DataOps: Companion to DevSecOps for Reimagining Applications

RFG Perspective: DataOps methodologies give business leaders a new set of tools designed to rapidly marshal corporate data from across their organization – and to prepare data, making it business-ready for new uses. DataOps is a necessary companion competency to DevSecOps for enterprise developers and data scientists. DataOps allows companies to rapidly build data pipelines for directed projects and rapid innovation. It's an approach that prepares a company's data for faster analytics that inform business outcomes. Business and IT executives need to adopt DataOps so that they can reengineer their data discovery, ingestion, cleansing and delivery processes -- and make new datasets available in a timely manner.

In the age of COVID-19, businesses processes and technologies must move forward quickly – accelerating development of key applications to meet new demands. The shape of business applications has changed – with built-in support for mobility and remote workers a given, and support for enterprise-level RAS a requirement.

Business processes and technologies – both get transformed as organizations move forward. DataOps gives IT managers and data scientists a path to untangle the complex patterns of data storage that were the inevitable result of years of application deployments across business units – each of them generating its own data pools. That's why DataOps efforts should be executed in parallel with associated DevSecOps efforts. This strategy is very different than using the widely adopted serial approach, which involves requests to DBAs for data at the time of test.

Business Rationale for Moving to DataOps

DataOps is an important part of reinventing enterprise IT – with the imperative of coding for simplicity, flexibility, private clouds and agile deployments. To get the most benefit from newly imagined, or reimagined, business applications developed with DevSecOps processes, we should now consider how data – in all its aspects – will play into the New Normal business picture. Much of our enterprise data is stored in separate data silos: It is scattered like islands across the broad sea of data inside the corporation as a happenstance of its earlier development within business units or newly acquired companies.

As businesses evolve, their data architecture needs to adapt to rapid change, too. DataOps allows Chief Data Officers (CDOs), data stewards, and data scientists a way to address the data architecture of our evolving business organizations. Importantly, it gives them a process to eliminate or minimize the islands of data in the enterprise data center – and in the clouds – that were created over time in previous development efforts.

How DataOps is Done

The two processes – DevSecOps and DataOps – should be coordinated, so that new applications can benefit from data pipelines optimized for those workloads. To accomplish this, we must break down the development process into its logical steps and find a way to automate and orchestrate them better for rapid execution.

46 Kent Hills Lane, Wilton, CT. 06897; (203) 429 8951;
http://www.rfgonline.com/; *Contact: inquiry@rfgonline.com*

But first, we must work with the data that's stored across the enterprise and scattered across the hybrid cloud. That data must be defined for its usefulness, discovered, identified, and then tagged with metadata, which allows applications to tap that data more quickly. Some customers have applied AI/ML and automated software to this process, allowing them to search through data lakes and repositories more quickly.

The Hunt for Relevant Data

The hunt for relevant data begins with understanding the business requirements for the new desired application. If it is not needed for the application at hand, then it is not relevant – and it should not be included in a focused DataOps project. We take this approach to avoid "boiling the ocean" of large data repositories, or data lakes. We need to find only the relevant data, to discover its location and storage method, and to tag it for later use – or possible storage in another location.

- **Definition.** Defining the data we need to run specific projects – many of which are being transformed, or redeveloped to meet new business needs.
- **Identification.** Labeling the data to state its origins, method of storage, its location – and its use for new projects is a vital stage for DataOps processes.
- **Automation.** Many organizations apply AI/ML for the automation steps, to reduce search time for IT staff, and to apply metadata tags that will help us apply the data to new projects.
- **Orchestration.** The use of orchestration is another secret ingredient of DataOps and should be thought of as the complement to the CI/CD process that is used by the DevSecOps team.

By taking the above steps we avoid time-consuming and costly, mostly manual ETLs (Extract, Transform, Load) processes – which soak up too much data administrator time, and may not yield the results we want. The objective here is to acquire and use only the data that will be needed for the immediate, high-priority DevSecOps project. Our business goal should be to deliver that data as rapidly as possible, while ensuring data currency, integrity, protection, privacy and security.

Business Benefits of DataOps

If we do DataOps right, we have the opportunity to meet accelerated project deadlines by gathering up the relevant datasets – and not spending our time reshuffling legacy data in a process that does not improve results. And when we make major changes, we should architect the data such that its logical views can be used multiple times. The objective should be to build a data architecture and infrastructure that is extensible for future use.

The immediate advantage of DataOps is the ability to create the needed datasets in a short period of time, usually a few weeks, versus the six to nine months traditionally associated with the standard ETL cycle. Moreover, the data has been created with an architecture in mind, ensuring that future islands of data need not be created. Instead, multiple metadata catalogs can provide

different logical views of the data. Each of those metadata catalogs will define the elements and rules for specific analytics uses and business cases.

Keys to DataOps Effectiveness

1) Get Customer Buy-In Early: Working with Your Customers and Showing Better Business Results

No project gets done without business-unit approval that pays for new projects that develop and deliver new business applications. Recognizing that fundamental principle of IT – getting projects funded – it becomes vital to gain and maintain the trust of your organization's business managers.

This much became clear in the cloud-computing era, when business units could opt to pay for cloud computing services rather than to build up physical infrastructure in company data centers. Developing and launching a series of focused projects will help to build momentum for DataOps projects – no matter how you end up staffing those projects (such as hiring inside providers or outside providers – and paying for cloud resources).

2) Limit The Project Scope

Focus on the business needs to be solved – and break the work down into steps that can be accomplished quickly. DataOps works best when it is highly focused on specific data goals. The smaller the data pools, the better to supply vital data to highly focused data projects, as accomplished by project teams' "sprints." Many companies organize multiple "sprints" to support well-defined data-based analytics projects, making the analytics process more efficient – resulting in better business outcomes.

3) Organize "Sprints" to Get the Work Done

Defining your project carefully – and breaking it up into multiple sprints – will make your project more manageable. This approach can allow your organization to break down larger data transformation efforts into multiple sprints of weeks, rather than months, of work before a new project can take flight.

Each sprint is planned to quickly deliver a usable business application within a defined time-period. The key is speed to market with the basic functionality. Then each subsequent sprint adds more functionality that can be leveraged for business purposes.

Organizing cross-discipline teams to hash out the use of datasets to achieve specific business goals is often done in "sprints." These sprints bring together decision-makers, data scientists, analysts, ML experts for prototyping code, and people with administrative IT skills to put the plan into action.

4) The Importance of Metadata

Metadata—the data about the data—is essential when finding, tagging and gathering up the data for a new project from data lakes, or across hybrid clouds. Many organizations are becoming adept at applying AI/ML software to automatically generate metadata for new data pipelines.

The AI/ML experts may be in-house, on your IT team – or they may work in a third-party partner firm. Cloud service providers (CSPs) are well-positioned to help with AI/ML, because they have the tools at-hand and the skilled people to apply them. Some organizations have found that leveraging AI/ML can reduce metadata generation time by 80 percent or more, compared with manual methods of generating metadata.

5) Providing Guardrails for DataOps Effectiveness.
The way we approach ETL and data cleansing will determine our success at deploying new applications that were developed with an agile DevSecOps method.

The need for data integrity, data protection and data security do not go away with new DataOps projects. We need to protect PII data (e.g., personal identification information, like Social Security numbers, birthdates and health data) and we need to satisfy all appropriate regulatory requirements (e.g., HIPPA for healthcare; GDPR in Europe; CCPA in California; and other regional data-privacy standards.)

Summary

We are Reimagining/Reengineering data center applications with DevOps, DevSecOps and DataOps. Why? We're working across data silos and hybrid clouds – gathering the datasets we need for high-priority fast-track development.

In the COVID-19 era, we stand at the intersection of modern applications and old ones – with all the enterprise requirements for reliability, availability, and security (RAS) that we've had before. We can't miss this opportunity to re-examine the data structures we have built across our organization – and to decide how to harvest them for well-defined, focused datasets that will help us to realize their business value for new projects.

We know we can't move large sections of enterprise data across the network, or across multiple clouds, without losing efficiency as we streamline the development process itself. From a people perspective, we must get our DevOps and DevSecOps teams to work with DataOps teams as we transform our business, and as we accelerate our projects to meet the pressing demands of New Normal business conditions.

46 Kent Hills Lane, Wilton, CT. 06897; (203) 429 8951;
http://www.rfgonline.com/; *Contact: inquiry@rfgonline.com*

RFG POV: To get things right for the next era of computing, we need to discover what we have – and then find and harvest the business data where, and when, it is best addressed to new workloads – and then start using it for new purposes. Business and IT executives need to employ the new development models – tapping DevSecOps and DataOps – if they want to remain competitive in the New Normal business world.

Additional relevant research and consulting services are available. Interested readers should contact Client Services to arrange further discussion or interview with Mr. Cal Braunstein, CEO and Executive Director of Research. Jean S. Bozman, President of Cloud Architects Advisors, co-authored this research report.

46 Kent Hills Lane, Wilton, CT. 06897; (203) 429 8951;
http://www.rfgonline.com/; *Contact: inquiry@rfgonline.com*