



1H 2026 Data Intelligence, Analytics, & Infrastructure Market Sizing & Five-Year Forecast Report

From Data Technician to AI Shepherd: The Shift to Production

Analyst(s): Brad Shimmin

Published: January 2026

Document #: 1H26MSFDAIBS202601

A global study of the nine sub-markets within the Data Intelligence, Analytics, & Infrastructure market.

This is a companion report to the entire dataset within Futurum's interactive visualization dashboard, where users can filter and drill down further into the study details - see app.futurumgroup.com.

Key Findings

- **Market Acceleration:** The global Data Intelligence, Analytics, & Infrastructure (DIAI) market is projected to reach US\$541.1 billion in 2026, growing at a 16.9% CAGR to surpass US\$1.2 trillion by 2031. This acceleration is driven by the mass transition of GenAI from experimental pilots to production-grade Agentic workflows, necessitating a significant re-platforming of the enterprise data estate.
- **The Data Technician to AI Shepherd Shift:** As Zero-ETL and automated pipelines commoditize manual data engineering (growing at a slower 12%), value is migrating rapidly to that of agentic oversight.
- **The Semantic Layer (growing 19% in 2026) and Data Observability (growing 22%)** have emerged as the critical control planes for preventing AI hallucinations and ensuring more deterministic outcomes.
- **The Two-Speed Economy:** A clear structural divergence has emerged between industries. Digital Native verticals such as Financial Services and Communications, Media & Tech (CMT) are accelerating immediately, capturing early ROI from automated code generation and process automation. In contrast, Physical verticals such as Resources and Utilities face a protracted integration curve, with their acceleration back-loaded to 2028+ as they tackle the complexity of converging IT with Operational Technology (OT).

Recommendations

To optimize their chance for success in this vast-ranging, competitive, fast-moving market, enterprise data, analytics, and AI practitioners should:

1. **Prioritize the Semantic Layer:** Enterprise data leaders must immediately treat the Semantic Layer as critical infrastructure, not just a BI user experience. Without a unified metric layer to ground LLMs, autonomous agents will fail to scale due to data ontology ambiguities and knowledge gaps, leading to model hallucinations. The Semantic Layer is the only mechanism to ensure that an AI agent defines Gross Margin the same way a human CFO does.
2. **Shift Budget to Day 2 Ops:** Organizations should reallocate budget from traditional Day 1 ingestion/storage maintenance to Day 2 operational tools such as data & AI observability. As data-hungry agentic systems go into production, the primary cost center shifts from building pipelines to monitoring semantic and data drift as well as performance and reliability. Companies must stop buying shelfware for ingestion and instead move 20% of the data engineering budget into data observability. If you can't detect a schema drift in 5 minutes, your autonomous agent may just make a million-dollar pricing error. In short, you cannot manage what you cannot see.

3. Adopt Composable Intelligence: Move away from monolithic black box platforms toward a composable intelligence stack based on open standards (e.g., Apache Iceberg, Model Context Protocol). This future-proofs the architecture against vendor lock-in and allows for the rapid swapping of model providers (e.g., swapping OpenAI for Anthropic) as the model landscape evolves.

TABLE OF CONTENTS

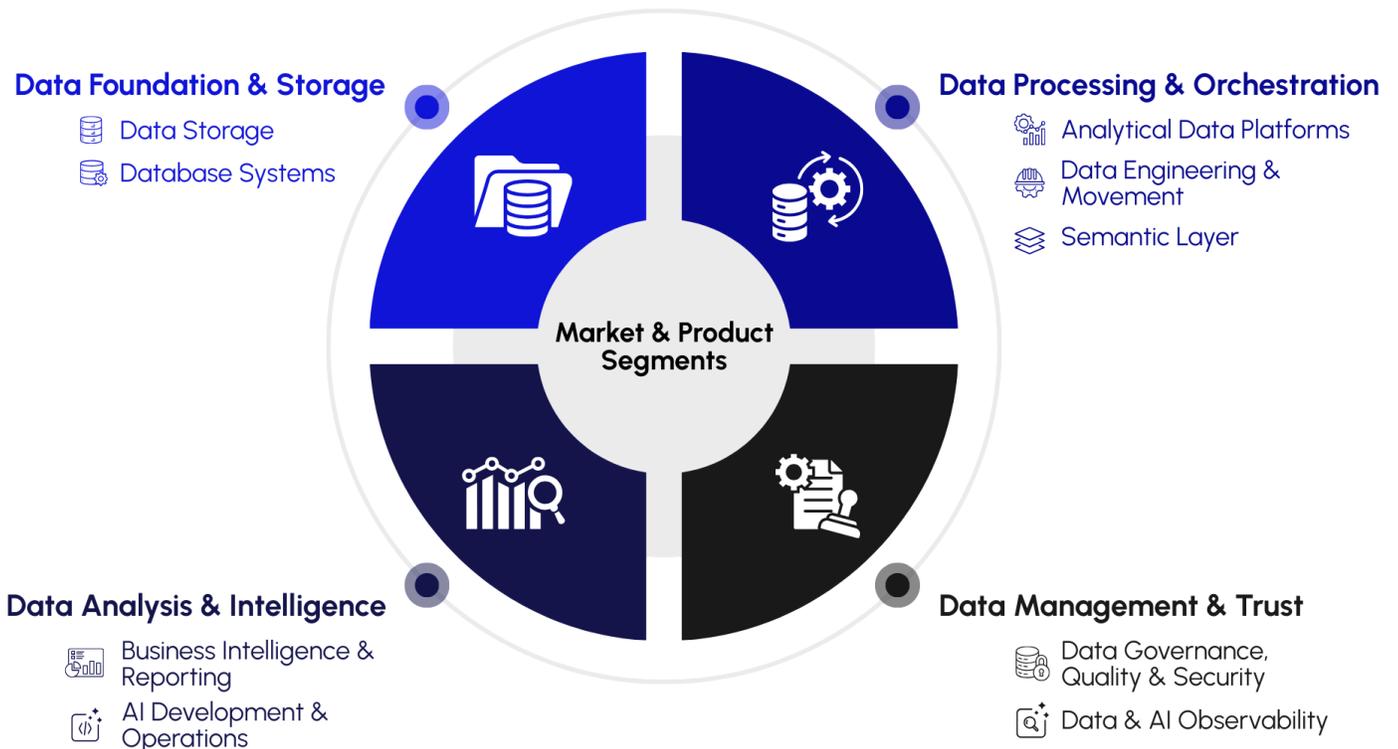
Key Findings.....	2
Recommendations.....	2
Market Model.....	4
Market & Product Segments.....	4
Industry Coverage.....	8
Regional Coverage.....	9
Executive Summary.....	9
Strategic Forecast Scenarios (2025–2031).....	10
Market Analysis: Sub-Segment View.....	11
Market Analysis: Regional View.....	12
Market Analysis: Vertical View.....	14
Methodology.....	16
About Us.....	17
About the Authors.....	17
About The Futurum Group.....	17
Copyright & Use License.....	17
Copyright Notice.....	17
License Notice.....	17
Limitation of Liability Notice.....	17

Market Model

Futurum defines the Global Data Intelligence, Analytics, and Infrastructure market not merely as a collection of tools, but as the converged operating system of the modern, AI-driven enterprise. It encompasses the end-to-end architecture and tooling required to turn raw digital assets into autonomous action.

This ecosystem has graduated from the often monolithic yet fragmented traditional data stack into a unified data intelligence platform, spanning the full lifecycle of data: from the immutable storage & foundation layer (now optimized for high-throughput AI workloads and vector search), through the processing & orchestration engines (powered by open lakehouse table format standards), to the high-level analysis & intelligence layers where human-derived insights and Agentic AI automation converge. Underpinning this entire stack is a management & trust layer that serves as the critical control plane, ensuring governance, security, and observability in an era where data quality determines AI value as well as safety.

Market & Product Segments



Data Foundation & Storage

- **Data Storage:** The immutable bedrock of the data stack, encompassing the raw physical and cloud infrastructure (Object, Block, and File) required to persist digital assets. While historically viewed as a dumping ground for commodities, this layer has evolved into a high-performance tier critical to AI. Modern object storage is no longer just for archiving; it is now optimized for high-throughput concurrency to feed hungry GPU clusters and data lakehouses.
 - Positioning/Vendors:
 - Cloud/Hyperscale: AWS S3, Azure Blob/Disk, Google Cloud Storage
 - Enterprise Infrastructure: Dell PowerScale, HPE GreenLake, NetApp, Pure Storage
 - AI-Ready/Object: MinIO, Cloudian, HammerSpace, NetApp AI storage, VAST Data, Weka

- **Database Systems:** The operational heartbeat of the enterprise, this segment captures the systems that power transactional applications across the line-of-business. This category is currently undergoing a massive convergence, with traditional RDBMS and NoSQL engines rapidly absorbing AI and analytics capabilities, blurring the lines between standard application backends and specialized AI memory systems.
 - Positioning/Vendors:
 - Operational/Relational: Oracle, Microsoft SQL Server, PostgreSQL, MySQL
 - NoSQL/Multi-Model: MongoDB, Redis, Cassandra, DynamoDB
 - Specialist: Neo4j, Redis, ElastiCache, InfluxDB
 - Vector-Native: Pinecone, Weaviate, Milvus (Zilliz), Qdrant, Chroma
 - Vector-Enabled: PostgreSQL (pgvector), MongoDB Atlas Vector, Azure AI Search

Data Processing & Orchestration

- **Analytical Data Platforms:** These are the heavy-lifting computational engines designed for complex querying, aggregation, and historical analysis. The market has shifted decisively here toward an open lakehouse architecture, which unifies the flexibility of data lakes with the governance of warehouses, increasingly supported by open table formats (such as Apache Iceberg) that enable data to be queried by multiple engines without being locked into a single vendor's silo.
 - Positioning/Vendors:
 - Cloud Data Warehouses and Lakehouses: Databricks, Microsoft Fabric, Snowflake, Google BigQuery, AWS Redshift, MotherDuck
 - Query Engines: Dremio, Starburst (Trino), Apache Hive, Apache Spark

- Emerging/Real-Time: ClickHouse, Firebolt, PuppyGraph, DuckDB, StarRocks (CelerData)
- Data Engineering & Movement: Functioning as the nervous system of the data stack, this layer handles the logistics of refining and transporting data from source to destination. It encompasses the full lifecycle of data plumbing: from Zero-ETL replication and high-velocity streaming to the Reverse ETL activation that pushes insights back into business apps, all governed by sophisticated orchestration frameworks that manage dependencies across complex workflows.
 - Positioning/Vendors:
 - Integration (ETL/ELT): Fivetran (dbt Labs), Airbyte, Informatica, Talend, Matillion
 - Streaming: Confluent (Kafka), Redpanda, Amazon Kinesis
 - Reverse ETL/Activation: Census, Hightouch
 - Data Orchestration: Airflow (Astronomer), Dagster, Prefect, Mage
 - Semantic Layer: With agentic AI taking center stage and demanding access to high-quality data from across the business, companies can no longer treat the semantic layer as a subsegment within Business Intelligence (BI). Still anchored in BI but increasingly marketed and managed as a distinct solution area, the Semantic Layer serves as the critical translation tier, defining business metrics (e.g., What is Revenue?) to ensure consistency across both visual dashboards and AI agents.
 - Positioning/Vendors:
 - Metric Stores: Cube, AtScale
 - Platform-native Layers: Databricks Unity Catalog, Microsoft Fabric, dbt (Fivetran) Semantic Layer
 - Knowledge Graphs: RelationalAI, Neo4j, Stardog

Data Analysis & Intelligence

- BI & Reporting: This segment is undergoing its most significant shift in two decades: moving from static dashboards to Generative BI. While traditional reporting remains for executive oversight, the modern experience is Agentic: users ask natural-language questions (Why did churn spike?) and receive auto-generated visuals and narratives. Success here now depends on integration with the Semantic Layer (see below) to prevent AI hallucinations, turning BI tools into trusted Data Agents rather than just visualization libraries.
 - Positioning/Vendors:
 - Visual and Embedded Analytics: Tableau, Microsoft Power BI, Google Looker, Qlik, Sisense, Domo, Yellowfin, Zoho Analytics

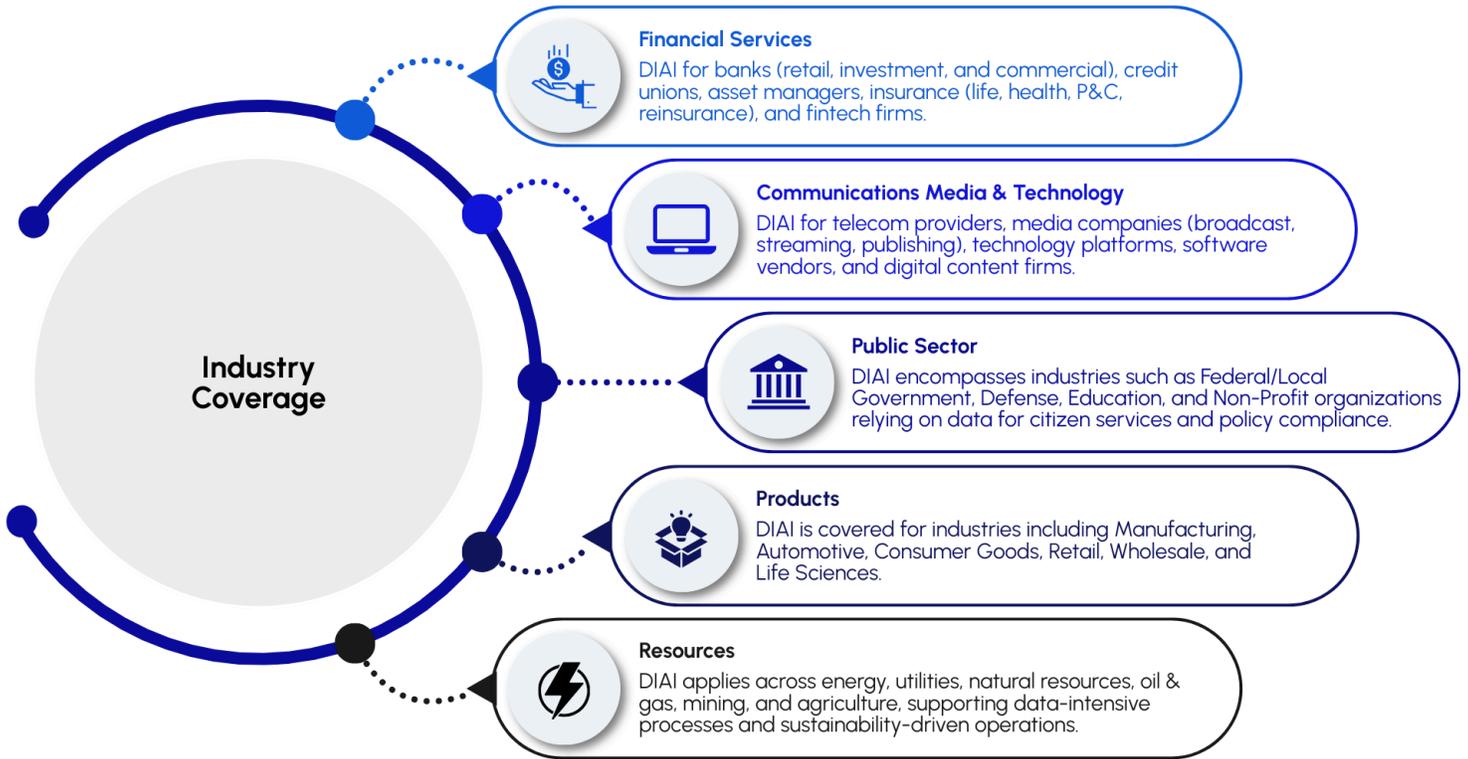
- Agentic and Semantic BI: ThoughtSpot, Databricks AI/BI and Genie, Snowflake Cortex AI, Amazon QuickSight, Julius, SynCFusion Bold BI, Omni, Tableau Next
- AI Development & Operations: This is the factory floor for AI, encompassing the end-to-end tooling required to build, deploy, and manage models. This segment has expanded beyond traditional predictive model training to focus on bringing data to bear on the GenAI stack. This includes the orchestration frameworks (such as LangChain), agentic memory management, inference serving infrastructure, and MLOps platforms necessary to run LLMs and agents in production at scale.
 - Positioning/Vendors:
 - Data Science Platforms: Dataiku, SAS, Alteryx, Databricks ML, Amazon SageMaker, Google Vertex AI, Microsoft Azure AI Studio
 - AI Data Orchestration (Agents): LangChain, IBM watsonx Orchestrate, LlamaIndex, Hex
 - Inference & Serving: Anyscale, Modal, Baseten, RunPod, Groq (together with NVIDIA), Tecton
 - MLOps: Weights & Biases, MLflow, H2O.ai

Data Management & Trust

- Data Governance, Quality, and Security: The static control plane responsible for defining the rules of the road: knowing what data you have, who owns it, and who is allowed to touch it. This segment focuses on compliance, cataloging, and policy enforcement. It serves as an essential license to operate that allows enterprises to democratize data access without risking regulatory breach or exposure to risk.
 - Positioning/Vendors:
 - Catalogs/Governance: Collibra, Alation, Atlan, Informatica (CDGC)
 - Access & Security: Immuta, Privacera, Okera (Databricks)
 - Data Quality: Monte Carlo (Quality features), Soda, Bigeye
- Data & AI Observability: The active radar system for the data stack, this segment focuses on detecting anomalies, pipeline failures, as well as model and semantic drift, often in real-time. This category is rapidly expanding to include Data FinOps (Cost Observability), becoming a critical defense mechanism for engineering teams to prevent silent data outages and control the spiraling compute costs associated with modern agentic AI and analytics workloads.
 - Positioning/Vendors:
 - Data Observability: Monte Carlo, Metaplane, Bigeye, Microsoft Purview
 - AI Observability: Arize AI, Galileo, Fiddler, HoneyHive
 - Data FinOps (Cost): CloudZero, Vantage, Finout, Chaos Genius

Industry Coverage

The Data Intelligence, Analytics, and Infrastructure model applies across the following five key vertical industries:



Deployment Type	Definition
Financial Services	DIAI for banks (retail, investment, and commercial), credit unions, asset managers, insurance (life, health, P&C, reinsurance), and fintech firms
Communications, Media & Technology	DIAI for telecom providers, media companies (broadcast, streaming, publishing), technology platforms, software vendors, and digital content firms
Public Sector	DIAI for industries such as Federal/Local Government, Defense, Education, and Non-Profit organizations relying on data for citizen services and policy compliance.
Products	DIAI for industries including Manufacturing, Automotive, Consumer Goods, Retail, Wholesale, and Life Sciences.
Resources	DIAI for energy, utilities, natural resources, oil & gas, mining, and agriculture, supporting data-intensive processes and sustainability-driven operations.

Each vertical is mapped to core Data Intelligence, Analytics, and Infrastructure use cases, including data modernization, real-time data activation, regulatory and data governance requirements, customer and operational intelligence, and AI-driven decision support.

Regional Coverage

The Enterprise Application model applies across the following four key regions:

Region	Scope
North America	<ul style="list-style-type: none"> Primary countries: United States, Canada
Europe, Middle East & Africa	<ul style="list-style-type: none"> Western Europe: United Kingdom, Germany, France, Italy, Spain, Netherlands, Switzerland, Nordics (Sweden, Denmark, Norway, Finland) Central & Eastern Europe: Poland, Czech Republic, Romania Middle East: United Arab Emirates (UAE), Saudi Arabia (KSA), Israel, Turkey Africa: South Africa, Egypt, Nigeria
Asia Pacific	<ul style="list-style-type: none"> Mature Markets: Japan, Australia, New Zealand, South Korea, Singapore Emerging Nations: China, India Southeast Asia: Indonesia, Malaysia, Thailand, Philippines, Vietnam
Latin America	<ul style="list-style-type: none"> Major Economies: Brazil, Mexico Emerging Markets: Argentina, Chile, Colombia, Peru, Costa Rica

Executive Summary

The era of the Data Technician—defined by manual pipeline plumbing and fragile ETL scripting—is effectively over. It is being replaced by the era of the AI Shepherd, where human expertise is focused on intent, defining business logic, governing data quality, and monitoring the behavior of autonomous AI agents.

This shift is more than just a narrative change. It is visible in the structural rotation of capital. Our 2025–2031 forecast reveals a structural rotation of capital away from legacy plumbing (Data Integration growing ~10–12%) toward intelligence infrastructure (Semantic Layer and Observability growing >20%).

While the Base Case forecast assumes a steady operationalization of GenAI, strategic risks remain. The Bull Case envisions a Singularity scenario where agentic AI drives ubiquitous adoption, pushing the market to US\$1.65 trillion by 2031. Conversely, the Bear Case accounts

for potential friction from energy constraints, storage and memory supply chain limitations, and regulatory fragmentation, which could cap the market at US\$822 billion.

Strategic Forecast Scenarios (2025–2031)

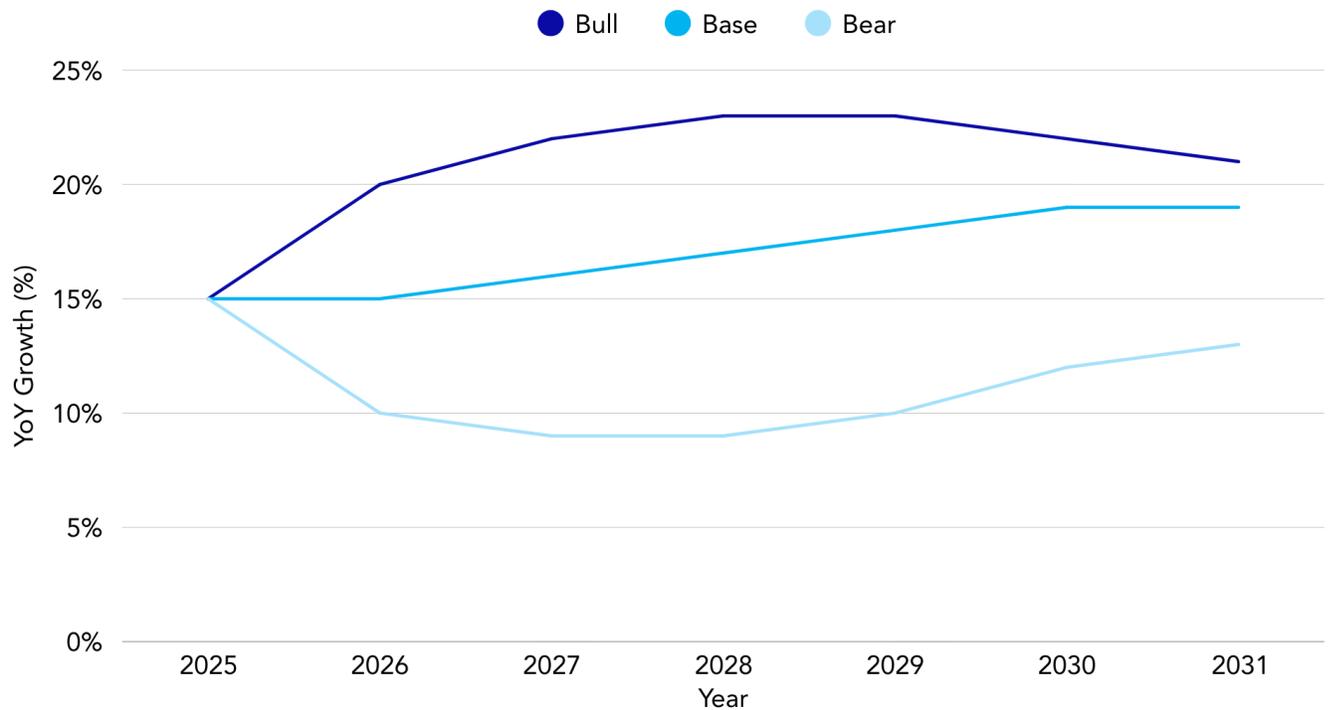
The following table outlines the three potential paths for the global market based on key swing factors such as AI adoption rates, regulatory harmonization, and energy availability.

Table 1: Global Market Forecast Scenarios (US\$ Billion)

	2025	2026	2027	2028	2029	2030	2031	CAGR
Bull	\$469.4	\$578.3	\$705.6	\$867.9	\$1,067.5	\$1,302.3	\$1,575.8	21.8%
Base	\$469.4	\$541.1	\$628.2	\$735.1	\$867.1	\$1,027.7	\$1,220.3	17.3%
Bear	\$469.4	\$498.7	\$543.6	\$592.5	\$652.9	\$728.0	\$822.7	10.4%

Source: Futurum Research, January 2026

Figure 1: Global Market Forecast Scenarios – Bull vs. Base vs. Bear Case Revenue (2022–2031)



Source: Futurum Research, January 2026

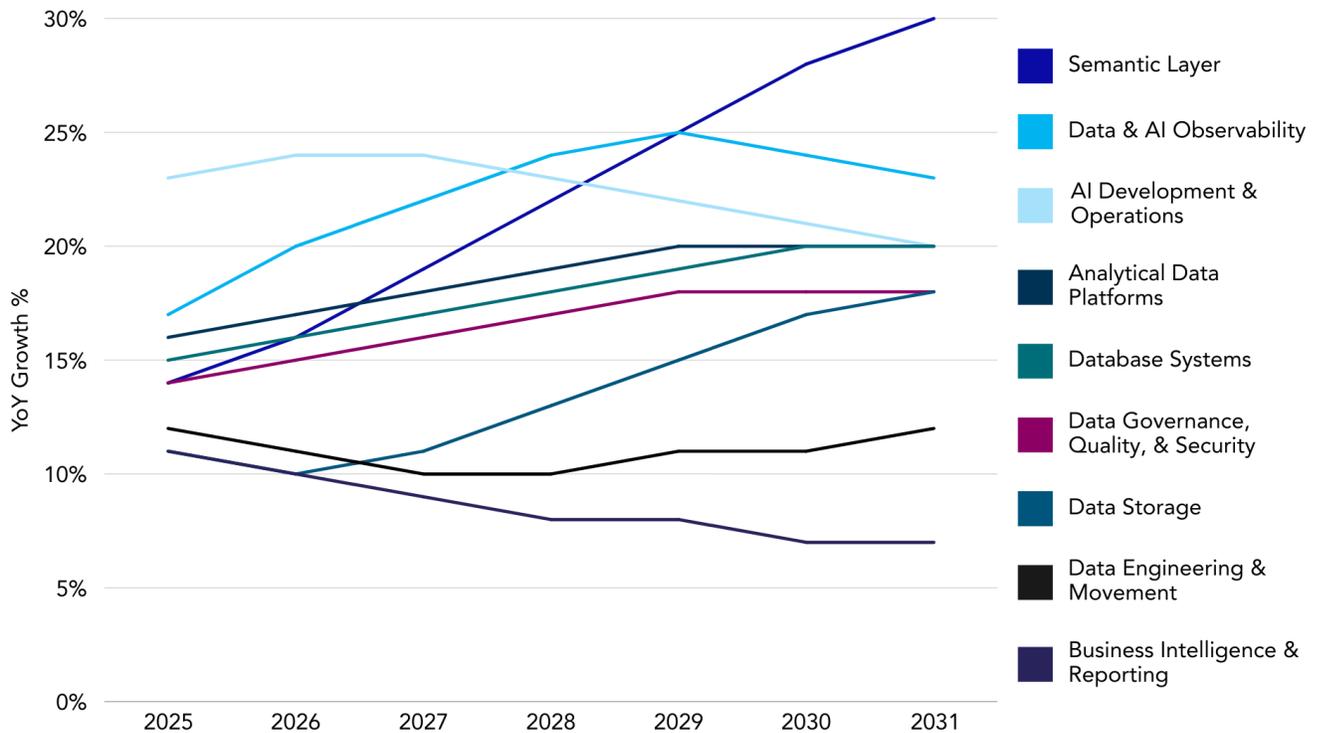
Market Analysis: Sub-Segment View

Table 2: Global Market by Sub-Segment (US\$ Billion)

Sub-Segment	2025	2026	2027	2028	2029	2030	2031
AI Dev & Ops	\$103	\$128	\$158	\$195	\$238	\$287	\$345
Data Storage	\$140	\$154	\$171	\$193	\$222	\$259	\$306
Database Systems	\$68	\$79	\$93	\$109	\$130	\$156	\$187
Lakehouse Platforms	\$37	\$43	\$51	\$61	\$73	\$88	\$106
Governance & Security	\$33	\$38	\$44	\$51	\$60	\$71	\$84
Data Engineering	\$38	\$42	\$46	\$50	\$56	\$62	\$70
BI & Reporting	\$29	\$32	\$35	\$38	\$41	\$44	\$47
Semantic Layer	\$13	\$15	\$18	\$22	\$27	\$35	\$46
Data Observability	\$9	\$11	\$13	\$16	\$20	\$25	\$31
Total	\$469	\$541	\$628	\$735	\$867	\$1,028	\$1,220

Source: Futurum Research, January 2026

Figure 2: Selected Sub-Segment Growth Rates (Semantic vs. Storage vs. Observability)



Source: Futurum Research, January 2026

The Unique Growth Curves of the market highlight distinct profiles of value creation over the next five years. We have identified three primary patterns that define this forecast period.

First, the Semantic Layer is exhibiting a Rocket Ship trajectory, undergoing a fundamental repricing in the market. Historically viewed as a niche governance tool, this segment is projected to accelerate from 16% growth in 2025 to 30% by 2031. This convex curve is driven by the rise of Autonomous Agents; without a semantic layer to define metrics, agents are prone to hallucination. As AI moves to production, this layer transforms from optional to critical infrastructure.

Second, the Data Observability market follows an S-Curve pattern, reflecting the industry's entry into the Messy Middle of AI adoption. As companies rush to production, pipelines will inevitably break, and cost models will break down, driving a surge in Observability spend that peaks at 25% growth in 2028. The curve stabilizes after 2029 as these tools become standard features across broader platforms, marking a transition from crisis response to operational maturity.

Finally, the foundational Data Storage layer demonstrates a Dip & Rip pattern. We forecast a temporary deceleration in 2026 (11%) as the need for workload consolidation and storage efficiency will grow, owing to ongoing supply chain issues. However, growth re-accelerates to 18% by 2030, driven by the rapid growth of model context and exhaust data along with the sheer volume of AI training data, which eventually overwhelms these early efficiency gains.

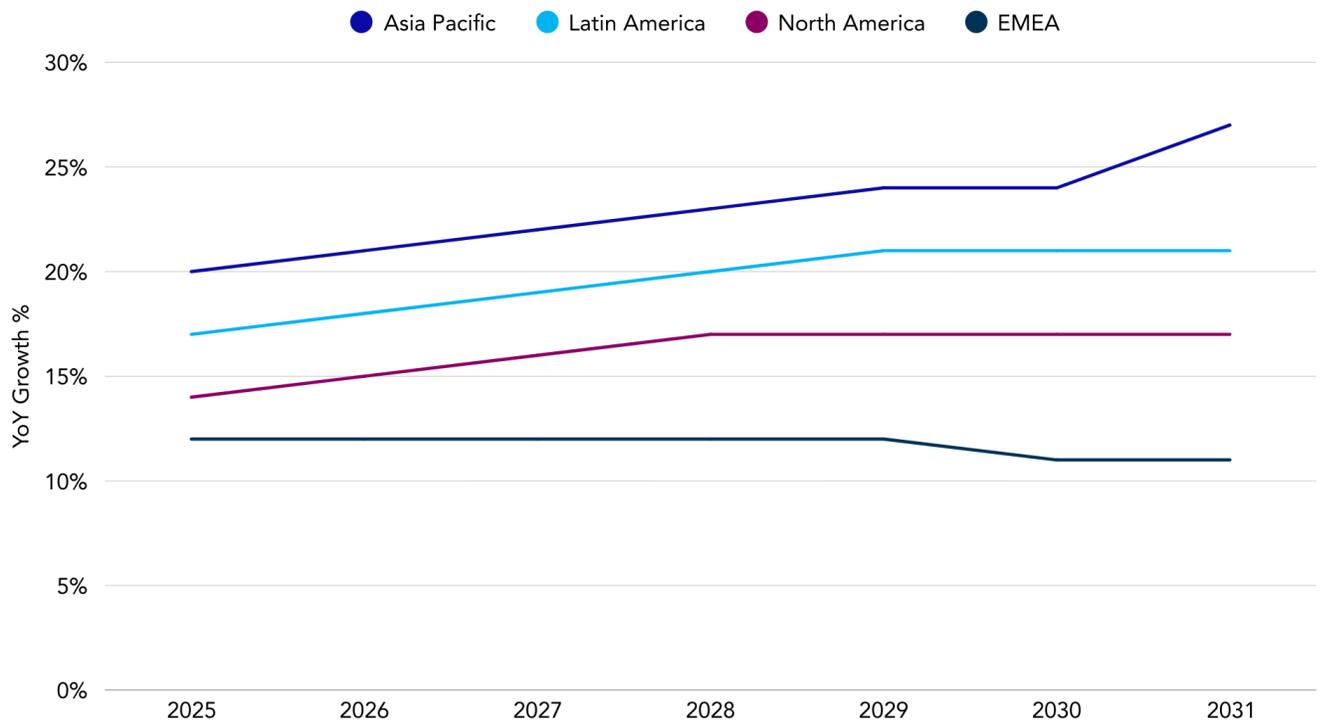
Market Analysis: Regional View

Table 3: Global Market by Region (US\$ Billion)

Region	2025	2026	2027	2028	2029	2030	2031
North America	\$222	\$254	\$292	\$339	\$395	\$463	\$541
Asia Pacific	\$101	\$122	\$149	\$184	\$230	\$291	\$369
EMEA	\$122	\$137	\$153	\$172	\$193	\$216	\$240
Latin America	\$25	\$29	\$34	\$40	\$48	\$58	\$71
Total	\$469	\$541	\$628	\$735	\$867	\$1,028	\$1,220

Source: Futurum Research, January 2026

Figure 3: Regional Growth Rate Divergence (APAC vs. EMEA)



Source: Futurum Research, January 2026

A distinct divergence in regional trajectories highlights a shifting global landscape. North America serves as the steady Anchor of the market. As the epicenter of GenAI innovation, it maintains consistent, high-value growth of around 15–17%. It remains the dominant region by market share (~45%), fueled by the massive capex cycles of the hyperscalers and the early adoption of AI-native software.

In contrast, Asia Pacific (APAC) is behaving like a Rocket, acting as the primary engine of global acceleration. We forecast growth to surge from 20% in 2025 to nearly 27% by 2031. This reflects a massive Digital Leapfrog effect, where rapidly developing economies focused on AI innovation bypass legacy software and infrastructure entirely to focus on cloud-native and AI-first opportunities. Sovereign AI initiatives in India, Japan, and Singapore are further amplifying this public-sector spend.

Meanwhile, EMEA faces a Regulation Saturation plateau. Growth is forecast to hover around 11–12%, significantly below the global average. While the EU AI Act drives necessary governance spending, strict data residency laws and a slower macroeconomic environment create structural friction that limits the explosive hyper-scale adoption seen in other regions.

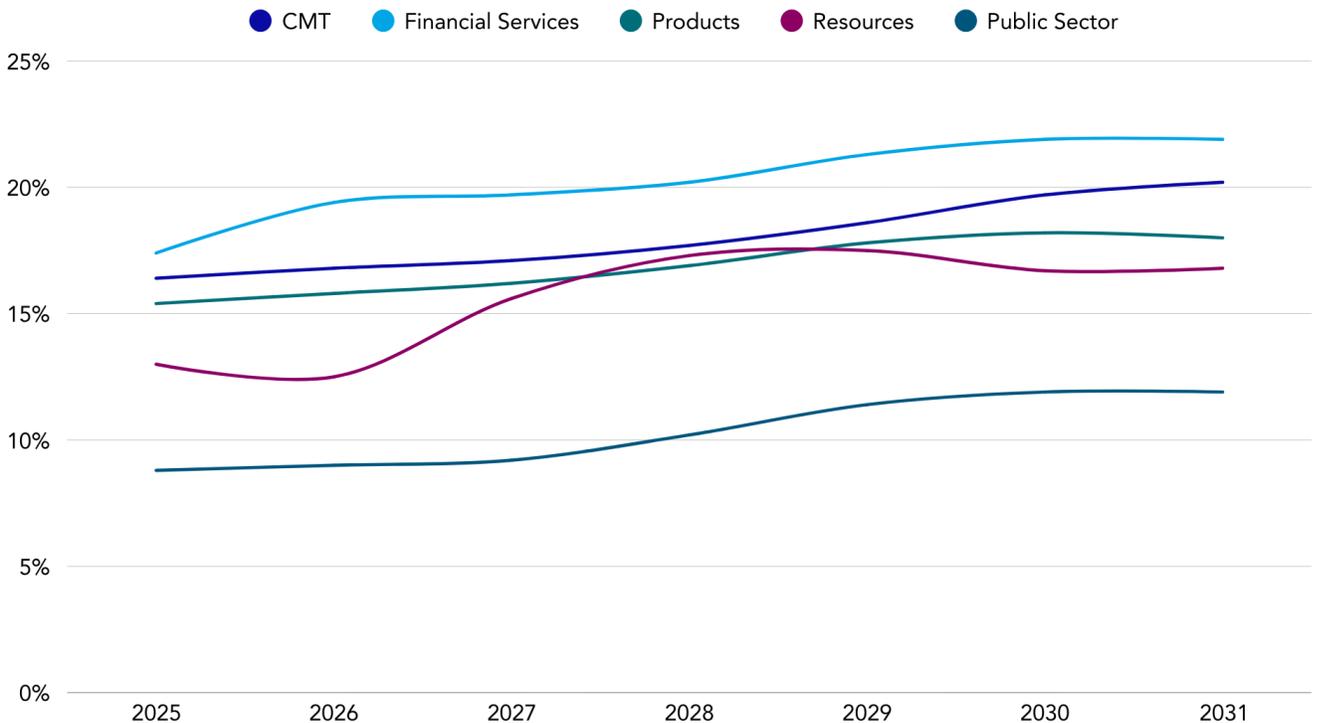
Market Analysis: Vertical View

Table 4: Global Market by Vertical (US\$ Billion)

Vertical	2025	2026	2027	2028	2029	2030	2031
CMT	\$139	\$162	\$190	\$224	\$265	\$318	\$382
Financial Services	\$99	\$118	\$141	\$170	\$206	\$251	\$306
Products	\$87	\$101	\$117	\$137	\$161	\$191	\$225
Resources	\$70	\$79	\$91	\$107	\$126	\$147	\$171
Public Sector	\$74	\$81	\$88	\$97	\$109	\$121	\$136
Total	\$469	\$541	\$628	\$735	\$867	\$1,028	\$1,220

Source: Futurum Research, January 2026

Figure 4: Vertical Growth Divergence (Two-Speed Economy)



Source: Futurum Research, January 2026

The market is splitting into a Two-Speed Economy, defined by the friction of physical assets.

In the fast lane, Financial Services has decoupled to become the market's primary growth engine, overtaking the technology builders themselves. As a purely data-native sector with zero physical latency, banking and insurance are the Power Users of this cycle, aggressively

operationalizing GenAI to address fraud and risk and drive immediate ROI. CMT remains the massive industrial base of this wave, maintaining high sustained velocity, but it is the Financial Services sector that is setting the pace for peak acceleration.

In the slow lane, Physical industries such as Resources and the Public Sector face a lagged integration curve. Their acceleration is back-loaded, picking up meaningfully only in 2027/28. For a mining company or a utility grid, adopting AI requires complex integration with OT and legacy hardware. While the long-term opportunity for Autonomous Operations is massive, the ramp-up is structurally slower due to these physical constraints.

Methodology

The data and insights in this report are derived from The Futurum Group's proprietary market sizing model, which triangulates data from three primary vectors to ensure accuracy and strategic relevance:

1. **Bottom-Up Financial Analysis:** We conducted a detailed line-of-business analysis of the Top 50 vendors in the space, ranging from hyperscalers (AWS, Microsoft, Google) to pure-play platforms (Snowflake, Databricks) and emerging ISVs. We isolated specific Data & AI revenue streams from general IT spend to create a clean baseline of addressable market revenue.
2. **Top-Down Demand Validation:** These supply-side figures were validated against The Futurum Group's *2025 Decision Maker Survey*. This study polled over 1,000 enterprise IT leaders on their committed investment intentions, specifically measuring the shift in budget allocation from maintenance (ETL) to innovation (GenAI/Observability) over the next 12–36 months.
3. **Taxonomy & Scope:** The market definition aligns with The Futurum Group's *Data Intelligence, Analytics, and Infrastructure* taxonomy. This encompasses 9 sub-segments across the Data Foundation (Storage/DB), Processing (Engineering/Semantic), Analysis (BI/AI), and Management (Governance/Observability) layers.

About Us

About the Authors

Brad Shimmin is Vice President & Practice Lead, Data Intelligence, Analytics, & Infrastructure at Futurum, where he provides strategic direction and market analysis to help organizations maximize their investments in data and analytics. Currently, Brad is focused on helping companies establish an AI-first data strategy.

About The Futurum Group

The Futurum Group's analysts, researchers, and advisors help business leaders anticipate tectonic shifts in their industries and leverage disruptive innovation. Unlike traditional analysts, The Futurum Group works in analysis and research and takes that insight and knowledge even further, engaging through the go-to-market process.

Futurum Research provides in-depth research and insights on global technology markets using advisory services, custom research reports, strategic consulting engagements, digital events, go-to-market planning, and message testing. It also creates, distributes, and amplifies rich media content that all stakeholders read, watch, and listen to.

See more details on The Futurum Group at futurumgroup.com.

Copyright & Use License

Copyright Notice

Copyright ©2026 by The Futurum Group, LLC. All rights, including translation into other languages, are specifically reserved. No part of this publication may be reproduced in any form, stored in a retrieval system, or transmitted by any method or means, electrical, mechanical, photographic, or otherwise, without the express written permission of The Futurum Group futurumgroup.com. United States copyright laws and international treaties protect this publication. Unauthorized distribution or reproduction of this publication, or any portion of it, may result in severe civil and criminal penalties and will be prosecuted to the maximum extent necessary to protect the publisher's rights.

License Notice

This document may be distributed only to the licensed organization.

The following acts are prohibited:

Transmittal to others outside your immediate organization, including partners, resellers, external consultants, etc., in any media format

Posting on a website that is accessible to others outside your immediate organization

The possession or use within an unlicensed organization

Limitation of Liability Notice

The information contained herein has been obtained from sources believed to be reliable. The Futurum Group shall have no liability for errors, omissions, or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for selecting these materials to achieve their intended results. The opinions expressed herein are subject to change without notice.