# CONSTRUYENDO RAIDS DISTRIBUIDAS Y ACELERADAS DE MANERA RENTABLE Y FIABLE

—

Aleksandr Khas Niskim. Mustafa Raíique, Ali R. Butt, Sudarshan S. Vazhkudai y Dimitrios S. Nikolopoulos

Estudiante: María Fernanda Díaz

Universidad Distrital Francisco José de Caldas

April 29, 2016

# Introducción

Almacenar y recuperar de una manera fiable y rentable grandes cantidades de datos como los producidos por instrumentos científicos como el Gran Colisionador de Partículas LHC,

Figure: Ilustración. Tomado de: http://www.taringa.net/posts/info/16118346.html

# Introducción

o por observaciones detalladas como el Mapa Tridimensional del Universo SDSS, puede desbordar la capacidad de cualquier sistema de cómputo/almacenamiento o elevar sus costos a alturas inalcanzables; con la complicación de que entre más datos haya, mayores serán las cantidades de pérdidas o errores

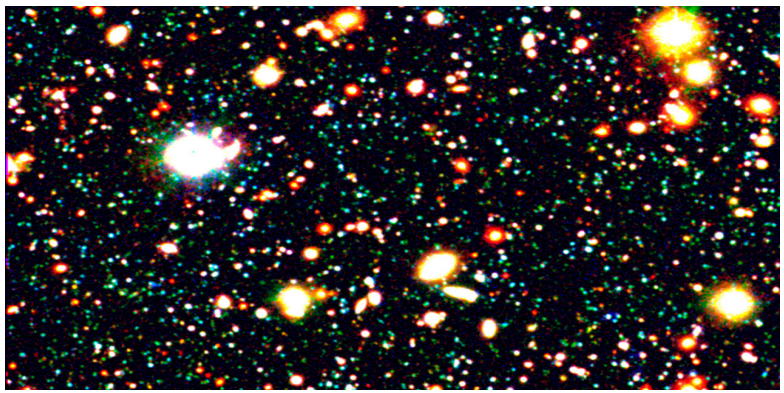

Figure: Universalidad.Tomada de: https://www.fayerwayer.com/2012/12/las-mejores-fotos-del-universo-de-2012/

# RAID(matriz redundante de discos independientes)

"Aumentar la tolerancia a fallos"
"Conjunto redundante de discos independiente"
"Que permite la reconstrucción de datos después de un fallo"

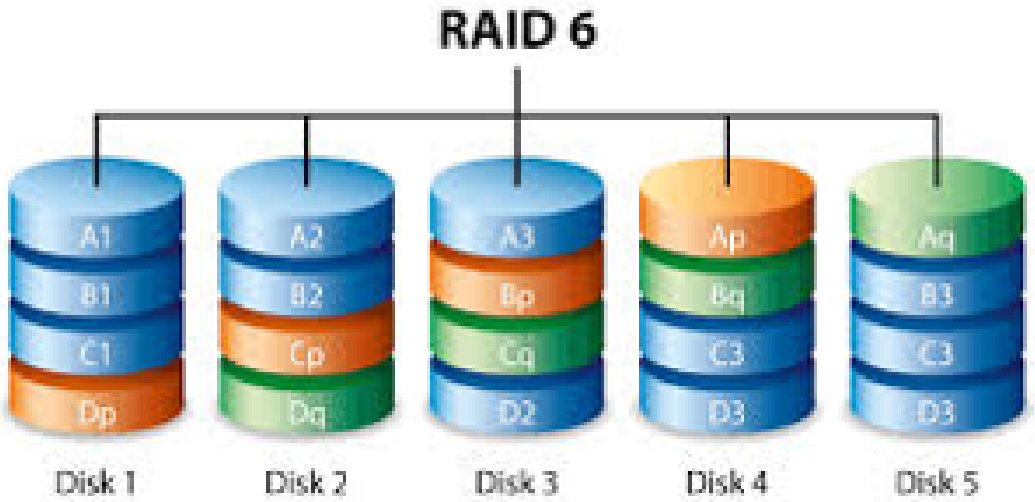


Figure: RAID: sistema de almacenamiento de datos en tiempo real que utiliza múltiples unidades de almacenamiento de datos (discos duros o SSD). Tomado de: http://www.seagate.com/la/es/manuals/network-storage/business-storage-nas-os/raid-modes/

# RAID (matriz redundante de discos independientes)

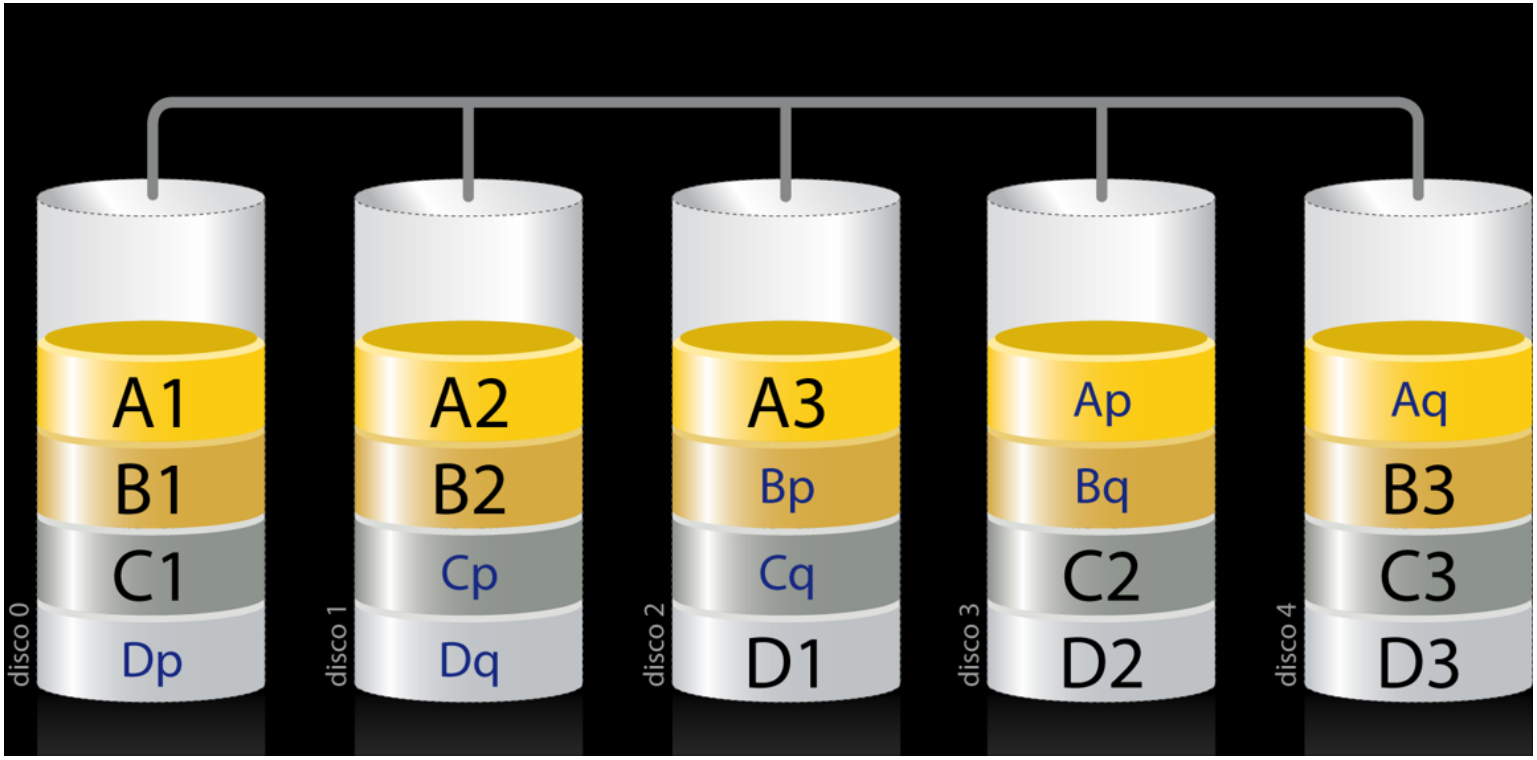


Figure: Diagrama de una configuración RAID 6. Cada número representa un bloque de datos; cada columna, un disco; p y q, códigos Reed-Solomon. Tomado de: http://www.abdata.es/sistemas%20raid.html

# RAID (matriz redundante de discos independientes)

Permite varias unidades de trabajo

Mejora el rendimiento

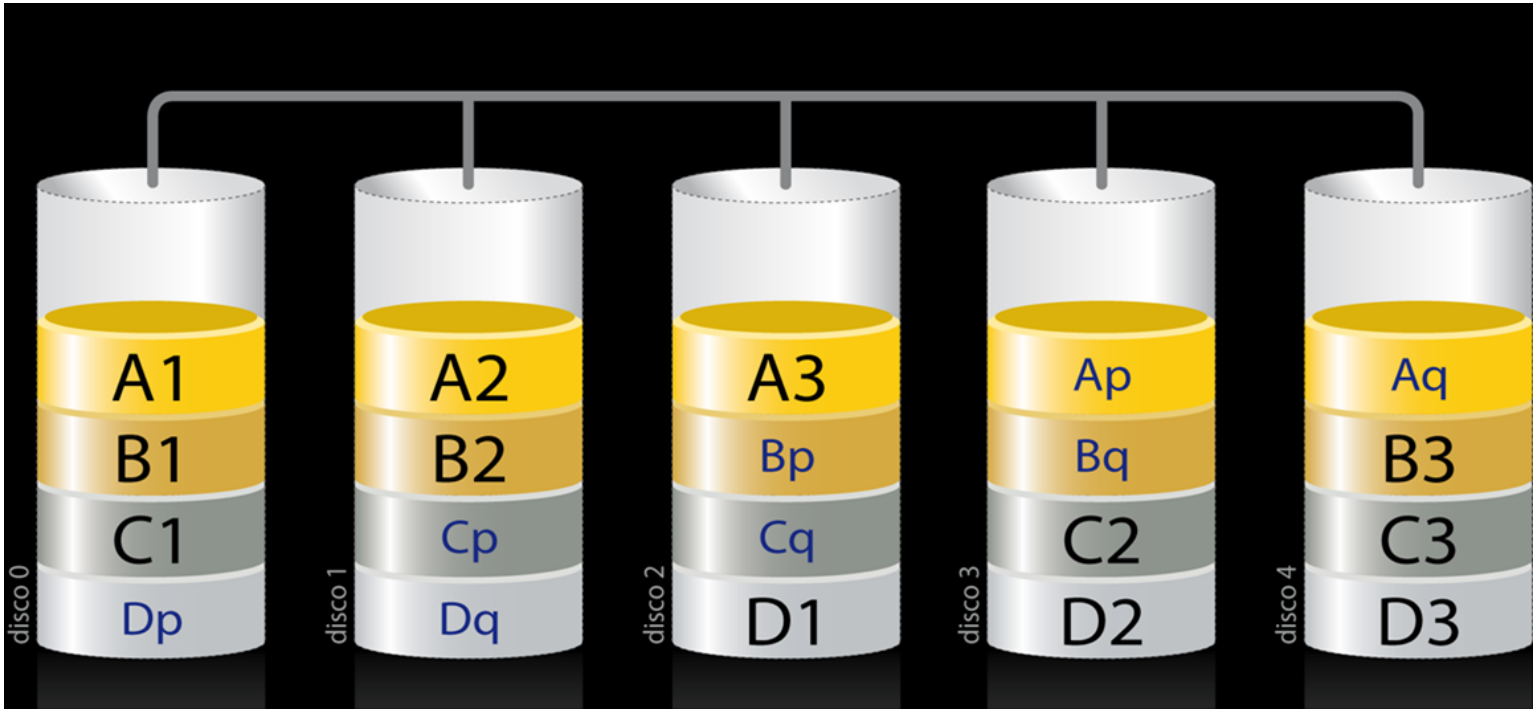


Figure: Ej. RAID 6 contiene k discos de datos y m (m=2) discos de paridad y puede recuperar datos del fallo simultáneo de m discos.

Tomado de: http://www.abdata.es/sistemas%20raid.html

# Graphics Processor Unit

Por otro lado, las GPU Graphics Processor Unit / Unidades de procesamiento gráfico, dedicadas originalmente a dicho procesamiento, se han convertido en una valiosa ayuda al dedicarse a otros propósitos, pues contienen grandes cantidades de chips dedicados.


Figure: Procesadores Nvidia

Tomado de: http://www.retrevo.com/content/blog/2010/05/five-laptop-features-matter-most

# Graphics Processor Unit

NVIDIA y AMD han pasado de ser aceleradores paralelos y dispositivos para el cálculo de paridad en proyectos de bajo costo.



Figure: Placa de video Nvidia GPU

Tomado de: http://www.akshatblog.com/amd-apu-processors-with-best-gpu-or-integrated-graphics/

# Contribución de la investigación

Este documento presenta los resultados de la construcción de un modelo de procesamiento de grandes cantidades de datos, que utiliza GPUs junto a los núcleos de procesamiento de datos de matrices RAID, dirigido al manejo de archivos, mediante la utilización del software libre (y potente) Lustre PFS.
La investigación aporta nuevo conocimiento para el manejo eficiente de datos, con bajo costo y de manera fiable.

# Contribución de la investigación

El aporte mas significativo es el de poder manipular la gran cantidad de chips dedicados de las GPUs para un trabajo nuevo, como es la E/S de datos de archivos, lo cual baja costos. Por otra parte las RAID se hacen funcionar de manera flexible bajo demanda, proporcionan paridad (fiabilidad) y recuperación rápida en caso de fallos (reconstrucción de la matriz).
Se avanza en el conocimiento de las posibilidades de los cluster manejados por Linux (Lustre PFS), que almacenan la información en bandas de archivos en varios discos independientes o distantes.

# Evidencias de soporte

Los autores presentan un análisis detallado de todas las herramientas tecnológicas utilizadas en su propuesta, las matrices RAID-6 con códigos de liberación, el sistema de archivos paralelos o en bandas de Lustre (mundialmene utilizado), con sus componentes: cliente, metadatos del servidor, servidores de almacenamiento de objetos OSS y de targets OST; y su capacidad de almacenar archivos de objetos del mismo tamaño en bandas apiladas sobre los discos, además de su gran tolerancia a fallos.

# Evidencias de soporte

Hacen hincapié en la referencia KGPU de baja latencia. Presentan el diseño de su arquitectura de alto nivel que une las GPU a los sistemas RAID. El uso de una configuración RAID-1 (gama baja) para los metadatos, que representan el 1% del almacenamiento, lo que reduce costos. Durante un fallo, el sistema es capaz de utilizar datos sobrevivientes para reconstruir los objetos perdidos.

# Evidencias de soporte

Se integra todo el proceso de paridad en el módulo del cliente. De manera flexible, el sistema intercambia configuraciones RAID-1 y RAID-6 de acuerdo al tamaño de los archivos. La reconstrucción de archivos afectados se hace en paralelo, utilizando el espacio del cliente.
En resumen se hace una buena economía de las técnicas elegidas, maximizando su desempeño. Finalmente los autores miden el rendimiento de su sistema en banco de pruebas de E/S y bajo una carga de trabajo real.

# Comentarios de árbitro

El documento está ampliamente referenciado, contiene fundamentación teórica, presenta de manera clara las herramientas tecnológicas y se observa un estudio detallado, por parte de los autores, de cada uno de sus componentes. Este documento aporta nuevos conocimientos con relación a las RAID_6 y su posibilidad de almacenar y recuperar de una manera fiable y rentable grandes cantidades de datos como los producidos por instrumentos científicos.

# Bibliografía


1. M. D. R. Alex Osuna, Siebo Friesenborg, "Considerations for raid-6 availability and format/rebuild performance on the ds5000," 2009, document Number: REDP-4484-00.
2. B & H Foto & Electronics Corp., "Active Storage 16TB ActiveRAID Hard Drive Array," 2011, http://www.bhphotovideo.com/c/product/697437-REG/Active_Storage_AC16SFC02_16TB_ActiveRAID_Hard_Drive.html.
3. J. Michalakes and M. Vachharajani, "Gpu acceleration of numerical weather prediction," in IEEE International Symposium on Parallel and Distributed Processing (IPDPS), april 2008, pp. 1–7.
4. C. Trapnell and M. C. Schatz, "Optimizing data intensive gpgpu computations for dna sequence alignment," Parallel Comput., vol. 35, pp. 429–440, August 2009.

# Bibliografía


1. M. Fatica, "Accelerating linpack with cuda on heterogenous clusters," in Proceedings of 2nd Workshop on General Purpose Processing on Graphics Processing Units, ser. GPGPU-2. New York, NY, USA: ACM, 2009, pp. 46–51.
2. T. D. Hartley, U. Catalyurek, A. Ruiz, F. Igual, R. Mayo, and M. Ujaldon, "Biomedical image analysis on a cooperative cluster of gpus and multicores," in Proceedings of the 22nd annual international conference on Supercomputing, ser. ICS '08. NewYork, NY, USA: ACM, 2008, pp. 15–25.
3. M. M. Rafique, A. R. Butt, and D. S. Nikolopoulos, "A capabilities-aware framework for using computational accelerators in data-intensive computing," J. Parallel Distrib. Comput., vol. 71, pp. 185–197, February 2011.

# Bibliografía


1. M. Curry, A. Skjellum, H.Ward, and R. Brightwell, "Arbitrary dimension reed-solomon coding and decoding for extended raid on gpus," in Petascale Data Storage Workshop, 2008. PDSW '08. 3rd, nov. 2008.
2. D. A. Alcantara, A. Sharf, F. Abbasinejad, S. Sengupta, M. Mitzenmacher, J. D. Owens, and N. Amenta, "Real-time parallel hashing on the gpu," ACMTrans. Graph., vol. 28, pp. 154:1–154:9, December 2009.