

The IT Manager's Guide to Simplifying Microsoft® Outlook® Backup



A new approach to securely and reliably backing up .PST and other large files

About this paper

This guide covers the challenges of backing up Outlook .PST files for Windows® and Mac®, and how these challenges can be addressed using endpoint backup technology.

Backing up Outlook .PST files for Windows and Mac is a challenge for IT managers, and needs to be addressed with an endpoint backup solution instead of a generic backup tool.

IT managers in large enterprises are familiar with the rich feature set offered by Microsoft Outlook. One such feature is the ability to create local folders known as Personal Storage Tables (.PST), which archive email messages, contacts, and more within a user's Outlook setup. These .PST archive files can be used to restore or move Outlook data in the case of hardware failure, unexpected data loss, or when you need to transfer data from one computer to another in a hardware refresh.

Yet, these large files are among the most difficult types of data to manage in a large organization. A .PST file often exceeds 15 GB per user, which means increased storage management costs for IT administrators. It's common to store .PST files on network servers, but as file sizes increase, it becomes unwieldy for IT managers to move the files around on local or remote servers. The result is the file transfers clogging up the network and impeding end user productivity.

On top of this, employees may create their own .PST file archives for various reasons, such as to take their archived email from one job to another. These "ungoverned" .PST files pose a risk to companies that need to locate and audit archived email for eDiscovery and to address compliance requirements.

What's Unique About .PST files

More than 30% of laptops in a large company environment contain .PST files. These large, uncompressed files have minimal relative change on a daily basis, making them a great candidate for eliminating redundant data through deduplication techniques and compression. .PST files contain content common across the organization; for example, the same email message is sent to a dozen recipients, or the entire sales team gets the same PowerPoint® deck. With


so many shared emails, attachments, and address books, global deduplication at a message level can reduce storage requirements.

Outlook email users typically archive .PSTs on their local drives. .PST files are almost always open during a backup, unlike a lot of other files, so supporting open file backup and consistency is critical. In addition, .PST folders managed by end users are dynamic in nature. Some archives are stagnant and hardly change, while others are actively used to file all email to avoid using up a quota on the mail server.

Common Outlook .PST Challenges

The simplicity of creating a .PST archive on an end-user system is the very source of today's IT challenge of managing them. End users are highly motivated to keep their older email messages, especially when they are faced with storage quotas on how much they can store, or how long the company retains the data (90 days in many cases). As a result, end users use .PSTs as an archive solution, to the detriment of your bandwidth and storage space.

Yet, many companies don't want to keep .PST folders/files due to their cumbersome size and the liability of keeping data beyond a required period. To end users, saving .PST files locally sounds like a simple fix; but most IT support see it as a major headache. That's not only due to storage requirements, but also because for some companies, the "stale" data, invisible to IT and legal, creates a legal liability. Email scattered across multiple computers in thousands of files makes eDiscovery efforts much more complicated, time consuming, and expensive.

Did you know?  *Employees sometimes create their own .PST files, which they store on USB drives.*

More fundamentally, when end users or IT use .PSTs as an archiving solution, they run into limitations of how the .PST file itself was intended to be used. For example, .PST files were not designed to be used for archiving over a network. When used in this manner, file corruption is common. Corruption is also a strong possibility when compressing .PST files. And if the .PST file grows over 2GB, the file in some cases becomes unusable.

In other words, .PST files were not designed to be a long-term, continuous-use method of storing messages in an enterprise environment.

A New Approach: Using Endpoint Backup for Managing .PST Files

So what is an IT manager to do when faced with an ever-increasing set of email archives across thousands of end users, and the manual approach is no longer effective or safe?

One approach for simplifying .PST backup involves non-intrusive endpoint backup. A large majority of data backed up in enterprise organizations is Outlook .PST files; it makes sense to use specialized endpoint protection software like Druva inSync to centralize .PST file backup across thousands of laptops and endpoints.

While you may be tempted to use legacy products used to backup desktops for this purpose, they will lack advanced features (such as application-aware global deduplication) to address backups across endpoints, including laptops and mobile phones. They also make the needed frequent backup of .PST files unviable, increasing both backup storage and bandwidth costs multifold.

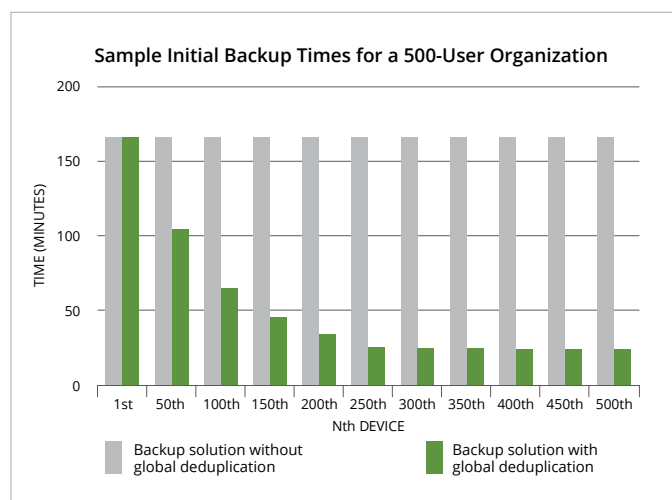
How does backing up .PST files with endpoint backup work? First, the endpoint client is deployed on every endpoint; each endpoint starts its backup but its data is deduplicated against every other endpoint in the process (decreasing the volume of data transferred for each additional endpoint). Then, periodic backups of .PST files are done by intelligently using Druva's advanced application-aware deduplication, backing up only

what's needed to handle increasing file sizes, and working across file types and bandwidths without impacting end user productivity.

Even better, endpoint backup solutions like Druva inSync solve the dilemma of .PST files scattered across the organization out of IT's visibility. Using federated search (built into products like inSync), IT staff can know where .PST files reside throughout the organization on any endpoint device.

How global deduplication impacts data transfer

Consider this real-world example of large-sized Outlook backup. A typical email system might contain 100 instances of the same 1 MB file attachment. If the email platform is backed up or archived, all 100 instances are saved, requiring 100 MB of storage space. With data deduplication, only one instance of the attachment is actually stored; each subsequent instance just references the one saved copy, reducing storage and bandwidth demand to only 1 MB. This reduction in backup size saves more than 99% of bandwidth, storage, and time.



Global Deduplication and its Impact on Data Transfer

And that's just with a tiny message. Imagine how that storage and networking need can snowball with an important large file, such as a sales presentation or a large video file.

This reduction in backup is achieved at two levels. First the system busies itself with the first backup activity, which it does across all available .PST files. The files are inspected at a granular level; then a powerful global deduplication algorithm is unleashed to capture non-repeated, or unique, blocks of data to form the first backup set.

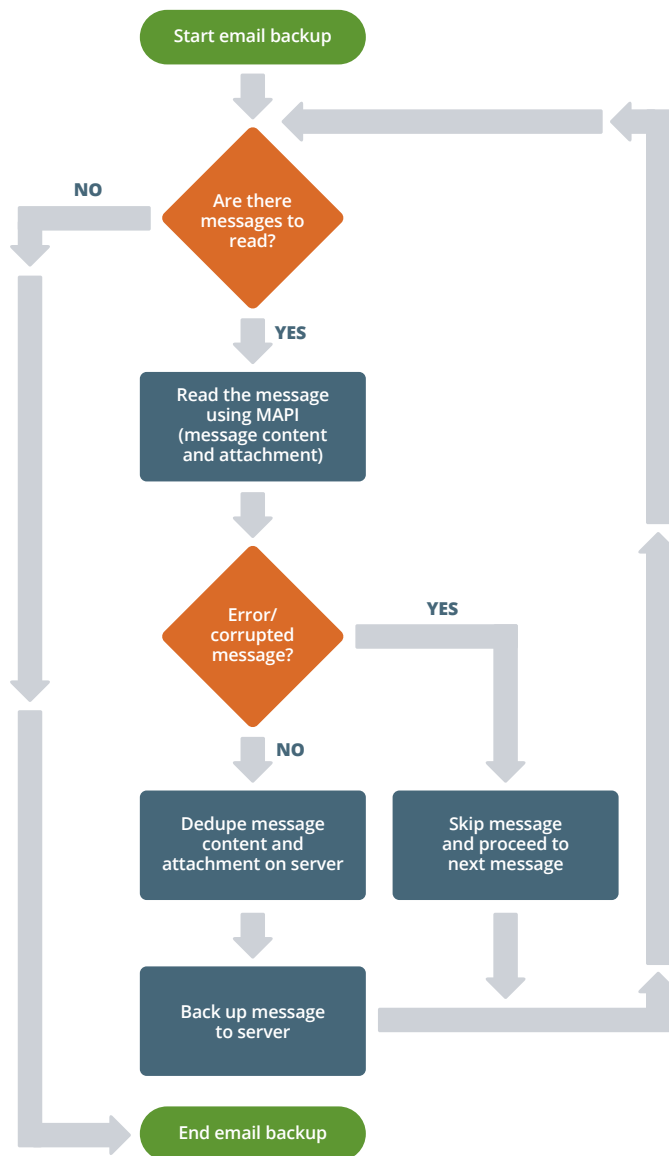
To transform .PST files into chunks of intelligent data, the endpoint backup client communicates utilizing the MAPI. Together with MAPI, Druva inSync accurately identifies structured email content within the files. What this means is that the .PST, hitherto a big (and ever-growing) blob of data, is transformed into meaningful content, including email headers, bodies, and attachments. For subsequent backups, Druva inSync queries MAPI to read freshly-added or updated email content from within the files. Again: Druva inSync tracks and protects only the unique blocks of email content.

Backing up Microsoft Outlook Email files Using MAPI

Let's look at how endpoint backup software such as Druva inSync works for backing up Outlook .PST files on Windows.

At its simplest: The backup client reads the .PST file and streams data to the backup server. That's easy enough. The challenge is doing this every day, or more frequently. If the user gets a few new email messages every hour, these need to be backed up regularly. However, it would not be advisable to back up that huge file every day, when hardly 1% of it is fresh, new data. How would one find the 1% delta in a 4GB file and transfer just those parts? Unless this problem is solved, backing up Outlook is an expensive proposition in terms of bandwidth (and nearly every other measure). In addition, a single email changes the checksum of the entire file, so a 4GB

.PST file over a week can add up to 28GB of transferred and stored data if new data is being added to it. By using MAPI, this problem can be circumvented.



Using MAPI to Backup Outlook .PST Files

This works because endpoint backup software like Druva inSync stores the checksums of file blocks locally on the client in a lightweight database. During incremental backups, the local client software scans the .PST file, looking at the updated messages, and identifies the changed blocks by referring to the checksums. Thus, the data that changed is identified at the client itself, and it transfers only this new data set to the server. This provides quick, efficient backup of large .PST files, making it feasible to back them up over a WAN. This works cross-platform, be it Windows, Mac, or Linux® endpoints.

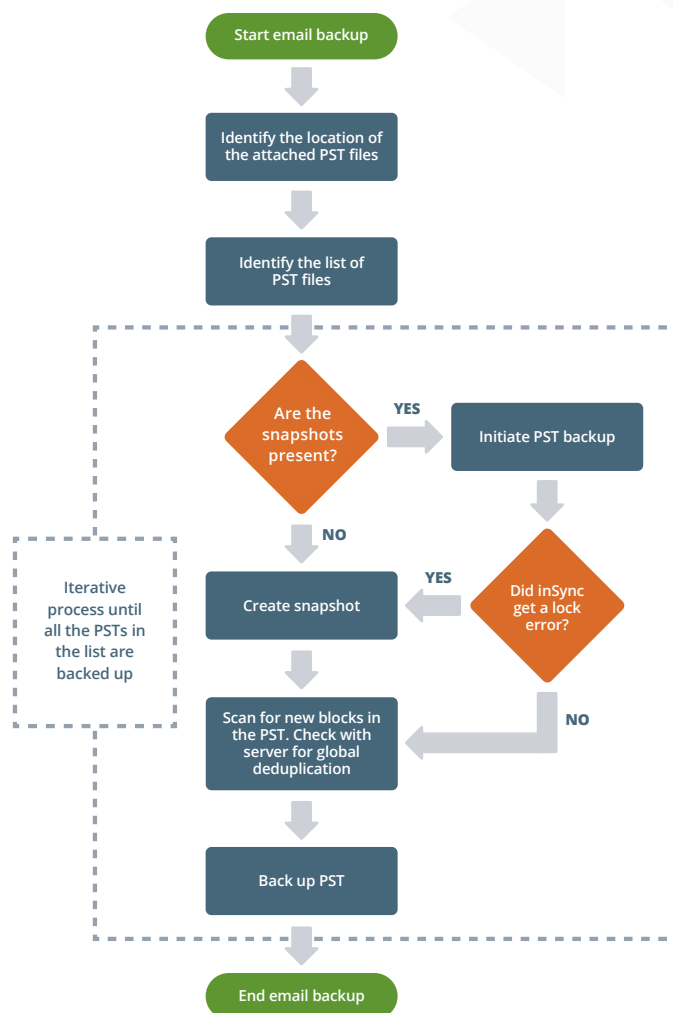
With application-aware deduplication, inSync backup Outlook data as proper email messages, not as blocks in a file. This makes it content-aware as far as email is concerned. MAPI brings structure to commonly unstructured and onerous data sets. This circumvents the ever-growing nature of .PST files.

Backing Up Microsoft Outlook Email Files on a Mac using FSEvents Framework

Microsoft Outlook is different on Mac OS. Instead of a Windows' single .PST file, on the Mac, Outlook stores email data in multiple small files. The number of files in this dataset can easily run into the hundreds of thousands. As a result, we have equally difficult datasets to handle.

The challenges are the same, however. Backing up this huge number of files once might be easily accomplished by copying them, one by one, to the backup server. The trick is in handling the incremental backups. Can we transfer so many files over the network every day when only 1% of the files change?

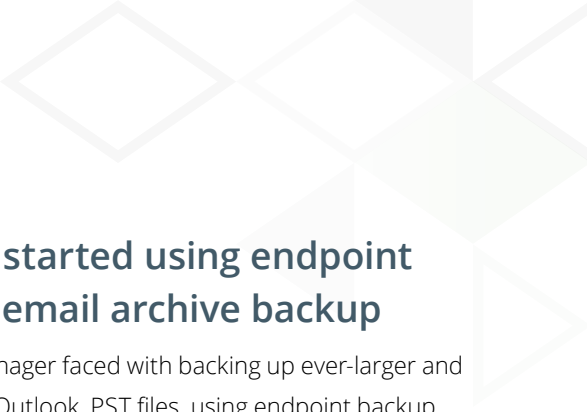
Druva inSync addresses this issue by integrating with Mac's FSEvents framework. The framework can identify files that changed since a previous checkpoint, and backs up only those files. This leads to quick incremental backups over a wide-area network, even with huge mailboxes.



Backing up Outlook .PST Files on a Mac

Deduplication across applications and other file types

While our focus here has been on Outlook, many other large files benefit from the techniques meant to ensure the data is regularly backed up without overloading servers or in-house connectivity. Faster backup due to less data being transferred and stored has obvious wins.



Since backup technologies like inSync understand the logical view of the data, it is much more efficient in discovering duplicates across applications and eradicating these than other deduplication based backup options. For example, Druva inSync can handle intelligent boundary slicing for deduplication for computer-aided design (.CAD) files, as well as large Microsoft Office® files or .PDFs. Each application stores and indexes its data differently. For example, consider an image file which is embedded in a Word® document as well as sent in a .PST file as an attachment. Block-based deduplication is often unable to identify such duplicates across applications, as the data itself has changed.

Choosing when to include or exclude .PST files from backups

When it comes to backing up .PST files, IT can decide which file extensions are excluded and included in the backup. For organizations with a no-.PST policy, the files can be excluded from the regular backup set, while also enabling search across all endpoints to find where they exist to aid in enforcing the policy. Of course, some organizations are fine with .PSTs; endpoint backup makes the backup and management of them efficient by leveraging deduplication at the block level, as well as allowing easy adjustments of file types in the data set. In other words, an IT manager can decide to initially exclude .PST files from backup, then add these file types in at a later date when other, perhaps more mission-critical, files sets are protected.

How to get started using endpoint backup for email archive backup

If you are an IT manager faced with backing up ever-larger and more voluminous Outlook .PST files, using endpoint backup to solve this may look pretty appealing. The first step, before you deploy endpoint backup software like Druva inSync, is to define your business need. Is the goal of improving .PST file backup to improve end user performance, or to lower overall capital expenses and save on storage costs? If storage costs are not the issue, but end-user uptime is, regular variable length deduplication can be used that compresses the .PST without regards to its content. If storage is a driving concern, the IT team can select application-aware deduplication using MAPI, which requires a longer transfer time, but delivers less data for storage and capital expense savings.

Once your business goals are clear, you are ready to plan the deployment, based on total number of sites, number of users, bandwidth limitations at sites, length of time to retain the data, and desired backup schedule.

With this new approach to .PST backup in place, it is now possible for IT managers to nimbly address the inevitable increase in email archive size within a growing organization, removing a once painful obstacle to efficiency and security, and keeping productivity at its peak.

About Druva inSync

Druva InSync software consists of four components: Backup, Share (file sync and share), Data Loss Prevention and Governance. Druva's inSync software is a storage and server agnostic platform for endpoints that provides secure backup, file sync and share/collaboration (optional – not included), Data Loss Prevention (optional – not included) and Governance (optional – not included).

To learn more, visit [Druva.com/insync](https://druva.com/insync).

Thank you to Anant Apte, Ashish Saxena and Amruta Phansalkar of Druva for their contributions to this paper. Microsoft®, Outlook®, PowerPoint®, Windows®, Office® and Word® are either registered trademarks or trademarks of Microsoft Corporation in the U.S. and other countries. Mac® is a registered trademark of Apple Inc. in the U.S. and other countries. Linux® is a registered trademark of Linus Torvalds in the U.S. and other countries.

About Druva

Druva is the global leader in Cloud Data Protection and Management, delivering the industry's first data management-as-a-service solution that aggregates data from endpoints, servers and cloud applications and leverages the public cloud to offer a single pane of glass to enable data protection, governance and intelligence—dramatically increasing the availability and visibility of business critical information, while reducing the risk, cost and complexity of managing and protecting it.

Druva's award-winning solutions intelligently collect data, and unify backup, disaster recovery, archival and governance capabilities onto a single, optimized data set. As the industry's fastest growing data protection provider, Druva is trusted by over 4,000 global organizations, and protects over 40 petabytes of data. Learn more at www.druva.com and join the conversation at twitter.com/druvainc.



Druva, Inc.

Americas: +1 888-248-4976

Europe: +44 (0) 203-750-9440

APJ: +919886120215

sales@druva.com

www.druva.com