Enabling security, speed, and flexibility for AI data ingestion

Deliver and secure training, fine-tuning, and RAG pipelines with F5 technologies.
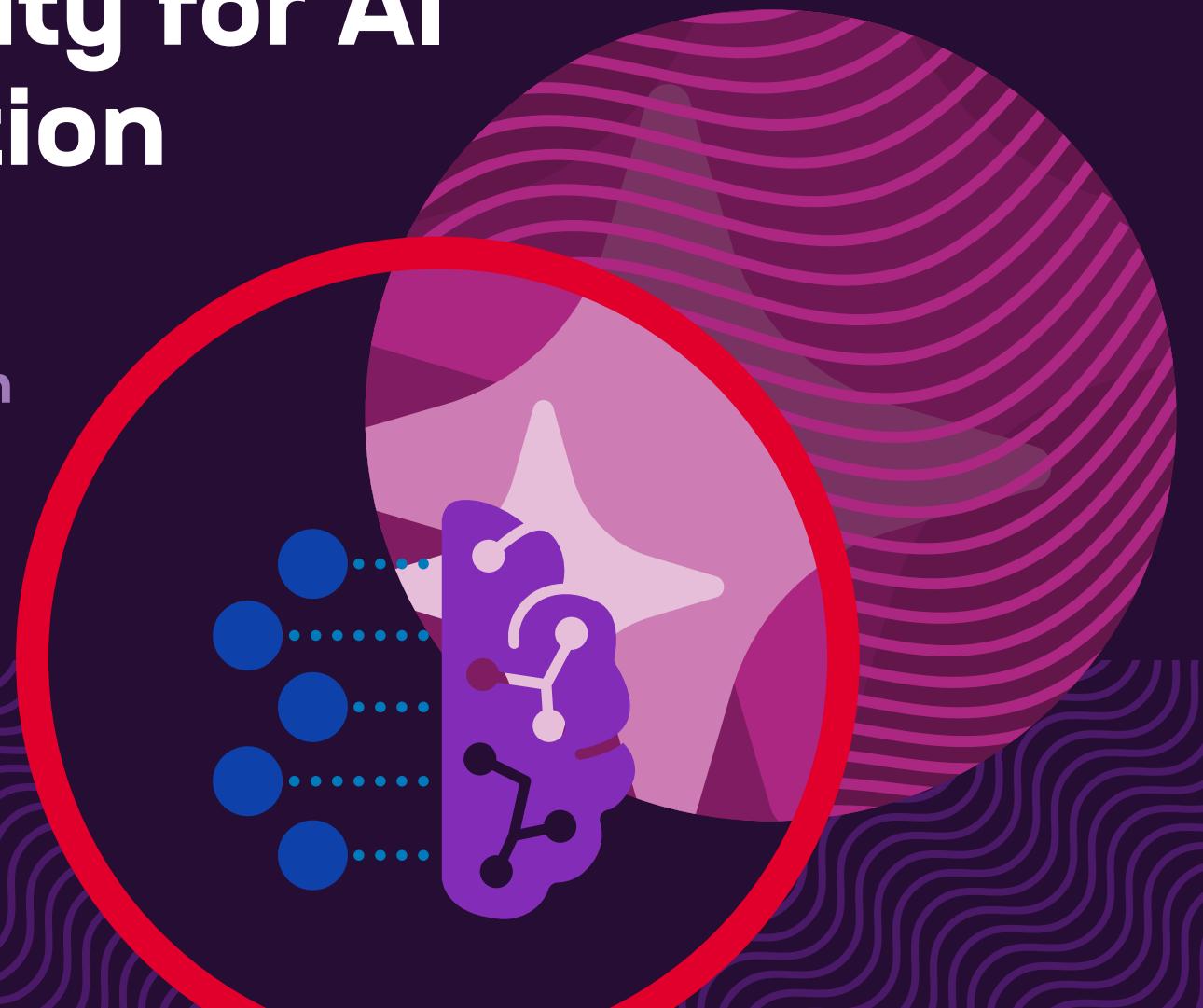
Contents

Ensure your AI pipelines are up to par

Secure, efficient, and scalable data ingestion is essential to enterprise AI initiatives. But many organizations struggle to unlock it, whether they're training from scratch, fine-tuning an existing model, or integrating external data sources via retrieval-augmented generation (RAG).

Today's data ingestion challenges

Inflexible and expensive repatriation and storage changes: Changing vendors, moving between clouds, or adding on-premises capacity requires costly application redesign, lengthy cutovers, and high-risk migrations.

Unchecked data security and compliance risk: Without centralized policy enforcement, privileged or poisoned training data can enter pipelines uninspected compromising the value of model outputs can exit the consequences of not having safeguards. Fines, brand damage, poor customer experience, data breach.

Performance collapse under load: Training jobs, parallel fine-tuning tasks, and RAG index builds can overwhelm storage backends, causing unpredictable ingest speeds, unreliable workload performance, and missed SLA targets.

Zero visibility into data movement: When applications connect directly to object stores, teams lose the ability to see and control who is moving what data, when they're doing it, and where the data is going.

Fragile data availability: Reliance on a single storage endpoint or region means that a failure or slowdown brings data ingestion to a halt—stopping training runs and delaying production timelines.

KEY FACT
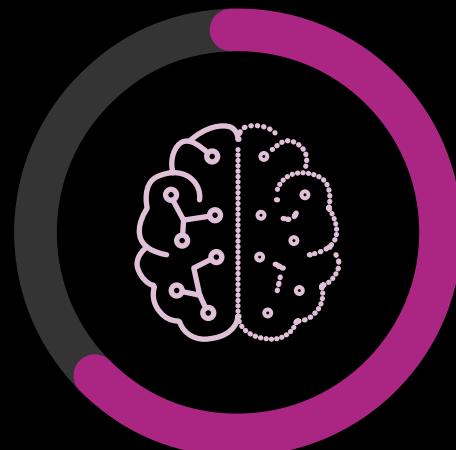
74%

of business leaders say they can't effectively scale AI because of infrastructure costs, disconnected data silos, or slow data ingestion.[1]

Accelerate AI data ingestion and delivery

Powered by F5® BIG-IP, the F5 Application Delivery and Security Platform (ADSP) provides a scalable, secure, and high-performance entry point for ingesting datasets of any size for use in training, fine-tuning, and RAG workflows.

Combining policy-driven security, intelligent traffic management, and resilient failover, the F5 ADSP helps you deliver AI pipelines that are reliable, predictable, compliant, and governable.

Maximize agility, security, and speed with the F5 ADSP

Enhanced flexibility
Uncouple applications from specific storage APIs and endpoints for seamless vendor changes, hybrid cloud expansions, and multicloud operations.

Deep observability and governance
Gain granular telemetry on data movement to support compliance reporting, security audits, and precise cost allocation.

Centralized management
Enable a single control point for all AI data ingress and egress, inspecting traffic and enforcing policies to block data poisoning, prevent unauthorized access, and stop model exfiltration before it happens.

Resilient multi-site availability
Monitor storage endpoints in real time and automatically route traffic to healthy, performant locations, ensuring uninterrupted ingestion even during outages or degradations.

Intelligent traffic optimization
Shape, prioritize, and load balance massive dataset transfers to maximize throughput, minimize latency, and enhance object store reliability.

Programmable delivery policies
Use iRules and automation to dynamically adjust routing, security controls, and performance parameters in response to changing workload patterns and business priorities.

# Securely power your AI use cases

Training

- Ensure uninterrupted access to training datasets by monitoring storage health and automatically routing to available, performant endpoints with F5 BIG-IP Local Traffic Manager (LTM) and F5 BIG-IP DNS.

- Enforce secure, policy-driven ingress using F5 BIG-IP Advanced WAF to protect against poisoned training data and ensure compliance with data governance requirements.

- Optimize data movement across hybrid and multi-cloud storage and avoid bottlenecks via BIG-IP LTM.
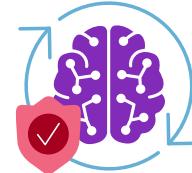
RAG

- Deliver low-latency ingestion of knowledge corpus data to keep RAG responses accurate, up-to-date, and responsive using F5 rSeries, F5 VELOS, and BIG-IP LTM.

- Implement egress safeguards via BIG-IP APM to prevent the unauthorized retrieval of sensitive data during augmentation processes.

- Unlock deep observability into ingestion patterns, enabling optimization of query performance and governance of dataset usage.

Fine-tuning

- Maintain predictable fine-tuning throughput by shaping and balancing traffic across storage backends via BIG-IP LTM and BIG-IP DNS.

- Block unauthorized access to sensitive datasets with F5 BIG-IP Access Policy Manager (APM).

Transfer massive datasets with confidence

F5 delivers robust distributed data transport to facilitate the seamless movement of AI training data across scalable infrastructures. By leveraging adaptive traffic management and intelligent routing, the F5 ADSP enhances data transfer efficiency and reliability. Whether for training, fine-tuning, or RAG data ingestion, your organization can take advantage of F5 capabilities to minimize bottlenecks, preserve data sovereignty, and cut operational costs.

Secure inputs and protect data movement

The F5 ADSP equips your organization to move sensitive data securely and safeguard the inputs that are essential to your AI outcomes:

Mitigate data poisoning risks with advanced data classification partnerships and behavioral defense mechanisms.

Enforce centralized policy-based inspection at all points of ingress to identify and block malicious or corrupted datasets.

Apply end-to-end encryption so your organization can maintain compliance while optimizing workloads.
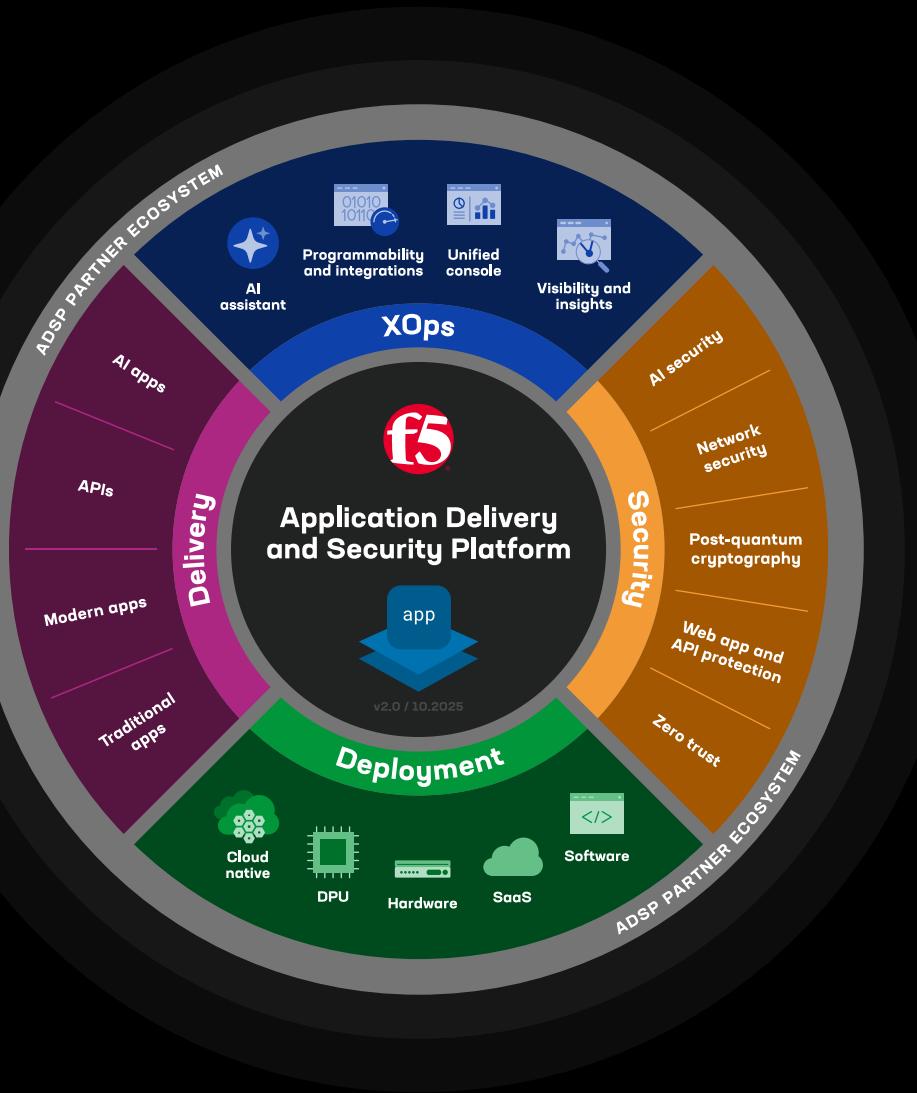
Strengthen resiliency and ensure performance

The F5 ADSP enables distributed loose coupling to separate applications from data layers and enhance flexibility across today's evolving storage solutions. By utilizing intelligent DNS and reverse proxy technologies, backend changes such as data repatriation or storage provider migration occur without disrupting applications. This approach strengthens endpoint resilience and reliability while safeguarding performance and ensuring regulatory compliance. The end result is a dynamic infrastructure suited to changing AI demands.

By taking advantage of F5 capabilities, your team can deliver uninterrupted AI pipeline availability while simplifying storage migrations. Through BIG-IP LTM and F5 BIG-IP DNS, the F5 ADSP dynamically redistributes traffic to healthy endpoints, providing failover and load balancing that prevent disruptions during ingestion and delivery. Your organization can also use F5 technologies to eliminate costly, high-stakes transitions by incrementally repatriating AI workloads from public cloud environments to on-premises resources.

Simplify migrations with F5

Eliminate vendor lock-in and adopt the right storage deployments at the right time

Easily add on-premises capacity without application redesign

Embrace multicloud architecture with less friction and enhanced performance

Streamline operations and maximize efficiency

Take advantage of unparalleled programmable traffic management capabilities to deploy customizable, policy-driven routing for AI data delivery. F5 empowers your team to fine-tune traffic flows and meet diverse processing and performance requirements.

Through advanced programmability, you can ensure scalability, operational efficiency, and cost optimization for heterogeneous AI deployments. Whether you're supporting on-premises, cloud, hybrid, or Kubernetes-based systems, F5 adapts to your requirements without compromising security.
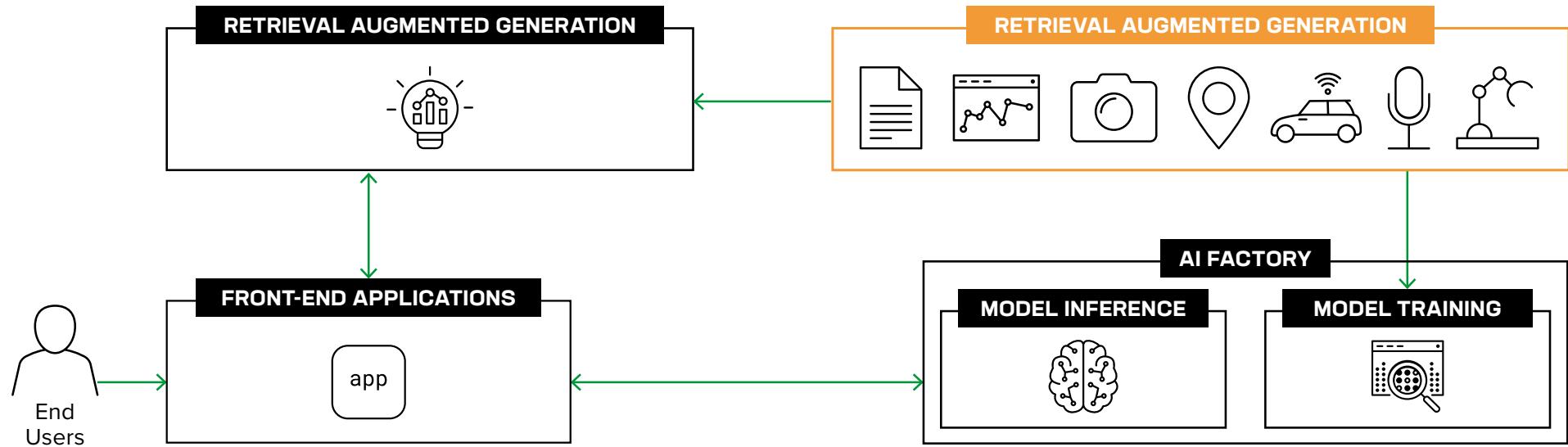

Figure 1: The F5 AI Reference Architecture provides a complete toolkit for protecting AI models, datasets, and interactions

Interested in taking a deep dive?

Explore the interactive AI reference architecture to simplify and scale AI deployments with best practices, security insights, and tools for hybrid multicloud innovation.

Accelerate AI ROI with F5

To enable secure AI outcomes with maximum efficiency, enterprises need to strengthen, streamline, and accelerate data ingestion for training, fine-tuning, and RAG. The F5 ADSP provides intelligent protection, simplified flexibility, enhanced resiliency, and real-time optimization to help you navigate today's data management challenges and power successful AI initiatives.

Learn more about how you can accelerate and protect your AI data ingestion workflows.

Deploy the customized intelligent features you need
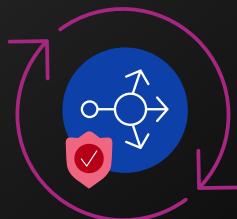
F5 iRules

Your team can use F5 iRules to enable highly customizable traffic routing, including tailored security controls and performance optimization in hybrid, multi-cluster, or AI-enabled storage environments. From shaping traffic flows to enforcing complex access policies, F5's programmable architecture empowers enterprises like yours to meet diverse workload demands with surgical precision.

Appendix

[1] Qlik, _AI Is Core to Strategy for 86%—But Most Are Stuck in Data Complexity_, May 2025.

ABOUT F5

BRINGING A BETTER DIGITAL WORLD TO LIFE

F5, Inc. (NASDAQ: FFIV) is the global leader that delivers and secures every app. Backed by three decades of expertise, F5 has built the industry's premier platform—F5 Application Delivery and Security Platform (ADSP)—to deliver and secure every app, every API, anywhere: on-premises, in the cloud, at the edge, and across hybrid, multicloud environments. F5 is committed to innovating and partnering with the world's largest and most advanced organizations to deliver fast, available, and secure digital experiences.

Together, we help each other thrive and bring a better digital world to life.

For more information, go to f5.com.

Ready to scale your AI applications?

Start the journey at **f5.com/ai**