What Do We Mean When We Talk About WebData

Avanite.Ltd

Introduction

The term 'WebData' is applied to any data which is created when accessing the internet. The data comes from two different sources:-

- The application being used;
  For example a web browser maintains a history of websites visited.
- The internet site being visited;
  For example when accessing a website with a web browser the site may download cookies to your computer.

Usage of the internet is ever-increasing with the number of websites available doubling from 2016 to 2017. The number of web based applications is also increasing. Most applications which can be downloaded from the Windows store communicate to servers on the internet, with many applications relying on third parties such as Facebook or Google for authentication. Add to this the rate of adoption of cloud technologies and it is understandable how WebData can be responsible for 85% of the size of user profile. This not only causes issues for the administrator with greater storage and network requirements, but also for the user themselves suffering from extended login times and poor browser and application performance. This document will briefly explain the function of differing types of WebData.

The WebCache Database

The release of Microsoft Windows 8.0 and Internet Explorer v10 brought with it changes to the ways in which the computer communicates to the internet. The WinInet subsystem is responsible for user facing applications which require web access. Whereas on earlier systems applications and browsers directly accessed the filesystem and registry, they now all communicate directly to a new database, the WebCache, which can be found in %localappdata%\Microsoft\Windows\WebCache. Opening this folder reveals not only the WebCache itself but other transaction logs that have yet to be committed to the database. Care must be taken as these transactions are written lazily and therefore can lead to data loss, especially between sessions.

Figure 1 : Accessing WebData

With initial releases of the new technology much of the data was still located on the filesystem and the database held references to them allowing the browser access to the

data. As patches and service packs have been released more data is moving completely into the database, such as browsing history and cookies.[1]

The WebCache Database is named WebCacheV01.dat and can be found in %localappdata%\Microsoft\Windows\WebCache.

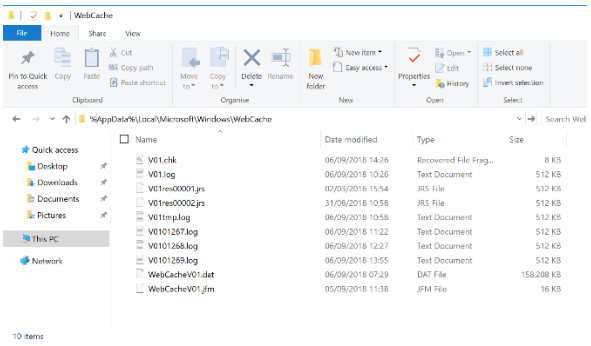

Figure 2 : WebCache Folder

As shown in 'Figure 2 : WebCache Folder' the folder contains other files alongside the database. Many of these files contain transactions that have yet to be played into the WebCache database. As data is written lazily care should be taken with these transaction files as they can lead to data loss, especially between sessions.

The WebCache database spawns at 24/32MB dependent on operating system and grows as users use the system. There is no functionality within Windows to reduce the size of the database, even deleting the caches via browser options does not shrink the database. As the browser only talks to the database this MUST be stored between sessions to allow webdata to roam between user sessions. Sizes in the order of 100s of Megabytes are common and there are many reports of multiple Gigabytes.

WebData – Cookies

When a user browses a page on a website (first party) data can be downloaded to the machine. The webpage may contain links to other websites (third party) including embedded images and social media icons. This forces a direct link from the computer to the third party website as if you had browsed the site via the browser, so giving the third party the same functionality, including the ability to deliver data.

Cookie are an example of data which is delivered both by first and third-party websites. Cookies are small files which are responsible for identifying the user and the computer to the website being visited and are used for a number of reasons, including: -

- Authentication Cookies
  Many websites that require authentication (such as shopping websites) also prompt the visitor to remember their credentials. A cookie is placed on the machine which tells the website that you do not want to be prompted for credentials. It is important

[1] Cookies are within the database on all Windows 10 operating systems post release 1709 inclusive.

to note that the username and password is not stored within the cookie but access is granted purely on the presence of the cookie. This can lead to security issues due to session or cookie hijacking.[2] If a cookie from User1 is taken and deployed by User2 then they will have full access to their account.

- Tracking Cookies
  Tracking cookies are placed on computers so that the website can track the user's online behaviour and are mostly delivered from third party websites. They commonly gather information such as:-
  - User interaction on websites – what are they clicking on;
  - Computer details including IP addresses;
  - Filesystem and Registry are scanned and details uploaded;
  - Personal details including email addresses and usernames;

  Tracking cookies tend to have long lifespans and therefore allows the website to build up a complete picture of the user, gathering more data each time the website is accessed. Large marketing companies are linked to many websites so ensuring a constant stream of data which can then be aggregated together.[3]
- Advertising Cookies
  Many web pages contain a number of advertisements to generate revenue. These adverts are usually links to third party advertising companies. Advertising companies purchase the data gathered by tracking cookies and can then identify the person directly and display adverts based on that data. Removing the cookie will not stop advertisements being displayed to the user.

Many cookies are beneficial and improve the online experience for the user. Web based applications such as Microsoft Office 365 use cookies extensively. They are used to tailor the experience for the user and so are essential. However with non-essential cookies catering for up to 80% of the cookies on your machine and with no-one following any standards defining good and bad is no easy task.

WebData – Browsing History

Browsing history is generated by the browser itself rather than being downloaded from an external party.

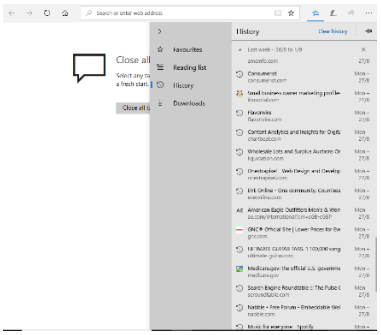

Figure 3 : Browsing History in Microsoft Edge

[2] https://en.wikipedia.org/wiki/Session_hijacking
[3] http://www.allaboutcookies.org/cookies/cookie-profiling.html

From within the browser it is possible to view each web page of each website that has been visited. This can be sorted in numerous ways included date and time visited. In pre-WebCache days these entries were stored as simple URL shortcuts within a folder and so could be easily managed. Today the browsing history is stored totally within the WebCache database or corresponding databases of other browsers. It should be noted that in many cases, especially with web applications, other URLs are accessed, for example where redirections occur. To successfully use Office 365 access to www.bing.com is required. This information is not displayed to the user but much of it is still captured within the browser databases.

WebData – DOMStore

Similar to cookie storage, DOMStore (Document Object Model Store) is a dedicated area allowing websites to place client side data for use by their websites. Whereas cookie files are limited to circa 4KB the files placed in DOMStore can house files up to 10MB[4]. This allows websites to be much more creative as can be seen by the scripts that are housed in DOMStore. As with other web data the purpose of DOMStore was to allow greater functionality and performance of web pages though it should be noted that this is also a common location for malware.

Although the actual files are stored on disk references are still held in the databases and so it is necessary to store the folder and subfolders alongside the WebCache to successfully roam this data between sessions.

WebData – Temporary Internet Files

Temporary Internet Files are created by the web browser as web pages are visited. Items such as images and video files are cached locally to improve performance. Initially created because of slow internet links these files this technology is not as effective as it used to be and many administrators choose to delete the files between sessions rather than store them. However, it should be noted that as with all internet data there are references stored within the WebCache database and as such these entries become redundant and are not cleared from the database. Unfortunately simple file manipulation is no longer possible as a way of managing internet data.

[4] https://en.wikipedia.org/wiki/Web_storage

WebData – Other Data

Although browsers work in similar ways to interpret the web page they all have their own features and functionality which adds to the amount of WebData.

- Internet Explorer
  Internet Explorer has long included a ‘compatibility mode’ feature designed to support some of the older websites. This effectively allows IE11 to mimic earlier versions of itself. This functionality was expanded with the introduction of ‘Enterprise Mode’, allowing administrators to configure site to IE versions mappings which are downloaded into the WebCache database. In addition there is a default list of mappings that are downloaded into the WebCache from the internet.
- Google Chrome and Mozilla Firefox
  Many people prefer to use Chrome or Firefox as their browser of choice because of the large number of addons and extensions that extend the capabilities. All of these are downloaded to the user’s profile and so forms part of web data. These addons often have multiple language files which are also downloaded and inflate the profile.