



The Cost of Internet Browsing and How to Combat It

Abstract

With the introduction of Windows 8 and Internet Explorer 10 in 2012, Microsoft made changes to the core subsystem of Windows that handles internet communications, namely WinInet. The change introduced a database as the central store for all internet data rather than simple files and registry entries. This had the desired effect of making the browser experience faster and allows for portability, but introduces new problems for business, especially those who have virtual environments or roaming users. The footprint size of the data has massively increased from 1 or 2 Megabytes to now hundreds of Megabytes, which all reside within the user profile.

The browser of choice for business has long been Internet Explorer but this is also a changing landscape. The consumer market favors Google Chrome, installed on each Android device as standard. Consumers will install Chrome to their Windows devices as it brings not only familiarity but is simpler and faster than Internet Explorer. Today Chrome is also a major player in business where companies are forced to install a second or even third browser as no one browser is compatible with all website and web applications. Historically these other browsers have always used databases to store their information but not utilizing WinInet they have their own. This leads to multiple copies of the data residing on the same machine, only exacerbating the problem.

All this data has an effect on both the business and the user. For the user they must now contend with much greater logon times, slower browser launch times, and slower web browsing with some pages timing out. For the administrator he must now address the issue where there is a much greater demand on network resources and for storage. In a time where more people are looking at cloud solutions and functionality, storage and network are the two prime concerns, and both can be costly to solve.

Examining the data reveals that much of it is unnecessary, with most cookies being delivered from third-parties and nothing to do with the website actually visited. Most of this data is for the express purpose of tracking users' online presence and activities. This is all uploaded to the third party which builds up a complete picture of the person which is then sold on to advertising agencies.

Whereas the problems could be previously ignored as the effects were minimal, today they are of prime concern. Web data is widely believed to be the number 1 cause of slow logons, in some cases representing up to 90% of the logon time. Tracking cookies means that personal or sensitive data is leaving the business and there is no visibility of the data being sent.



With web data seen to contain names, address, account details and other sensitive information it is important that this is protected. The obvious answer is not to persist the data between sessions, but this again effects the user.

Some data is beneficial to browsing and allows for users to customize websites for location, or to remember credentials for websites. If the data is not persisted each website works as if visited for the first time. Web based applications follow the same rule and will only lead to an increased number of support calls.

What is needed is a way of examining the data, deleting everything that is not required and keeping only what is wanted. There are solutions available today which are discussed later in this paper and which can delete the data, but with an all-or-nothing approach. Avanite WebData Control is the only software available today that can address the issues. It was designed for the enterprise and to directly address the problems discussed. It is administratively configured via standard group policy and supports all the common Windows browsers.

Avanite customers have realized: -

- An average of 60% savings on login times;
- A reduction of 85% of profile space saved;
- The number of cookies reduced from tens of thousands to a few hundred.

An evaluation of the software is available from www.Avanite.com and includes full reporting. Data before and after is recorded and presented back via a website in Azure, allowing administrators to graphically see the savings that are to be had on their own systems, and download a report that can be presented as a business case for the software.

Introduction

The birth of the internet can be defined as the first message to be sent from one network node to another. In the 1960s the US Department of Defense awarded a contract to develop the ARPANET project. In 1969 the first message was sent over the ARPANET network from a computer in the university of California to a computer in Stanford Research Institute¹. The World Wide Web was developed in the 1980s and the first hypertext document library created. The first web page was published by CERN on 6th August 1991 and can be seen today at its original address :- <http://info.cern.ch/hypertext/WWW/TheProject.html>.

As shown in figure 1 below in 2017 there were nearly 1.8 billion websites though only 200 million of them are active².

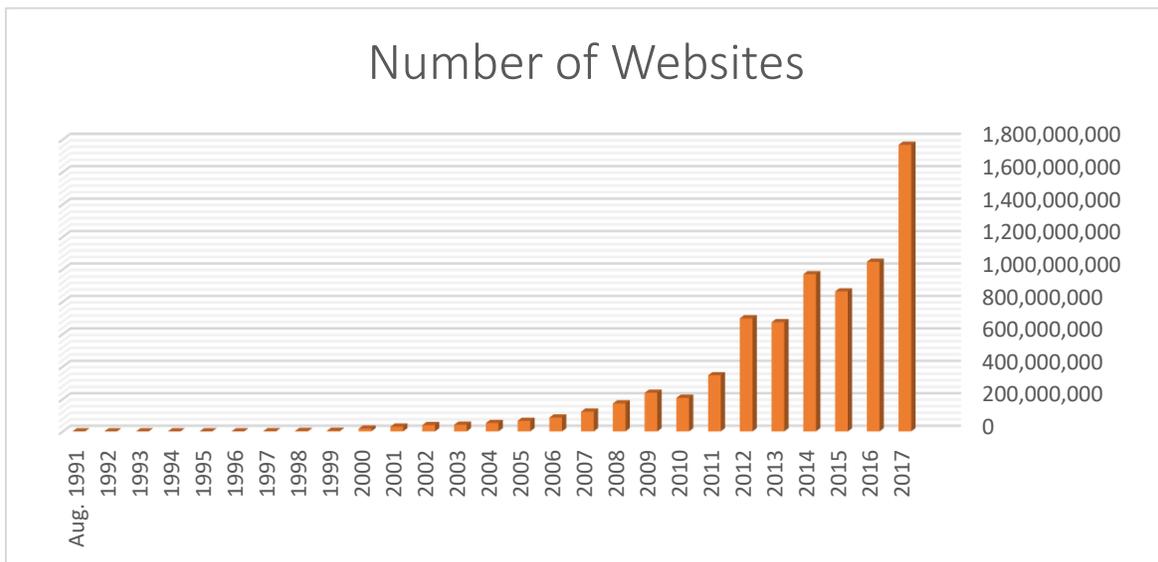


Figure 1 : Number of websites per year

1 https://en.wikipedia.org/wiki/History_of_the_Internet

2 <http://www.internetlivestats.com/total-number-of-websites/#trend>



Avanite.Ltd

Historically websites could be likened to a library with users visiting web pages for information but no real feedback going back to the website. In 1994 a programmer working for Netscape named Lou Montulli invented the web cookie; a small text file which websites could place on a website visitors machine allowing the website to identify the computer between sessions and introduced stateful information to websites, allowing them to identify and therefore interact with the visitor³. Today cookies are commonplace and are used for a variety of reasons, including authentication cookies to simplify access to protected areas such as shopping carts. Where a webpage has an embedded link to another (third-party) website a direct channel to the third-party is made and this website can also deliver cookies to the visitors' computer. It is not uncommon to visit a modern website and have over 100 cookies downloaded from simply hitting the homepage.

Along with cookies other data is generated from web browsers and web applications. Browsers maintain a browsing history which allow users to easily revisit websites. DOM (Document Object Model) storage allows websites to place much more data on the computer such as scripts which can then be executed locally, or data which can be shared between websites. And the list of web data goes on.

Web data can cause real problems for business today. This paper will look at the problems and then detail the common ways in which administrators can overcome these issues, along with pros and cons of each solution.

³ https://en.wikipedia.org/wiki/HTTP_cookie

Problem Definition

By far the largest, and only supported, business browser over recent years has been Microsoft Internet Explorer. Integrated into the operating system it was fully security patched and had defined support agreements. Storing web data was considered a necessary evil to maintain a rich browsing experience for the user. By limiting what was captured the data footprint was relatively small with most files being in the order of bytes.

Users today have multiple devices and computers sometimes more powerful than their work machines. Finding Internet Explorer slow and cumbersome most adopted Google Chrome or Mozilla Firefox as their primary browser. These proved more efficient, were faster to launch, and faster to render the web page. Web pages also became more complicated with multiple underlying technologies and not all websites and web applications rendered correctly in Internet Explorer forcing business to provide a second browser to the user.

To improve the speed of internet communications and in readiness of its new browser Edge, in 2012 Microsoft released Windows 8 and Internet Explorer 10, which included a change to the WinInet⁴ subsystem in Windows operating systems. This utilizes a database for the central store of web data, similar to Chrome and Firefox. Browsing speed and internet communications have increased, but the databases browsers use can grow to multiple hundreds of Megabytes and sometimes Gigabytes, meaning web data is now a real issue for administrators.

When designing any computer system, the user experience is of prime importance. Gone are the days when users came into work and simply sat down with the equipment on their desk to do the job at hand. Users now have multiple devices at home and are looking for work to provide for the same smooth experience they have on personal devices. This is especially difficult in a virtual environment where all desktops are built from a standard image and this is where the user profile comes in. As a user tailors their system and applications (e.g. the creation of an email signature), the customization is stored in the user profile.

⁴ For further information on WinInet see <https://docs.microsoft.com/enus/windows/desktop/wininet/about-wininet>

In a virtual environment, users may connect to different machines at each logon and therefore to make it feel and behave the same each time the user profile must exist on the machine the user is logged on to. This is achieved by copying the profile to a central store at logoff, and back into the local machine at logon. The time taken to logon is directly proportional to the amount of data in the profile and has led to the term “profile bloat” and many of the industry experts have written articles on how to keep sizes small⁵. On personal devices profile data is stored locally between sessions so the size has minimal impact to the user logon. That said, modern operating systems load browser databases at logon time and larger amounts of data still affect not only the logon time but also browser launch times and the load time of web pages.

The answer to any profile issues is usually to delete the profile, which effectively resets the user. Profile Management Solutions have been developed to help with the problem, they are able to target data contained in a profile and some will even stream the data as required. However, these simply capture the data in the form of files and registry entries and have no intelligence when it comes to the contents of the files themselves.

Web data comprises of many things such as: -

- Cookies
Small text files websites download to user’s machines to identify them between sessions.
- Browsing History
A history of all websites and pages within websites is kept. Originally stored in small files on disk Microsoft now stores this information in the webcache database.
- Compatibility Lists / Enterprise Mode
To ensure compatibility with older websites. When these sites are visited Internet Explorer can mimic older versions to render the page correctly.
- DOM Storage
An area on disk where websites can download data such as scripts for client-side execution, style sheets to reduce rendering times etc. amongst others.

The webcache database alone spawns at 24/32MB (dependent on operating system) and grows as users use the system. All applications that use WinInet, including Windows Store applications and Windows Explorer itself, all write data to the webcache database, which in turn grows in size. The graphs below illustrate the issue. To gather the data a new user was created and for each logon they simply browsed the landing page of 50 websites using Internet Explorer only. Testing was carried out using a Windows Server 2016 operating system.

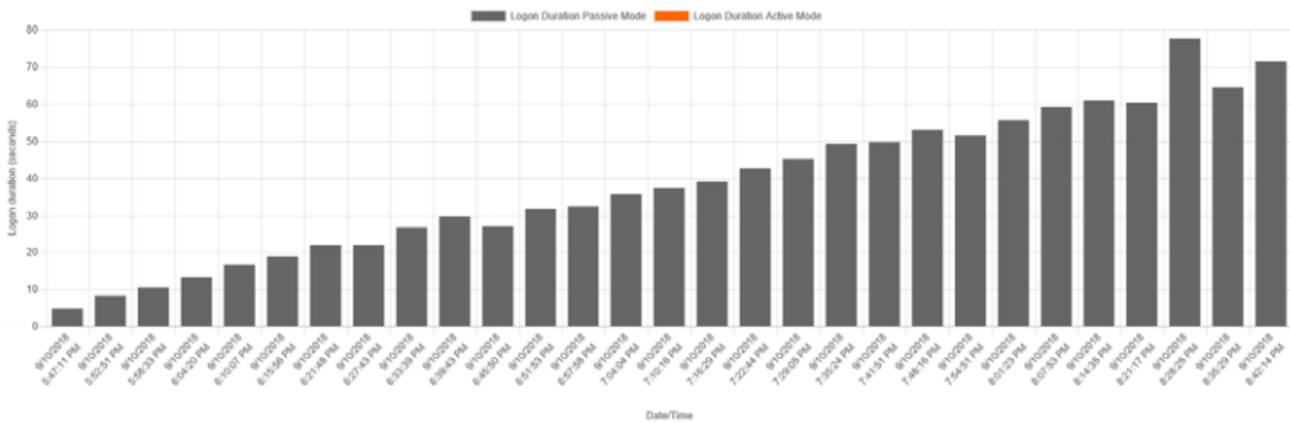


Figure 2 : Logon Time

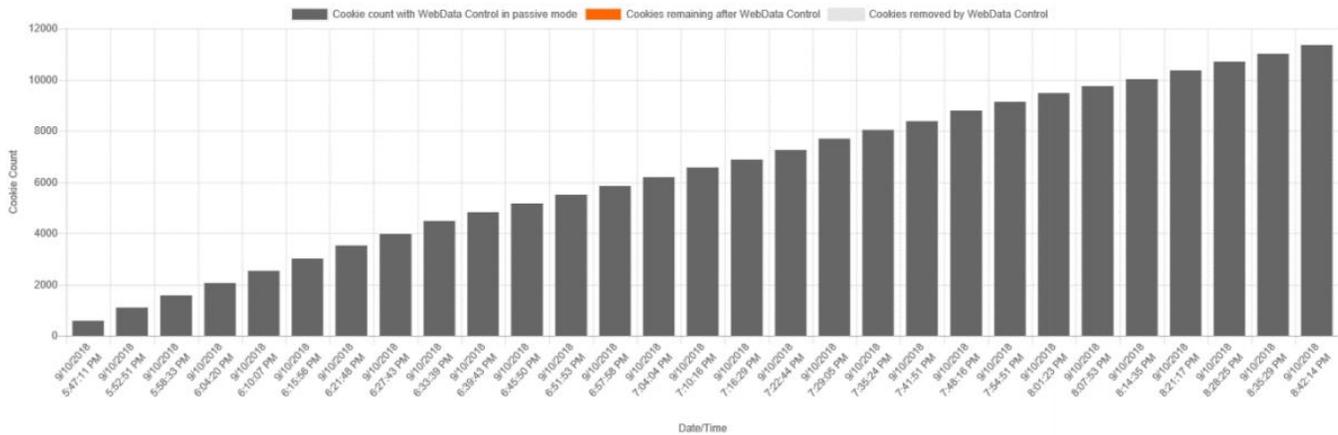


Figure 3 : Number of Cookies

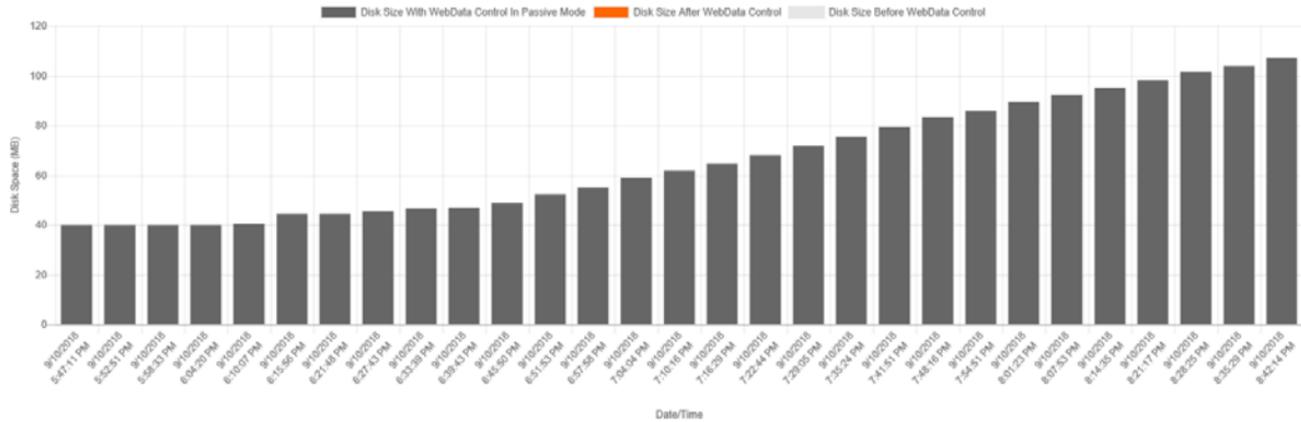


Figure 4 : Disk space usage

The graphs above clearly show how internet usage will affect the user experience over a relatively small amount of time and browsing. The graph showing the number of cookies is particularly interesting when you consider that no cookie prompts were accepted and there was no interaction with any elements on each web page. GDPR legislation dictates that companies should ask for permission prior to the placement of cookies and the collection of any data. Advertising revenue is huge, in Q1 2018 Google posted a revenue of \$31.1 billion of which \$26.6 billion was down to advertising⁶. Researching Facebook and other advertising companies yields similar results.

6 <https://www.investopedia.com/articles/investing/020515/business-google.asp>

High Level Solution

As shown in the problem section of this paper web data poses a very real problem to business today. The most common options to address the problem are discussed below.

Do Not Capture Web Data

One option the administrator has is to exclude capturing the data with the profile management solution of choice. By effectively 'dumping' the data at each logoff the problems of network usage and storage are dismissed. However, this does introduce further implications. A lot of effort has gone into making web browsing a rich experience for the user and web data forms a large part of that. A visit to the BBC website's weather page⁷ allows visitors to sign in and set locations to display the weather. Cookies are then placed on the machine to authenticate and identify the user to the website so that it can replay the settings previously configured. If the web data is deleted, then the user will have to configure preferences each logged in session. Business websites and web application rely heavily on the data to run efficiently and omitting it only leads to increased support calls and reduced productivity.

Delete the 'Old' Cookie File

Up until Windows 10 release 1709 all cookies were stored in small files on disk. Cookies have an expiry date, but browsers do not delete them when expired. This leaves a number of redundant cookies and users have been known to have 10s of thousands of cookie files. In an effort to reduce this, administrators sometimes run a script to check the date of each file and if older than x then delete it.

A common misconception is that each file represents a single cookie, whereas websites actually store multiple cookies within the same file. What's more if a cookie in the file is updated the modified time of the file does not change.

As detailed previously the browser does not communicate directly with the cookie files but to the webcache database which references the files on disk. If these files are manually deleted the references will still remain in the webcache database, leaving redundant data behind and will adversely affecting browsing performance.

All these factors mean that the process of deleting cookie files is flawed and may have the opposite to the desired effect and make the system less performant to the user. It is also worth noting that cookie files are only used by Microsoft Internet Explorer and only on older operating systems so cannot be implemented for other browsers which store all the data in their respective databases.

⁷ <https://www.bbc.co.uk/weather>

Advertisement Blocking Software

A number of companies such as AdBlock plus⁸ and Privacy Badger⁹ have produced browser extensions, herein referred to as 'Ad Blockers', in an attempt to reduce the number advertisement shown to users when browsing web pages.

They integrate into the browser and effectively block communications from the web sites displaying the advertisements.

Whilst this works for a number of sites it should be noted that: -

- Ad blocking solutions have an "acceptable ads" list to allow certain websites according to certain criteria set by the solution providers; the policy for Ad Block plus can be found at <https://adblockplus.org/en/acceptable-ads#criteria>. In many cases advertising companies pay to have their advertisements whitelisted and allowed by ad blockers.
- There are websites which will fail to load correctly or even at all with ad blockers installed. Many detect the software and will only allow users to progress to that website if the software is disabled. Looking at the cookie policy at the BBC website¹⁰ it clearly states that "If third-party cookies are disabled on your browser, some features on the BBC may not work quite as you'd expect."
- Ad block solutions are primarily targeted at the consumer market rather than business. They have no central administrative configuration and rely solely on the user to build and maintain the blacklist of sites. In virtual environments this configuration must be part of the 'roaming profile' for the user but there is little detail on how to achieve this.

For the home user these types of solution may well be acceptable and having to allow a degree of tracking and advertising is not an issue, but this is only a partial solution. Data will still continue to grow, and these products do not allow any control over what is already there. In fact, where large webcache databases are present, examining the data shows that much of it is down to other components of web data, most noticeably browsing history. One 'badly' written web application could be responsible for a large chunk of the data.

8 <https://adblockplus.org/>

9 <https://www.eff.org/privacybadger>

10 <http://www.bbc.co.uk/usingthebbc/cookies/what-happens-if-third-party-cookies-are-disabled-onmy-browser/>

'Cache' or 'Junk' Clean Up Utilities

For a number of years now there have been utilities available to people to help 'clean' their computer, one of the most notable being CCleaner from Piriform¹¹ though a simple google search will reveal many more. Microsoft Windows has long been known to become less performant over time due to things such as cache files getting larger, and data generated simply by using the system and applications. Cache cleaners are designed to tidy up these files which are slowing down your system by simply removing them. These applications are intelligent in that they know where cache and temporary files reside and so can be quite targeted, though they can be too aggressive or simply misconfigured which can cause the operating system or applications to stop functioning.

Where web data is concerned, they are able to clean up the data by simply deleting it. In the case of Microsoft Internet Explorer for example the applications may remove the files within the cookie folder, the browsing history, temporary internet files, and will also delete the webcachev01.dat database. This essentially takes us back to the solution of not capturing the data. It is easy to be confused given the name of the database (webcachev01.dat) though it should be remembered that this database is not a traditional cache (i.e. temporary storage), so deleting it deletes all web data. As discussed previously not only is this database used by Internet Explorer, but also Microsoft Edge, Windows store applications, and other applications which use the WinInet API within Windows, so deleting the database will also affect these other applications and may actually result in redundant files being left on the computer.

¹¹ <https://www.ccleaner.com/>

Avanite WebData Control

Avanite WebData Control is unique as it is the only software available today which has been designed specifically to address the problems discussed above. Avanite consultants were facing the problems on each engagement and unable to find a satisfactory solution they developed one. Version 1 of the software, previously known as WebCache Manager, was a simple command line utility aimed at the reduction in size of the webcachev01.dat file to bring back the user experience and reduce costs. Today WebData Control is at version 4.1 and has a rich feature set including: -

- **Centralized Configuration**
WebData Control works invisibly to the user and is configured administratively only via a standard group policy adm(x) which are provided. All settings can be configured at either a machine or user level allowing for full granular control.
- **Support For All Common Browsers**
Today all businesses are forced to employ more than one browser for their users because of incompatibilities with websites and/or applications. Google Chrome has become the number 1 browser of choice¹² with Internet Explorer (IE) coming second in the list. The figures do not take into account the type of device that is being used as IE is not available for mobile devices, but Chrome is starting to make tracks into business releasing enterprise supported versions of the product. WebData Control is designed to work with the top 4 browsers, Google Chrome, Internet Explorer, Microsoft Edge, and Mozilla Firefox. Each browser can be configured individually, or global settings applied across all of them.
- **Targeted Data Removal**
Each business and therefore each computer system is unique, so the definition of what data should be removed and what is kept is individual to each business. WebData Control can be configured in a number of ways.
- **Whitelist Support**
Always keep data for specific URLs.

12 <https://netmarketshare.com/browser-marketshare.aspx?options=%7B%22filter%22%3A%7B%22%24and%22%3A%5B%7B%22deviceType%22%3A%7B%22%24in%22%3A%5B%22Desktop%2F%22Laptop%22%5D%7D%7D%5D%7D%2C%22dateLabel%22%3A%22Trend%22%2C%22attributes%22%3A%22share%22%2C%22group%22%3A%22browser%22%2C%22sort%22%3A%7B%22share%22%3A%22id%22%3A%22browsersDesktop%22%2C%22dateInterval%22%3A%22Monthly%22%2C%22dateStart%22%3A%222017-08%22%2C%22dateEnd%22%3A%222018-07%22%2C%22segments%22%3A%22-1000%22%7D>

- Blacklist Support
Always remove data for specific URLs
- Targeted Removal of All Third-Party Data
- Removal of known advertising and tracking cookies.
- Removal of other excess data such as default compatibility lists which are repopulated each time
- Microsoft Store (Universal) Applications
As discussed above any application that makes use of WinInet will write data to the webcache database. An example of this is Windows 10 store applications who communicate to servers on the internet. These applications all generate the same web data and as such can be tidied up in the same way.

The list of features continues but the real strength of the product is in the results which it can achieve for business. Figures of 60% reduction in login times have been quoted but some customers have realized savings of up to 90%. Space saving figures are again large with most webcache databases reduced from hundreds of megabytes to between 8MB and 24MB; similar savings are reported for other web browsers also.

Avanite have also made evaluation of the software as simple as possible. A stand-alone analysis tool is available which is non-intrusive and runs on any Microsoft Windows system. It will detect which browsers are present and will graphically display what savings can be made with the inbuilt configuration. An evaluation version of WebData Control is available from the Avanite website (www.Avanite.com). To simplify the evaluation, it comes ready populated with a trial license and default configuration. Providing an email address at installation time automatically sets up Avanite's evaluation reporting where data is uploaded to a website hosted in Azure. Administrators can then login and see a true before and after view of things such as login times to see the effect of the software prior to any purchase.

Conclusion

The graphs on pages 6 and 7 of this document clearly demonstrate the effect of web browsing on the user; simply browsing an extra 50 websites can add 4 seconds to your login time. Internet usage is on the increase; the number of websites available doubled from 2016-2017, more applications are becoming web based and cloud adoption is at an all time high. Add to this the changes that Microsoft have introduced and administrators are now left with a difficult decision; either store web data, resulting in extended login times and poor browser and application performance; or delete the data, essentially resetting the user each time. This results in a poor online experience, reducing the functionality of both web pages and web applications. Neither of these black or white solutions are attractive and both have a negative impact to the business, the administrator, and the user.

What is required is an intelligent solution which allows for true management of the data. In most cases users can reduce the number of cookies from many thousands to a few hundred essential ones. As shown herein the webcache database itself can quickly grow to an unmanageable size so the solution needs to address this, reducing the size and optimizing the residual data. Most businesses employ two or more browsers and the data from all of these must be managed.

The IT world is changing at an ever increasing pace and IT departments are involved in large scale projects such as the rollout of Windows 10 which often means that smaller projects have to take a back seat. It is however worth investing the time to address this problem now. The gains far outweigh the effort.

- Login times reduced by 60%,
- Faster web browsing and web applications,
- Removal of all unnecessary data including tracking cookies which are responsible for data loss,
- Profile size reduced by 85%.