Six Strategies for Moving Enterprise Analytics to the Cloud

By Judith Hurwitz, Daniel Kirsch, and Fern Halper

Sponsored by: MicroStrategy
**FOREWORD**

**NUMBER ONE**
Think simplification for fast deployment

**NUMBER TWO**
Manage data placement based on governance

**NUMBER THREE**
Connect to multiple data sources from your analytics platform

**NUMBER FOUR**
Unlock insights from big data

**NUMBER FIVE**
Consider the ease of upgrade of the analytics platform

**NUMBER SIX**
Ensure that your platform is change ready

**CONCLUSION**
Take the leap

**ABOUT OUR SPONSOR**

**ABOUT THE AUTHORS**

**ABOUT TDWI RESEARCH**

**ABOUT TDWI CHECKLIST REPORTS**
In the era of analytics, more businesses are taking advantage of ever-increasing amounts of data to become more competitive. An emerging challenge for analytics has been the size and complexity of the data. The challenge is real; businesses are dealing with vast amounts of data across different departments, partners, and third-party data sources. As the amount of data explodes, organizations often discover that the costs of dramatically expanding their on-premises environment to support this scale of analytics are enormous. It is typically not feasible to simply add more servers to the data center.

Increasingly, business leaders are utilizing cloud services (including compute, storage, and analytics) to support their needs. For example, in a recent TDWI Best Practices Report about the cloud, over 90% of respondents reported either using or planning to use cloud services in the next several years. This trend of taking advantage of cloud services is important for all sizes of companies. In fact, large enterprises that are managing large volumes of data are increasingly taking advantage of a variety of cloud services including analytics services.

The cloud has numerous benefits.

- **Robust architecture:** The design and architecture of cloud services incorporates self-service, scalability, and integration. Cloud computing services are modular and configurable. Therefore, with well-architected APIs and cloud services, it is possible to apply these services to a variety of data in order to perform analytics without having to worry about scaling your physical infrastructure.

- **Scalability and flexibility:** When you need to perform analytics processing on a large data set, the cloud enables you to procure as much storage and compute services as necessary. When you are finished with the analysis, you can release these services so you are no longer paying for more capacity than you need. Thus, the ability to scale up or down on demand means your organization has the flexibility to respond to new initiatives at the right time and for the right price. It also provides the ability to experiment with analytics.

- **Agility:** Many organizations want to use cloud services to analyze data but then manage that data in their own private environments. Cloud computing provides the type of agility to support businesses that want to be able to react to the changing customer demand in an economical and practical manner. For instance, configuration automation and APIs can help organizations quickly launch many instances of analytics without manual configuration. Cloud services provide agility for analytics in that the platform is ready to go without weeks or months spent installing software and hardware.

This checklist report focuses on helping organizations build strategies for moving enterprise analytics to the cloud. It examines benefits to consider, issues to think about, and how the cloud can impact future plans.
Provisioning new hardware in a data center takes time, a large upfront investment, and expertise. It can take months to get servers ordered, up and running, and configured for analytics. The time, expense, and expertise needed to create an on-premises analytics environment are the primary reasons many organizations are turning to the cloud. Cloud analytics services that provide a self-service interface can enable you to provision the right amount of compute and storage service based on the job. If a project grows, cloud customers can easily scale up by accessing additional compute and storage resources. Likewise, during slower periods, customers can de-provision resources to reduce costs.

Self-service is one of the most important aspects of cloud computing because it streamlines the deployment experience. When you are tasked with analyzing large amounts of complex data, the last issue you want to deal with is the tedious process of procuring storage and compute capabilities in your data center. It is simply not practical to build your internal services based on the intermittent peaks of need.

When you are dealing with complex analytics, you might not know precisely how much compute or storage you will need. Using a public cloud environment, you can spin up an instance with all of the necessary services without having to worry about configurations. Based on the security and governance requirements, your organization can select a multitenant or a single-tenancy environment. Some organizations prefer a single tenant environment because they feel it can be more secure because the company has its own application instance along with the supporting hardware. If the workload is especially large, you can provision a clustered environment. If you are considering a clustered environment, make sure that your analytics platform supports high availability and redundancy required to support a cluster.

One of the most important benefits of the self-service environment is that users do not need to have deep technical knowledge to set up an analytics environment with the proper specifications. Self-service provisioning of a ready-to-go analytics platform in the cloud means that organizations do not have to spend additional time installing and configuring the environment. This means that it is less likely that someone will make a mistake that will cause performance problems and delays.

Placement of data is an important security and governance consideration for critical customer data. TDWI research shows that a majority of companies have an on-premises data warehouse. These same companies often have some public and private cloud services to support their data. In many cases, companies want to keep tight control of sensitive data, such as personally identifiable information (PII), because they consider that data private. In certain geographies companies must comply with laws that require personal data to be treated a specific way. For instance, the recently announced EU Privacy Shield imposes stronger obligations on U.S. companies to protect Europeans’ personal data. Emerging regulations such as the European Union’s General Data Protection Regulation (GDPR) dictate how and where customer data must be stored.

In reality, BI and analytics users will most likely need to analyze data that is both on premises and in the cloud. For instance, analysts may look to gain new insights into future customer buying patterns using an analytics platform in the cloud for the reasons cited above. They may wish to correlate different factors such as weather, influence of social media, and new innovative technology on future revenue. Some of this data might be external to the data warehouse and in the cloud (perhaps provided by a public cloud provider).

The rest of the data may be stored on premises in the warehouse. In this case, the organization might strip information to anonymize that data. In other cases, they may mask sensitive data in order to analyze that information together with the external data in a public cloud. Alternately, they might bring the third-party data into their data warehouse and move all of the data into a virtual private cloud where their analytics platform resides, and connect to the cloud analytics provider using a virtual private network (VPN) or a dedicated private connection.

One of the greatest benefits of the cloud is that it provides many options for both where data will be located for analytics and how the data will need to be governed. Sometimes organizations want to leave key data stores on premises because of important organizational imperatives. These organizations also want the flexibility to move that data into a cloud environment for analysis when conditions are right.

If you want to move sensitive data to the cloud, then make sure that the cloud vendor has the required set of cloud certifications to keep your data safe—even if you are temporarily moving it there. As your organization increases the use of cloud computing and becomes more mature, it will need to consider the impact on current governance policies and whether these may need to be updated.

As described above, data does not come from just one source. Organizations are more frequently using disparate data types from many sources. Some data might be generated internally while other data sources may be supplied by open-source data providers, third-party data aggregators, or social media companies. Data sources might be highly structured in relational databases; other data might be semi-structured such as data from sensors and system logs. Large amounts of data might be in managed in a Hadoop cluster. By leveraging multiple data sources from both your organization as well as third parties, you will be able to enrich your organization’s data and improve your insights. TDWI routinely sees that organizations using disparate data for analysis are more likely to measure tangible value than those that do not.

There are different techniques for integrating data sources. As mentioned in Number Two, some companies are using a dedicated connection or VPN. Other organizations may have data in disparate environments and are combining data into a logical data model for analysis. To satisfy requirements, companies use a variety of techniques to build unified views of data, depending on user requirements. For instance, some businesses implement data federation or virtualization software. This is often facilitated by using data preparation tools that can blend data from numerous sources using a drag-and-drop interface. This data is then analyzed live or in memory.

Whatever method you choose, it is important that users are always analyzing data in a secure and governed manner. If you control and govern your data, you can be assured that, teams are not analyzing out-of-date data, losing track of proprietary data, or replicating the same data repeatedly.

Big data is rapidly becoming a fact of life for many organizations. This data can be created in both the cloud and on premises. No matter where your data resides, you have to make sure that data is prepared for the type of analytics required to meet your business goals. How you prepare your data will depend on the nature of the analytics task.

For instance, movement of warehouse or factory employees can now be tracked using sensors on their badges. This capability increases productivity and safety on the factory floor. Badge readers can be placed throughout an area to pick up movement as well as unlock doors, which then track how many times a physical asset is used or not used. With many employees moving from place to place in these environments, managers can perform analytics on the data to determine the movement of employees and the best way to allocate resources. Trackable employee badges can also be used in emergency situation—for example, to make sure everyone is out of the building.

In dealing with any big data IoT application, it will be important to think through the kinds of data needed for analysis. Where should the collected data go? How often should the data be collected and analyzed? What is the frequency of all of the data sources? How can these be synched? These decisions might affect performance and cost.

The data will also have to be prepared for analysis. Data quality and issues such as outliers and missing data do not go away with big data. Likewise, the data will need to be transformed and normalized. A good data preparation tool that can handle big data sources for blending, data cleanliness, and transformation can be helpful.

Finally, when performing analytics on big data you need to consider how the results will be presented to constituents. Mobile phones and tablets (as well as generating a lot of data) have become a primary way many business users view and interact with analytics. Considerations for mobile analytics include limited screen real estate as well as the fact that applications need to perform quickly on mobile hardware. If you are interested in delivering big data analytics generated in the cloud to mobile devices, make sure your vendor can support this.
CONSIDER THE EASE OF UPGRADE OF THE ANALYTICS PLATFORM

One of the greatest challenges for an organization is the volume of changes as vendors add new capabilities to their platforms. Managing changes in software releases can take a substantial amount of time and can cause unnecessary outages. If you need these services, you face the challenge of upgrading your existing infrastructure as well as updating the existing applications themselves. In some cases, businesses are required to maintain multiple versions of an application to support teams and applications with different needs and requirements.

The need to manage change becomes painfully apparent when you upgrade enterprise software. Typically, organizations spend months determining how an upgrade should be done and managing the process. Cloud applications have drastically changed the way many businesses handle software updates. With cloud-based software, it is possible to use a provisioning tool to upgrade software with the click of a button. This process will automatically update software so your organization takes advantage of the latest innovations. Streamlining the upgrade abstracts the complexity so you are not responsible for all the detail (such as managing configuration files). This is a strategy that many organizations are beginning to deploy.

By adopting a cloud strategy for analytics, organizations can adopt the latest software releases as soon as they are available. Continuous updates help organizations to be more agile and take advantage of the latest technology.

ENSURE THAT YOUR PLATFORM IS CHANGE READY

Compare the analytics tools you might have been using three years ago with how you are analyzing your data today. There are new innovative tools available that will make analytics easier. Your organization might be facing new challenges that require new kinds of analytics, such as analyzing big data. You might be interested in analyzing this new data with new techniques (such as predictive analytics to understand churn or predict fraud). You may be performing predictive maintenance. The possibilities are wide and varied. Your analytics platform needs to be ready.

In addition, new analytics innovations are constantly being added to tools and solutions. Make sure your organization is ready for change. You may require Apache Spark for fast cluster processing or suddenly discover that developers require access to popular languages such as R or Python. Prepare for the unexpected.

Your company will benefit if the platforms and tools you use are flexible enough to support the tools that developers want to use. Ensure that your cloud analytics vendor is continually integrating new technologies and features to support changing needs.

In addition to the value of the agile data management platform, the cloud environment itself enables your organization to dynamically scale up or out. This level of flexibility means that your organization will be ready for the future and the changes that will inevitably happen.

In fact, at TDWI we often see that as organizations gain experience and reap success with one kind of analytics, they typically adopt more sophisticated analysis that drives additional value. A cloud analytics platform can be an important part of the strategy to help provide the agility and flexibility to succeed.
CONCLUSION
TAKE THE LEAP

Cloud-based analytics services are dramatically changing your ability to manage and analyze data about customers, partners, and operations. Businesses are increasingly bringing together their on-premises data environments with flexible cloud-based data services. Businesses do not have the time or money to be limited by traditional approaches to data analytics. These organizations simply move too quickly to use valuable time and resources to update software on their own. The cloud provides organizations with the latest innovations in analytics in a flexible and scalable manner. The time is right for organizations to evaluate cloud analytics offerings to determine if cloud services can help them meet business objectives.

ABOUT OUR SPONSOR

MicroStrategy

microstrategy.com

MicroStrategy (NASDAQ: MSTR) is a worldwide leader in enterprise analytics and mobility software. A pioneer in the BI and analytics space, MicroStrategy delivers innovative software that empowers people to make better decisions and transform the way they do business. We provide our enterprise customers with world-class software and expert services so they can deploy unique intelligence applications. To learn more, visit MicroStrategy online and follow us on Facebook and Twitter.
ABOUT THE AUTHORS

Judith S. Hurwitz is president and CEO of Hurwitz & Associates, Inc., a consulting, strategy, and research firm focused on emerging technology including machine learning and AI, advanced data analytics, hybrid cloud computing, and security. She is the co-author of eight books, including *Cognitive Computing and Big Data Analytics* (Wiley, 2015), *Big Data for Dummies* (Wiley, 2013), and *Hybrid Cloud for Dummies* (Wiley, 2014). In 2011 she authored *Smart or Lucky? How Technology Leaders Turn Chance into Success* (Jossey Bass, 2011). Judith provides strategy, planning, and thought leadership to companies making the transition to a new business model focused on the business value of emerging platforms. Judith holds BS and MS degrees from Boston University. She is a member of Boston University’s College of Arts and Sciences Dean’s Advisory board and a board member of the Boston University Alumni Council.

Daniel Kirsch is principal analyst and vice president at Hurwitz & Associates. Dan focuses on security and privacy, cloud, and advanced analytics. He worked as a consultant with Hurwitz & Associates in 2010 and agreed to join the team full-time in the fall of 2011. Dan has authored numerous reports, presented webinars, and performed strategy consulting with a variety of clients. In addition, Dan co-authored *Hybrid Cloud for Dummies* (Wiley, 2014). Dan has a JD from Boston College Law School where he focused on corporate law, emerging corporations, and intellectual property. As an attorney, Dan represented start-ups, cloud computing ventures, and companies going through financing. Dan graduated magna cum laude from Union College in Schenectady, NY.

Fern Halper, Ph.D., is vice president and senior director of TDWI Research for advanced analytics. She is well known in the analytics community, having been published hundreds of times on data mining and information technology over the past 20 years. Halper is also coauthor of several Dummies books on cloud computing and big data. She focuses on advanced analytics, including predictive analytics, text and social media analysis, machine learning, AI, cognitive computing, and big data analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead data analyst for Bell Labs. Her Ph.D. is from Texas A&M University. You can reach her by email (fhalper@tdwi.org), on Twitter (twitter.com/fhalper), and on LinkedIn (linkedin.com/in/fbhalper).

ABOUT TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on analytics and data management issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data management solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

ABOUT TDWI CHECKLIST REPORTS

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.