September/2020

A Data-Driven Approach to Your Cloud Migration Journey

How to Make All Analytical Data Available

Claudia Imhoff, Ph.D.

Sponsored By:

teradata. Alation

Table of Contents

Introduction

The long sought-after objective for analytical environments is to make *all* data useful for analytics available to *all* people who need it. While life was simpler when we had a single, on-premises source of data for analytics (the enterprise data warehouse or EDW), it was also frustrating in terms of the latency inherent in getting data into the warehouse and prepared for analytics. Fast-forward to today where analytical data can be stored in multiple databases using various deployment methods (e.g., cloud, on-premises, or hybrid environments). The cloud option, in particular, has become an appealing one for data storage and analysis because of its many promised benefits like cost savings, performance, elastic storage, and so on.

To help determine if this paper is right for you, here are three assumptions we made about you, the reader.

1. You and your company have made the commitment to become a data-driven organization,
2. You and your company have decided or are already on a cloud migration journey, and
3. Your company has been affected by the COVID-19 pandemic and economic downturn.

The goal of a journey to the cloud is to make things simpler for the ultimate users of analytical data. However, making things simpler for our business communities generally means making things more complex on the IT backend. That is, there will be more complexity in the technical infrastructure and more difficulty for the people supporting that infrastructure. It is important to remember, though, that this is a journey comprised of multiple initiatives, not a single project.

This paper is a thought-leadership piece to help companies with their journeys to the cloud and advise those responsible for these migrations. We start with a discussion of the three strategies that all organizations must develop to understand how their data and cloud deployments will help the business reach its objectives. We then continue with a section on avoiding the pitfalls of this journey: by analyzing the impact of moving data to the cloud; by reducing costs; by understanding what data should or should not be moved (usage analysis); by implementing an active data catalog to improve the access and understanding of critical analytical assets; and by ensuring that business users are provided with a single view of the data, regardless of where it resides.

Three Strategies for Data-driven Organizations

Becoming a data-driven organization is not simple – it takes significant time, effort and resources. However, becoming data-driven in your decision-making processes is immensely beneficial to the entire organization, especially in these uncertain times. Sound decision-making is the key to survival for most enterprises. To ensure a successful data-driven transition, these three strategies should be first on your list of program activities.

- Business Strategy – COVID-19 and its impact on global economic conditions have affected many organizations and individuals greatly. Returning to "normal" will not happen for quite a while. This new reality means that most organizations' goals, objectives and strategies have been completely upended and must be reconsidered. A good place to begin your organization's data-driven initiative today is to review and reformulate the business strategy for its data usage. Let's start with a fundamental point. The business must understand that data is an asset or something of great value to the organization, especially in its quest to become data driven. In creating a new business strategy, consider data utilization in the context of business activities, new goals and updated objectives. For example, how will the organization use its data to further its business strategies in today's environment? Changes to all aspects of the business should be considered: entering new markets, producing new products, attracting new customers as well as retaining existing customers, and so on. Some of the best indicators of updated business objectives and goals are in changes to executives' compensation plans, feedback from executives (e.g., through interviews), and quarterly reports on the financial well-being and future plans of the organization. The new Business Strategy will drive the Data Strategy.
- Data Strategy – Once the current Business Strategy is understood, you can turn your attention to how data and decision-making capabilities will support these strategies. A Data Strategy document must be able to answer all sorts of questions, including those resulting from the environmental changes to your organization. Here is a list of some important questions to address:
    - What data is needed for decision-making?

- Where is it located?
- How is it being used today and/or how will it be used in the near future?
- How frequently is it used?
- What data, reports, analytics are not being used (and can potentially be "retired")?
- Who uses the data?
- What data is sensitive and how secure is it?
- What are the privacy policies that impact its usage and how will the company enforce these policies?
- What compliance and regulatory restrictions are on the data and its usage?
- How are compliance and regulatory rules supported?

You must understand what data you have, how it is being used today, and how it will be used in the future before you can move forward with cloud migrations and technological decisions. The Data Strategy will then drive the Cloud Strategy.

- Cloud Strategy – It is important to note that the Cloud Strategy is the last strategy to be created. After all, cloud deployments support data-driven initiatives, which, in turn, ensure the Business Strategy is achieved. Migrating some or all of your analytical data to the cloud is a significant activity, however. The process is much more than a simple fork-lift of on-premises data to a cloud-based platform. As such, it is the perfect opportunity to reassess the current Data Strategy and build/improve upon it. Start with the data storage designs – what can be moved and what must stay where it currently resides? What analytical capabilities are in use now and what new ones will be needed for the future? What technology is appropriate to support all business users and all forms of analyses? It is the time to create an environment that ensures that, regardless of location, all data is made accessible to all users for all analyses. It is also important to note that many companies choose to keep some applications on premises. Therefore, when creating the Cloud Strategy, you should determine how the business community will access different data and applications, as well as how the IT implementation teams will support a hybrid environment.

All strategy documents should be considered living documents. That is, they will change over time as new information is made available, different technologies are adopted, new environmental impacts occur, and the company evolves its objectives or goals. Be sure to review these strategies periodically to ensure they remain aligned with the organization's data-driven directions.

Avoid the Potholes in Migrating to the Cloud

As mentioned, migrating to a cloud environment is more than simply moving your data from on-premises to a cloud-based platform. If that is all you do, you have missed a great opportunity to improve your data management environment. Likewise, there is a high likelihood that the migration will disrupt or destroy processes, activities, or analyses that worked on-premises, which you need to prepare for and address.

There are at least four key activities that should be performed to improve the overall environment and, thereby make the migration easier and more efficacious. Let's discuss these in more detail.

- *Determine the technological impact of moving data to the cloud.* An impact analysis is critical to the overall success of this initiative. Given that business practices and activities are likely to change, there undoubtedly will be impacts to the way that data and analytics are made available to analysts, executives, operational personnel, and front-line workers.

  A good place to start is with a study of all data transformations. Are there any that need to be rewritten? Are there transformations that are no longer needed? Which transformations will work or not work in the new environment? What new data and, therefore, new data transformations are needed in the new environment?

  Next, the migration team should determine the impact of the migration on analytical and reporting processes and procedures. From complex data science "experiments and explorations" to the everyday creation of critical reports and dashboards, the impacts of the migration may be felt. Are there algorithms that won't work or are no longer needed? An examination of reports and dashboards will likely uncover many that are no longer used and may be discarded or need to be retooled for the cloud environment. New algorithms, reports, dashboards, etc., may also need to be

created.

Finally, there should be an impact analysis performed on the security and privacy policies in place. Are these policies in need of updating? If so, how well will the new environment uphold the new policies? Is the overall security of the data equal to or better than the on-premises version? How will the new environment detect security or privacy breaches? Will all new data, processes, activities and data access be monitored for misuse or potential breaches?

A sophisticated data catalog can accelerate impact analysis though automated extraction of data lineage from diverse sources like databases, BI reports, and data flows.

- *Determine the business community's usage analytics.* Important and critical information for the success of the migration efforts comes from actual usage analysis. Who is using what data for what types of activities – reporting, data science, dashboard analytics? The popularity of certain data can change the storage, access and actual design of this data. It can be used to assist search and cohort types of queries (e.g., people using this data also use these other data as data products)

  What data is not being used and is no longer needed? This information can change all kinds of activities from data transformation, data design, and data access, to ultimately, security and privacy policies themselves.

  Understanding new data usage patterns is also important information as well. This can affect design and access as well as identify other adjustments that must be made based on these patterns.

- *Use a data catalog to keep data accessible and accurate.* The ability to catalog not only what data is available but also all of the characteristics (metadata) about it aids in so many areas of a data-driven organization. These include, but certainly are not limited to, an improved understanding of what data is available, where it is located, how to gain access to it, and even its level of quality. A sophisticated data catalog can speed data discovery through an NLP interface. It can eliminate redundant creation of similar

analytical assets by finding and presenting them to a business user before that user attempts to recreate an already existing wheel. For example, a simple and easy search can quickly present KPIs, reports, certain visualizations, statistical or other advanced analytics to the business community eliminating the duplicate effort and wasted time in recreating these. The data catalog can also monitor all sorts of usage statistics, perhaps leading to the creation of an even more streamlined environment.

- *Provide a single view of the data.* The holy grail of any analytics environment is to have a modern technology platform that ensures the availability of all data regardless of where it resides in the enterprise. Virtualization or other technological connectivity mechanisms mean that data does not have to be moved to the cloud in order for a businessperson or data scientist to use it. Your technology should enable businesspeople to access and analyze data not only from *any* location but also in *any* technology seamlessly.

  The follow on to the any location in any technology is that the data must be discoverable in *any* format – graph, data lake or NOSQL storage, relational, and other formats. These are not easy functions for a platform to perform, so ensure that your cloud platform can handle this level of complexity easily and consistently.

  Finally, the front end of the environment, the access point for most businesspeople, must be simple and easy. Without this simplicity, the adoption of analytics in decision-making will be greatly hampered. The simpler it is though, the more complex the technological architecture must be. Making something look easy and simple means that there is a lot of complexity in the backend. Again, make sure your technological infrastructure can handle the complexity and that those responsible for implementing the migration are not overwhelmed.

Summary

A data-driven organization must rely on a solid foundation of data, analytical assets, and the technologies to support these. Add to this the need to migrate to a lower cost, more flexible, and modern cloud

platform. To simplify this migration process, you must develop three strategies based on where your organization finds itself in today's business and environmental milieu.

These strategies start with the Business Strategy. The organization must understand where it is in today's economic environment and where it needs to go in the near-term. The business strategy guides all other strategies, including the Data and Cloud Strategies.

The Data Strategy is next for the migration to the cloud. Understanding the organization's analytical data needs is key to satisfying the businesspeople who use that data to make informed and timely decisions.

Finally, the Cloud Strategy is formed based on the first two strategies. What goes into the cloud, how to satisfy the need to access all data, regardless of where it resides, and to support any and all technologies that use that data is an important consideration in choosing the best cloud platform for analytics.

A cloud migration starts with four activities that help to ensure a successful implementation. These include determining the technological impact of moving data to the cloud, determining the business community's usage analytics, using active data cataloging techniques to keep data accessible and accurate, and providing a single view of the data. A data catalog can gather this information with these key features – log processing, ingestion of metadata, machine learning capabilities, and a behavioral analysis engine.

With this in hand, the migration to a cloud platform guarantees a simpler, easier environment for usage by businesspeople and maintenance by the technical implementers responsible for it.