

INGENIERÍA DE SISTEMAS DE INFORMACIÓN (2017-2018)
DOBLE GRADO EN INGENIERÍA INFORMÁTICA Y MATEMÁTICAS
UNIVERSIDAD DE GRANADA

Software científico: El experimento ATLAS

Luis Balderas Ruiz

20 de mayo de 2018

Índice

1	Introducción	3
2	¿Qué es el CERN?	3
3	El experimento ATLAS	4
3.1	Del colisionador al portátil	5
3.1.1	Colisión	5
3.1.2	Reconstrucción	6
3.1.3	Simulación	6
3.1.4	Análisis	6
3.1.5	El Framework de ATLAS	6
4	Bibliografía	8

1. Introducción

La ciencia es actualmente el motor de nuestra sociedad. La tradición científica, aunque en sus orígenes era más una práctica de eruditos deseosos de calmar la necesidad de respuestas a preguntas trascendentes, se ha ido transmitiendo de generación en generación y, con ella, la forma de investigar. Sin embargo, el desarrollo de la computación y el software ha sido el revulsivo imprescindible para llegar al mundo tal y como lo conocemos hoy. Este documento se centra en un ejemplo concreto y a la vez paradigmático que subraya la necesidad de un soporte informático con una potencia sin precedentes: el experimento ATLAS, capitaneado por el CERN (Organización Europea para la Investigación Nuclear) en Ginebra, Suiza. Para ilustrar la importancia y belleza de este proyecto sin precedentes, haré un breve recorrido por el CERN mismo y la parte físico-teórica de ATLAS para acabar centrándome en su alma: el sistema de información que hace posible recoger, almacenar y darle sentido a la cantidad desbordante de datos que genera cada instante.

NOTA: La información recogida se ha extraído de las citas siguientes: [1], [9], [5], [7], [6], [8], [2], [3] y [4].

2. ¿Qué es el CERN?

El nombre CERN proviene del acrónimo en francés *Conseil Européen pour la Recherche Nucléaire*". Es una organización fundada en 1952, afincada en Ginebra, con el mandato claro de establecer un centro de investigación especializado en física fundamental (especialmente de partículas) a nivel europeo. Los doce países miembro fundadores fueron Alemania, Bélgica, Dinamarca, Francia, Grecia, Italia, Noruega, Países Bajos, Reino Unido, Suecia, Suiza y Yugoslavia (se retiró en 1961) y luego se unieron más países, en particular España en 1961.

Hoy en día está formado por 21 estados miembro y 7 observadores. Entre sus mayores méritos ingenieriles se encuentra el LHC (Large Hadron Collider), el acelerador de partículas más largo del mundo (27 km), en el que colisionan hadrones con energías cercanas a la velocidad de la luz. Esta infraestructura hace posible el estudio de la famosa antimateria o el descubrimiento de partículas tan importantes como el bosón de Higgs, con el que se completa la teoría del Modelo Estándar de partículas fundamentado a mediados del siglo pasado.

Sin embargo, el gran potencial tecnológico y humano del CERN no han conseguido solamente avances en física de partículas. En el corazón del CERN, de la mano de Tim Berners-Lee, fue inventada la World Wide Web en 1989, primero como una red interna para resolver una necesidad imperiosa de compartir información entre los investigadores y que fue finalmente liberada en 1993 para disfrute de todos nosotros.

La idea de red interna se desarrolló hasta dar lugar a lo que hoy es el Worldwide LHC

Computing Grid (WLCG), que es una colaboración mundial entre centros de computación para dar almacenamiento, distribución y análisis a los datos generados cada año por el LHC. Este WLCG tiene 4 capas de componentes software:

- Software específico de física, como los programas ROOT o POOL (framework para almacenamiento de datos) u otros softwares para modelar la generación, propagación o interacción entre las partículas elementales.
- Middleware, dando soporte distribuido a computadores a través de la red. El middleware más importante es European Middleware Initiative (ARC, gLite, UNICORE, Globus Toolkit...).
- Hardware, con Quattor como software de mantenimiento para automatización de tareas de actualización o dCache system o CASTOR que proporcionan herramientas específicas para almacenamiento y acceso a los datos para distintas simulaciones y análisis (lo veremos más tarde en ATLAS).
- Networking. Debido a las grandes necesidades de conexión entre equipos vía Internet, se han desarrollado conexiones de fibra óptica que trabajan a una velocidad de 10Gbps (LHC Optical Private Network).

Entre los experimentos más importantes, situados como detectores o sensores dentro del LHC me atrevo a destacar 4: ALICE, dedicado a detectar el plasma de quarks-gluones (estado de la materia instantes posteriores al Big Bang), ATLAS y CMS, de propósito general y con intención de encontrar partículas fundamentales, dimensiones extras o materia oscura (cada uno con soluciones tecnológicas y diseño diferentes) y LHCb, dedicada a estudiar las diferencias entre la materia y la antimateria a través del quark b (beauty). A continuación, me centro en el experimento que nos ocupa: ATLAS.

3. El experimento ATLAS

Millones de haces de partículas corriendo a través del LHC colisionan en el centro del detector ATLAS generando nuevas partículas, las cuáles se alejan en todas las direcciones. En ATLAS, 6 subsistemas de detección organizados en capas se encargan de reconocer el recorrido, el momento y la energía de las partículas, permitiendo identificarlas de forma única.

Las interacciones en los detectores de ATLAS generan enormes flujos de datos. Para poder procesarlos, se usa un avanzado sistema de disparadores para comunicar al detector qué eventos almacenar y cuáles ignorar. Con 46 metros de largo, 25 metros de ancho y de alto, las 7000 toneladas de ATLAS conforman el mayor detector de partículas jamás construido. Desde febrero de 2012, en ATLAS trabajan más de 3000 científicos de 174 instituciones y 38 países.

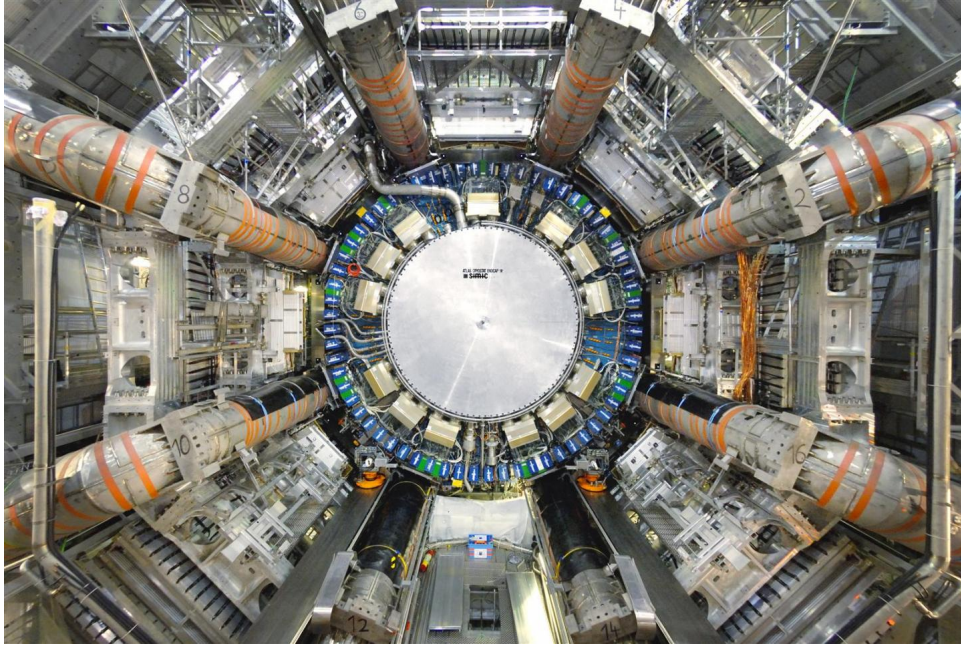


Figura 3.1: Detector ATLAS empotrado en el LHC

Veamos esto de forma más detallada:

3.1. Del colisionador al portátil

3.1.1. Colisión

Cada 50 nanosegundos se producen colisiones de protones en el LHC. La colisión de dos protones produce una pequeña explosión, donde parte de la energía cinética de los protones se convierte en nuevas partículas. La mayoría de las partículas resultantes son muy inestables y decaen muy rápidamente en una cascada de partículas más ligeras. Los ingenieros diseñaron el detector para identificar la energía y dirección de dichas partículas, de forma que es posible hacer el camino contrario en la cascada y obtener evidencias sobre las primeras partículas (las más interesantes). Es el caso del bosón de Higgs.

En cada colisión, un complejo disparador dividido en tres partes reduce el número de colisiones de 20 millones a cerca de 500 por segundo en busca de señales de partículas muy energéticas y descartando las de baja energía. Por cada colisión seleccionada, se escribe en disco 1MB (resultado de almacenar toda la información sobre las energías). Se almacenan más de un billón de eventos cada año, lo que es equivalente a un petabyte de información.

3.1.2. Reconstrucción

Después de que todos los distintos detectores estudien una colisión, el sistema reconstruye una colección de al menos 100 partículas de la forma más precisa posible a partir de su tipo, energía y dirección. En la práctica, la reconstrucción está formada por más de 200 algoritmos que necesitan 20 segundos en un core con 2GB de memoria. Para ello, una granja de 6000 computadores conforman el CPD del CERN. Los resultados de la reconstrucción son distribuidos a lo largo del planeta, de forma que los físicos pueden analizar los eventos en menos de una semana desde que se produjo la colisión.

3.1.3. Simulación

Como en todas las ciencias, la simulación de los experimentos nos ayuda a entender enormemente. En el caso de ATLAS, existe un simulador extraordinario donde se ponen en práctica todas las técnicas conocidas. Los resultados de la simulación son incluidos en un modelo virtual que ha resultado ser bastante veraz y que aún se sigue utilizando para analizar nuevas ideas y evaluar factores de corrección. Se pretende calcar las condiciones del simulador a las reales e incluso, mantener el mismo número de simulaciones que de eventos reales. A pesar de todo ello, la simulación es tremendamente costosa e intensiva desde el punto de vista de la CPU (1 segundo por evento).

3.1.4. Análisis

Tras la construcción del modelo y la obtención de los datos, los distintos grupos analizan los resultados. El input de datos a analizar suele estar alrededor de un petabyte de información real y simulada. Después, cada grupo generan conjuntos de datos propios con los que trabajan. Analizar el petabyte completo suele llevar una semana, así que para mejorar el rendimiento manteniendo la flexibilidad necesaria para crear nuevas ideas, se definen conjuntos de datos intermedios de un terabyte, el cual puede ser analizado en una noche. Se trata, por tanto, de un trabajo intensivo de minería de datos con grandísimas cantidades de información.

Como hemos podido ver, ATLAS sobrevive gracias a la gran capacidad de computación y el software que soporta esa inmensa cantidad de datos.

3.1.5. El Framework de ATLAS

Excepto la primera fase de reconstrucción, los diferentes procesos de ATLAS (incluida la simulación) se realizan en los clusters de ATLAS, que están formados por al menos 100 centros de computación alrededor del mundo. Cada centro tiene preinstalado el software ATLAS y reciben los distintos datasets para ser analizados. El sistema divide automáticamente las tareas en procesos en un solo core con un par de gigabytes de memoria y que trabaja algunas horas. El framework de todo el software de ATLAS sigue una arquitectura basada en pizarra e implementada en C++. En él, un evento es cargado en memoria y después los algoritmos se ejecutan secuencialmente, leyendo una o más colecciones de

objetos de la pizarra y escribiendo una o más salidas. Cuando todos los algoritmos han terminado, se escriben las colecciones de datos en el disco y se carga el siguiente evento. En definitiva, es un proceso en una hebra ejecutado en un sistema operativo Linux. El diseño en C++ para la transición entre objetos de la pizarra es relativamente sencillo y estable: los objetos tienen un significado físico bien definido y su complejidad reside en los algoritmos. La configuración se realiza con una capa en Python para ganar en flexibilidad, haciendo posible desacoplar un algoritmo tradicional y reemplazarlo por uno más moderno a través de un algoritmo basado en árboles de decisión.

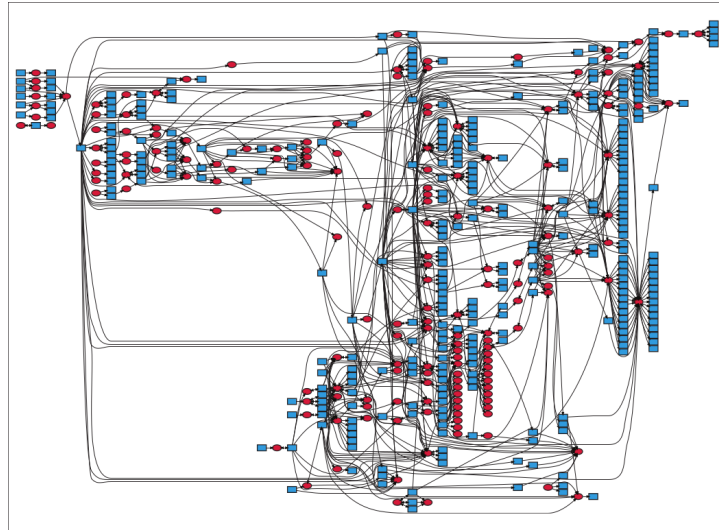


Figura 3.2: Flujo de datos de izquierda a derecha. Las elipses rojas representan algoritmos y los rectángulos azules, objetos de la pizarra

4. Bibliografía

Referencias

- [1] <https://home.cern/about>, Consultado el 17 de mayo de 2018.
- [2] <https://home.cern/about/computing>, Consultado el 17 de mayo de 2018.
- [3] <https://home.cern/about/computing/grid-software-middleware-hardware>, Consultado el 17 de mayo de 2018.
- [4] <https://home.cern/about/computing/worldwide-lhc-computing-grid>, Consultado el 17 de mayo de 2018.
- [5] <https://home.cern/about/experiments/alice>, Consultado el 17 de mayo de 2018.
- [6] <https://home.cern/about/experiments/atlas>, Consultado el 17 de mayo de 2018.
- [7] <https://home.cern/about/experiments/cms>, Consultado el 17 de mayo de 2018.
- [8] <https://home.cern/about/experiments/lhcb>, Consultado el 17 de mayo de 2018.
- [9] David Rousseau. *The Software behind the Higgs Boson Discovery*. *IEEE Software* 29, 5 (September 2012), 11-15. 2012.