Computing Deltas - File Attributes

Computing Deltas - File Attributes

Methods for Computing Deltas in Backup Applications

While there are hundreds of different backup applications all of them use one of several known methods for computing Deltas.

Deltas are simply defined as the data that has changed since the last backup run. Defining it any further than that depends on how the backup application computes deltas. A delta could be a raw disk block, a variables length portion of a file or even a complete file depending on the method.

File Attributes

The simplest form of computing Deltas for backup purposes is by using file attributes. All files have the date that in theory is the date they were last modified. In Windows there is a concept of an archive bit that is set and cleared during backup operations that can also be used. When a file is backed up the archive bit is set true and next backup operation the file is skipped.

The biggest disadvantage with this approach is that large files always have their complete contents backed up as if the entire file has changed. This method also provides no guarantee that all changes are detected as last modified dates and archive bits can easily be changed by any user or application to appear as if there have been no changes. In Windows a user can simple set the archive bit by right clicking on a file and changing its properties. In Linux and Unix a user can set the last modified date to any time they desire using the touch utility. Windows has several utilities available for overriding the last modified date and an overview is available here.

Requires Time Consuming Walk of Entire File System Tree to Compute Deltas	Yes
Delta Granularity	File
Accuracy in Identifying Changes	Not Guaranteed
Disk I/O Impact for Small Files	Small only while files are read or skipped based on attribute
Disk I/O Impact for Large Files	Very bad for performance as whole files are backed up for even a small change