

# ATLAS EventIndex y EventWhiteboard. Un sistema de indexado y catalogado de eventos para experimentos con grandes cantidades de datos.

*martes, 13 de marzo de 2018 11:45 (10)*

El EventIndex de ATLAS corre actualmente en producción indexando todos los eventos reales y simulados, a lo largo de sus distintos reprocesados, para obtener un catálogo completo de los eventos.

Actualmente ATLAS procesa del orden de 30 billones de eventos al día (350 Hz de media). Estos procesos leen de media 8 Millones de ficheros y producen 3 Millones de nuevos ficheros cada día.

Para cada uno de los eventos indexados se recolecta un pequeño conjunto de metadatos que incluye sus identificadores de evento (run y event numbers, trigger stream, luminosity block), patrones del trigger (l1, l2, ef), así como otros datos que caracterizan el evento y referencias (punteros) a las localización física en ficheros en el GRID para cada fase del procesado.

Esta contribución describe la arquitectura general del EventIndex, la producción de datos con la infraestructura distribuida del GRID y su recolección a través de una infraestructura distribuida usando tecnología de mensajería y un ObjectStore para almacenamiento temporal.

Los datos recolectados, del orden de TeraBytes, son finalmente consolidados en una instancia HADOOP en el CERN, y

una parte de los mismos se copia también a una base de datos Oracle, para proveer los casos de uso que necesitan menor latencia.

Los casos de uso actuales incluyen la selección de eventos individuales en base a criterios de búsqueda (event picking para RAW y AOD data), cálculos de estadísticas para grandes cantidades de eventos (trigger counting para cada run de datos reales), cálculo de eventos duplicados para distintos datasets, y cálculo de overlaps entre datos derivados (DAOD). También se usa para hacer crosschecks de completitud y consistencia en la producción de datos.

Revisamos los retos que impone el incremento esperado del rate de producción que alcanzará los 35 Billones de eventos reales por año en el Run 3, y los 100 Billones en el Run 4. Para eventos simulados los números son incluso mayores, con 100 Billones de eventos por año en el Run 3 y 300 Billones en el Run 4.

También analizaremos los retos que enfrentamos para acomodar el futuro Event WhiteBoard de ATLAS, una herramienta cuyo objetivo es el marcado de eventos con metadatos totalmente flexibles y la creación de datasets virtuales por parte de usuarios finales y grupos de producción, integrado con las capacidades que proporciona el EventIndex.

## Summary

**Primary author(s) :** FERNANDEZ, Alvaro (IFIC); Sr. GARCÍA MONTORO, Carlos (IFIC); FULLANA TORREGROSA, ESTEBAN (IFIC); Sr. SANCHEZ MARTINEZ, Fco. Javier (IFIC); Dr. SALT, JOSE (IFIC); Sr. APARISI POZO, Javier (IFIC); Dr. LOZANO BAHILLO, Julio (Instituto de Física Corpuscular (CSIC-UV)); Dr. GONZALEZ DE LA HOZ, Santiago (IFIC)

**Presenter(s) :** FERNANDEZ, Alvaro (IFIC)

**Clasificación de la sesión :** Informática y procesado de datos

**Clasificación de temáticas :** Informática y procesamiento de datos