

Venti: a new approach to archival storage
Sean Quinlan and Sean Dorward (Bell Labs)

One-liner: The authors present Venti, a “building block” for constructing applications that frequently require an efficient, reliable and accessible backup mechanism.

The authors begin by talking about how archival storage is a “second class citizen,” implying that they aim to fix this problem with Venti. In a certain sense, they do this by developing a system that places the archive and the active data on the same system. One of the strengths of this paper is the use of a data block's hash to identify the location of the block. I am sufficiently convinced that the computational overhead is not significant, especially compared to disk access. I also like the idea that the authors propose concerning the “write-once” structure of the file system. The authors successfully demonstrate that this technique reduces wasteful redundancy in the archive and also provides the kind of reliability that one would expect from archival storage because data cannot be overwritten.

There were a number of points that bothered me about this paper. First, I felt that there were a large number of claims that were not substantiated—they were simply declared. Also, I get the impression that authors were so hell-bent on promoting the status of archival storage that they turned active storage into a second-class citizen. The disk performance that they show, which takes advantage of very expensive disk technology to eek out bearable performance, is several times worse than what one would expect in a traditional file system. The authors also seem to confuse whether Venti is a “building block” or whether its only effective use is in the Plan 9 file system, where even there it requires a number of hacks to make performance bearable. Another issue I had was that the authors suggest dividing physical storage into a log and an index region, but make no comment about the size for each of these regions, nor what happens when you've used the system for a while only to find out that you chose the wrong sizes. Finally, even though the probability is insignificant, the fact that the same data can hash to the same fingerprint bothers me.

Personally, I don't think that this work will really have influence on the average user or even the average corporation. My opinion is that access to archival data for most systems is sufficiently rare that the monetary and performance cost of using this system is not outweighed by the access simplicity. The authors mention that they expect performance to improve as the system matures, but that seems like an empty promise.