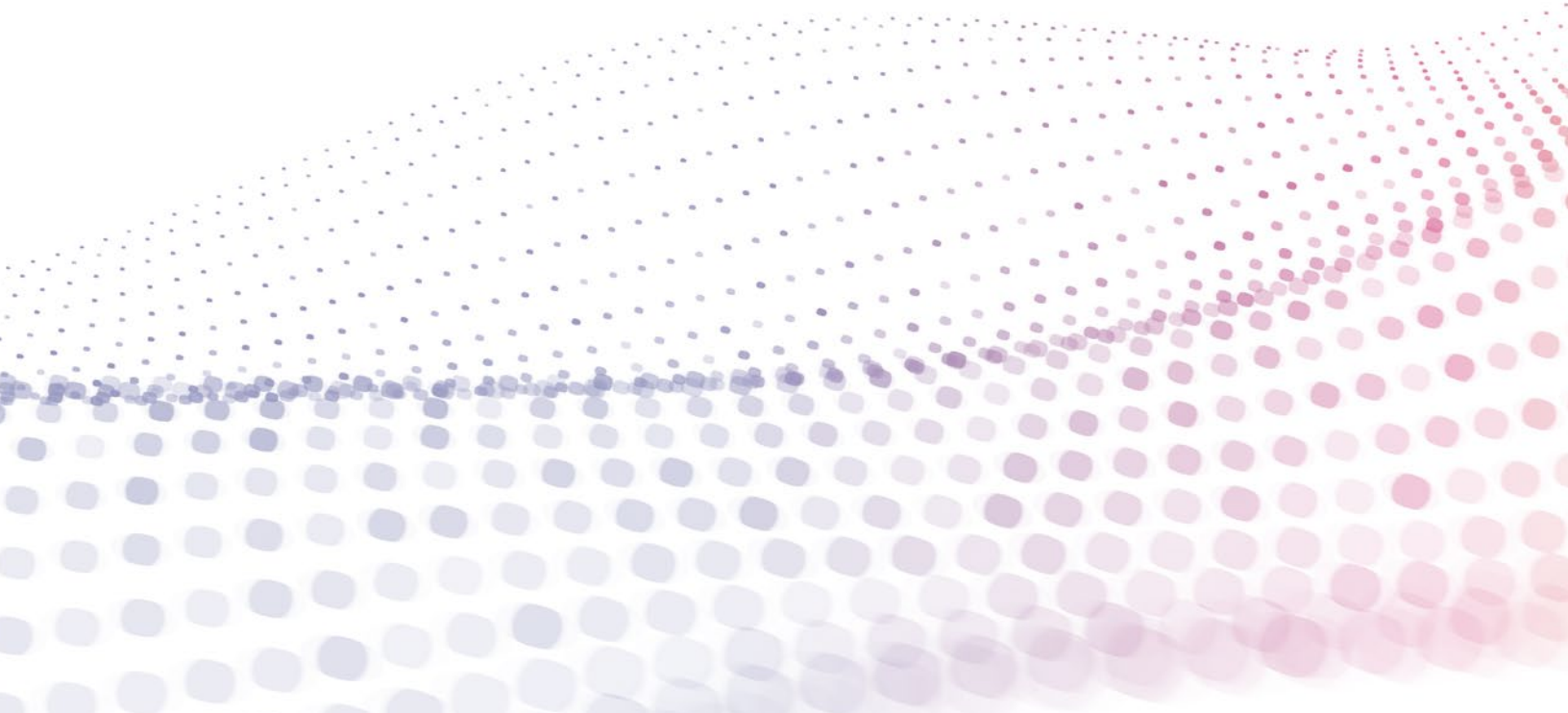
MarkLogic

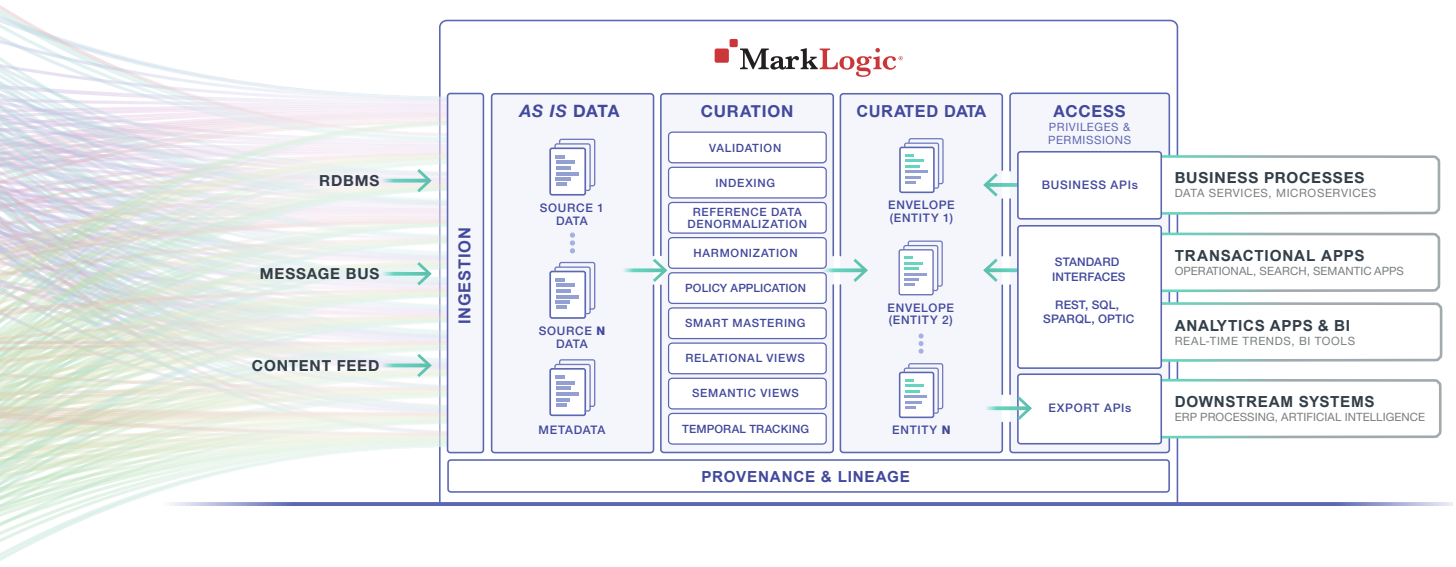
MarkLogic Operational Data Hub

MarkLogic Technology and MarkLogic Expertise to Integrate Your Data, Faster, with High ROI

QUICK START GUIDE · JUNE 2018

If you spend more time fighting your data than benefiting from it, you aren't alone. Enterprise data integration eats up 60 percent of a typical IT department's budget — and it still leaves a gap between the operational and analytical needs of your business. MarkLogic can quickly and easily help stand up an efficient Data Hub alternative to integrate and operationalize data.





What's an Operational Data Hub?

An Operational Data Hub (ODH) is a single integrated system that stores, indexes and serves all your data, so it's available as enterprise data services. MarkLogic is the best platform for an ODH.

The image above shows how an ODH works:

- Data is loaded *as-is* and made immediately available. MarkLogic is "multi-model," so handles almost all data formats natively, without modeling or transformation, and makes them immediately available at a base level.
- Data is then *Curated*, to be more highly structured, and leverage some (minimal) canonical models for key fields – but still without requiring a slow, expensive modeling process for all the data.
- Finally, MarkLogic projects, masters, cleans and exposes the data as REST API endpoints, SQL views and bulk data exports for use throughout the enterprise.

Throughout the Enterprise?

This last point about enterprise use is critical. A Data Hub is an enabling enterprise technology. Most data integrations are for narrow business purposes such as an analytic data warehouse, or a data science repository. An ODH is different in that it is intended to help you "run your business" not just report on it:

> *An Operational Data Hub is an enterprise technology that allows real-time, universal, scalable, secure and governed data access for use throughout the enterprise.*

Flexible Like a Data Lake, Operational Like a Transactional Database, Governed Like a Hub

Another approach – Data Lakes – also hold disparate data from many sources, but are not governed, not indexed, and not real-time. In contrast, an Operational Data Hub (ODH) is all these things.

Data is governed: Security and temporal tracking policy ensures your data is tracked, and secure. **Data is indexed** to allow complex queries spanning structured data, unstructured text, geospatial data, graph and semantic data, and binary metadata. Lastly, and ODH is **real-time**: they can simultaneously ingest and serve thousands of API requests per second, and scale out on cloud platforms or your own servers.

Customer ODH Example

So what might that look like? An ODH for a retail business holding Customers with related Address and Purchase data could serve all of:

- The customer portal where people check on their purchase history
- The Shopping Cart system where new orders are added
- The fulfillment system that updates order status and schedules shipments
- A bulk-mailing system that needs current addresses to print envelopes
- A predictive analytics system to identify potential churn, or target marketing campaigns

These are **Enterprise needs**. These varied data consumers need **operational**, **unified**, **high-quality**, **real-time** data.

And it's secure and governed. MarkLogic does all this with industry-leading **data governance** by tracking, securing and auditing data, rather than becoming a new collection of unreliable, unsecure data with questionable provenance.

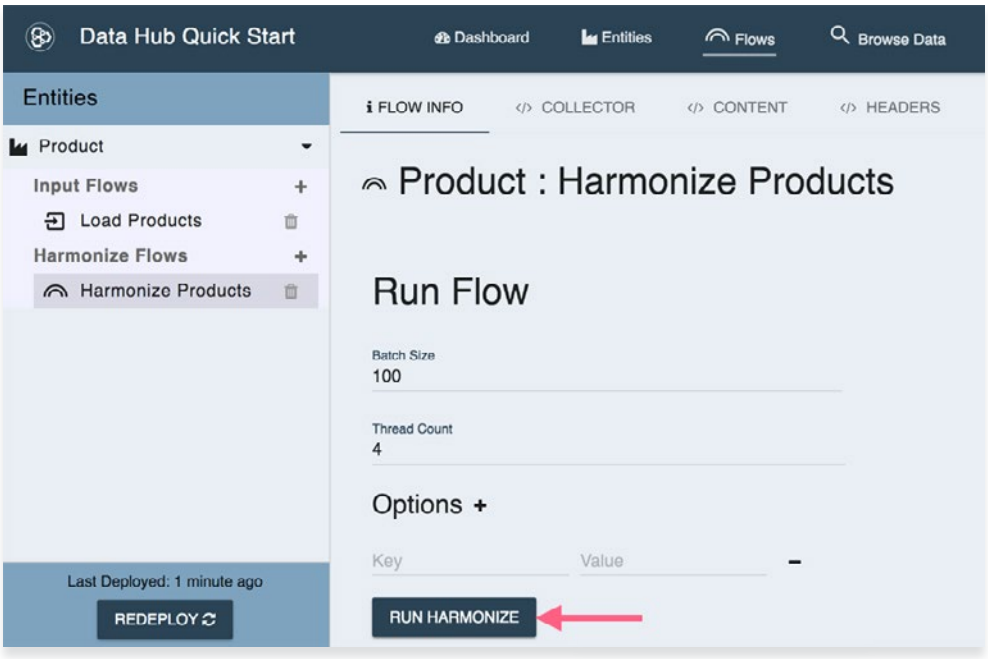


Figure 3: This screen capture from the Data Hub Framework illustrates the organization of flows for various entities on the left, and the operations, job tracking and transform step management on the right.

The Data Hub Framework – a MarkLogic Toolkit for an ODH

MarkLogic's Quick Start engagement uses our Data Hub Framework to quickly deliver an initial data hub. Our core product is our database management system (MarkLogic Server). But we have also distilled our best practices into the Data Hub Framework which runs on top of MarkLogic.

The Data Hub Framework organizes data into business "entities" and groups processes into "data flows" to keep your team organized and productive.

The Data Hub Framework also supports batch operations, exposes real-time ingest and output APIs, and automates common DevOps processes such as creating databases and deploying code.

45 Days to a Working Hub

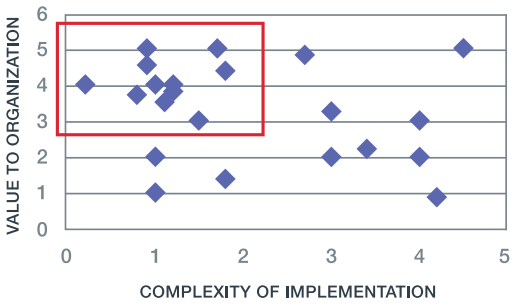
MarkLogic Consulting has been building data hubs with and for our customers for over ten years. While digital transformations are large and disruptive to most organizations, they are commonplace to us. This is our day job.

How to Get Your Data Integration Project Started

MarkLogic Consulting can accelerate your team's data integration project with a 45 day quick start engagement. Our Quick Start engagement has four steps:

1. Identify High-value Data and APIs

In our Quick Start engagement, we help you identify high-value data sources to ingest, and high-value services to produce to satisfy a real need in your enterprise. This high-value, low-cost set of services is illustrated by the red box in this diagram.



By focusing on a few "quick wins" we find that the ODH approach can be validated early, giving fast feedback to stakeholders. By choosing the most important APIs, stakeholders actually want or need the new APIs or data exports (quickly) and typically drive urgency to keep things on track and on schedule

After a Quick Start proves the technology and educates the team, other APIs, exports and data sources are then brought into the Data Hub in the next iterations.

2. Scope and Plan the Initial Data Services

Step 1 produces important APIs, and we then work backwards from those APIs to identify which data sources have the required inputs, and what data cleaning, curating, transforming and mastering processes are needed to turn that data (the raw material) into a working API (the finished product).

3. Divide and Conquer

The Data Hub Framework toolkit we use decomposes the most difficult tasks into Entities and Data Flows. The data inputs and API outputs are now decomposed into specific tasks organized by Entity and Flow. Each entity will progress through well-defined steps to be sufficiently curated to serve the required APIs.

Notice that "extensive data modeling" is NOT on this list. Minimal, agile, "just-in-time" modeling is performed to get the APIs working. An agile process can later refine and extend these models, but most data can be stored, queried and transformed "as-is" depending on API needs. Pull in **all your data** "as-is" first, and evolve later – that Agile Modeling saves time and reduces risk.

4. Implement Each Entity Flow as a Series of Plugins

After those first three steps, you know the key requirements, and have a breakdown of the work into Data Flows. Next, these data flows are planned, scheduled and divided up among the implementation team. Building them out constitutes the remainder of the Quick Start engagement.

MarkLogic Consulting typically leads this initial 45-day effort, with your staff involved and learning to do the work "on the job" so they are ready to take over for future deliveries.

Why MarkLogic Consulting

The MarkLogic Consulting Services team are experts who focus on data integration – all day, every day. We have insight into common challenges and their solutions. From change data capture to DevOps, team structure, hardware selection, cloud migration and performance tuning – we have relevant experience and methodologies needed to smooth the path forward and make your effort successful.

In addition to standing up a system quickly, we aim to build required skills in your own team. Even the best team of data and systems professionals may need a little while to get their heads around new, agile, MarkLogic approaches – where you can ingest data without modeling it first, drive agility down to the very core of the data integration effort, switch from data-first (and modeling focused) integration to API based development.

There simply isn't time for a completely new group to think all this through, adapt to a new paradigm, AND develop an impressive system in 45 days.

MarkLogic Consulting have individually and collectively gathered skills and best practices from earlier projects, and look forward to supporting you on your initiative.

Contact Us

Contact us at consulting@marklogic.com to learn more and request an Operational Data Hub Quick Start engagement.

MARKLOGIC CORPORATION
999 Skyway Road, Suite 200 San Carlos, CA 94070
+1 650 655 2300 | +1 877 992 8885 | www.marklogic.com | sales@marklogic.com

999 Skyway Road, Suite 200 San Carlos, CA 94070
+1 650 655 2300 | +1 877 992 8885
www.marklogic.com | sales@marklogic.com