

## ATLAS DATA CHALLENGES

Santiago González, Vicente Lara, Eduardo Ros, José Salt, Francisco Javier Sánchez, Marcel Vos

*Instituto de Física Corpuscular (CSIC-Universitat de València) Apartado de Correos 22085 E-46071 (Valencia)*

Con los experimentos del LEP (Large Electron Positron) ha finalizado una era en la investigación de las Altas Energías y ahora se abre otra caracterizada por la preparación de cuatro experimentos – ATLAS, CMS, ALICE y LHCb – en el LHC (Large Hadron Collider) del CERN (Laboratorio de Física de Partículas, Ginebra).

La comunidad de Físicos de Partículas que estamos preparando estos experimentos y que analizaremos los datos que se obtengan de ellos nos enfrentamos a un problema técnico de envergadura: la inmensa cantidad de datos que se van a recolectar que ya sabemos que no se va a poder procesar con el sistema informático y de cálculo que hemos utilizado hasta ahora.

Tres de los cuatro experimentos del LHC generarán al menos 1 Petabyte (PB), de datos en bruto cada año durante un periodo de 10 años (la frecuencia de lectura de los detectores de ATLAS y CMS se espera que sea de 100 Hz, como muestra la figura 1, y el tamaño de datos brutos para cada suceso es de 1MB). Así pues, el reto tecnológico [1] es suministrar un acceso rápido a muestras del tamaño del TB y un acceso transparente a los recursos de cálculo distribuidos por el mundo, para posibilitar el trabajo de análisis. Por lo tanto, la investigación científica de la próxima generación de experimentos requerirá cálculo intensivo y análisis de bases de datos compartidos y de gran escala.

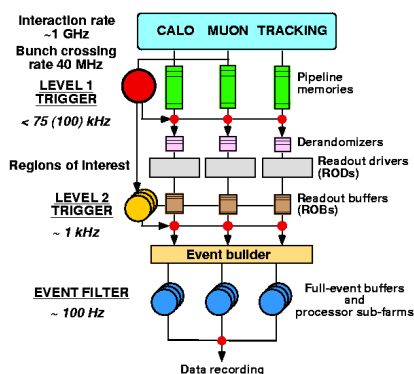


Figura 1.- Arquitectura de los tres niveles de trigger del detector ATLAS

La iniciativa GRID [2] es una solución al problema formulado aportando un middleware orientado a resolver muchos de los problemas de un modelo distribuido, con la provisión de herramientas de publicación de recursos, distribuidores de trabajos, monitoraje, replica de bases de datos, control de acceso, etc... El desarrollo de un testbed en el campo de Altas Energías dentro de los proyectos DataGrid [3] coordinado por el CERN y CrossGrid [4] va a suponer un ensayo definitivo del modelo de computing en LHC.

Con el objetivo de validar el modelo de computing, de datos y el software del detector ATLAS, la colaboración ha desarrollado un programa de “Data Challenges” (DC) [5]. Estos DC permitirán la elaboración de un Interim MoU sobre computing. Este MoU es necesario para trabajar en el desarrollo de software, y plantear el desarrollo de un prototipo de computing. Los resultados que se obtendrán se utilizarán para la elaboración del Technical Design Report (TDR), el cual contará con una simulación realista basada en Geant4, con producción masiva de sucesos y debe ser capaz de preparar a la colaboración para el análisis físico.

ATLAS trabaja en contacto directo con los proyectos GRID. Los científicos participantes en los ATLAS DC están integrando el software desarrollado en estos proyectos con el objetivo de utilizar las herramientas y el middleware de acceso remoto a datos y computing distribuido en desarrollo en estos proyectos GRID.

En el caso del IFIC (Valencia) está participando en los DC de ATLAS con el objetivo de:

- Proveer a la comunidad científica sucesos producidos en nuestra granja (GoG, ver figura 2) para la realización del TDR (mitad 2003).
- Probar en nuestra granja el nuevo software desarrollado, como por ejemplo: Geant4, los nuevos modelos de datos, etc...
- Utilizar herramientas – tecnología GRID y el software desarrollado en el proyecto LHC Computing Grid (LCG) en la generación de sucesos.



Figura 2.- Grupo de Ordenadores para el Grid (GoG) en el IFIC de Valencia

Los sucesos producidos por la colaboración se generaron con PYTHIA y se simularon a través de ATLSIM (Dice, GEANT3). 37 institutos y 18 países están participando en los ATLAS DC debido a la gran cantidad de tiempo de CPU necesaria para el procesamiento de dichos sucesos. En total alrededor de 50 millones de sucesos en 35000 ficheros han sido generados por toda la colaboración, utilizando unos 3200 procesadores (Pentium III 500 MHz), 1722K horas de CPU y con una capacidad de 100 TB. El IFIC de Valencia ha contribuido hasta el momento con 2.2 millones de sucesos en 2350 ficheros, utilizando 100

procesadores (Athlon 1.4 GHz), 30K horas de CPU y 2.5 TB de almacenamiento.

La tabla 1 muestra el tamaño de los sucesos y el tiempo de CPU necesitados, para las muestras de datos simulados en el IFIC. Para el procesamiento de dichos datos se han utilizado aproximadamente 100 Athlons a 1.4 Ghz, con 1GB de memoria.

REFERENCIAS

- [1] S. L. Lloyd, P. W. Jeffreys, *The LHC Computing Challenge*, Internal note (1999).
- [2] I. Foster, C. Kesselman, *The GRID: Blueprint for a New Computing Infrastructure* (1999).
- [3] EDG/DataGrid: <http://www.eu-datagrid.org>
- [4] CrossGrid: <http://www.eu-crossgrid.org>
- [5] Atlas Data Challenges Collaboration, *ATLAS DATA CHALLENGES*, Document presented for the ATLAS Executive Board (2001)

Tipo de sucesos	Tamaño por suceso (MB)	Con pile-up a baja lumi (MB)	Con pile-up a alta lumi (MB)	Tiempo por suceso (s)	Tiempo por suceso a baja lumi (s)	Tiempo por suceso a alta lumi (s)
Di-jets	2.5	3	-	250	30	-
WH→μνgg	2.5	3.6	10	200	30	130
WH→μνbb	2.5	3.7	7.5	200	30	150

**Tabla 1.-** Valores medios para diferentes muestras de sucesos producidos en el IFIC.

---