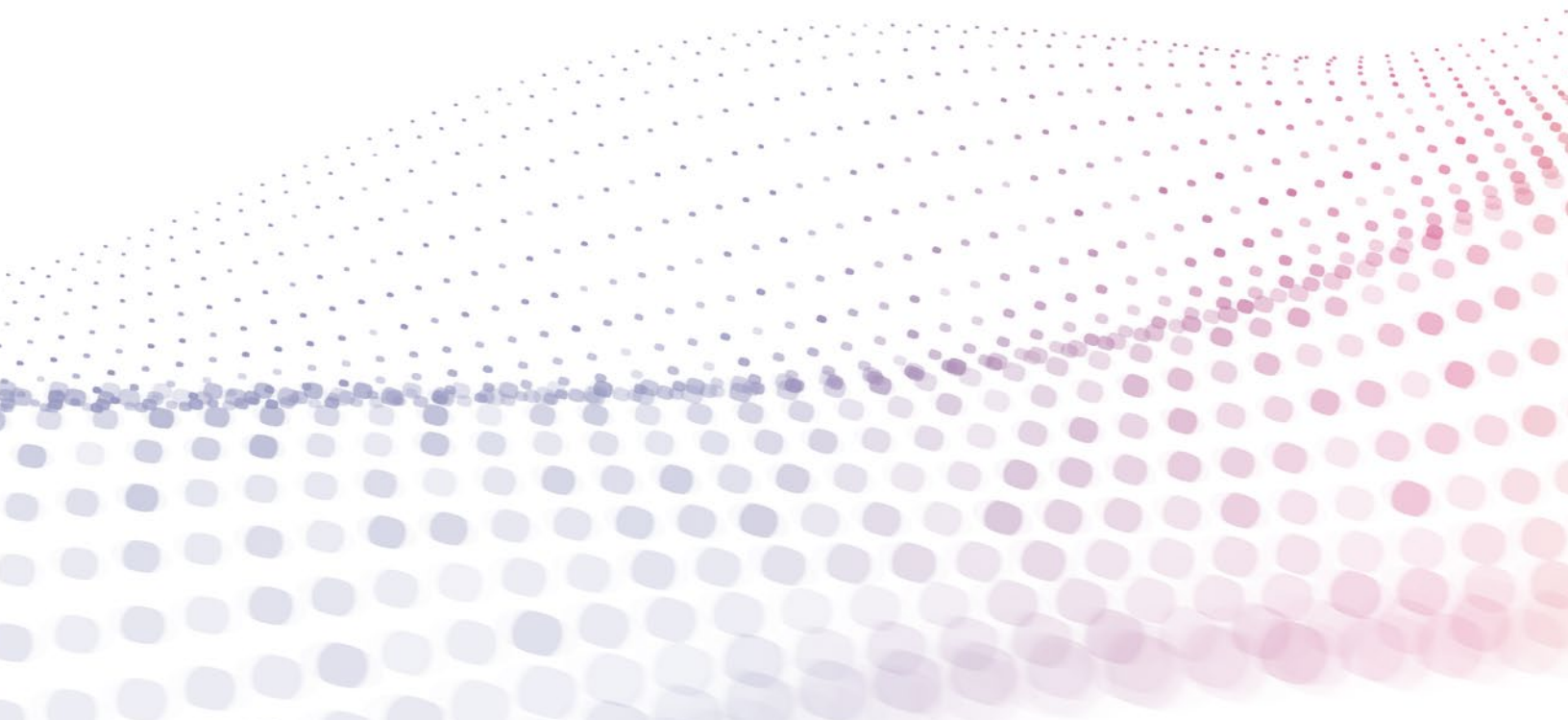


Resourcing Your MarkLogic Implementation

MARKLOGIC WHITE PAPER · JUNE 2019

Implementing a MarkLogic project is a fairly straightforward process, but there are some important differences from the legacy relational approach. In this guide we will detail how to structure a MarkLogic project, how to identify the right roles, and how to ensure that your existing team(s) have all that they need to succeed.



Contents

Executive Summary	1
Project Approaches	1
Relational Projects	
MarkLogic Projects	
The MarkLogic Project Team.....	5
Form Your Approach	
Recommended Deployment Models	
Team Roles	6
Resources to Support Your Implementation.....	8
MarkLogic University	
Mentorship	
Center of Excellence	
Get in Touch	12

Executive Summary

Expertise in legacy technology does not ensure successful delivery of a modern technology project. It doesn't matter how good your team is if the technology is hindering your chances of success. Relational databases were built for a different era when organizations could weather multi-year, big-bang IT projects. Back when the rate of change in requirements could be minimized.

The inherent inflexibility of the relational database stands in the way of successful adoption of modern software development methodologies. Trying to deliver to the higher expectations of customers using decades-old tools is a losing proposition. Many years of trying to make these technologies fit into the new paradigm, where the only constant is change, have largely failed.

But adopting new technology can also be an inherently risky move. If you have to bring in a whole new team, a group of untested and unproven resources, you have simply replaced one risk profile for another.

Ideally, the next-generation technology that you adopt should greatly reduce both sources of risk to your projects. MarkLogic was built for modern software delivery practices. And the good skills and foundations that your existing teams have acquired through their enterprise implementations are fully transferable.

In this guide we will detail how to structure a MarkLogic project, how to identify the right roles, and how to ensure that your existing team(s) have all that they need to succeed.

Project Approaches

How many projects have you run where the schema design and ETL migration steps have consumed too much of your allocated time, budget, and the patience of your customers? These are steps that provide no business value, but which cannot be avoided when working with relational technologies. With a MarkLogic project these steps are greatly reduced or skipped altogether. You and your team can jump straight into delivering high-business value features and functionality. In the current world with the pressures of Agile delivery this decreased time-to-value is an essential differentiator of MarkLogic Projects.

Relational Projects

Projects using relational technologies must be done in rigid order, starting with schema design. Getting the schema design right requires understanding the state that the data as is in. Obtaining a comprehensive view of the current state legacy systems is often difficult or impossible. And if the project involves migrating data from more than one source system, designing a single canonical model can be time-consuming, contentious, and often results in data being abandoned or modified in unwanted compromises.

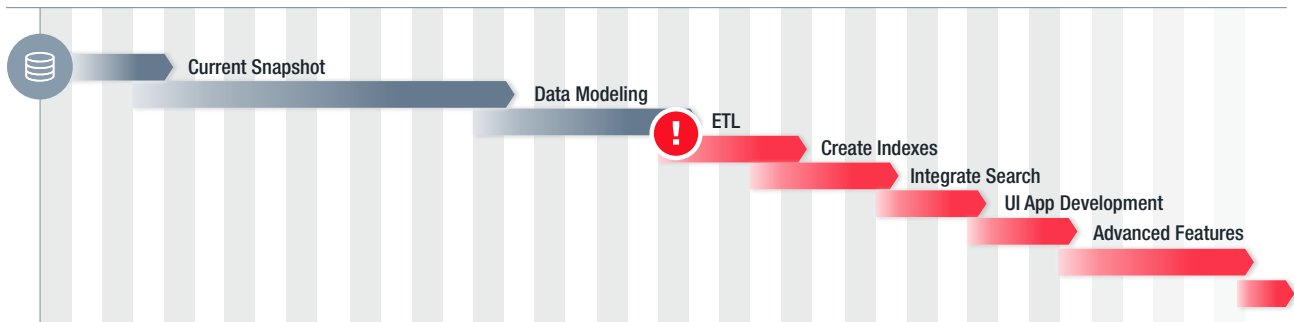


Figure 1: Traditional Software Projects

Though it is understood that these steps cannot be skipped, teams face pressure to start development work as soon as possible after a project officially commences. The most common approach is to create synthetic test data for the developers to write their code against. But synthetic test data never manages to cover all of the strange cases and quality issues to be found in real data. And once the new application and data are married up, problems often arise, and rework is required.

The disconnect between development and data migration is exacerbated when employing Agile delivery methodologies. Agile projects are most successful when all tasks can be broken down to the right size for teams to deliver in sprints. Because of the limitations of relational technologies, breaking the data modeling and migration tasks down into increments that fit into sprints is impractical and often impossible. Teams try to cope by running a separate stream for data modeling and migration in parallel.

The diagram below shows a development team working in a typical sprint delivery process. And the data stream is running more or less in a traditional waterfall model. This mismatch is a common cause of failure in projects.

The costs to the project of getting the schema design and data migration wrong are high. Mistakes at the data migration step have impactful effects all the way down the project timeline. In our example project above, we see that when the data and development streams merge their work and code is tested against real data for the first time, re-work will be required. The forward momentum of the project is affected, and unplanned delays occur. Problems later in any project are always more expensive than those that are tackled early on.

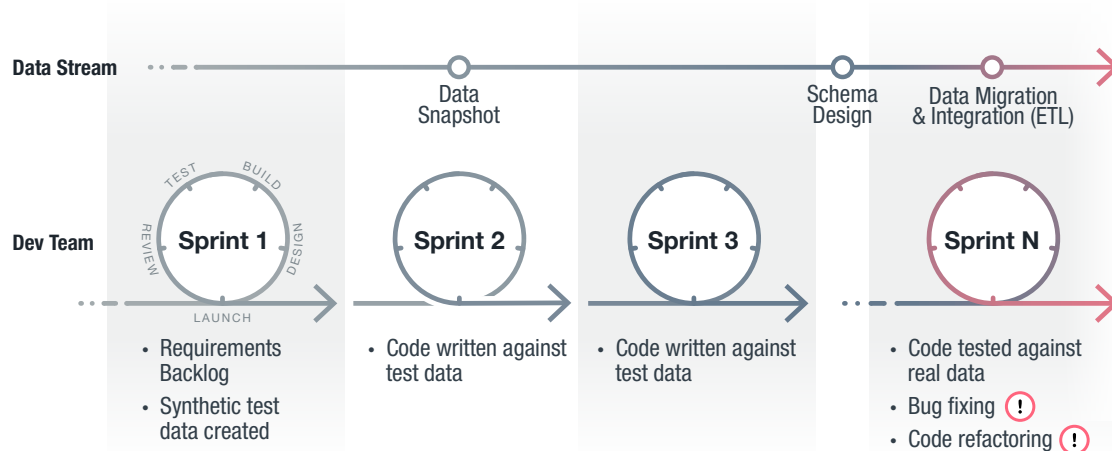


Figure 2: Agile Project Using Relational Technology

MarkLogic Projects

MarkLogic facilitates the Agile delivery model by taking a different approach to data. There is no need for — and, in fact, it is not desirable to have — a single canonical schema. Therefore, schema design is not where you should start your implementation. Rather a “Data Services First” approach is recommended. The fundamentals of project delivery using Data Services First, e.g. the skills needed, are not different from those in relational projects. But the recommended steps and their order reflect a better alignment with the goals of Agile: progressive, iterative, and fast delivery.

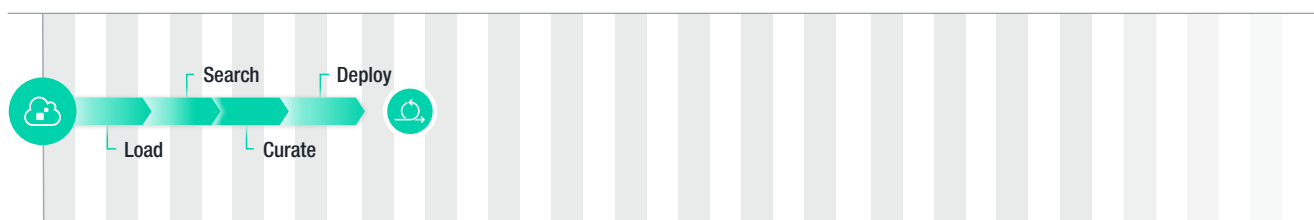


Figure 3: MarkLogic Projects

Data Services Approach

Experience has taught that focusing on delivering the most high-value functionality early in a project increases the chances of a successful delivery. Delivering a minimum viable product (MVP) into production quickly is the goal of most modern software projects. An MVP that is scoped to provide high value with low risk, and with progressive enhancements is the exemplar of Agile methodologies. But delivering incrementally with relational technologies is challenging and fraught with the risk of rework. Using MarkLogic while adopting a Data Services Approach allows project teams successfully deliver on the promise of the MVP.

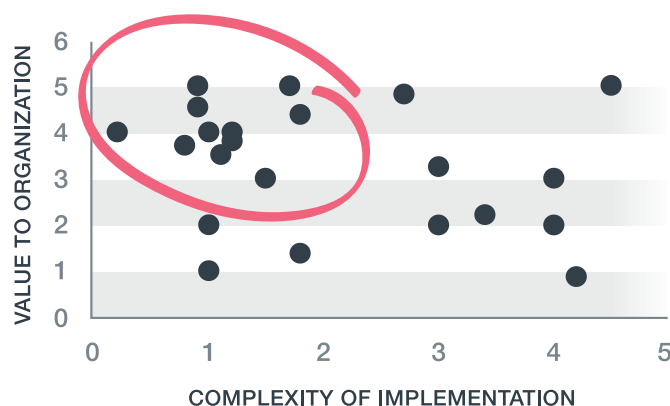


Figure 4: MarkLogic Allows Confidence In Agile Delivery

Once you have defined your MVP requirements, you can roll up your sleeves and get to work on developing to those requirements straight away. You’ll do this by focusing on those data services required for those features that deliver the highest value features first.

Below are the basic steps for implementing a MarkLogic MVP using a Data Services First approach.

Define the data service

Define what the output requirements are based upon what the service will be used for.

A data service is a fixed interface to the data managed in MarkLogic expressed in terms of how the data will be consumed by the application tier. For example, data services can run queries ("Find eligible insurance plans for an applicant"), updates ("Flag this claim as fraudulent"), or both ("Adjust the rates of plans that haven't made claims in the last year"). The data service allows the service developer to obscure the physical layout of the data and constrain or enhance queries and updates with business logic. This allows application developers to contribute fully to the project without needing expertise in MarkLogic.

Ingest data

Ingest your source data as is into MarkLogic. The harmonization steps defined by your data service model(s) will transform those pieces of the data needed for the defined data service(s). Any un-harmonized data can be processed later if and when it is needed by a newly defined data service. This is a truly progressive approach to data modeling.

Model only what you need for each data service

Create a model with the minimum set of output properties required. Modeling only what you need, when you need it, provides the opportunity to ensure that the model is created collaboratively, that it serves the business requirement and works well for developers to interact with.

Allow application developers to access the data service(s)

Now that one or more data services have been created they can be exposed so that your application developers can begin writing code against the real data. Code is tested against real data. This close alignment between data modeling and development significantly reduces the risk that code will need to be refactored later. No more surprises in the middle of your project!

Iterate

While your application developers are busy writing and releasing features with their existing set of data services, the next set of data services can be defined and built. Data Services facilitates a truly iterative methodology.

Adopting the Data Services First approach reduces the complexity of your project. Dispensing with the need for separate data and development streams reduces risk and administrative oversight requirements.

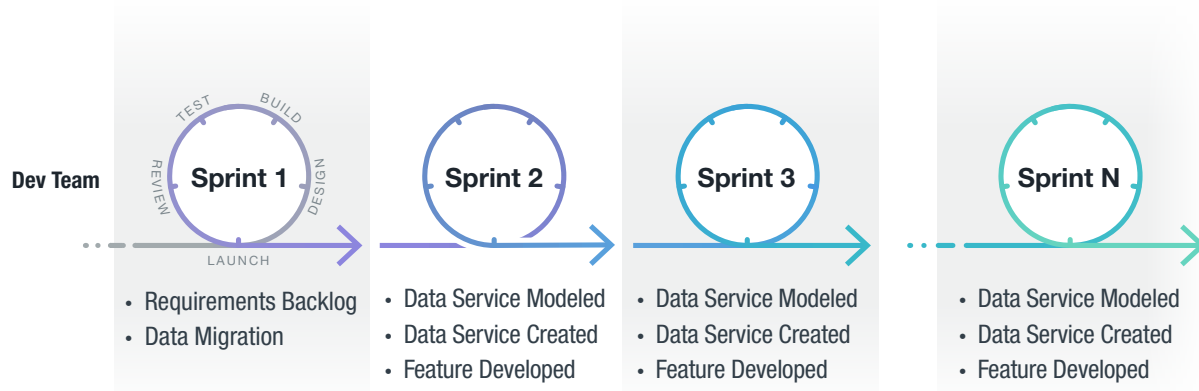


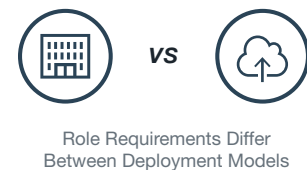
Figure 5: Data and Development Work Together

The MarkLogic Project Team

Resourcing your MarkLogic projects does not require reinventing the wheel. You will be able to leverage the skills and talents that already exist in your organization. The people don't have to change – the order and timing of their input will be adjusted to take advantage of the increased agility and speed of the data services approach.

Form Your Approach

First and foremost, determine your deployment model. Will your application be hosted in the cloud or on-premises? Will you manage your own cloud instance, or will you take advantage of the MarkLogic Data Hub Service? The depth of expertise you need in your team will vary depending upon which model you select.



Recommended Deployment Models

You can run MarkLogic wherever you want. You can start on-premises and migrate to your chosen cloud provider later on. You are not locked into any one deployment model. It is important, however, to consider what deployment model you will be using now, as this will influence the number of resources and their depth of needed expertise.

Data Hub Service

Deploying your application using the MarkLogic Data Hub Service lowers the operational burden to your team. As a fully automated cloud service, it's effortless to run with no infrastructure to buy or manage. This deployment model requires fewer administrative resources as the Data Hub Service provides on-demand capacity, auto-scaling, automated database operations, and proven enterprise data security via the managed service.

MarkLogic Managed Service

MarkLogic Managed Services allows our customers to outsource the day-to-day MarkLogic database and data management operations to our experts, allowing your IT staff to spend more time supporting your core business needs.

MarkLogic keeps your business-critical workloads running in the cloud as expected through round-the-clock management and monitoring. We manage your data and application in your cloud-hosted MarkLogic deployment. MarkLogic Managed Service is available for implantations that do not utilize Data Hub.

Self-managed Cloud Deployment

In a self-managed cloud deployment your team will need to manage the cloud instances, demand, scaling, and database operations. More MarkLogic expertise is required to administer your application than with one of the managed services options.

On Premise

For on-premise or hybrid deployments all aspects of the database must be managed by your team. This deployment model requires more in-depth expertise for the management of your application.

Team Roles

Once you have identified your deployment model(s) you can identify the right people to fill the roles in your project team.

MarkLogic is a cutting-edge platform that allows your teams to build innovative new solutions. But a successful MarkLogic project does not require a new group of employees with scarce or specialized expertise. The ideal team for a MarkLogic project will, in fact, look very familiar to you.

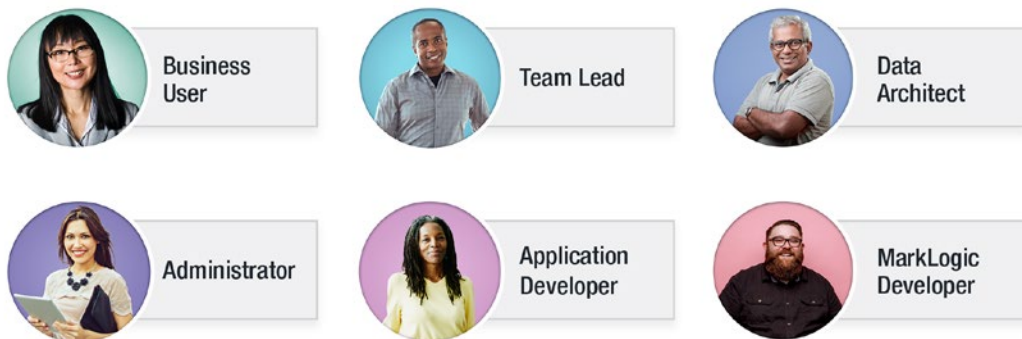


Figure 6: Typical Roles in a MarkLogic Project

Project Governance Roles

Good project governance is a universal requirement for successful delivery of an IT project. And while governance activities are not technology dependent, our experience has shown that team members in the governance track perform better in their roles and with more satisfaction if they have some understanding of what MarkLogic is and the role it will play in the application.

Project Sponsor/Business Owner

An effective project sponsor/business owner is key to the success of any software development project. They control the budget and the business inputs, and remove barriers as needed. And most importantly they understand and can communicate the goals and objectives to the project team. This is the definition of a good project sponsor for any type of software development project.

Your sponsor's primary role is to know the business needs, and no explicit MarkLogic expertise is needed to fulfill that role. However, we do find that some grounding in the basics helps sponsors to better understand the context of what, why, and how the technical staff are working.

Team Lead

This person is the central hub of the team. They need to be able to communicate well with both business and technical team members and be comfortable running cross-organization integration projects. In addition, they need to know how to manage requirements, capabilities, and budgets.

Like the sponsor, they do not require in-depth technical expertise in MarkLogic. But it is essential that they understand the art of the possible, so that they are able to participate in the decision-making processes of the technical team.

Business User

We recommend between 3-7 key business stakeholders on the team, depending upon the size and complexity of the project. Direct interaction with the day-to-day business users and their domain expertise provides in-depth knowledge of the use cases to the team. They can provide guidance during development and testing through continuous feedback.

The business users primary role is to provide their expertise on the business requirements and current as-is-state, and they do not require any technical expertise in MarkLogic to function well in that role. But similar to the business sponsor and the team lead, it is ideal to have a basic foundation.

Technical Roles

Administrator

The number and expertise level required of your database administrators (DBAs) will depend upon the deployment approach. The number of persons required to administer even the more complicated on-premise hosted applications is generally lower than those required to administer traditional relational technologies.

For those projects utilizing MarkLogic Data Hub Service, DBAs will need to know the basics of administration and how to deploy code and load data, but will not perform the day-to-day database administrative tasks. For on-premise or self-managed cloud deployments, more in-depth training is recommended. Training courses for administrators are tailored to the different deployment models.

MarkLogic Developer

These are the developers that will work on the core MarkLogic database. They will design and build the data services on the MarkLogic database platform, including data hubs, RESTful services, and data flows. Only a small subset of developers on the team will need to do direct development in the MarkLogic database itself.

Developers of any background can be trained as MarkLogic Developers. A grounding in JavaScript is helpful, though not required. These developers will gain a deeper understanding of MarkLogic than the other members of the team. Training and certification resources are available, as well as mentoring from experts at MarkLogic.

Application Developer

Application developers, who make up the majority of most project teams, will not need to become experts in MarkLogic development. The underlying complexities of the database and data structures should be abstracted away by data services created by the smaller number of MarkLogic developers on your team. They will need to know how to access data via web services — but they don't have to know a lot about the specifics of MarkLogic.

Application developers should have a good understanding of how to develop applications with Web Services, APIs, and middle tiers (e.g. Node.js, .NET, Python, Java).

Data Architect

One of the biggest differences in a MarkLogic project is how you and your team interact with the data and how and when data should be migrated to MarkLogic. The data architect is the person who understands your data environment, the footprint, and where the data needs to come from. They will need to know enough about MarkLogic to understand how the ingestion and data harmonization flows work. But they will not need to have as much in-depth knowledge as the MarkLogic Developers.

Resources to Support Your Implementation

MarkLogic provides resources and services to help make sure your team ramps up quickly and effectively. Every organization has its own ways of learning and working. And your needs will change over time.

As you embark on your very first MarkLogic project — training, certification, and mentorship are likely to be more important. Having your developers and administrators complete the certification process provides hands-on experience to facilitate rapid ramping. With more experience under your belt, a MarkLogic Center of Excellence (COE) inside your organization will help to establish a repeatable model for success.

MarkLogic University

MarkLogic University (MLU) is more than just a course catalog. Attending lectures on their own is no guarantee that students can translate knowledge into hands-on proficiency. Therefore, in addition to the wide range of courses offered, MLU provides hands-on video tutorials, project simulations, and mentorship to help accelerate skills acquisition.

Students are guided through the process from initial course work through to certification as a MarkLogic Developer or Administrator. Holding certifications demonstrates real-world expertise and provides confidence to you and your customers that your team can deliver a successful MarkLogic implementation.

To get started students should select the right track for their role in the team.

Role-Based Learning Tracks

Training courses are delivered in a variety of formats, languages, and geographic locations to provide flexibility for each student's learning styles. They are tailored into role-based tracks to ensure each attendee gets the most from their time invested.

Tracks for the Business User and Data Architect are designed to provide a solid foundation of knowledge needed to work successfully on a MarkLogic project. The Developer and Administrator tracks help students get started with confidence, and guide them all the way through to full self-sufficiency.

For administrators, the training needs will depend upon your chosen deployment model. There are two administrator tracks, one for self-managed deployments (either on-premise or cloud-based) and one for administering the Data Hub Service.

MarkLogic Developer	Application Developer	Administrator	Data Architect	Business User
Bootcamp	Java Client API	MarkLogic Server Administration	Data Modeling	MarkLogic for Business
Data Hub Flight School	Node.js Client API	Data Hub Service Administration	Data Hub	
Relevant Specialty Courses	Relevant Specialty Courses	Relevant Specialty Courses	Relevant Specialty Courses	
<i>Continuous learning through Community</i>	<i>Continuous learning through Community</i>	<i>Continuous learning through Community</i>	<i>Continuous learning through Community</i>	
Certification	Certification	Certification		

Figure 7: MLU Role-Based Tracks

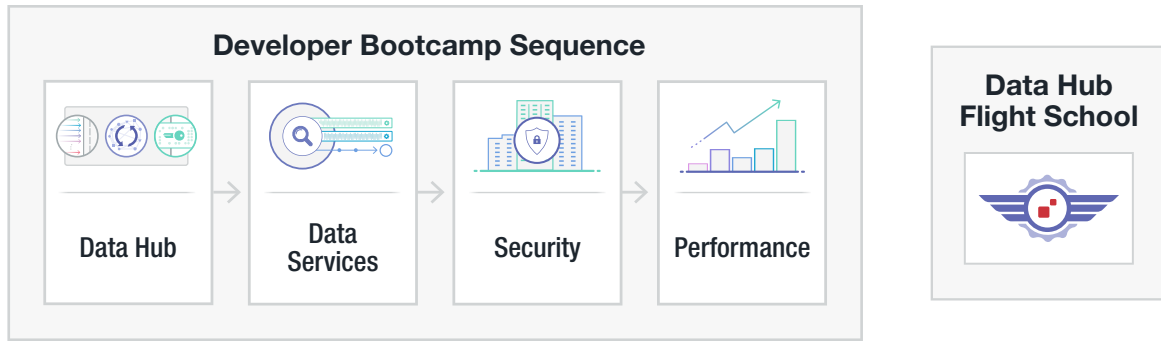


Figure 8: Developer Bootcamp Sequence

MarkLogic Bootcamp is the ideal training sequence for the MarkLogic Developers on your team. The Bootcamp provides a new developer with a focused track that prepares them to work on data hub use cases through the widely used lens of JavaScript programming.

Completion of Bootcamp training provides a developer with a host of new skills. But completing a training course and delivering a project are different. And to ensure that developers are able to synthesize and apply those new skills, we offer the Data Hub Flight School. The Data Hub Flight School is designed to lead a developer through a simulated real-world project.

There is also a track for Application Developers.

Course Formats

Courses are offered in a variety of formats to support different learning styles and geographic locations. Regardless of format, students are able to validate their level of achievement against course objectives via learning assessments. Students with a more advanced skillset may also use the learning assessments to test out of pre-requisite courses.

Self-paced

A large number of the courses and tutorials are available as on-demand, self-paced modules that can be taken on their own or in combination with instructor-led courses.

Instructor-led

Instructor-led courses are delivered via a live online format that enables students to attend the course from anywhere in the world. Courses are offered on a worldwide timetable and delivered in English, French, and Japanese.

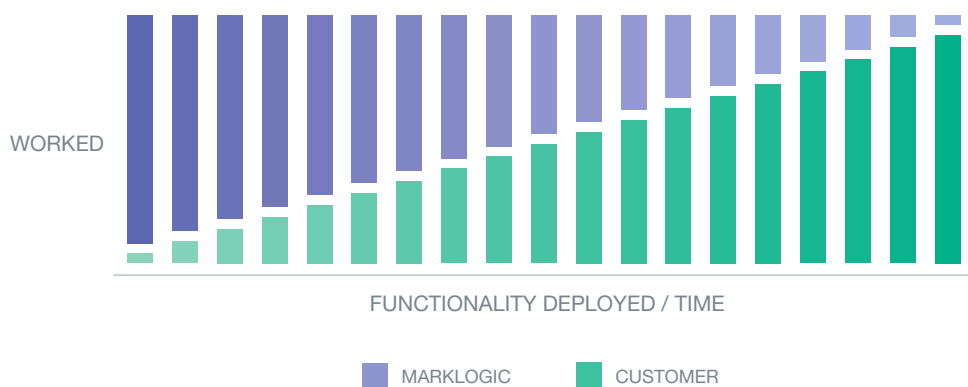
Instructor-led training courses utilize an interactive virtual classroom experience that includes a hosted lab machine for the student to use for hands-on work.

The full course catalog and schedule is available online at <https://www.marklogic.com/learn/university/>

Mentorship

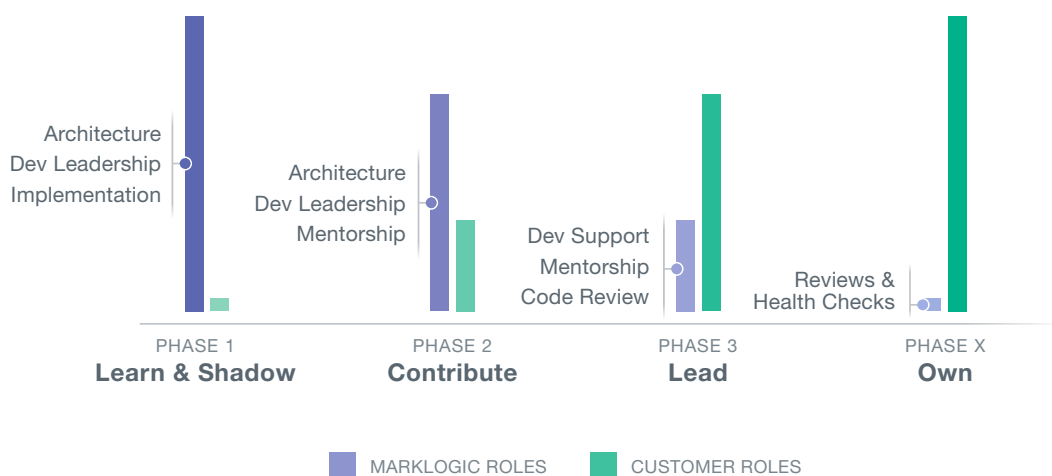
Training is the ideal place to start. But jumping into your first project on your own can be daunting. Having experts on hand to support your team through their initial project greatly improves your opportunity for success. MarkLogic consulting services provides flexible engagement models to ensure that your team gets the support it needs.

The consulting services team puts an exceptional focus on customer success. With vast experience solving big data challenges for some of the world's most complex data projects, we're committed to helping you scope and select the best application of MarkLogic technology.



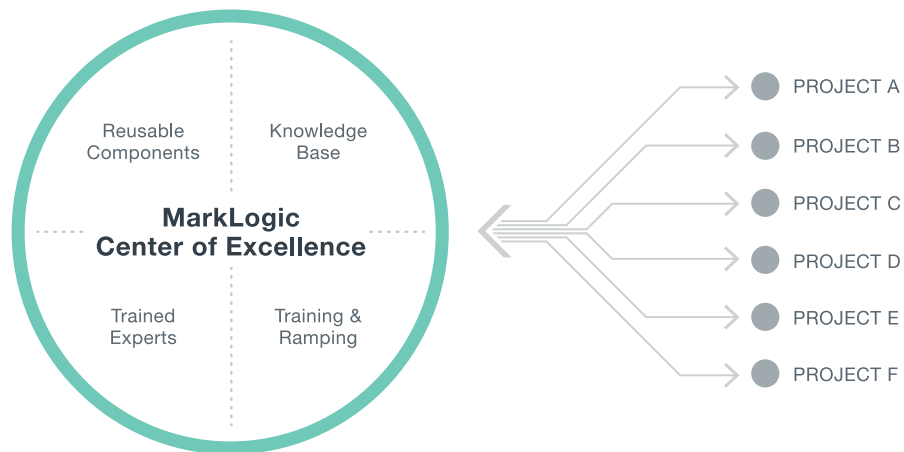
We can help you identify and implement an MVP to provide high value with low up-front risk. We can train and mentor your team members over the course of project iterations. This co-implementation model allows your team to transition into a technical leadership role over time.

As your team grows in confidence, the MarkLogic consultants can be available in an advisory capacity, as needed.



Center of Excellence

You've invested in a well-trained team, and you have completed your first MarkLogic implementation with support from the MarkLogic consulting team. Establishing a MarkLogic center of excellence (COE) is the best way to make the most of that investment by building a repeatable model for success. Forming a COE establishes skills development and consistency of adoption across your organization.



Centralizing people, processes, and standards of best practice will allow each of your projects to leverage the skills and capabilities acquired via successful implementations. The COE model facilitates the most efficient and effective communication across teams working on or planning MarkLogic implementations.

MarkLogic will support your COE from initiation through operation. From the baseline assessment through to standing up, each COE is tailored to the specific needs of your business.

At every stage of your journey from training through to COE, MarkLogic provides resources to support your organization's delivery capabilities.



Get in Touch

The IT landscape has changed dramatically and with it, so have the expectations of your customers, both external and internal. The tools and technologies that have constituted the bedrock of the enterprise application are not capable of supporting the pace of change required to support the challenges facing business today.

Successful delivery is measured in days and months, not years. MarkLogic supports successful delivery by providing the flexibility and agility without sacrificing any of the enterprise functionality that your customers will take for granted.

To discuss more about how to successfully form your team, contact consulting@marklogic.com.

About MarkLogic

Data integration is one of the most complex IT challenges, and our mission is to simplify it. The MarkLogic Data Hub is a highly differentiated data platform that eliminates friction at every step of the data integration process, enabling organizations to achieve a 360° view faster than ever. By simplifying data integration, MarkLogic helps organizations gain agility, lower IT costs, and safely share their data.

Organizations around the world trust MarkLogic to handle their mission-critical data, including 6 of the top 10 banks, 5 of the top 10 pharmaceutical companies, 6 of the top 10 publishers, 9 of the 15 major U.S. government agencies, and many more. Headquartered in Silicon Valley, MarkLogic has offices throughout the U.S., Europe, Asia, and Australia.

© 2019 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED. This technology is protected by U.S. Patent No. 7,127,469B2, U.S. Patent No. 7,171,404B2, U.S. Patent No. 7,756,858 B2, and U.S. Patent No 7,962,474 B2. MarkLogic is a trademark or registered trademark of MarkLogic Corporation in the United States and/or other countries. All other trademarks mentioned are the property of their respective owners.

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