



White Paper

Adapting to the Cloud for New Data and Analytic Demands

Sponsored by: IBM

Dan Vesset
February 2017

IDC OPINION

One of the biggest changes facing organizations making purchasing and deployment decisions about analytic databases – including relational data warehouses – is whether to opt for a cloud solution. A couple of years ago, only a few organizations selected such cloud analytic databases. Today, according to a 2016 IDC survey, 56% of large and midsize organizations in the United States have at least one data warehouse or are deploying in the cloud.

One of the organizations that has adopted a cloud-based data warehousing solution from IBM is Roskilde Festival Charity Society, which runs Denmark's Roskilde music festival – Northern Europe's largest cultural and music festival that hosts 130,000 visitors annually. Roskilde was looking to mine valuable insights from volumes of real-time data to enhance planning and running the eight-day festival. The project's lead, Per Østergaard Jacobsen, external lecturer at Copenhagen Business School, envisioned an opportunity to engage in this project under greater interest in sustainability and Smart Cities research. At the onset of the project, Jacobsen said, "One day we had a room with chairs; the next day we had a full cloud-based data laboratory."

For Roskilde, the speed of deployment of the new analytic environment based on IBM's cloud solution was a key factor in the success of the project. The biggest recent change in the adoption of cloud analytic databases is the view among technology professionals that these solutions are no longer only about lowering technology and related maintenance and administration costs, although that view remains one of the key decision variables for today's technology purchasers. Cloud analytic databases provide their users with other important benefits that help in promoting more pervasive use of business intelligence (BI) and analytics technology throughout the organization.

The move to the cloud includes a better ability to address changing system scalability and performance requirements, greater opportunity to monitor technology usage patterns and encourage greater utilization of the technology, an opportunity to more rapidly take advantage of new functionality, easier integration of cloud data sources, more efficient information sharing with external parties, and faster development cycles for new analytic applications. These and other benefits of cloud analytic databases are enabling organizations to ensure that any and all of their employees have their decision-making processes supported with the right data at the right time.

However, "cloud" can mean several different technology deployment options. The reality for most organizations is a business analytics technology environment that will be partially deployed on the public cloud and partially deployed on-premise using a range of deployment options. This hybrid environment will continue for the foreseeable future and will require organizations to address the challenges that come with more deployment options. But it will not end there. What we mean by hybrid

can no longer be constrained to on-premise and off-premise deployment of storage, servers, and information management software. Hybrid must also be defined by the specialization and diversification of the information management technology stack that is deployed based on a broad range of information management workloads.

One of the leading vendors in the analytic database management market is IBM, which has a range of cloud and on-premise databases of various types to address today's varied data processing workloads. One of the key products in IBM's portfolio is Db2 Warehouse on Cloud.

IN THIS WHITE PAPER

In this white paper, IDC describes the current and emerging states of business analytics solution deployments. The demand trends, based on IDC's market research, are being addressed with an emerging new data management and analytics architecture that emphasizes a hybrid approach to these analytics technologies. Relying on surveys and case-based research, this white paper cites several organizations that have embraced a hybrid data management and analytics architecture based on technology from IBM. The document focuses on IBM's Db2 Warehouse on Cloud-managed service and provides readers with recommendations to consider when assessing new analytic data management technologies.

SITUATION OVERVIEW

Is your organization in the midst of a digital transformation journey? Are you assessing the value your organization is able to extract from data? Is there a new executive mandate to embrace data-driven decision making? If you answered "yes" to any of these questions, your organization is among the majority of those in the midst of various business analytics, big data, and cognitive/AI initiatives.

Drivers of Change

The investments include reassessment of business analytics strategy and staffing, data, process, and technology requirements. For years, typical business analytics technology has consisted of a three-layered stack of data extraction, transformation, and loading (ETL) tools feeding transactional data into an on-premise data warehouse, which in turn was accessed by a few business analysts and managers using business intelligence tools. This architecture was sufficient for many organizations for over three decades. But over the past few years, this state has changed drastically due to:

- Greater availability of various types of data from both internal and external sources
- Commercialization of a variety of database types, in addition to the RDBMS, and availability of databases on the cloud
- Demand for self-service information access and analytics by a broad range of user groups/types

Recent IDC market research shows that in the past 24 months, 76% of North American large and midsize organizations have expanded the number of users with access to business analytics solutions. It is no longer only a few select analysts and top executives who have access to data, key performance indicators (KPIs), or recommendations from business analytics solutions. Today, in more analytically oriented organizations, managers at all levels, business analysts, data scientists, and operational and customer-facing employees benefit from data-driven decision making.

But it's not about only the number of users of business analytics solutions. 77% of the same organizations cite having expanded the type and sources of data being analyzed. At the same time, 91% of these organizations have started to use new analytic techniques. Most importantly, 74% of these organizations have started using new metrics and KPIs – an indication of a willingness to question the status quo of decision processes and performance management practices.

RSG Media Systems, a software company focused on the media and entertainment industry, and its clients are examples of such change. RSG Media uses various IBM and non-IBM data management and analytics technologies to address clients' needs to understand audience behavior across fragmented viewing platforms. RSG Media's Media Mantra data science platform was developed based on the IBM platform to help media companies optimize content scheduling. The base technologies for Media Mantra are IBM Cloudant NoSQL document store and IBM Db2 Warehouse on the IBM Bluemix cloud platform. In this environment, semistructured data from dozens of sources is ingested into Cloudant and transformed into a schema for Db2 Warehouse. RSG Media's data scientists are using open source R, which is built into Db2 Warehouse to do statistical modeling and analysis. According to RSG Media, IBM was able to package its technology in a way that supported the data pipelining requirements in this complex market. The resulting solutions from RSG Media have shown documented business benefits for clients in the form of lifts in viewership and additional advertising revenue.

However, there are many other organizations that continue to struggle to derive value from data and face challenges in addressing the analytics and decision support requirements of their internal and external stakeholders. According to IDC's surveys conducted in 2015 and 2016:

- Only 30% of organizations cite having an enterprisewide analytics strategy
- Only 38% of organizations indicate that executives emphasize data-driven culture and mandate and incentivize the use of data and analytics
- Only 45% of users agree that the speed of IT's response to business users' or analysts' analytics-related requests meets expectations

These are not indicators of a market where business analytics are pervasively available.

The New (Hybrid) Analytics Data Management Architecture

As a result of the demand trends and challenges discussed previously, organizations have begun to expand their existing business analytics architecture with the guiding principle of a hybrid technology environment. In this case, the new hybrid analytics data management architecture is about not only on-premise and cloud technology deployment but also effective use of a broad range of purpose-built and optimized data management technologies.

The new hybrid architecture is based on technologies for federating nonrelational and relational, transactional, behavioral, descriptive, attitudinal, interactional, and performance data. The reality is that it is no longer practical to have all organizational data in one type of a database. Another two examples of organizations embracing the new, hybrid analytics data management architecture are Roskilde Festival Charity Society and BPM Northwest (a ProKarma company).

Denmark's Roskilde music festival is Northern Europe's largest cultural and music festival. The eight-day event, run by the nonprofit Roskilde Festival Charity Society, hosts around 130,000 visitors annually. Roskilde had the goal and the challenge of mining valuable insights from volumes of real-time data to enhance planning and running the festival and to contribute the findings to better

understanding of Smart Cities efforts around Denmark. The organization was in need of technology that could be deployed rapidly, with minimal up-front costs and appropriate services support. Roskilde selected IBM as its primary technology vendor and used IBM Db2 Warehouse on the Bluemix cloud platform, IBM SPSS Modeler, and IBM Watson Analytics. This ongoing multiyear project began in 2013 and will end in 2018. With this solution and cooperation with a team of researchers from the Copenhagen Business School, Roskilde's decision makers were able to get faster insights for strategic, operational, and tactical decisions, enabled by a flexible, scalable, and cloud-based analytics and data management platform. For more information about Roskilde's use of IBM Db2 Warehouse and other related technologies, see IDC's Customer Spotlight *Experimentation, Innovation, Sustainability: Driven by Cloud Analytics and Data Management* (IDC #US41334116, June 2017).

BPM Northwest, an IT service provider with digital, analytics, big data, and DevOps expertise, has helped clients with a range of hybrid analytics data management initiatives. For one manufacturing customer, BPM was engaged in helping the finance department define strategy to address its data and analytics needs. Part of the solution was to ensure that a database existed that allowed the finance department to pull together just the enterprise and external data it needs and have this be something the finance staff members can maintain themselves. BPM recommended IBM Db2 Warehouse running on IBM Bluemix for this departmental cloud-based data mart. With many other customers, BPM has seen increased use of cloud enterprise applications that provide best-of-breed functionality, but not the data integration capabilities needed to support analytics and decision support needs. A financial services firm with only cloud applications and no on-premise IT infrastructure needed a solution to move general ledger data to its cloud financial planning system. In the process of developing a self-service solution for this client, BPM began the process of collecting this data and other data in the central, cloud-based database that provides that client with consolidated, enterprise data warehouse capability based on IBM Db2 Warehouse.

More broadly, these trends are supported by IDC's market research, including a data integration survey published in 2016, in which we found that 37% of the 651 respondents from large and midsize North American organizations have hybrid (on-premise and cloud) data environments. Furthermore, of the same set of organizations, only 23% had data environments that consisted of only relational databases. The other three-quarters of the organizations also had a range of NoSQL databases and Hadoop, among other data management technologies (based on IDC's survey data, 34% of organizations with a Hadoop deployment have such deployments with a cloud service. According to IDC's data, public cloud spending on analytic data management technology grew 88% in 2015 (growth rate of full year 2016 is not available at the time of writing this white paper, but all indications suggest continued high growth in public cloud spending).

IBM Db2 Warehouse on Cloud

IBM Db2 Warehouse on Cloud is part of a family of SQL cloud database products optimized for either analytics or transaction processing. In this white paper, we focus only on the fully managed (public cloud) version of IBM Db2 Warehouse for analytic workloads, although a client-managed (private cloud) configuration of the product exists when data must stay more directly under business control or on-premise because of internal requirements and other mandates. Organizations can deploy Db2 Warehouse on Cloud as a standalone cloud data warehouse or integrate with on-premise data warehouse for a hybrid solution.

Db2 Warehouse on Cloud for analytics is a relational database that fuses IBM DB2 and Netezza technologies, offering in-memory columnar processing and in-database analytics, with MPP scalability. IBM Db2 Warehouse's in-memory processing can be used on data organized in both columnar and row-based tables. Db2 Warehouse is compatible with a range of IBM and other vendors' BI and analytics software tools, which enables data in Db2 Warehouse to be analyzed with familiar SQL tools or built-in analytics to conduct predictive analysis (including with Netezza, R, and Spark) and geospatial analytics.

For cloud migration, Db2 Warehouse works with supported Db2 drivers and is compatible with IBM PureData for Analytics (PDA) and other databases. For example, a PDA application can federate queries to Db2 Warehouse using built-in IBM Fluid Query to quickly enrich analysis.

IBM Db2 Warehouse is available on IBM Bluemix and Amazon Web Services cloud platforms. Db2 Warehouse's analytics engine is shared with IBM's client-managed private cloud version of the same database, ensuring a hybrid deployment option. The Db2 Warehouse on Cloud is a fully managed service available using the typical cloud pay-as-you-go pricing model.

The Impact of Hybrid Analytics Data Management

Today, the hybrid analytics data management technology supports the following three primary use cases:

- **Development and deployment of new cloud applications, with the option to integrate with on-premise systems.** RSG Media's Media Mantra applications is an example of a new application developed using cloud data management and analytics technology. This data science platform that supports highly demanding customers in the media and entertainment industry has the scalability, performance, and flexibility to support complex preparation and analysis of large and diverse sets of internally generated and externally procured data.
- **Provisioning of self-service, business-driven sandboxes.** Roskilde's use of IBM Db2 Warehouse on IBM Bluemix highlighted the organization's initial need to quickly provision an analytical environment for researchers to start collection of data and its analysis with tools such as IBM SPSS Modeler and IBM Watson Analytics. Self-service analytics is one of the important drivers of technology purchase and deployment, and thus this use case is one of the most frequent ones among organizations across industries. One of BPM Northwest's clients, a large transportation company, has a data science group using SPSS in the cloud to run predictive models. This client needed a data preparation area to feed data into SPSS and pull back the output to use in its analyses. The IT department said that it did not have time or resources to create this environment, and the department was more than happy to have the data science team use a cloud database solution that IT didn't have to support.
- **Migration of existing data warehouse workloads.** Matt Mercier, cofounder of BPM Northwest said, "There is a lot of interest among our clients in the ROI of moving away from traditional on-premise database platforms and into less expensive cloud database environments." This use case works best when the goal is to "lift and shift" a data warehouse as is. However, many organizations use the opportunity of evaluating cloud options to also reassess the entire analytic database environment to ensure that existing shortcomings of a data warehouse are not simply replicated in the cloud. This review of the overall analytics environment is itself a benefit.

These and other use cases in production across industries and various company sizes illustrate the impact that hybrid analytics data management environments have on improving the agility and

adaptability of enterprises looking to digitally transform themselves to provide highest value to their stakeholders.

CHALLENGES/OPPORTUNITIES

In the current environment, as the adoption of cloud business analytics has begun in earnest, IBM has a tremendous opportunity to innovate and ensure that its current and new clients are able to fulfill their cloud technology strategy and plans. Yet the company (along with its competitors) also faces certain challenges. Chief among the challenges is to balance R&D and customer service investments for both on-premise and cloud-based business analytics clients. As the great mathematician Alfred North Whitehead said, "The art of progress is to preserve order amid change and to preserve change amid order." IBM must ensure that it addresses the reality of a hybrid business analytics market that allows clients to subscribe to cloud-based solutions while enabling many thousands of clients to maintain some applications off the cloud – a decision dictated by financial, risk, compliance, staffing, and competitive factors for each individual organization. As Mercier said, "We're still seeing some concern from IT departments that they will be giving up the detailed level of control over hardware, security, and configurations they are used to with on-premise enterprise installations."

For technology users, hybrid analytics data management solutions provide opportunities such as

- Rapid (data-driven) application development (e.g., development and test in the cloud)
- Flexible scalability via hybrid on-premise/off-premise architecture
- Specialized/optimized data management software based on data types and query workloads

However, these users also need to address the following top challenges associated with these solutions:

- Governance around self-service data access
- Data integration, federation, virtualization, and quality issues
- More technology choice requiring more knowledge and new skills (IT and business)

RECOMMENDATIONS

IDC recommends considering the following as your organization contemplates deployment of new analytic database technology:

- Don't let perceived risks of cloud data warehousing solutions deter you from trying such technology options. The risks associated with cloud data warehousing have dissipated. The adoption of cloud data warehousing is now well rooted in market supply and demand trends and should be part of any organization's analytics and data management strategy. The early concerns about system performance and security deficiencies of cloud-based data warehousing services compared with their on-premise equivalents have been laid to rest. There are now reference cases and publicly available studies of cloud data warehouse deployments across industries and geographic regions.
- You don't have to go "all in" on the cloud. Select a pragmatic approach that meets your organization's business requirements and analytic workload types. Especially, if you already have a robust on-premise data warehousing solution, consider cloud deployments to support new analytic applications, provision sandbox environments for analysts engaged in ad-hoc

analytics, or migrate one of the existing data marts rather than a centralized enterprise data warehouse to the cloud. Lessons learned from such projects will dictate the pace of further cloud data warehouse adoption.

- Don't expect any cloud solution to automatically address all the perennial challenges with data quality and data governance. IBM's analytic data management (including Db2 Warehouse), business intelligence and analytics tools, and cognitive computing technology provide functionality to address a broad range of workloads and to help derive value out of data. Standards and integration among software products can enhance collaboration between IT, business analysts, data scientists, and line-of-business staff. Yet, appropriate time and staff will need to continue to be allocated to data governance, data integration, and data integrity issues.
- Embrace the hybrid analytics data management architecture. The reality for the foreseeable future for the vast majority of organizations will be a hybrid analytics and data management environment. Consider "hybrid" to be the basis for a new architecture that is about not only on-premise and off-premise technology deployments but also effective use of a broad range of purpose-built technologies and optimized data management technologies. The relational data warehouse has existed for decades and will continue as an integral part of any analytic data management architecture, but it will need to be augmented with new data management technology. Therefore, the need for expert analytic data management technology architects will rise and command premiums similar to those for expert data scientists.
- Assess benefits of cloud data warehousing and the broader hybrid analytics data management environment based on a number of metrics that go beyond initial cost considerations. Although subscription-based pricing of a cloud solution is attractive to most organizations, the measurement and monitoring of benefits should also include labor costs (with efficiencies favoring the cloud solution), greater flexibility in addressing scalability requirements, greater speed of deployment (and hence speed to insights), easier data integration with a growing number of cloud-based source systems, and easier ability to monitor technology usage and adoption.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2017 IDC. Reproduction without written permission is completely forbidden.

