

Solution Showcase

Unlocking the Power of Analytics with an Always-on Data Lake

Date: November 2017 **Author:** Edwin Yuen, Analyst

Abstract: Digital transformation is impacting the world of business, both in terms of the technology available for the transformation and how that technology is leveraged to deliver business outcomes. As more companies look to analytics and data science for effective decision making, a new generation of integrated, web-scale, highly durable, resilient, and secure storage is needed to support their digital transformation journey.

- Businesses are using data from both internal and external sources to drive digital transformation.
- Leveraging cloud-based analytics for effective decision making is one of the highest priorities for businesses today.
- Cloud-based object storage can provide a persistent data store for analytics, delivering scale, flexibility, and cost savings beyond traditional database and file systems.

Digital Transformation and the Need for an Always-on Data Lake

Digital transformation is driving change, both in the world of IT and in the greater world in which we live. Data and information are now fundamental parts of our lives, from the devices we interact with to the services that we leverage. Data is now being gathered not just from our smartphones and tablets but also from the cars we drive and the buildings in which we live and work. This data can be internal to the organizations we work for, such as email archives, end-user home directories, content repositories in NAS filers, departmental data-marts, IoT devices, and corporate data warehouses. Its source can also be external to the business, from social media, weather, and 24-hour news feeds, to governmental and other origins, such as data from traffic systems, the electrical grid, and public health systems. This new wealth of data contains information that holds a significant amount of business value for companies, much of which was not available before.

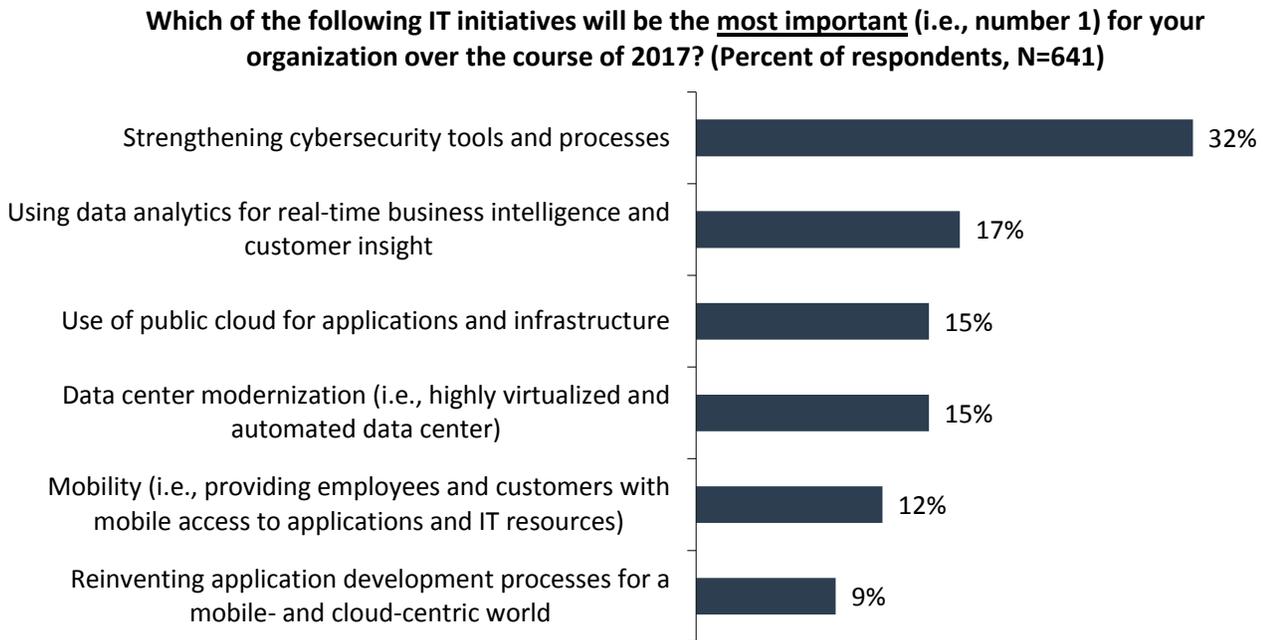
But digital transformation goes beyond just the scope of gathering and cleaning data from internal and external sources. Today, business users in all organizations expect access to data and analysis tools for effective decision making with great speed, and depend on data availability 24x7x365. It is no longer sufficient to provide only a select few decision makers with access to data and only during “9 to 5” business hours. Users now have an expectation of “always-on IT,” a model that supports their needs by offering the right data and the right analysis tool, on demand, when they need it. IT staffs are now inundated with requests from throughout the business, across functions and even from external users, to slice and dice this vast data in millions of ways to create custom reports.

So, how can businesses both leverage the vast amount of data becoming available, and process and deliver the needed results through analytics? With digital transformation, the days of merging spreadsheets and data-marts to produce analytics reports are well past us. What is needed is an integrated, advanced analytics and artificial intelligence (AI) assisted solution that combines a self-service data analytics platform with an always-on data lake to support it. Such a solution would create a common data analytics language, which both data specialists and business users could use to communicate and work with one another.

Businesses Are Looking Toward Leveraging Data Analytics and Public Cloud

Companies are looking to take advantage of the data in their digital transformation journey, using data analytics to drive real-time insights and intelligence to improve and optimize their internal business processes. Another factor that companies must grapple with, as they begin their data analytics and digital transformation journey, is the location and integration of the resources used for these services. These decisions have become complex with the emergence of public cloud services and the scale and economic advantages they provide over traditional, on-premises infrastructures. In ESG’s *2017 IT Spending Intentions Survey*, we surveyed 614 IT decision makers and asked them what the most important IT initiatives were for their organizations. The respondents were asked to choose only one highest-priority response.

Figure 1. Most Important IT Initiative for 2017



Source: Enterprise Strategy Group, 2017

As can be seen in Figure 1, two of the top three results are related to data analytics and the use of public cloud services.¹ The combination of respondents selecting either data analytics or public cloud usage was nearly one-third of the total survey respondents and equal to the overall highest priority, cybersecurity. Businesses are taking advantage of growth opportunities by leveraging data analytics to drive business intelligence and are using the public cloud for applications and infrastructure. The use of public cloud services is notable because it shows that companies are actively looking to leverage off-premises infrastructure to support their IT growth, especially if it provides capabilities they don’t have on-premises. These services are not just extensions of existing on-premises services but relatively new technologies that enable insights

¹ Source: ESG Research Report, [2017 IT Spending Intentions Survey](#), March 2017.

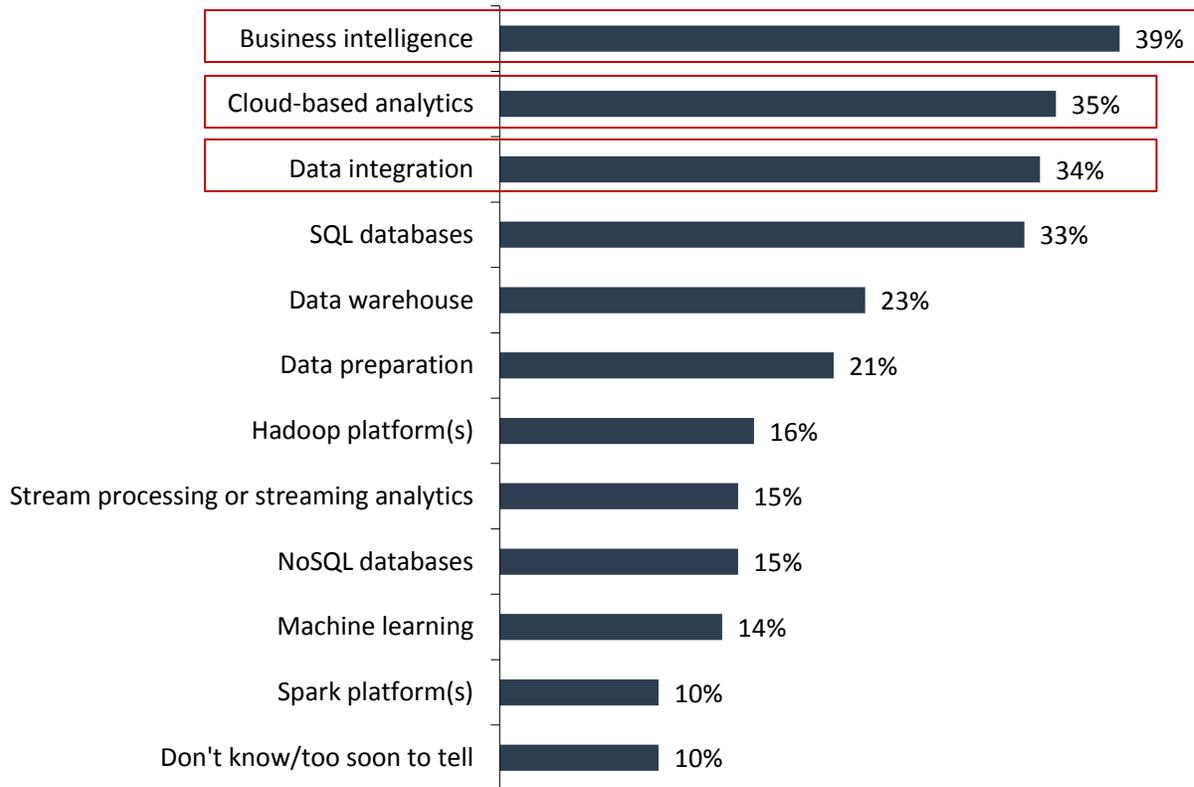
and capabilities not available before. Leveraging new capabilities created by new technologies is at the heart of digital transformation.

Cloud-Based Analytics and Data Integration Are Key Priorities for Enterprise Data Analytics

The use cases for public cloud services are well known to IT today. The same cannot be said about data analytics. The world of data analytics can span services such as business intelligence, cloud-based analytics, SQL and NoSQL databases, data warehouses, real-time and streaming data, and artificial intelligence, machine learning, and deep learning. So, what are companies looking for when they invest in data analytics? ESG asked companies to identify the areas of data analytics in which their organizations would make the most significant investments over the next 12-18 months. The respondents could choose up to five of the responses listed in the survey.

Figure 2. 2017 Business Intelligence, Analytics, and/or Big Data Spending Priorities

We would like to learn more about your specific spending plans for data analytics. In which of the following areas will your organization make the most significant investments over the next 12-18 months? (Percent of respondents, N=337, five responses accepted)



Source: Enterprise Strategy Group, 2017

The two most-cited areas of data analytics in which organizations expected to make significant investments were **business intelligence** and **cloud-based analytics** (see Figure 2). More than a third of the respondents cited those two spending trends as areas for significant spending by their organizations. Other key spending trends included data warehousing, data preparation, and the use of Spark platforms.²

Leveraging business intelligence and cloud-based analytics is important as businesses adapt to digital transformation and the always-on world, and need to further differentiate themselves from the competition by enabling their businesses with

² Source: ESG Brief, [2017 Data Analytics Spending Trends](#), January 2017.

data analytics. Cloud-based analytics solutions such as IBM Watson Data Platform are becoming increasingly popular for businesses that need on-demand, always-on analytics.

Companies also find it challenging to integrate all the different sets of data to drive cloud-based data analytics, where the supporting data lake needs to have the availability and scalability to support cloud-scale analytics usage. As with the top two responses in Figure 2, more than one-third of respondents cited data integration as a top data analytics spending priority. Without data integration, businesses can have difficulty accessing all the required data and leveraging the full capabilities of the analytics platform. IT organizations need the ability to connect multiple internal and external data sources in a single, secure, scalable data lake in order to take full advantage of cloud-based analytics.

Set Free Your IT Using IBM Watson Data Platform with Cloud Object Storage

The explosion in volume and variety of data sources, both internal and external, has far exceeded the capacity of internal IT organizations to ingest and process this data to provide dashboards and reports for effective decision making. The inability to fully meet the on-demand and always-on needs of today's business users and decision makers in all organizations—large and small—can often undermine the benefits of data analytics. Data management is further complicated by the quality of data in existing data lakes, where slow and incomplete data sets make users question whether they can trust the data for their reports. This means that, to take advantage of the analytics to acquire new customers, generate new insights, and create new business models, businesses must offer a self-service analytics and BI platform that puts highly available data and powerful tools at end-users' fingertips, and that enables collaboration.

This is where IBM can step in. IBM offers a combined solution on IBM Cloud, with Watson Data Platform, that provides tools and services for data persistence, governance, exploration, advanced analytics and AI, with a data lake based on Cloud Object Storage. This tightly integrated solution helps organizations become more agile for effective decision making. The key is that the combined solution provides an always-on, advanced analytics and AI platform that can empower business analysts to create data visualizations and deploy analytics applications, enabling data scientists, developers, line-of-business users, senior management, and other executives to ask and answer questions, and accelerate effective decision making.

Cloud Object Storage is a web-scale, unstructured data storage service designed for high durability, resiliency, and security. Users of Cloud Object Storage can get up and running in minutes, with the ability to deploy data buckets across multiple storage classes to optimize their budget and data access. Cloud Object Storage also secures the data store with automatic server-side encryption and integration with IBM Identity and Access Manager, ensuring granular access controls down to the data bucket level and according to user role.

Availability and security in Cloud Object Storage is ensured with a patented technology where data is encrypted, sliced, and distributed across multiple devices in multiple facilities. Cloud Object Storage also offers both regional resiliency, where data is stored in multiple IBM data centers within a single geographic region, and cross-regional resiliency, where the data is stored across three geographic regions, providing dispersed access and protection to data should a regional outage occur. This means that a single copy of data is kept across multiple data centers, as opposed to more expensive solutions that must keep complete data copies in multiple data centers.

Cloud Object Storage works with the Watson Data Platform to provide a cloud-based, self-service data analytics and business intelligence solution. This solution integrates a flexible, consistent, and always-on data lake based on Cloud Object Storage, with advanced analytics and AI tools to deliver composable data and analytics services on the IBM Cloud. IBM allows businesses to leverage both an analytics platform and a data lake that goes well beyond the scalability, security, and cost efficiencies of traditional on-premises solutions.

Watson Data Platform can ingest all the available data sources into the Cloud Object Storage data lake, providing governed enterprise data and trusted user access across the business, including data stewards, data scientists, data engineers, and business analysts. These users can then, in a self-service manner, produce data analytics projects that analyze and deploy the data for use by all business users in the organization. The combined solution also improves the productivity of cloud-native LOB application developers by providing a Spark-based SQL service to query data in Cloud Object Storage. It also combines self-service advanced analytics, AI, collaboration tools, and data ingest and visualization tools and makes them available with rich APIs to program custom applications. All the data analyzed in projects doesn't go away. The insights persist in the data lakes to be re-ingested for future analysis.

It is in this area that traditional data lakes have been unsuccessful when used with cloud-based analytics. Traditionally, data lakes lack the tight integration with the analytics platform, making it difficult to save and reuse the data and models. Cloud Object Storage works with the IBM Data Catalog, with data profilers that segment and protect data, allowing for better governance of data such as personally identifiable information or other private data. By implementing a metadata catalogue, IBM's solution has a fundamental understanding of what the data is and what data policies may apply to it, and then implements those policies. And the data doesn't need to be moved in and out of the Cloud Object Storage-based data lake. Rather, the Cloud Object Storage-based data lake is the persistent store for the analytics.

Unlike other cloud-based analytics solutions, the IBM solution keeps compute and storage infrastructure separate, which gives greater flexibility and reliability. The data is stored in Cloud Object Storage and the analytics clusters connect to the object storage repository when they need to access it. Because the clusters themselves are not involved in data storage, users can spin up a cluster environment for the duration of a single job, and delete it on completion—with no risk of data loss. Moreover, the configuration script can be written once and be reused to create new clusters with the exact same configuration as previous clusters.

All these features are available from IBM, in a service with integrated provisioning, billing, support, and processing, simplifying operations. With the Watson Data Platform and IBM Cloud Object Storage, companies can reduce the IT burden of creating custom data analysis reports for data professionals, business analysts, and company executives. Only a tightly integrated, cloud-based solution can deliver the data analytics businesses want, with the cost-effective, scalable, secure, and always-on capabilities of an object storage-based data lake they need to support digital transformation.

The Bigger Truth

As companies look to accelerate their digital transformation, they must analyze and leverage the vast amount of data that has become available for effective decision making. By leveraging cloud-based analytics and a highly scalable data lake, companies can unshackle their IT departments from traditional on-premises analytics and develop business insights from enterprise data that were not possible just a few years ago. Businesses that leverage a solution like IBM Watson Data Platform with IBM Cloud Object Storage can move beyond just trying to keep up with existing data requests to leveraging data analytics to acquire new customers, gain new insights, and create new business opportunities. A tightly integrated analytics and data lake, with a metadata catalogue, enables deep insights for all business users in a company, creating a common data analytics language they can use to communicate and collaborate with each other. IBM Cloud has created a solution that meets the priorities companies are looking for to leverage data analytics to drive their organization's goals and differentiate themselves in today's competitive marketplace.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.