IBM Technology Atlas

Data roadmap

Business innovation driven by generative AI is fueled by open data stores, formats, and engines; a product-oriented data fabric; and the infusion of AI at all levels to radically improve data consumption.

Updated May 2024

- ✓ completed
- → pushed to next year
- ↻ on target

	2023	2024	2025	2027	2029	2030+
Data journey	✓ ***Increase data products by orders of magnitude as open formats become mainstream.***	↻ ***Make data consumption and management smarter with generative AI.***	***Enable ubiquitous multicloud workloads with open engines/ formats attaining parity.***	***Democratize governed data consumption via generative AI to bolster productivity.***	***Leverage composable hardware going mainstream and plummeting egress costs.***	***Make instant permissioned data easily accessed anywhere.***
Strategy overview	✓ In 2023, the number of trusted, discoverable data products (high-quality, centrally governed enterprise datasets) will increase 10-100x via a hybrid data fabric. Many of these data products will be stored in open table formats (e.g., Iceberg), which have become mainstream.	↻ In 2024, AI will be used in data consumption and management, and open metadata will become a differentiator. The line will blur in management of structured and unstructured data driven by AI.	By 2025, the performance and management of pluggable open engines processing open-format data will reach parity with proprietary online analytical processing (OLAP) solutions, enabling secure, multicloud, data-centric workloads to become ubiquitous.	By 2027, all forms of data consumption will be democratized, from basic reporting to predictive insights, sophisticated analytics, and decision-making. Employees regularly driving actions with data will triple, boosting productivity.	In 2029, composable hardware will go mainstream. This will transform networks, drive a 3x cost reduction due to disaggregated hardware, and push egress costs down.	Beyond 2030, the number of permissioned data products that any enterprise can instantly find and access regardless of the product location—on devices, edge, data center, SaaS, etc.—will grow by orders of magnitude.
Why this matters to our clients and the world	✓ The data product paradigm will set enterprise standards for data quality, governance, semantics, and observability. ✓ AI will scale data production and management of workflows. ✓ Open formats will enable interoperability.	↻ Self-service data consumption will increase return on investment. Less reliance on expert data skills will lead to more and more diverse data consumers. Secure, cost-efficient data operations will reduce silos and raise productivity.	Enterprises will be able to prepare and access data products simply and securely, independent of location. The parity of open formats and exchangeable engines will eliminate lock-in via approaches like OneTable.	True data-driven enterprises will empower employees in every role, function, and skill level to leverage data and its insights to drive efficiencies, increase competitiveness, and gain advantages.	Major hardware trends will disrupt data stores and processing engines. Lower egress fees will remove vendor lock-in and change the cost dynamics of multicloud computing.	Enterprises will have a true 360º view of all data assets and the ability to leverage all the information to radically improve business insights and actions.
The technology or innovations that will make this possible	→ Advances in AI will power semantic augmentation, management, and governance. ✓ Formats like Iceberg and Parquet will provide an open approach for storing data for analytics and AI. → On-demand, policy-driven data replication and caching will address sovereignty, compliance, and performance.	↻ AI-based agents will help users to easily find and create high-quality data views and custom pipelines. Open metadata enables integration and interop through the data ecosystem. Multicloud data observability and transparent access to data will emerge. ↻ Unstructured data processing will improve significantly. SQL generation will improve via large language models.	AI will optimize where and how data is stored, moved, and transformed. Building on open formats, automated composition of the data stack, and query components will enable custom-fit solutions. Multi-modal handling techniques will unlock value in more unstructured data. Multi cloud technologies will enable transparent, governed access to unstructured and structured data regardless of location.	Natural language interactions and intelligent visualizations will be scaled to large dynamic data sets. Almost all analytic tasks will be facilitated by new foundation model customization capabilities to scale in the field to the task at hand. 360º view of governed data and replicas will support scaling an AI-driven data consumption landscape. Data exchanges will go big.	Compute Express Link (CXL), with its ability to disaggregate memory from processor, will be used to deliver faster memory and compute. New types of accelerators like data processing cores will be commonly used. Cheap, fast networking will bring the next wave of data into most enterprises. Competitive pressure will sink cloud egress fees.	A universal, location-independent method to find and identify data will be realized. We will produce and use semantically rich metadata that is interoperable and federated across any set of organizations and repositories. Security mechanisms will be developed to ensure permissioned data access in this widely distributed, decentralized setting.
How these advancements will be delivered to IBM clients and partners	✓ Watsonx.data, db2, and NZ will support integration with open table formats. ✓ Data marketplaces will bring data products to consumers and have a unified, virtualized online analytic processing “front door” with graph capabilities. ✓ Watsonx.ai will power semantic automation. Open table formats enter the enterprise mainstream.	↻ Watsonx.governance will power insights on data and AI attributes like fairness and drift. Watsonx.ai will be fueled by data prepared and managed on watsonx.data, and watsonx.data will be infused with watsonx.ai models. Lineage will be critical and will be tracked via Manta/IKC. Open technical and business metadata will thread through watsonx and IKC.	z/OS data will be available through cloud services and secured with fully homomorphic encryption, multi-party computation, etc. Compliance with data governance will be offered using code-driven processes instead of post-fact forms. Twinning/mirroring across IBM data stores will improve the access to data. Global caching will reduce replication penalties.	Democratized access to data products will be provided with a 360º view of governance that includes infrastructure, data, and AI. Watsonx.data cleanroom for sharing will be available.	Data stores and processing engines, scaled with CXL, will be offered for major cost and performance wins. There will be more focus on secure, software-defined networking.	Computing platforms will be furnished with mechanisms for distributed data identity and full permissioned network connectivity.