

Ingeniería de Servidores

Big Data y TPCx-HS

Resumen

En este texto puedes incluir un resumen del documento. Este informa al lector sobre el contenido del texto, indicando el objetivo del mismo y qué se puede aprender de él.

Índice

1. Introducción	2
2. Big Data	3
3. Map Reduce. Hadoop	4
4. Spark y Flink	4
5. Benchmarks: TPCx-HS	4
5.1. Carga de trabajo de TPCx-HS	5
5.2. Fases de ejecución del benchmark	5
5.3. Medida del rendimiento	6
5.4. Comparación de la medida	7
6. Conclusión	7

1. Introducción

Desde hace miles de siglos el ser humano ha investigado la manera de almacenar y recopilar información. Durante muchos siglos la escritura y la pintura eran los únicos mecanismos existentes. Posteriormente surgió la fotografía, los discos de vinilo... Sin embargo, poca información seguía ocupando mucho volumen físico. Gracias a los avances tecnológicos de las últimas décadas, hoy en día disponemos dispositivos electrónicos para el almacenamiento de datos binarios. Además, la evolución de estos dispositivos ha sido frenética. IBM comercializó el primer disco duro en 1956. Este constaba solamente de 5 mega bytes de capacidad [3] mientras que actualmente podemos utilizar discos duros con más de 1 tera byte.

La capacidad de cómputo y procesamiento de los computadores también ha crecido de forma exponencial. El primer ordenador comercial se presentó en 1951 y se conoce como UNIVAC 1 [7]. Este computador realizaba n cuentas por segundo y tenía m kilo bytes de memoria principal o RAM. Actualmente utilizamos procesadores con más de 2 giga gercios de frecuencia de reloj, esto es, realizan hasta dos mil millones de operaciones por segundo. Además, es habitual utilizar ordenadores con 8 o más giga bytes de memoria principal, lo que permite trabajar con bastante información de forma eficiente.

Estas nuevas tecnologías han posibilitado que el almacenamiento de datos sea mucho más sencillo. Podemos guardar multitud de archivos en un dispositivo de unos centímetros y compartirlos con cualquier usuario. Además, el procesamiento de estos archivos e información es eficiente gracias a la capacidad de los computadores actuales.

El mayor flujo de datos es producido en Internet. Aunque es relativamente joven, se hizo público en 1993, actualmente existen más de mil millones de páginas webs [10]. Además, multitud de dispositivos electrónicos se conectan e interaccionan con Internet (lo que se denomina Internet de las cosas [11]). Los usuarios de estos dispositivos utilizan aplicaciones web y redes sociales, publicando textos y archivos multimedia.

Todo este cúmulo de tecnologías y actividades ha dado lugar a que hoy en día haya más de 10 zeta bytes de información almacenados ($1 \text{ ZB} = 10^{12} \text{ GB}$). La Figura 1 muestra la evolución histórica de la cantidad de información acumulada por el ser humano. Podemos observar que cada año se generan varios zeta bytes de información, el crecimiento es exponencial. Hasta 2003 se habían almacenado en total 5 exa bytes de información ($1 \text{ EB} = 10^9 \text{ GB}$). Actualmente, generamos esta cantidad de datos en dos días [9].

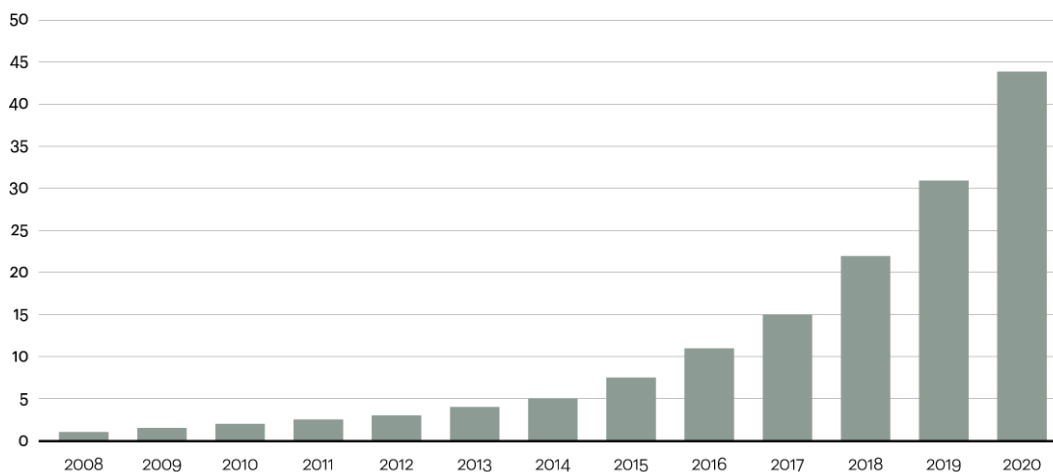


Figura 1: Evolución histórica del número de Zeta Bytes de información almacenados y predicción para los próximos años [2].

Toda esta cantidad de información necesita ser procesada y clasificada. Por ejemplo, encontramos empresas como Facebook y Google cuyos usuarios generan una gran cantidad de datos diariamente. Estos datos deben ser tratados en tiempo real para poder mantener los sistemas de recomendaciones asociados. En estos ejemplos la cantidad de datos a procesar supera con creces la capacidad de un computador usual. Se denominan conjuntos de datos masivos. Estos conjuntos de datos requieren nuevas tecnologías y algoritmos que permitan tratarlos eficientemente. El desarrollo de nuevas tecnologías, algoritmos y herramientas para tratar conjuntos de datos masivos es lo que se conoce como Big Data [4].

En este trabajo presentamos una introducción a Big Data y a las diferentes herramientas software existentes, destacando el papel de la ingeniería de servidores en este contexto. En particular, mostramos la importancia del desarrollo de nuevos benchmarks que permitan evaluar estas tecnologías de forma clara y objetiva.

El resto del texto se organiza como sigue. La Sección 2 contiene una mayor descripción del concepto de Big Data. En la Sección 3 introducimos el paradigma de programación map reduce y la tecnología que lo implementa, denominada Hadoop. En la Sección 4 describimos las nuevas tecnologías que han surgido para cohesionar la filosofía map reduce con el procesamiento iterativo. Estas se denominan Spark y Flink. En la Sección 5 explicamos uno de los benchmarks existentes para las tecnologías Big Data, denominado TPCx-HS. Por último, en la Sección 6 presentamos las conclusiones obtenidas.

2. Big Data

A pesar de la evolución de los computadores en todas sus facetas, la cantidad de datos e información a procesar y almacenar crece incluso a mayor velocidad. En muchas ocasiones un ordenador normal no es capaz de tratar tantos Giga Bytes de datos. Estos conjuntos de datos se denominan masivos. Además, en múltiples aplicaciones se necesita aplicar algoritmos sobre los conjuntos de datos recopilados, requiriendo mucho tiempo de cómputo en el caso de que estos sean de gran tamaño.

Big Data engloba el tratamiento de datos masivos desde el punto de vista tecnológico y algorítmico. En palabras de Michael J. Franklin, profesor de informática en la universidad de Berkley [12]:

“Un problema sobre datos entra en el ámbito de big data cuando la aplicación de las actuales tecnologías no permite al usuario obtener soluciones rápidas, efectivas en costo y de calidad”

En la literatura especializada se destacan las siguientes características de Big Data, que se denominan las 3 V's de Big Data [9]:

- **Volumen.** El tamaño de los conjuntos de datos a procesar es cada vez mayor, por ejemplo, facebook procesa cada día 500 TB de información. Este volumen de datos requiere tecnologías específicas para que los servidores de altas prestaciones puedan manejar la información con éxito.
- **Velocidad.** Necesitamos herramientas que permitan procesar y analizar conjuntos de datos masivos en poco tiempo. Además, es habitual que el procesamiento de los datos deba ser incluso en tiempo real, esto es, los datos llegan al sistema de forma continua y este debe agregar la información de los mismos.
- **Variedad.** Los datos a tratar provienen de una gran variedad de fuentes. Por tanto, las herramientas Big Data deben permitir procesar a la vez datos de diferentes características y tamaños. Es más, habitualmente encontramos datos de tres tipos: estructurados, semi estructurados y sin estructurar. Los datos estructurados son sencillos de clasificar. Sin embargo, los datos sin estructurar son aleatorios y difíciles de analizar. Por su parte, los datos semi estructurados requieren técnicas avanzadas para poder clasificarlos correctamente.

Algunos autores han extendido la definición hasta utilizar un total de 9 V's: veracidad, valor, viabilidad y visualización entre otras [17].

Los conjuntos de datos contienen conocimiento que necesitamos extraer. Por ejemplo, todas las empresas almacenan información sobre sus clientes, la actividad y transacciones realizadas. Esta información necesita ser analizada en tiempo real para poder actuar en consecuencia. Aquella empresa que mejor conozca el mercado y actúa rápidamente obtendrá mejores resultados. La ciencia de datos es la rama de la inteligencia artificial que se encarga de tratar y extraer conocimiento de los datos (referencia). Se han desarrollado múltiples algoritmos para ello (referencia machine learning), que permiten resultados tan impactantes como el reconocimiento de voz o los sistemas de recomendación. Big Data Analytics [5].

3. Map Reduce. Hadoop

Por tanto, necesitamos recurrir a servidores de altas prestaciones y procesamiento distribuido para poder aplicar estos algoritmos.

Los servidores de altas prestaciones (alguna descripción). Se han desarrollado herramientas de cómputo en paralelo y distribuido, como OPEN MP y MPI (referencias), que permiten implementar algoritmos distribuidos sobre estos. Sin embargo, estas implementaciones dependen del servidor y son a bajo nivel. Una determinada implementación sobre un esquema hardware puede funcionar bien en determinado momento pero al año tendrá que ser capaz de trabajar con el doble de datos. Esto probablemente suponga la necesidad de ampliar el hardware y rehacer la implementación. Necesitamos pues nuevos paradigma de programación que permitan abstraer el desarrollo de software para plataformas distribuidas del hardware y proporcionen capacidad para el tratamiento de datos masivos.

4. Spark y Flink

5. Benchmarks: TPCx-HS

La variedad de computadores es bastante heterogénea. Cada uno realiza de forma eficiente un determinado conjunto de operaciones a consta de presentar peores resultados en otros factores. Por tanto, la comparación entre diferentes modelos de computadores es compleja. Consecuentemente, se han creado benchmarks con el objetivo de aportar elementos de juicio con los que se discernir entre el uso de un computador u otro para una determinada aplicación. Habitualmente el benchmarking se define como la obtención de información útil mediante pruebas empíricas que ayude a una organización a mejorar sus procesos [1]. Sin embargo, el benchmarking de computadores es un proceso costoso computacionalmente. Gasta tanto energía como mucho tiempo de cómputo. Por tanto, se ha de reservar para cuestiones sean importantes y no para evaluar tareas simples [8].

Una técnica utilizada para ver qué sistema nos da más prestaciones son los benchmarks, uno de ellos es TPC. TPC es una organización sin ánimo de lucro que estudia el proceso de transacción y los benchmarks para las bases de datos. El término transacción se suele atribuir a aspectos bancarios, pero si pensamos en este concepto como una función que tienen los ordenadores, una transacción puede ser un conjunto de operaciones que incluyen lectura y escritura en disco, llamadas a funciones del sistema operativo, o cualquier forma de transferencia de datos de un sistema a otro. Este es el ámbito en el que se mueve TPC, que produce benchmarks que miden el proceso de transacción y el rendimiento de las bases de datos en términos de: datos un sistema y una base de datos, ¿cuántas transacciones pueden hacer por unidad de tiempo?[13]

TPCx-HS se desarrolló para proveer de un rendimiento fiable, rendimiento relación precio, disponibilidad y, opcionalmente, datos de consumo de energía de los sistemas de Big Data. TPCx-HS fue el primer benchmark objetivo que permitía medir tanto hardware como software, como el Hadoop Runtime. [16]

5.1. Carga de trabajo de TPCx-HS

La carga de trabajo de TPCx-HS consiste en los siguientes módulos [14]:

- HSGen: es un programa que genera los datos según factor de escala, que suele estar entre un 1TB y 10000TB(se denota TB como terabytes).
- HSDataCheck: es un programa que comprueba el cumplimiento del conjunto de datos.
- HSSort: es un programa que ordena los datos según un orden total.
- HSValidate: es un programa que valida la salida, es decir, los resultados obtenidos.

5.2. Fases de ejecución del benchmark

Una ejecución válida consiste en cinco fases separadas que corren secuencialmente. Estas fases no se solapan en su ejecución, es decir, el comienzo de la fase 2 no puede darse hasta que la fase 1 esté completa. Para que comience cada fase se necesita de un script llamado ¡TPCx-HS-master¿, que es el que inicia cada fase y que puede ser ejecutado desde cualquier nodo del sistema que se está bajo test [14].

Las fases de ejecución son las siguientes:

- Fase 1: se generan los datos de entrada usando HSGen. Se han de copiar en un soporte duradero y hacer la copia replicada en tres discos (llamado *"3-ways replication"*), que suelen ser el principal, el secundario y uno de respaldo[6].
- Fase 2: se verifica la validez del conjunto de datos usando HSDataCheck. El programa sirve para verificar la cardinalidad, tamaño y el factor de réplica de los datos generados. Si el programa reporta un fallo, entonces la ejecución no es válida y se deberá volver a empezar.
- Fase 3: se lanza el programa HSSort con el que se ordena los datos de entrada. Esta fase muestra los datos de entrada y los datos de salida (los datos ya ordenados). Al igual que en la fase 1, se han de copiar los datos en un soporte duradero y hacer el *"3-ways replication"*.
- Fase 4: se comprueba la viabilidad del conjunto de datos usando HSDataCheck. El programa comprueba la cardinalidad de los datos, tamaño y el factor de replicación de los datos ordenados. Si el programa reporta un fallo, entonces la ejecución no es válida y se deberá volver a empezar.
- Fase 5: validación de los datos con HSValidate. Como su nombre indica, HSValidate comprueba que los datos de salida sean correctos y si reporta el fallo consisten en que el HSSort no generó el correcto orden de salida, la ejecución se considerará inválida.

El benchmark consiste en dos ejecuciones (ver la Figura 2) y cada vez que se ejecuta una fase se ha de especificar el tiempo que ha llevado la ejecución. Entre la primera ejecución de las cinco fases y la segunda, hay una fase intermedia que sirve para limpiar el sistema, conocida como *"file system cleanup"* y, lógicamente, no se permite ninguna actividad durante dicha fase. Una vez realizadas las dos ejecuciones se obtiene el tiempo total de ejecución que servirá para calcular datos en la fase de medición.

Como consideraciones finales en la ejecución, el sistema bajo test no puede ser reconfigurado o cambiado durante o entre cualquiera de las fases de la ejecución ni tampoco entre la primera y la segunda ejecución. Cualquier cambio que se haga en el sistema se deberá realizar antes del comienzo

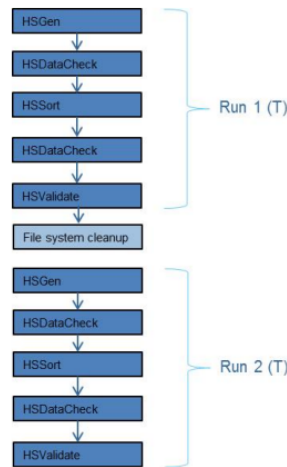


Figura 2: Gráfico de las ejecuciones de TPCx-HS

de la fase 1 de la primera ejecución. El factor de escala usado en el conjunto de datos del test debe ser escogido del conjunto de factores de escala definido como sigue: 1TB, 3TB, 10TB, 30TB, 100TB, 300TB, 1000TB, 3000TB, 10000TB.

5.3. Medida del rendimiento

Para hacer el análisis de la medida del rendimiento, TPCx-HS define las siguientes medidas:

- HSph@SF: refleja la medida del rendimiento de TPCx-HS.
- \$/HSph@SF: es la medida del rendimiento-precio.
- Si se escoge la opción de "TPC-energy", la medida de la energía que hace TPCx-HS informa de la potencia por rendimiento.

La medida del rendimiento se representa con HSph@SF, que se mide con la siguiente expresión:

$$HSph@SF = \frac{SF}{T/3600}$$

donde:

- SF es el factor de escala escogido.
- T es el tiempo total que se obtuvo al sumar el tiempo de las dos ejecuciones.

La medida del rendimiento-precio se puede calcular con la siguiente fórmula:

$$$/HSph@SF = \frac{P}{HSph@SF}$$

donde P es el costo del sistema en el que se han hecho las ejecuciones.

5.4. Comparación de la medida

Un resultado dado por la ejecución de TPCx-HS sólo puede ser comparado con otro resultado que provenga de la ejecución de este benchmark y que tenga el mismo factor de escala. Otras consideraciones a tener en cuenta son:

- Los resultados producidos por test que tienen diferente factor de escala no son comparables, debido a los *retos* computacionales encontrados en volúmenes de datos de distinto tamaño.
- Si los resultados medidos con diferentes factores de escalas aparecen impresos o en algún documento electrónico, entonces cada referencia que se haga a uno de estos resultados se ha de especificar claramente el factor de escala que se usó para obtener esos resultados. Si los resultados aparecen de forma gráfica, el factor de escala en el que se basó dicha medición se deberá poder discernir, por ejemplo usando una etiqueta en uno de los ejes.

Se presenta a continuación una tabla con los resultados para algunas compañías donde se puede ver el factor de escala usado, el sistema y los resultados obtenidos tras la ejecución del benchmark.











Date Submitted	Scale Factor	Company	System	HSph	Price/HSph	Watts/KHSph	System Availability	Apache Hadoop Compatible Software	Operating System	Cluster
03/30/16	1 TB		Cisco UCS Integrated Infrastructure for Big Data	10.12	38,168.98 USD	NR	03/31/16	MapR Converged Community Edition Version 5.0	Red Hat Enterprise Linux Server 6.7	Y
03/30/16	10 TB		Cisco UCS Integrated Infrastructure for Big Data	12.02	32,135.61 USD	NR	03/31/16	MapR Converged Community Edition Version 5.0	Red Hat Enterprise Linux Server 6.7	Y
03/22/16	10 TB		Cisco UCS Integrated Infrastructure for Big Data	11.56	37,066.53 USD	NR	03/23/16	MapR Converged Community Edition Version 5.0	Red Hat Enterprise Linux Server 6.5	Y
10/23/15	30 TB		Cisco UCS Integrated Infrastructure for Big Data	23.42	36,800.52 USD	NR	10/26/15	Cloudera Distribution for Apache Hadoop (CDH) 5.3.2	Red Hat Enterprise Linux Server 6.5	Y
10/23/15	100 TB		Cisco UCS Integrated Infrastructure for Big Data	21.99	39,193.64 USD	NR	10/26/15	Cloudera Distribution for Apache Hadoop (CDH) 5.3.2	Red Hat Enterprise Linux Server 6.5	Y
10/16/15	1 TB		Dell PowerEdge 730/730xd	7.39	46,762.93 USD	NR	10/19/15	Cloudera Distribution for Apache Hadoop (CDH) 5.4.2	Red Hat Enterprise Linux Server 6.5	Y
10/16/15	3 TB		Dell PowerEdge 730/730xd	8.25	41,888.25 USD	NR	10/19/15	Cloudera Distribution for Apache Hadoop (CDH) 5.4.2	Red Hat Enterprise Linux Server 6.5	Y
10/16/15	10 TB		Dell PowerEdge 730/730xd	9.07	38,101.22 USD	NR	10/19/15	Cloudera Distribution for Apache Hadoop (CDH) 5.4.2	Red Hat Enterprise Linux Server 6.5	Y
10/16/15	30 TB		Dell PowerEdge 730/730xd	8.38	41,238.43 USD	NR	10/19/15	Cloudera Distribution for Apache Hadoop (CDH) 5.4.2	Red Hat Enterprise Linux Server 6.5	Y
09/24/15	3 TB		Cisco UCS Integrated Infrastructure for Big Data	11.76	44,052.98 USD	NR	09/25/15	Cloudera CDH 5.3.0, HDFS API ver 2, Map Reduce API ver 1	Red Hat Enterprise Linux Server 6.5	Y

Figura 3: Tabla de resultados obtenidos con TPCx-HS [15].

6. Conclusión

Referencias

- [1] Confederación Granadina de Empresarios. *¿Qué es el Benchmarking*. URL: <http://www.cge.es/portalcge/tecnologia/innovacion/4111benchmarking.aspx>.
- [2] Hugo Evans y et. al. *Big Data and the Creative Destruction of Today's Business Models*. URL: http://www.atkearney.es/paper/-/asset_publisher/dVxv4Hz2h8bS/content/big-data-and-the-creative-destruction-of-today-s-business-models/10192.

- [3] Rex Farrance. «Timeline: 50 Years of Hard Drives». En: *PCWorld* (2016). URL: <http://www.pcworld.com/article/127105/article.html>.
- [4] F. Herrera. *Inteligencia artificial, inteligencia computacional y Big Data*. Servicio de Publicaciones, Universidad de Jaén, 2014.
- [5] Karthik Kambatla y col. «Trends in big data analytics». En: *Journal of Parallel and Distributed Computing* (2014), págs. 2561-2573.
- [6] Linbit. *Three-way replication*. URL: <https://www.drbd.org/en/doc/users-guide-83/s-three-way-repl>.
- [7] Computer History Museum. *Timeline of Computer History*. URL: <http://www.computerhistory.org/timeline/1951/>.
- [8] Universidad Técnica José Peralta. *Ventajas, desventajas y causas posibles de fracasos del benchmarking*. URL: <http://www.buenastareas.com/ensayos/Ventajas-Desventajas-y-Causas-Posibles-De/3531988.html>.
- [9] Seref Sagiroglu y Duygu Sinanc. «Big data: A review». En: *International Conference on. IEEE* (2013), págs. 42-47.
- [10] Internet live stats. *Total number of Websites*. URL: <http://www.internetlivestats.com/total-number-of-websites/>.
- [11] Mario Tascón y Arantza Coullaut. *Big data y el Internet de las cosas*. Catarata, 2016.
- [12] Steve Todd. *AMPed At UC Berkeley*. URL: http://stevetodd.typepad.com/my_weblog/2011/08/amped-at-uc-berkeley.html.
- [13] TPC. *About TPC*. URL: <http://www.tpc.org/information/about/abouttpc.asp>.
- [14] TPC. *TPC EXPRESS BENCHMARK™ HS, Standard Specification Version 1.3.0*.
- [15] TPC. *TPCx-HS - Ten Most Recently Published Results*. URL: http://www.tpc.org/tpcx-hs/results/tpcxhs_last_ten_results.asp.
- [16] TPC. *Transaction Processing Performance Council (TPC) Launches TPCx-HS, the First Vendor-Neutral, Industry Standard Big Data Benchmark*. URL: http://www.tpc.org/information/other/tpcx-hs_press_release_final.pdf.
- [17] P. Zikopoulos y C. Eaton. *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.