

# Sistema de Análisis Distribuido de Datos Procedentes de ATLAS Utilizando el GRID

F. Fassi<sup>1</sup>, J. Lozano<sup>1</sup>, M<sup>a</sup>. D. Jordán<sup>1</sup>, L. March<sup>1</sup>, J. Salt<sup>1</sup>, S. González de la Hoz<sup>1</sup>, J. Sánchez<sup>1</sup>, A. Fernández<sup>1</sup> y M. Kaci<sup>1</sup>

<sup>1</sup> IFIC –Instituto de Física Corpuscular (Centro Mixto CSIC – U. València), Edificio de Institutos de Investigación, Apartado de Correos 22085, E-46071 Valencia.

## I. INTRODUCCIÓN

ATLAS es un experimento de Física de Altas Energías que adquirirá datos de las colisiones protón-protón que tendrán lugar en el acelerador LHC (Large Hadron Collider). Dichos datos serán del orden del Petabyte al año, y se producirán además cantidades similares de datos simulados con el sistema de producción de ATLAS<sup>1</sup>. Con el fin de facilitar a los físicos de la colaboración de ATLAS tanto el procesamiento como el análisis de tal volumen de datos, ATLAS está desarrollando un sistema de análisis distribuido denominado ADA<sup>1</sup> (Atlas Distributed Analysis). El principal enfoque del proyecto ADA es identificar los componentes necesarios, contando con otros proyectos con el objetivo de obtener un sistema de análisis de datos útil basado en las tecnologías GRID. Entre los proyectos se encuentran: GANGA<sup>1</sup>, el sistema de producción de ATLAS<sup>1</sup>, ARDA<sup>1</sup> y DIAL<sup>2</sup>. Gracias a ésta estrategia, ha sido posible para ADA ensamblar todos estos componentes, aunque está aún en su fase de desarrollo.

## II. SISTEMA DE ANÁLISIS DISTRIBUIDO DE ATLAS

Utilizando ADA se espera llevar a cabo la manipulación y la explotación de los datos de acuerdo con el actual modelo establecido por ATLAS<sup>3</sup>. Para ello, los objetivos principales de ADA consisten en proporcionar un sistema fácil de acceder y con una funcionalidad que permite realizar el análisis en un entorno, donde los usuarios, los datos y el procesamiento están distribuidos geográficamente.

ADA (ver figura 1) está basado en una colección de Servicios Web, *High-level services*, (middleware necesario para implementar aplicaciones cliente-servidor) mediante los cuales se suministra el acceso a los catálogos, el envío de los trabajos y a su monitorización.

ADA, además tiene otros servicios que proporcionan el acceso a los ficheros y al software necesario para las aplicaciones. Los servicios de análisis de ADA (figura 1) están implementados de acuerdo con el modelo que está siendo desarrollado por el proyecto ARDA<sup>1</sup> de LCG. Mientras que el entorno del usuario está desarrollado por GANGA y ROOT<sup>4</sup>. La función fundamental de los servicios de análisis consiste en recibir una petición genérica de un trabajo

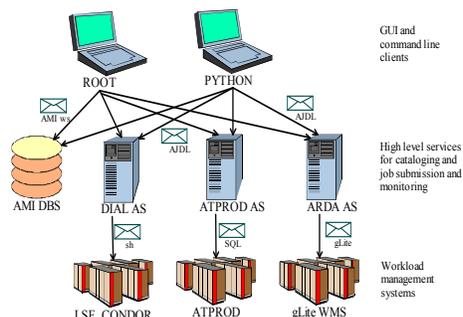


Figura1. El sistema de ADA consiste en 3 niveles: los servicios de análisis y de los catálogos residen en el nivel intermedio, mientras que las interfaces del usuario y los sistemas de procesamiento de los trabajos se encuentran en el primer y el último nivel respectivamente.

creando y ejecutando su correspondiente trabajo. El usuario (responsable de dicho trabajo) es capaz, desde su entorno de análisis, de monitorizar el progreso de su trabajo, incluso puede obtener un resultado parcial del mismo. Los servicios de análisis distribuido y de los catálogos están integrados al entorno del usuario mediante una interfaz denominada AJDL<sup>2</sup> (Abstract Job Definition Language). Ésta interfaz permite describir los datos y las definiciones genéricas de los componentes fundamentales del procesamiento (ver figura 2). Tales componentes vienen citadas a continuación:

- Datasets*<sup>1</sup>: describen los datos
- Transformación: está dividida en dos componentes: la aplicación y la tarea que transporta los datos necesarios para la configuración de la aplicación.
- Trabajo: es un ejemplo particular de una transformación que actúa sobre un dataset
- Task (Tarea): especifica la configuración de los usuarios para una aplicación.

### III. EL SERVICIO DIAL Y LAS APLICACIONES

Actualmente ADA tiene funcionando un solo servicio de análisis basado en DIAL<sup>2</sup>, el resto de los servicios de análisis se espera integrarlos pronto en ADA. El servicio DIAL se utiliza, también, para validar la arquitectura de ADA y la funcionalidad de sus componentes. Para ello, DIAL proporciona un servicio de análisis interactivo y no-interactivo usando colas rápidas de LSF y lentas de Condor respectivamente. El usuario (desde su entorno de análisis) interactúa con DIAL especificando la aplicación, el dataset y el task (figura 2). La aplicación y el task se someterán a DIAL formando un trabajo. DIAL, por su parte, dividirá los datasets en sub-datasets creando y ejecutando un sub-trabajo para cada uno de los sub-dataset. Finalmente concatenará el resultado de cada sub-trabajo dando lugar al resultado completo.

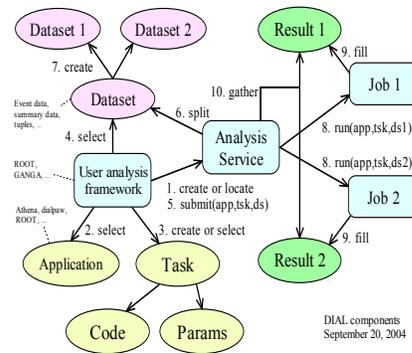


Figura 2. Los componentes que describen AJDL y sus conexiones con el entorno de análisis del usuario y con el servicio de análisis.

ADA es capaz de proporcionar el entorno apropiado para ejecutar un amplio rango de aplicaciones, pero el énfasis se concentra en desarrollar aquellas aplicaciones que permiten la generación, la simulación, la reconstrucción y el análisis de datos. El presente sistema de ADA dispone de tres aplicaciones<sup>5</sup>: *atlasopt*, *atlasreco* y *aodhisto*. Ésta última está dedicada exclusivamente para el análisis de datos<sup>6</sup>

### Referencias:

- <http://www.usatlas.bnl.gov/ADA>.
- <http://www.usatlas.bnl.gov/~dladams/dial/>.
- <http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/OO/domains/Reconstruction/AODESD/AODESDtaskforce.htm>.
- <http://root.cern.ch/>
- <http://www.usatlas.bnl.gov/~dladams/dial/releases/1.00/apps.html>.
- "ATLAS Distributed Analysis" presentaciones de F. Fassi, ATLAS software web, 8 de Diciembre de 2004 y 21 de Febrero de 2005.