MD5 – ¿Cómo funciona?

Las siglas MD5 se utilizan a menudo en la industria de las telecomunicaciones ya que se convirtieron en un término genérico como pasó con Escalator o Frigidaire. Lo que no es tan conocido es lo que esconden esas siglas. Siguiendo con los cálculos polinómicos del mes pasado, tenemos un documento que nos introduce los conceptos básicos de este sistema y como utilizarlo en la solución de control de calidad basado en archivos de Aurora.

MD5 = Message Digest Version 5

El MD5 es básicamente una función hash criptográfica que permite obtener una firma digital única para archivo en cuestión y de tamaño fijo independientemente del tamaño y complejidad de los datos, creando de esta manera una especie de huella digital del archivo procesado.
Este algoritmo fue inventando por el criptologista americano Ron Rivest en 1991. Fue seguido por la versión 6 (MD6) pero fue rechazada tiempo después. El MD5 es una de las funciones criptográficas más usadas hoy en día, aunque existen muchas otras con diferentes grados de complejidad.

La función Hash

La función Hash, replica de alguna manera la consistencia de los datos de un archivo en un formato predefinido, pero con un tamaño mucho menor. Un método muy sencillo de Hashing es, por ejemplo, la simple encriptación del tamaño de un archivo. Si el contenido del archivo cambia, el volumen de datos probablemente también cambiara, y este, es un primer nivel de verificación, fácil de experimentar en nuestro uso diario con Windows.
La función Hash se dice que es criptográfica cuando es fácil y rápido conseguir la firma digital de un archivo, pero el proceso inverso es imposible de calcular.
Eso significa:

- Una firma dada solo puede corresponder a un conjunto particular de datos (Por ejemplo, esto no sucede con el tamaño de un archivo),
- Es posible diferenciar dos series de datos muy parecidas entre sí,
- Y, por el contrario, es posible confirmar que dos archivos son absolutamente idénticos (si tienen el mismo Hash MD5),

Sin embargo, desde su creación, se ha probado que el sistema tiene vulnerabilidades, haciendo que este algoritmo solo sea útil para la detección de errores involuntarios durante la transmisión.

Utilización del MD5 en la transferencia de ficheros

A partir de un número indeterminado de datos, el algoritmo calcula una única firma de 128 bits (32 caracteres codificados en hexadecimal). Por ejemplo, dos cadenas de datos similares, tienen MD5 muy distintos (ver el ejemplo abajo). De manera que, si un archivo con código MD5 sufre algún tipo de cambio durante la trasferencia, dará lugar en recepción a un código MD5 diferente. Al igual que los códigos CRC en video, el MD5 está incluido dentro del contenedor del archivo, para que después de la transferencia, sea posible realizar el mismo cálculo y comparación del resultado y así garantizar que el archivo recibido es absolutamente idéntico al archivo enviado.
El test MD5 es una buena manera de validar una transferencia de archivos y está disponible en la suite AURORA.
Un error de MD5 indica que el archivo ha sido modificado o dañado probablemente durante la transmisión. En estos casos, se debe pedir el envío una nueva copia del archivo y no seguir adelante con el proceso hasta que los MD5 coincidan.

Ejemplo simple:

Este es el resultado MD5 para dos palabras muy similares:

Tektronix = 5b7cebd92f48f197e31d10ea605e9ba4
tektronix = c07b8df0376cf4e8f7b768c3a1d048df

A pesar de la mínima diferencia entre las dos palabras, el resultado de los códigos MD5 es muy distinto.

The MD5 code – How it works

MD5 acronym is very often used in the telecommunication industry since it became a generic term similar to an Escalator or a Frigidaire. But it is less known what is hidden behind these letters. Following the polynomial calculations of last month, here is a paper introducing the basics of this system and how to use it in the Aurora file-based QC solution.

MD5 = Message Digest 5

MD5 is basically a data **hash cryptographic function** that allows to get a unique digital signature of the data having always the same size regardless of the amount of data. Therefore, it creates a kind of fingerprint of the input.
This hash algorithm was invented by the American cryptologist Ronald Rivest in 1991. It was followed by a version 6 (MD6) that has been rejected later. MD5 is today the most used hash function, but many others exist with varying degrees of complexity.

The Hash Function

A hash function somehow replicates the consistency of data in a format known in advance and much smaller. A very simple method of hashing is the simple encryption of the file size. If the file content changes a bit, the data volume will also probably change and this is a first level of verification that is easy to experience in Windows on a daily basis.
The hash function is called **cryptographic** when it's fast and easy to get a fingerprint of the data while the reverse process is absolutely impossible by calculation.
This means that :

- A given signature can only correspond to a particular set of data (which is not the case of the size of a file for example),
- It is possible to differentiate two sets of very close data stream
- and with the opposite effect, it is possible to confirm that two objects are absolutely identical (if they have the same MD5 hash).

Since its creation, however, it was proved that very many MD5 collisions are possible, making this algorithm useful only for involuntary transmission error detection.

How to use MD5 in file-based transfer

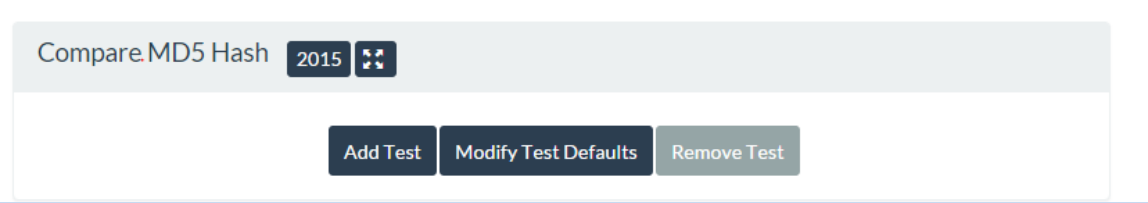
From an unknown number of data, the algorithm calculates a unique 128 bit signature (32 characters coded in hexadecimal) such as two sets of data streams almost similar have very different MD5 (see example below). MD5 code thus identifies a file which, if it undergoes changes during transport for example, will lead to a different MD5. As the CRC codes in video, MD5 is embedded in the file container such as after transfer, identical calculation and comparison, it is possible to guarantee that the file received is absolutely identical to the sent file.
The MD5 test is a good way to validate a file transfer. So, this is a test available the AURORA software.
A MD5 mismatch indicates that the file has been modified or corrupted since the MD5 hash was created, likely occurring during network transport. A new copy of the file must be requested with a matching MD5 hash before proceeding further.

Simple example

Here are the MD5 signature of two words quite similar :

Tektronix = 5b7cebd92f48f197e31d10ea605e9ba4
tektronix = c07b8df0376cf4e8f7b768c3a1d048df

Despite the very slight difference between the two words, the 2 MD5 fingerprints are very different.



Aurora container menu to select MD5 test

ABACANTO SOLUCIONES
T +34 911 591 213 | Caldereros, 13, Bj. B. 28370 Chinchón, Madrid | SPAIN
abacanto@abacantosoluciones.com

Tektronix®