

10 Signs You Need a Big Data Retention Solution

Overview

What constitutes Big Data is relative, and varies by organization. For a large enterprise, Big Data may be in the petabytes or more, while for a small or mid-size enterprise, data volumes that grow into tens of terabytes can become problematic. To deal with this, a lot of investment has been made in solutions that handle unstructured data elements such as archival of emails and other documents, and from a physical storage perspective millions has been poured into binary compression and de-duplication technologies. Meanwhile much of the focus on structured data or semi-structured data has been primarily on Big Data analytics. The reality, driven by more stringent legislation, governance and extended on-demand accessibility to historical data, is that structured data retention is now fast becoming the #1 imperative for businesses worldwide.

This document aims to provide some guidelines for organizations who are currently struggling with large growing data sets that need to be retained in the most cost effective way and at the same time meet demanding business needs to access that data at any point in time. It highlights why traditional online transaction processing (OLTP) and online analytic processing (OLAP) offerings are not the answer, and shows how specialized online data retention (OLDR) solutions provide the best capabilities and economics for retaining Big Data.

10. You agonize over when to keep or purge data

Traditional RDBMS or analytics repositories do not have any features to support the enforcement of retention or expiry policies dictated by industry regulations or business process data governance. As such, you may experience sleepless nights with trying to balance the ongoing cost of storage to retain growing volumes of data, fines through non-compliance and unnecessary liabilities due to holding data beyond legal expiry timeframes.

An OLDR solution can complement your existing OLTP and OLAP systems, providing you with reduced physical structured data storage (40 to 1 or more) through data value de-duplication, and offering built-in configurable rules to enforce retention policies, allowing you to rest easy without counting storage arrays.

9. Your data volumes exceed human comprehension

Systems that track human-generated activity such as records of every medical interaction, stock trades, call data records, webpage clicks; and direct “machine generated” data such as IT log files, barcode scans, RFID tag reads, GPS location entries, industrial automation control and environmental sensor outputs are examples of data sets that once “transacted” don’t change. The update capabilities of traditional systems are overkill for this type of data, and it can be a major IT headache to have to continuously add expensive hardware and memory to keep up.

An OLDR solution can be your primary repository for immediately historical data, specializing in ingesting and storing billions of records per day on low cost commodity servers, allowing you to avoid serious hardware headaches.

8. Your production database arteries are clogged

If your production application system diagnosed as constantly having “performance pains”, it is likely that the RDBMS holding your data is suffering from excessive volumes, resulting in unhealthy enlarged indexes. Studies show that a significant proportion (up to 90%) of a production application’s data set doesn’t require constant updates and is therefore static in nature. A best practice approach to solving this would be to put your production system on a data “diet”.

An OLDR solution can be a complementary repository that holds limitless volumes of static data while providing continued on-demand accessibility, allowing you to benefit from a slimmed down and ongoing healthy production database and application.

7. You’re addicted to hardware purchases

You wished your local hardware shop carried fibre connected SAN’s in 12 petabyte packs and you’ve added more memory than you can remember. If this sounds like you then you may be addicted to using hardware to compensate for the deteriorating performance of your critical production OLTP and OLAP systems caused by growing data volumes. If only you could get this growing enterprise carbon footprint under control.

An OLDR solution can hold massive volumes on low cost commodity servers, allowing you to defer costly hardware or memory upgrades for your production systems, thereby kicking your hardware addiction.

6. You have more DB admin specialists than end-users

While this is a highly unlikely IT doomsday scenario, it is indicative of a trend towards increasing numbers of specialized administrators needed to support, tune, backup, migrate and manage systems across a wide range of heterogeneous repositories that are likely doubling in size each year. Aside from the shrinking time window to backup large volumes, migrate to new application versions and arcane processes to move data offline to tape, the cost of highly skilled DBA's constitute a major portion of the total cost of ownership (TCO) per terabyte of data retained.

An OLDR solution is a low to zero administration data repository, allowing you to allocate next to zero specialized resources to big data retention.

5. You've heard of Hadoop and want to vote NoSQL

With major web properties handling Google-scale data volumes, Hadoop and MapReduce have been popularized as open source saviors to the Big Data problem. Additionally NoSQL is a grouping of various technologies that provide alternatives to traditional relational repositories. Many of these technologies are early and relatively immature in their enterprise business use. There are vendors offering commercial variants of these technologies today but focused on high performance analytics, not retaining and keeping Big Data online accessible and immutable.

An OLDR solution is specialized for enterprise and business-class structured data retention, and makes use of many of the concepts found in techniques such as MapReduce. This allows you to solve your retention problem today while benefiting from leveraging new techniques and technologies in the future.

4. You believe old dogs don't have to learn new tricks

The common, yet cyclical view of IT's evolution says that everything old is new again. Many of the technologies such as Hadoop and NoSQL have roots that can be traced back to days prior to commercial RDBMS offerings. It is a costly and time consuming endeavor to retool and train a new wave of experts in order to leverage these new technologies. The reality is that a large majority of the enterprise business world uses SQL for data access and the learning of "new tricks" are viewed as an added integration and roll-out expense.

An OLDR solution "speaks" SQL and supports standard ODBC/JDBC, allowing you to rapidly integrate it into your architecture by returning the exact same results from queries and reports requested through your application and BI tools.

3. Your head is in the Clouds, your body on-premise

You've been asked to consider Cloud options. Should you leverage data retention through a hosted offering, public, private or hybrid cloud? You'd like to deploy an offering that solves data retention problems today without precluding a hosted or cloud strategy in the future.

An OLDR solution works equally as well on-premise or in any form of hosted/cloud environment, allowing you to deploy what makes the most sense for your business.

2. You don't want to go through this every year

Not having to say "sorry back to the drawing board" means an architectural solution that will scale indefinitely with your data volumes. Whatever your retention requirements today, you should plan for multiples of those volumes in the near future. Predicting the TCO of any solution you deploy today is key to managing for the unexpected.

An OLDR solution scales proportionately with your data volume growth, it can parallelize ingestion and query, allowing you incremental scale as volumes dictate, while continuing to benefit from attractive economics.

1. You want to retain more AND pay less

Ultimately you want to view the Big Data retention problem as an opportunity, not a challenge. You want the highest possible de-duplication and compression, the ability to avoid fines and liabilities through configurable retention policies. You want healthy production systems without the dependency on costly hardware upgrades or additional DB administrators. You would rather not risk leveraging immature open source technologies, and fast integration, compatibility and time to market are critical for your business. You're testing the waters with Cloud architecture, and feel that it is somewhere in your future. Finally you must benefit from the lowest possible TCO per terabyte retained.

An OLDR solution can offer all of the above with concrete ROI and proof points, allowing you to justify savings and/or new business opportunities.

The Big Data Retention Calculator

The benefits of an OLDR solution can be quantified through the Big Data retention calculator. The calculator allows you to customize inputs based upon your current costs and anticipated data growth. It provides default values reflecting market/industry statistics as well as feedback from organizations that have benefited from online data retention solution deployments. Areas covered typically include:

1. **Data Reduction (Storage) Savings** - Reducing production storage and moving static data to lower cost tiered storage while maintaining online access
2. **Infrastructure** - Reducing table and index sizes which frees up RAM and improves performance/response times without having to upgrade
3. **Administration & Maintenance** - Smaller production databases reduce specialist admin time for scale-up, backup, test/QA sync, application migrations
4. **Compliance, Fines, Reduced Liability & Efficiency** - Configurable retention rules guarantees compliance, eliminates fines and removes liability through auto expiry and delete.

The result is a detailed executive report highlighting savings and benefits of deploying an OLDR solution (section from a sample report below).

Realize Your Data Retention Savings?

Run your calculations and determine how significant cost savings can be realized by deploying RainStor as the unique specialized data repository for your leading industry solutions. RainStor is extremely flexible and can be deployed as a primary repository for immediately historical high volume data or as a secondary, augmented repository to manage the historical data that continues to grow within your business applications. To use the calculator and learn how you can save significantly, please contact: 10xless@rainstor.com and our business development representative will contact you to set-up a call and run the calculations.

About RainStor:

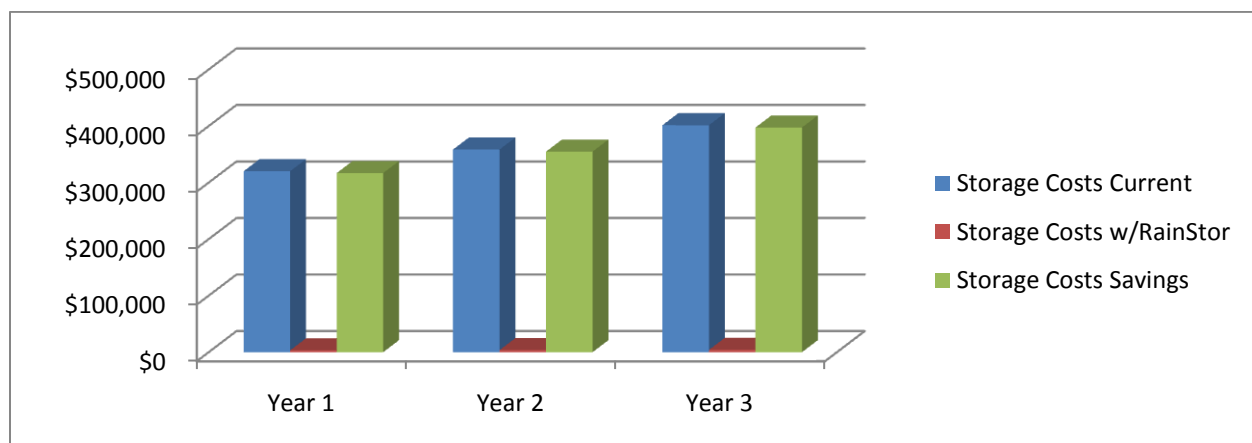
RainStor is a technology pioneer in online information preservation whose specialized data repository significantly reduces the total cost of retaining data through extreme data reduction, simplified data management and near-perfect scalability on commodity hardware. RainStor solutions are deployed by technology partners across multiple industries and include Informatica, AdaptiveMobile, Group2000 and On Point Technology to reduce the cost and complexity of preserving information in the enterprise or the cloud. By embedding RainStor in industry leading solution offerings, customers achieve a 10x less total cost of ownership for long term data retention. RainStor is a privately held company with offices in San Francisco, USA and Gloucester, UK. For more information, visit www.rainstor.com or info@rainstor.com

Sample Calculation:

| | Year 1 | Year 2 | Year 3 | Total |
|--------------------------------------|------------------|------------------|------------------|--------------------|
| 1. Data Reduction Savings: | \$316,800 | \$354,816 | \$397,394 | \$1,069,010 |
| 2. Infrastructure Savings: | \$180,000 | \$158,400 | \$139,392 | \$477,792 |
| 3. Admin & Maintenance Savings: | \$197,000 | \$216,700 | \$238,370 | \$652,070 |
| 4. Compliance & Efficiency Benefits: | \$75,000 | \$82,500 | \$90,750 | \$248,250 |
| | <u>\$768,800</u> | <u>\$812,416</u> | <u>\$865,906</u> | <u>\$2,447,122</u> |

Data Reduction Savings:

Taking a closer look at the data reduction savings, a sample company has a 5 TB production database with a total of 20 TB when all supporting production databases are included. 80% of the total data set can be moved out of production and into RainStor (16 TB). Assuming an achieved compression ratio of 98%, RainStor would store all this data in 0.32 TB of lower cost commodity storage, while still keeping this data online and accessible factoring in a drop of storage costs of 30% per year countered by 60% data growth per year. The table below just shows the storage cost savings by using RainStor. Of course depending on the solution requirements, further savings are seen around infrastructure, admin and resource maintenance and avoidance of fines for compliance.



Conclusion

The 10 signs listed above were developed based on the experiences of leading organizations that have successfully evaluated and deployed OLDR solutions.

RainStor provides an OLDR repository available exclusively through technology partners. RainStor has been deployed within partner applications at over 100 major organizations worldwide.

If you are an Independent software vendor (ISV), Managed Service Provider (MSP), or System Integrator (SI) with customers facing data retention challenges, the Powered by RainStor program provides you with the tools and support to rapidly integrate RainStor into your offerings. When you join the program, you receive a software development kit (SDK), training, and consulting to minimize development effort and accelerate time to revenue.

“Using the SDK, our developers were able to embed RainStor within the Message Retention module in a matter of weeks with minimal support. The ability for us to bring this new service to market so quickly was an important driver in our buy-versus-build decision when evaluating whether to OEM the RainStor software.”

Gareth Maclachlan, Vice President of Products at AdaptiveMobile, embeds RainStor in their application that retains tens of thousands of SMS /MMS records per second for communication service providers.

RainStor's Big Data Retention Calculator allows partners to quickly assess if an OLDR solution can significantly reduce costs to improve margins or generate new revenue opportunities. If you would like to create a custom report for your organization to see how RainStor could benefit you, please email us at info@rainstor.com or call us at +1 415-402-0110 (US) or +44(0)845 456 3590 (Europe).