

PATRONES DE CRECIMIENTO EN LA GENERACIÓN DE INFORMACIÓN EN DISCOS DUROS

José Alfredo Victoria Barbosa

Manuel Ávila Aoki

*Centro Universitario UAEM Valle de Chalco, Universidad Autónoma del
Estado de México*

Patrones de crecimiento en la generación de información en discos duros

RESUMEN

Las tecnologías para el almacenamiento de información en discos duros, se caracterizan por un crecimiento en el tiempo exponencial de su capacidad. Actualmente, dichas capacidades están en el rango de los Terabytes (2^{40} Bytes). Por otra parte, la capacidad de los equipos para almacenamiento masivo alcanza los Petabytes (2^{50} Bytes). Asimismo, la información que en un año genera la humanidad es del orden de los Exabytes (2^{60} Bytes) y crece de manera polinomial al cubo. Hasta el momento no existe la tecnología capaz de almacenar Exabytes, lo que ha propiciado su búsqueda al interior de los gobiernos y las Organizaciones Internacionales.

Palabras clave: Petabytes, disco duro, almacenamiento, Exabytes ¿Terabytes no?

Introducción

La era del Disco Duro comienza el 4 de septiembre de 1956, cuando IBM¹ lanzó el modelo IBM 350, el cual podía almacenar 5 millones bits (Perenson, 2006). EL IBM 350 fue incorporado a la computadora IBM 305 RAMAC², que tenía 50 discos de 24 pulgadas de diámetro, con 100 superficies de grabación. Cada una tenía 100 pistas y los discos giraban a 1200 revoluciones por minuto (rpm). La tasa de transferencia de datos era de 8800 caracteres por segundo (Kozierok, 2005).

En 1961, IBM anunció la serie IBM 1301, que tenía un módulo con la capacidad para almacenar hasta 28 millones de caracteres, así como 24 discos de 24 pulgadas de diámetro. En cada pulgada se podían almacenar 50 pistas y una pista podía almacenar 520 bits. Los discos tenían una rotación de 1800 rpm. Tenía un costo de 2100 dólares mensuales y 115500 dólares por su adquisición completa (Pugh, 1991).

1 Por sus siglas en inglés, International Business Machines.

2 Por sus siglas en inglés Random Access Method of Accounting and Control (Método de acceso aleatorio para contabilidad y control).

1973	IBM presentó la unidad de disco IBM 3340, con capacidad de 30 Megabytes. Por esta razón fue apodado el "Winchester" 30-30 (López, 2004).
1993	El Integral Peripherals podía almacenar hasta 5 Gigabytes (Andrews, 2008).
1997	IBM obtiene el modelo Deskstar 16GPTitan que podía almacenar hasta 16.8 Gigabytes (Shilov, 2007)
2000	IBM logra almacenar 75 Gigabytes con el modelo IBM Deskstar 75GXP DTLA-307075 (Porter, 2005), en ese mismo año Maxtor Corporation supera la marca con 80 Gigabytes con el modelo Maxtor DiamondMax 80 98196H8 (Herrería, 2000).
2002	IBM presenta su modelo DeskStar 180GXP, con una capacidad de 180 Gigabytes (Schmid, 2002).
2003	Maxtor Corporation saca a la venta un disco duro de 250 Gigabytes con tecnología SATA, con 72000 rpm, 3.5 pulgadas y caché de 8 Megabytes (Cortés, 2003).
2005	Hitachi saca a la venta el modelo Hitachi Deskstar 7k500 con una capacidad de 500 Gigabytes, tecnología SATA, 8 Megabytes de caché, 7200 rpm (Kohrs, 2005).
2006	Seagate Technology alcanza 750 Gigabytes con el modelo Barracuda 7200.10 (Shilov, 2007).
2007	Salen a la venta varios modelos con una capacidad de almacenamiento de un Terabytes por ejemplo los modelos; HP Hasrd Disk 1 TB SATA 300 (HP, 2008), Hitachi Deskstar 7J1000 (Hitachi, 2008) y WDH1CS10000 (Western Digital).
2008	Seagate Technology logra almacenar 1.5 Terabytes con el modelo Barracuda 7200.11 (Eugen, 2008). Se ha publicado que la empresa Seagate Technology en el presente año sacará a la venta un modelo con capacidad de 2 Terabytes (Ranchal, 2008).

De acuerdo con la tabla anterior, es evidente que la capacidad de almacenamiento de información en discos duros, va escalando con el tiempo. La pregunta que surge en este punto, entonces es: ¿Existe un algoritmo de crecimiento temporal de dicha capacidad? Otra pregunta pertinente en el mismo contexto es: ¿Existe un patrón de crecimiento en la cantidad de información generada por la humanidad en un año?

El objetivo del presente trabajo es investigar los patrones de comportamiento de la razón temporal de crecimiento del almacenamiento de la información, así como la cantidad de información generada por la humanidad por año.

Para encontrar la ecuación que ilustre la evolución en la capacidad de almacenamiento en discos duros, se utilizará mínimos cuadrados con una ecuación exponencial.

$$Y = Ce^{At} \tag{1}$$

C y A son constantes a determinar, t es el tiempo y Y es la capacidad de almacenamiento en discos duros. Para hallar el valor de las constantes A y C, transformamos la ecuación exponencial a lineal.

$$\begin{aligned} \ln(Y = Ce^{At}) \\ \ln Y = \ln C + At \\ Y' = C' + At \end{aligned} \tag{2}$$

En la tabla 1, se ilustran los datos necesarios para obtener los valores de las constantes. Es importante afirmar tres puntos: i) 1024 Bytes = 1 Kilobyte = 2^{10} , por lo tanto 1024 Kilobytes = $2^{10} \times 2^{10} = 2^{20}$, por lo tanto cada 1024 Bytes = 2^{10} . ii) Al origen Y, se hizo una división de 2^{30} , por ejemplo 5 Gigabytes = $5 \times 2^{30} = 5.36879$ Bytes dividido entre $2^{30} = 5$. iii) Al origen del tiempo se restaron 1990, ya que al utilizar números tan grandes resultaría difícil y bastante engorroso utilizar mínimos cuadrados.

X=tiempo	Y=Capacidad	Xi	Yi	XiYi	Xi ²
1993	5 Gigabytes	3	1.6094379	4.8283137	9
1995	9 Gigabytes	5	2.1972246	10.986123	25
1997	16 Gigabytes	7	2.7725887	19.408121	49
2000	80 Gigabytes	10	4.3820266	43.820266	100
2002	180 Gigabytes	12	5.1929569	62.315482	144
2003	250 Gigabytes	13	5.5214609	71.778992	169
2005	500 Gigabytes	15	6.2146081	93.219121	225
2006	750 Gigabytes	16	6.6200732	105.92117	256
2007	1024 Gigabytes	17	6.9314718	117.83502	289
2008	1536 Gigabytes	18	7.3369369	132.06486	324
$\sum_{i=1}^N$		116	48.778786	662.17747	1590

Tabla 1. Evolución de la capacidad de almacenamiento en Disco Duro de 1993 a 2008.

$$a = \frac{\sum_{i=1}^N XiYi - \frac{1}{N} \sum_{i=1}^N Yi \sum_{i=1}^N Xi}{\sum_{i=1}^N Xi^2 - \frac{1}{N} (\sum_{i=1}^N Xi)^2} \tag{4}$$

$$a = \frac{662.17747 - \frac{10}{48.778786 * 116}}{1590 - \frac{1}{10} (116)^2} = 0.3942044$$

$$b = \frac{\sum_{i=1}^N Yi \sum_{i=1}^N Xi^2 - \sum_{i=1}^N Xi \sum_{i=1}^N XiYi}{N \sum_{i=1}^N Xi^2 - (\sum_{i=1}^N Xi)^2} \tag{5}$$

$$b = \frac{48.778786 * 1590 - 116 * 662.17747}{10 * 1590 - (116)^2} = 0.3051074$$

Al sustituir los valores de a y b, obtenemos la ecuación (6).

$$\ln Y = \ln C + \ln At$$

$$\text{Almacenamiento} = 2^{30} * 1.3567707e^{(0.3942044*(t-1990))} \tag{6}$$

En la tabla 2, aparece, en la primera columna, la capacidad real de almacenamiento en Disco Duro, mientras que en la segunda columna aparecen las predicciones de la ecuación. Como se puede ver en la tabla 2, el ajuste es bastante bueno.

Evolución del Disco Duro 1993-2008		
Año	Capacidad real en Gigabytes (2 ³⁰ Bytes)	Capacidad con ecuación 6 (2 ³⁰)
1993	5	4.427
1997	16	21.42
2000	80	70
2002	180	153.78
2003	250	228.1
2005	500	501.78
2006	750	744.25
2007	1024 ó 1 Terabyte (2 ⁴⁰)	1103.87
2008	1536 ó 1.5 Terabytes (2 ⁴⁰)	1637.27
2009	2048 ó 2 Terabytes (2 ⁴⁰)	2428.4

Tabla 2. Evolución de la capacidad de almacenamiento en Disco Duro de 1993 a 2008, predicha por la ecuación (6).

En la figura 1 se ilustra el comportamiento geométrico real y simulado de la capacidad de almacenamiento en los discos duros. Como se puede observar, la ecuación (6) modela bastante bien la realidad. De aquí se puede concluir que dicha capacidad muestra el comportamiento exponencial.

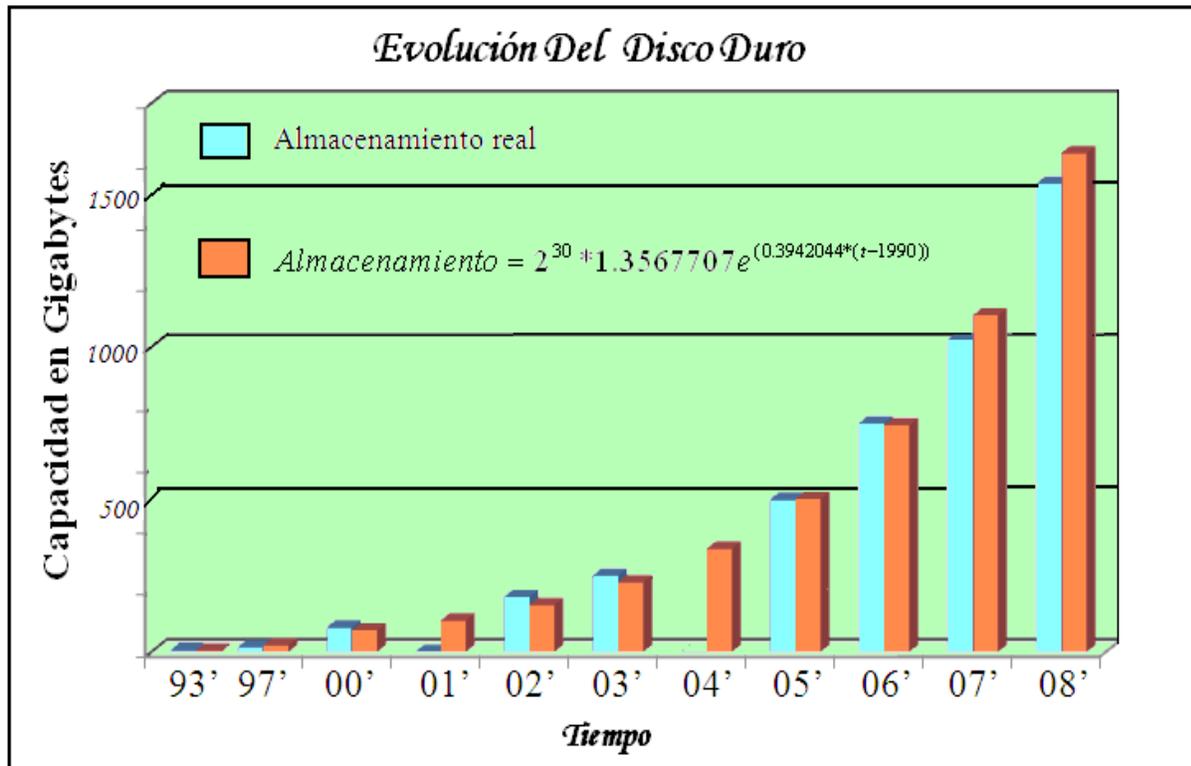


Figura 1. Curva resultante de la capacidad de almacenamiento en disco duro tanto real como la simulada por la ecuación (6).

Aumento de información

Actualmente casi toda la información está grabada digitalmente y es generada diariamente por todo el mundo, sin cesar, es decir, no se detiene el crecimiento excesivo de la información. IDC pronostica que en el año 2010 se generarán aproximadamente 988 Exabytes ($988 \cdot 2^{60}$) de información (Gantz, 2007).

Asimismo, la información guardada en aparatos digitales aumentará considerablemente año con año. En 1999 se generaron 1.5 Exabytes ($1.5 \cdot 2^{60}$) de información a nivel mundial (Rementeria, 2002). Más tarde, en el 2003 se generaron 5 Exabytes ($5 \cdot 2^{60}$). Tres años después, la cantidad de información se incrementó de una manera sorprendente, al generarse 161 Exabytes ($161 \cdot 2^{60}$) de información a nivel mundial (Gantz, 2008). De acuerdo con un ejecutivo de Microsoft, en el 2007 se dio una primera muestra de lo rebasada que está la industria del almacenamiento ante la información digital generada, y ese es uno de los retos que enfrentará el sector, ya que se generaron 281 Exabytes ($281 \cdot 2^{60}$) (Quiñones, 2008).

Para hallar la ecuación geométrica que modela el crecimiento de la cantidad de información generada por año a nivel mundial, se utilizará la ecuación

$$W = At^3 + Bt^2 + Ct + D. \tag{7}$$

Esta es un polinomio de tercer grado, donde A, B, C y D, son constantes a determinar, t es el tiempo en años y W es la cantidad de información generada a nivel mundial.

Utilizando los cuatros puntos correspondientes a los años 1999, 2003, 2006 y 2007 que tenemos para el análisis, de la ecuación (7) se obtiene el siguiente sistema de ecuaciones. Es importante señalar que la cantidad de información generada a nivel mundial se dividió entre 2^{60} .

$$\begin{aligned} A(1999)^3+B(1999)^2+C(1999)+D &= 1.5, \\ A(2003)^3+B(2003)^2+C(2003)+D &= 5, \\ A(2006)^3+B(2006)^2+C(2006)+D &= 161, \\ A(2007)^3+B(2007)^2+C(2007)+D &= 281. \end{aligned} \tag{8}$$

En la figura 2 se muestra la hoja de trabajo de Maple 9.5 (Waterloo Inc), que se usa para resolver la ecuación (8). En dicha hoja de cálculo se tomó el origen en 1990, fecha del primer registro histórico de la cantidad de información generada por la humanidad.

```
> eqns_2 := {A*(9.0*9.0*9.0)+B*(9.0*9.0)+C*(9.0)+D=1.5,
A*(13.0*13.0*13.0)+B*(13.0*13.0)+C*(13.0)+D=5.0,
A*(16.0*16.0*16.0)+B*(16.0*16.0)+C*(16.0)+D=161.0,
A*(17.0*17.0*17.0)+B*(17.0*17.0)+C*(17.0)+D=281.0 };
eqns_2 := {729.000 A + 81.00 B + 9.0 C + D = 1.5,
2197.000 A + 169.00 B + 13.0 C + D = 5.0,
4096.000 A + 256.00 B + 16.0 C + D = 161.0,
4913.000 A + 289.00 B + 17.0 C + D = 281.0}

> sols_2 := solve( eqns_2 );
sols_2 := {A = 1.212053571, B = -38.75446429, C = 408.6495536, D = -1420.821429}

> t:=2006;
año := 2006

> inf_gen_mudial:=1.212053571*((t-1990)*(t-1990)*(t-1990))-38.75446429*
((t-1990)*(t-1990))+408.6495536*(t-1990)-1420.821429;
inf_gen_mudial = 161.0000
```

Figura 2. Hoja de trabajo de Maple 9.5 para resolver la ecuación (6).

Por lo tanto, la ecuación geométrica que modela el crecimiento de información a nivel mundial es la siguiente:

$$W = 2^{60}(1.212053571 (t-1990)^3 + (-38.75446429 (t-1990)^2) + (408.6495536 (t-1990)) - 1420.821429) \tag{9}$$

En la Tabla 2 se muestra, en su segunda columna, la cantidad de información física generada por la humanidad por año, mientras que en la tercera columna aparecen los resultados de la ecuación polinómica de grado cúbico, dado por la ecuación (9).

Información generada a Nivel Mundial

Año	Información Real en Exabytes (2^{60})	Información Simulada ecuación 9 (2^{60})
1999	1.5	1.5
2003	5	5
2005	-----	79
2006	161	161
2007	281	281
2008	-----	447
2009	-	666
2010	-	946
2025	-	17374

Tabla 2. Información generada a nivel mundial, experimental y simulada.

Como se puede observar en la tabla 2, la ecuación (9) modela bastante bien la realidad. Por ello se espera que sus pronósticos futuros sobre la información generada a nivel mundial sean aceptables. Así, en el 2011 dicha ecuación pronostica que el mundo generará 1236 Exabytes ($1236 \cdot 2^{60}$). Con esto la ecuación (9) predice que el mundo entrará dentro de 2 años en la nueva era del Zettabyte (2^{70} Bytes).

Los equipos más actuales para cubrir la demanda en el crecimiento de la información son: Universal Storage Platform™ V de Hitachi (Hitachi, 2009) y HP XP24000 de HP (HP, 2009) mismos que poseen cualidades similares. Estos equipos logran almacenar hasta 247 Petabytes, utilizando tecnología Storage Area Network (SAN), que es una red dedicada a datos de alta capacidad y velocidad, que conecta diferentes tipos de unidades de almacenamiento, utilizando el estándar Fibre Channel, haciendo posible que la información se consolide y sea compartida entre diversos y diferentes servidores de red sin ningún impacto en la red LAN. A medida que se agregan dispositivos adicionales a la SAN, éstos son accesibles desde cualquier servidor en la red. En suma, la tecnología de máximo almacenamiento de información disponible, hoy en día, está en el rango de Petabytes, lo cual no satisface los Exabytes y Zettabytes que demanda las necesidades de la humanidad.

Conclusiones

En el presente trabajo se ha modelado la evolución, al paso de los años, de las capacidades de almacenamiento de discos duros producidos por las empresas más importantes a nivel mundial en las respectivas tecnologías. Se demostró que dicha capacidad obedece, de manera notable, a un crecimiento de tipo exponencial. Esta función exponencial, que está dada por la ecuación (6), se ajustó matemáticamente, y es tan sorprendente su fidelidad con la realidad, que sus pronósticos para años futuros podrían resultar altamente confiables. Por ejemplo, se estima que la nano-tecnología empezará florecer de lleno en el 2030 (Moore, 1965). Luego entonces la ecuación (6) pronostica que el almacenamiento en dicho año será de 9561769 Gigabytes = 9.11 Petabytes ($9.11 \cdot 2^{50}$) en un solo disco duro. También se modeló el crecimiento, en el tiempo, de la información total generada en el mundo. Para dicho propósito se propuso que esta razón de crecimiento fuese de tipo polinómico de grado tres, tal como lo muestra la ecuación (9). De la Tabla 2 podemos concluir que la simulación propuesta en la Ec. (9) es bastante buena. Así, los pronósticos de esta ecuación, para la era nano-tecnológica, indican que la información total generada a nivel mundial será de 30489.446 Exabytes = 29.7 Zettabytes ($29.7 \cdot 2^{70}$).

Aquí surge la pregunta de si existe un servidor que pueda almacenar la información total generada a

nivel mundial en un año. La respuesta es que no, no existe tal aparato. La razón es que hasta el 2009 los equipos para máxima capacidad de almacenamiento es el XP24000 Disk Array de HP y el Universal Storage Platform™ V de Hitachi, con una capacidad de almacenamiento de 247 Petabytes cada uno. Así, lo máximo que se podría guardar no superaría ni el 1% de la información total generada por la humanidad en un año. Esta investigación sirve para hacer un llamado a los organismos internacionales y las empresas especializadas, para emprender el diseño y la implementación de discos duros con una capacidad de Zettabytes, es decir, un millón de Petabytes. Al generar tanta información a nivel mundial, se tiene como consecuencia una pérdida de tiempo para la información que deseamos utilizar, y más aún, día con día los servidores gratuitos dejan de almacenar la información de cualquier persona sin una supervisión previa. Esto hace que la información se esté duplicando y sólo se copie y pegue, lo que ocasionaría el grave problema de pérdida de información original.

Bibliografía

ANDREWS Euan. 2008. *Hard Drive*, en http://www.htpcnz.net/Hard_Driver, consultado el día 11 de Diciembre de 2008.

CORTÉS Ángel. 2003. *Maxtor presenta disco duro de 250 GB*, en www.noticiasdot.com/publicaciones/2003/0603/3006/noticias300603/noticias300603-9.htm, consultado el 5 de Enero de 2009.

EUGEN, 2008, *Hard drives hit 1.5 TB*, en http://www.storagereview.com/hard_drives_hit_1_5_tb, consultado el 5 de noviembre de 2008.

GANTZ F. John. 2008. *An Updated Forecast of Worldwide Information Growth Through 2011*, en <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>, consultado el 4 de Noviembre de 2008.

HERRERÍA Rey Juan. 2000, *Conozca su Hardware*, en http://www.conozcasuhardware.com/noticias/30_2000.htm, consultado el día 7 de noviembre de 2008.

HITACHI. 2009. *Hitachi Global Storage Technologies*, en <http://www.hitachigst.com/portal/site/en/menuitem.b5bc67ba7b48099056fb11f0aac4f0a0/>, consultado el 10 de Enero de 2009.

HP. 2008. *HP StorageWorks XP24000/XP20000 Disk Arrays*, en www1.hp.com/products/storageworks/xp24000/software.html, consultado el 9 de Enero de 2009.

KOHRS Jason. 2005. *Hitachi Deskstar 7K500 500GB SATA-II Hard Drive*, en <http://www.bigbruin.com/reviews05/hitachi7k500/index.php?file=1>, consultado el 27 de Diciembre de 2008.

KOZIEROK Charles M. 2005, *A Brief History of the Hard Disk Drive*, en <http://www.storagereview.com/guide2000-ref-hdd-index.htm>, consultado el 1 de diciembre del 2008

LÓPEZ C. P. A. 2004. *Manual fundamental de Hardware y componentes*, Edit. Anaya, Madrid.

MOORE G. E, 1965, "Cramming more components onto integrated circuits", *ELECTRONICS*, Numero 8, Volumen 38.

PERENSON Melissa J. 2006, *The Hard Drive Turns 50*, *PC WORLD*, Septiembre 13.

PORTER N. James. 2005. *An Historical Perspective of the Disk Drive Industry*, en <http://www.thic.org/pdf/April05/disktrend.jpporter.050419.pdf>, consultado el 20 de Noviembre de 2008

PUGH W. Emerson, Lyle R. Johnson and John H. Palmer. 1991. *IBM 1301 disk storage unit*, en http://www-03.ibm.com/ibm/history/exhibits/storage/storage_1301.html, consultado el día 25 de Octubre de 2008.

QUIÑONES Azcárate Ernesto. 2008. *Petabytes de información: Repensando el modelamiento de base de datos*, en http://www.eqsoft.net/presentas/Petabytes_de_informacion_repensando_el_modelamiento_de_datos.pdf, consultado el 3 de Noviembre de 2008.

RANCHAL Juan, 2008, *Seagate anuncia la comercialización de SSDs y discos duros de 2 TB en 2009*, en http://www.theinquirer.es/2008/05/30/seagate_anuncia_la_comercializacion_de_ssds_y_discos_duros_de_2_tb_en_2009.html, consultado el 23 de Diciembre de 2008.

REMENTERIA Piñones José Ariel. 2002. *COMUNICACIÓN, DERECHO DE LA INFORMACIÓN Y ÉTICA EN ESPAÑA Y AMÉRICA LATINA*, en <http://www.cabid.cl/doctos/cabid2.pdf>, consultado el 5 de Noviembre de 2008

SCHMID Patrick. 2002. *The Next Generation: IBM DeskStar 180GXP with up to 180 GB: Introduction*, en <http://www.tomshardware.com/reviews/generation,551.html>, consultado el 26 de Octubre de 2008.

SHILOV Anton. 2007. *Hitachi Unveils World's First 1TB Hard Disk Drive*, en <http://www.xbitlabs.com/news/storage/display/20070108103913.html>, consultado el día 23 de Diciembre de 2008.