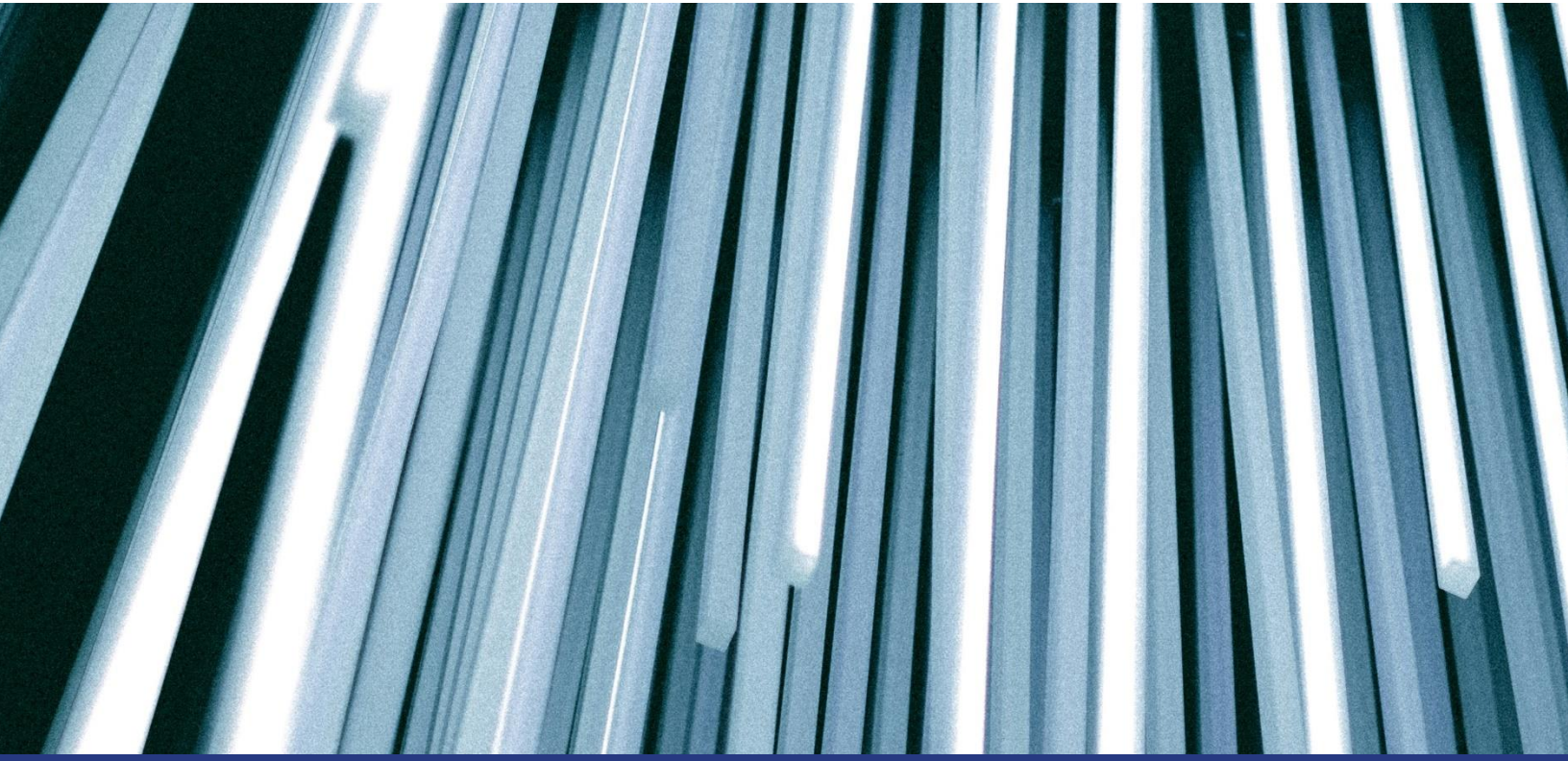


Data Lake Strategy & Design

Overview and Early Adoption Playbook

IMS Enterprise & Technology Architecture

August 2025

Executive Summary

Harnessing the immense value that AI and analytics have unlocked over the past few years is pushing organizations to move towards flexible & resilient data architectures that can scale to meet the requirements of tomorrow's data hungry platforms. Siloed data locked away in legacy systems can hamper even the best organizational efforts to transform and adopt modern technologies.

As we enter a world of increasing complexity and uncertainty, technology and IT executives will face immense pressure over the coming years to deliver solutions to support functions and operations across the business as they navigate market challenges. Whether its optimizing supply chains and managing risks through analytics or enabling the finance team's forecasting with AI/ML models – data remains critical to delivering the needs of business tomorrow.

Enter the data lake – a centralized store of unprocessed data which pulls from systems from across the enterprise to create an easy and scalable way to break down data barriers across your organization. The flexibility of a data lake architecture to support different systems, interfaces, file and data types as well as scale as needed makes it perfectly suited to deliver the use cases of the future – reporting, analytics, machine learning, etc.

No matter your environment and host strategy, the advances in the data and analytics space have opened up new opportunities to take advantage of modern solutions. For companies in the cloud – data storage costs have fallen around 80% and compute costs have halved over the past decade. For those on-prem the improvements and proliferation of new tools and approaches has made a self-hosted data lake more accessible than ever.

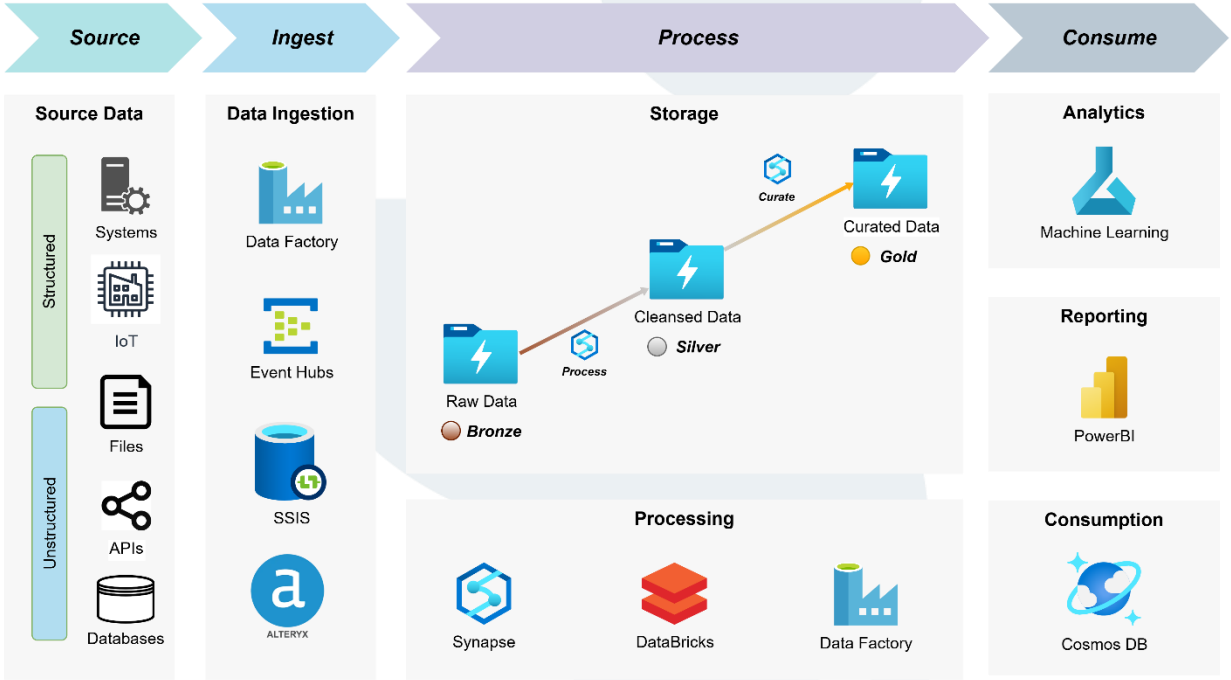
How We Got Here

While the data lake is a relatively new concept which has evolved over the past decade, the fundamentals of what sets a data lake apart from other data solutions remain the same: the ability to handle large amounts of unprocessed data in various formats without rigid pre-defined schemas.

A data lake's defining feature is the ability to ingest and store in its 'raw' state directly from your source systems and not be constrained to a pre-defined rigid schema structure.

Data Lakes

In contrast to data warehouses and other relational data stores, data lakes don't require a pre-defined schema to ingest and load data from source systems. This affords the data lake a level of flexibility to retain data in its native formats and shifts towards a dynamic schema assignment as needed per the use case. The gradual move away from ETL towards adoption of ELT has opened new opportunities to process and explore data faster, cheaper and more effectively than ever.



High Level Basic Data Lake Architecture Based on Azure Services

Medallion Architecture

One of the best practices for setting up a data lake environment is employing a medallion architecture to ensure you progressively increase the quality of (and trust in) your data as it moves from ingestion through the consumption.

By initially landing everything in the 'Bronze' layer, you preserve history and enable fast onboarding; Silver standardizes, de-duplicates, and enforces business logic; Gold delivers analytics-ready models and metrics for BI, AI, and APIs. This separation improves governance (lineage, access control), accelerates change (iterate upstream without breaking reports), and reduces cost by keeping heavy transforms where they add value.

Bronze	**Unprocessed / raw data** Loaded from source systems. Various file formats supported and usually retained in blob storage
Silver	**Cleaned / structured data** Processed from Bronze layer Assigned a schema and put through data quality tests & standardization
Gold	**Curated / Consumption Ready** Processed from silver and assign dimensions and factors Ready for business consumption and reporting

Example Medalion Architecture Setup

This approach is baked directly into the data lake storage design and allows you to leverage data within the same environment to cut down on data ingress and egress costs and enable data consumption more easily.

Data Lake as an Enabler

Unlocking value from your data will become key as we enter an even more data intensive environment. The increasing number of interconnected systems and vendors providing services at companies will only increase – increasing the complexity of the landscape as well. A data lake can help solve some of these pressing issues to unify your data strategy and get data in the hands of those who need it most.

Some of the areas where we see the biggest value for adopters include:

1. Unlock the ability to create comprehensive and complex AI/ML algorithms
2. Create consumable data products compiled from various sources
3. Serve other downstream consuming systems with prepared data products
4. Simplify reporting and generate cross functional insights faster
5. Create the single source of truth from your systems of record

Common Challenges

Opportunity and new initiatives which span across functional and operational lines often bring challenges impacting business adoption and operational sustainability.

Some of the below pitfalls and challenges can delay or even derail your vision from taking shape – navigating these challenges can ensure you set yourself and your teams up for success.

	PITFALL	IMPACT	SOLUTION
Not Treating Data Strategically	Treating the data lake as a data 'dumping ground' Undefined vision for governance, integration, or use cases	Finding accurate, usable data is slow and costly Frustrated business users and a loss in confidence in the solution	Taking a structured iterative approach based on concrete use cases combined Adopting a 'data-first' mindset across the
Over-Complicating From The Start	Targeting too many use cases & initiatives with long lead times Building without ensuring compatibility with existing tools, PLM, ERP, or MES,	Limited adoption; teams continue using legacy data silos because the lake isn't part of daily workflows	Focusing on a subset of use cases with clearly defined success metrics for initial launch Deliver up front value and confidence in solution
Not Designing For Scale	Overinvesting into proof of concepts without path to move to scalable production Overreliance on existing tools & technologies	Performance bottlenecks & compute inefficiencies Consistently high compute consumption Slow response to new business use cases	Design core solution around scalable services Commit to a target platform or leverage platform-agnostic tools
Ignoring Process & Governance	No data source, pipeline & rules governance Ingesting incomplete. dirty, or duplicate data without standardization or cleansing processes	Unreliable data; business & production decisions are delayed or misinformed Inconsistent metadata, no lineage tracking, and weak access controls.	Start small and apply basic data & build governance to initial use cases Align on organizational best practices and guardrails up front

Starting the Journey

Looking for a way to explore the value that a data lake can bring to the table? Here are some of the first steps in kicking off your big data journey.

Our approach: start small, pick a use case and pain point for the business that you can solve through data.

Demonstrating value up front to the business with shorter lead times is key to getting stakeholder buy-in while building the internal team capabilities to support the tools and technologies in the long run are key to a sustainable solution.

Tactical Actions to Kickstart your Data Lake Journey

Identify High Value Use Cases

Initial buy-in by your stakeholders is critical to maintaining momentum with any big data initiative. Identifying a narrow, high value use case which solves a perpetual pain point for the business can help win over skeptical and resistant partners while demonstrating the opportunities that a data lake can bring to the enterprise.

Select and solve for a use case that has a visible impact (saves money, improves uptime, speeds reporting) to build builds stakeholder confidence and keep the business engaged.

Ideally this should be a use case your teams and partners can solve with minimal upskilling or upfront investment into learning. Keeping things simple

Setup a Small & Scalable Data Environment

Setting up your initial environment which can support your selected use cases without requiring a huge upfront capacity investment by your teams is key. Selecting a target platform or defining an architecture which can be ported over to another platform or hosting solution can ensure that you avoid future architectural debt and rework.

This early stage is where the YAGNI principle (‘you aren't gonna need it’) should hold – don’t boil the ocean and solve for every requirement. New features, functionalities and

governance can be included in your backlog as your data lake gains critical mass and new requests from the business come in.

Ingest an Appropriate Data Sub-set

Start with the sources and data required to enable your use case end-to-end. Depending on data sources and size of the data, decide on whether to bring in the full source system dataset or just a subset.

Keep pipelines simple for the time being with a focus on the use case, showing the value that centralized and unified data can bring to the table.

Build Basic Analytics and Reporting

Once you've setup your environment and ingested your dataset to support your use case – think about how the end user will interact with the data product or output. This is a critical step in getting the user to understand the value of the solution you and your team have delivered and demonstrate the art of the possible.

Use simple queries, BI tools, or AI/ML notebooks to demonstrate how the lake adds value for your business users and solve the use case at hand. Integrating your data outputs and products into other consuming systems or data stores is an additional boost and value add.

Here is where you and your team can shine and show the art of the possible first hand.

Evaluate & Iterate

Completing the Proof of Concept is not the end, but just the start of the data lake journey. Being to plan the next phase of your data roadmap: adding governance, cataloging, and growing your data sources.

The path forward can be broken into high level growth areas:

- Productionalizing infrastructure and processes to support scaling
- Continuing to identify new use cases and grow the pool of data sources

Formalizing data and operational governance processes, making pipelines more resilient and implementing robust security controls are some of the areas to handle early on before scaling too quickly.

Using the experience and feedback from your PoC is a good way to implement a future solution tailored to your organization – ensuing adoption.

Conclusion

As technology progresses and the size and scope of data continues to increase, there has never been a more relevant time to take advantage of the flexibility and depth of information that a data lake can provide. Whether standalone or used as the basis for a lake-house or feeding into an existing data warehouse, a well designed and sustainable data lake can bring business insight and operational value that can take your business to the next level. Bringing together data from across your system landscape can help reduce those system siloes and barriers to fuel analytics, AI, and reporting initiatives that drive real business impact.

Looking for additional insight? Reach out to IMS and we can show you why our clients trust to help define their vision and to help deliver on their mission critical projects.

About IMS

Intra Management Solutions is a premier consulting, program execution management and technology implementation services provider.

IMS' Enterprise & Technology Architecture practice brings together our functional, system implementation and technology experience to help companies in their digital transformation efforts.

Our experience ranges from design through to implementation across different platforms and technology stacks. IMS' pool of functional SMEs also serves as an expert body to ensure what works on paper will work in the real world for end users.

For over 15 years we've focus on delivering value through sustainable solutions which scale with the needs and ambitions of our clients.

For more information visit us at intramanagement.us

Intra Management Solutions, Inc.

6136 Frisco Square Blvd, Suite 400

Frisco, TX 75034

Tel: (214) 444-6544

Fax: (214) 785-2951

Web: intramanagement.us

Email: info@intramgmt.com