



October 2017

Moving Your Data Infrastructure to the Cloud – Things You Should Know

Claudia Imhoff, Ph.D.

Sponsored By:

WhereScape[®]

Table of Contents

Introduction	1
Challenges and Benefits of the Cloud	1
Dispelling the Myths	4
Automation is Mandatory for Cloud Implementations	5
Summary	6

Introduction

Let's start with a quote from 2010:

“There's no way that company exists in a year.”

Tom Siebel, Founder of Siebel CRM Systems, speaking about Salesforce.com.¹

Yes, we all laugh at this quote now but, truth be told, cloud computing has had a rocky road to general acceptance despite Salesforce.com's remarkable success. Today acceptance is now a given after Amazon, Google, Microsoft, IBM, and other well-known companies began to offer their version of cloud computing.

You may ask, “What is cloud computing”? The simple way of explaining it is from Whitfield Diffie: “Cloud computing means you are doing your computing on somebody else's computer.”²

The category of business technology that has heartily embraced cloud computing is Business Intelligence (BI) and analytics. The massive increase of both data volumes and the complexity of these environments have made the move to cloud implementations necessary for many companies. Certainly, there are clear benefits to this move but the cloud does not remove or eliminate the basic design principles so necessary to have a sustainable and maintainable analytics ecosystem.

This paper will discuss benefits and challenges in the move to cloud computing. It will also dispel several myths that have sprung up when the move involves business intelligence (BI) and analytics capabilities. It ends with a discussion of how data warehouse automation is a great boon to the team performing the migration.

Challenges and Benefits of the Cloud

Like any new initiative, there are both challenges and benefits. It is advisable to understand both of these when deciding whether cloud computing is suitable for your company's analytic environment. Let's start with the challenges:

¹ <https://www.forbes.com/forbes/2010/1206/features-technology-salesforce-marc-benioff-web.html>

² American cryptographer and computer scientist, winner of the Turing Award, https://en.wikipedia.org/wiki/Whitfield_Diffie

- IT governance and control - IT departments are still leery of letting go of their data. There are many reasons but certainly job loss and the concerns about security and privacy over data rank high on the list. IT is generally responsible for corporate data assets being implemented and used according to agreed-upon corporate policies and procedures. This means that service level agreements between the company's IT department and the cloud provider are critical to ensure acceptable standards, policies and procedures are upheld. The IT personnel may also want insight into how the data is obtained, stored, and accessed by its business personnel. Finally, it is recommended that IT determine whether these cloud-deployed assets are supporting your organization's strategy and business goals.
- Changes to IT workflows – IT workflows dealing with compliance and security become more complicated in hybrid environments (those consisting of both on-premises **and** cloud deployments). The workflows must take into consideration the need of advanced analysts and data scientists to combine data that is on-premises with data in various cloud computing sites. Keeping track of where the data resides can be quite difficult if good documentation and lineage reports are not available.
- Managing multiple cloud deployments – Often, companies have more than one cloud computing implementation; they may use a mix of both private and public deployments – maybe even multiple ones in each category. The company must determine if each cloud provider is in compliance with regulatory requirements. Also, when considering your cloud provider(s), determine how security breaches are prevented or detected. If data security concerns are great, it may make sense for the corporation to maintain highly sensitive data (like customer social security numbers, medical health records, etc.) within their premises rather than deploying them to cloud computing.
- Managing costs – the on-demand and scalable nature of cloud computing services can make it difficult to determine and predict all the associated costs. Different cloud computing companies have different cost plans. Some charge by volume of data stored, others by the number of active users, and others still by cluster size. Some have a mixture of all three. Be sure to watch out for hidden costs like requested customizations, database changes, etc.
- Performance – It is clear that if your provider is down, so are you. All you can do is wait for the provider to come back up. A second

concern is your internet bandwidth. A slow internet means slow connectivity.

Now let's turn to the many benefits of migrating to a cloud computing environment:

- Lowered operating costs – This is perhaps the first benefit that companies realize when considering a move to the cloud. There is a significant difference between capital expenses and operating expenses. Basically, you are “renting” the infrastructure rather than bearing the costs upfront of building your own environment. The cloud computing provider bears all the system and equipment costs, the costs of upgrades, new hardware and software, as well as the personnel and energy costs.
- No maintenance or upgrade hassles – These are again the headaches for the cloud computing provider. This frees up all resources to have a laser focus on obtaining, accessing, and using the data, not on managing the infrastructure.
- Ease of implementation – For most companies, purchasing a cloud computing environment is as easy as swiping your credit card. It takes only minutes to access the environment because the technological infrastructure is all ready to go. This must be differentiated from the data infrastructure that must also be established. Whether you implement a data lake, a data vault, or a data warehouse, design and development work must be performed in addition to the technological set up.
- Innovation from new cloud companies – Cloud technologies have been “born” from very innovative new companies. They make full use of all the advantages that the cloud has to offer. These technology companies can also add new features, functions, and capabilities, making them available to all customers immediately.
- Elastic scalability – Many customers say this is the most appealing attribute of cloud computing. You can quickly scale up **and down** based on real needs. There is no need to buy extra computing capacity “just in case” you may need it at a later date. Cloud data warehouses can increase or decrease storage, users, clusters with little or no disruption to the overall environment.
- Ability to handle the vast diversity of data available for analytics – Cloud computing providers can handle both well-structured data (like from operational systems) as well as the “unusual” data so popular today (like social media, IoT, or sensor data). Cloud implementations can support both fixed schemas and dynamic

ones, making it perfect for routine production analytics like Key Performance Indicators or financial analyses as well as unplanned, experimental, or exploratory analyses so popular with data scientists.

Dispelling the Myths

Unfortunately, with new technological advances comes the mistaken idea that they must also be the long sought-after “silver bullet” that makes all problems go away. These myths are discussed and dispelled here:

1. *Just throw all your data into the cloud and start analyzing it – no design or architecture is needed.* Nope – sorry, not going to happen. The data and access methods will not just magically be understood and usable. You will just make a data dump or data swamp in your cloud implementation. And that is simply a big waste of money, time and effort. An analytics environment is planned and architected so that all users can understand and use it. The manipulation of the data and its lineage must be documented; its components and data schemas must be known so the analytical personnel can easily use the environment.
2. *Just forklift all your data warehouse into the cloud – there is no need to redesign it.* Negative – just not so. Your multi-year-old data warehouse has grown some barnacles along the way or is in need of being updated with new requirements. This is your chance to blow the dust off, remove inefficient processes, wasted space from unused assets (old reports, visualizations, analyses no longer used), and excess workspace for users who no longer use the environment. This is a perfect opportunity to automate many processes to make them far more efficient. It is also a time to reassess the original requirements, perhaps bringing in new ones that are waiting in the wings.
3. *Just by changing to a cloud deployment, your implementers will be more productive.* Again no – migrating to the cloud will most likely change your entire process and methodology. Infrastructure gains do not necessarily equate to productivity gains if the team continues to develop and operate using the same outdated means, methods or processes. Productivity gains may actually be negative at first if moving to the cloud invalidates current methodologies and/or productivity tools. The team will need time to retool and learn the new methodologies and processes. Then their productivity will improve.

Automation is Mandatory for Cloud Implementations

Migrating to a cloud data warehouse can be a very successful endeavor for many organizations. One critical success factor ensuring a satisfactory conversion is the utilization of data warehouse automation technology. Here are a few of the many justifications you can use to warrant the usage of automation technology.

- The first thing to think about is that hybrid environments – those consisting of on-premises and cloud implementations – are much more complex than single location ones. The need for consistency of design, development standards and documented process controls across all environments is much greater. We know that analytics environments must be built iteratively – that is, each project is built upon the foundation of the previous projects, reusing the designs, standards and knowledge. These projects combine into a data warehouse program and to ensure consistency, the implementers must use the same standards and conventions for all projects. Automation technology employs the best practices and strengths from leading data warehouse methodologies, making it the best way to support necessary consistency and reliability across the program, regardless of the cloud infrastructure platform selected.
- Second, migration to the cloud can involve the movement of massive volumes of data. The team must ensure that all migration mechanisms preserve the structure and integrity of data. Automation again can rapidly guarantee that all structures follow documented guidelines. It also sets up the proper data quality and integrity processes in a repeatable and reusable way.
- A third rationalization is support for the ever-changing data and analytics requirements. One thing is guaranteed for implementers creating an analytical environment – it will change. Changes are a sign of healthy analytics usage but the implementation team must be equipped to handle the changes quickly without disrupting other analytic functions. Data warehouse changes require agility in terms of fast prototyping and multiple iteration support. Automation makes the implementers so much more efficient and effective, improving their ability to deliver reliable additions to the cloud data warehouse very quickly.
- The fourth justification for using data warehouse automation is the mandatory requirement for reliable, up-to-date documentation.

Because data warehouse projects are part of a program, documentation becomes critical to the overall maintenance and sustainability of the environment. Developers, data modelers, architects come and go in projects but understanding why they did what they did must remain. We all know that documentation and impact analysis capabilities help lower risk for future changes to the data warehouse environment. Unfortunately, documentation is the last thing most team members want to do – it is difficult, not very interesting to do, and gets out of sync very quickly. Removing the drudgery of creating documentation is certainly welcomed. That is the beauty of automation – documentation is automatically created and maintained, leading to better sustainability and maintainability of the future data warehouse configurations.

- The final justification must be the cost factor. Any technology that increases the productivity and efficiency of the implementation team results in reduced costs and lower risk for the overall implementation. Data warehouse automation results in a team that can turn on a dime and be far more innovative. The ability to fast-track migration and new cloud-based data infrastructure projects not only reduce implementation costs and risk, it ensures that companies are in a better position to reap the ongoing benefits the cloud provides sooner.

Summary

The migration of many data warehouse capabilities to cloud computing environments seems inevitable but it doesn't have to be onerous. Just remember – there is no silver bullet that replaces sound design and deployment practices. However, the complexity involved with these migrations can be greatly simplified by using data warehouse automation technology.

With automation, the entire analytics environment development becomes more consistent, reliable, and reusable. Workflows are more efficient and effective, leading to faster deployments and higher overall satisfaction rates from the business community.

Because documentation of the entire environment is painlessly created and updated, new additions as well as maintenance and sustainability of existing functionalities are easier, even in the face of constantly changing requirements. The team gains a sense of confidence and competence in the face of what could be a daunting effort.