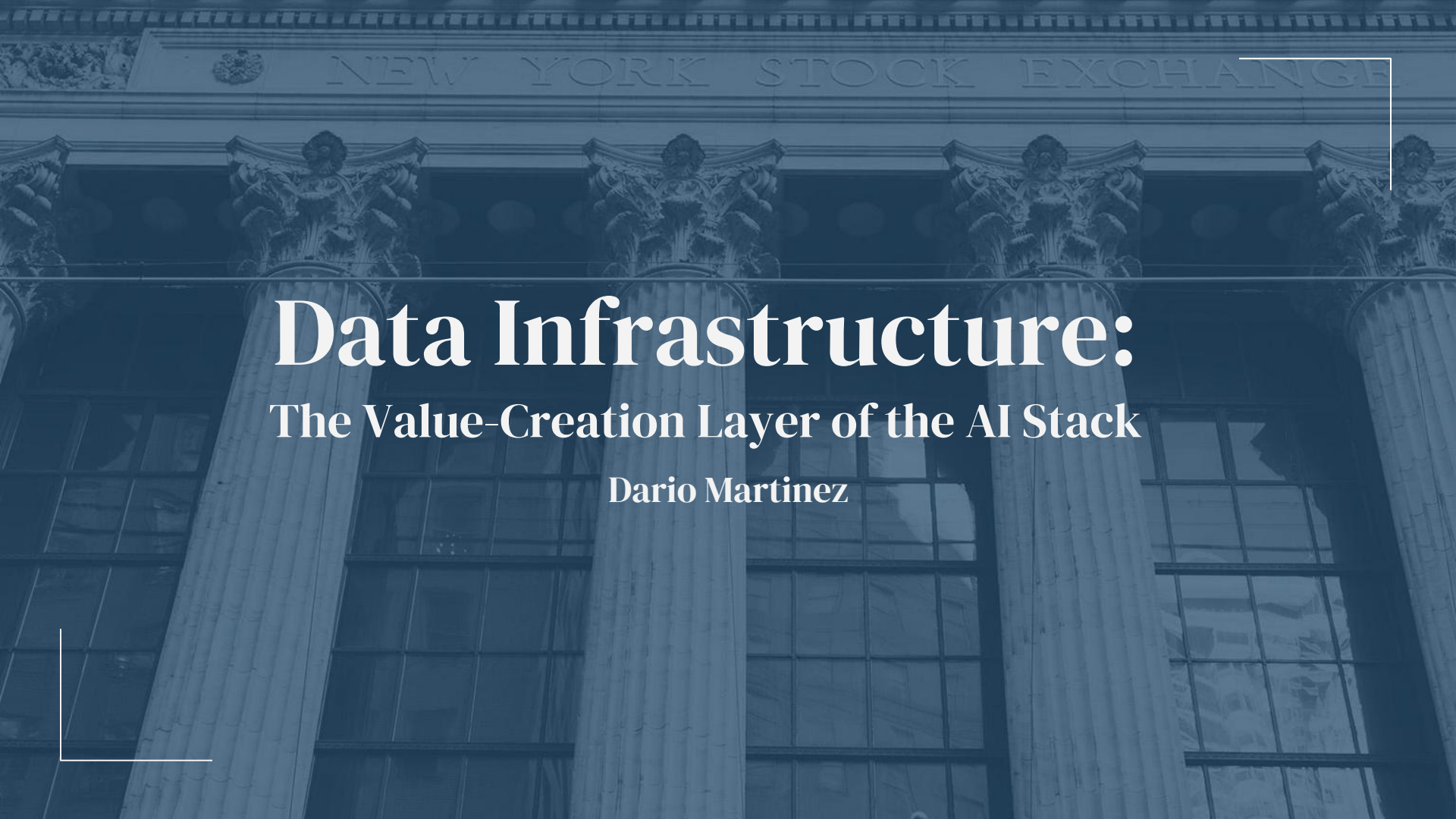


Data Infrastructure:

The Value-Creation Layer of the AI Stack

Dario Martinez

Executive Summary

What is Data Infrastructure?	The software layer above where data is physically stored, turning raw information into reliable, actionable and enterprise usable data
Why This Matters?	Allows companies to make decisions quicker and automate work, improving margins and converting AI spend into tangible results
M&A Implications	Demand for fewer vendors is accelerating consolidation in a fast growing market, concentrating value and creating winners and losers similar to the dot com era

Path From Data Creation Into Actionable Insights

Data is the New Asset and Enterprises Generate Massive Volumes

Generating more raw information across operations, customers, and machines as digital and AI adoption accelerates

Digital Infrastructure Physically Stores and Moves the Data Created

Provides the compute, storage, and connectivity required to house data and make it accessible at scale

Data Infrastructure Converts Inputs Into Usable State

Standardizes data, controls access, and prepares it for enterprise systems

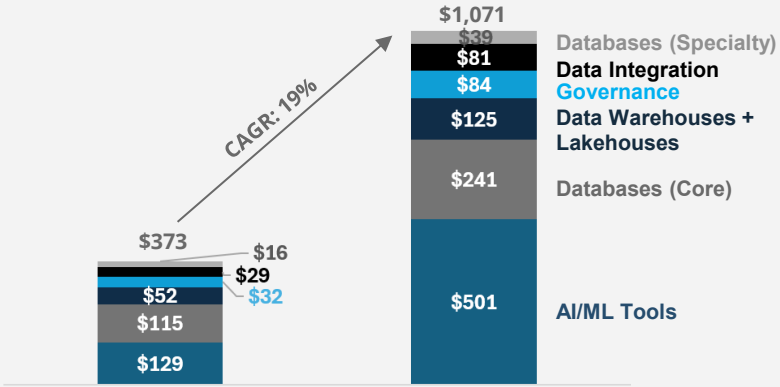
Modern Data Stack Creates Control Points With Sub-Verticals

Sub-Vertical	What is the Problem Solved?	How Does This Create Value?
Ingestion	• Moves data from original systems into centralized storage	• Feeds downstream teams with current data without need of re-platforming • Becomes anchor for tools to depend on
Storage and Lakehouse	• Centralizes structured/unstructured data into a single repository for workloads	• Supports scale without linear cost increase • Preserves data as long-term asset
Transformation	• Cleans, models, and prepares raw data into analytics ready datasets	• Eliminates constant reconciliation of data • Makes outputs repeatable by defining information to be used later in querying
Query and Complete	• Allow teams quickly analyze large amounts of data at scale	• Increases data usage across enterprise • Converts stored data into faster real-time business insights
AI enablement	• Makes data usable for machine learning models and AI applications	• Shortens gap between data access and dependable LLM operations • Reduces failure rates and hallucinations
Governance	• Ensures regulatory compliance through safe access to enterprise data	• Data sharing at large scale • Reduces legal and security exposure

Market Overview

Market Size
($ in Billions)

Businesses rely more on data to operate and compete, driving strong market growth



90x Increase in Data Created Annually from 2010 to 2025	**80%** Of Global Data Created by Enterprises by 2028 Versus 60% in 2023	**$496.8B** By 2035 Data Center Market Is Projected to Expand Driven by Cloud Computing Demand
60% Of Organizations Are Limited by Data Silos in Deriving Value From Their Data	**3.5x** From 2025 to 2030 AI Workloads Require Increase in Gigawatt Supply	**$5.2T** Required CapEx Investment by 2030 to Support AI Compute Infrastructure

Themes Driving Investment

Platform Consolidation

- Companies consolidate around fewer platforms to simplify operations, lower total cost, and speed execution across teams
- Platforms that own multiple parts of the workflow expand in relevance and spend as usage grows across the organization

Trust, Control, and Reliability at Scale

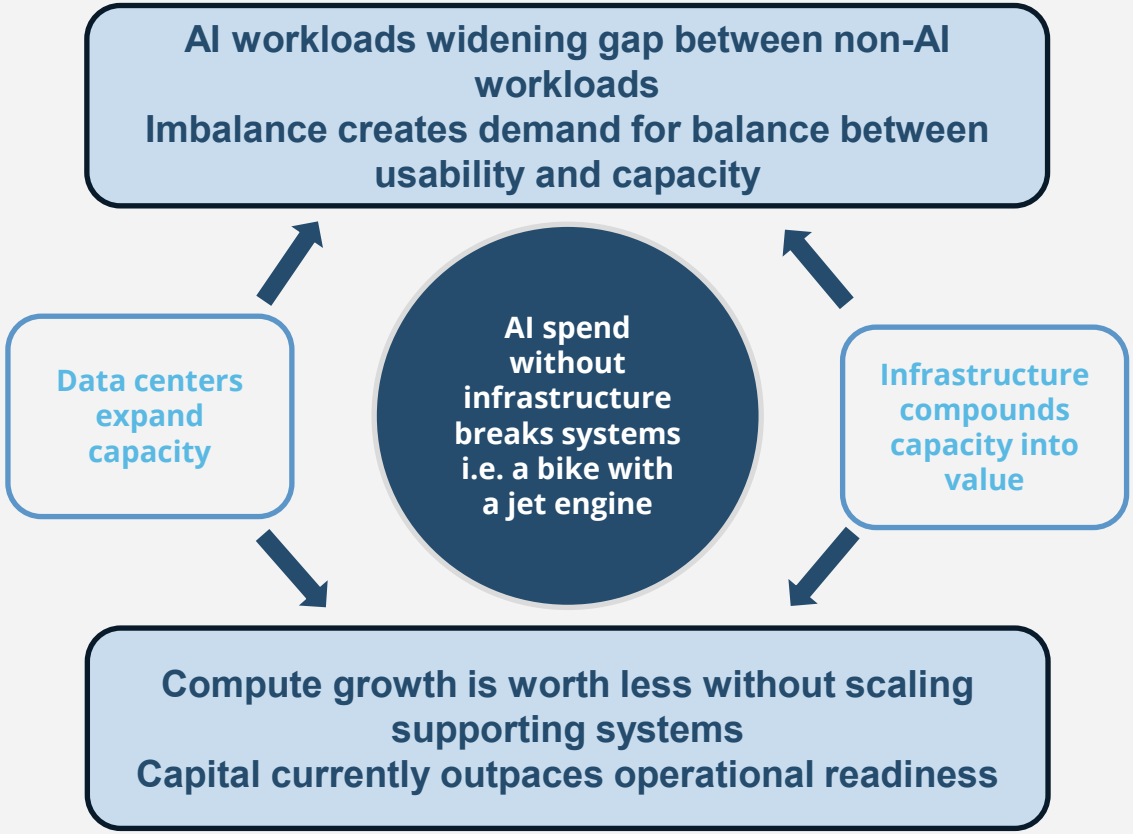
- Consistent data enables dependable reporting and decisions that directly support revenue and customer experience
- As data supports more automated workflows, companies require systems that enforce accuracy, permissions, and uptime by design
- Platforms embedded as the source of trusted data become harder to replace and expand in value as business reliance increases

AI Enablement

- Isolated datasets waste resources by lacking consistency, limiting model accuracy, increasing hallucination risk
- Clean data enables repeatable results, reliable inference, and scale from experimentation into operational workflows that reduces labor cost and time
- Platforms above AI workflows capture more activity as companies move from pilots to core operations

Sources: Houlihan Lokey, McKinsey, Yahoo Finance

Sustained AI CapEx Forces Infrastructure Scaling



Company	2025 CapEx	YoY Inc. %
Alphabet	$82.4B	57%
amazon	$117.0B	41%
	$12.7B	35%
Meta	$72.2B	94%
Microsoft	$91.3B	42%

Sources: Bloomberg, CNBC, SEC

Current Headwinds for AI Investment

Key Risks

Heavy CapEx Requirements Don't Justify Returns

- AI infrastructure requires asset-heavy upfront spend, while only a few companies have realized tangible earnings stemming from AI spend
- Present day investment case is underwritten on expected demand growth rather than proven workloads

Energy and Grid Restraints Obstacles to Growth

- Power demand exceeds existing grid, creating strain and limiting development
- Energy projects require long lead times, increasing risk as tech companies become involved in energy sourcing

Financial Engineering and Transparency Risk

- SPVs and non-recourse financing limit risk, reduce visibility into debt exposure and return profiles for investors

Why Long-Term Fundamentals Remain Intact

Earnings Lag Adoption in Every Compute Cycle

- Early buildouts show weak near-term profitability as capacity is built ahead of demand, with returns improving as utilization scales and workloads mature

Capital Constraints Drive Consolidation

- High capital requirements favor well-funded platforms, accelerating consolidation as undercapitalized vendors struggle

Infrastructure Bottlenecks Attract Policy and Capital

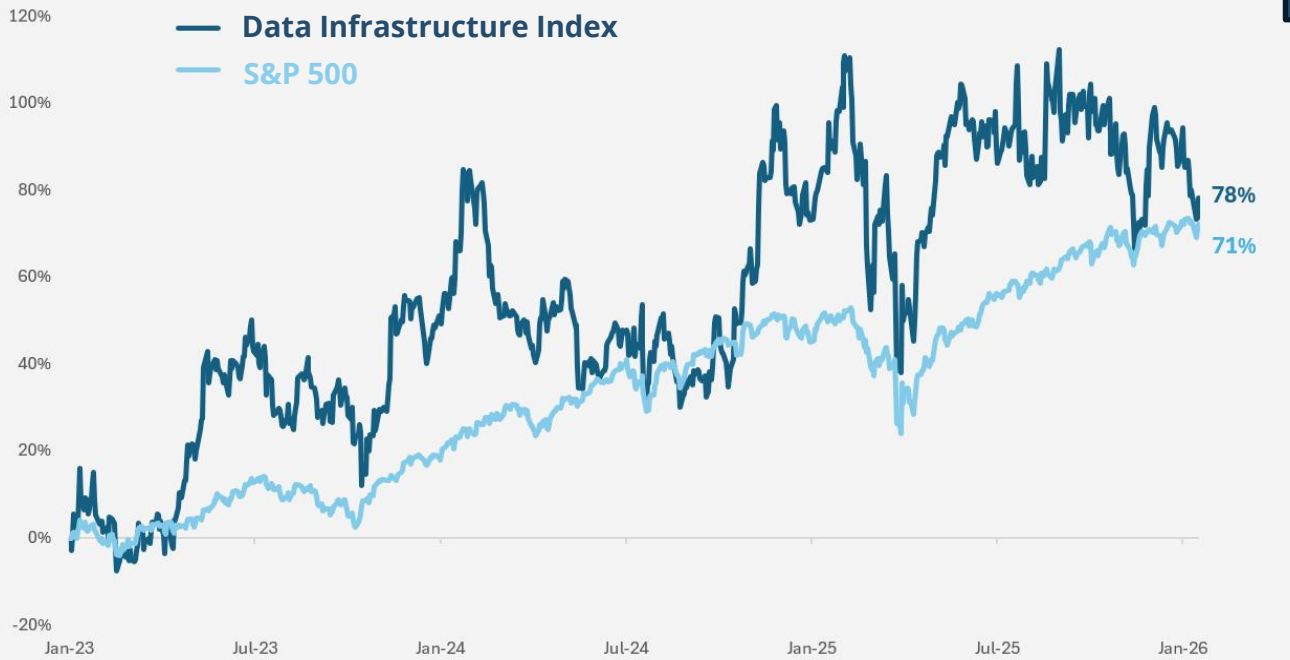
- Power constraints historically trigger regulation support, private capital deployment and accelerated investment

Risk Properly Allocated With Sophisticated Buyers

- Non-recourse financing isolates project risk, underlying returns are still supported by creditworthy parties

Sources: Bloomberg, CNBC, WSJ

Data Infrastructure Tracks the S&P With Higher Volatility



Data Infrastructure

Commvault
CONFLUENT
elastic
mongoDB
rubrik
snowflake

Sources: S&P Capital IQ as of January 23, 2026. Note: Indexed at 0 as of January 23, 2023

M&A Activity and Public Market Valuations

Platform Expansion

May 2025

databricks

Acquired

NEON

TV: $1.0B | TV/Rev: ND

May 2025

salesforce

Acquired

Informatica

TV: $8.1B | TV/Rev: 4.97x

Vertical Data and Operational Software

Aug. 2024

PLAYGRAM

Acquired

MDSTECH

TV: $3.7M | TV/Rev: 0.78x

March 2025

EMERSON

Acquired

aspentech

TV: $7.2B | TV/Rev: 14.84x

Sponsor-Led Consolidation

Oct. 2024

THOMABRAVO

Acquired

DARKTRACE

TV: ND | TV/Rev: ND

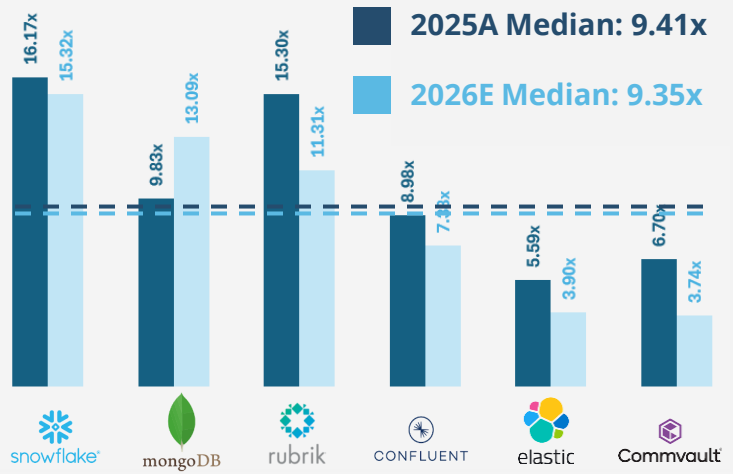
Jan. 2025

Qlik

Acquired

upsolver

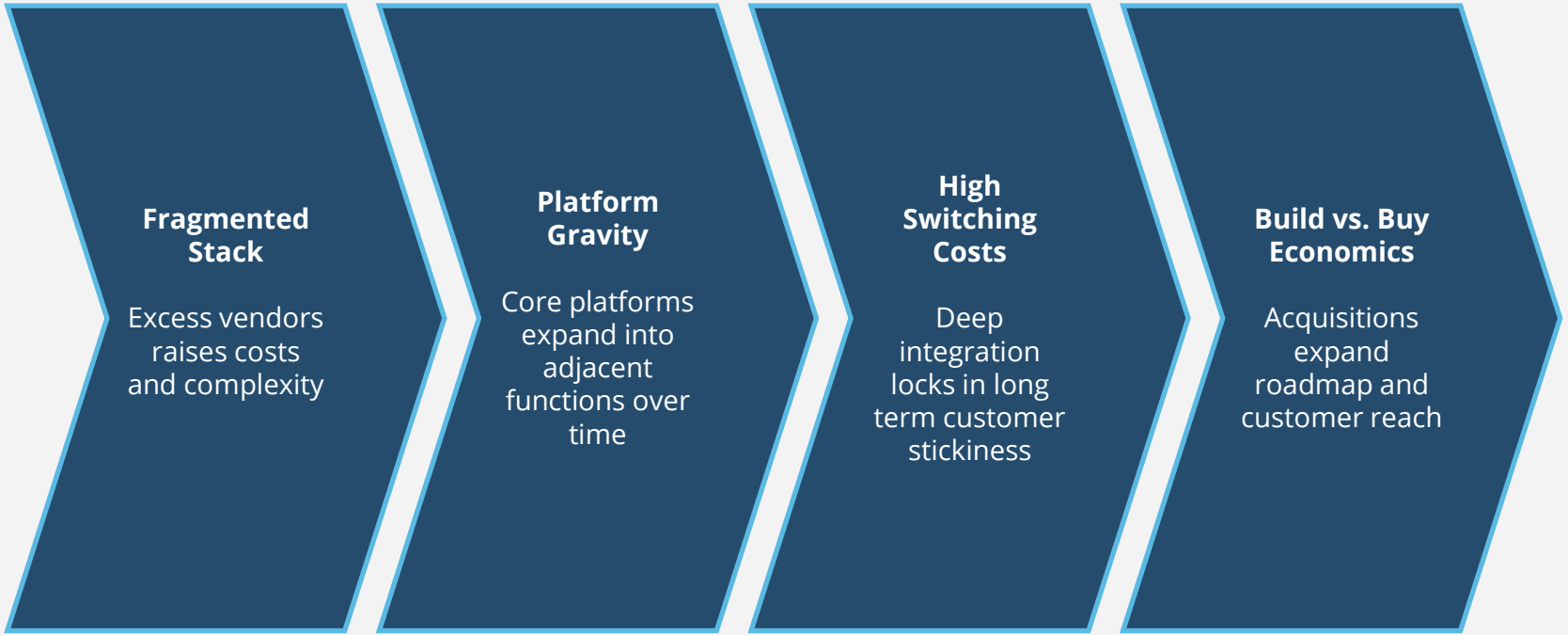
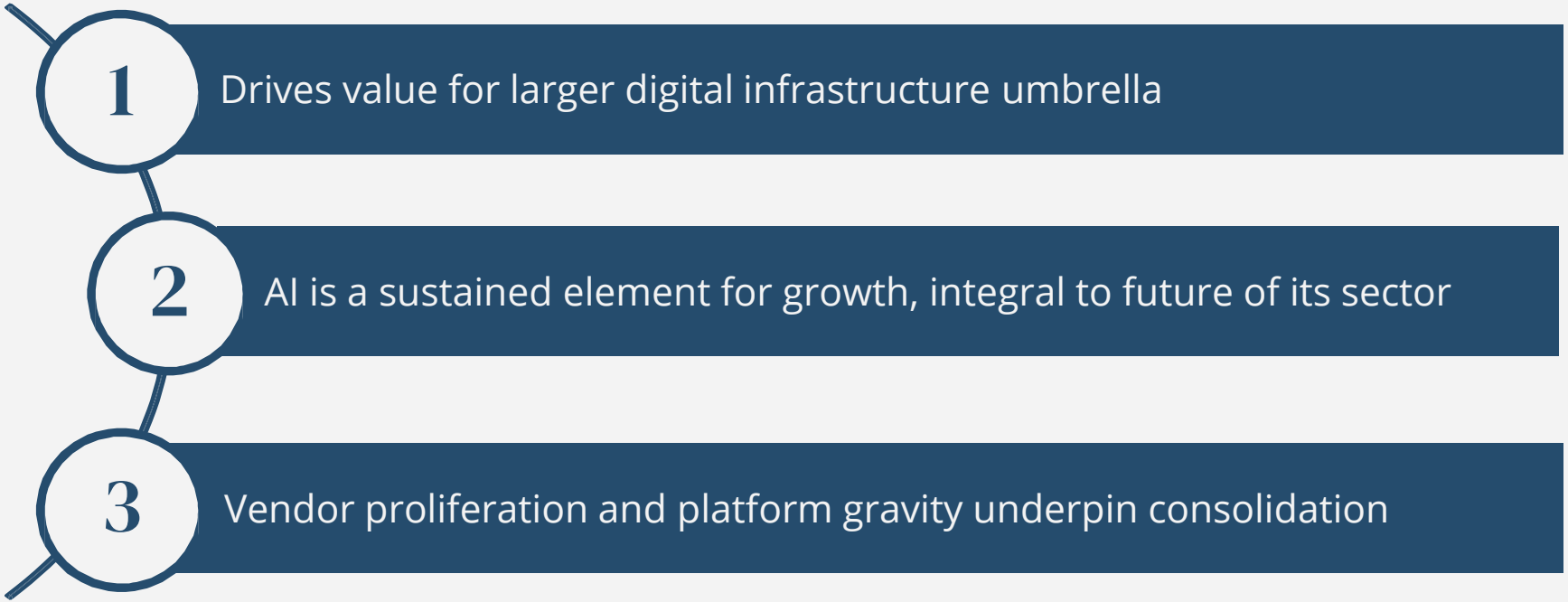
TV: $5.1B | TV/Rev: 9.36x



		snowflake	mongoDB	rubrik	CONFLUENT	elastic	Commvault
Rev Growth	25A	29%	19%	41%	22%	17%	19%
	26E	28%	22%	44%	17%	16%	17%
EV / Rev	25A	16.17x	9.83x	15.30x	8.98x	5.59x	6.70x
	26E	15.32x	13.09x	11.31x	7.38x	3.90x	3.74x
EV / EBITDA	25A	NM	NM	NM	NM	NM	63.61x
	26E	110.26x	66.30x	NM	63.52x	23.27x	18.66x

Sources: Bloomberg

M&A Creates Path Towards Integrated Platforms



Data Infrastructure: Key Takeaways