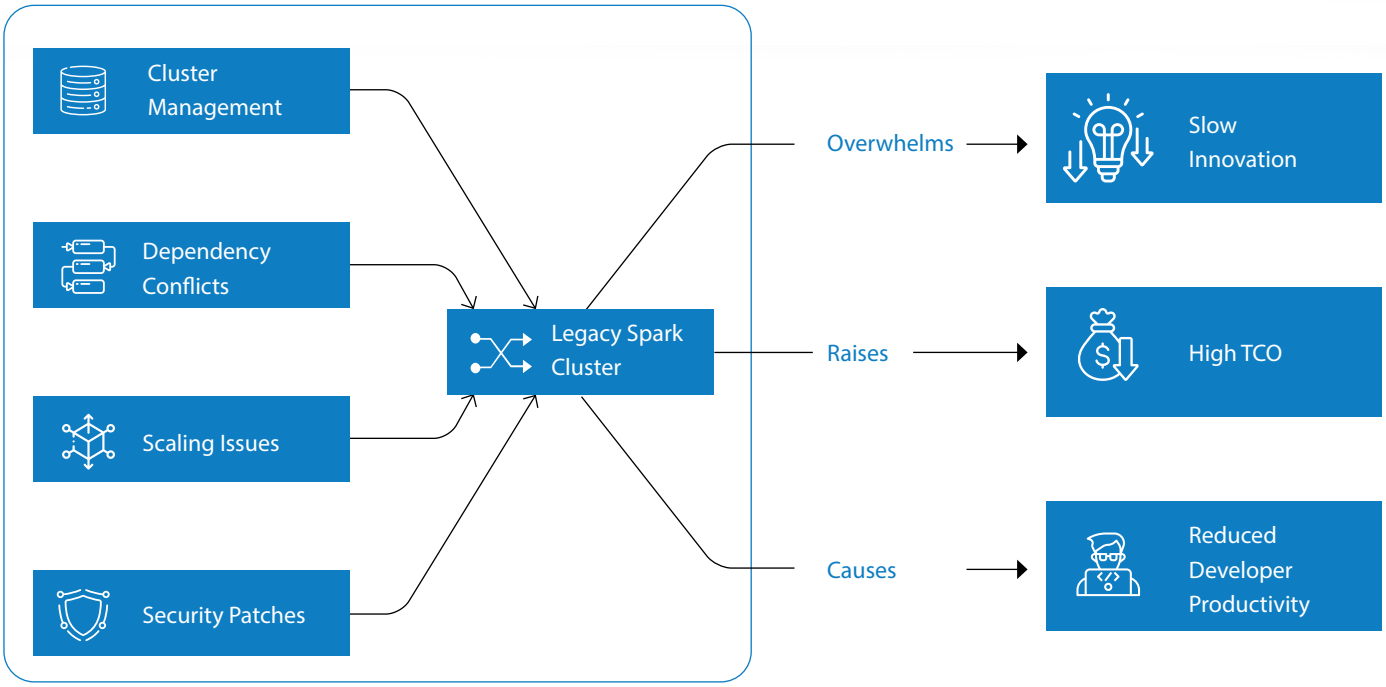
The Legacy Cost: Why Traditional Spark Limits Modern Data Initiatives

Managing traditional Spark environments is resource-intensive and riddled with inefficiencies:

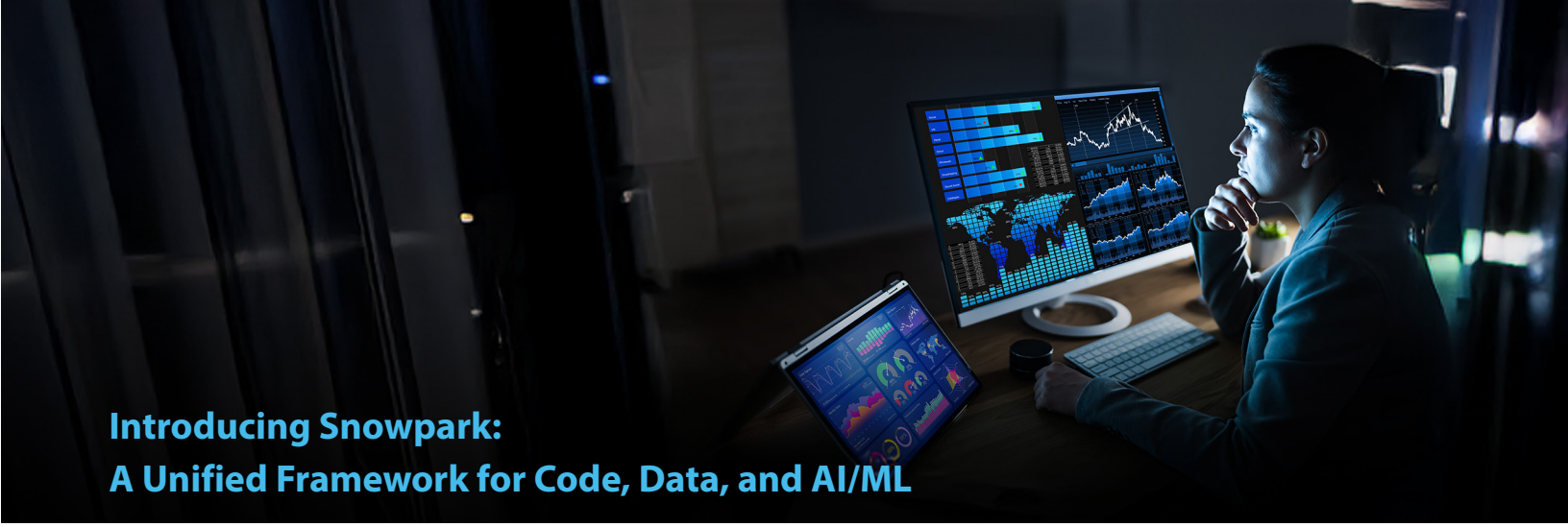
Operational Overhead: Spark workloads often encounter resource contention, suboptimal cluster sizing, and poor data locality, resulting in prolonged processing times and inefficiencies in critical tasks such as model training. This directly hinders productivity and accelerates technical debt, where the cost of maintaining infrastructure outpaces the value derived from analytics initiatives.

Hindered Data Science: This complexity translates directly to slower **Time-to-Insight** and hampered **ML iteration cycles**. Data must constantly be moved or copied, increasing latency, cost, and governance risk.

Unpredictable Economics: Provisioning for peak capacity leads to substantial idle resource costs, resulting in unpredictable bills and a high **Total Cost of Ownership (TCO)**.



Summary of the challenges posed by traditional Spark environments



Introducing Snowpark: A Unified Framework for Code, Data, and AI/ML

Snowpark is a set of libraries and code execution environments that empowers developers to securely deploy and process **code (SQL, Python, Java, Scala)** with the **Snowflake vectorized engine**. This transformative approach abstracts infrastructure complexity and enables seamless execution of workloads on Snowflake's cloud-native platform.

Snowpark fundamentally shifts the paradigm by allowing developers to ship their code to the data rather than exporting large datasets to external environments. This is achieved through three core components:

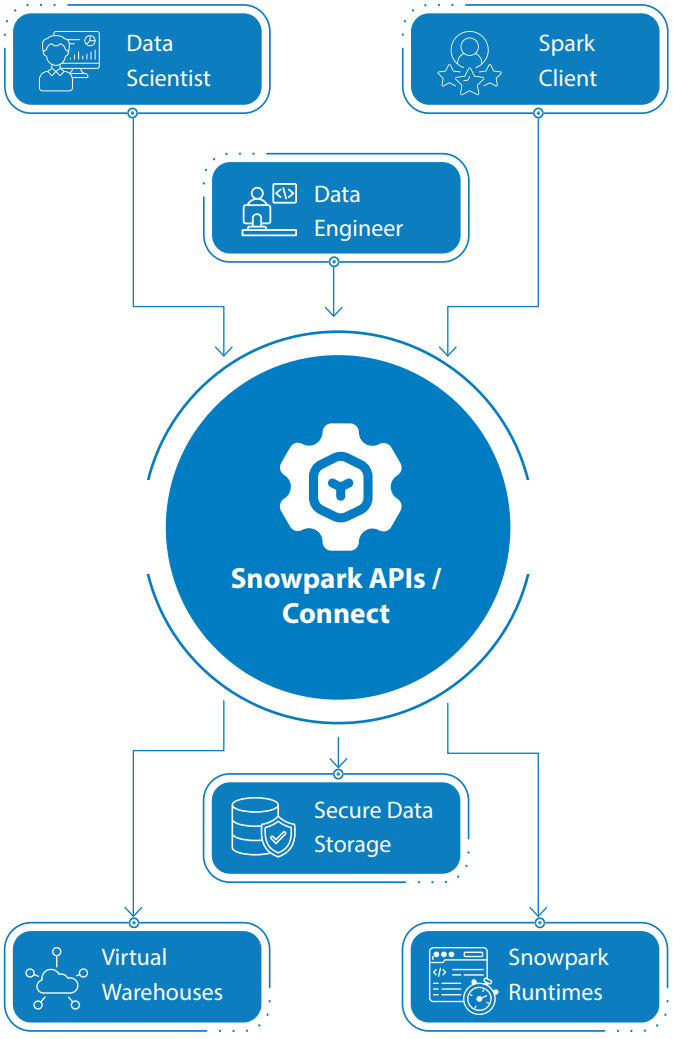
Snowpark APIs (Client Side): Developers use familiar programming languages (like Python) and libraries, including a DataFrame API (similar to Spark DataFrames) and native ML APIs, to write complex data transformation logic using familiar tools such as Jupyter, VS Code, or Snowflake Notebooks.

Runtimes (Server Side): Code is executed securely inside Snowflake's managed computes engine. Developers can deploy logic as User-Defined Functions (UDFs) and Stored Procedures (SPROCs), running custom logic directly in the data warehouse.

Snowpark Connect for Apache Spark (The Bridge): This specific component utilizes Spark 3.4's decoupled client-server architecture, enabling existing Apache Spark applications to use the Snowflake Data Cloud as the execution engine. This solution separates development from execution, allowing data teams to continue using familiar Spark interfaces while Snowflake manages orchestration, scalability, and resource allocation behind the scenes. Snowpark Connect provides secure and trusted connectivity for external access and API interactions, leveraging Snowflake's robust security features.

By tapping into Snowflake's readily available virtual warehouses, Snowpark eliminates the **long startup times** associated with traditional distributed clusters, leading to proven performance and cost efficiencies. This architectural shift eliminates the need for traditional cluster management, significantly reduces operational overhead, and provides transparent cost control through per-second billing.

How Snowpark unifies code execution within Snowflake



Strategic Value by Role: Why Snowpark is Recommended

Snowpark delivers unique, targeted advantages for every key stakeholder in the modern data ecosystem:

For Data Scientists: Accelerating Model Delivery

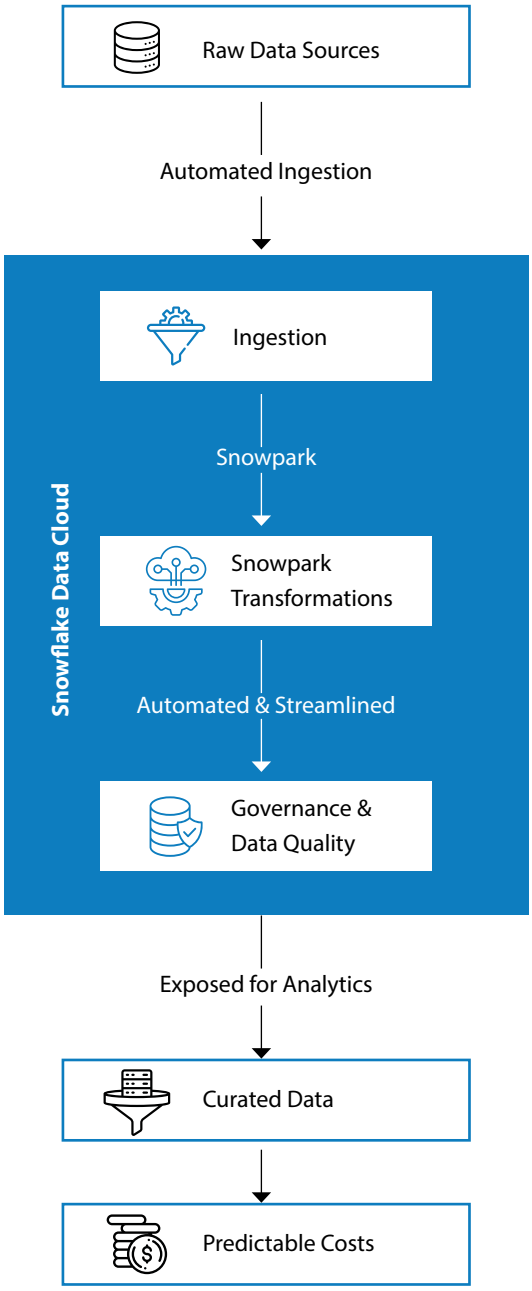
Snowpark is an ideal framework for large-scale Data Science and Machine Learning development and deployment because it provides immediate scale and eliminates data friction:

Massive Scale on Demand: Provision massive compute instantly for complex feature engineering and high-volume training tasks. Data Scientists focus solely on the algorithm, leveraging their preferred Python libraries (e.g., scikit-learn) without infrastructure constraints. Snowflake's intelligent query optimizer and auto-scaling capabilities enable interactive processing of large datasets, accelerating time-to-insight and reducing operational friction.

Eliminating Data Movement: Execute model training and batch inference directly on governed, live data in Snowflake using Snowpark's DataFrame API. This is significantly more efficient, dramatically reducing job runtime for complex workloads by removing delays caused by provisioning or resource contention.

Secure Deployment: Seamlessly integrate and deploy trained models as UDFs or Stored Procedures for secure, real-time inference within the platform, maintaining reliability, scalability, and security.

Here's how Snowpark empowers Data Scientists to accelerate their workflows:



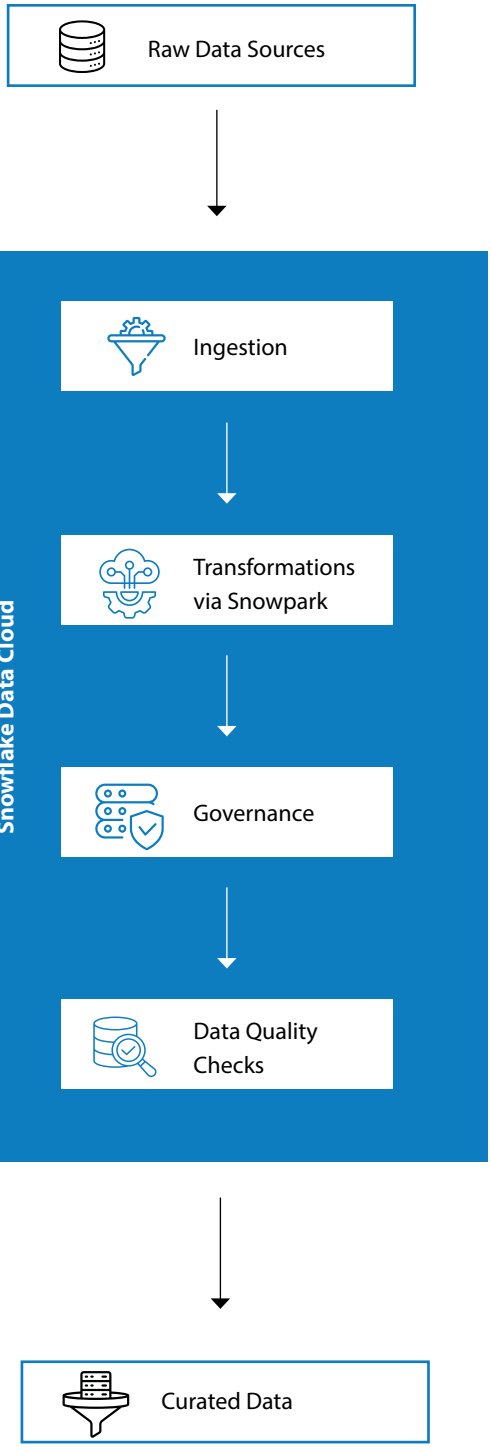
For Data Engineers, Snowpark replaces operational headaches with robust, production-ready capabilities for complex data transformations and ELT workloads:

True Operational Simplification: Achieve 100% compute-as-a-service. Engineers focus solely on pipeline logic, data quality, and building robust, well-engineered pipelines, leveraging the functional programming paradigm for better code readability and reuse. This directly addresses industry-wide challenges related to talent constraints, performance bottlenecks, and unpredictable infrastructure costs.

Unified Governance and Security: Inherit Snowflake's enterprise-grade security, governance, and auditing. All data remains within Snowflake's secure perimeter, removing the need for complex network configurations or encryption management. Uniform role-based access controls across environments further simplify governance and regulatory compliance.

Predictable FinOps: Snowpark introduces a transformative economic model for data processing. It replaces inefficiency with a true consumption-based pricing model, where compute resources are billed per second of actual usage. This elastic approach enables organizations to dynamically scale compute resources in response to workload demands, eliminating the need for manual intervention or long-term infrastructure commitments.

See how Snowpark simplifies the operational landscape for Data Engineers



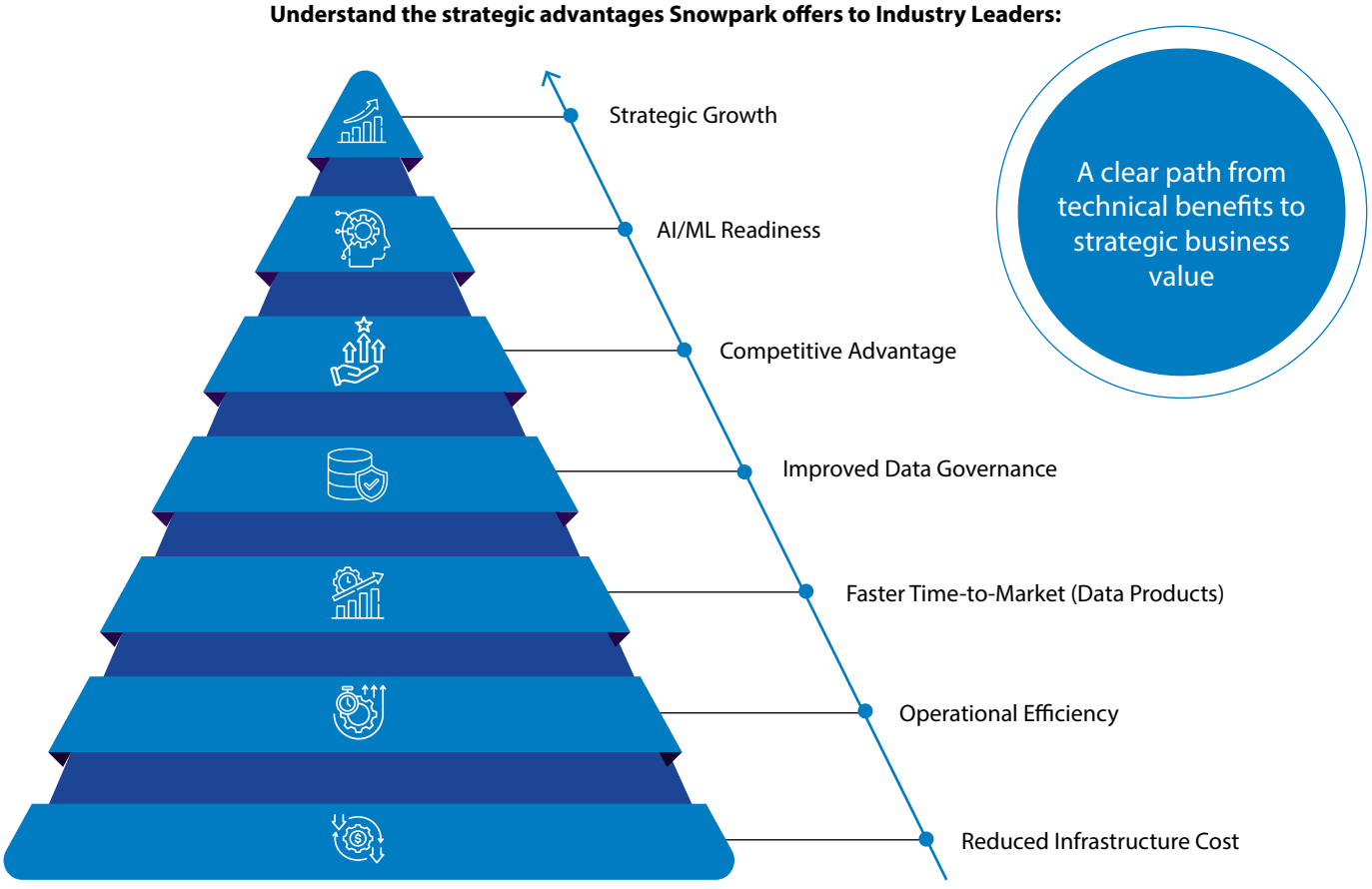
For Industry Leaders: Strategic TCO and Time-to-Market

For CxOs and business leaders, Snowpark is a strategic investment that de-risks modernization and accelerates data product delivery:

Significant TCO Reduction: Beyond infrastructure savings, Snowpark reduces operational overhead by removing the need for dedicated DevOps teams, accelerating development cycles, and minimizing system maintenance. These efficiencies contribute to improved project economics, allowing data teams to focus on innovation and strategic outcomes rather than infrastructure management. This model redefines the total cost of ownership, aligning technology investments more closely with business value.

Future-Ready Data Foundation: Snowpark enables the development of Data-Intensive Applications and Snowflake Native Apps that run directly on Snowflake compute, providing a unified, secure platform ready for AI/ML workloads. It offers a future-ready platform that integrates the power of Apache Spark with Snowflake's scalability, security, and operational simplicity.

Understand the strategic advantages Snowpark offers to Industry Leaders:



Infosys: Your Trusted Partner for Accelerated Snowpark Migration

The shift to Snowpark is a critical step in your data modernization journey. It requires deep technical expertise and a proven strategic roadmap.

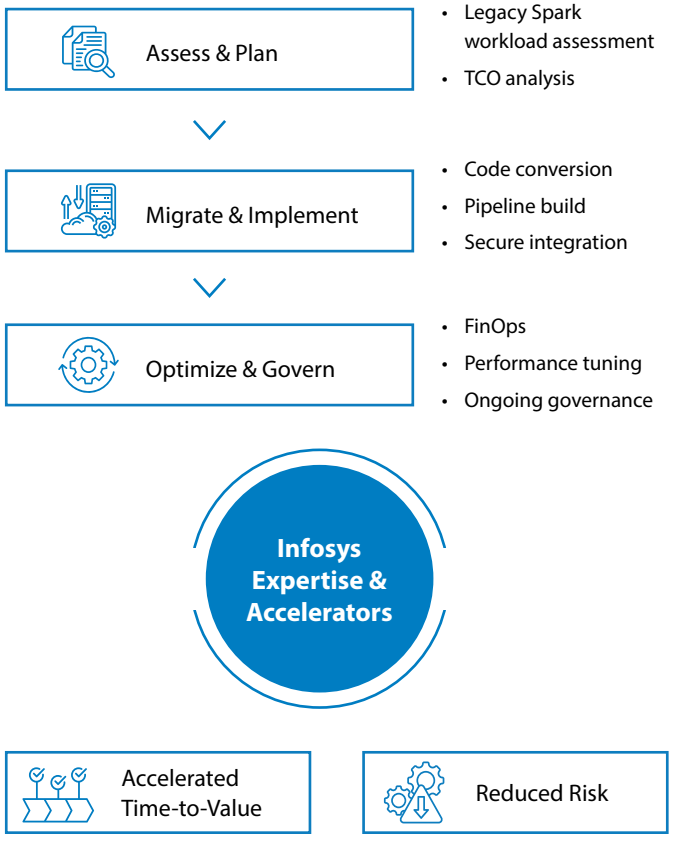
Infosys, as a trusted Snowflake Elite Services Partner, is uniquely positioned to de-risk and accelerate your Snowpark adoption:

Global Scale and Expertise: We operate a global Snowflake Center of Excellence (CoE) with certified architects and specialists experienced in large-scale data and cloud migration across complex enterprise environments.

De-Risked Migration Methodologies: We utilize proprietary frameworks and accelerators, such as the Infosys Data Modernization Utility, to perform rapid Legacy Spark Workload Assessments and execute phased migration planning. This minimizes downtime and ensures a smooth, outcome-driven transition. We focus on warehouse sizing, query optimization, and structured data organization, avoiding traditional Spark tuning techniques in favor of Snowflake's intelligent processing capabilities.

Full Lifecycle TCO Optimization: Our support extends beyond migration. We provide post-deployment FinOps and performance optimization services, ensuring you implement Snowflake's best practices to maximize cost management and sustain the TCO benefits of Snowpark long-term. We leverage built-in tools such as query history, performance metrics, and cost tracking, eliminating the need for external monitoring infrastructure.

Here's an overview of Infosys's approach to Snowpark migration



Snowpark is not just a technology upgrade; it is a strategic inflection point. By partnering with Infosys, you can confidently transition to a unified, scalable, and secure data platform, enabling your teams to focus on innovation that drives real business results.

Ready to transform your data engineering and accelerate your data science capabilities? Contact Infosys today for a complimentary Snowpark Readiness Assessment.

Snowflake Snowpark Connect implementation outcome and impact

With the advent of Snowpark Connect for Apache Spark, Infosys is uniquely positioned to help enterprises modernize their data platforms by eliminating the overhead of managing Spark clusters and enabling native execution of Spark workloads within Snowflake's secure, scalable compute engine.

Infosys brings deep expertise in enterprise data modernization and large-scale Spark implementations, making us a trusted partner for organizations transitioning to cloud-native architectures. Leveraging our proven frameworks, accelerators, and domain knowledge, Infosys leads to the migration of Apache Spark workloads from on-premises infrastructure to Snowflake. This capability enables clients to simplify their data architecture, reduce operational complexity, and accelerate time-to-value—while ensuring performance, scalability, and governance.

Outcomes

Significant Performance Gains: Spark code execution on Snowflake virtual warehouses is expected to deliver up to 5.6× improvement in performance, enabling faster analytics and more responsive decision-making.

Cost Optimization: As per industry benchmarks, organizations can achieve approximately 41% reduction in total cost of ownership (TCO) by eliminating Spark infrastructure and its associated operational overhead—resulting in more efficient resource utilization.

Developer Productivity: Snowpark Connect enables Spark code execution using familiar tools such as PySpark, Jupyter, and Airflow, allowing development teams to maintain existing workflows with minimal code changes—accelerating adoption and enhancing developer velocity.

Authors

Priya Joseph
Partner Solution Engineering,
Snowflake

Dhanus Kodi Paulraj
Principal Technology Architect, Data Analytics and AI,
Infosys

Marutha Pandian Jagadesan
Technology Architect, Data Analytics and AI,
Infosys

For more information, contact askus@infosys.com

Infosys®
Navigate your next