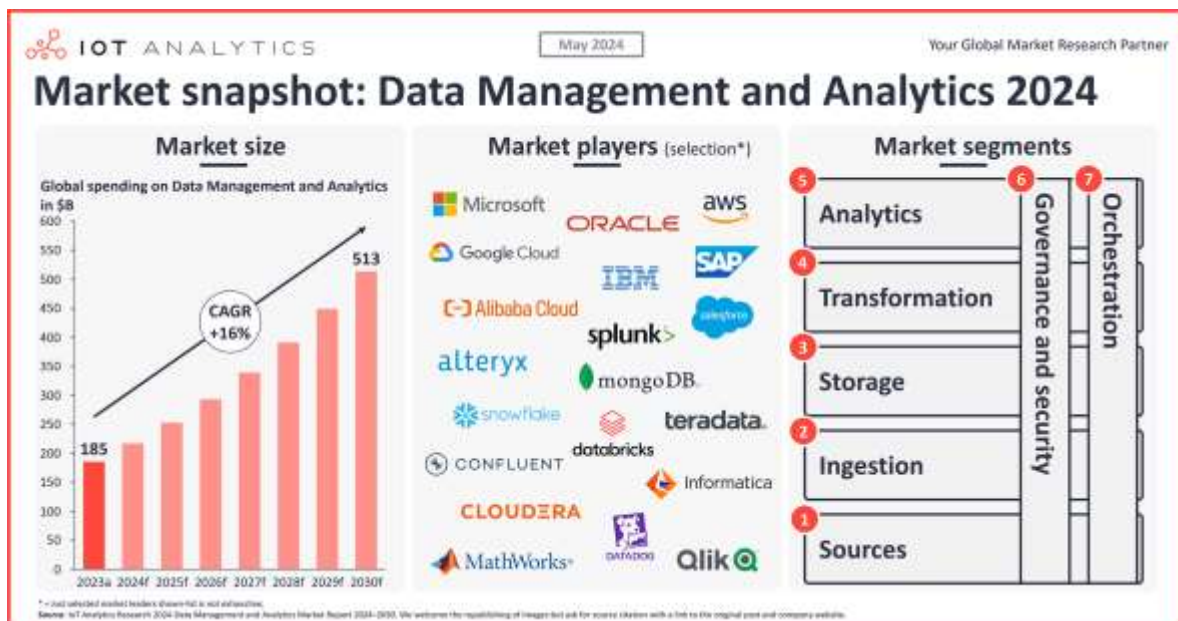


FOR IMMEDIATE RELEASE

New research from IoT Analytics highlights 7 Data Management components that companies must evaluate for building effective AI Models



[Hamburg, Germany] – [May 29, 2024] – As the field of Artificial Intelligence (AI) continues to experience explosive growth, a new research article from IoT Analytics sheds light on an overlooked aspect: data management. The research article, based on findings from the recently published "Data Management and Analytics Market Report 2024-2030," reveals 7 key data management components that are critical for building effective AI models. AI is projected to drive significant growth in the data management market, which is expected to reach US\$513.3 billion by 2030.

KEY INSIGHTS

- AI is projected to drive significant growth in the data management market, which is expected to reach \$513.3 billion by 2030 – according to a new **report** by IoT Analytics.
- AI relies on robust data management across 7 key components to build effective AI models: 1. sources, 2. ingestion, 3. storage, 4. transformation, 5. analytics, 6. governance and security, and 7. orchestration.
- Hyperscalers dominate the market at >50% market share but a number of promising data management vendors are championing individual segments of the market.

SELECT QUOTES

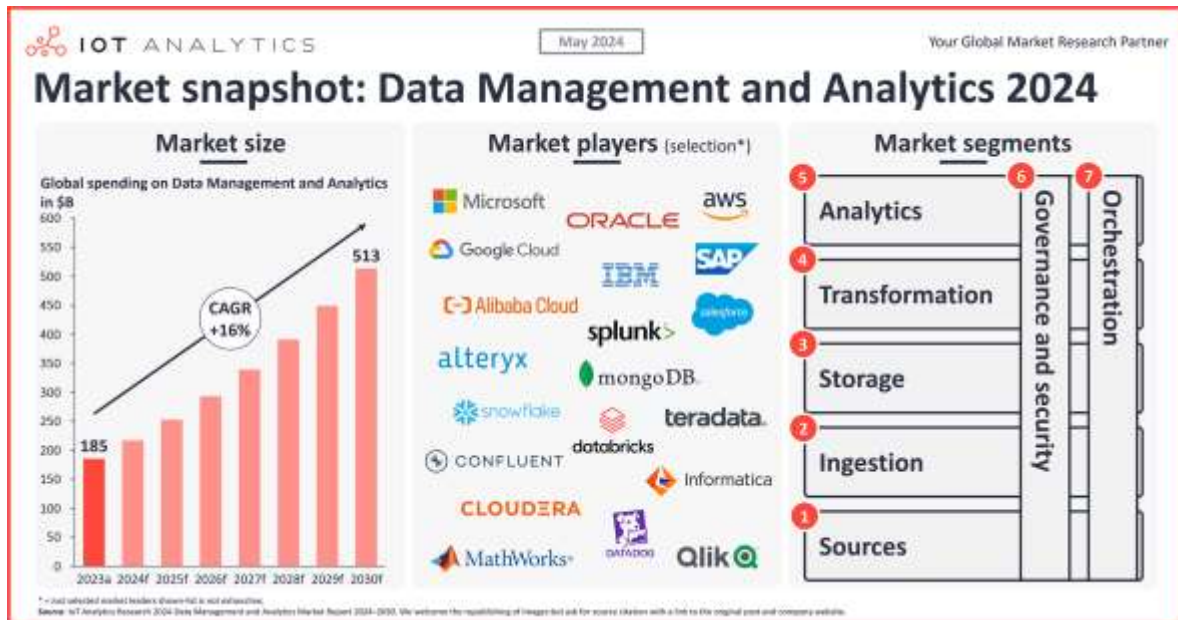
Knud Lasse Lueth, CEO at IoT Analytics, remarks *"Hyperscalers like AWS, Microsoft, and Google dominate the data management market with highly integrated portfolios across all major market segments. There are also some quickly growing data management scale-ups that are regarded as having a best-in-class offering and are subsequently enjoying strong market traction. It will be interesting to see whether companies will opt for the convenience of having everything from one vendor or settle on three to five main data management solutions on top of their cloud architecture."*

Oktay Demir, COO at IoT Analytics, adds that *"C-level executives often overlook the critical importance of data management for AI. Strong data management is the foundation for successful AI implementation. AI holds transformative power, and their successful implementation brings significant prestige to leadership. However, the less glamorous groundwork—a well-executed data management strategy—is frequently neglected. The "Data Management and Analytics Market Report 2023–2030" highlights how excellent data management strategies enable successful AI initiatives and the adverse effects of poor data management."*

Mohammad Hasan, Analyst at IoT Analytics, adds *"In my opinion, data fabric is still not very popular in terms of adoption as it can come with a heavy price tag due to unfit data architectures. However, given the increase in data complexity because of the exponential growth in big data, propelled by hybrid cloud, AI, IoT, and edge computing, there seems to be a good opportunity for vendors in this scenario."*

[The full research article is attached below]

How global AI interest is boosting the data management market



The bigger picture: No AI strategy without a data strategy

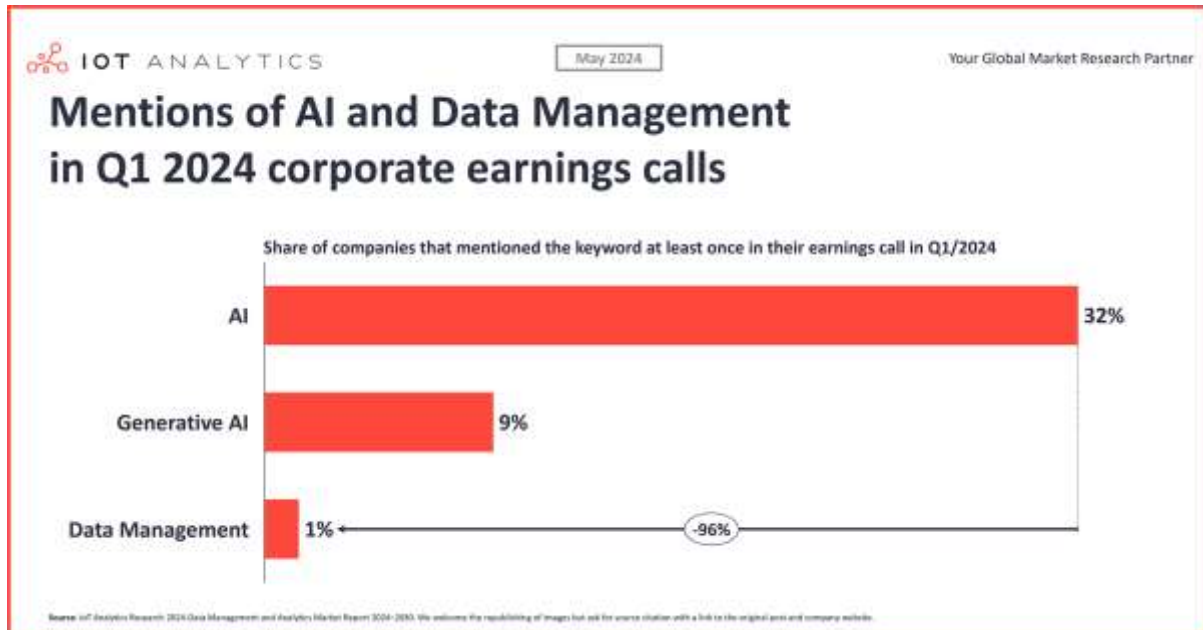
In Q1 2024, the CEOs of nearly 1/3 of all companies discussed AI in their earnings calls, but only a very small portion (1% of all earnings calls) discussed data management. However, as IoT Analytics' inaugural Data Management and Analytics Market Report 2023–2030 report notes, proper data management is the foundation of AI.

This raises the question: **Are companies overlooking the need for data management investments to succeed in AI?**

"We said it many times: There's no AI strategy without a data strategy. The intelligence we're all aiming for resides in the data, hence the quality of that underpinning is critical."

Frank Sloopman – (former) CEO, Snowflake (Nov 2023)

With modern data management crucial for AI success, the data management report forecasts strong growth in the data management market over the next six years. Overall growth is expected to be 16% per annum from 2023 to 2030, with the market anticipated to reach \$513 billion by the end of 2030. The renewed strategic importance of AI and ML (including both predictive and generative AI) is the key driver for this market expansion.

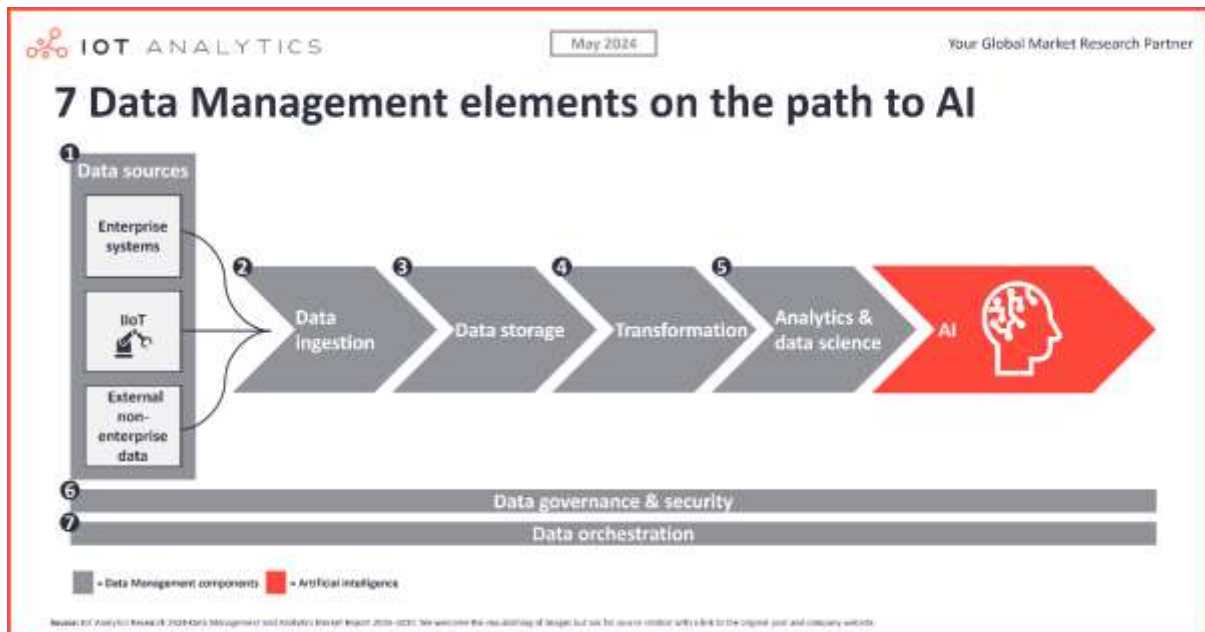


32% of CEOs mentioned AI in their Q1/2024 earnings call, while only 1% mentioned data management

The 7 key components of data management and why they matter for AI

AI models rely heavily on data for training and operation, making robust data management essential. To train models that are specifically designed to work within the context of their particular business processes, and environments, companies must evaluate the **7 key components of their data management tech stack**:

1. **Sources**
2. **Ingestion**
3. **Storage**
4. **Transformation**
5. **Analytics**
6. **Governance and security**
7. **Orchestration**



The data management stack at a glance

Data management component 1: Sources



Description

Sources encompass diverse data formats from multiple repositories. Data sources may include enterprise systems such as enterprise resource planning (ERP) or customer relationship management (CRM), IIoT data, coming from devices such as programmable logic controllers (PLCs) or sensors, or other external data, such as social media or governmental data.

Note: IoT Analytics does not count sources as part of the data management market.

Importance for AI

Data sources are the foundation for AI. Data sources provide the foundation for AI training, as sophisticated AI often requires a large and diverse mix of data from different sources. The more data sources connected, the more powerful and versatile AI models can become. Therefore, it is crucial to identify what data is needed, where it can be obtained, and how it will be collected. For instance, generative AI works primarily with unstructured data—information that doesn't fit neatly into relational databases, like text or images. This unstructured data must be identified, organized, and seamlessly integrated into the data storage system to maximize AI's potential.

Report insight(s)

There are three main types of data. The report identifies three distinct types of data, each with different characteristics that determine the appropriate data management strategies and database technologies.

1. **Structured data:** Typically sourced from enterprise systems like ERP and CRM, structured data is characterized by its organized, tabular format and defined relationships between data points. Relational databases, such as SQL databases, excel with structured data, handling well-defined schemas and relationships effectively.
2. **Semi-structured data:** Includes data such as JSON or XML files, which do not conform strictly to relational database schemas but maintain a hierarchical structure. NoSQL databases are well-suited for managing semi-structured data, as they handle formats that do not fit neatly into traditional rows and columns, providing flexibility in data management.
3. **Unstructured data:** Includes text, images, and videos. It lacks a predefined format, making its management more complex. Distributed file systems and NoSQL databases are ideal for unstructured data.

The type of data structure guides the selection of database model. If a project mostly relies on unstructured data, it requires a data management strategy that emphasizes advanced search capabilities to catalog and retrieve unstructured information efficiently. Distributed file systems and NoSQL databases offer methods for managing large volumes of diverse content types, enabling fast data processing and meaningful insights extraction.

Data management component 2: Ingestion

2

Data ingestion

Description

Ingestion channels data from sources into storage. It collects data from primary sources (ERP, CRM, PLCs, or external feeds) and unifies it into a storage system using connectors to ensure compatibility and proper format handling.

Importance for AI

Ingestion assures continuous intake of large data volumes. Data from various sources must be continuously collected and fed into the AI algorithm. To ensure AI models function, it is crucial to avoid connection problems that could lead to data gaps. Continuous data streams are especially vital for AI applications requiring real-time data, where delays could lead to missed opportunities or increased risks.

Report insight(s)

Real-time data ingestion is on the rise. The report discusses 8 key trends, one of which is the increasing importance of real-time data ingestion for immediate decision-making. Technologies like US-based nonprofit **Apache's** Kafka, a distributed event streaming platform, facilitate real-time data collection and processing with high throughput and low latency. This enables organizations to act on data as it is generated, improving responsiveness and operational efficiency. For instance, online streaming service **Netflix** leverages Kafka to manage over 700 billion daily events, ensuring smooth data flow and real-time processing to maintain a high-quality user experience across its +260 million subscribers.

Data management component 3: Storage

3

Data storage

Description

Storage uses technologies and architectures to safeguard, organize, and store data. There are two main components:

1. **Storage technologies** – includes hardware (HDDs or SSDs), and software (database management systems (DBMS)) used for data storage.
2. **Data architectures** – includes data warehouses, data lakes, or cloud and used as a blueprint for data architecture and data access.

Importance for AI

Data storage ensures efficient access to required data. Storage is crucial because it provides the necessary infrastructure to centrally organize and manage the vast amounts of data required for AI models. Storage technologies ensure quick access to data, directly impacting the performance of AI applications. Additionally, scalable storage systems support the growing data needs as AI projects expand and evolve.

Report insight(s)

Data storage market is driven by data architecture segment growth. As data volumes continue to grow, the report forecasts an 18% CAGR for the data architecture sub-segment until 2030, highlighting the increasing importance of organizing data to generate valuable insights. In contrast, storage technologies are expected to see a below-average growth rate of 8% CAGR during the same period. The report notes that hardware storage costs have significantly declined over the past decade. For example, in 2016, the cost of memory was \$203 per terabyte, which has since decreased to \$49.50 per terabyte for solid-state storage. This trend is expected to contribute to the below-average growth rate in the storage technologies market sub-segment.

Vector databases are rising in popularity for advanced generative AI applications. Additionally, the report notes the growing popularity of vector databases for generative AI use cases. These databases are crucial for indexing and searching high-dimensional vectors used for similarity searches and pattern matching. This trend indicates a shift towards new, advanced data management systems for specific AI applications (in this case Generative AI).

"We're seeing significant interest in our vector search offering from large and sophisticated enterprise customers, even though it's only still only in preview. As one example, a large global management consulting firm is using Atlas Vector Search for an internal research application that allows consultants to semantically search over 1.5 million expert interview transcripts."

Dev Ittycheria, president and CEO at MongoDB, 23 September 2023

Data management component 4: Transformation

4

Transformation

Description

Transformation refines and restructures data into formats suitable for detailed analysis. This part of the stack involves cleaning, integrating, and modifying data to ensure quality and compatibility with analytical tools and storage structures. ETL (extract, transform, load) plays a crucial role in this stage by extracting data from various sources, transforming it into a standardized format, and loading it into target storage. This process ensures that data is clean, structured, and ready for analysis.

Importance for AI

Transformation prepares the data for AI. Data transformation is essential for AI because it converts raw data into clean, structured formats, making it digestible for AI to train and operate. This process includes file format conversions, data cleansing, securing sensitive data (especially important for generative AI), and aggregating data to support frequent queries. Both predictive AI and generative AI require preprocessed data to maintain data quality and usefulness.

Report insight(s)

Reverse ETL is important for integrating AI-generated insights into business processes. Unlike traditional ETL, which moves data into a centralized storage system (e.g. centralized data warehouse, data lake, or cloud) for analysis, reverse ETL extracts data from these systems and syncs it back to operational applications. By moving AI-generated insights to systems like ERP, reverse ETL enables organizations to integrate AI findings into the business processes, ensuring that these insights can be promptly applied to enhance business operations, decisions, and many more.

Data management component 5: Analytics

5

Analytics & data science

Description

Analytics converts data into meaningful and actionable information. This part of the stack consists of:

- **Business intelligence tools.** Translates data into visual reports, dashboards, and metrics, making it easier to understand and communicate insights.
- **Data science tools.** Enhances analytics by identifying deeper patterns, trends, and correlations that may not be immediately visible through traditional methods.

Importance for AI

Analytics tools help build and maintain AI models. Analytics is crucial for AI because it provides the necessary tools to develop and refine AI models. By leveraging techniques such as data mining, statistical analysis, and machine learning, analytics helps uncover patterns and trends, extracting insights and knowledge from both structured and unstructured data.

Report insight(s)

Analytics is the fastest-growing data management market segment. The report forecasts a 20% compound annual growth rate (CAGR) for the analytics market segment until 2030. Among the two sub-segments, the data science segment is expected to grow the fastest, with a 27% CAGR, while the business intelligence segment is projected to grow at a 16% CAGR. These projections underscore the increasing significance and investment in data-driven decision-making.

To illustrate this trend, the report showcases a case study on how US-based online homestay marketplace **Airbnb** leverages AI to improve its host-guest matching process, using techniques like A/B testing, image recognition, and predictive modeling to enhance user experience and increase bookings. Their models predict booking probabilities based on user searches, and their price tip feature advises hosts on optimal pricing. Additionally, to address high bounce rates among certain Asian visitors, Airbnb made site modifications, boosting conversion rates by 10%.

Data management component 6: Data governance and security

6

Data governance & security

Description

Data governance and security ensures organizational data's integrity, usability, and consistency through policies, processes, and roles, underpinning its trustworthiness for business operations.

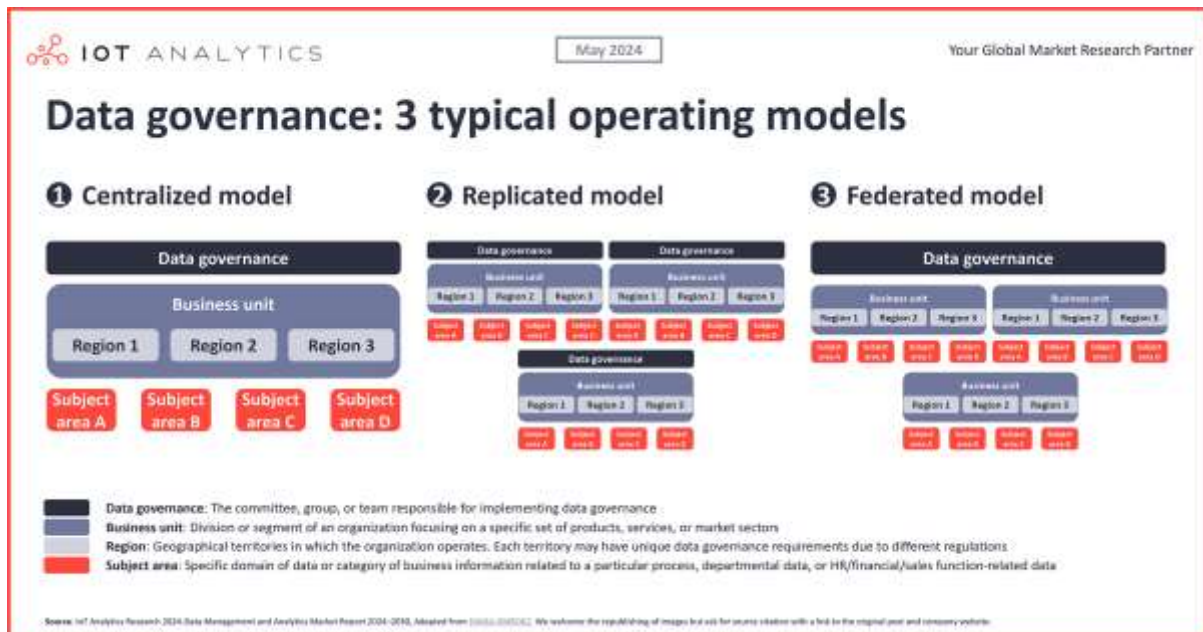
Importance for AI

Data governance is important for data integrity. Data governance and security are critical for AI because they protect both the data and AI models. This is essential for developing accurate, ethical AI models and safeguarding the intellectual property invested in these models. Poorly governed data risks leaking proprietary and private information, potentially resulting in fines and negative publicity. Additionally, robust measures ensure that the data used are accurate and uncompromised, enhancing the quality of AI models. They also prevent information breaches and unauthorized access to the AI and their underlying model code.

Report Insight(s)

The report identifies three typical operating models for implementing data governance.

1. **Centralized model:** A single data governance entity oversees and standardizes activities across all subject areas within the organization.
2. **Replicated model:** Each business unit independently adopts and implements the same data governance model and standards.
3. **Federated model:** The data governance body coordinates with multiple business units to ensure uniformity in definitions and standards across the organization.



Overview of 3 typical operating models: centralized, replicated, and federated.

Data management component 7: Data orchestration

7

Data orchestration

Description

Data orchestration is the systematic management and coordination of data flows across different systems and services. It involves automated data movement, ensuring availability in the appropriate format and location for analysis and decision-making.

Importance for AI

Data orchestration coordinates the movement of data. Data orchestration ensures the seamless integration, coordination, and flow of data across various systems and facilitates AI model training, deployment, and refinement. It enables the automated and coordinated movement of data across various systems, ensuring their availability in the appropriate format and location when needed for pre- and post-data processing.

Report insight(s)

The goal is to automate the data flow. The report showcases a case study on how US-based software company **Cox Automotive** leverages **AWS** Step Functions to automate model deployment, enhancing model creation, team productivity, and precision. This automated workflow has allowed their data scientists and decision teams to focus on higher-value tasks, improving overall performance and quality. Additionally, by reducing the manual overhead in model retraining and deployment, Cox Automotive has increased the speed and accuracy of its AI-driven initiatives, leading to more reliable and timely decision-making.

Analyst takeaways on the current data management market

Takeaway 1: Hyperscalers dominate the data management market with integrated, cost-effective services despite less specialization.

The 3 main hyperscale's together, **AWS**, **Microsoft**, and **Google**, dominate the data management market with a combined 52% market share in 2023, and leading services in each of the market segments outlined above. It is important to note that in many cases, these companies do not offer the most sophisticated and best-rated solutions, but they provide cost-effective and highly integrated services that are easy to scale for their customers. Companies face an important choice: They can opt for several best-of-breed data management solutions from smaller, specialized firms, or they can take advantage of the convenience and comprehensive offerings of one or two hyperscale's, which deliver all necessary services under one umbrella.

"Hyperscalers like AWS, Microsoft, and Google dominate the data management market with highly integrated portfolios across all major market segments. There are also some quickly growing data management scale-ups that are regarded as having a best-in-class offering and are subsequently enjoying strong market traction. It will be interesting to see whether companies will opt for the convenience of having everything from one vendor or settle on three to five main data management solutions on top of their cloud architecture."

Knud Lasse Lueth, CEO at IoT Analytics

Takeaway 2: Some executives fail to see the complete picture: There can be no winning AI strategy without a winning data management strategy.

In Q1 2024, the CEOs of nearly one-third of all companies discussed AI in their earnings calls, yet only a very small portion (1% of all earnings calls) mentioned data management. While AI holds transformative potential, its success is deeply intertwined with comprehensive data management practices. My observation is that some C-level executives exhibit a short-sighted approach, focusing on the prestige of AI without recognizing the essential role of a robust data strategy. As this article highlights, winning AI strategies rely on strong data management as their foundation, where all seven components—data sourcing, ingestion, storage, transformation, analytics, governance and security, and orchestration—work seamlessly to support the AI strategy.

"C-level executives often overlook the critical importance of data management for AI. Strong data management is the foundation for successful AI implementation. AI holds transformative power, and their successful implementation brings significant prestige to leadership. However, the less glamorous groundwork—a well-executed data management strategy—is frequently neglected. The "Data Management and Analytics Market Report 2023–2030" highlights how excellent data management strategies enable successful AI initiatives and the adverse effects of poor data management."

Oktay Demir, COO at IoT Analytics

Takeaway 3: Data fabric is emerging as an advanced evolution of the data lake.

Data fabric is a relatively new term, describing a comprehensive data integration and management framework, encompassing architecture, data management tools, and shared datasets, designed to assist organizations in handling their data. It differs from data lakes in that it goes beyond storing raw data, and it differs from data warehouses in that it handles only processed or refined data. Data fabrics offer a cohesive, consistent user interface (UI) and real-time access to data for all members of an organization, regardless of their global location.

"In my opinion, data fabric is still not very popular in terms of adoption as it can come with a heavy price tag due to unfit data architectures. However, given the increase in data complexity because of the exponential growth in big data, propelled by hybrid cloud, AI, IoT, and edge computing, there seems to be a good opportunity for vendors in this scenario."

Mohammad Hasan, Analyst—software and cloud at IoT Analytics

For more information or media inquiries, please contact:

Hoang Pham Van

IoT Analytics
+49 (0) 40 6391 1891
[press\(at\)iot-analytics.com](mailto:press(at)iot-analytics.com)

For further reading please visit:

www.iot-analytics.com/research-blog

About IoT Analytics

IoT Analytics, founded and operating out of Germany, is a leading global provider of market insights and strategic business intelligence for the IoT, AI, Cloud, Edge, and Industry 4.0.

Our key workstreams across the tech stack include IoT applications, IoT platforms and software, IoT connectivity and hardware, and industrial IoT.

We are trusted by 1000+ leading companies around the world for our market insights, including globally leading software, telecommunications, consulting, semiconductor, and industrial players.

###