



Agile Data Delivery with Talend Data Fabric



Table of contents

- Executive Summary 2**
- Agile Data Integration with Talend Data Fabric..... 2**
- Implementation Success 3**
 - Selecting the Initial Implementation Scope..... 3
 - Agile Data Integration Sprints 4
 - Agile Data Management 5
- Choosing the Right Team..... 6**
 - Business and Data Subject Matter Experts..... 6
 - Program Governance 6
 - Architecture..... 6
 - Engineering..... 7
- Resources to Support Your Implementation..... 7**
 - Customer Success Management 7
 - Talend Academy..... 7
 - Strategic Architects 7
 - Technical Consultants 8
 - Talend Partners..... 8
 - Customer Support..... 8
 - Forming a Centre of Excellence for your DataOps..... 8
- Get In Touch 8**
- About Talend..... 9**



Executive Summary

It is unquestionable that in order to create and maintain competitive advantage, all businesses need digital capabilities embedded into the core of their organizations. But digital transformation does not just mean automating a few of your business' processes. The most successful businesses are data driven.

If the efficient exploitation of data is the differentiator between success and failure, then your organization cannot afford to be slowed down by legacy technologies and legacy methodologies. 60% of organizations report that they are behind on their digital transformation journey¹. In the face of this pressure, the pace of change will only increase, and adaptability in the face of digital disruption and transformation is crucial to success.

In a data-driven business, data integration is the most important thing to get right. Yet most organizations avoid tackling it head on due to the perceived impossibility of doing it right. They pick around the edges of what seems to be a problem that is too large and too complex to solve. It is one of the last practices in software development resistant to change. How many times have you heard "it's not possible to do Agile data integration"?

Those businesses that have successfully exploited advancements in DevOps and cloud computing are beating their competitors who have not adapted quickly enough. Data integration will be significantly impacted by these advancements. The next wave of transformation will be to DataOps. The ability to adapt and transition to DataOps will separate the next group of leaders from the pack. Businesses that are still dependent upon legacy data integration tools that only support legacy methodologies, are poorly positioned to successfully make this transition.

In this guide, we will detail how to implement Agile data integration and achieve DataOps using Talend Data Fabric, and how to identify the right roles and resources to ensure success.

Agile Data Integration with Talend Data Fabric

Legacy data integration technologies tend to focus primarily on one aspect of data integration, ETL (extract, transform, and load). Tremendous resources are consumed just for the lifting and shifting of data using inflexible tools and processes. Because of this, businesses have to spend their remaining resources trying to make critical decisions using data that is incomplete, noncompliant, untrustworthy, and often incorrect².

Yes, it is incredibly difficult to do Agile data integration with yesterday's tools. This is why at least 45% of data transformation projects fail³. Your data integration strategy must be about more than just moving the data from point "A" to point "B". Effective Agile data integration incorporates governance, compliance, quality, security, sharing, support for real-time processing, and application integration capabilities. Trying to bolt these new processes onto a fragile legacy data stack only increases the complexity and risk. The DataOps winners will have the capability to execute these tasks in continuous, automated processes.

Designed for the DataOps revolution, Talend Data Fabric provides a unified, cloud native platform that allows businesses to achieve Agile data integration success. Built for flexibility, it allows connection of any data type or source for deployment anywhere; on-premises, cloud, or hybrid. Talend Data Fabric scales without penalty and will integrate easily with tomorrow's innovative technologies. It allows you to build iterative, CI/CD (Continuous Integration / Continuous Deployment) data integration capabilities without sacrificing the tenets of data governance such as data quality, security, and interoperability.

¹ Forrester, The CIO and the Emerging Digital Crisis, 2017

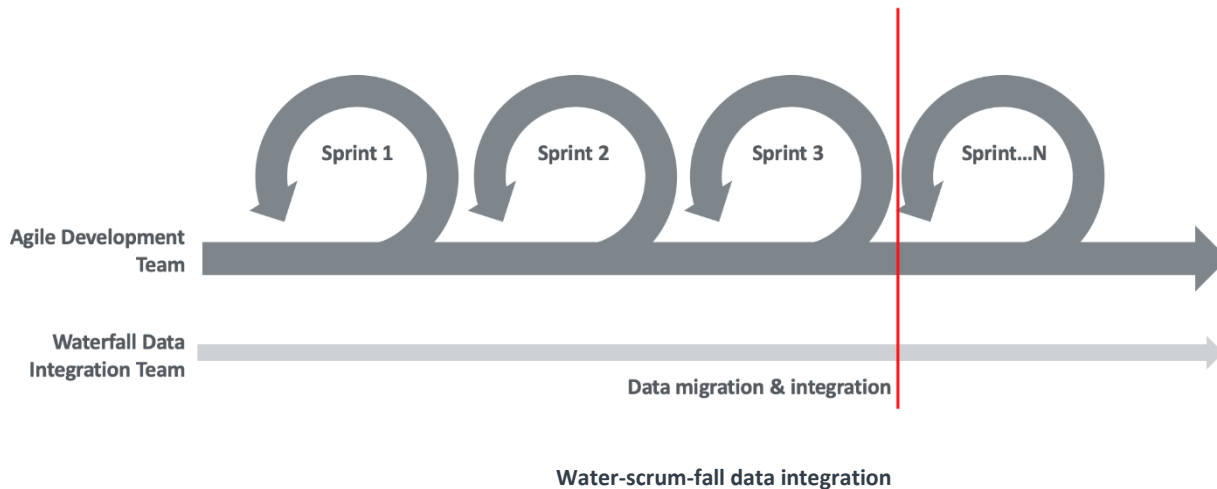
² Zoominfo, 35+ Statistics About Dirty Data, 2018

³ McKinsey Digital. Digital Transformation: Improving the Odds of Success, 2019

Implementation Success

Agile methodologies have enjoyed wide adoption across the software development landscape due largely to faster delivery of time to value. The ability to deliver working software and analytics results early in the project helps to secure business buy-in and deliver value. Despite this, organizations that use Agile methodologies for their application development programs often still rely on waterfall or water-scrum-fall when it comes to their data integration workloads. The conventional wisdom is that this is the least risky way to work.

In a water-scrum-fall software project, the development team is working iteratively using Agile iterations/sprints, while the data integration team is working to a rigid waterfall timeline. The data migration team is working to deliver a final set of perfect data in a large batch process. Because of this, mismatch of timings, the development team won't get the chance to integrate their code with real data until late in the project. Experience tells us that large amounts of refactoring late in a project can stall momentum and cause unplanned delays. Trying to manage two separate sets of processes and methodologies leading to the same project go-live date increases the risk to project delivery success.



Ideally, you want your data integration team to work hand in hand with the rest of your development team in the most efficient way possible. But that doesn't mean you have to run 30- to 45-day development sprints, with weeks and months before feature releases, to accommodate the waterfall style of the data integration team.

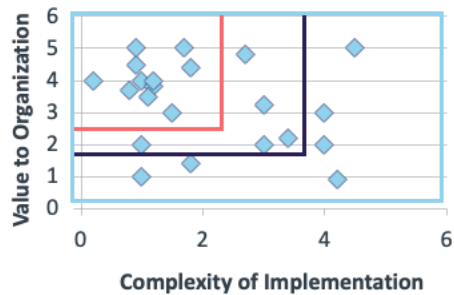
Talend Data Fabric facilitates the closer alignment of the data integration team with the developers and data users. It eliminates the need to kick-off a separate data integration workstream months or years in advance. Run your data integration tasks in your sprints alongside the developers, data scientists, and business users who will consume the data for software development and analytics.

Selecting the Initial Implementation Scope

Running successful Agile data integration projects with Talend Data Fabric should look quite familiar to anyone using Agile for software development. Because of its flexibility, Talend Data Fabric enables your team to focus on identifying and assessing the most important use cases for development. Business requirements can drive the data integration timeline, rather than the difficulty of data integration dictating the order in which work is carried out.

Phase 1 is best scoped to target the high value use cases where data subject matter experts (SMEs) are available and the identified data sources are easily accessed. Subsequent deliveries will build upon the solution already established and incrementally add more data sources, enhanced data

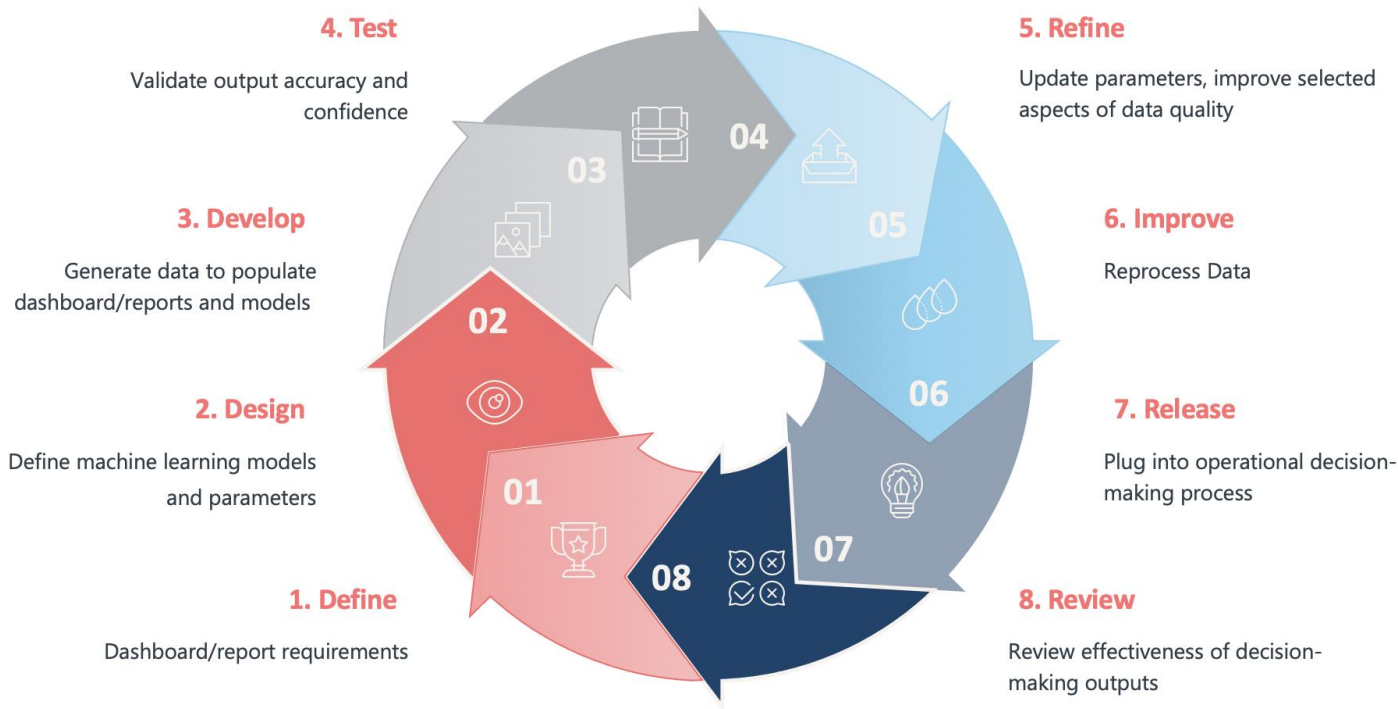
flows, and application capabilities. This approach, as shown in the diagram below, allows more time for more complex cases to be tackled in an order that reduces burden of risk to the project.



A timeboxed implementation is the recommended starting point and lays the foundation for project success. It drives focus from the team, and that strategy will drive critical early success and buy-in from the business. This incremental approach to data integration removes the penalties associated with change in a traditional data integration project. Flexibility is built into the project from the start.

Agile Data Integration Sprints

With the scope for the phase 1 delivery identified, the data integration work can now be run in sprints alongside the development team.



Example data integration sprint

The diagram above shows an example sprint for a data integration project. In this example, the team is tasked with developing a report dashboard.



Define

The team works directly with business SMEs, data scientists, and users to define the requirements for the dashboard screen.

Design

The team designs the machine learning models and parameters for the report data. The data can be previewed, profiled, and scored to aid in the design process.

Develop

The data for the dashboard screen is generated and the screen is populated. The development team members building the reports have been able to build those new screens using real data instead of synthetic test data.

Test

Testers and business users test the validity and accuracy of the data by interacting directly with the data in the dashboard. This provides an opportunity for feedback on both the data and the new dashboard screen together. This vital input is given to both the data integration and software developers at the same time.

Refine

Feedback on the data can be used to update parameters to improve data quality, governance, and security.

Improve

Data is reprocessed with the updated parameters. The newly reprocessed data can be integrated with the dashboard screens that have also been refactored in the sprint.

Release

Now that the screens and their data have been designed, built, tested, and improved they can be released. The dashboard can be plugged into operational decision making and can start to deliver business value. The team can move on to the next dashboard screen or feature.

Review

The business users can track the impact the new data has on decision making processes, and can feed their findings back to the Agile team to help improve the design of future sprints.

Agile Data Management

Delivering iteratively and incrementally each sprint does not mean that you can dispense with critical data management tasks. It is essential to perform these tasks iteratively throughout the project lifecycle. Talend Data Fabric makes it easy to govern your data and ensure its integrity with embedded, machine learning-aided data quality functionality throughout.

Data Architecture

In traditional data integration projects data architecture, modelling, and design are seen as a monolithic set of steps that have to be completed for all data before the project can progress and data can be moved. Talend Data Fabric's schema-agnostic capabilities support true Agile data modelling. Your team does not have to get everything right the first time or risk disruption to the project. Data architecture can be done successfully when underpinned by modern data integration tools that support schema on read, live previews, data discovery, and profiling. These tools allow data to move more frequently and faster without sacrificing understanding and control of the architecture.

Data Governance

Maintaining pervasive data quality and governance at speed may sound like an impossibility. Maintaining end-to-end lineage and provenance allows data to flow faster while keeping the risk profile low. Using data integration tools that provide data quality and trust scoring allows quality to be improved and maintained as it flows through pipelines. Robust monitoring allows you to track progress and improve control over your data.



Data Security

The ability for non-technical team members to interact directly with the data is a large part of the reason that Talend Data Fabric fits so well into Agile development processes. The activities of these users require data to be shared safely by default. More frequent processing of smaller amounts of data creates more entry points into the data. This can increase the risk to any data integration project. Talend Data Fabric's robust security architecture ensures that the democratization of data integration activities is not accompanied by a higher security risk.

Data Interoperability

DataOps cannot be realized unless the data can flow in a frictionless manner. Out of the box connectors and advanced API development capabilities reduce cycle times. This allows data to flow faster, and for pipelines to be refactored more easily.

Choosing the Right Team

Frictionless data integration and management becomes achievable when the right tools are paired with the right methodologies. Agile data integration done right makes the move to DataOps a practical reality for your customers. But the journey to DataOps is not just a technology and process change, it is also a mindset change. That doesn't mean you need a whole new workforce. It means that you can make better use of the range of skills and roles that you already have available to you.

As the data workstream is going to be integrated into the wider team, it will be easier to bring expertise into closer interaction with your data integration experts. Close alignment with your end customer and their objectives is more important than technology factors in determining the success of any implementation project. For example, it's important to have strong executive sponsorship, regardless of the technology. Including business users in the data integration work will lead to increased chances of success.

Business and Data Subject Matter Experts

The iterative nature of the process means it is critical to have direct engagement from SMEs and end users right from the beginning. These roles build the models, run the analysis, and generate the reports developed during each sprint. Performing these tasks in every iteration is the key to true agility.

Business Roles: Data Steward, Data Scientist, Data Worker, Data Privacy Specialist

Program Governance

Program Governance by your program manager ensures stakeholders are well heard and well informed, and that there is proper prioritization of requirements. Domain expertise provided by business analysts and business line users allows the solution to be properly targeted for the business need.

Governance Roles: Program Manager, Project Manager, Scrum Master, Business Analysts

Architecture

Your architects will be working directly with the project team during every sprint to further refine the models based upon the input from the business and data SMEs. They oversee the data management tasks ensuring that governance, security, and interoperability are maintained and improved during every sprint.

Architecture Roles: Solution Architect, Data Architect



Engineering

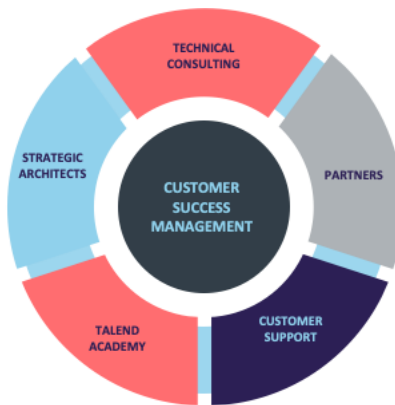
The engineers on your team will work closely with both the architects and the business SMEs to deliver the data needed in each sprint. They will be working on smaller, more easily managed tasks than if they were working in a legacy data integration project.

Engineering Roles: Integration Developer, Automation Developer, Quality Engineer

Resources to Support Your Implementation

Even the most effective teams, using the best tools and the right methodologies, need help sometimes. Talend resources are there to help make sure your team ramps up quickly and effectively, and will be there to support you every step of the way.

Customer Success Management



Customer Success Management Model

Engaging a Customer Success Manager (CSM) means having an advocate to help you plan and coordinate how you'll achieve your objectives with Talend. Your dedicated CSM will orchestrate all of the available Talend resources to ensure your success.

Talend Academy

The best technology tools in the world are nothing without the trained and skilled staff who make them work. Experience tells us that the most successful implementation teams start with training. Putting this at the very beginning improves your teams' understanding and efficiency.

Talend Academy provides a platform for your team to ramp-up their Talend knowledge, access implementation accelerators, and update their skills as we release new product capabilities. Training can be done self-paced online, or instructor-led.

Talend Academy is also the platform through which our Delivery Innovation team releases their content. Our Delivery Innovation team is a key asset to your success as we aggregate best practices from our own employees, partners, and customers to provide solution templates, frameworks, and reference architectures among other assets.

Strategic Architects

Strategic Architects form trusted, long-running relationships with their customers by providing thought leadership on how to achieve the optimal architecture.



Technical Consultants

Technical Consultants provide the expertise to help you get your solution into production, as well as assisting with major new initiatives, hands-on problem-solving, and periodic technology refreshes.

Talend Partners

Talend Partners provide key industry knowledge, complementary technology, and skilled development and operations capacity.

Customer Support

The Customer Support team is there to help you get answers to your product-related technical questions as well as to interface with our Research & Development team on your behalf.

Forming a Center of Excellence for your DataOps

A Center of Excellence (COE) centralizes efficiencies and aligns your organization to a consistent best practices approach. A successful COE typically includes a pool of experts, standardized learning paths tailored for your organization, organization-specific best practices and development methodology, and standardized development tools and frameworks.

Talend will support your organisation's DataOps COE from inception to operation, allowing you to leverage Talend technology across projects or workstreams without incurring the same initial start-up costs on each new initiative.

Get in Touch

Data integration technologies and methodologies are evolving at an ever-accelerating pace. The winners will be those businesses that embrace the coming DataOps wave. Talend Data Fabric provides the technology to underpin a successful transition to truly Agile data integration methodologies, and you don't need to reinvent the wheel. You already have the skills and capabilities in your teams to embrace this change successfully.

To learn more about how Talend can support your implementation success please contact us at: <http://www.talend.com/contact-sales/>



About Talend

Talend, a leader in data integration and data integrity, is changing the way the world makes decisions.

Talend Data Fabric is the only platform that seamlessly combines an extensive range of data integration and governance capabilities to actively manage the health of corporate information. This unified approach is unique and essential to delivering complete, clean, and uncompromised data in real-time to all employees. It has made it possible to create innovations like the Talend Trust Score™, an industry-first assessment that instantly quantifies the reliability of any data set.

Over 6,500 customers across the globe have chosen Talend to run their businesses on healthy data. Talend is recognized as a leader in its field by leading analyst firms and industry media

For more information, please visit www.talend.com and follow us on Twitter: [@Talend](https://twitter.com/Talend).

