



PST Migration/Elimination Considerations

Background

This document looks at the key areas and considerations of a PST Migration and Elimination project and compares the design, scalability and functionality of C2C's PST Enterprise solution with Microsoft's PST Capture utility.

Scenario

This document was written in early Feb 2012. At the time the Microsoft PST Capture utility had just been released and was at version 1.0. The C2C PST Enterprise solution had been released for many months and was at version 2.1.

All information regarding features and capabilities are based on these versions.

Executive Summary

The Microsoft PST Capture tool is best suited to small scale projects of up to 200 PST files where the organization is already using Exchange 2010 and Outlook 2010, and the PST files are located on a network file server, close to the Exchange infrastructure.

C2Cs PST Enterprise solution is a considerably more scalable option, with a well thought through feature set that allows administrators to comfortably manage a project of up to 5 million PST files. This solution allows automated granular control of the processes, allowing administrator to remove old or non-valuable data before importing, and fully manages the removal of the original data from the client computers after it has successfully been migrated into their mailboxes.

Copyright: C2C Systems, 2012

Disclaimer of Liability: While every precaution has been taken in the preparation of this document, C2C Systems assumes no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.



Introduction

PST migration projects are a very serious undertaking. From the 'reasoning' perspective the benefits are clear – risk reduction (for litigation), file contents corruption (common in older or large PST files) and simply loss of data (e.g. loss of the laptop they reside on).

From a project or operational perspective the implications are enormous. Not only are you likely to affect every user in your organization, you also impact all the key elements of your IT infrastructure including: network bandwidth, Exchange Servers, Storage, Firewalls, Home office users, Branch office users and more.

The risks of something going wrong can be relatively high, and the consequences of something going wrong are an administrator's nightmare – loss of email data, on a massive scale (i.e. affecting lots of users) or worse still perceived wanton destruction of email, something courts impose severe penalties for.

The main considerations, which need careful attention, for a PST migration / elimination project can be categorized as follows:

- Client Software / Agents (how to deploy and configure, if needed)
- Understand the scope (volume of data, number of files etc.)
- Bandwidth and Network implications
- Reducing the scope (removing older data if possible)
- End user experience (training, interaction required)
- Coupled and Uncoupled PSTs (those open in Outlook and those that have been 'lost' or 'orphaned')
- Scalability.

Client Software/Agents

PST Migration / Elimination solutions will generally utilize some kind of client component. There are some solutions that use Windows WMI and administrative shares, but these technologies do not normally work well in a corporate environment with firewalls etc.

The main consideration related to Client Software / Agents is how you get it deployed. Manually installation by end users themselves or IT staff visiting desktops (the method employed by the Microsoft PST Capture tool) is not a workable option for anything other than a few users. Having the software installed and packaged into the corporate workstation images is a better solution but also takes a long time to get fully rolled out.

Good solutions allow the client software to be installed by Group Policy or other standard software deployment tools, but the best solutions provide clients that require no software install, but instead simply run in memory – these can then be initiated automatically by a login script or manually by a desktop shortcut or start menu entry. PST Enterprise provides a network share from which clients can be executed (initiated by a login script or the like), making it the simplest and easiest to manage solution.



Understanding scope

After clients are deployed and the 'discovery' phase has started, data is gathered about the PST files in the organization.

It is imperative at this stage that the information available gives a full understanding of the scope of the problem. While the 'PST Size' (on disk) is a good measure, as it gives an understanding of how much data will be saved on workstations and file servers after the migration, it does not give an understanding of the actual volume of email data that needs to be migrated as the 'size on disk' and 'size of data' in a PST file can be vastly different, depending on many factors such as whitespace.

Only, when you have a good, clear understanding of the size of data, size on disk, number of PSTs, volume of PST data by user etc. can you really begin to project network utilization, project Exchange server storage requirements and to accurately plan timescales for your migration project.

The 'size on disk' provided by the Microsoft PST Capture tool only provides one angle of the scope, the additional data reported by PST Enterprise gives a much fuller view of the project scope.

Reducing scope (delete)

After gaining a good, accurate understanding of the scope, many organizations realize that all the PST data will simply not be able to be migrated. This could be for a number of reasons including project deadlines, bandwidth availability, Exchange storage infrastructure and many more.

At this stage it is often prudent to investigate how the scope can be reduced. This should be done even if there are no limitations – for example it is much better, quicker and more cost effective to delete in place rather than bringing all data back into Exchange only to delete it from there.

PST Enterprise provides the ability to delete data that exceeds the corporate retention policy (e.g. anything older than X years) or, to only migrate the last X years of data, or even to only migrate items in particular folders. Often organizations find this can reduce the scope of the project from migrating hundreds of terabytes to only a fraction of that, therefore saving time, costs and resources.

Bandwidth/network implications

The crux of any PST migration project is the implications it will have on the network bandwidth. Users have been accumulating PST data for many years and they will generally have a lot of data – multiple gigabytes. At some point some or all of that data will need to move across your network, and bear in mind that much of this data could be in Branch offices with considerably slower network links.

Both the PST Enterprise solution and the Microsoft PST Capture utility allow the migration to happen between certain times, but only PST Enterprise provides a range of other 'throttling' controls to ensure that the effect on network bandwidth is minimized and LAN, WAN and branch office links are not completely swamped with PST data to the exclusion of everything else.

'Throttling' controls available in PST Enterprise include configurable delays between sending messages, limiting the number of clients that transfer data at any one time* and limiting the amount of data transferred by each client in any one processing run*.

*Due in version 2.2 – March 12



End user experience

During any project which interacts with the end users workstations and/or data it is imperative that the end users productivity is not affected and that their normal working process is not changed. The last thing you want in a large scale project is hundreds or thousands of calls to the IT helpdesk demanding to know why their machine has 'hung' or why their data has 'disappeared'.

Both PST Enterprise and the Microsoft PST Capture tool take measures to ensure that end user productivity is not affected – with the Microsoft tool this is mostly around scheduling the processing to take place outside hours (as it cannot access PST files that the user has open in Outlook anyway), with PST Enterprise the strategies include ensuring the process utilization is kept low and ensuring the network bandwidth impact is also kept low and PST Enterprise can manage both open and closed PST files.

In terms of not confusing the end user, PST Enterprise takes steps to ensure that the data the user views inside Outlook is migrated into a highlighted, configured folder and that the original hierarchy etc. of the PST file(s) is maintained. PST Enterprise also allows for the removal and deletion of the original PST files/data after the migration is successfully completed – this is not something that is available in the Microsoft utility, which moves the content into the mailbox but then leaves the original PSTs connected to the mailbox also – giving a risk that the user gets confused by seeing two copies of their data.

Coupled/Uncoupled PST files

In any PST migration project it is necessary to address both 'coupled' PST files (those which a user has open in Outlook and is actively using), and 'uncoupled' PST files (those which are just stored on disk and are not in use, or are possibly from an ex-employee).

There are two challenges to consider with Coupled/Uncoupled PSTs. The first is that if a PST file is open in Outlook (coupled), then Outlook places a 'lock' on that file and it cannot be moved, edited or opened by any other application. This is an area that Microsoft has chosen not to address in their PST Capture tool – they simply raise an error that the file could not be opened for processing – instead relying on that fact that the migration should be completed outside hours, with (hopefully) all the end user PCs still switched on.

PST Enterprise faces this particular challenge head on and overcomes it by ensuring that the client can work in a mode which allows it to hook into the Outlook session and access the open PST file through the Outlook interface (transparently to the end user) as well as in the 'uncoupled' mode where it simply searches drives to look for the uncoupled PST files.

The second challenge that is faced is determining who the owner of the PST file is. Microsoft's PST Capture tool takes the approach of using the NTFS file 'owner', unfortunately this is generally set to 'BUILTIN\Administrators' and therefore is of little value, and meaning that a manual assignment of owner is needed – difficult, time consuming and certainly not scalable.

PST Enterprise uses some neat custom built technology to analyze all the data inside the PST file against 4 separate algorithms to find the best match, and present the admin with the 'determined owner' as well as a 'confidence factor' percentage of how accurate the determination has been.



Administrators can set a minimum confidence level, above which the determined owner is automatically assigned as the owner of the PST file. Of course this ownership can also be manually set or overridden if required. This technology allows a fully automated solution to the question of PST ownership.

Scalability

For scalability, the solution needs to be as 'hands off' as possible. There should be minimal user or manual intervention required and the system should function seamlessly without needing constant attention.

PST Enterprise is designed in just this manner – it is a simple web based tool that allows for simple, automated deployment of clients, which are automatically configured from the central server and which automatically execute policies defined on the central server.

All the processing happens in the background without the end user being involved (or in many cases, even being aware) and the system automatically retries when/if it gets interrupted.

PST Enterprise is stress tested to a level of 5 million PST files.

The Microsoft utility on the other hand requires manual installation of agents on workstations, manual configuration of the client agents and the operation for getting policies/migrations to run requires a lot of administrator interaction.

The possible confusion that the end user can experience due to having the same data in PST files in their profile and in a folder in their mailbox can also result in many IT helpdesk calls which would prevent the tool being used in any sizable organisation.

C2C Systems
February 2012