Cloud Storage: File Hosting and Synchronisation 2.0

MATTHIAS GEEL – THE RAINMAKER

*«**One Cloud** to rule them all,*
***One Cloud** to find them,*
***One Cloud** to bring them all*
and in the database bind them»
*- Lord of the **R**esources*

(Un)fortunately, we are not quite there yet, but the cloud(s) are everywhere. For those of you who have managed to remain oblivious to the cloud hype, here the most important points in a nutshell. Similar to other buzzwords such as Web 2.0 and HTML 5, the "Cloud" is merely an umbrella term used to describe a diverse range of technologies, architectures and approaches that are closely related to the world of distributed computing, clusters and grids. It is actually very simple. Connect a few computers to a cluster and perform some voodoo that allows you to add and remove computers at will. Add a layer that allows you to automagically (sic!) distribute the workload of running applications over the physical hardware. Voilà, scalable infrastructure. Offer a web interface to access those applications. Boom! Cloud computing. Add an array of storage devices that can be extended at runtime. Boom! Cloud storage. Virtualize your infrastructure so that you can sell storage and computing resources in manageable pieces. Boom! Cloud infrastructure provider. Buy, rent or host your own cloud infrastructure and provide a useful service on top of it. Boom! Cloud service provider. It is that simple.

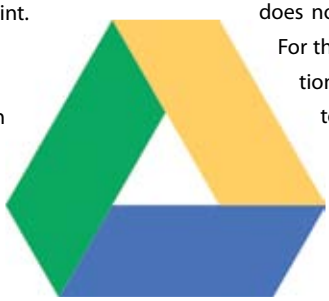
In this article however, we focus on a particular type of cloud service, namely cloud storage solutions for end-users. These services basically combine file hosting and synchronisation with cloud storage. So who are those cloud storage providers and what can they offer us?

The Competitors

The area of cloud storage services has generated quite some attention recently with several major vendors having overhauled their services in the last couple of months. While we can choose from dozens of cloud storage providers today, this article concentrates on the three major players: Dropbox, Microsoft's Skydrive and Google Drive.

Dropbox, Inc. is a startup company founded in 2007 by Drew Houston and Arash Ferdowsi, two former MIT students. According to Dropbox, the service is currently being used by over 50 million people, making them one of the largest competitors in the business of cloud storage. Microsoft was also part of the game early on and made their file hosting➔

solution Microsoft SkyDrive available in most countries in spring 2008. While the service was initially only accessible as a web application, Microsoft has just released desktop applications for Windows and Mac OS X in April 2012 to allow the synchronisation of files across devices. Interestingly, Google only recently entered the market with its own product Google Drive in April 2012. It is basically an extension of Google Docs and it stands to reason that the two products will be merged at some point.



The Promises

Cloud storage services aim to provide effective solutions to synchronise personal resources (e.g. documents, photos, music, arbitrary files etc.) and/or personal settings (e.g. application preferences, browser history, save games etc.) across many devices, to access personal data from anywhere and to easily share data with friends, co-workers and the world.

Platform Integration

A crucial feature of any cloud storage solution is how well the client applications integrate themselves with the underlying platform. Dropbox, SkyDrive and Google Drive offer desktop applications for Microsoft Windows and Mac OS X that integrates with the file system in some way. But desktop integration can also be taken a step further by exploiting various extension mechanisms the underlying operating systems offer. Microsoft Windows for example allows applications to register so-called Windows Shell Extensions handlers that hook into many facilities provided by the Windows Shell. Using those hooks, the desktop versions can offer a tighter integration with the Windows platform by extending the context menu with custom entries, providing icon overlays to indicate the synchronisation status or by adding custom property sheets to display service-specific properties. On Mac OS X, the clients integrate with the Finder application.

The synchronisation itself is automatically carried out by a background service and does not involve any user interaction. For that reason, the desktop applications rely on the operating system to be notified when files are being created, changed or deleted in order to propagate those changes to the cloud storage service.

In contrast, files are usually not automatically synchronised on mobile platforms, mainly due to storage constraints of those devices and also to limit the bandwidth needed by the cloud storage application. Instead, the users have to manually select the files they want to download and store offline. As a result, users might not always be aware of the latest version of a file they have downloaded earlier.

Privacy and Security Concerns

Before you start using a cloud storage service, you should ask yourself a few questions. First, who can access the data? Well, the cloud storage vendors mentioned in this article can and do access your uploaded data, but insist to do so only to provide you with the service. Moreover, since those companies are located in the United States and therefore required to comply with their law and regulations (includ-

ing the DMCA Act), they may be forced to disclose your data to US law enforcement agencies. The most effective way to prevent that is to follow Microsoft advice as written in their services agreement: "If you don't agree, don't use the service. Thanks." However, a slightly more pragmatic approach might be to use an additional tool like TrueCrypt to encrypt the files locally before they are being uploaded.

Second, and equally important, is the question of who owns the data? At least Dropbox and SkyDrive clearly state in their respective service agreements that "you retain full ownership to your stuff" (Dropbox) and that "we don't claim ownership of the content you provide on the service" (Microsoft). However, the situation with Google Drive is a little bit more complex. After they unified their terms of services and privacy policies in March 2012, the same rules also apply to Google Drive. Not surprisingly, those policies are formulated in a very broad sense. On the one hand, "you retain ownership of any intellectual property rights", whereas on the other hand "you give Google a worldwide license to use, host, store, reproduce, modify, create derivative works, communicate, publish, publicly perform, publicly display and distribute such content." However, both Dropbox and SkyDrive have similar terms that basically give them the right to use your content to the extent necessary to provide the service.

As with all terms of service and privacy policies, these statements are subject to interpretation and there have been several heated debates about what these statements allow the service provider to do. In any case, it is advisable to read the corresponding terms of service and privacy policy before deciding whether to use a particular cloud storage service. Luckily, most of them are rather short and there has been a trend to use a language which can be understood without a law degree.

Open Challenges and Opportunities

One immediate problem is interoperability. Since there is no common interface for consumer-oriented cloud storage services, each vendor has to offer its own client applications for all supported platforms. As a consequence, it is currently not possible to share data from one service with another, other than manually copying the files back and forth. A related challenge is the question of how cloud storage services interact with third-party applications. On the desktop, the situation is straightforward. Modern operating systems abstract the individual file systems by providing a single API to access the data in a file system agnostic way. It is therefore able offer facilities such as common file open and file save dialogs that can be used by all applications built on top of that operating system. And since the cloud storage applications integrate with the file systems, they can be used immediately by all existing desktop applications. However, such a layer is completely missing in the current architecture of the web. Instead, the prevalent practice is, that third-party web applications need to use a vendor-specific API for every service they want to connect to. One possible solution might be the development of a meta cloud service that abstracts the concrete cloud storage services from the actual applications and mediates all request through a unified API. ➔

Another interesting question is also whether cloud storage solutions can partly solve the problem of "information fragmentation". Information fragmentation describes the fact that we use an increasing number of different devices, applications and services to create, share and distribute personal information. Whether we upload a video on Facebook, edit images on the go using our tablet, use Google Docs to create documents or use our smartphone to take a picture, more and more personal data is spread across several different devices and shared with numerous online services. This makes it increasingly difficult for people to keep track of their personal information and to be able to control who can access which resource. It is therefore likely that we will see a consolidation of service providers and applications in the future. A few signs of that development are already noticeable today with Google starting to consolidate and interconnect their services and Facebook expanding their portfolio with the latest acquisition of Instagram and a partnership with Spotify.

But the journey has just begun and the one cloud to rule them all does not exist yet. There are still plenty of challenges to be solved and a lot of opportunities that await you, my fellow student. Take them and become rich. That is the plan.