DataOps is a new discipline but is catching on quickly because it reduces data errors and application errors and enhances the speed and quality of data, giving businesses a competitive edge.

The Business Value of DataOps

October 2021

Questions posed by: GRAX
Answers by: Archana Venkatraman, Associate Research Director, Cloud Data Management, IDC Europe

Q. What is DataOps, and what is its significance in helping organizations become intelligent digital businesses?

A. We are in a digital economy. By 2022, more than 65% of global products and services will come from digitally transformed organizations that are heavily reliant on data and data-native workers for business success. According to an IDC survey, 80% of CEOs emphasized that using data in advanced decision models for performance and competitive advantage is "extremely important" to their organizations. This is because they see a strong correlation between data-driven insights and business outcomes.

As the value of data and data-native workers grows within enterprises, many are turning to DataOps strategies to improve the outcomes of data analytics, data science, artificial intelligence (AI), and machine learning (ML).

DataOps is a new discipline but is catching on quickly because it reduces data errors and application errors and enhances the speed and quality of data, giving businesses a competitive edge. According to IDC's research, effective DataOps methods in use among enterprises today include data sandboxes, version control, feedback loops, and logic and data testing. As DataOps borrows the principles of agile and DevOps methodologies, it lends itself to data science, data visualization, and data warehousing use cases by breaking down the silos in data pipelines from ingestion to analysis and visualization. DataOps is a set of best practices enabled by platforms such as data-enabling backup platforms that help break down data silos.

By 2023, 60% of organizations will start implementing DataOps programs to reduce the number of data and analytics errors by 80% and to boost trust in analytics outcomes and efficiency of data-native workers. The DataOps focus on quality will improve the level of trust in data, data analytics, and data science as it helps organizations adapt to new business needs.

DataOps is not one tool but an entire pipeline made up of multiple tools and technologies that are connected or coordinated in DataOps processes. The key to success is to support more integrations and automate testing at various points within the data life cycle.

Q. Where does valuable business data reside? How can businesses best capitalize on this data?

A. Data gravity is beginning to shift to the cloud as cloud is everywhere for everything.

Business data is growing at an exponential rate, and more organizations are having to manage petabyte-scale data. In fact, according to IDC's research, twice as many organizations in 2021 admit to having petabyte-scale data compared with 2019. This data is fragmented across core datacenters, SaaS applications, cloud infrastructure, and edge locations.

SaaS adoption is accelerating as these applications bring a modern UI, enable better collaboration, and transform legacy business processes. According to IDC research, mature business processes such as CRM, ERP, finance, HR, email, and collaboration are all predominantly becoming SaaS based. This in turn means that business-critical data such as sales, customer, marketing, and even financial data increasingly resides in SaaS applications.

In conversations with IDC, more customers have said that they rely on data in Salesforce to inform their business road map, customer experience management, and marketing strategies. About 52% of large enterprises have customized or added development on CRM applications to meet their business needs, making data in SaaS CRM applications business critical and relevant for analytics outcomes.

Multiple trends are colliding:

» There is pressure to become data driven. According to IDC research, 87% of CXOs have said that being a more intelligent enterprise is their top priority for the next five years.

» AI and ML are moving beyond being just "nice to have" programs to becoming ubiquitous across all business processes, especially in the post-pandemic world. Investment in AI services, software, and hardware will grow from $50 billion in 2020 to $110 billion by 2024 (source: *IDC FutureScape: Worldwide Artificial Intelligence 2021 Predictions,* #US46917020).

» Business-critical data is increasingly stored in SaaS environments with control of, access to, and use of data becoming very manual, laborious, time consuming, and costly for organizations. IDC estimates that automating SaaS data backup and sandbox environments can give large enterprises an average annual savings of $250,000 that would be spent on highly trained CRM administrators executing manual tasks spread over 100 hours a month.

Customers have been relying on Salesforce for the past 10–15 years, and there is increasing pressure to unlock the data, including historical data, in these environments to create a holistic analysis for maximum impact.

In conversations with IDC, large enterprises have indicated that there is growing value in accessing and using historical data in Salesforce for analytics purposes to inform changes in forecasts, deal value changes, new personas, and new opportunities.

To capitalize on SaaS data, especially historical data, customers need to own and control this data; have unrestricted and speedy access to the data without impeding, deleting, or corrupting production environments; and be able to integrate the data with broader analytics programs.

In IDC's opinion, over time the value of historical data will become significant as AI programs take center stage and as organizations look to build trust and transparency in their AI models. IDC predicts that by 2022, over 60% of consumer-focused AI decisioning systems in finance, healthcare, government, and other regulated sectors will include provisions to explain their analysis and decisions.

Q. How can organizations be in control of historical data, and what are the best practices?

A. With SaaS applications such as Salesforce becoming central to business processes, leveraging historical data to identify business patterns/trends, anomalies, and potential is key. But accessing and using this data with traditional approaches will not yield the desired results.

Customers looking to leverage SaaS application data today are following inefficient practices:

» **Using historical reporting functionalities in the SaaS platform that have limited features.** The Salesforce platform itself offers the Historical Trend Reporting feature, which is very popular. But it is subject to certain limitations, such as a cap of 5 million rows, limited ability to customize, and complexity in filtering data. In addition, the complexities of data APIs and extract, transform, and load (ETL) mean that users typically analyze data only within the Salesforce platform instead of combining or ingesting historical sales/customer data into modern data warehousing platforms such as Snowflake or Amazon Redshift. They are also not able to combine data from external sources into Salesforce for richer analytics. This limits success in DataOps strategies.

» **Storing all the history in the SaaS vendor's production environment.** This practice is not cost efficient, and it also increases the risks of errors/corruption and application overload in the production environment. Being in control of SaaS data by backing it up in the customer's cloud environments of choice can allow reuse of SaaS data in a way that mitigates risk. Customers can take this valuable SaaS data closer to their core analytics environments and use it in downstream environments without ETL, heavy customizations, or homegrown connectors. Modern cloud strategies are all about adopting cloud on customers' terms for business outcomes that matter to customers.

» **Migrating historical data to cold storage archives on premises or in the cloud.** This practice is cost effective, but retrieving data from traditional cold storage is difficult and time consuming and incurs high egress charges. The secondary use cases of data such as analytics and data mining as well as for training AI/ML models are growing rapidly, and having easy access to archive data is critical to augment these use cases.

The first principle for analytics and AI success is having the ability to control, access, restore, and integrate all the relevant data into the analytics platform or data warehouse in an automated, integrated fashion. This can help maximize the value, speed, and quality of data programs.

Archiving cold data is a good starting point, but today's cold data can be tomorrow's active data for AI and analytics. This means having flexibility of data access is a paramount consideration when archiving SaaS data.

It is not too late for any organization to invest in the future of intelligence. Over the next five years, the ability to synthesize data, the capacity to learn, and the capability to apply insights at scale will define an organization's success.

SaaS data, including historical data, is at the heart of this intelligent future. But companies need to own this data and be able to access it per their needs.

This is where modern backup platforms that go beyond the static role of storing backup data become a powerful enabler of organizations' AI and analytics journeys.

Q. Are you recommending a paradigm shift in the role of backup and viewing backup through a new lens as a data source?

A. Absolutely. Backup solutions have never been seen as revenue generating. They are considered an "insurance policy" or a "necessary cost center" for disaster recovery, compliance, and business continuity needs. Most organizations didn't even consider SaaS workloads in their data protection strategies until 2019.

More recently, the rise of business continuity, cyber-resilience, and data governance as boardroom priorities has piqued interest in backup, particularly SaaS backup, as part of efforts to mitigate ransomware risks and to show regulatory compliance. But backup still remains in the realms of niche storage and backup administrators and not part of the data life-cycle strategy. In IDC's 2021 multicloud research, 58% of organizations said they have a dedicated backup strategy for SaaS data and another 33% said they are "considering" SaaS data backup in the next 12 months. This is an improvement from 2019 when 6 out of 10 organizations didn't consider backing up SaaS data.

However, backup is still siloed from archiving strategies and disconnected from broader data-driven initiatives. It is also the last environment to be modernized with automation, APIs, DataOps, and cloudlike flexibility.

We have argued the value of DataOps, and it starts with a modern backup platform as the foundational pillar. Data protection strategies should move away from traditional, infrastructure-defined focus to become part of the data ecosystem and DataOps processes.

There is a proliferation of SaaS applications, and the need to be in control of business-critical data in these environments requires a paradigm shift. Organizations have to develop a unified strategy of four stages — back up, archive, activate, and monetize — to get the most value from their SaaS data.

For example, many organizations reusing SaaS data for innovation are following manual processes and going through multiple steps such as requesting data, configuring data, loading data, and anonymizing data as silo processes. This approach undermines their speed of innovation and increases the risk of corrupting data in production platforms. Other organizations rely on homegrown applications that are costly to manage and don't scale easily with changing needs.

It is time to move from a single-purpose, static backup and restore approach to a holistic data management and enablement platform view. Having a unified platform for the complete life cycle of mission-critical data can be a game changer.

Modern backup platforms that focus on the data ecosystem and help maximize the value of SaaS data through backup, recovery, compliance, archiving, access, integration, and reusability are redefining the role of backup.

These solutions bring a cloudlike experience as well as cost efficiency, automation, and flexibility to backup but also go far beyond in the data value chain. They offer the simplicity of a single platform that captures SaaS backup data and archives

that data to free up the core production environment but still gives the ability to access the data instantly for analytics. These platforms also help maintain the history of the data and capture all the changes for business insights, improve the quality of data for AI and ML models, and facilitate integration of SaaS data into data warehousing or analytics platforms in a seamless way.

SaaS data is a valuable and reliable source of historical data captured on time. Modern backup platforms can be a compelling repository of valuable historical data and allow businesses to unleash its potential.

A data optimization journey starts with a platform that facilitates data-driven innovation on the customer's terms and at the speed and scale the customer requires.

Modern backup can be both:

» A strong defensive strategy

» A powerful data source for DataOps

Q. Do you have any examples that illustrate how some forward-thinking organizations are using a modern platform to realize the values of SaaS data?

A. The value of SaaS data is only growing. As this data becomes pivotal for building an intelligent enterprise, leveraging a modern backup platform as a data source can make all the difference. It will help reskill backup and CRM professionals to DataOps roles, and automation will help the teams focus on strategic tasks such as data quality, metadata management, and data as a service.

More importantly, a modern backup platform delivers value to all the stakeholders of DataOps — IT operations (backup and governance professionals), business owners (CRM professionals), data engineers, data architects, and data scientists. It empowers them with data ownership, availability, and flexibility.

Data engineers and architects are at the heart of any business' data-driven strategy because they are responsible for data quality, integration, data flow, ETL, and data warehousing. They develop the data pipelines and are the bridge between IT operations, software development, and data scientists who are responsible for analytics, data mining, and AI/ML strategies. Data engineers face multiple challenges such as inconsistent data APIs, fragmented data sources, and multiple data types. Addressing these issues is increasing the technical debt and making data engineering a bottleneck in data-driven strategies.

Having access to historical data from SaaS applications within organizations' own cloud environments eliminates ETL complexities. The data teams within organizations can ingest the data into downstream environments in a consistent manner and accelerate the speed of data programs. Using historical data from modern backup to ingest into data warehouses in an automated way helps data engineers focus on more strategic tasks such as ensuring data quality, monitoring access controls, and quickly meeting data science needs. The core functionality of backup shouldn't be compromised. It should still improve data resilience through swift restores and recovery in case of cyberattacks or

accidental deletions. The archiving capability should help organizations adhere to compliance regulations via defined policies for data retention and a mechanism to deliver on their shared responsibility mandates. But viewing backup from the new lens of DataOps can bring much more value. The platform can help maintain history and data lineage. It can provide instant access to any data, blurring the lines between hot and cold data, and improve the trust in data and AI strategies.

It is time to take control of SaaS application data with modern backup and data-enabling platforms.

Consider the following example: A multinational pet healthcare company, with strong data-driven ambitions, was storing connected customer device data in AWS. It wanted to capitalize on features such as Amazon CloudWatch and ElasticSearch for intelligence, but it couldn't integrate business-critical data in Salesforce and ERP environments with the connected device data. As a result, it couldn't yield quality insights.

The company used a modern backup and analytics-enabling platform to capture and archive historical sales and service data and subsequently reused its historical Salesforce data alongside cloud data to achieve 360-degree customer visibility and inform key business decisions. The winning strategy was using a backup platform that acted as a data hub for data engineers to use standard Parquet formats to ingest data into existing analytics pipelines without custom engineering or dealing with complex data APIs. The company's ecommerce revenue grew exponentially from 0% to 70% of total revenue in less than 18 months. The backup platform not only protected critical SaaS data but also facilitated the reuse of that data with minimal engineering and coding needed to deliver data-driven business outcomes.

About the Analyst

Archana Venkatraman, *Associate Research Director, Cloud Data Management, IDC Europe*

Archana's primary research coverage is cloud data management. She covers multiple topics including data protection, edge-to-cloud data trends, application and data availability, compliance, data integration, intelligent data management, DataOps, data quality, and multicloud priorities and trends. Archana is also a co-lead of the cloud practice and an active contributor to IDC's Europe's DevOps and AI research practices.

MESSAGE FROM THE SPONSOR

About GRAX

More and more business is conducted in SaaS applications like Salesforce. The strategic value of that data is enormous -- especially when leveraged in other enterprise analytical and operational systems. However, many enterprises don't take full advantage of it because of the complexities inherent in protecting, accessing and integrating that data into their ecosystem. To minimize SaaS app data loss and simplify reuse, best practices include frequent back-ups to the customer's own cloud data lake, where complete historical data sets can be flexibly, securely and easily accessible for downstream consumption.

GRAX's solution helps organizations transform SaaS data backup from a simple insurance policy to a true business accelerator. To see GRAX in action, click here.

IDC Custom Solutions

IDC Research, Inc.

140 Kendrick Street

Building B

Needham, MA 02494

T 508.872.8200

F 508.935.4015

Twitter @IDC

idc-insights-community.com

www.idc.com