The Integrator's Dilemma

Rethinking application and data integration in the era of Big Data, Cloud Computing, Social Media and Mobile Computing

WHITEPAPER

snapLogic

Introduction

"If you only do what worked in the past, you will wake up one day and find that you've been passed by."
- The Innovator's Dilemma, Clayton Christensen

Businesses are increasingly relying on data to devise new go-to-market strategies, cultivate and attract new audiences and streamline operations. Medicine and research are relying on crunching large data sets to simulate clinical trials. Data has been referred to as "the new oil" and with new sources, types and locations, it has never been more abundant in the enterprise. IDC reports that from 2005 to 2020, the digital universe will grow by a factor of 300, from 130 exabytes to 40,000 exabytes, or 40 trillion gigabytes (more than 5,200 gigabytes for every man, woman, and child in 2020).[1]

Not surprisingly, the need to integrate and harvest this data to ensure it is trustworthy and ready for the analytics that promise such game-changing insights and impact has never been more critical. With the onslaught of unstructured and semi-structured data resulting from connected objects and devices, a network that's being referred to as the "Internet of Things," the situation is only going to get worse. The challenge, however, is that the legacy set of tools, processes and organizing principles that enterprise IT organizations have relied upon to connect and manage disparate data sources were not designed to keep up with this data deluge.

This is what we call the "Integrator's Dilemma." This paper will examine the common challenges faced in today's multi-cloud hybrid enterprise IT environment, and recommend solutions to address them.

1 The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

Rethinking Application and Data Integration

Drawing parallels with the hugely popular book "The Innovator's Dilemma," the Integrator's Dilemma is faced by enterprise IT organizations as their legacy integration technologies[2] struggle to keep up in the new world of big data, cloud computing, social media and mobile computing. The following section explains both the promise and the dilemma that these trends pose for traditional integration technologies.

1. Moving Applications and Platforms into the Cloud

Legacy on-premises application suites, such as ERP, CRM, Supply Chain, etc., are being replaced by hundreds of point software-as-a-service (SaaS) applications that focus on a small subset of functionality. While there is great value in moving to best-of-breed cloud-based applications, there are also challenges:

> **More Endpoints:** The explosion of SaaS applications means you need to handle an order of magnitude larger number of endpoints - both in the cloud and on-premises. A toolset that is geared towards multi-point orchestration and data integration is required and the ability to address both cloud-to-cloud and cloud-to-ground use cases is essential.
>
> **Cloud Service Level Agreements:** Accessing your information that resides outside the firewall while traversing the public internet means that security and reliable access concerns are top of mind as the burden is shifted to the cloud service provider. Integration technologies must also be able to take advantage of modern data formats and protocols like JSON and REST.

Cloud applications are extremely dynamic, with multiple releases per year and constantly changing fields, objects and application programming interfaces (APIs). As Ventana Research notes, "cloud-based data integration often involves many data sources, with new sources being added frequently. This integration challenge is likely to intensify."[3] Legacy middleware technologies were conceptualized and built in the 1990s and were not designed to handle an IT ecosystem that expands beyond the traditional confines of the firewall.

2. The Big Data Management Challenge

Legacy EAI or ESB technologies were built for handling individual business events such as newly created purchase orders, but are not designed for large data volumes. On the other hand, legacy ETL technologies were designed to move large batches of data, but only structured information - rows and columns. Both of these technologies are inadequate when it comes to handling the petabyte scale big data volumes, not to mention the variety and velocity of these unstructured data sets.

3. Sentiment via Social Channels

Social media has become a primary source of information consumption. Unfortunately for traditional integration technologies, social data is unstructured data and the majority of enterprise IT data stores are relational databases. For example, a structured table made up of a rows and columns is not flexible enough to handle free-form text and images that make up data created on social channels such as Facebook and LinkedIn. Traditional integration technologies were either purpose-built for

2 Examples include enterprise application integration (EAI), enterprise service bus (ESB) and extraction transformation and loading (ETL) tools

3 Ventana Research: Selecting a Platform for Cloud Integration

moving large data sets in scheduled batches (ETL) or moving small volumes of data in real-time (EAI). Neither technology is suitable for this new world of unstructured information.

4. The Impact of the Mobile Revolution on Your Data

In the new world of bring your own device (BYOD), there is a growing need to harvest machine data generated by these devices so that technologies such as Splunk can be used to process and detect anomalous behavior. This analysis needs to happen in near real-time so your IT organization can immediately respond to and mitigate emerging breaches and threats. Most IT organizations are accustomed to moving such large data sets in batches on a nightly or periodic basis. However, in order to deliver such data in near real-time, the integration technology must also support streaming. Most legacy integration technologies were not built to deliver on this new requirement of streaming machine data in real-time.

The Integrator's Solution

Enterprise IT leaders must embrace new tools and techniques to address the Integrator's Dilemma that Big Data, cloud computing, social media, and mobile computing represent. The old debate between application versus process versus data integration is clearly no longer relevant. All of the above are required. But trying to solve this dilemma with traditional tools and approaches won't allow you to meet the need for flexibility, agility and responsiveness that your business requires to ensure data is an asset instead of a liability. The Integrator's Solution consists of the following three strategies:

1. Avoid a Rigid Integration Layer

SaaS application endpoints are designed to scale on demand. It does not make sense to rely on an integration layer that is rigid or of limited scale. A rigid integration layer quickly becomes the weakest link in the entire IT infrastructure chain. Social media channels such as Twitter can generate unexpected spikes in traffic, quickly turning a low volume data traffic into an ETL-style workload. An elastic integration layer can handle such data traffic spikes seamlessly and cost-effectively. Elasticity becomes a key requirement where the lines between EAI-style workloads and ETL-style workloads are fast blurring.

Choose a platform that delivers comprehensive pre-built connectivity to SaaS and on-premises databases and applications. With the right platform that can orchestrate across a multitude of these systems. And with APIs fast becoming the defacto method of accessing information, the cloud integration platform needs to not only handle APIs natively, but also needs to protect you from the frequent API versioning changes made by the application provider.

2. Focus on Data Streaming

With today's unprecedented speed of business, freshness of information is critical. With traditional methods of business intelligence, data gets refreshed into a data warehouse on a nightly basis. The expectation today is that relevant information will be streamed from one or many sources when an event

occurs. Inconsistencies across streams must be resolved while this data is in-flight. An integration platform that can stream data in real-time from APIs, applications, and other data sources can be used to build an architecture that delivers the freshest information to business users.

It is worth mentioning that there will always be use cases where you need to move petabyte scale structured data in small time windows. ETL or ELT technologies will still remain suitable for such solutions. However, as Hadoop gains mainstream adoption, fewer and fewer use cases fit in this standalone category. IT organizations need a fast, multi-point and modern cloud integration platform that can handle both.

3. Empower The Citizen Developer

IT professionals are often overwhelmed by ad hoc, one-off data integration requests.
A cloud integration platform must empower the "citizen developers" within the business units with self-service to access the freshest information, while ensuring central IT can retain governance and control. The cloud integration platform must be simple yet powerful. It must abstract out the technical complexities such as data models and proprietary protocols that are typically associated with endpoint applications. Recognizing that there are repeatable integration patterns between systems, look for pre-built application and data flow templates that can be configured, shared and reused.

Next Steps

To respond to the new hybrid reality, look at your next-generation data management challenges with a fresh perspective. Replacing existing infrastructure overnight can be an expensive task that is fraught with risk. The ideal approach is to attack the Integrator's Dilemma head on and embrace the Integrator's Solution. In practical terms, this means getting your arms around the number of SaaS applications already being deployed in the enterprise. Identify use cases where your organization can benefit from streaming data as opposed to moving it on a periodic basis and start a pilot project to tackle cloud integration with a faster, multi-point and more modern technology.

To learn more about the SnapLogic Integration Cloud and how we solve the Integrator's Dilemma, visit www.SnapLogic.com.

Integrator's Dilemma
2013

Headquarters

2 West Fifth Avenue
Fourth Floor
San Mateo, California, 94402
United States

email: info@snaplogic.com
tel: +1-888-494-1570