Data Preparation - Market Share Analysis, Industry Trends & Statistics, Growth Forecasts (2025 - 2030)

Market Report | 2025-06-01 | 120 pages | Mordor Intelligence

AVAILABLE LICENSES:

- Single User License $4750.00
- Team License (1-7 Users) $5250.00
- Site License $6500.00
- Corporate License $8750.00

Report description:

Data Preparation Market Analysis

The data preparation market size stands at USD 6.95 billion in 2025 and is projected to reach USD 14.71 billion by 2030, expanding at a 16.2% CAGR. This expansion mirrors the surge in AI-ready infrastructure as enterprises embed generative AI into day-to-day workflows; adoption has reached 83% of organizations in China and full production roll-outs in 24% of United States companies. Proliferating data-governance programs, now present in 71% of organizations compared with 60% in 2023, reinforce spending on systematic data preparation tools. Deployment choices continue to diverge: on-premises solutions controlled 65.7% of 2024 revenue, while cloud deployments are scaling fastest at 17.8% CAGR, a pattern shaped by sovereign-cloud regulations such as Vietnam's Data Law, effective July 2025, that restrict cross-border transfers. Large enterprises held 68.9% revenue share in 2024, yet small and medium enterprises (SMEs) show the strongest momentum at 18.1% CAGR as low-code analytics and consumption-based pricing lower entry barriers. Data-ingestion modules retained the top 24.3% slice of 2024 revenue; however, governance-centric solutions are rising fastest at 17.3% CAGR, pushed by greenhouse-gas-reporting mandates emerging from the EU Corporate Sustainability Reporting Directive. IT and telecommunications contributed the largest 22.8% vertical share in 2024, while healthcare and life sciences climbed at a 16.8% CAGR through 2030 as AI enters diagnosis, patient-workflow and life-science research and development. Regionally, North America led with 37.1% revenue in 2024, yet Asia-Pacific will outpace all others at 17.5% CAGR, underpinned by expanding data-center capacity-12,206 MW active and 14,338 MW in development. Mergers and acquisitions activity signals intensifying competition: Salesforce agreed to purchase Informatica for USD 8 billion in May 2025, and Alteryx was taken private for USD 4.4 billion in March 2024.

Global Data Preparation Market Trends and Insights

Accelerated Shift to Low-/No-Code Self-Service Analytics Tools

Low-code interfaces are redefining the data preparation market by enabling business specialists to build pipelines via drag-and-drop designs rather than scripts. Google Cloud's BigQuery data preparation illustrates the trend, offering AI guidance that cleans, profiles and transforms data with natural-language prompts. The approach reduces reliance on scarce data engineers, shortens development cycles and aligns analytics delivery with domain expertise. GenAI-powered augmentation is spreading quickly; industry forecasts suggest nearly all BI platforms will embed GenAI by 2026. Adoption, however, requires diligent governance to keep proliferating citizen-built flows aligned with enterprise quality and security standards.

Surging Cloud Adoption Among SME Analytics Teams

SMEs are scaling cloud-native pipelines to close capability gaps with larger rivals, driving incremental demand across Asia-Pacific where 60% of firms plan AI language-model implementation by 2025. Cloud elasticity and consumption pricing let smaller firms avoid capital expenses while accessing advanced data-prep functions. UK research shows sub-1% of SMEs exploit big-data analytics today, underscoring runway as cost and complexity hurdles fall. Yet skills shortages persist; managed service providers are stepping in to configure pipelines and enforce compliance, particularly around emerging data-localization rules.

Skills Gap for Complex Data-Governance Configuration

Nearly one-third of CIOs cite data-management complexity as a critical obstacle, and shortages of governance specialists delay the rollout of scalable pipelines. The challenge intensifies where legislation such as California's climate-disclosure rule mandates automated capture of Scope 1-3 emissions. Emerging markets face deeper shortages as academic programs lag, pushing firms toward external consultants and managed-service contracts that inflate deployment budgets.

Other drivers and restraints analyzed in the detailed report include:

Integration of GenAI Copilots Inside Data-Prep Workflows / Vendor Bundling of Data-Prep Modules into Broader Data-Fabric Suites / Steep Total Cost of Ownership for Multi-Cloud Data Pipelines /

For complete list of drivers and restraints, kindly check the Table Of Contents.

Segment Analysis

The data preparation market size for on-premises platforms totaled USD 4.57 billion in 2024, translating to 65.7% data preparation market share, a reflection of enterprise demand for direct control amid tougher localization rules. Vietnam's Data Law and India's Digital Personal Data Protection Rules reinforce on-prem and sovereign-cloud models that keep sensitive records within national borders. Cloud services, though smaller, are projected to compound at 17.8% through 2030 as SMEs and digitally native units prioritize agility. In North America, hybrid blueprints predominate, fusing local clusters for regulated data with hyperscale reservoirs for lower-risk workloads. Cloud providers respond with dedicated regional instances and encrypted-key control to offset compliance fears, widening adoption beyond traditional tech hubs as smaller cities gain direct-connect fiber.

The economic calculus hinges on workload variability: steady ETL batches and predictable enrichment jobs remain on-prem due to licensing amortization, while bursty AI inference and citizen-developer sandboxes migrate to pay-as-you-go clouds. Over half of multinationals are expected to run sovereign-cloud instances by 2029, creating demand for seamless policy enforcement across private, public and edge nodes. Vendors now emphasize unified control planes that propagate data-quality rules and lineage

graphs no matter the substrate.

Large corporations generated USD 4.79 billion revenue in 2024, equal to 68.9% of the data preparation market, supported by dedicated governance teams and global footprints. Their spend skew favors platform bundles that integrate catalog, lineage and observability into existing data fabrics. Conversely, SMEs contributed USD 2.16 billion yet will outgrow other cohorts at 18.1% CAGR, lifting the data preparation market size for SME solutions to a projected USD 5.6 billion by 2030. Consumption billing and automated schema-detection reduce capital obstacles, enabling regional retailers, fintechs and SaaS start-ups to achieve parity with incumbents.

A Small Business Institute Journal survey shows 70% of U.S. SMEs acknowledge analytics value, but only a minority has in-house talent to execute end-to-end pipelines. Low-code cloud workbenches and managed-service ecosystems fill gaps, while industry associations offer modular training to accelerate citizen usage. Challenges persist in developing policy frameworks that map to emerging AI-act obligations, creating openings for channel partners specializing in compliance overlays.

The Data Preparation Market Report is Segmented by Deployment (On-Premises and Cloud), Enterprise Size (Small and Medium Enterprises (SMEs) and Large Enterprises), Solution Type (Data Ingestion, Data Cataloging, and More), End-User Vertical (BFSI, Healthcare and Life Sciences, and More), and Geography.

Geography Analysis

North America's USD 2.58 billion spend in 2024 reflected 37.1% data preparation market share, an outcome of early AI experimentation and dense vendor ecosystems. California's climate-disclosure statute compels companies above USD 1 billion revenue to publish Scope 1-3 emissions, reinforcing governance-tool demand across the continent. Multinationals headquartered elsewhere yet active in the United States must still report, extending influence beyond borders. Canada advances parallel frameworks through Bill C-27's Consumer Privacy Protection Act, while Mexico's data-localization proposals are prompting hybrid-cloud blueprints for cross-border maquiladora supply chains. The region's investment focus has pivoted from initial ingestion capabilities to advanced observability and automated remediation that reduce operational toil.

Asia-Pacific is the fastest climber, expanding 17.5% annually as public-cloud growth surpasses other regions. China's 83% GenAI adoption manifests in aggressive pipeline modernization, while South Korea and Japan allocate national AI funds to health-record digitization and smart-factory programs. Vietnam's Data Law and India's DPDP Rules trigger data-residency layers within multinational stacks, increasing on-prem edge deployments and stimulating demand for integrated policy engines. Australian enterprises face new Critical Infrastructure Security obligations that require real-time anomaly detection in upstream data-prep stages. Meanwhile, Singapore's IMDA grants push SMEs to cloud services, reinforcing the region's mass-market momentum.

Europe posts steady mid-teens growth as ESG mandates drive □report-ready□ pipeline investments. The EU Corporate Sustainability Reporting Directive forces roughly 50,000 firms to log greenhouse-gas metrics using consistent taxonomies, elevating data catalog and quality tooling to the executive agenda. Germany and France lead spend, though momentum accelerates in Italy and Spain as Recovery and Resilience Facility grants underwrite digital-transition projects. The EU AI Act requires transparency, bias monitoring and human-oversight logs, deepening the need for secure lineage archives that span edge nodes and hyperscaler zones. Eastern European states ramp local-cloud capacity to keep citizen data domestic, encouraging partnerships between regional telcos and global hyperscalers.

List of Companies Covered in this Report:

Alteryx Inc. / Informatica LLC / IBM Corporation / Microsoft Corporation / Tableau Software LLC (Salesforce) / SAP SE / SAS Institute Inc. / QlikTech International AB / TIBCO Software Inc. / Talend SA / Oracle Corporation / Trifacta Inc. (Google) /

Databricks Inc. / Snowflake Inc. / Dataiku SAS / MicroStrategy Inc. / RapidMiner Inc. / Paxata Inc. (DataRobot) / Unifi Software Inc. / Denodo Technologies Inc. /

Additional Benefits:

 The market estimate (ME) sheet in Excel format /
3 months of analyst support /

Table of Contents:

1 INTRODUCTION
1.1 Study Assumptions and Market Definition
1.2 Scope of the Study

2 RESEARCH METHODOLOGY

3 EXECUTIVE SUMMARY

4 MARKET LANDSCAPE
4.1 Market Overview
4.2 Market Drivers
4.2.1 Accelerated shift to low-/no-code self-service analytics tools
4.2.2 Surging cloud adoption among SME analytics teams
4.2.3 Integration of GenAI copilots inside data-prep workflows
4.2.4 Vendor bundling of data-prep modules into broader data-fabric suites
4.2.5 Rapid rise of domain-specific 'vertical AI' data-prep pipelines
4.2.6 Sovereign-cloud rules fuelling on-prem / hybrid repatriation
4.3 Market Restraints
4.3.1 Skills gap for complex data-governance configuration
4.3.2 Steep total cost of ownership for multi-cloud data-pipelines
4.3.3 Escalating data-sovereignty penalties in emerging markets
4.3.4 Carbon-footprint quotas pushing back on compute-heavy prep jobs
4.4 Value Chain Analysis
4.5 Technological Outlook
4.6 Porter's Five Forces Analysis
4.6.1 Bargaining Power of Suppliers
4.6.2 Bargaining Power of Buyers
4.6.3 Threat of New Entrants
4.6.4 Threat of Substitutes
4.6.5 Intensity of Competitive Rivalry
4.7 Assessment of the Impact of Macroeconomic Trends on the Market

5 MARKET SIZE AND GROWTH FORECASTS (VALUE)
5.1 By Deployment
5.1.1 On-premises
5.1.2 Cloud
5.2 By Enterprise Size
5.2.1 Small and Medium Enterprises (SMEs)

5.2.2 Large Enterprises
5.3 By Solution Type
5.3.1 Data Ingestion
5.3.2 Data Cataloging
5.3.3 Data Quality
5.3.4 Data Governance
5.3.5 Data Wrangling
5.3.6 Data Enrichment
5.4 By End-user Vertical
5.4.1 BFSI
5.4.2 Healthcare and Life Sciences
5.4.3 Retail and e-Commerce
5.4.4 Manufacturing and Industrial
5.4.5 IT and Telecommunications
5.4.6 Government and Public Sector
5.4.7 Others (Energy, Education, Media)
5.5 By Geography
5.5.1 North America
5.5.1.1 United States
5.5.1.2 Canada
5.5.1.3 Mexico
5.5.2 Europe
5.5.2.1 Germany
5.5.2.2 United Kingdom
5.5.2.3 France
5.5.2.4 Italy
5.5.2.5 Spain
5.5.2.6 Russia
5.5.2.7 Rest of Europe
5.5.3 Asia-Pacific
5.5.3.1 China
5.5.3.2 Japan
5.5.3.3 India
5.5.3.4 South Korea
5.5.3.5 Australia and New Zealand
5.5.3.6 Rest of Asia-Pacific
5.5.4 South America
5.5.4.1 Brazil
5.5.4.2 Argentina
5.5.4.3 Rest of South America
5.5.5 Middle East and Africa
5.5.5.1 Middle East
5.5.5.1.1 Saudi Arabia
5.5.5.1.2 United Arab Emirates
5.5.5.1.3 Turkey
5.5.5.1.4 Rest of Middle East
5.5.5.2 Africa

5.5.5.2.1 South Africa
5.5.5.2.2 Nigeria
5.5.5.2.3 Rest of Africa

6 COMPETITIVE LANDSCAPE
6.1 Market Concentration
6.2 Strategic Moves
6.3 Market Share Analysis
6.4 Company Profiles (includes Global level Overview, Market level overview, Core Segments, Financials as available, Strategic Information, Market Rank/Share for key companies, Products and Services, and Recent Developments)
6.4.1 Alteryx Inc.
6.4.2 Informatica LLC
6.4.3 IBM Corporation
6.4.4 Microsoft Corporation
6.4.5 Tableau Software LLC (Salesforce)
6.4.6 SAP SE
6.4.7 SAS Institute Inc.
6.4.8 QlikTech International AB
6.4.9 TIBCO Software Inc.
6.4.10 Talend SA
6.4.11 Oracle Corporation
6.4.12 Trifacta Inc. (Google)
6.4.13 Databricks Inc.
6.4.14 Snowflake Inc.
6.4.15 Dataiku SAS
6.4.16 MicroStrategy Inc.
6.4.17 RapidMiner Inc.
6.4.18 Paxata Inc. (DataRobot)
6.4.19 Unifi Software Inc.
6.4.20 Denodo Technologies Inc.

7 MARKET OPPORTUNITIES AND FUTURE OUTLOOK
7.1 White-space and Unmet-need Assessment

Data Preparation - Market Share Analysis, Industry Trends & Statistics, Growth Forecasts (2025 - 2030)

Market Report | 2025-06-01 | 120 pages | Mordor Intelligence

To place an Order with Scotts International:

- - Print this form
- - Complete the relevant blank fields and sign
- - Send as a scanned email to support@scotts-international.com

ORDER FORM:

Select license	License	Price
	Single User License	$4750.00
	Team License (1-7 Users)	$5250.00
	Site License	$6500.00
	Corporate License	$8750.00
	VAT	
	Total	

*Please circle the relevant license option. For any questions please contact support@scotts-international.com or 0048 603 394 346.

** VAT will be added at 23% for Polish based companies, individuals and EU based companies who are unable to provide a valid EU Vat Numbers.

Email*		Phone*	
First Name*		Last Name*	
Job title*			
Company Name*		EU Vat / Tax ID / NIP number*	
Address*		City*	
Zip Code*		Country*	
		Date	2026-02-28
		Signature	