

Desarrollos y herramientas para el almacenamiento y procesamiento de grandes cantidades de datos aplicado en ATLAS Tier2 y EventIndex EI3

Wednesday, 14 October 2020 12:05 (10)

En esta presentación hablaré de los desarrollos realizados y herramientas utilizadas para el almacenamiento y procesamiento de grandes cantidades de datos, y aplicado en particular en el EventIndex EI3 dentro del experimento ATLAS.

Usamos un Object Store basado en CEPH para el almacenamiento temporal de información sobre los eventos producidos a nivel mundial en el GRID y el Tier0, que se consolidan finalmente en HADOOP.

Los nuevos desarrollos para el próximo run3 se centran en mejorar la usabilidad, y satisfacer los rendimientos necesarios. En el nuevo prototipo usamos HBase y una estrategia columnar para el almacenamiento de los datos, así como una capa unificada para el acceso basada en Apache Phoenix.

Se utilizan herramientas como InfluxDB y Grafana para la monitorización de datos basados en series temporales, así como GitLab para el versionado y control de código.

El procesamiento de los datos consolidados se realiza con entornos como MapReduce y Spark, que proveen grandes capacidades de análisis en paralelo.

Primary author(s) : FERNANDEZ CASANI, Alvaro (IFIC)

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

Session Classification : Informática y procesamiento de datos