



Summary Insights: Structuring Hybrid Clouds with a Data Architecture

RFG Perspective: The shift to the cloud created a paradigm change for data architectural considerations. Companies were just beginning to address the challenges created by the islands of data that existed within their data centers when the cloud movement began. Clouds have exacerbated the data architectural challenges, because the drive to cloud is led by development teams – and, in many cases, data architects are not consulted. As a result, data tends to be an afterthought rather than architected. Therefore, the ephemeral nature of applications -- combined with the permanency and volume of data, data gravity issues, cost of data exports, and other elements -- is causing new data architectural issues that need immediate consideration. Enterprises need to understand the data architectural concepts that are now available and develop methodologies to address the data challenges they are encountering.

Introduction

On August 19, 2020, an **RFG 100 video conference** brought top IT executives together for a frank discussion of techniques to map out an organization's data resources – and to manage and control them better in the future. Participants included CTOs and top business and IT executives that have been addressing the hybrid multi-cloud data architecture. The presenters on the call were:

- Terri Sage, CTO, 1010data
- Brian Binovsky, VP, Barclays
- Stephen Harris, Cyber Security Architect, FiServ
- John Downey, Field CTO, HammerSpace

In the COVID-19 era, it's more important than ever to take stock of your data resources – and to manage them to support your organization's key applications. Once you discover where the data is, you can turn your attention to securing it, encrypting it, protecting it, managing it – and making sure it's available to your applications, new and old. That due diligence – with better data preparation and better data governance -- will help your enterprise succeed in a financial recovery from the crises of COVID-19, business shutdowns and financial disruption.

Video-conference Survey Results



The following survey questions summarize the enterprise positions from the conference call.

Does your organization have a coherent cloud data architecture strategy?:

- Yes, completely 29%
- Yes, mostly 14%
- Yes, partially 43%
- No 14%

Who is responsible for the data architecture?:

- Head of Data or Data Services 36%
- Chief Data Officer 14%
- Data Steward 7%
- No one 21%
- Not sure 21%

Are you confident that you can store sensitive/PII data in a public cloud with an acceptable level of risk?:

- Yes 86%
- No 7%
- Undecided 7%

Things are changing – rapidly. Recent customer interviews reveal that the challenges in data architecture are as much in the “people” issues as they are in “storage” issues. IT staffs are rapidly losing a common historical memory of where data is located, and how to manage all of it. They can no longer rely on an apprenticeship system of transferring knowledge from the programmers who designed the older data-stores to the new generation of DevSecOps staffers and data-management specialists who are running their IT operations now.

Add to this the second wave of cloud migrations, which brings its own set of challenges that are centered on managing distributed data in



hybrid multi-clouds. Taking inventory of the data silos, identifying key data-sets, correcting inaccurate data – and managing all of the data in a consistent way – are all keys to building out a modern data architecture. With the emphasis on application modernization, containers and orchestration, the sheer volume of data that resides in data silos and data lakes is an important IT challenge that must be answered – right away.

To achieve better business outcomes that boost business growth, data acquisition, cleansing, and analysis must happen faster, and data searches must center on key datasets properly outfitted with metadata tags.

RFG100: Summary of Key Findings:

Key findings from the RFG 100 conference are shown here. As your organization reimagines its workloads for the post-COVID-19 era, consider the following factors that emerged from the **RFG 100** survey and discussion:

- 1) Establish a Process for creating a Cloud Data Architecture.** A process for creating a coherent cloud data architecture needs to be developed. Very often, DevSecOps teams are requesting data at the “last minute” of a software project, resulting in a reactive response -- rather than a structured cloud data architecture that can be used for multiple applications across multiple clouds. To avoid this problem, there must be an owner somewhere within the organization who is responsible for the cloud data architecture.
- 2) Map the Data Architecture.** Knowing what you already have is essential before preparing to manage and control data in a multi-cloud environment. Enterprise applications require data



that is secure, resilient (supporting high availability and disaster recovery), and compliant with data-protection regulations.

- 3) **Discover your organization's data resources.** Discover your organization's data resources – both on-prem and off-prem. Use data identification software to “find” data throughout the organization. No data silo should go unnoticed – or unsecured. Silos that grew up around specific business units – such as specialized departments within a hospital or health-care organization, must now be viewed as inputs to a wider view of enterprise-wide data. While the sensitive data, with personal information or HIPAA data, should be encrypted and anonymized or pseudonymized to protect identities, the data must be treated in a consistent way across the hybrid cloud – and multi-clouds.
- 4) **“Clean up” the data.** This is about removing incorrect or non-relevant data. Incorrect or incomplete data leads to biased or incorrect analysis. That is counter-productive to the goal of producing actionable insights for better business outcomes.
- 5) **Planning: Bring Everyone to the Table:** Bring all teams “to the table” to work out Best Practices for Data Management across the organization. The teams include all the data architects, application architects and networking architects. Without a complete holistic enterprise-wide view, structural flaws in the data architecture will persist. *This is the most important finding of the RFG discussion: Organizations must update a new set of Best Practices for Data Identification, Data Protection and Data Management, across many geographic locations worldwide.* This is no easy task. But without the direct input from many teams across the organization, key data resources may be overlooked – and important considerations may be missed.
- 6) **Discover and manage your organization's “data patterns.”** It turns out that multi-cloud computing is driving IT experts to re-discover their organization's data patterns. That is, the way



these organizations use data today – and how those usage patterns may have to be adapted for multi-cloud operations. Architectural patterns are reusable data-based solutions. Examples include accessing data lakes; accessing in-memory data, accessing streaming data, and accessing a variety of data types as stored in different formats (e.g., block, file, objects). Having a centralized and integrated view of an organization's data resources, accessible by a global namespace, supports fast access to high volumes of data – and supports a wide range of data analytics.

- 7) **Plan for serverless workloads – and stateful ones.** The first wave of cloud migration included many stateless scale-out applications, such as those for search or scalable compute. The second wave of cloud migration is underway, bringing more enterprise applications to hybrid clouds and multi-clouds. Both stateful and stateless enterprise applications can require persistent data-stores and highly resilient applications to produce correct and meaningful results.
- 8) **Prepare for the “democratization” of data.** This process – “data democratization” – brings fast and easy access by data scientists and analysts throughout the organization. Having a holistic approach to the way all data-stores are housed, replicated and managed, is essential for ensuring that data scientists will perform data analysis that brings productive, and useful, results.

Principles of Data Management for the “New Normal”

Discover the Data – then Map It

Data silos and data lakes pose significant challenges to mapping – and managing – an organization's data. They must take an inventory of all the significant data repositories across data centers and the hybrid



cloud. To do that, they'll need software that can discover and identify the data resources – and then generate metadata that will help them manage the data – no matter where it resides.

Do the Homework: Generate the Metadata

Metadata — the data about the data — is vital to any data architecture. You cannot control what you cannot see. Generating metadata for all of a company's databases is extremely time-intensive. That's why many IT organizations are looking to AI and automation to speed the process of metadata generation for many TB of data on-prem and off-prem.

It's worth noting that cloud service providers (CSPs) offer services that speed up data-tagging and metadata generation – and these can help with building up metadata that will help IT staffers manage and control data across sites –in a multi-cloud world.

Create a Global Namespace for Universal Management and Use

The best way to get a single, unified view of your organization's data resources is to create a global namespace that lists all relevant data repositories. This helps IT staffers to view all data resources – and to manage them better. In turn, this helps business managers understand the data resources already in place that should be applied to new data analysis – and to new business strategies. A thorough mapping will support preventive and proactive data analysis across the multi-cloud.

Impact on Reimagining Applications: DevSecOps and DataOps

DevSecOps, the combination of application development, security, and operations, is an accelerator for a development of new cloud-native applications. It combines the programming and operational skill-sets to speed development, testing and deployment for new applications.

Now, DataOps is being used to identify key data – and to apply new management and control capabilities. As we wrote in our June report: **[“DataOps: Companion to DevSecOps in Reimagining Applications”](#)**



“To get the most benefit from newly imagined applications, developed with DevSecOps processes, we should now consider DataOps as a companion competency for enterprise developers and data scientists. Here, we see technology requirements and people requirements. Both must interact in new ways to address an entirely new set of business conditions brought about by COVID19, remote working and the New Normal business world that’s taking shape.”

Summary

It is becoming mantra in the Business world: “Data is the New Oil” – i.e., your data resources are an important asset that your organization owns and controls. If your data is “clean” and up-to-date, then tapping data resources on-premises and off-premises will add to your organization’s potential for growth. Incorrect data that is aging-in-place is a deterrent for achieving those better business goals and will cause future cloud technical debt.

RFG POV: Hybrid multi-cloud ecosystems have created new data architectures, challenges and risks that many enterprises have not encountered before. These challenges must be understood and cannot be ignored. Business and IT executives should ensure their teams remain current on the ever-changing data architectural concepts – and are building solutions that address their near-term and long-term requirements for agility, compliance, cost, data integrity, and security requirements.

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 LLC, contributed to this report.