

# Efficacité du stockage dans un VCS

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 4 septembre 2009

<https://www.bortzmeyer.org/time-and-space-in-vcs.html>

---

Un intéressant article d'Eric Sink, « *Time and Space Tradeoffs in Version Control Storage* » <[http://www.ericssink.com/entries/time\\_space\\_tradeoffs.html](http://www.ericssink.com/entries/time_space_tradeoffs.html)>, expose de manière très simple et très compréhensible (même par un programmeur peu expérimenté), les choix à faire lors de la conception du stockage des fichiers d'un VCS.

L'auteur décrit et met en œuvre successivement plusieurs systèmes pour stocker les différentes versions d'un fichier dans un VCS. La plus évidente est de garder chaque version complète : cela rend la récupération d'une version donnée très rapide, mais consomme beaucoup de disque. Plus sophistiqué, le stockage des différences uniquement (par exemple en suivant le RFC 3284<sup>1</sup>). On occupe moins de place mais on est plus lent. Bref, il n'y a pas de méthode idéale, juste des compromis.

L'intérêt de l'article est de tester systématiquement chaque approche et les résultats ne sont pas forcément conformes à l'intuition. Par exemple, l'une des solutions qui prennent le moins de place est... de stocker chaque version complète mais en la comprimant. Plusieurs algorithmes très astucieux que teste l'auteur sont moins bons, à la fois en espace et en temps, que le bon vieux zlib. Cela illustre bien le fait, qu'en matière d'optimisation, il **faud** tester et ne pas se fier à des impressions.

(Par contre, l'auteur, après avoir très justement noté que la première qualité d'un VCS était de ne **pas** perdre les données qu'on lui a confié, oublie ce critère par la suite et ne prend pas en compte la fragilité de chacune des solutions testées. Par exemple, la chaîne des différences - on stocke la première version, puis la différence entre la version 1 et la 2, puis celle entre la 2 et la 3, etc - est très fragile car la perte d'une des différences fait perdre toutes les versions ultérieures.)

---

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc3284.txt>