



# **8 REASONS WHY YOUR LEGACY DATA INTEGRATION PLAN MUST CHANGE**

DAVID S. LINTHICUM



*Most legacy data integration technology is outdated, and can't keep up with the emerging patterns of data use. The old technology needs to be replaced as soon as possible with data integration technology that can support the "8 reasons" presented in this paper.*





## Table of Contents

Executive Summary	1
The 8 Reasons	3
1. Application and Data Convergence	3
2. Need for Self Service Data Integration	3
3. Rise of “Big Data”	4
4. The Rise of the Data Lake	5
5. Applications Moving to the Cloud	6
6. Data Moving to the Cloud	7
7. Rise of the Internet of Things (IoT)	7
8. Rise of Microservices	8
Call to action	9
About SnapLogic	10



## Executive Summary

The growth of data in enterprises is a well-documented concept. Indeed, every analysts' report shows the massive growth of data within enterprises. Not surprisingly, most analysis includes reports of enterprises' inability to manage and make use of that data.

The problem, may be not in the growth of data, which will always be constant, but in the inability to properly integrate the data. Data growth is occurring within existing silos. Worse, the number of silos continues to increase with the addition of new systems, including big data systems, cloud-based data, and the emerging use of the Internet of Things (IoT) data automation.

We will continue to see overall data growth for the 8 major reasons outlined in this paper, which include:

**The evolving use of data.** This is also a constant, meaning that we'll deal with this issue forever. As technology changes over time, the need to maintain data around these new systems pushes many enterprises to just leave the data within silos.

**Diverse data types.** Once we could assume that data existed in traditional relational databases, or behind well-defined application interfaces (e.g., SAP). These days, data exists in object-based databases, cloud-based databases, and data streams from IoT devices, and the data may be structured or unstructured. The lack of a common expandable and adaptable integration layer makes most of this data useless outside of its home systems.

**The desire to finally go "real time" with data delivery.** The use of data from system to system in real time or near real time was once a utopian dream. Now the dream is reality, and the ability to see up-to-date data within all systems is becoming table stakes for most modern applications. Enterprises see the benefits, but struggle to deliver this up-to-date data to internal and external customers.

Now the bad news, which isn't really news to most of us: Existing "legacy" data integration practices and technologies are woefully lacking. While many enterprises partnered up with state of the art data integration players at the time, the issues with the growth and evolution of data, as described above, have stymied the enterprise. Indeed, traditional data integration approaches actually hinder the strategic use of data. These approaches need to be corrected before significant change can occur, such as cloud enablement, IoT, and the use of data lakes.

In this paper, we lay out the case for a data integration upgrade by helping you understand the "8 Reasons Why Your Legacy Data Integration Plan Must Change." While a single reason would be compelling, so much is changing inside of the enterprise that it creates an overwhelming business case to fund the updating of your data integration technology. In many cases, data integration should become the number one priority before building or migrating any additional systems that are data dependent.

Conclusions reached include:

- Today there are significant opportunities within enterprises that drive new uses for data. The inability to provide sound data integration practices and technology hinders the use of data.

- Data integration is a strategic advantage, directly tied to the success of the business. Those who can't correctly practice data integration will likely fall by the wayside within their industry.
- The strategic use of data integration technology should rise above other IT priorities. Consider the importance of data, and new systems built on the use of data, and thus data integration technology.
- Most legacy data integration technology is outdated, and can't keep up with the emerging patterns of data use. It needs to be replaced as soon as possible.

## The 8 Reasons

The fact that there are 8 major reasons to strongly consider moving away from your traditional data integration approaches should be enough to get your attention. Although all of these reasons may not relate to your particular situation, the idea is to think about these concepts in order to define the changes your enterprise needs to make.

Take these as concepts to explore and build upon, and include an ROI (return on investment) model to present to stakeholders within the organization. The idea is to build a framework of best practices and technology that will provide a foundation for most change and growth that will occur in the next 10-20 years. In other words, the next wave of data integration technology should build toward change and growth, and not just solve the problem for a single instance in time, as we did in the past.

Let's start our review with reason number one to change your data integration plan:

### 1. Application and Data Convergence

In a January 2016 report, Gartner predicts “35 percent of enterprises will consolidate their data integration and application integration competencies as one team for aligning disciplines and technologies.”<sup>1</sup>

“*Integration leaders are increasingly recognizing the value of leveraging integration competencies in a consistent way across both application and data integration disciplines. Today's digital business requires application and data integration to work closer together to enable more optimized integration delivery. Converging both will drive staffing optimization and collaboration between practices and teams to align skills for common goals.*”

In other words, there should be no distinction between application data and stored data. We made distinctions in the past between application integration and data integration, and even addressed each problem domain with its own data integration technology. That has proven to be folly.

Data is information, and data resides everywhere. The wide distribution of data makes the ability to deal equally with all data an imperative. However, most within enterprise IT dealt differently with the different types of data. This approach created silos of systems that now need to be broken apart. Consolidation is a fundamental task that most of us must complete on the journey to 100 percent integrated data.

### 2. Need for Self Service Data Integration

In an article by John Whittaker, published in The Data Warehouse Institute, he defined the concept of self-service data integration.<sup>2</sup>

---

<sup>1</sup> <https://www.globalscape.com/blog/how-converge-application-and-data-integration>

<sup>2</sup> <https://tdwi.org/Articles/2013/01/08/Self-Service-Data-Integration.aspx?Page=1>

“Essentially, self-service data integration allows IT to partner with business analysts to provide actionable information to the enterprise's executive officers in a more responsive and timely manner. It releases IT to focus on adding value to the enterprise with such activities as improving performance levels, boosting availability, and developing new applications that better align with business activities. Self-service data integration also is a cost saver, reducing the need for an army of IT specialists down to just one data analyst to cleanse and integrate the data.”

The story here is that IT needs to remove itself from the data integration use cases. Those who consume the data need to understand how they are empowered to create the data integration solution they need, on-demand, and existing within a sound security and governance framework.

This approach provides a few primary benefits:

- The ability to create a data integration abstraction layer, or layers, that's customized for the specific users and use cases.
- The ability to create a governance and security layer to insure that the data integration technology is not used improperly or insecurely.
- The ability to remove IT from all steps in this process, placing the power of data integration into the hands of those who can most benefit from it.

For most in traditional IT who leverage legacy data integration technology, this approach is an impossible task. The technology is not conducive to allow end users to leverage it because of the complexity and programming that is typically a part of “old school” data integration technology.

Updates to data integration technology and approaches should include the selection of technology with end-user interfaces that can be driven by those outside of IT. Again, this is prepping for growth and change, and the ability to empower those who need the data versus making them go through a long and time consuming process each time that they need data.

### 3. Rise of “Big Data”

Two recent surveys shed new light on the state of big data, showing growing adoption but persistent data management and integration challenges.

Earlier this year (2016), management consulting firm, NewVantage Partners (NPV)<sup>3</sup>, released the results of its 4th annual Big Data Executive Survey. Survey respondents were Fortune 1000 senior business and technology executives. Last month, IT industry association, CompTIA, released its Big Data Insights and Opportunities report<sup>4</sup>, based upon an online survey of 402 business and IT professionals.

---

<sup>3</sup> <http://newvantage.com/>

<sup>4</sup> <https://www.comptia.org/resources/big-data-insights-and-opportunities-2015?cid=download>

NVP found that 62.5% of firms have at least one instance of big data in production, up from 31.4% in 2013 and 48.2% in 2014. Only 5.4% of firms reported that they had no big data initiatives planned or underway. 51% of the companies participating in the CompTIA survey said that they have some form of big data initiative in place, up from 42% in 2013.

Basically, any survey done in the past 5 years shows the growth of big data. While we've made the case for big data for years, now it's here. Those who don't have big data programs cite the complexity and the cost of the technology as a major hindrance. This includes the addition or the expansion of existing data integration practices and technology.

A few conclusions can be drawn here.

- Big data works, and can deliver the value. 69.6% of firms view big data as very important or critical to their business success, up from 54.5% in 2014 (NVP). 72% of companies that have launched some form of big data initiative say that their results have exceeded expectations (CompTIA).
- However, data integration is still lacking. Firms continue to report that variety is the primary technical driver behind big data investments (40.0%), with volume (14.5%) and velocity (3.6%) lagging well behind (NVP). 45% of companies say that a high degree of their data is fragmented, and another 42% say their data fragmentation is moderate (CompTIA).
- Use of real-time data is lacking. 42% of companies don't have real-time analytics skills, and 41% lack relational database skills. As a result of poor data management and usage, 29% suffer from inefficient decision-making and inability to reach new customer segments, while 38% suffer from wasted time (CompTIA).

## 4. The Rise of the Data Lake

Data lakes (also called big data lakes) can meet growing data challenges and provide services to your business around the use of data. By collecting a wide variety of data sets relevant to the business, all in one place, and enabling multi-talented analytics based upon big data approaches, this become a source of data analytics that's both easy-to-scale and adaptable.

It's adaptable because the schemas do not need to be bound to the data; schemas can be bound at the time of the data result set extraction. The total potential value of a data lake grows with the amount of useful data it holds that is available for analysis.

So, how do you define the value of a data lake, and what does this have to do with data integration? Pretty much everything. The value of the data lake is its ability to bring together diverse data. Stored data, unstructured or structured, does not mean data integrated. Thus, data lakes are highly dependent upon good data integration technology.

## 5. Applications Moving to the Cloud

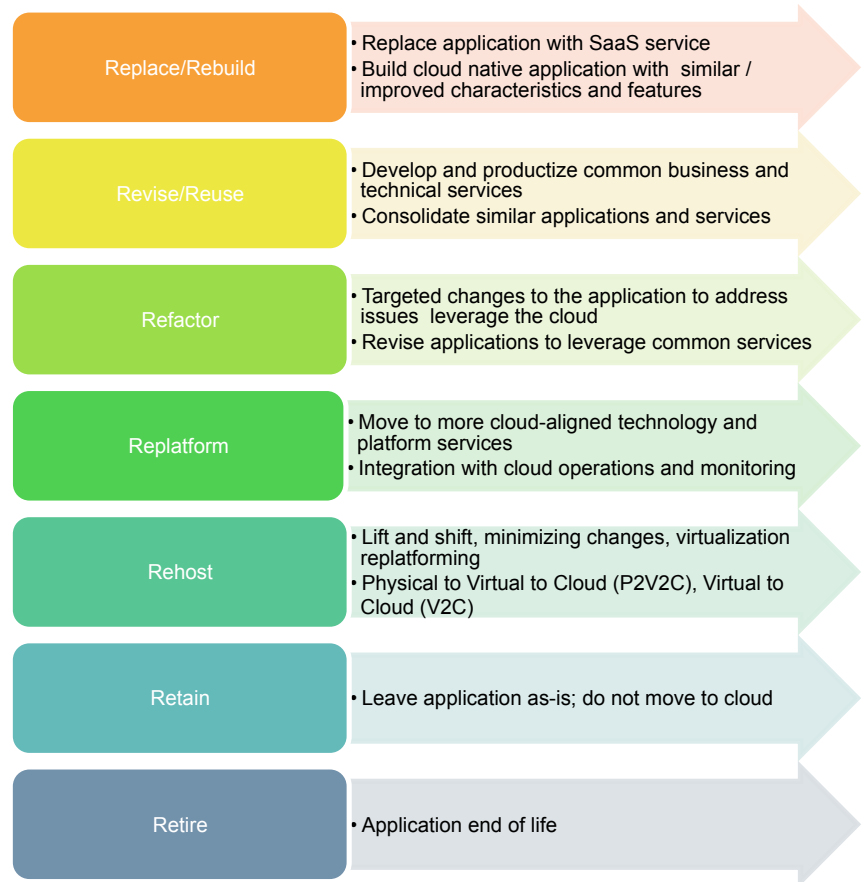
A few years ago, only 10 to 20 percent of enterprises had a cloud migration strategy. Nowadays, nearly all do. There are many good reasons to move to the cloud, but for many enterprises, the path to cloud is not paved with gold, nor is it easy to navigate.

Cloud-based application migration is often so complex that many enterprises leave out some critical processes and technology, such as data integration. This comes back to bite them when they finally understand what they have done. By migrating applications and data to private or public clouds, they create new silos of data that must be integrated.

However, the problem remains that legacy data integration technology is not up for the job. Enterprises often resort to ad-hoc data integration practices, such as simple FTP file transfers, or even shipping data via FedEx directly to the public cloud provider.

When migrating existing applications to the cloud, enterprises typically have hundreds or thousands of applications from which to choose for migration. When making these decisions, there are a few basic rules to follow:

- It is technically impractical to relocate certain applications and data to the cloud, since they are based on traditional technology. This would include moving old COBOL systems.
- Many applications and data are not cost justifiable due to the amount of changes that need to occur within the application that would allow them to be hosted in the cloud. This is the case with new or dated applications.
- Applications and data should be placed in priority order, from those that will provide the most value to the business when migrated, to those that will provide the least amount of value.
- Applications and data should be analyzed as to the amount of work needed to meet the requirements,



**Figure 1:** When migrating applications and data to the cloud, there are a variety of migration choices.



including a direct port (lift-and-shift<sup>5</sup>), to different degrees of refactoring the applications.

Figure 1 depicts the ways that you can move applications to the cloud, from replacing them with SaaS-based analogs (such as Salesforce.com), to refactoring them (which means rebuilding them for the target cloud platform), to retiring them, or making a choice not to move the application or data.

Having said all that, it's also clear that more than 50 percent of application workloads, and data bound to those applications, will be in public or private clouds during the next 5 years. The data integration problem that most enterprises now face is huge. Consider the distribution of the data between private and public clouds. This alone will drive most of the interest in updating enterprise data integration strategies and technology stacks.

## 6. Data Moving to the Cloud

Just as with the growth of cloud computing, cloud computing and data is also on the same trajectory. This is a bit different than application workloads, where the data is bound to the application. Data moving to the cloud means that the data may be decoupled from the application, or the data may exist on its own, or the data may be a combination of data types, such as data used in a data lake (see above).

What is distinctive about data in the cloud is that it, unto itself, may serve both on-premises and cloud-based applications, or data users (such as analytics). Moreover, the types of database technology may vary greatly, meaning non-structured, to object, relational, columnar, etc.. Data in the cloud is typically diverse, which means data integration intra-cloud is a critical need. This is the ability to integrate data inter-cloud (cloud to cloud), but, most important, it's the ability to integrate the data between the public cloud and on-premises systems.

In 2015 we seemed to move to a point where data that exists in the cloud is actually safer than data that exists on-premises, thus the green light is now on for migration. Public cloud-based databases, such as AWS's Redshift, are now cheaper and a higher-performing alternative to traditional database technology, but there is more to migrating the data that needs to be addressed.

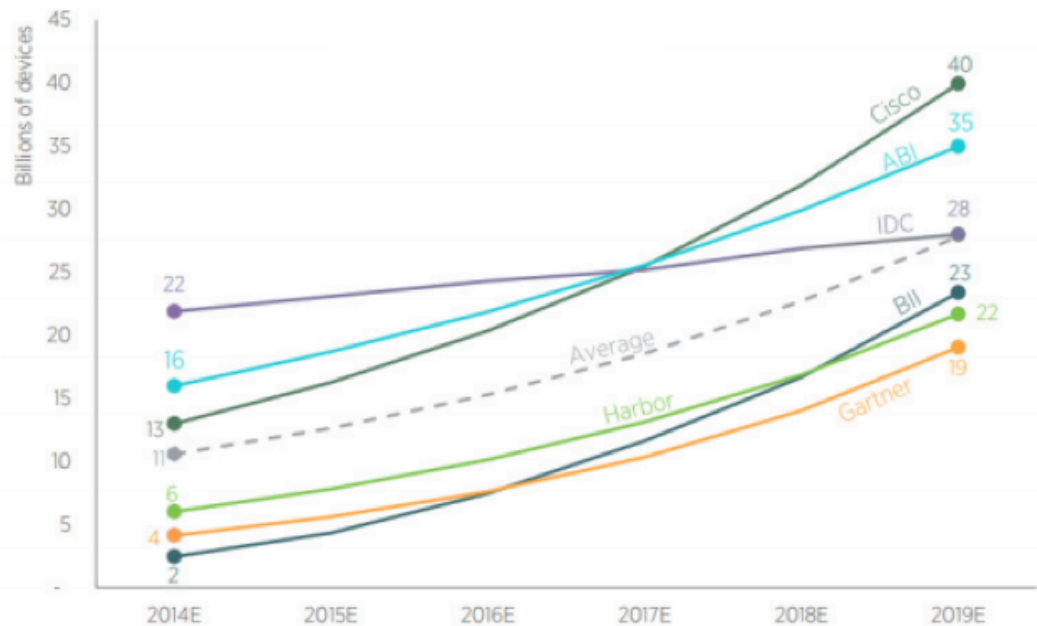
Aside from security and governance is the need for a sound data integration and data integration technology strategy, created along with the data migration and re-hosting plan. Again, the same issues with application migration to the cloud, meaning that traditional data integration tools (legacy) can't provide the same flexibility as more modern data integration tool sets. You need to rethink your data integration approaches and technology at the same time you move your data.

## 7. Rise of the Internet of Things (IoT)

Figure 2 shows the rapid growth of the IoT from both analyst firms and technology providers. The average is 28 billion devices by 2019 that will be in place and spinning off data.

---

<sup>5</sup> Lift-and-shift means moving the applications from traditional platforms to the cloud, with little or no modifications to the code or the data.



Source: John Greenough, "The Internet of Things is Rising: How the IoT Market Will Grow Across Sectors," *Business Insider Intelligence*, October 8, 2014. Produced by Adam Thierer and Andrea Castillo, Mercatus Center at George Mason University, 2015.

Figure 2: The growth of IoT has led to a new need to deal with new sources of data. Most enterprises are ill-prepared.

IoT, data management, and data integration are tightly coupled. Indeed, you can't have an IoT strategy and technology set without having a sound data management and data integration strategy. There are a few things to consider when you formulate this strategy:

- IoT will be systemic by 2020. Consider the cost of the devices and the value that IoT can bring. Take the growth of wearable fitness devices, for instance. These are spinning off megabytes of data a day from device owners, and that data must be transported, analyzed, and centrally maintained.
- IoT is all about the data. Devices themselves do nothing but gather and/or generate the data. It's the use of the data that brings the real value to IoT. The potential for this data, either atomically or aggregated, is virtually limitless.
- IoT is all about delivery of real time data. IoT instantly processes data coming from devices or sensors, and that allows desired outcomes to occur. No matter if it's reducing the RPMs of a jet engine that's overheating so it does not catch on fire, or dynamically adjusting irrigation based upon water saturation readings, it's about the real time delivery of data and the ability to act upon the data. Thus, data integration is key.

## 8. Rise of Microservices

Microservices are an architecture as well as a mechanism. Enterprises looking for a next generation approach to service oriented architecture (SOA) have become interested in them, as well as the use of microservices within containers.

Microservices are an architectural pattern in which complex applications are composed of small, independent processes that communicate with each other using language-agnostic APIs (services). This means that it's service-oriented computing, which decomposes the application down to the functional primitive, and builds itself as sets of services that can be leveraged by other applications or the application itself.

The benefits of this microservices approach includes efficiencies through reuse of microservices. As we rebuild applications, we modify the applications to expose services that are accessible by other applications. More importantly, we can consume services from the rebuilt application, so we don't have to build functionality from scratch.<sup>6</sup>

So, what are the benefits?

By breaking down functionality into sets of microservices and abstracting it, development teams can focus on updating just the relevant pieces of an application or system. We can place volatility into a single domain, and then abstract things out that most often change from things that don't. This means that data integration needs to be a key component, considering that the services will be widely distributed, and data is a core component of the services.

The proper approach to microservices can result in far more efficient use of code and underlying infrastructure, such as storage and compute. Users already report cost savings—in some cases, reducing the amount of infrastructure required to run a given application by 50%.

The separation of functionality across services should result in no single point of failure, and thus the systems built using microservices will be more efficient. The systems perform better, with very limited downtime, and have the ability to scale to the needs of the business.

There are a few core needs for data integration with microservices, including:

- To leverage distributed microservices, the data needs to be brought together around this distribution.
- The data is tightly bound to the microservices, and thus services interfaces may exist to extract the data and change it for delivery to the target.
- Centralized tracking of very complex data sources is a requirement to keep microservices working and playing well together. Thus, a robust data integration layer is needed.

## Call to action

As we can see from the data presented in this paper, there are a significant number of opportunities within enterprises today that drive new uses for data. While cloud and IoT drive advances in this new technology, it's also driven by a need to strategically leverage data.

---

<sup>6</sup> <http://techbeacon.com/containers-microservices-how-modernize-legacy-applications>

Data integration is becoming a competitive advantage that is directly tied to the success of the business. Those that can't practice data integration correctly and rely on legacy technologies will likely fall by the wayside within their industry. The ROI from data integration is huge, considering the upside to the business.

Consider the importance of data, and new systems built on the use of data, and thus data integration technology. The strategic use of data integration technology should rise above other IT priorities. This means data integration projects should be the first ones to be funded, not the last. Enterprises can realize significant cost savings with subsequent application development and big data projects, that is, if the data integration infrastructure is done properly.

Most legacy data integration technology is outdated, and can't keep up with the emerging patterns of data use. The old technology needs to be replaced as soon as possible with data integration technology that can support the "8 reasons" presented in this paper. Data integration technology updates should be IT's top priority for the next few years.

## About the Author

David S. Linthicum is an internationally recognized industry expert and thought leader in the world of cloud computing and the author or co-author of 15 books on computing, including the bestselling Enterprise Application Integration, and his latest book, Cloud Computing and SOA Convergence. He is a blogger for InfoWorld, Intelligent Enterprise, eBizq.net, and Forbes, and he conducts his own podcast, the Cloud Computing Podcast. His industry experience includes tenure as the CTO and CEO of several successful software companies, and upper-level management positions in Fortune 100 companies. In addition, Linthicum was an associate professor of computer science for eight years and continues to lecture at major technical colleges and universities.

## About SnapLogic

SnapLogic is the industry's first unified data and application integration platform as a service (iPaaS). The SnapLogic Elastic Integration Platform enables enterprises to connect to any source, at any speed, anywhere — whether on premises, in the cloud or in hybrid environments. The easy-to-use platform empowers self-service integrators, eliminates information silos, and provides a smooth onramp to big data. Founded by data industry veteran Gaurav Dhillon and backed by investors including Andreessen Horowitz, Ignition Partners, Microsoft Corp., and Silver Lake Waterman, SnapLogic is helping companies across the Global 2000 to connect faster. Learn more at [www.snaplogic.com](http://www.snaplogic.com).

