

F R O S T & S U L L I V A N

沙利文



头豹
LeadLeo

2023 China Data Management Solutions Market Report

Data Lakehouse、Data Intelligence、Cloud Native、Data Governance

2023.6

LeadLeo Research Institute
Frost & Sullivan (China)

Table of Content

◆ Technological trends of data management tracking	-----	5
<ul style="list-style-type: none">• The two major trend themes revolving around technological dynamics• Data and intelligent collaboration drive technological innovation• Cloud native become more accessible in data management solutions• Methodology helps improve data management solutions• Build secure data flow and sharing capabilities		
◆ Market Analysis of Data Management Solutions in China	-----	18
<ul style="list-style-type: none">• Opportunities and Challenges in Market Implementation• Analysis of the market deployment layout on the supply side• Industry coverage map of suppliers		
◆ Competitive landscape in China's data management solutions market	-----	27
<ul style="list-style-type: none">• Assessment Scoring• Comprehensive Vendors Assessment – Frost Radar• Leading competitors		
◆ Terms	-----	36
◆ Methodology	-----	37
◆ Legal Disclaimer	-----	38

Figures & Charts

• The two major trend themes around which technological dynamics revolve	-----	5
• The relationship between humans and AI, data and AI	-----	8
• The intelligent capabilities that AI brings to data management solutions	-----	9
• AI is proficient in basic capabilities to perform tasks but lacks the ability to autonomously handle complex processes	-----	10
• Traditional cloud hosting data management solutions and cloud native data management solutions	-----	12
• Value of cloud native data management solutions	-----	12
• The dynamics of cloud-native and serverless aspects in the supply and demand side	-----	13
• Three architectures in the way of migrating to the cloud	-----	14
• Introduction to the three major data management methodologies	-----	15
• DataOps and Data Fabric can serve as the implementation support for Data Mesh	-----	16
• The Data Security Law and the Personal Information Protection Law relevant content	-----	17
• Data value to economy, society and businesses has been enhancing, driving the development of data management	-----	19
• Proportion of Participating Units by DCMM Capability Level and Changes in the Number of Participating Units	-----	20
• Data Management Policy Compilation	-----	21
• User concerns regarding serverless can be addressed with suitable solutions	-----	22
• Different industries have varying demands due to their distinct business requirements and foundational capabilities	-----	23
• The price reduction of cloud computing products will enhance technology promotion	-----	24
• Representative suppliers that have achieved enhanced industry-specific capabilities include	-----	25
• With the joint efforts of suppliers, data management solutions have achieved coverage in multiple industries	-----	26
• Assessment Scoring	-----	28
• Comprehensive Vendors Assessment – Frost Radar	-----	30
• Leading competitors	-----	31

1.1 The two major trend themes revolving around technological dynamics

Key findings

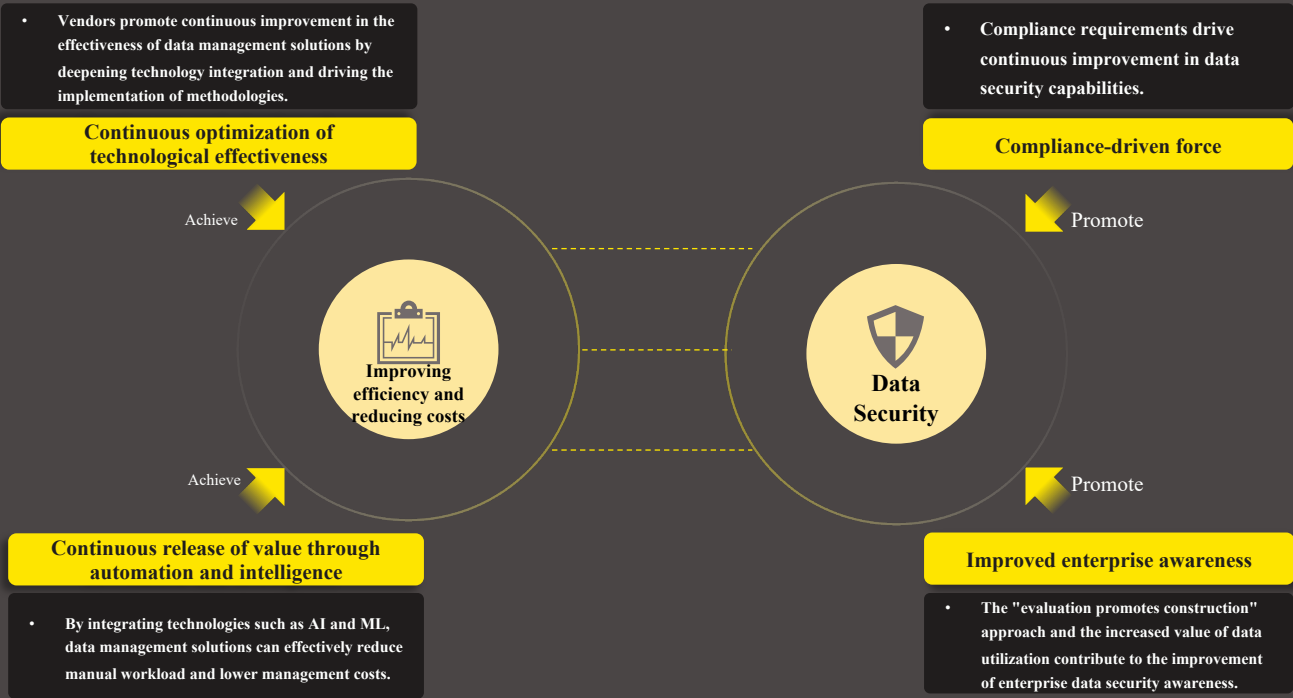
Vendors' technological dynamics revolve around two main themes: improving efficiency and reducing costs, and ensuring data security. This is done to enhance the commercial implementation effectiveness of data management solutions, strengthen the alignment between data management solutions and the sustainable development of enterprise data management capabilities, and facilitate the flow and release of data value.

□ Improving efficiency and reducing costs, and ensuring data security

With the continuous optimization of data management technology in recent years, coupled with the increasing attention from the government on the development of the data field, Chinese enterprises have strengthened their practical implementation of data management technology, and the level of data management is gradually improving. However, enterprises still face the challenge of not fully unleashing the value of data even after implementing the technology. This is primarily due to the fact that in the process of digital transformation, enterprises need to iterate their technology while considering their existing IT architecture and organizational structure. Merely pursuing new technologies without considering the inherent attributes of the enterprise can result in a failure to connect data value with the company's business development. For example, technological complexity exacerbates data silos, and increased communication costs between data teams and business teams.

In this context, we observe that vendor's' technological dynamics revolve around two main themes: improving efficiency and reducing costs, and ensuring data security. By implementing technology integration and methodologies, they enhance the performance and implementation effectiveness of solutions in specific scenarios while reducing costs. Additionally, by enriching data security functionalities, they facilitate data circulation. These strategies will strengthen the alignment between data management solutions and the sustainable development of enterprise data management capabilities, thereby improving the effectiveness of solution implementation.

The two major trend themes around which technological dynamics revolve



01 Improving efficiency and reducing costs: continuous release of value through automation and intelligence

□ Advancements in AI technology have expanded its role in the field of data management

In the early stages, automation and intelligence were primarily based on predefined rules and models, which could alleviate some of the pressures on data management teams. However, the functionalities and impact were limited. With the improvement of AI technology, the range of achievable intelligent capabilities has expanded. These include predictive functions, adaptive functions, natural language understanding, and more. These advanced capabilities not only help businesses further alleviate team pressures and gain better data insights, but also reduce data management and operational costs. They also lower the barriers to data usage and expand the range of data users. This in turn helps businesses address both technology implementation and organizational collaboration challenges, making data more usable, valuable, and accessible. Ultimately, it empowers business development by leveraging the full potential of data.

02 Improving efficiency and reducing costs: continuous optimization of technological effectiveness

□ By strengthening cloud native development and integrating methodologies into solutions, DMS technical performance has been continuously optimizing

Technologies such as data lakes, data warehouses, and HTAP have gained traction and implementation in recent years as they can address challenges brought by business growth. However, the full potential of these technologies has yet to be fully realized. The main obstacles lie in the need to integrate new technologies with existing IT and organizational architectures. Common challenges include the learning curve required for the collaboration between new technologies and existing IT architectures and the increased operational pressure after technology implementation. Additionally, the complexity of the system is heightened with technology implementation, while the data usage processes remain unchanged or difficult to update, resulting in technology implementation that falls short.

To address these challenges, vendors offer solutions based on cloud-native approaches combined with methodologies to unlock the value of technology. For example, cloud-native solutions can make the implementation more lightweight and provide higher performance ceilings, reducing operational pressures and improving performance. Methodologies provide guidance from a holistic perspective to build new architectures, thus enhancing data management efficiency and organizational collaboration. By choosing these approaches, businesses can gradually overcome obstacles and fully leverage the efficiency of the technology.

03 Data security: The drive for compliance and increased awareness among enterprises has elevated the importance of data security capabilities

□ The data security capabilities directly impact whether the value of data can be fully explored and unleashed

The core issue of data security lies in ensuring the compliant and secure flow of data internally and externally, and the extent of this flow determines whether the value of data can be truly explored and unleashed. With the increasing awareness of enterprises in data management and the drive for compliance, the importance of data security has been recognized, leading to an increased willingness to invest in this area. Vendors assist enterprises in building data security capabilities by providing methods and tools for data security management and data sharing. This facilitates the compliant, secure, and smooth flow of data, enabling its full utilization and value extraction.

1.2 Data intelligence collaboration drive technological innovation

Key findings

AI and data have a natural positive synergy, which makes the integration of AI and data management essential for technological innovation. By leveraging AI, data management solutions can not only optimize the effectiveness of technology implementation but also facilitate collaboration between teams to enhance organizational synergy, thereby fully harnessing and unleashing the value of data.

AI and data management have a natural mutually empowering relationship

In an environment characterized by data complexity, large volumes of data, and rapid business changes, enterprises require more efficient data management solutions. AI brings technological innovation to data management, and the relationship between the two is not unidirectional but rather a synergistic collaboration, further strengthening the inevitability of their integration.

AI empowers data stakeholders with powerful tools and decision support, reducing the cognitive bandwidth consumed by complex and mundane tasks for data teams. This enables the team to focus more on business delivery rather than tool management, resulting in improved effectiveness and efficiency in data management. Simultaneously, optimized data management leads to improved data quality and effective management of diverse data from multiple sources, thereby supporting the performance and training efficiency of AI. As a result, AI and data management have a natural mutually empowering relationship.

With the development of large AI models, the capabilities of scenario-based applications are enhanced, and new technological functionalities arising from the integration of data intelligence are being realized. Sullivan believes that the current trend of “data intelligence fusion” has reached its peak in terms of attention, and the next stage in this trend will be the transition from “AI for DMS” (Data intelligence integration) to “AI in DMS” (Data Management Systems with AI-enabled Unified Governance).

The relationship between humans and AI, data and AI



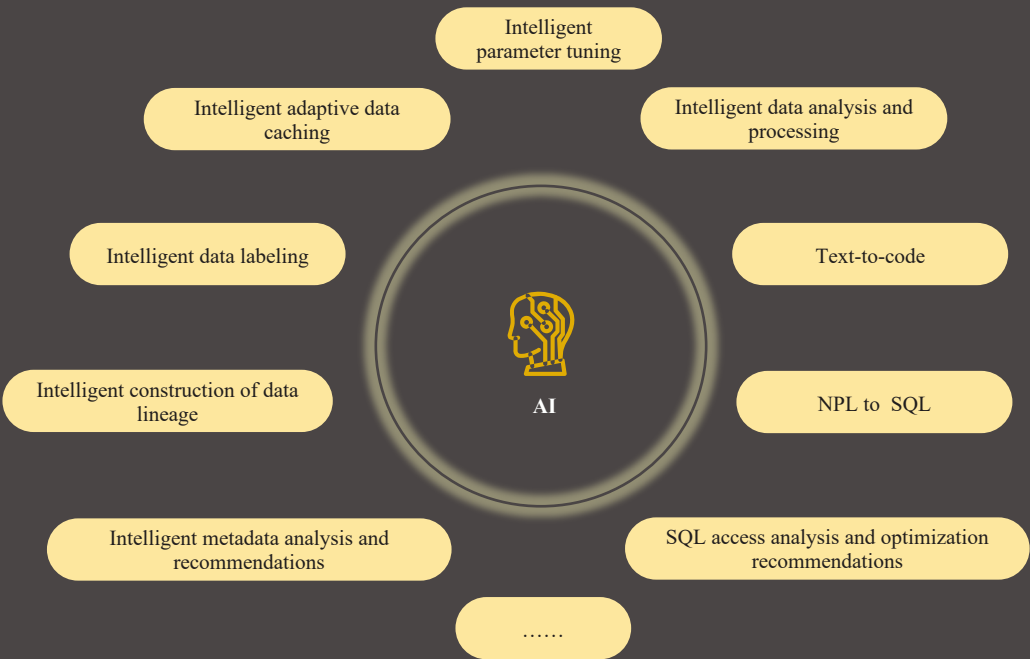
Source: Sullivan

❑ AI can not only optimize technology but also promote organizational collaboration

The integration of AI into data management solutions brings value primarily in two aspects: enhancing the effectiveness of technology implementation in data management and promoting organizational collaboration.

- **Enhancing the effectiveness of technology implementation:** AI extends and expands the functionality of data management, enabling capabilities that were previously unachievable or difficult to achieve manually. For example, AI can analyze SQL access patterns and habits to troubleshoot performance issues in databases and related products, providing optimization recommendations. In environments without AI, these performance issues may be challenging to detect or not significant enough to warrant immediate attention, resulting in the accumulation of problems. Additionally, AI technology can continuously learn and optimize through data-driven approaches, adapting to evolving data management needs and challenges, and improving the practical application of data management technology.
- **Organizational collaboration:** The introduction of AI technology reduces the barriers to data utilization, promoting data democratization. For example, business users can leverage natural language to SQL conversion capabilities for data retrieval, reducing communication costs and process time, and improving data analysis efficiency. Previously, the data retrieval process might have required individual requests, leading to inefficiencies caused by misunderstandings or queuing multiple requests from different teams. Furthermore, AI technology can enhance collaboration between data engineers and data scientists. For instance, AI-based data flow monitoring enables timely and accurate issue detection, reducing friction between the two parties and facilitating efficient data development.

The intelligent capabilities that AI brings to data management solutions

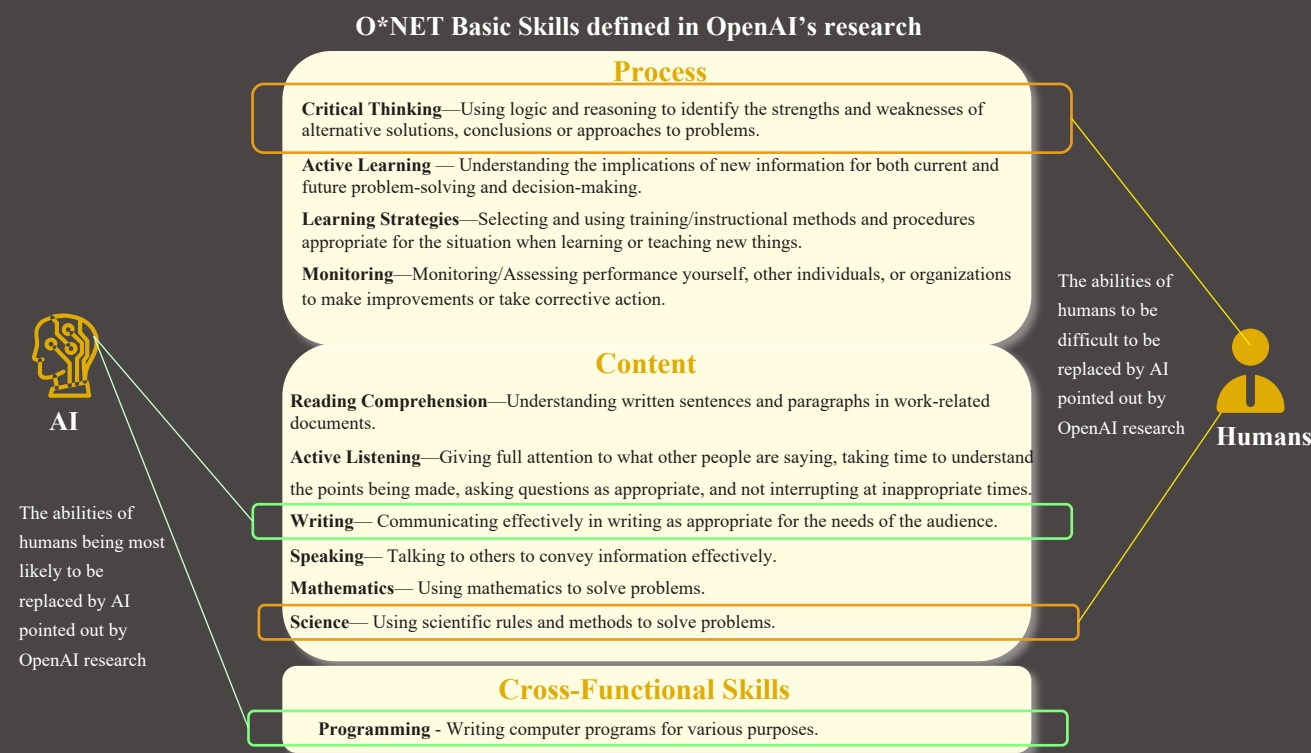


❑ Impact of AI on Data Management Practitioners

Different from the use of ML and AI to enhance data management technologies, the emergence of powerful models like GPT-4 poses certain challenges for data analysts, data scientists, DBAs, data development engineers, and other data management practitioners. How will they be impacted?

In the field of data management, the focus on large language models should be placed on "task augmentation" rather than "human replacement." This means that AI can take over tasks that humans don't want to do or aren't skilled at, thereby freeing up human cognitive capacity to concentrate on more important business delivery. From the perspective of required skills for the profession, after training and dynamic optimization, AI can essentially perform tasks that require basic competencies. However, it faces challenges in tasks that require deep processing capabilities and lack human guidance and supervision. This is because such tasks are influenced by multiple factors, such as varying standards of right and wrong, communication, collaboration, empathy, and the application of multidimensional knowledge, experience, and skills.

AI is proficient in basic capabilities to perform tasks but lacks the ability to autonomously handle complex processes



The current data management processes and architectures will undergo upgrades with the integration of AI. What humans need to do is understand and learn AI, leveraging its capabilities to empower decision-making. This requires:

- 1. Enterprises need to strengthen training to improve employees' AI literacy, with a focus on critical thinking and problem-solving skills.
- 2. Data teams and AI teams should learn how to collaborate effectively, optimizing workflows and organizational structures to achieve fruitful data and AI synergy.
- 3. Data stakeholders should learn to leverage AI to continuously optimize their data management practices, leveraging their experience and knowledge while embracing the evolution of AI.
- 4. Enterprises should further advance high-quality data governance transformation initiatives, breaking down "information silos," harnessing the value of "dark data," and transforming data into valuable assets. "No governance, no analysis": Trustworthy AI requires access to massive amounts of high-quality data.

Source: "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models", LeadLeo

01 Key trends brought about by data intelligence integration: The positions in the field of data management will be further specialized

❑ Humans remain crucial in the era of data intelligence and need to adapt to the ever-changing landscape of data management

Taking cues from the software engineering field, where the role of "software engineer" has evolved into more specialized roles such as DevOps engineers or Site Reliability Engineers (SRE), it is anticipated that the data engineering field will also witness the emergence of more positions defined based on business needs or objectives. These may include:

- Data Reliability Engineers: Responsible for ensuring data quality.
- Data Product Managers: Drive the application and monetization of data.
- DataOps Engineers: Focus on data governance and efficiency improvement.
- Data Architects: Focus on addressing data silos and long-term investments.

Existing positions such as data analysts, data scientists, database administrators (DBAs), and data development engineers may undergo some modifications due to the impact of AI-driven self-service processes. However, we believe that these positions will not disappear entirely. While ChatGPT has received significant attention, the process heavily relies on the expertise of professional data practitioners for review and adjustments. Given the challenges posed by large-scale language models, such as low interpretability, ethical and compliance concerns, uncertain data privacy security, and low accuracy in quality control, human involvement remains crucial in data management. Although infrastructure may change, and automation technologies may alter how we spend time and attention, in the foreseeable future, professional data management practitioners will continue to play a leading role in the data management process.

02 Key trends brought about by data intelligence integration: Intelligent Unified Governance - AI in DMS

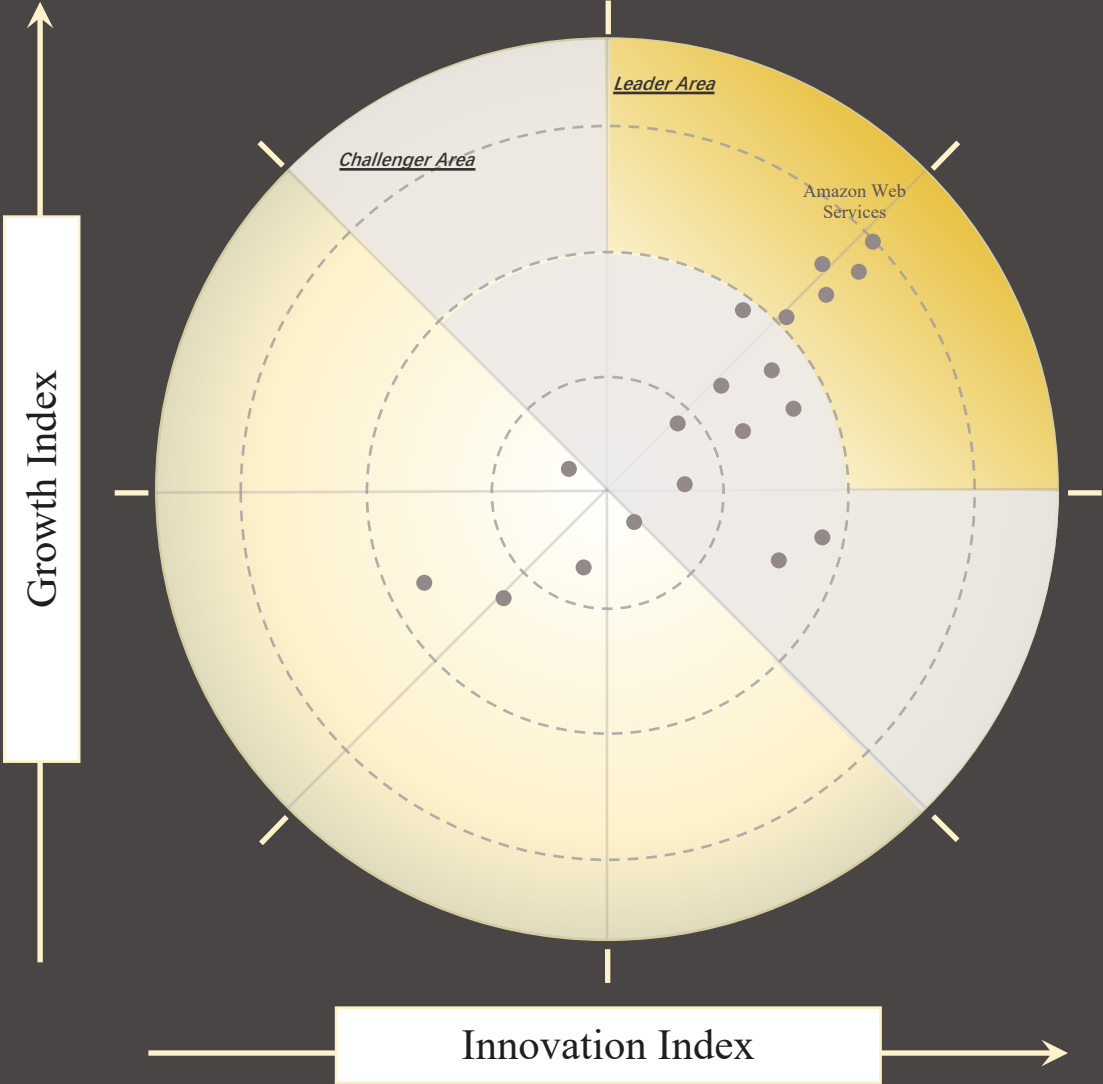
❑ AI in DMS will significantly enhance the user experience of data management solutions.

AI in DMS (Artificial Intelligence in Data Management Solutions) involves not only the use of AI technology in data management systems but also the integration of AI platforms as built-in features of data management systems, shifting AI from an auxiliary tool to a core middleware. This deep integration allows data management systems to better process and understand large volumes of data, thereby improving the efficiency and accuracy of data analysis.

In the environment of this intelligent unified platform, the entire lifecycle of data value extraction can be covered, from data foundation to AI development and various data analysis applications. This unified production on the same platform not only enhances the efficiency of data processing but also ensures data consistency and accuracy.

AI in DMS will become a new paradigm for data management. By deeply integrating AI technology into data management systems, it enables a profound integration of data and intelligence, thereby enhancing the overall efficiency and accuracy of data management systems. With the integration of intelligent dual systems, experiences such as intelligent application of streaming data, real-time data visualization, and business-driven data and AI development will become new norms.

Comprehensive Competitive Performance of China’s Data Management Solutions Market——Frost Radar™



Note: The circle corresponds to the comprehensive competitive performance from low to high according to the logic of increasing from the inside to the outside, and the competitiveness is obtained by combining the “innovation index” and “growth index”. The conclusions only apply to the development of the data management solutions market at this stage.

- ❑ **The vertical coordinate represents the “growth index” :**
 - It measures the competitiveness of competitive entities in the data management solutions growth dimension, the higher the position, the stronger the market growth capability and level of data management solutions in terms of performance and compatibility, security capabilities, services support, ecological construction, maturity of solutions for meeting market’s needs, etc.
- ❑ **The horizontal coordinate represents the “innovation index” :**
 - It measures the competitiveness of competitive entities in the data management solutions growth dimension, the more the position is to the right, the stronger the innovative technologies or capabilities of data management solutions such as data warehouse & data lake basic technology improvement, cloud native innovation, data lakehouse innovation, data management methodology innovation, etc.

Leader: Amazon Web Services

Key findings

Amazon Web Services is intensifying its cloud-native data strategy, continuously enhancing the technological advancement of data management solutions. By strengthening the performance of product integration, combining automation and machine learning to provide intelligent services, and optimizing the user experience and effectiveness of solution implementation, Amazon Web Services is a Leader of China data management solutions.

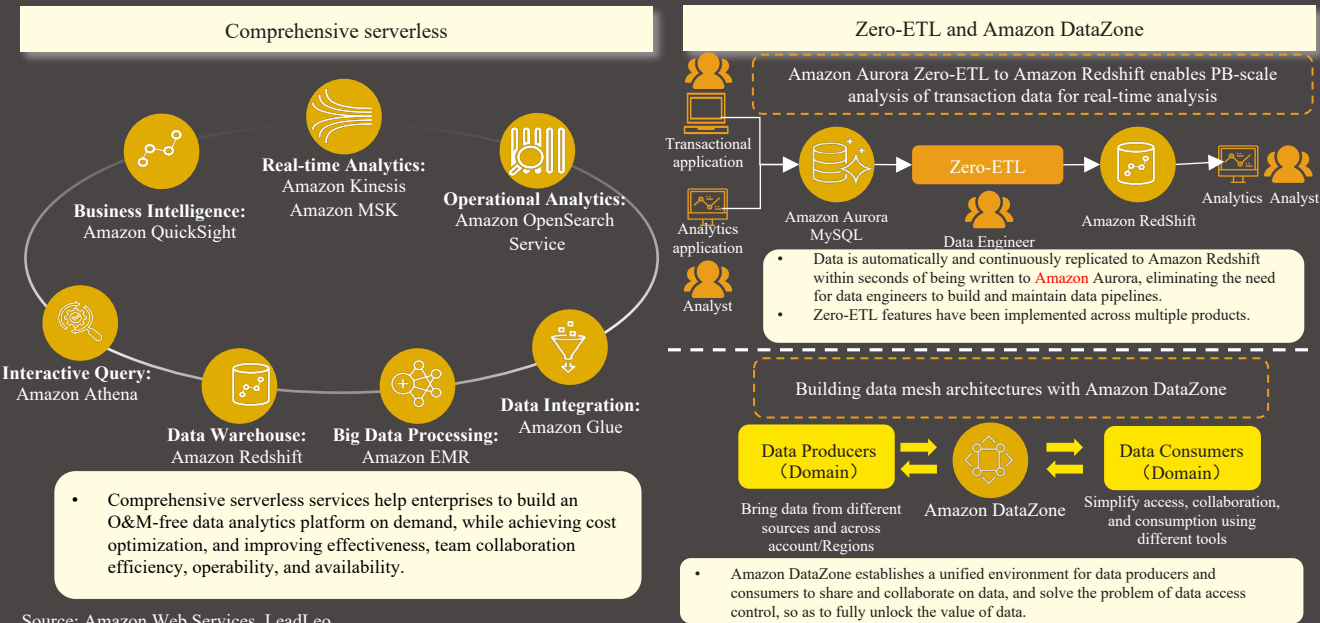
❑ Improving cloud-native product portfolio and continuously optimizing the user experience and effectiveness of solution implementation.

The analysis and management of data within enterprises dynamically evolve over time, accompanied by challenges such as the difficulty of introducing data technologies and the increased cost of data management. Amazon Web Services has established a comprehensive cloud-native product portfolio to help enterprises address these challenges. One notable aspect is that the analytics services provided by Amazon Web Services have achieved full serverless capabilities this year, which enables enterprises to achieve cost efficiency and increased productivity through the integration of managed services, while also being able to quickly adapt to changes in business scale, logic, and requirements. Additionally, the level of intelligence in Amazon Web Services products plays a crucial role in optimizing the user experience. For example, AWS Glue FindMatches utilizes machine learning capabilities to automatically identify duplicate or matching records within datasets, enabling functionalities such as associating customer records, deduplicating product catalogs, and fraud detection.

❑ Achieving cross-organizational and efficient data integration

To address the challenge of improving enterprise data analytics efficiency, Amazon Web Services proposes the solution of 'Zero-ETL,' which reduces the manual migration or transformation of data between different services. This helps enterprises save time by minimizing or eliminating the need to build and maintain ETL (Extract, Transform, Load) data pipelines. Additionally, the efficiency of data sharing across organizations can be affected by issues such as data unavailability and overly strict or lenient data access controls. Amazon DataZone, by establishing a unified environment, assists enterprises in data classification, discovery, management, sharing, and analysis. It significantly reduces collaboration barriers among organizations and improves overall efficiency.

Amazon Web Services have been augmenting its products for better user experience



Source: Amazon Web Services, LeadLeo

Terms

- ◆ **Data Warehouse:** A strategic collection of all types of data to support the decision making process at all levels of the enterprise.
- ◆ **Data Lake:** centralized repository that allows you to store all your structured and unstructured data at any scale.
- ◆ **Data Lakehouse:** a new open architecture that combines the flexibility and scalability advantages of data lakes with the data management capabilities of data warehouses.
- ◆ **Microservices:** It is an architectural style and design pattern that advocates for breaking down applications into a series of fine-grained services, with each service focusing on a single business function and running in independently deployable processes. The boundaries between services are clear.
- ◆ **Monolithic Architecture:** It is a traditional architectural style and design pattern where all business modules run, develop, and deploy in a single point.
- ◆ **Serverless Deployment:** Serverless deployment provides services through Functions-as-a-Service (FaaS) and Backend-as-a-Service (BaaS), allowing users to develop, run, and manage applications without the need to build and maintain complex infrastructure.
- ◆ **Containers:** Containers are lightweight, portable, self-contained software packaging technologies that allow applications to run in the same way almost anywhere. Developers create and test containers on their own laptops and can run them on virtual machines, physical servers, or public cloud hosts without any modifications.
- ◆ **DataOps:** DataOps is a concept aimed at achieving the maximum automation of the entire data pipeline, from data collection and cleansing to analysis. It is not a specific technology or toolset but a philosophy. Nowadays, pure business applications no longer exist, and data-driven applications are beginning to impact the experience of enterprises. DataOps provides an opportunity for development. DataOps consists of three main components: methodology, technical tools, and organizational culture. The combination of these three components can improve data management efficiency.
- ◆ **Data Fabric:** Data Fabric is an emerging data management design concept that enables enhanced data integration and sharing across heterogeneous data sources. It supports cross-platform design, deployment, and usage of data systems by continuously analyzing existing discoverable and inferable metadata assets, thereby enabling flexible data delivery.
- ◆ **Data Mesh:** Data Mesh is a data management approach using a distributed architectural framework. It aims to address the bottlenecks and challenges of traditional data management. Data Mesh advocates for transforming data from a single centralized management model to decentralized and autonomous data units through Data Mesh, enabling data decentralization, collaboration, and autonomy.

Methodology

- ◆ Frost & Sullivan has conducted in-depth research on the market changes of 19 major industries and 532 vertical industries in China with more than 1,000,000 industry research samples accumulated and more than 10,000 independent research and consulting projects completed.
- ◆ Rooted on the active economic environment in China, the research institute, starting from data management and big data fields, covers the development of the industry cycle, follows from the enterprises' establishment, development, expansion, IPO and maturation. Research analysts of the institute continuously explore and evaluate the vagaries of the industrial development model, enterprise business and operation model, Interpret the evolution of the industry from a professional perspective.
- ◆ Research institute integrates the traditional and new research methods, adopts the use of self-developed algorithms, excavates the logic behind the quantitative data with the big data across industries and diversified research methods, analyses the views behind the qualitative content, describes the present situation of the industry objectively and authentically, predicts the trend of the development of industry prospectively. Every research report includes a complete presentation of the past, present and future of the industry.
- ◆ Research institute pays close attention to the latest trends of industry development. The report content and data will be updated and optimized continuously with the development of the industry, technological innovation, changes in the competitive landscape, promulgations of policies and regulations, and in-depth market research.
- ◆ Adhering to the purpose of research with originality and tenacity, the research institute analyses the industry from the perspective of strategy and reads the industry from the perspective of execution, so as to provide worthy research reports for the report readers of each industry.

Legal Disclaimer

- ◆ The copyright of this report belongs to LeadLeo. Without written permission, no organization or individual may reproduce, reproduce, publish or quote this report in any form. If the report is to be quoted or published with the permission of LeadLeo, it should be used within the permitted scope, and the source should be given as "LeadLeo Research Institute", also the report should not be quoted, deleted or modified in any way contrary to the original intention.
- ◆ The analysts in this report are of professional research capabilities and ensure that the data in the report are from legal and compliance channels. The opinions and data analysis are based on the analysts' objective understanding of the industry. This report is not subject to any third party's instruction or influence.
- ◆ The views or information contained in this report are for reference only and do not constitute any investment recommendations. This report is issued only as permitted by the relevant laws and is issued only for information purposes and does not constitute any advertisement. If permitted by law, LeadLeo may provide or seek to provide relevant services such as investment, financing or consulting for the enterprises mentioned in the report. The value, price and investment income of the company or investment subject referred to in this report will vary from time to time.
- ◆ Some of the information in this report is derived from publicly available sources, and LeadLeo makes no warranties as to the accuracy, completeness or reliability of such information. The information, opinions and speculations contained herein only reflect the judgment of the analysts of leopard at the first date of publication of this report. The descriptions in previous reports should not be taken as the basis for future performance. At different times, the LeadLeo may issue reports and articles that are inconsistent with the information, opinions and conjectures contained herein. LeadLeo does not guarantee that the information contained in this report is kept up to date. At the same time, the information contained in this report may be modified by LeadLeo without notice, and readers should pay their own attention to the corresponding updates or modifications. Any organization or individual shall be responsible for all activities carried out by it using the data, analysis, research, part or all of the contents of this report and shall be liable for any loss or injury caused by such activities.