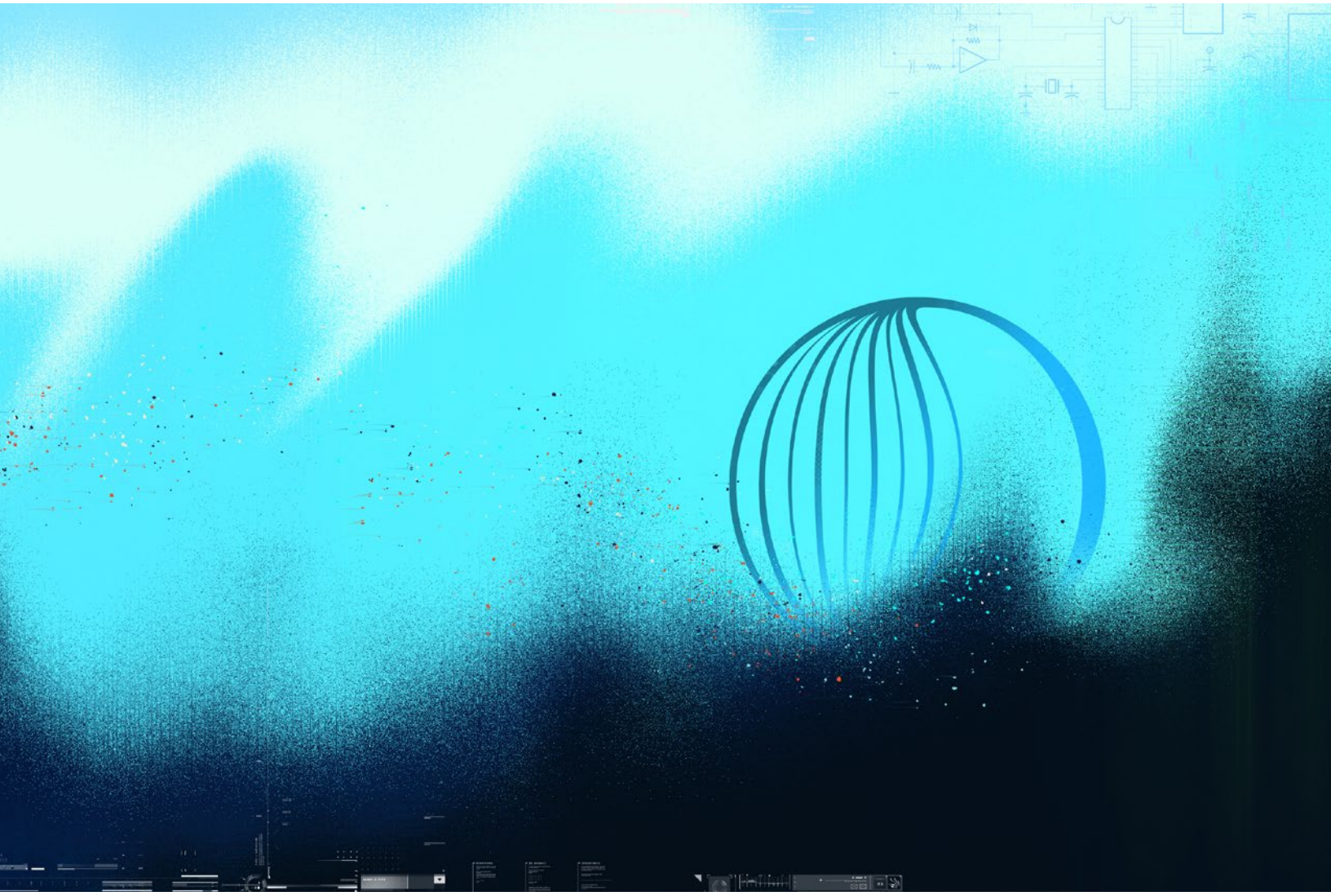
signal65



EXECUTIVE SUMMARY

Data Preparation for Enterprise AI

Compute-Ready Data with TopicLake Insights and Dell AI Infrastructure

AUTHORS

Matthew Goldensohn
Performance | Signal65

Brian Martin
AI Data Center Performance | Signal65

JANUARY 2026

IN PARTNERSHIP WITH

DELL Technologies

GADGET SOFTWARE

Introduction

As enterprises accelerate the deployment of generative and agentic AI systems, the limiting factor has shifted from model capability to data readiness. Many organizations have heavily invested in large language models and GPU infrastructure but continue to provide these systems with raw or minimally processed documents. This results in inconsistent output quality, hallucinations, security risks, unpredictable latency, and rising operational costs. The key to successful enterprise AI is the ability to transform unstructured data into **compute-ready assets** that are secure, governed, and optimized for AI.

On-premises data preparation, implemented with Gadget Software's TopicLake Engine on purpose-built Dell AI infrastructure with Broadcom high-performance networking, creates a robust foundation for enterprise AI. By applying semantic reasoning at ingestion and persisting that intelligence as metadata, organizations enhance AI accuracy, performance predictability, energy efficiency, and return on investment. This approach elevates input documents into reusable, governed digital assets that support AI applications, business intelligence, and operational systems.

Key Highlights

Data Sovereignty
Complete control over document processing, model selection, and access governance

Higher Token Throughput
On-premises compute-ready document generation on Dell platform outperforms public cloud alternatives

Energy Efficiency
Lower power consumption per token and reusable source data across applications

A Data Preparation Imperative

AI systems are only as effective as the data they consume. Raw documents such as PDFs, reports, and emails are poorly suited for computational reasoning due to inconsistent formats and unclear structure. Such issues force AI systems to repeatedly "re-understand" documents, leading to hallucinations, poor citations, higher latency, and increased security risks.

Compute-ready documents resolve these problems by transforming unstructured content into **structured, semantically dense, governed data objects**. During ingestion, analytical reasoning is applied, and each document is decomposed into topic-aligned units and enriched with summaries, keywords, entities, and validated question-and-answer pairs. Governance metadata—including security classification, access controls, and citation lineage—is inherited by every derived artifact, resulting in a data asset that AI can consume efficiently.

Why On-Premises Data Preparation Matters

The location of data transformation is strategic. While cloud services offer convenience, they introduce latency variability and potential security risks. On-premises pipelines directly address these issues.

The Anatomy of a Compute-Ready Document

On-premises solutions ensure that proprietary and regulated data remains under organizational control. Companies can enforce zero-trust principles, align workflows with internal risk policies, and maintain an audit-ready chain of custody.

Deterministic Performance and Throughput

Purpose-built platforms like the Dell PowerEdge XE7745 provide the compute density necessary for efficient document transformation. High-memory, multi-GPU configurations allow for massive parallelism, while Broadcom 400 GbE networking guarantees low-latency data movement. In contrast to cloud environments, on-premises solutions maintain predictable processing times regardless of document complexity.

Energy Efficiency and TCO

Local processing eliminates repetitive data transfers and API calls common in cloud workflows. By transforming documents once and reusing enriched semantic twins, enterprises reduce energy consumption and operational costs. Though on-premises solutions require initial investment, they lead to quicker ROI and enhanced sustainability.

Compute-Ready Data - A Strategic Advantage

The insights in this report converge on one conclusion: **compute-ready data is essential for enterprise AI readiness**. On-premises infrastructure allows for higher token performance and consistent latency while reducing third-party data exposure and energy usage. This transforms unstructured data into a strategic asset. Organizations adopting a topic-oriented data lake architecture can further benefit by normalizing diverse content into enriched semantic twins stored in a governed, write-once-read-many repository, creating a shared intelligence layer that serves AI systems, analytics, and operational tools.

Conclusion

The success of enterprise AI initiatives relies more on solid data preparation than on the latest model. On-premises compute-ready document generation offers advantages in sovereignty, performance, energy efficiency, and cost predictability. TopicLake Engine running on Dell PowerEdge infrastructure, combined with Broadcom high-performance networking, provides the necessary computational density, memory capacity, and reliability to scale AI operations.

Enterprises that invest in compute-ready data architectures can adopt future AI capabilities without re-engineering their pipelines. Those that don't will face hidden costs from unreliable outputs, security risks, and inefficiencies. In the age of generative AI, competitive advantage goes to organizations that understand that **AI systems are only as valuable as the data they utilize**.

Important Information About this Report

CONTRIBUTORS

Matthew Goldensohn
Performance | Signal65

Brian Martin
AI Data Center Performance | Signal65

PUBLISHER

Ryan Shrout
President and GM | Signal65

INQUIRIES

Contact us if you would like to discuss this report and Signal65 will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "Signal65." Non-press and non-analysts must receive prior written permission by Signal65 for any citations.

LICENSING

DISCLOSURES

Signal65 provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

ABOUT SIGNAL65

Signal65 is a leading research organization specializing in enterprise AI infrastructure optimization and deployment strategies. Our lab focuses on evaluating and optimizing AI hardware and software solutions for real-world enterprise applications, with particular expertise in large language models, retrieval-augmented generation systems, and distributed AI architectures.

For more information, visit signal65.com
or contact research@signal65.com

IN PARTNERSHIP WITH

Read the full report on the Signal65.com website.

CONTACT INFORMATION
Signal65 I signal65.com