

WHITE PAPER

Simplifying complex data integration

An overview of the MarkLogic Data Hub Platform and how global organizations are using it to solve the complex data silo problem.

Why data integration?

Whether it is a unified view of customers, patients, financial trades, manufacturing parts, or anything else—every business benefits from integrating their data and achieving a unified 360° view. What organizations are now realizing is that their 360° view must also be actionable, real-time, and governed.

The benefits are numerous. Here are a few examples:

Better serve customers – We've all had that experience of being bumped around to different people when you call the service hotline. A 360° view of customer data prevents that experience from happening.

Improve operations – A 360° view of assets with supply chain lineage tracking provides a unified view of operations. Not only can you optimize processes and keep equipment humming, but you can also keep everyone safe.

Answer regulators – A 360° view helps manage regulated data and address increasing regulatory scrutiny. Rather than missing deadlines and getting fined, you can share the exact data a regulator needs at any time and can trace back how that data changed and where it came from.

Every organization can benefit from having a 360° view.

But, why is it so hard to move from a world of silos to one that is integrated?

The complexity of data silos

Despite many attempts at solving it in recent years, the data silo problem has become worse, not better.

Existing data integration products and strategies are not working. Today, IT departments spend most of their time just keeping their heads above water, managing a complex web of data silos and frequent ETL processes to shuffle data around. In recent years, new sources of data like IoT feeds, message feeds, AI/ML tools, and other new systems have increased the number of silos. It's clear now that Hadoop was never the right answer.

The problem of a bad architecture impacts everyone. The business cannot understand why it's taking IT so long – they needed their app delivered last year. Meanwhile, developers are frustrated trying to explain why the ETL alone will take a year or more, and are also upset because they are unable to focus on building apps. Meanwhile, the architects are stuck in the middle because they know they need agility, but they still need safety, security, and resiliency.

What if you had a simplified world in which you removed the friction of data integration and the problems it causes from all aspects of your business, architecture, and development?

“ 60% of the cost of data warehouse projects is on ETL.”

– Report sponsored by Informatica

A simplified world

We designed the MarkLogic Data Hub Platform to simplify data integration. With the MarkLogic Data Hub Platform, you ingest your data as *is*, curate it, apply security and governance, and make it accessible.

And, the platform is flexible so you can avoid having to model everything at once, and you don't have to change the model or manually do ETL every time the data or business needs change. The platform is designed to give the business, architect, and developer what they all want.



Fast pipelines – Instead of waiting on complex ETL, data ingestion is immediate. Ingest structured and unstructured data *as is* with a flexible data model that adapts to both changing data and changing data structure. MarkLogic natively stores JSON, XML, text, geospatial, and semantic triples.



Unified platform – Instead of stitching together a bunch of separate products, everything is already unified in an integrated and single platform that provides a consistent, real-time view of data. With AWS for example, you'll need a dozen different products (various databases plus ETL, security, monitoring products), each with different compute, storage, management, and billing needs. MarkLogic gives you a unified platform that's not only simpler, but significantly less expensive.



Smart curation – Instead of worrying about mapping schemas together, integrating an MDM tool, writing custom algorithms, and other non-value adding tasks, you can leverage built-in, smart and automated capabilities to enrich, harmonize, and master data easier and faster. With Embedded Machine Learning, MarkLogic is getting even smarter to help you with your machine learning and artificial intelligence initiatives.

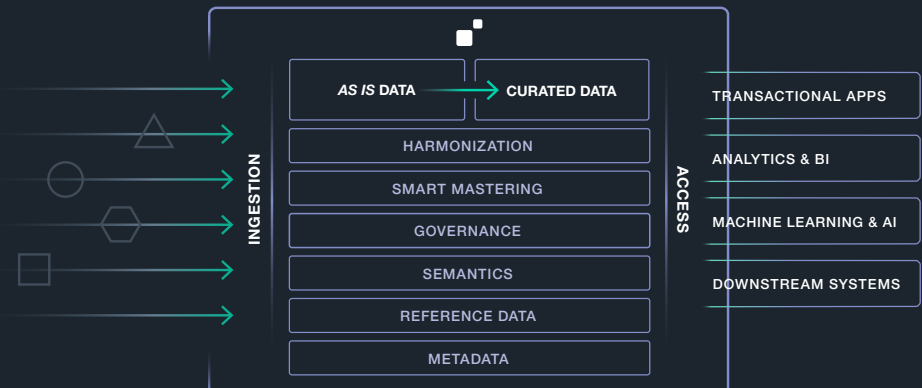


Advanced security – Instead of disjointed data lineage, the database tracks that metadata right alongside the data itself. Instead of handing over security to app developers to worry about, the database manages the roles, permissions, privileges, etc. And, most important, instead of worrying whether to lock data up or risk sharing it, you have extremely granular, tight control over exactly what data gets shared with whom.



Simple development – Instead of waiting for ETL to complete or having to learn some proprietary language, developers start building data services APIs as soon as the initial data is loaded. Instead of having to learn a proprietary language, developers can rely on industry standard languages and open APIs to build apps.

MarkLogic Data Hub Platform



- ✓ **Ingest and discover** – There is no need to define a schema in advance. We're not throwing structure and transformation out the window—we're just saying you don't have to do it all in advance. Index and search through your data immediately.
- ✓ **Curate** – By skipping up-front data modeling before ingesting, you can focus on harmonizing and enriching your data for your first use case. Then iteratively harmonize more data over time for future use cases.
- ✓ **Access** – Easily access data for your downstream systems and applications with full transactional consistency and enterprise-grade security. Our platform is DevOps focused with standardized APIs and other tools to enable fast application development and deployment to any environment.

The best database for a data hub

Why is the MarkLogic Data Hub superior? It's powered by the MarkLogic multi-model database, a modern and trusted database that is proven in mission-critical environments around the world.

A modern database

Multi-model database – You can store and query documents, graph data, or relational data from a single database which provides incredible flexibility.

- Documents – The document model is incredibly flexible, which is critical for data integration so you do not have to worry about schema complexity.
- Semantic graph – You can manage the relationships and context surrounding the entities you are integrating.
- Relational – MarkLogic gives you the ability to create relational views on top of documents, so you can still query with standard SQL. MarkLogic is the only multi-model database with this capability.

Built-in search – MarkLogic has a search engine built-in to its core. This results in less time and effort to build and configure indexes for standard queries, and does not require a bolt-on search engine for full-text search like other databases.

This is incredibly helpful for data integration because it saves time during the curation process, provides users immediate access to their ingested data, and enables users to ask more complex questions of integrated data.

Elastic scalability – We have customers storing upwards of 2.5 Petabytes of data in MarkLogic. When integrating massive volumes of data, you want a database that scales quickly, easily, and at low cost. MarkLogic is a massively scalable database that scales horizontally in clusters on commodity hardware. And, when demand dissipates, MarkLogic can scale back down without having to worry about complex sharding.

A trusted database

Secure and governed – MarkLogic is the most secure modern database, with more security certifications and granularity than any competitor. Security – and governance more broadly – is critical in order to safely integrate data, track it through that process, and safely share its curated form.

High availability and ACID transactions – MarkLogic is built for the enterprise and is a hardened platform with uncompromising data resiliency. This is not true of any other modern database, which promise high availability but lack full scale ACID transactions. In the event a node fails, how can they guarantee consistency after recovery? To be resilient, you need enterprise-grade HA/DR and ACID Transactions. And, ACID Transactions must be proven at scale in the enterprise, not a marketing claim.

**“ We’re putting
in an event-
based real-time
infrastructure
and building a
foundation for
the future. The
heart of that is
MarkLogic.”**

- SVP, Director of Architecture
and Innovation
Northern Trust

Proven results

ERP integration – Unified view across all ERP systems. Using MarkLogic, [Eaton](#) integrated data 5x faster than they could with Oracle to get a 360° view of their 200+ ERP systems.

Regulatory hub – A data platform to safely share unified data, whether it's for safeguarding government secrets or keeping regulators up to date. [ABN Amro](#) used MarkLogic to quickly assemble regulatory data reports on GDPR and MiFID II, and also handle all future regulatory requirements (whatever they may be).

Advanced analytics – To get meaningful algorithmic analysis and training data for AI, you have to have trusted data assets. [AIRBUS](#) built a search application that would enable them to find a historical test that matched the exact parameters.

Research and development – Centralized knowledge sharing hubs to enable faster business results. With MarkLogic, [AbbVie](#) was able to consolidate five data sources, an ontology database, and a search engine.

Unified 360° view – A data platform for smarter, more agile MDM (Master Data Management). Using MarkLogic, [Aetna](#) successfully managed HR data for around 35,000 of their own employees.

Content monetization – A content platform to get value fast. In four months, [NBC](#) built an award winning, cloud-based app on MarkLogic to help fans find comedy gems from the past 40+ years.

MarkLogic Data Hub Service

Interested in a fast, serverless deployment?

MarkLogic Data Hub Service is a fully automated cloud service you can use to deploy the MarkLogic Data Hub in under 15 minutes. Based on MarkLogic's Data Hub, the service enables agile teams to immediately start integrating and curating data for both operational and analytical use.

- **Fast deployment** – There's no lag time to get infrastructure up and running. It removes skills, hardware, config, and ops barriers that slow implementation.
- **Simplicity and reliability** – You get a full-stack Data Hub and can avoid the complexity of managing and paying for a dozen different components. And, it's enterprise-grade with 100% ACID transactions and certified security.
- **Cost-effective** – You only pay for what you need. Spikes in demand don't cause spikes in cost. MarkLogic's elastic cloud credit pricing works like rollover minutes on your mobile phone plan, keeping costs consistent and predictable.

SONY

CASE STUDY

Mainframe migration for “sales-to-finance” Data Hub

Sony Pictures Television was using a legacy mainframe system that was becoming increasingly complex and difficult to maintain. They no longer had anyone in the organization who understood how every dollar goes from the sales to finance system and were facing issues with the rigid nature of relational databases. The existing system couldn't keep up with the growing business, was prone to error, and computationally expensive.

It was becoming increasingly difficult to forecast sales and much of the validation to finance processes was manual. Sony needed to modernize.

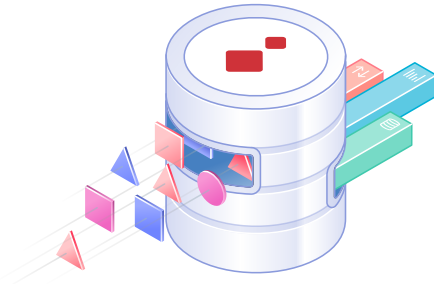
With MarkLogic, sales data is made available to users in just three months. Now faceted searches are completed in seconds vs the previous 40+ minutes. Sony attained faster sales, increased revenue, and lowered costs.

“ We needed a flexible way to just throw this in – whatever data we found – because some of it would be spreadsheets, some of it was core from the mainframe and all that. And so we needed ways to be flexible with this, and MarkLogic was our choice.”

– VP, Worldwide IT, Sony

Data integration. Simplified.

Data integration is one of the most complex IT challenges and the MarkLogic Data Hub Platform simplifies it. MarkLogic removes friction at every step in the process so that organizations can achieve a 360° view faster than ever. Complex and slow ETL? Eliminated. Simple and fast data curation? Implemented. Modern database capabilities? Utilized. Total data integration success? Achieved. With MarkLogic, it only takes days or weeks – not months or years – to integrate data and build secure data services that enable IT to keep pace with the speed of business.



Increased agility

4-12x faster results with less risk when things change



Lower costs

Millions saved with lower IT, development, and operational business costs



Secure

Sharing Petabytes of data around the world every day

More Information

What is MarkLogic?

Read more about MarkLogic's unique set of features at marklogic.com/what-is-marklogic

Get a deeper dive in our eBook for architects

marklogic.com/resources/introducing-the-operational-data-hub/

Start building with MarkLogic Data Hub for free

marklogic.com/product/marklogic-database-overview/database-features/data-hub-framework/

Schedule a Meeting

Discuss your particular use case with a MarkLogic sales representative by contacting us at sales@marklogic.com

About MarkLogic

By simplifying data integration, MarkLogic helps organizations gain agility, lower IT costs, and safely share their data. Headquartered in Silicon Valley, MarkLogic has offices throughout the U.S., Europe, Asia, and Australia.

999 Skyway Road, Suite 200
San Carlos, CA 94070

+1 650 655 2300

+1 877 992 8885

www.marklogic.com

sales@marklogic.com

© 2019 MarkLogic Corporation.

MarkLogic and the MarkLogic logo are trademarks or registered trademarks of MarkLogic Corporation in the United States and other countries. All other trademarks are the property of their respective owners.

 MarkLogic®