Sauvegarder ses données hébergées sur un site extérieur

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 2 février 2007. Dernière mise à jour le 18 février 2008

https://www.bortzmeyer.org/sauvegarde-donnees-distantes.html

Il est fréquent, surtout depuis la mode du Web 2.0 et de ses services stockant **vos** données (Gmail,del.icio.us, etc) que des données vitales soient hébergées sur un site extérieur. Et si elles disparaissaient tout d'un coup ?

La disparition peut avoir plusieurs causes : pannes de disque dur, fausse manipulation, bogue du logiciel, ou bien tout simplement arrêt du service (oui, les services gratuits comme Gmail peuvent s'arrêter du jour au lendemain) ou encore faillite de l'entreprise. Même les plus sérieux peuvent perdre des données <http://groups.google.com/group/Gmail-Problem-solving/browse_thread/thread/e19d6ab5d41e58eb/bd2a9386c2a1ad41>.

Il **faut** donc copier en local régulièrement. Notez que c'est vrai que l'hébergeur soit une grosse boîte capitaliste états-unienne, une coopérative écolo tiers-mondiste ou bien un réseau P2P anonyme avec chiffrement. Dans tous les cas, vous ne pouvez pas demander à cet hébergeur (surtout s'il est gratuit) de se préoccuper de la pérennité de vos données plus que vous ne le faites vous-même.

Cela n'est pas secret : il est fréquent que l'hébergeur ne garantisse rien. Dans les CGU (Conditions Générales d'Utilisation), on voit souvent l'hébergeur se décharger de toute responsabilité quant aux sauvegardes du site. Mais, même si l'hébergeur garantit des sauvegardes, cela ne **vous** garantit pas de pouvoir les récupérer, par exemple en cas de faillite. Il est donc impératif de sauvegarder en local votre site ou vos données. Autrement, même les utilisateurs avertis se font parfois avoir <http://www.wired.com/politics/security/commentary/securitymatters/2008/02/securitymatters_0221>.

Avant de choisir un service distant, avant d'y avoir déposé des mois de travail, pensez à toujours vérifier qu'on pourra sauvegarder facilement.

La façon de faire des sauvegardes lorsque le site Web et/ou la base de données est hébergée à l'extérieur dépend de l'hébergeur. Elle est (ou devrait être) documentée sur son site. Par exemple, pour l'hébergeur grand public Free, un mode d'emploi de la sauvegarde des bases de données <http://www1.assistancefree.fr/v1/documentation/?forfait=axl&rac=367/365> est fourni. Même chose <http://sourceforge.net/docman/display_doc.php?docid=30227&group_id=1> pour le site d'hébergement de projets de développement de logiciels libres, Sourceforge.

Avec del.icio.us, il existe plusieurs méthodes offertes par l'hébergeur (regardez dans la rubrique "*Settings*"). Sinon, une simple ligne utilisant wget suffit :

```
wget -O bookmarks-at-delicious --http-user=YOURNAME--http-passwd=YOURPASSWORD 'api.del.icio.us/v1/posts/...
```

Elle est facile à mettre dans cron ou équivalent. Le résultat est un fichier XML (sauf erreur, on peut aussi avoir du JSON) facile à rétro-ingénierer (pour ceux qui s'inquiéteraient d'avoir juste les données et pas le logiciel).

Pour ceux qui ne sont pas sur Unix, Foxylicious <http://dietrich.ganx4.com/foxylicious/>, une extension Firefox fait également cette sauvegarde en synchronisant avec les marque-pages locaux (renseignement donné par une utilisatrice de Foxylicious, Ève Demazière).

Parfois, c'est le logiciel qui fournit ce mécanisme de sauvegarde en local. Par exemple, l'excellent logiciel SPIP, très utilisé chez les hébergeurs grand public comme Free, a son propre mécanisme de sauvegarde <http://www.magusine.net/formation/article.php3?id_article=63>. Il est dans le menu "Configuration" puis "Maintenance" puis Sauvegarde/restauration de la base". La base est sauvegardée dans un répertoire tmp/dump/, compressée, au format `.xml.gz` : il faut aller la chercher et la stocker en local.

Même chose pour Dotclear qui documente « Sauvegarde et restauration <http://fr.dotclear.org/documentation/2.0/admin/backup> ».

Les sauvegardes doivent être faites régulièrement. Chez certains hébergeurs, et avec des machines Unix, cela peut s'automatiser complètement. Sinon, il faut se fixer des procédures comme « tous les lundis, à 9 h, je sauvegarde ».

Il est prudent de tester les sauvegardes régulièrement, par exemple en tentant de les relire ou bien de les installer sur un autre site : il arrive qu'elles soient mal faites ou invalides.

Merci à David Larlet pour une entrée très intéressante <http://www.biologeek.com/journal/index.php/reve-de-geek> sur son blog, qui m'a donné envie de revenir sur ce problème. On notera également la passionnante étude "*Evaluating Personal Archiving Strategies for Internet-based Information*" <http://arxiv.org/pdf/0704.3647.pdf> de Catherine Marshall, Frank McCown et Michael Nelson, qui ont étudié les croyances des utilisateurs quant au risque de perte de leurs données, les stratégies de sauvegarde utilisées et les efforts qui avaient été fait pour récupérer ses données. Cette étude montre que le problème est réel : la grande majorité des utilisateurs ne prend pas la question suffisamment au sérieux, même pour des sites Web commerciaux.

Évidemment, cet article ne traite que la question de la pérennité, pas celle de la confidentialité, autre gros problème de certains sites Web 2.0... (Merci à Karl Dubost pour son intéressant article <http://www.la-grange.net/2006/03/29.html> à ce sujet.)