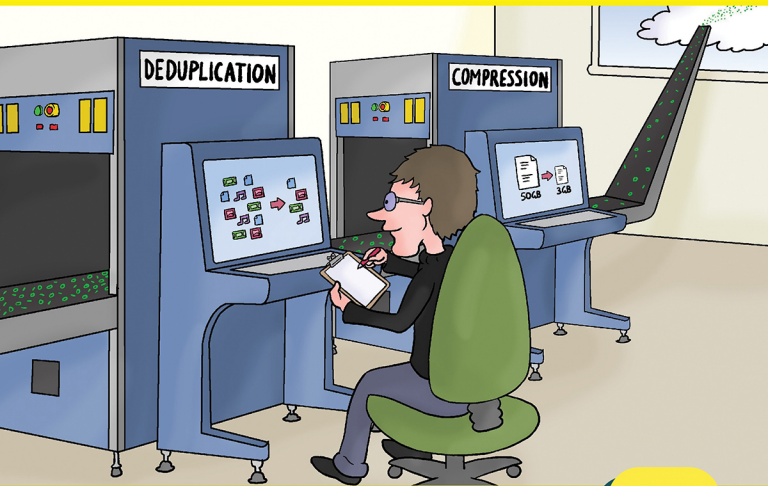


Conversational Cloud Storage and Data Security Optimization

By Greg Altman (Enterprise OCI Architect)



Learn about:

- Why backing up data to the cloud is a continually growing challenge
- Ways to optimize backups to increase speed, lower costs and ensure data security

MINI
Edition

Sponsored by

Quest

Sponsored by Quest®

Quest Software delivers cybersecurity, data performance, directory and identity, and modernization solutions that simplify enterprise IT management. Quest solutions protect and empower users and data, streamline IT operations and harden cybersecurity from the inside out. Quest has the experience to help you manage and monitor your databases, take control of your hybrid environments and secure them all with security solutions that ensure you are ready for the next threat.

We're Quest – Where Next Is Now.

The Quest logo is displayed in a bold, orange, sans-serif font. The letter 'Q' is significantly larger than the other letters. A registered trademark symbol (®) is positioned to the upper right of the 't'.

For more information, visit www.quest.com.

Conversational Cloud Storage and Data Security Optimization (Mini Edition)

by Greg Altman

© 2024 Conversational Geek



ConversationalGeek®

Conversational Cloud Storage and Data Security Optimization (Mini Edition)

Published by Conversational Geek® Inc.

www.ConversationalGeek.com

All rights reserved. No part of this book shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher. No patent liability is assumed with respect to the use of the information contained herein. Although every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions. Nor is any liability assumed for damages resulting from the use of the information contained herein.

Trademarks

Conversational Geek, the Conversational Geek logo and J. the Geek are trademarks of Conversational Geek®. All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. We cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Warning and Disclaimer

Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied. The information provided is on an “as is” basis. The author and the publisher shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or programs accompanying it.

Additional Information

For general information on our other products and services, or how to create a custom Conversational Geek book for your business or organization, please visit our website at www.ConversationalGeek.com.

Publisher Acknowledgments

All of the folks responsible for the creation of this book:

Author:	Greg Altman
Project/Copy Editor:	Ian Whiteling
Content Reviewer(s):	Blathnaid O'Brien Adrian Moir

The “Conversational” Method

We have two objectives when we create a “Conversational” book. First, to make sure it’s written in a conversational tone so that it’s fun and easy to read. Second, to make sure you, the reader, can immediately take what you read and include it into your own conversations (personal or business-focused) with confidence.

“Geek in the Mirror” Boxes

We infuse humor and insight into our books through both cartoons and light banter from the author. When you see one of these boxes it’s the author stepping outside the dialog to speak directly to you. It might be an anecdote; it might be a personal experience.



Within these boxes I can share just about anything on the subject at hand. Read 'em!

Optimizing and Securing Your Cloud Storage Backups



“I need you to back all that up into this please.”

Data, the invaluable asset of business in the 21st century, is surging at an exponential rate due to digitization. This rapid increase in data volume often leaves businesses struggling to keep pace with the need to securely and effectively back up all of their data. The total global data storage is projected to

surpass 200 zettabytes by 2025¹, encompassing data on public and private clouds, utilities, personal devices like PCs and phones, and IoT (Internet of Things) devices.

To put this into perspective, a zettabyte is a one followed by 70 zeros or a billion terabytes. If you stacked a billion one-terabyte hard drives, the stack would be almost 16,000 miles high – 15,782 miles high, to be precise. I'm assuming your backup jobs aren't quite this scale, but they're still big enough to be a problem.

The data avalanche is not just a challenge in terms of volume. It's a complex issue. It's slowing down backup processes and escalating storage costs. If you've been in the technology game for more than a week, you've probably heard all of the old advice about backups. "Keep everything; storage is cheap," they said. But that's an oversimplification, to put it mildly. If it was as simple as plugging in a tape drive and writing a script to make a copy of your data, the

¹ The World Will Store 200 Zettabytes Of Data By 2025
(cybersecurityventures.com)

global data backup and recovery market wouldn't be exploding now, would it?



Globally, the data backup and recovery market grew from \$12.18 billion in 2022 to \$14.15 billion in 2023. The future growth is expected to reach \$23.64 billion by 2027².

That's real money where I come from.

In the modern era, the challenges of data backup have evolved. We no longer just worry about hardware failures or human error. Today's interconnected world presents a much more complex landscape for our data, with the threat of humans accidentally reformatting data drives or deleting files, hardware failures, ransomware attacks, and natural disasters. This complexity underscores the importance of strategic data backup and recovery.

² Forecast: Worldwide Trends in Data Restoration, 2023 (reportlinker.com)

The Data Backup Challenge

There are several aspects of the challenge.

1. Not only do you need to keep a backup copy of your business data, you really should have three copies! Have you heard of the 3-2-1 rule of backups? That is three copies of your data on two different media types, and one should be offsite. There is a new addition that adds “immutable” to that mantra.
2. This backup has to be recoverable as fast as possible. Recovery Time Objective (RTO) is a crucial data backup and recovery metric. It is the maximum time an application or process can be unavailable before significantly damaging the business. TechTarget’s Enterprise Strategy Group’s research indicates that 57% of respondents reported that they cannot tolerate the loss of mission-critical applications for even one

hour, and 15% said their organizations cannot tolerate any downtime, ever!³

3. Sadly, your backup and storage budget will not increase at the same rate as your data growth. Whilst backups are essential, they do not generate revenue, unless you work for a backup company.

So, to recap – Keep three backup copies securely, be able to restore them quickly, and keep costs down. No problem, right?

To meet all of these needs, let's discuss cloud storage. "But cloud storage is too expensive!" Well, it doesn't have to be prohibitively costly if properly optimized. This is where deduplication and compression come into play.

³ Enterprise Strategy Group Research Report, Real-world SLAs and Availability Requirements, October 2020

Optimizing Backup Data = Optimized Backups

How do you optimize backups so that you can meet the goals mentioned? There are three main methods you should explore:

- **Deduplication** – This technology reduces the size of your backup by eliminating redundant data and storing only unique data sets.
- **Compression** – This reduces backup size by encoding the data using fewer bits than the source data being backed up.
- **Storage Tiering** – A strategy of optimizing the use of different cloud storage classes to optimize spending.

Let's examine each of these methods a bit deeper to see how to use them and where they yield the most benefit in the backup process.

Data Deduplication

Imagine your data was a physical library of books. Some of the books are duplicates, like the spreadsheet that everyone has a copy of. These duplicates take up unnecessary space. Data deduplication is a way to clear the clutter. There are two main methods of deduplication: basic block and variable block:

- **Basic Block Deduplication** – In this method, data is broken into smaller equal-sized blocks (usually 4KB to 128KB) that are then compared to each other. The process begins by scanning these blocks to identify duplicates. If two or more blocks are identical, only one instance is retained, and subsequent occurrences are replaced with pointers or references to the original block. Basic block data works well with structured data where the data has the same ‘size.’ Depending on how it’s implemented and the data being deduplicated, deduplication can save space into the 90%+ range.

- **Variable Block Deduplication** – This method takes a more adaptive approach. Algorithms determine the optimal size for each block. If the chunk of data is unique, it gets its own block. If it's similar to existing blocks, it shares a block. This can lead to more efficient storage utilization as some data needs big blocks, and some will fit into smaller blocks. Variable Block Storage really shines with unstructured data because it can adapt to the data's natural patterns.

The percentage of savings will vary between deduplication technologies, but a good rule of thumb on the space-saving is:

File < Fixed-Block < Variable-Block

Because deduplication can be processor-intensive, how you approach the solution can significantly impact your time to back up and recover. Here are a few decision points to consider while planning:

- **Source or Destination?** – This one is pretty straightforward. Do you deduplicate the data before you send it to the cloud (source) or after it's uploaded (destination)? There are two very good reasons to do source-side deduplication. The biggest is that system performance (CPU, Memory, etc.) improves faster than networking speed, available bandwidth, and latency. Nearly as important is the “ingest” or upload cost of your cloud storage. Simply put, it costs less to upload one terabyte of data than 10 terabytes.
- **Job vs. global level** – Is your deduplication “comparison set” restricted to a single job, or is it global across all of your backup jobs? The larger the sample, the more effective the deduplication, so it's generally considered best practice to go with the global approach.
- **Software Defined vs Appliance-Based** – It's true that most backup software offers some

deduplication, but most organizations prefer a dedicated deduplication solution. These usually are either hardware-based or virtual appliances. Hardware appliances bring their own complications like hardware lifespan, warranties, power consumption, rack space, etc. Virtual appliances run like a VM in your virtual environment, either on-prem, in the cloud, or both. This approach allows you to tune the CPU and RAM allocation to achieve the data deduplication performance needed for your environment.



Don't confuse incremental backups with deduplication. Incremental backups only reduce the size of the existing job; deduplication reduces the size of every job.

Compression

This one has been around for a long time as a way to squeeze more data onto a disk. In a backup to the cloud scenario, it's a good idea to compress the data before it's sent to the cloud for many of the same reasons to deduplicate before doing so. Specifically, the less data that has to traverse the network, the faster your restores will be. However, a balancing act needs to happen as compression is also processor-intensive. Attempting too much compression will introduce latency into your backup and recovery process.



Deduplication and compression can be used together!

It's like packing for a trip. You don't take three of the same shirts (dedupe) and then put the shirts you do take into a vacuum packing cube (compression).

For reference, compression ratios under 1:10 (compressed size vs original size) are considered good performance. Different types of data will compress differently, largely depending on whether the file type has compression built in,

like JPG files. Anything better than 1:10 (like 1:12) is considered excellent.

Tier the Storage!

Both deduplication and compression methods can be applied to on-prem or cloud-based sources and targets. However, cloud storage has other features that you may or may not be fully utilizing.

Every major cloud provider offers several tiers of storage. These tiers are usually designed around how quickly you need the first byte of data during a recovery operation. Recent critical backups are needed instantly (or as close as possible), while recovering something from several years ago can likely wait a bit for the restoration process to begin. Other factors that come into play between the different tiers are availability, storage charges per object, and storage duration minimums. Generally, as you move toward the slower, less easily accessible tiers, the price per GB drops significantly.

So why not always use the cheapest tier? Think back to RTO (Recovery Time Objective). Remember how we wanted that number to be measured in

minutes? The coldest tiers of cloud storage have the time to recover the first byte of data listed as “hours” – one or two, eight or even more.

A better solution is to work the tiers with a policy-based automated method of moving older backups to slower, colder tiers to reduce costs. This can be extended to a hybrid solution where backups are kept locally for 48 hours, then moved to the fast (expensive) cloud tier, where it stays for 30 days. Then, it’s moved to the middle tier for three to six months, after which it moves to the lowest, cheapest tier for long-term storage. Don’t forget that you can always pull from any tier; the difference is how fast you can access the data. The balancing act of cost versus speed is unique to every organization and to even different data sets within an organization.

Pulling It All Together – A Practical Backup

First, take your business objectives and translate them into technical requirements in order to get the most from your backup plan. Remember the two goals from earlier:

1. Be able to recover any part of the environment as quickly as possible.
2. Do so while keeping costs as low as possible.

This is the balancing act we mentioned a bit ago.

So how do you do this?

An excellent place to start is with the things you can likely do immediately for little or no cost, then move into more sophisticated solutions that may cost money or take time to implement.

- 1. Update Your Data Retention Policy** – I'm assuming that you have written retention policies dictating how long backups of certain workloads or data types should be kept and available. Do your policies specify a time frame for retrieval? They should. This ensures that expectations are aligned between IT and the rest of the business. Nothing ruins an evening like a 5pm request from Legal to have all of the documents

related to a job or order from five years ago available first thing in the morning.

2. **Use Tiered Storage** – Research and understand the various storage tiers offered by your cloud provider. Know the different costs per GB, the first-byte latency, and what minimum storage durations are, if any. Then, map this to your various workloads and data held in your backup jobs. Add this to your data retention policy so your organization has a well-defined strategy, not just about how long to keep backups but what storage tiers are used over time.

Of course, to properly use storage tiers, you must move data between them. The worst-case scenario is to use the cloud provider's native tools to manually do the moves. This is tedious but do it anyway. Older data stored "just in case" and unlikely to be needed at a moment's notice does not need to be in the most expensive "fast" tier.

Some cloud providers have storage management tools to automate this process

somewhat, which is a solid option. Most enterprise backup solutions have some tiered storage management capabilities built in. If yours has the ability – use it!

3. **Dedupe the Data** – Some backup solutions have deduplication included or as an add-on feature. This is a start, but understand that there may be limits on how much deduplication you can achieve. For example, if your backup dedupes at the VM level and you have 100 servers built from the same base image, there is an opportunity for massive savings by deduplicating their common parts. Unfortunately, that's not always the case, and usually, the OS part of your servers isn't where the bulk of your data lies. But some dedupe is better than no dedupe, so read the documentation of your backup software carefully and understand the details of how any deduplication features work. Remember, variable block deduplication will be the best option, but any deduplication functionality will help so

long as it balances overall backup and recovery performance.

4. **Compress the Data** – This can be a bit of a sticking point. Most backup solutions offer some compression, but this is not their core competency. The core focus is on backing up and recovering data. That means that while compression may be offered, it's not likely the best. The big downside to compression is how CPU-intensive it is and the diminishing returns of chasing more and more compression. In some cases, achieving even 10% more compression results in 10x CPU usage! So yes, use the compression, but understand you won't get the full benefits of what can be accomplished with backup solution-based compression. The answer to this problem is to move the compression out of the backup software to a third-party storage solution. Doing this lets you choose the best software for backups and still reap

the rewards of a dedicated storage optimization solution.

5. **Protect the Data** – Backup data is the most common entry point for ransomware attacks. While you are ensuring data is deduplicated, compressed, and organised, attackers are looking for a way in. You need to organise your protection perimeters to block this path by considering implementing the following techniques:
 - **Anomaly Detection** – Detect unusual deviations in backup data during backup, such as file sizes, excessive deletions, etc.
 - **Data Immutability** – This prevents the deletion or altering of your backups by an attacker.
 - **Encryption** – Actively prevent an attacker from being able to read your data – in flight and at rest.
6. **All of the Above** – This is where you can really reduce cost while increasing speed

and ensuring protection. Just remember the old saying about “Jack of all trades, but master of none.” Today's best backup software can't be the best at backup and recovery *and* at compression, deduplication, ransomware protection and tiered storage management. So, what's the answer?

Consider looking for a solution that complements your backup solution and offers:

- Best-in-breed compression and dedupe,
- Ransomware protection
- Anomaly Detection, and
- Cloud tier management

This allows you to optimize your backup strategy and ensure fast, accurate, secure backup and recovery.

It's going to mean spending some of your budget on a third-party solution, but that should be offset

by the savings in storage costs while gaining speed and security.

Use the best tool for each job rather than adding overhead to your backup solution.

The Big Takeaways

Let's face it: organizational data only grows and almost never shrinks. Additionally, the growth rate can increase exponentially. On the other hand, IT budgets do not grow at the same rate. Therefore, it's necessary to leverage optimization tools to increase the speed of backup and recovery while controlling costs and ensuring data is protected.

Organizations can get proactive by:

- Optimizing their backup strategy using deduplication and compression
- Ensuring they are using the least expensive cloud tier possible
- Leveraging the fastest rate of backup and recovery
- Protecting their data against ransomware attacks using cutting-edge technologies.

Quest® QoreStor®

Make your backup solution better!

Quest QoreStor will make whatever backup solution you use even better. Exponential data growth continues to be a major challenge but now you can fully optimize your backups and recover faster in both the data center and in the cloud.



Accelerate backup and recovery



Reduce backup storage requirements and cost



Connect to the cloud for backup, disaster recovery and long-term data retention



Gain ransomware protection via immutable backups and anomaly detection

[Learn More](#)

Quest
Where Next Meets Now.

The rate at which corporate data is growing is putting a strain on IT's ability to cost-effectively back it up and recover it quickly when using cloud storage. I'll explain what's the root cause of the problem and set out ways to optimize backups to lower storage costs, speed up the backup process and keep data secure.



About Greg Altman

Greg Altman has 30 years of experience in infrastructure management, certifications in Microsoft server and Azure, a blog at TheHoundTech.com, and two Iron Butt certificates for long distance motorcycle rides (so far).



ConversationalGeek®

For more content on topics geeks love, visit

conversationalgeek.com