

E-Guide

# BI in the Cloud: Overcoming Obstacles in Data Integration

*An essential guide to help you transition into the cloud*

## Contents

Stampede to cloud  
presents data  
integration problems

---

BI in the cloud: Find  
your bliss

*Faster project deployment, zero dollars invested in hardware and less IT staff to hire are among the reasons the cloud is a huge home run for companies. But you can strike out if you move your BI into the cloud without developing a data integration strategy.*

*Learn from Geezeo Inc.'s data integration experience with using Amazon Web Services and find out what items industry expert Wayne Eckerson says should be on your checklist before jumping into a BI cloud project.*

---

### Stampede to the cloud presents data integration problems

By: Alan Earls

The rapid adoption of the cloud by many organizations, for a growing array of use cases, is perhaps the biggest change in IT in a generation. Not surprisingly, that transition comes with plenty of challenges.

Among the largest: data integration problems. Determining how to allow multiple application programs to share data in cloud-based networks, either directly or through third-party software, has become increasingly burdensome.

For instance, James Elwood, CTO of Geezeo Inc., a Tolland, Conn.-based software services provider for banks and credit unions, ran into difficulties when his company started using cloud networking with Amazon Web Services in 2009. "We needed to build a quasi-tunnel or switching tunnel architecture between our applications and customer data centers," he said.

---

## **Contents**

**Stampede to cloud  
presents data  
integration problems**

---

**BI in the cloud: Find  
your bliss**

The wild card was that Geezeo was purely in Amazon Elastic Compute Cloud (EC2). "Without the infrastructure, being able to set up multiple [Internet Protocol Security (IPsec)] tunnels to various data centers and the ability to keep them separate cleanly is a bit of a technical challenge, as well as an ongoing management challenge," Elwood said.

Geezeo looked into the costs of buying a traditional data center, installing physical Cisco networking devices and creating a way to bridge its infrastructure cleanly over to Amazon EC2. Then the company spent two months cobbling together solutions that involved 15 hours of employee time each week for monitoring and maintenance -- which company officials considered too costly.

Geezeo eventually ended up finding a non-traditional data center product: VNS3 from CohesiveFT. VNS3 directs and manages traffic with the insight and control Geezeo needed, Elwood said, adding that if IPsec tunnels go down, it means the company isn't gathering transactional data. Adding VNS3 Manager allowed the core team to "manage all of their connections from a single control point and offer reliable, manageable and secure tunnels," he added.

### **Challenges and opportunities**

In the past, integration architecture approaches such as the enterprise service bus and message brokering services managed to pull together disparate systems for various integration scenarios, said Jiten Patil, principal cloud expert and technology consultant for Persistent Systems Ltd., a global

---

## **Contents**

**Stampede to cloud  
presents data  
integration problems**

---

**BI in the cloud: Find  
your bliss**

technology services company. However, in the evolving cloud era, cloud integration services are responsible for making devices, cloud applications, on-premises systems, hybrid and social solutions work seamlessly to deliver heterogeneous business use cases. "Cloud integration is replacing on-premise integration middleware and moving integration scenarios to cloud-based multi-tenant and self-service [Software as a Service (SaaS)] applications," he said. Furthermore, he noted, by pre-provisioning integration services such as connectors to various endpoints, it is helping drive the emergence of integration Platform as a Service (iPaaS).

Joe McKendrick, analyst at McKendrick & Associates, points out that the idea of having many applications sharing one set of data has always been the holy grail of IT. It was the idea behind relational databases when they first came to market in the early 1980s, namely, to have one repository of data which many applications can access. "Relational databases were partially successful in accomplishing this ... because there are two issues that keep getting in the way: organizational silos and data variety," said McKendrick. These challenges will also vex cloud-based data integration unless the organization makes a strong commitment to an enterprise data architecture.

Solving data integration problems requires that the business be put in charge of the data, McKendrick said. "The first step in any cloud process is going to be deciding which data sets will be going to the cloud, and which will remain within their current locations," he noted. In addition, it's important -- probably doubly important in the cloud -- to assure the data's trustworthiness. "Trust is the foundation of data management and analytics. Everyone needs to have complete trust in the information being captured and presented," he said.

---

## **Contents**

**Stampede to cloud  
presents data  
integration problems**

---

**BI in the cloud: Find  
your bliss**

McKendrick also pitches the idea that a move to the cloud means it's a good time to consider a move to master data management (MDM). "MDM is essential to any data integration project, as it establishes a single master 'gold copy' of data, versus separate, siloed data sets," he said.

Similarly, according to McKendrick, a move to the cloud also means a natural move to Data as a Service.

The advantages of cloud integration are twofold, according to McKendrick. First, business end users are able to get at information they need on a moment's notice. Second, IT and data management staffs will see a boost in their own productivity, because they aren't consumed in writing scripts or code to achieve manual integration.

"You don't have to tear up your existing data center and databases to achieve greater integration," McKendrick emphasized. What is needed, however, is an architectural approach to integration, one that addresses requirements across the enterprise. "The practice to date has been 'one-off'-type data integration efforts, done with manual scripting, on a project-by-project basis," he said. The problem with that approach: It doesn't allow for predicting who is going to need which data source or predetermining where the next data source will come from. "Instead, we need to make it easy -- in almost a self-service way -- for decision-makers to identify the data source they need and be able to bring it into the decision-making environment through a well-architected flow, with no need to go through IT and set up reporting or dashboards," he said.

## **Contents**

**Stampede to cloud  
presents data  
integration problems**

---

**BI in the cloud: Find  
your bliss**

McKendrick says companies that are putting their data out with third-party cloud vendors need to take a hard look at the contractual terms. Scrutiny needs to go beyond just security from hackers. For example, what happens to your data after your contract is terminated? How long does the vendor hang on to it? What about an outage or data loss? "Most vendors will state right in their contract they're not liable for lost data. You need to make sure you have either a secondary cloud or an on-premises backup site," he said.

"It's all about architecture," McKendrick continued, adding that companies want to avoid what he called "JBOD -- just a bunch of data" in an architecture sprawled across their organizations. Instead, "you want a flexible architecture that accommodates any and all changes, such as new data sources being added, older ones being removed, or interfaces changing," he said.

Deploying enterprise applications in the cloud, whether moving a legacy application or launching a "greenfield" project, requires understanding how the cloud environment differs from the physical co-located or hosted state, said Michael Higgins, manager of enterprise solutions architecture at CloudSigma, an Infrastructure as a Service (IaaS) provider.

Companies must be willing to review architectures that were created for hardware solutions, Higgins said. "The cloud has a rich set of features to offer that are not available in a typical hardware environment," he noted. Companies should understand the differences between "burst" and long-term commitment billing models and add cost control as a key metric in project planning. "Solutions that can accept single, simple failures or which can self-

heal -- roll back, fail over, 'round-robin,' etc. -- work amazingly well in the cloud," he added.

## **Contents**

**Stampede to cloud  
presents data  
integration problems**

---

**BI in the cloud: Find  
your bliss**

---

### **BI in the cloud: Find your bliss**

**By: Wayne Eckerson**

Hosting business intelligence off-premises might seem scary, but soon it will be a necessary and normal part of corporate life. For the biggest benefit, learn the lay of the land.

Companies have embraced the public cloud for many operational applications, but they've been less than eager to put business intelligence systems in the cloud. Why?

BI holds the keys to corporate data, and for security and compliance reasons many companies don't want that information to pass beyond the firewall. Plus, robust BI applications require custom development, which undermines one of the main benefits of using the cloud -- fast deployment.

Nonetheless, BI managers will have to face the cloud. There are simply too many advantages: Aside from deploying quickly, there's no hardware to buy or IT people to hire, and customers pay a subscription based on usage instead of a licensing fee.

So how do you begin? If you're a startup, the answer is simple: Go all-in with BI in the cloud. Most likely, a majority of your operational applications will

## Contents

Stampede to cloud  
presents data  
integration problems

---

BI in the cloud: Find  
your bliss

also be cloud-based. And, if all your data is already in the cloud, just run your BI environment there too. If you already run your BI environment on-premises, however, you need to be more tactical.

Say you have a business unit that has data it doesn't need to share with other departments. That would make a good test bed for cloud BI -- for example, a bank implementing a cloud-based analytics platform for its wealth management division.

Or perhaps you have business analysts who need to explore data in the enterprise data warehouse along with data outside the corporate firewalls. You could use the cloud to create a data sandbox, which could replicate the information in the data warehouse and allow users to upload personal data sets and connect to other cloud-based or third-party data.

Before jumping into a cloud BI project, there are several things you should do:

**Evaluate your data security and privacy requirements.** If your company has strict regulations regarding data privacy and security, BI in the cloud may not be a good fit. But don't rule it out. You might be able to use the cloud for prototypes, test beds or temporary applications with nonsensitive data. Or, you can request that your cloud provider put your data on a separate server in its data center.

**Know your sources.** Your cloud BI system will need to hook into your source systems, and requisite data integration programs will have to be built



## Contents

Stampede to cloud  
presents data  
integration problems

---

BI in the cloud: Find  
your bliss

to move, clean, integrate and aggregate information. Most cloud BI tools offer connectors to the major packaged business applications. If you're lucky, you can reuse custom extractors built for your data warehouse. You can also simply create a custom data mart from your data warehouse and use that as your source.

**Scrutinize data volumes.** If your cloud BI system needs to ingest large quantities of data every day, you'll need to implement a trickle-feed architecture that captures changes in source systems and routes them into the cloud BI environment in near real time. Make sure your cloud BI provider can implement the appropriate hooks or agents in your on-site systems and that it supports change data capture and trickle feeds.

**Estimate costs.** Once you know projected data volumes, you're ready to estimate ongoing operational costs. Fortunately, the costs of storing data in the cloud have dropped precipitously, thanks to Amazon, whose new Redshift data warehousing service charges about \$1,000 a year to store a terabyte of data. This will exert significant downward pressure on cloud BI fees, making the cloud even more attractive.

**Know your users and their needs.** Typical business users need reports and dashboards that access a fully dimensionalized data model and support a range of queries and drill-downs in a guided environment. Power users need an ad hoc environment that lets them upload and merge data and create their own reports and dashboards. Before getting started, understand which types of users your cloud BI provider is best suited to support.

## **Contents**

**Stampede to cloud  
presents data  
integration problems**

---

**BI in the cloud: Find  
your bliss**

**Assess your BI skills.** If you have people with BI or data warehousing skills on staff, choose a Platform as a Service offering that enables you to build your own BI environment. If not, you'll have to get a vendor to build and operate a cloud BI environment for you. If you support only power users, pick a Software as a Service offering that lets those savvy users upload their own data to the cloud to conduct analyses and build ad hoc reports and dashboards for colleagues.

Analytics applications have always closely trailed their operational brethren. Today, most new operational applications are designed to run in the cloud. Thus, it stands to reason that most new analytics applications will also run in the cloud. The ease, convenience and lack of capital outlays will soon make BI in the cloud inevitable. The time to start evaluating BI in the cloud is now.



## Contents

[Stampede to cloud  
presents data  
integration problems](#)

---

[BI in the cloud: Find  
your bliss](#)

## Free resources for technology professionals

TechTarget publishes targeted technology media that address your need for information and resources for researching products, developing strategy and making cost-effective purchase decisions. Our network of technology-specific Web sites gives you access to industry experts, independent content and analysis and the Web's largest library of vendor-provided white papers, webcasts, podcasts, videos, virtual trade shows, research reports and more—drawing on the rich R&D resources of technology providers to address market trends, challenges and solutions. Our live events and virtual seminars give you access to vendor neutral, expert commentary and advice on the issues and challenges you face daily. Our social community IT Knowledge Exchange allows you to share real world information in real time with peers and experts.

## What makes TechTarget unique?

TechTarget is squarely focused on the enterprise IT space. Our team of editors and network of industry experts provide the richest, most relevant content to IT professionals and management. We leverage the immediacy of the Web, the networking and face-to-face opportunities of events and virtual events, and the ability to interact with peers—all to create compelling and actionable information for enterprise IT professionals across all industries and markets.

## Related TechTarget Websites

