

# Achieving Data Resilience at Scale Requires a New Approach to Backup and Recovery

by DCIG Lead Analyst Storage, Ken Clipperton



**COMPANY**

Arcitecta  
 LOUISVILLE OFFICE  
 357 South McCaslin Blvd.  
 Suite 200  
 Louisville CO 80027  
 USA  
 +1 720 583 4006  
 talk@arcitecta.com

MELBOURNE OFFICE  
 15 Little Bakers Lane  
 Northcote, Victoria 3070  
 Australia  
 +61 3 9029 3437  
 talk@arcitecta.com

[arcitecta.com](http://arcitecta.com)

**INDUSTRY**

Data Management Software

**SOLUTION**

Mediaflux

“Business continuity in today’s world starts with IT. Many organizations, regardless of their vertical, can only continue operations if their IT systems consistently remain available during any type of crisis.” says Jerome Wendt, Founder of the Data Center Intelligence Group. “Everything from hospital emergency room check-ins to power from utility companies to tellers in banks to burgers in fast food restaurants to you name it; they need IT all the time.”

Achieving continuous availability of applications and the data that drives operations is an extreme challenge that not even the largest cloud and SaaS providers have overcome. Therefore, every organization needs the ability to recover data and applications promptly when the inevitable failure occurs.

The data environments of many businesses have become so large and complex that their backup infrastructures can no longer protect all their data or meet their recovery time requirements. Consequently, those businesses are at significant risk of harm through data loss and service outages.

Unfortunately, many business leaders discover this lack only **after** they have suffered a data loss incident or service outage.

***“Many organizations can only continue operations if their IT systems consistently remain available during any type of crisis.”***

— **Jerome Wendt**  
*Founder of the Data Center Intelligence Group*

## Protecting All of an Organization’s Data is Complicated

Today, protecting all of an organization’s data can be very complicated. Data formerly used exclusively by staff during regular business hours now drives customer-facing applications with expectations of 24x7x365 availability. Enterprises have also increased the complexity of their data environments by adding data silos on-premises or in the cloud. Furthermore, the realities of a distributed



workforce may result in corporate data being stored on personal computers around the globe.

In today’s typical business environments, even knowing where all of a corporation’s data resides is a challenge, much less ensuring all that data is protected and recoverable.

## Backup is Broken at Scale

Whether or not a business has intentionally pursued digital transformation, many businesses can only continue operations if their IT systems consistently remain available.

As they expand the use of information technology and embrace new opportunities to create value from data, many businesses have crossed boundaries of capacity and complexity beyond which even well-designed and managed backup infrastructures can meet business data protection and availability requirements.

When I served as an IT director in higher education, our data environment grew to the point that some backups were no longer finishing inside the backup window (remember those?), and some backups began to take more than 24 hours to complete. This hurt the performance of our production systems, resulted in missed incremental backups, and increased the university’s risk of data loss and lengthy recoveries. Ultimately, the university had to upgrade its backup infrastructure to resolve these issues.

Many organizations have attempted to work around their backup problems by creating high availability (HA)

***“Many businesses have crossed boundaries of capacity and complexity beyond which even well-designed and managed backup infrastructures can meet business data protection and availability requirements.”***

infrastructures. A peer at another university that had outgrown its backup system installed identical storage arrays at two locations and then used synchronous replication between the arrays to enable the secondary location to take over in the event of a failure in the primary location.

However, even such environments can experience data loss and downtime. My friend’s nightmare began when a storage controller failed and corrupted substantial amounts of data on the primary array. The initial failover to the secondary location worked as planned. Then the storage array vendor replaced the failed controller and brought the primary array online. Unfortunately, a mistake by the vendor’s technician caused the newly repaired array to replicate the corrupted data to its twin at the secondary location. It was a mess.

Getting the university’s systems online again required the IT staff to restore from tape backups. The restore process succeeded, but the university lost access to its core business suite (recruiting, admissions, registration, financial aid, student billing, accounts receivable, accounts payable, and fundraising) for more than 48 hours.



**The Ideal Recovery Time is Zero**

The ideal for IT systems is continuous availability, but it is nearly impossible for systems to be available 100% of the time. Given that some disruptions are unavoidable, the ideal time to recovery and the amount of data lost is zero.

Traditional approaches to backup have never come close to achieving these ideals. It has been either technically impossible or more costly to achieve than it is worth. Thus, business leaders have had to accept some data loss, as defined by recovery point objectives (RPO), and some downtime, as defined by recovery time objectives (RTO.)

The example cited above involved less than 100TB of data. Today, many organizations are storing 10x or 100x that capacity. They have crossed the boundary beyond which the compromises associated with traditional backups are tenable.

Money math is driving changes. The cost of data loss and downtime is increasing, even as capacity growth and a worsening threat environment make extended downtimes more likely. Responsible IT professionals are exploring alternative approaches and solutions.

**We Need to Change our Thinking**

Happily, new approaches to data protection make it possible to approach the ideal of zero RPO and zero RTO, even at scale. But getting these results requires more than a rearchitecting of enterprise backup infrastructures. It requires a change of thinking:

- from data storage to data management
- from backup and restore to rapid recovery of data and applications to an operational state

Old Approach	New Approach
Data Storage	Data Management
Backup and Restore	Rapid Recovery to Operational State
Expand Backup Infrastructure	Alternative Solutions

Achieving data resiliency at scale also means exploring alternative approaches and solutions that are designed to manage:

- billions of files or objects
- petabytes or exabytes of data
- data spread across multiple storage systems and clouds

***“New approaches to data protection make it possible to approach the ideal of zero RPO and zero RTO, even at scale.”***

Beyond rapid recovery, the next-generation enterprise data platform must enable the business to derive more value from data while reducing business risk. To meet these requirements, the solution must feature the following:

- a global namespace enabling global views of the organization’s data, with identity and policy-based access controls
- data provenance
- global, distributed access that aligns with the needs of a distributed workforce
- intelligent data placement and movement so that the data is in the right place at the right time and at the right cost
- multi-factor authentication, including approval workflows for administrative actions
- multi-protocol data access to enable the data to be processed by any application
- policy-based data movement controls, including support for sovereign data regulations
- self-service data management tools for end users, application owners, and developers that enables them to continue working without interruption and frees IT from routine data recovery tasks



## Achieving Data Resilience at Scale Requires a New Approach

Achieving data resilience at scale requires a new approach. IT leaders must change their focus from successful backups to successful recoveries. Beyond meeting

recovery time requirements, a truly scalable solution must also enable effective self-service for its data users, automated data placement for optimal business results, and enterprise-class data governance.

The time has come to explore alternative approaches and solutions that are designed to manage data at scale.

Arcitecta is addressing the challenges of data resiliency and effective data management at scale. Its Mediaflux product is based on an XML-encoded object database (XODB), initially developed by Arcitecta to meet the data management requirements of research computing organizations. XODB is hyperscale database technology that integrates and operates on rich, arbitrary metadata.

Arcitecta and Mediaflux are proven in demanding at-scale data environments. Jason Lohrey founded Arcitecta in 1998. Arcitecta serves more than 300,000 active users at organizations in the life sciences, research, clinical, geospatial, media and entertainment, and defense industries. Mediaflux users generate multiple petabytes of new data each month and move many petabytes of data for analysis each month... with no data loss or outages exceeding 3 minutes. ■

### About Arcitecta

*Arcitecta is a creative and innovative data management software company. Founded in 1998, Arcitecta builds the world's best data management platforms, enabling thousands of users worldwide in some of the most demanding data-driven environments. Arcitecta's flagship Mediaflux platform began with the vision to provide organizations with extraordinary technology for handling all forms of data, from small to very large and complex. Today, it forms the foundation for managing the simplest and the most complex data for all sizes of organizations and global enterprises, empowering them to simplify data-intensive workflows and accelerate time to insight from their data to improve business and research outcomes*

### About DCIG

The Data Center Intelligence Group (DCIG) empowers the IT industry with actionable analysis. DCIG analysts provide informed third-party analysis of various cloud, data protection, and data storage technologies. DCIG independently develops licensed content in the form of TOP 5 Reports and Solution Profiles. Please visit [www.dcig.com](http://www.dcig.com).