**Windows файлови системи**

4. Опишете механизмът за компресиране в NTFS

NTFS поддържа компресиране на файловете. Това се прави от алгоритъм за компресия без загуба, наречен LZNT1 - вариация на популярния алгоритъм LZ77. За целта файловете се четат на парчета (chunks) по 16 клъстера и всяко празно място на диска в рамките на едно парче (chunk) се освобождава, а във файла остава единствено информация, че има определено количество празно място (sparse file). При четене, операционната система "разширява" sparse file-a, като на мястото на референциите попълва нужния брой нули. Така например, при cluster size 4 KB имаме 4 \* 16 = 64 KB за всеки chunk. Ако файлът има 50 "пълни" KB, се записват само те, като на местата на останалите 14 KB се записва единствено колко голямо празно пространство има, напр. две места по 7 KB. Така на диска може да се използват тези 14 KB. При четене, файлът ще бъде върнат в нормалния си вид, като 14-те KB се заместят с нули (т. е. празно място). Тъй като не се правят промени по "реалното" съдържание на файла, алгоритъмът работи без загуба и може да се използва доста успешно за компресия на всякакви файлове.

Ако файлът е по-малък от 900 B, съдържанието му се записва в Master File Table (МFT) и по принцип не се компресира.

LZNT1 има няколко предимства пред LZ77: на първо място, алгоритъмът за указване и кодиране на дължините е по-бърз; освен това броят битове, в които се запазват дължините може да варира (при LZ77 този брой е фиксиран); групира данните по байт наведнъж, а не по 4 байта наведнъж.

Особеност на този алгоритъм е, че компресира данните като серия от парчета (chunks), като всяко парче в последствие се декомпресира поотделно, а не целият файл наведнъж (както се случва при LZ77). Това понякога (при по-големи файлове) може да доведе до значителна фрагментация.

Дали този начин на компресия води до увеличаване на производителността е спорно. Като цяло твърдите дискове са достатъчно големи и мястото за съхранение на данни не е значително ограничено. Файлове, компресирани по този начин действително спестяват място на диска, но при четене и (особено) при запис се увеличават значително заявките към процесора. Като цяло, процесорът отговаря много по-бързо от твърдия диск (хиляди, до стотици хиляди пъти) и това по-скоро води до увеличаване на производителността, но от друга страна, при много заявки към един и същи файл, процесорът се претоварва.

Както се препоръчва от Microsoft, в идеалния случай компресията трябва да се прилага към файлове, от които се чете сравнително рядко и се записва много рядко и които не са нито прекалено малки (такива файлове почти не се компресират), нито прекалено големи (такива файлове може да се фрагментират и да затруднят четенето от тях още повече, вместо да ускорят работата на машината).