Convocatorias 2016
Proyectos EXCELENCIA y Proyectos RETOS
Dirección General de Investigación Científica y Técnica
Subdirección General de Proyectos de Investigación

AVISO IMPORTANTE

En virtud del artículo 16 de la convocatoria **NO SE ACEPTARÁN NI SERÁN SUBSANABLES MEMORIAS CIENTÍFICO-TÉCNICAS** que no se presenten en este formato.

Es obligatorio rellenar los tres apartados (A, B y C). La parte C de la memoria no podrá exceder de 25 páginas.

Lea detenidamente las instrucciones para rellenar correctamente esta memoria, disponibles en la web de la convocatoria.

Parte A: RESUMEN DE LA PROPUESTA/SUMMARY OF THE PROPOSAL

A.1. DATOS DEL PROYECTO COORDINADO

INVESTIGADOR/ES COORDINADOR/ES

INVESTIGADOR COORDINADOR PRINCIPAL 1 (Nombre y apellidos)**:**

Santiago González de la Hoz

INVESTIGADOR COORDINADOR PRINCIPAL 2 (Nombre y apellidos)**:**

José Francisco Salt Cairols

TÍTULO GENERAL DEL PROYECTO COORDINADO: Hacia un genuino Tier-2 Federado Español de ATLAS para afrontar el reto de la gestión y procesado del Big Data del LHC

ACRÓNIMO DEL PROYECTO COORDINADO: ES-ATLAS-T2

RESUMEN DEL PROYECTO COORDINADO Máximo 3500 caracteres (incluyendo espacios en blanco):

Este proyecto propone una contribución al computing del experimento ATLAS (LHC) mediante una infraestructura GRID de almacenamiento y proceso de datos y a través de la participación en el desarrollo de software y en tareas de computing del experimento.

Es objetivo por un lado, proveer los servicios de proceso de datos de manera estable durante 24 horas al dia todos los días del año con una fiabilidad superior al 95% en las condiciones que necesita el experimento y, por otro lado, encaminarse hacia una genuina federación con un único punto de entrada que reúna los recursos de computación Tier-2 de ATLAS en España favoreciendo el uso de software y metodologías comunes y facilitando la comunicación con ATLAS, de manera que se mejore la eficacia y visibilidad en el experimento.

Como resultado se espera producir miles de millones de sucesos simulados anuales para diferentes procesos de fisica, almacenar grandes cantidades de datos necesarios para los análisis de física del experimento, pertenecer al núcleo de centros GRID suficientemente fiables para albergar datos críticos y proveer un primer nivel de soporte a los físicos locales de ATLAS. Además al final de proyecto estará disponible una infraestructura de TIER-2 en el GRID visible como una única entidad.

Gracias a la contribución de este proyecto al computing GRID del LHC, los físicos de altas energías españoles y del resto del mundo podrán almacenar, procesar y analizar los miles

de millones de sucesos producto de las colisiones protón-protón del acelerador LHC del CERN registradas por el experimento ATLAS. El IFIC y la UAM han venido participando activamente durante los últimos 10 años en una de las infraestructuras distribuidas de cálculo científico más grandes del mundo que cuenta con más de medio millón de núcleos de CPU y más de 300 Pbytes (1 Pbyte = 1000000 Gbytes) de almacenamiento en disco en más de 150 centros.

PALABRAS CLAVE DEL PROYECTO COORDINADO: CERN, LHC, ATLAS, TIER2, GRID, DISTRIBUTED COMPUTING, LCG, EGI, DATA ANALYSIS, BIG DATA

TITLE OF THE COORDINATED PROJECT: Towards a True Federated Spanish ATLAS Tier-2 to face the LHC Big Data management and processing challenges.

ACRONYM OF THE COORDINATED PROJECT: ES-ATLAS-T2

SUMMARY OF THE COORDINATED PROJECT Maximum 3500 characters (including spaces):

This project proposes a contribution to the ATLAS experiment (LHC) computing effort providing a GRID data storage and processing infrastructure and participating in the experiment computing tasks and software development.

The main goal is twofold: providing data processing services uninterrupted 24 hours a day, 365 days a year with a reliability greater than 95% in the conditions required by the experiment, and on the other hand, moving towards a true federation which will concentrate all the spanish Tier-2 resources in a single contact point favouring the usage of common software and methodologies making the communication with ATLAS easier in order to increase its efficiency and visibility.

As a result, it is expected to produce billions of monte carlo events every year for different physics channels, store huge amounts of data needed to pursue the analyses, belong to the core of GRID centers reliable enough to store critical data and provide a first level support to the ATLAS local physicists. Moreover, at the end of this project, there will be a GRID Tier-2 site visible and accessible as a single entity.

Thanks to this project's contribution to the LHC GRID computing, high energy physicists from Spain and their colleagues around the world will be able to store, process and analyze billion events produced by the LHC proton-proton collider at CERN registered by the ATLAS experiment.

IFIC and UAM have been participating actively during the last 10 years in one of the biggest distributed computing scientific infrastructure around the world composed by more than half a million CPU cores, 300 Pbytes (1 Pbyte = 1000000 Gbytes) of disk storage in around 150 centers.

KEY WORDS OF THE COORDINATED PROJECT: CERN, LHC, ATLAS, TIER2, GRID, DISTRIBUTED COMPUTING, LCG, EGI, DATA ANALYSIS, BIG DATA

A.2. DATOS DE LOS SUBPROYECTOS

SUBPROYECTO 1 *(el investigador o investigadores principales del subproyecto 1 son los coordinadores del proyecto coordinado)***:**

TÍTULO: Hacia un genuino Tier-2 Federado Español de ATLAS (site IFIC) para afrontar el reto de la gestión y procesado del Big Data del LHC

SUBPROYECTO 2:

INVESTIGADOR PRINCIPAL 1 (Nombre y apellidos)**:**

José del Peso Malagón

TÍTULO: Hacia un genuino Tier-2 Federado Español de ATLAS (site UAM) para afrontar el reto de la gestión y procesado del Big Data del LHC

Parte B: INFORMACIÓN ESPECÍFICA DEL EQUIPO

B.1. RELACIÓN DE LAS PERSONAS NO DOCTORES QUE COMPONEN EL EQUIPO DE TRABAJO (se recuerda que los doctores del equipo de trabajo y los componentes del equipo de investigación no se solicitan aquí porque deberán incluirse en la aplicación informática de solicitud). Repita la siguiente secuencia tantas veces como precise para cada uno de los subproyectos.

1. Nombre y apellidos: Carlos García Montoro
   Titulación: ingeniero
   Tipo de contrato: otros (actualmente desempleado)
   Duración del contrato: no procede
   Subproyecto al que pertenece (nombre y apellidos del investigador principal): Santiago González de la Hoz y José Fco. Salt Cairols

B.2. FINANCIACIÓN PÚBLICA Y PRIVADA (PROYECTOS Y/O CONTRATOS DE I+D+I) DEL EQUIPO DE INVESTIGACIÓN (repita la secuencia tantas veces como se precise en cada uno de los subproyectos participantes hasta un máximo de 5 proyectos y/o contratos por cada subproyecto)

SUBPROYECTO 1: Financial support of the IFIC staff participating in the Project

1. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Jose Fco Salt Cairols, Santiago González de la Hoz, Fco. Javier Sánchez Martínez y Álvaro Fernández Casaní
   Referencia del proyecto: FPA2013-47424-C3-1-R
   Título: Tier-2 Distribuido español para el experimento ATLAS (LHC) Fase 3 y su papel en la gestión y procesamiento de grandes cantidades de datos.
   Investigador principal (nombre y apellidos): Jose Fco. Salt Cairols
   Entidad financiadora: MINECO
   Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 01/01/2014-31/12/2016
   Financiación recibida (en euros): 1.252.350 euros
   Relación con el proyecto que se presenta: mismo tema
   Estado del proyecto o contrato: concedido

2. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Santiago González de la Hoz y Fco. Javier Sánchez Martínez
   Referencia del proyecto: FPA2012-39055-C02-01
   Título: Contribuciones al Experimento ATLAS en el colisionador hadrónico LHC
   Investigador principal (nombre y apellidos): M. Carmen García García
   Entidad financiadora: MINECO
   Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 01/01/2013-31/12/2015
   Financiación recibida (en euros): 910.000 euros
   Relación con el proyecto que se presenta: está muy relacionado
   Estado del proyecto o contrato: concedido

3. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Jose Fco Salt Cairols, Santiago González de la Hoz, Fco. Javier Sánchez Martínez y Álvaro Fernández Casaní
   Referencia del proyecto: FPA2010-21919-C03-01
   Título: Tier-2 Distribuido Español para el experimento ATLAS (LHC) fase 2
   Investigador principal (nombre y apellidos): Jose Fco. Salt Cairols
   Entidad financiadora: MICINN
   Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 1/1/2011-31/12/2013
   Financiación recibida (en euros): 711.900 euros
   Relación con el proyecto que se presenta: mismo tema
   Estado del proyecto o contrato: concedido

4. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Santiago González de la Hoz y Fco. Javier Sánchez Martínez
Referencia del proyecto: FPA2009-13234-C04-01
Título: Contribuciones al Experimento ATLAS en el colisionador Hadrónico LHC
Investigador principal (nombre y apellidos): M. Carmen García García
Entidad financiadora: MICINN
Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 1/1/2009-31/12/2012
Financiación recibida (en euros): 1.610.510 euros
Relación con el proyecto que se presenta: está muy relacionado
Estado del proyecto o contrato: concedido

5. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): José Fco. Salt Cairols
Referencia del proyecto: SEV-2014-0398
Título: Acreditación Centro de Excelencia Severo Ochoa (IFIC)
Investigador principal (nombre y apellidos): Juan José Hernández Rey García
Entidad financiadora: Ministerio de Economía y Competitividad, Secretaría de Estado de I+D+I
Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 1/7/2015-30/6/2019
Financiación recibida (en euros): 4.000.000 euros
Relación con el proyecto que se presenta: está muy relacionado
Estado del proyecto o contrato: concedido

SUBPROYECTO 2: Financial Support of the UAM staff participating in the Project

1. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Jose del Peso, Claudia Glasman
Referencia del proyecto: FPA2013-47424-C3-2-R
Título: Tier-2 Distribuido español para el experimento ATLAS (LHC) Fase 3 y su papel en la gestión y procesamiento de grandes cantidades de datos.
Investigador principal (nombre y apellidos): Jose del Peso
Entidad financiadora: MINECO
Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 01/01/2014-31/12/2016
Financiación recibida (en euros): 210.000 euros
Relación con el proyecto que se presenta: mismo tema
Estado del proyecto o contrato: concedido

2. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Jose del Peso, Claudia Glasman
Referencia del proyecto: FPA2010-21919-C03-03
Título: Tier-2 Distribuido Español para el experimento ATLAS (LHC) fase 2
Investigador principal (nombre y apellidos): Jose del Peso
Entidad financiadora: MICINN
Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 1/1/2011-31/12/2013
Financiación recibida (en euros): 330.000 euros
Relación con el proyecto que se presenta: mismo tema
Estado del proyecto o contrato: concedido

3. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Jose del Peso, Claudia Glasman
Referencia del proyecto: FPA2007-66708-C03-03
Título: Tier-2 Distribuido Español para el experimento ATLAS
Investigador principal (nombre y apellidos): Jose del Peso
Entidad financiadora: MICINN
Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 1/1/2008-31/12/2010
Financiación recibida (en euros): 542.800 euros
Relación con el proyecto que se presenta: mismo tema
Estado del proyecto o contrato: concedido

4. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Claudia Glasman, Jose del Peso
Referencia del proyecto: FPA2011-26774
Título: Estudio de las colisiones protón-protón con el detector ATLAS en el LHC
Investigador principal (nombre y apellidos): Juan Terron
Entidad financiadora: MINECO
Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 1/1/2012 - 31/12/2014
Financiación recibida (en euros): 332.000 euros
Relación con el proyecto que se presenta: está muy relacionado
Estado del proyecto o contrato: concedido

5. Investigador del equipo de investigación que participa en el proyecto/contrato (nombre y apellidos): Claudia Glasman, Jose del Peso
Referencia del proyecto: FPA2008-00601
Título: La estructura del protón y la física en el LHC
Investigador principal (nombre y apellidos): Fernando Barreiro
Entidad financiadora: MICINN
Duración (fecha inicio - fecha fin, en formato DD/MM/AAAA): 1/1/2009 - 31/12/2011
Financiación recibida (en euros): 790.000 euros
Relación con el proyecto que se presenta: está muy relacionado
Estado del proyecto o contrato: concedido

Parte C: DOCUMENTO CIENTÍFICO. Máximo 25 páginas.

C.1. JUSTIFICACIÓN DE LA COORDINACIÓN

Both IFIC and UAM contribute to the storage and processing of data coming from the ATLAS experiment at CERN with substantial resources, both human and material. Nevertheless, current ATLAS job flow and data processing occurs in both sites independently adding an extra effort in all parties. We want to coordinate our work to converge through a true federation which will allow ATLAS to send data and processing jobs to a single end point and on the other hand, will force both projects to share knowledge in order to solve problems more efficiently as well as to increase the visibility within the ATLAS experiment. The participants have shown sufficient experience in computing for particle physics experiments to engage in this new challenge.

The coordination of the Project will be the responsibility of the IFIC centre.

Interaction between the objectives, activities and subprojects of each subproject

These common objective implies that both subprojects participate actively in all tasks. These can be summarized as: provision of infrastructure and services, operation, user support, contribution to the ATLAS computing tasks, usage of opportunistic resources and convergence to a true federation.

Coordination mechanisms:

There will be regular meetings to cover different aspects, namely:

- TIER-2 Operation Meetings, with a weekly frequency. Short meetings (30 minutes - 1 hour) by videoconference. Very focussed on operational status report and problem solving.
- Face to face meetings four times a year, one per quarter. One day meeting taking place in Valencia or Madrid. Taking advantage of the train high speed connection between both cities (AVE). We will meet to present the status report of the project for the previous quarter, discuss and agree directions for the next period and share knowledge and strategies. Cost will be kept low going and returning in the same day.
- Project Leader meetings which will be held on demand, having at least twice a year.

In addition, there are certain meetings of the ATLAS Collaboration which are perfectly linked to the coordinated project, as for example:

- Software and Computing weeks at CERN every 4 months
- ATLAS Computing Institute Collaboration Board (ICB) (In average there are 3 or 4 meetings per year).
- Many other short videoconference meetings, within the scope of the ATLAS computing activities, that regularly occurs two or three times a week (Not all the project members participate in all the meetings, though).

C.2. PROPUESTA CIENTÍFICA

C.2.1. Antecedentes y estado actual

The Large Hadron Collider (LHC) experiments, in particular ATLAS, use a network of computer centres to process the billions of events coming from the proton-proton collisions. This collection of centres are seen as an enormous distributed machine thanks to the usage a software layer called “GRID middleware”. The Worldwide LHC Computing Grid (WLCG), which have its headquarters at the CERN laboratory (Geneva), coordinates all the experiment’s computing resources and promote general policies and tools to be used among all participating centres.

The amount of data transferred to the GRID for processing is huge: dozens of PB (1 PB = 1000 Tbytes) a year, which is the most challenging data processing of the history of science. There are almost a hundred computer centres on the ATLAS GRID distributed all over the world. These centres are classified in two categories at present, according to their services: Tier-1 (keeping early stages data) and Tier-2 (mainly for simulations and storage of final stages data) centres.

The second period of data taking (Run2), started in June 2015, brought a large increment in the data volume due to an increase on both beam luminosity and an energy. Software has changed to adapt the data processing to the new requirements. In particular, jobs evolved from sequential to parallel execution to take profit of new multi-core CPUs while maintaining memory footprint. These modifications imply a change in the local batch system configuration to allow for both type of jobs running simultaneously (single and multi-core) to allow efficient resource usage.

The ATLAS software was optimized to run faster and cope with the increase in the number of events and data volume with respect to the previous run (Run1). Also a new production system (Prodsys2) and a new data management system (Rucio) were developed for Run2, which implied some activities at the sites, not only at the level of computing services but also in the user support of local physicists. New data indexing at event level, the Event Index, was developed and put into production. Development is still ongoing to increase its performance and overcome detected limitations on current implementation. Single event dispatching and processing using the Event Service and the object store to collect the output opens the possibility to use opportunistic resources (for example Supercomputers) more efficiently reducing dependencies on local storage allowing unexpected job termination due to cpu reclaim without losing large processing times. A new analysis model was put in place for Run2 using a new data format (xAOD) readable through both ATLAS Athena software and the general analysis tool ROOT.

The Run2 conditions will remain stable for the period of this FPA project request. However, new developments have already started to adapt the ATLAS computing to the next big challenge (Run3) where there will be another increase in luminosity hence, in the number of events for processing.

Monte Carlo production is still the dominant activity in terms of CPU consumption in 2015-2016, as foreseen [2], many billions of events have been produced in a year in the IFIC + UAM sites. In addition, it is expected that these Tier-2 sites keep two copies of reconstructed data for certain processes and trigger selections.

This project is a continuation of the present TIER-2 project FPA2013-47424-C3.

Nationals and International groups that work in the same subject:

Nationals: Group of High Energies of the University of Santiago de Compostela (US), Group of High Energies of the Universitat of Barcelona (UB), Group of High Energies of the

CIEMAT of Madrid, Group of High Energies of the IFCA of Santander, PIC of Barcelona, IFAE of Barcelona.

International: FZK of Germany, RAL of United Kingdom, Italian groups of the INFN (Milan, It is above, Rome, Bologna, etc.), Computing Centre of the IN2P3 of France, NIKHEF of Holland, etc.

Bibliografía más relevante

[1] Report on Project Status, Resources and Financial Plan, CERN-RRB-2015-087, October 2015
[2] ATLAS Computing Status and Plans Report to the C-RSG, April 2016.
[3] V. Sánchez et al, 2013, Lessons learned from the ATLAS performance studies of the Iberian Cloud for the first LHC running period. International Conference on Computing in High Energy and Nuclear Physics (CHEP13). Journal of Physics: Conference Series 513 (2013) 032082.
[4] S. González de la Hoz et al, 2015, Spanish ATLAS Tier-2 facing up to Run-2 period of LHC. International Conference on Computing in High Energy and Nuclear Physics (CHEP15) Journal of Physics: Conference Series 664 (2015)
[5] 'Market, Technology and Costs Trends, 2012, CERN T0 and CAF Computing Facility'. Bern Panzer-Steindel, CERN-IT,may 2012
[6] Update of the Computing Models of the WLCG and the LHC Experiments, LCG-TDR-002, WLCG-ComputingModel-LS1-001, April 2014

C.2.2. Hipótesis de partida y Objetivos Generales (Reto cuyo estudio se pretende abordar)

IFIC and UAM Tier-2 sites have been receiving, storing, processing, producing and sending data for the ATLAS experiment contributing to the worldwide global data processing for more than 10 years. Both sites have performed so reliable that they are included in a privileged category called T2D, which are Tier-2 centres suitable to receive critical data and are candidates for the ATLAS Core of computer centres. Presently the resources of the two centres are: 21225 HEP-SPEC06 and 2700 Tbytes (66.7% IFIC, 33.3% UAM). This represents the 75% of all ATLAS Tier-2 resources in Spain and the 3.75 % of all ATLAS Tier-2 worldwide.

The basic goal of this project is to keep contributing to the ATLAS GRID with data storage and processing capabilities according to the global needs of the experiment which is steadily increasing every year as new data is coming, more montecarlo is being generated and new analyses from physicists are using more and more data. To cope with the new experiment needs for the next 3 years, IFIC and UAM main goal is to increase their participation to the 5% of the worldwide Tier-2 new additional resources needed by the ATLAS experiment. Both sites are capable of managing the new resources and operate them efficiently. IFIC and UAM commitment level was proved in the past years providing even more cpu that required for some periods, and increasing their participation minimises the risks of inefficient usage and operation if new additional resources are spread over more sites.

Correct maintenance and operation of present and new resources in each site is a must, since the success of the whole ATLAS scientific program is tied to the local and global data processing capability. The objective to converge towards a single virtual site with a single entry point for ATLAS will increase the overall efficiency and will reduce the maintenance effort for both sites and ATLAS.

The project wants also to contribute to the general ATLAS computing operation tasks requested by the experiment for all the participating institutes.

These objectives implies to deal with very advanced computing technologies and work within a large collaboration of computer centres interconnected through the LHC computing GRID network. The project involves the continuity and consolidation of a scientific facility which has providing a large impact in particle physics: with no such facilities the Higgs boson discovery, for example, would not have been possible. The challenge to deal with such unparalleled huge amount of data and the tools being developed to store and process them will led to innovations useful to the information society. Emphasis has been given to e-infrastructures both in R&D Spanish and UE policies as an essential tool for scientific and technological research. This project also contributes to push up further the performance of the National Research and Academic Networking, which will benefit other scientific areas.

C.2.3. Objetivos específicos de cada uno de los subproyectos participantes.

1. Provide a Tier-2 infrastructure for data analysis and Monte Carlo simulation production for the ATLAS Experiment, offering the mandatory contribution to the Atlas computing per site as reported in previous section. The exact requirements are released in a year-to-year basis and they have to be approved by a supervisor committee named Resources Review Board (also named RRB, http://cds.cern.ch/record/2054031/files/CERN-RRB-2015-087.pdf). We expect that the Spanish Atlas Tier-2 will have to provide the resources shown in table 1 during the period 2017-2020.

	2017	2018	2019	2020
CPU (HEP-SPEC06)	42300	50800	61000	73200
Disk (Tbytes)	3900	4500	5200	6000

Table 1: Resources to be provided by the Spanish ATLAS Tier2 in the period 2017-2020.

CPU and disk requirements for the period 2017-2018 were taken from the most recent ATLAS collaboration report. A 5 % was applied to get the Spanish contribution. Since there is not an official request from the experiment for period 2019-2020 we have estimated the resources needed as an steady yearly increase of 20% for CPU and 15% for disk, after consultation with the ATLAS computing management board. Renewal of equipment older than 4 years was also taken into account in the tables.

	2017	2018	2019	2020
CPU (HS06)	42300	50800	61000	73200
ΔCPU (HS06)	14000	8500	10200	12200
Renew CPU (HS06)	6250	8900	8300	4700
TOTAL CPU requested (HS06)	20250	17400	18500	16900
DISK (Tbytes)	3900	4500	5200	6000
ΔDISK (Tbytes)	300	600	700	800
Renew Disk (Tbytes)	470	410	1010	1650
Total Disk requested (Tbytes)	770	1010	1710	2450

Table 2: Profile of the computational and storage resources for the period 2017-2019.

It is a specific objective to provide and operate the following new resources in addition to those that are already in production at each site:

	2017	2018	2019	2020
IFIC CPU requested (HS06)	13500	11600	12333	11267
UAM CPU requested (HS06)	6750	5800	6167	5633
IFIC Disk requested (Tbytes)	513	673	1140	1633
UAM Disk requested (Tbytes)	257	337	570	817

Table 3: Profile of the computational and storage resources share IFIC-UAM for the period 2017-2020.

In this document we use a cost model agreed by all LCG projects in Spain. The final values are shown in table 4 which are used to estimate the total infrastructure funding cost.

	2017	2018	2019	2020
CPU €/HS06	6.1	5.0	4.2	3.5
DISK €/Tbyte	178.9	151.4	128.6	111.9

Table 4: Raw prices for CPU and disk including taxes (VAT) [5,6]

2. Provide user support for the Spanish physicists at each site, members of the Atlas collaboration concerning both the use of the grid infrastructure and the use of the Atlas software according to the Atlas Computing Model. Involvement in Data analysis/processing to test new developments in a realistic environment will improve the User Support.

3. Participate in different common activities defined by the Computing ATLAS Collaboration during the duration of the project. We want to participate on Regional TIER-2 Operations, Distributed Analysis Shifter Team (DAST), ATLAS Distributed Operations shifts (ADCOS), Distributed Analysis, Grid Data Processing (GDP), Cloud computing and Networking, and the development and deployment of an Event Index.

4. Contribute to the ATLAS detector manpower and operation budget. The Management and Operation (M&O) Budget absorbs the daily operating costs of the ATLAS experiment. The 2018 and 2019 ATLAS M&O Category-A and Category-B for Spain is requested in this project. The dominant part of the cost in Category-A is providing the required ATLAS technical services (eg. detectors, gas system, heavy handling, cooling, etc). The main costs of Category-B are related the detector modules electronics and controls, maintenance of detector structures and technical manpower to run the facilities. Taking into account the payment commitments and the administrative constraints spelled out by the Ministry, the Spanish ATLAS Community in agreement with the Spanish HEP Coordinator have agreed to request this budget in this ATLAS Computing Project.

C.2.4. Metodología

The main tasks to be performed by this project are:

1. Project Management

The project management coordinators define the objectives, organize the tasks and strategies and decide about policy and execution plans. They will be the responsible to follow up all the activities and elaborate all necessary reports related with the project.

The coordinators (project management board) will represent the Tier-2 in different committees and forums (except the ATLAS International Computing Board designated by the Contact Physicist) and establish the coordination links with the related tiered computing centres and the LCG project.

2. Federation operation and infrastructure support.

A TIER-2 site is a Computer Centre with very demanding requirements. According to the prerequisite established in the Computer Model of the ATLAS Experiment a regular TIER-2 site must run round-the-clock every day of the year and must be attended to by personnel 12

hours a day during working days (12h/5d). The personnel must have enough expertise to solve problems and guarantee the proper operation of the center. The requirements are even more stringent for the so called TIER-2D. These are TIER-2 sites which may receive data from any other ATLAS Grid centre in contrast with a regular TIER-2 which receives data from centres that belong to the same Cloud. A TIER-2D site must be stable enough to guarantee the process of the data that houses. A TIER-2D site must be properly operational for more than 90% of the time. There are regular monthly reviews about the availability of the centres. A site failing more than 3 days in a month is automatically degraded. This implies that a TIER-2D centre must be watched continuously, every day of the year.

Some of the tasks designated to the technical personnel to operate and support the TIER-2 infrastructure include hardware installation and maintenance, installation and configuration of the operating system releases, Grid middleware updates, monitoring the center, operation of the distributed storage system, user administration, etc.

This commitment requires 2 FTEs at each site.

The Operation Manager (OM) will be the responsible of the global performance of the federated TIER-2 and will interact with the personnel of the sites in order to establish the technical coordination to ensure the proper operation of the TIER-2 and the well-functioning of the link with the TIER-1. The OM has been assigned to IFIC and one FTE for this task is needed. The OM is the convener of the TIER-2 Operation Meeting.

3. User support and Data Analysis

One of the goals of this project is to support end users, that is, the local physicists doing ATLAS data analysis in our institutes. An analysis usually has two steps: i) process ATLAS data on the GRID and produce as output a set of n-tuples with selected events, ii) process these n-tuples on a local computing infrastructure. For the latter, end users own a Tier-3 facility of resources. The installation, configuration, operation and management of this facility is done by the personnel of the Tier-2, as part of the user support. The Tier-3 facility must be very reliable and have a low latency in order that users get physics results in time. Since this facility has a smaller size than the Tier-2 centre, its operation can be done by the Tier-2 staff as an economy of scale.

Another way to support local users is by solving problems they may have when doing data analysis. The ATLAS Collaboration has a service to answer questions about analysis jobs, which is called “DAST”. However, this service is overloaded, since it is provided by a few experts and there are several thousands of potential users. The local user support is due to solve many local user problems, which otherwise would scale to the DAST service. Hence, the local user support is a strategy to avoid saturation of the DAST. On the other hand, there are issues which are not covered by the DAST as, for example, those related to certificates or to running jobs on a local Tier-3. In addition, for some type of problems a personal meeting is in order, which is only feasible from the local support team. As a coordinated project, the Spanish Tier-2 centres have developed a “how-to” documentation for users, which can be accessed on the web. In addition to this global initiative, at least one person in each centre is responsible for given support to their local users.

Some of the members of the research team have manifested their interest to perform data analysis. This can be very useful for user support, since many questions of local users are related to ATLAS software for analysis or to errors on analysis programmes. In addition, those who become experts on data analysis procedures will be candidates to join the ATLAS DAST team which is overloaded at present, contributing in this way to provide a more efficient service to the ATLAS Collaboration. Besides, the expertise in computing and physics may lead the research team to use the latest ATLAS tools for analysis and pass their experience and recommendations to other end users. On the other hand, this kind of activities yields to physics results for the experiment and to PhD thesis for students; hence, enhancing the research and educational capabilities of the project. Data analysis on two

physics topics are foreseen, one per institute, namely: prompt photon cross sections (UAM) and top quark production (IFIC).

Prompt photon cross sections (UAM).

The analysis goal is the measurement of the inclusive prompt photon differential cross section as a function of the photon transverse energy in proton-proton collisions at 13 TeV using the full integrated luminosity collected by the end of 2016. It is expected that by the end of 2016 the integrated luminosity will be approximately 28/fb, which is a factor of 9 compared to that collected in 2015. This increase in luminosity will allow a significant reduction of the statistical uncertainty in the high photon transverse energy region, which is currently a major limiting factor, and to extend the measurement up to 2 TeV. Furthermore, improvements are expected in the performance of the photon reconstruction, identification and calibration that will also reduce significantly the systematic uncertainties and, thereby, will allow a more precise comparison to the predictions of the Standard Model (SM). The high photon transverse energy region is especially important since it is the one most sensitive to possible deviations with respect to the SM due to new phenomena as well as to electroweak effects not accounted for in pure fixed-order QCD calculations. The measurements will be compared to the predictions of state-of-the-art generators as well as to calculations at fixed-order in perturbative QCD such as those provided by the program JetProx. Furthermore, to gauge the importance of electroweak effects, comparisons will be made to calculations which include resummation of QCD threshold logarithms and electroweak Sudakov logarithms such as those provided by the program PeTeR. One of the senior members of the group, Claudia Glasman, has a long experience in analyses with photons using the ATLAS detector and is currently contributing to the measurement of the inclusive prompt photon cross section with 3.2/fb.

Top quark production (IFIC).

The Standard Model of particle physics (SM) provides an excellent description of the interaction between the fundamental constituents of matter. The LHC experiments have achieved the discovery of the Higgs (SM-like) particle in 2012. However, several good reasons exist to search for signals of physics that goes beyond the known set of particles. The ATLAS group at IFIC has performed several searches and precision measurements on the LHC data. Most relevant for this proposal are the analysis of top physics in events collected by the ATLAS experiment in Run 1. In particular, we would like to emphasize the scrutiny of highly boosted top quark production, both in t-tbar resonance searches and in the first charge asymmetry measurements in this regime taking profit of the new data coming from the Run 2.

These studies will be done considering several benchmarks coming from different BSM (Beyond the Standard Model) alternatives. On top of that, we can focus on physics channels of our interest. This research line needs a preliminary work in more technical aspects as the delivery of a top tagger, the study of Scale Factors, systematic uncertainties, etc.

The supervisors of both groups (UAM and IFIC) for the physics topics mentioned previously will be in the best position to help PhD students become familiar with the new analysis model in use for Run 2 as well as in-depth know-how specific to the proposed analysis.

4. ATLAS Core Computing Participation

According to the ATLAS policy every funding agency must fulfill a minimum of common tasks to operate the detector, which include the computing. A natural way to cover these activities is through the people of the TIER-2 project. These activities comprise computing shifts, software support, production validation, management of the condition database, among others. The TIER-2 members have been contributing to these type of tasks, covering the percentage of participation required by the ATLAS Collaboration.

4.1 Development, deployment, maintenance and monitoring of an EventIndex (EI) for experiments collecting large amounts of data

The ATLAS experiment had an "EventTag" database that was designed before the start of data taking in 2009 and includes various physical quantities for each event, as well as references to files that contain the event, but was stored in Oracle databases with separate tables for each cycle of reprocessing. The Tag system is potentially very useful, but was very little used, at least in its database format.

It was possible to design a more agile system that keeps the functionality of skimming by keeping only the pointers (currently GUIDs) to the files where the event in question can be found at every stage of processing and eliminating other variables of lower importance.

This work, aimed at creating an "EventIndex" to catalogue all the collected events, started from the analysis of use cases and the actual experience of the last years (since the LHC started taking data). This work examined how best to store the necessary information using modern technologies of data storage ("NoSQL" databases) that allow to store key-value pairs in memory and selecting the best Tools to support this application from the point of view performance, robustness and ease of use. During the project various computing tools were tested that are currently in use by large international companies (Google, Yahoo!, Amazon, etc..) in order to select the system that results the most stable and scalable for this application. The tests took place on a large scale using the existing data, and those collected in 2012, by the ATLAS experiment.

Using NoSQL technologies it is possible to store information for each event in a single record. The record is created upon recording of the event from the online system, with initially only limited information on the event immutable characteristics, such as event number, run number, time stamp, luminosity block number, which trigger selected the event, and the identifier of the file that contains the event in RAW format. Two main applications has been developed simultaneously to the database: the infrastructure to populate the database and the search for the records and extracting the identifiers of the files that contain the events of interest. IFIC people have worked and will work in the development of the infrastructure to populate the database. IFIC group is contributing with 1.5 FTE to the project. The ATLAS EventIndex has been running in production since mid-2015, reliably collecting information worldwide about all produced events and storing them in a central Hadoop infrastructure at CERN. A subset of this information is copied to an Oracle relational database for fast datasets discovery, event-picking, cross-checks with other ATLAS systems and checks for event duplication.

The system design and its optimization is serving event picking from requests of a few events up to scales of tens of thousands of events, and in addition, data consistency checks are performed for large production campaigns. Detecting duplicate events with a scope of physics collections has recently arisen as an important use case.

Next developments are in the way to improve the throughput of the overall system. IFIC team is designing and testing a new architecture approach for the Distributed Data Collection of all needed data, to overcome the limitations of the current implementation. It has been seen that in some situations, the messaging system architecture that we are currently using is not able to cope with some peaks in the production rates. While this is not a problem now because the produced backlog is consumed in small periods, it might suppose a limitation with high sustained rates in the future. We are testing a new transient collection and storage system based on Object Stores, to convey the information of EI distributed produced data, to the final storage at Hadoop and Oracle database.

At the same time we are adapting the data transfer protocol specification to be optimized for the new approach and the use cases that are most used, simplifying the current implementation. In parallel we will explore new storage options for the Hadoop backend that could improve the performance in the ETL (Extract, Transform and Load) procedures, but also to the overall response of the system for final users.

The EventIndex initiative has been a good opportunity to involve our computing specialists in an interesting R&D project which is being very useful to Physics Analysis procedures. The first prototype was already tested and is now in production, whereas the tool needs to be monitored and followed from the operational point of view. The current and future efforts will be devoted to increase the performance, and the level of automation of the overall system.

4.2 Distributed Analysis Shifter Team (DAST)

The Distributed Analysis Support Team (DAST, formed in 2008) is a team of experts and shifters who provide first direct support for all help requests on distributed data analysis by monitoring a mailing list called the distributed analysis help forum. There are two shifters on duty during working hours; one in the North American time zone and one in the European time zone, covering 15 hours a day on weekdays. The DAST traffic, from 1,320 users, was exchanged 90,478 emails since Oct. 2008, more than 10,000 a year. Common problems include authentication difficulties, finding interesting data, unstable or misconfigured computing sites, complexity of the software being used inside the Grid environment (ATLAS standard code - Athena and user analysis code) and retrieving job outputs. Since 2012 Farida Fassi at Mohammed V University in Rabat, and linked membership at IFIC co-coordinates the overall activity of the team.

4.3 Distributed Analysis

ATLAS Distributed data Analysis (DA) is based on a complex and evolving system with many components interacting with each other. Users can submit their analysis jobs to the system using the friendly Front-End tools. The ATLAS software Athena is available at all ATLAS Grid sites through a web-based distributed file system. PanDA handles the matchmaking of analysis jobs to computing resources. Jobs are executed on the Grids, which provide the computing resources. ATLAS jobs on distributed Grid resources are typically classified into Production and Analysis activities. Production jobs consist of processing the raw data from the ATLAS detector to produce the reconstructed data for physics analysis as well as Monte Carlo (MC) generation, simulation, and reconstruction. Then, individual physicists submit tasks to analyse the centrally produced data. The tasks are routed to sites based on the availability of relevant data and processing resources. The data distribution is optimized to fit the resource distribution, and it is dynamically changed to meet rapidly the evolving requirements for analysis use cases. Distributed analysis tools used to analyze the data are reliable and fast to work with. The Distributed Data Management (DDM) system consists of a set of resilient services for data location, data movement and data deletion. Users directly interact with the DDM system too, when querying the data locations or uploading/downloading datasets. PanDA manages centralized Monte Carlo (MC) production; simulation and user analysis.

The involved people in the Tier-2 project aim to work on the testing, validating and deployment of DA system at the Tiers-2 infrastructure, using a common analysis framework and considering a real analysis example. In addition, the local Tier-3 provides an interactive analysis mode based on the grid tools. This functionality would be the best proof of the good throughput and performance of the ATLAS Computing system, including Tier-2 and Tier-3 facilities. Moreover, this real analysis example would help to ensure that we have the required Grid tools components deployed. Run-2 presents new opportunities and challenges. Based on the learned experience during Run-1 new developments for a more dynamic system are considered necessary, and also to sustain the load in the coming years with new technologies and concepts. We aim to be strongly involved in these activities.

4.4 Cloud Squad Support

Previous years' experience running the distributed TIER-2 has shown the necessity of monitoring at the cloud level both job execution and data transfers. A shift crew can be organized to detect problems as soon as they appear and to take action to solve them. The sources of failures could appear at different levels, the TIER-2, the TIER-1 or the ATLAS system. The crew must identify the source and contact the proper experts in case it does not know how to solve the problem. It should also make computer programs to facilitate the task of both monitoring and action taken. This activity complements the ATLAS operation shifts and increase the quality of the TIER-2 service. All technical tasks of the coordinated Tier-2 belong to this category as well. It helps also sites on ATLAS-specific issues and plays an interface role between sites and the ATLAS central operation. The Iberian Tier-2s are composed by the three Spanish centres (IFAE, IFIC, UAM), two Portuguese centres

(Coimbra and Lisbon) and a centre in Chile (UTSM). The coordination has been done by a person from Coimbra in the past years.

4.5 Experiment Computing Shifts

Common computing operation represents about 20%-30% of the whole ATLAS experiment operation. Keep the distributed computing machinery working is a responsibility of all participants. There are two shift categories: Class 1 shifts and Class 2 shifts. The firsts have the requirement of being 'in person' at CERN to carry them out. This is not needed for the Class 2 shifts.

In period of Run 1 (2010-2012), Tier-2 IFIC team involved approximately 1.5 FTE for the LS1 and Run 2 the Tier-2 team involved approximately 0.5 FTE. Part of the contribution in this second period has moved to another responsibilities related to Event Index activities (Class 3 in ATLAS OTP). IFIC team is going to continue participating in this ATLAS Computing shifts for next years

5. Opportunistic use of resources

To compare the detector measurements with theoretical predictions and extract conclusions about the nature of particle physics, ATLAS must produce simulated events. The simulation of physics events works in three steps:

- Event Generator. Events of certain physics processes are generated and stored in files. This step has low CPU consumption.
- Detector Simulation. Events from event generator files are read and passed through a detector simulation to obtain the detector response. These events are stored in files. The CPU time needed for this step is very large.
- Reconstruction. Events from the previous step of detector simulation are used to create physical objects such as electrons, muons, etc. This step has low CPU consumption.

For the second step, the GEANT computer programme is used, which is very costly in terms of CPU time. This is due to the complex geometry and fine granularity of the ATLAS detector as well as the good precision required for these simulations: physics results should be dominated by statistical errors on the measurements and never by systematic uncertainties from the simulation.

The pledged CPU power requested by ATLAS is not enough during some periods of special simulation campaigns for precision studies. One possible way to satisfy these needs is the use of extra resources in an opportunistic way. Recently, the ATLAS Collaboration started to develop software tools to go in this direction. Instead of processing files of events, as it is the case in the standard production procedure, an Event Server is in place to process isolated events. Once an event is processed on the extra resources, it is transferred to an Object Store located on the Tier-2 site (or other pledged centre) and next event enters the extra resources for processing. In the end, some particular jobs merge the events from the Object Store to produce the final files of events which are distributed to the GRID. This procedure is not intrusive to the extra resources and is very robust to failures: only one event can be lost in case of problems.

Work will be done to implement and test this procedure on a realistic environment. Initially, it is expected to use the huge amount of UAM computers utilised to teach students in practical lectures. In a later stage, some resources of a UAM computer centre will be used: those arranged for researchers. The task implies a collaboration with the UAM-IT Department, ATLAS computing experts on the topic and, of course, the personnel of UAM and IFIC sites.

6. Site convergence towards a true federation.

The IFIC and UAM sites currently provides loosely coupled resources to the ATLAS GRID infrastructure and both are seen as a different sites with separate access points (IFIC-LCG2, UAM-LCG2). The aim of this task is to converge towards a single virtual site with a single entry point for ATLAS to be more efficient and reduce the maintenance effort.

In this task, we will define the requirements that must fulfil this virtual site and we will build a testbed to set up possible solutions to provide the necessary services.

Both IFIC and UAM will unify the management of their infrastructures adopting common agreed software tools and procedures. After carefully evaluation of different alternatives to provide services, both sites will implement the same solutions for storage, computing, fabric management and monitoring. These will be tested in the testbed and, when possible, will be implemented in current production infrastructure without disrupting current operation.

The process will result in a true and more efficient federation by the end of the project, ready to face the RUN-3 of the ATLAS experiment.

Work will proceed in different areas:

6.1 Local and federated storage

Storage services are provided by IFIC using Lustre + StoRM and by UAM using d-Cache. Both alternatives are still valid and provide good performance. Work will be done to converge to a single solution, after careful reevaluation of the present systems, with a unique entry point for data.

6.2 Local cpu workload resource management.

CPU workload is managed by TORQUE/MAUI in both sites. Some limitations have been detected and there is a common consensus in the WLCG that another system has to be used. One candidate is HTCondor that will be evaluated by both IFIC and UAM. Changes in the local cpu resource management will imply the setup of a different Computing Element solution. We will define a common set of requirements for both sites and a single procedure for installation, configuration and maintenance.

More work will be needed to define and test how cpu workload can be routed between both sites to reach a single entity.

6.3 Service aggregation

IFIC and UAM present two end points for data transfer and job management. Work has to be done to allow federation of both storages and proper job routing between sites to provide a single end-point. The Information System has also to deal with it to present a unified view of the available resources and their status.

6.4 Networking

Present network configuration provided good connectivity and transfer rates for Run1 and the first period of Run2. Both sites have 10 Gbps connection to their respective campus network. Data transfer uses general R+D network (regional and national) which started to show some bottlenecks under heavy load as our traffic has to compete with the general one.

We will work to get IFIC and UAM inside LHCONE infrastructure. This is a requirement to cope with the needs of latest Run2 period and Run3, and it will be absolutely required to build a true federated site since both sites can only be seen as one if transparent access to data from both sites is provided and fast data movement between sites is possible. Fast (and low latency) data access from jobs to both sites has to be offered irrespective of the place where they are running.

IPv6 will play an important role in the future. Work will be done to provide IPv6 stack on all services.

6.5 Installation and configuration systems.

Both IFIC and UAM use similar systems to install and configure the machines based on PXE, Anaconda, quattor and puppet. Work will be done to share experience and procedures to get similar configuration profiles based on common templates. Documentation and guidelines will be developed

6.6 Virtualization

Both sites currently use visualization to provide some services, but we do not have a specific virtualization platform such a OpenStack. We will select and implement a proper solution to define, start, configure and operate virtual machines on demand. This will allow us, eventually, to provide in the future an IaaS service. Work will be done at IFIC and UAM to reach this objective.

6.7 Monitoring tools

Monitoring is an essential activity to get a good level of site availability and reliability. Work will be done to unify monitoring systems and to build a single monitoring point for both sites using tools like Nagios.

6.8 Common methodologies and procedures. Documentation elaboration

We will develop common methodologies and procedures for all services installation, maintenance and operation and we will create proper documentation. This will allow managers from both sites to share their experience, have a common knowledge and eventually operate both sites.

6.9 Federation portal

A web portal will be developed to collect all the federation information which will at least contain:

- Federation general information.
- Unified federation status, both, per site and aggregated.
- Detailed monitoring information
- Historic views.
- Documentation (procedures, scripts, configuration templates...)

Information will be private or public depending on their sensitivity.

C.2.5. Medios materiales, infraestructuras y equipamientos singulares

One of the goals of this project is to become an ATLAS Tier-2 virtual centre composed by two sites located at IFIC and UAM. The resources available at IFIC at present are 1800 Tbytes of disk storage and 14150 HEP-SPEC06 of CPU power, whereas the resources available at UAM are 900 Tbytes and 7075 HEP-SPEC06. Both sites are well-functioning computer centres set up as ATLAS Tier-2 sites in 2005; hence, with more than 10 years' experience. They both have UPS protected power lines and efficient air conditioning systems. Their network bandwidth is 10 Gpbs, enough to provide a good service for the ATLAS data transfers. However, the network link is in the academic backbone; hence, shared by many universities and institutes. Even though this link satisfy the ATLAS requirements for data transfers, we have as another goal to set up exclusive links to face up RUN 3 data taking properly. There are some differences though in the storage and configuration systems between both sites. The IFIC site uses Lustre+Storm for their storage and Puppet+related-tools for installation and configuration of servers and services. In contrast, the UAM site uses dCache for storage and Quattor for installation and configuration of servers and services. All these systems work perfectly, however we would like to converge to a common solution to simplify the operation of an unique virtual centre.

C.2.6. Cronograma

O1: Project Management
Responsible: Santiago González de la Hoz (IFIC)
Participants:José Salt (Co-IP, IFIC), Jose del Peso (UAM)
Periodo de ejecución (en trimestres); T1 - T12 (Activity to be done in a continuous way)
H1: Face to face meetings to decide about policy and execution plans, T1,T2,T3,T4
H2: Face to face meetings to decide about policy and execution plans, T5,T6,T7,T8
H3: Face to face meetings to decide about policy and execution plans, T9,T10,T11,T12
E1: Mid project report, T6
E2: Final project report, T12
E3: Report the International Computing Collaboration ATLAS Board (ICB), T4
E4: Report the International Computing Collaboration ATLAS Board (ICB), T8

O2: Federation operation and infrastructure support
Responsible: Jose Salt (IFIC)
Participants: Javier Sánchez (IFIC), Alvaro fernandez (IFIC), Julio Lozano (IFIC), Carlos García (IFIC), Contratado IFIC1, Contratado IFIC2, Jose del Peso (UAM), Contratado UAM1, Contratado UAM2
Periodo de ejecución (en trimestres); T1 - T12 (Activity to be done in a continuous way)
H1 - Pledge update every year, T1
H2 - Pledge update every year, T5
H3 - Pledge update every year, T9
H4 - Weekly Federation Operation meetings, T1-T12
H5 - Baseline Software versions check and update, T3, T7, T11
H6 - Service status review. Installation and deletion, T3, T7, T11
E1 - Spreadsheet report to WLCG project, T4
E2 - Spreadsheet report to WLCG project, T8
E3 - Spreadsheet report to WLCC project, T12
E4 - Availability, reliability report, T4, T8, T12
E5 - Accounting report, T4, T8, T12

O3: User support and data analysis
Responsible: Claudia Glasman (UAM)
Participants: Santiago González (IFIC), José Salt (IFIC), Farida Fassi (U. Mohammed V)
Periodo de ejecución (en trimestres); T1 - T12 (Activity to be done in a continuous way)
H1: Computing topics for qualification of new students, T1-T4
H2: Analysis of data, T5-T10
H3: Keep user support documentation up-to-date, T1-T12
H4: Solve problems of local users, T1-T12
H5: Operation of Tier-3 system, T1-T12
E1: Report of results, T11-T12
E2: PhD thesis, T12
E3: Physics publication, T12

O4: ATLAS Core Computing Participation

Responsible: Álvaro Fernández (IFIC)

Participants: Santiago González (IFIC), José Salt (IFIC), Javier Sánchez (IFIC), Julio Lozano (IFIC), Farida Fassi (U. Mohammed V), Carlos García (IFIC), José del Peso (UAM), Claudia Glassman (UAM)

Periodo de ejecución (en trimestres); T1 - T12 (Activity to be done in a continuous way)

H1: Implement distributed new Event Index data collection prototype based on Object Storage and new protocol for data transfer. T4
H2: Evaluate new Event Index prototype with production environment rates. T5
H3: studies on new Hadoop storage format approaches for the Event Index. T5
H4: participation in the testing, validation and deployment of the Distributed Analysis system. T4, T8, T12
H5: participation in the experiment computing shifts T1-T12
H6: participation in the distributed analysis support team T1-T12
H7: participation in the cloud squad T1-T12
H8: Event Index system operation T1-T12
E1: New Event Index prototype in production using the new data transfer protocol and the Object Storage. T6
E2: New Event Index prototype in production using the new Hadoop storage format. T8
E3: New system for the Event Index application monitoring T6

O5: Opportunistic use of resources

Responsible: Jose del Peso (UAM)

Participants: Javier Sánchez (IFIC), Julio Lozano (IFIC), Carlos García (IFIC), Contratado IFIC2, Contratado UAM2

Periodo de ejecución (en trimestres); T1 - T12 (Activity to be done in a continuous way)

H1: Setup UAM PCs to be useful for ATLAS simulation processing, T1-T3

H2: Setup and Event Server resource and install an Object Store, T4-T5

H3: Test the system thoroughly, T6-T7

H4: Set up the general use UAM Computer Centre to process ATLAS simulations, T8-T9

H5: Test the system, T10

H6: Transfer setup to another centre and test it, T11-T12

E1: Extra resources for ATLAS production jobs, T12

O6: Site convergence towards a true federation

Responsable: Javier Sánchez (IFIC)

Participantes: Álvaro Fernández (IFIC), Julio Lozano (IFIC), Carlos García (IFIC), Contratado IFIC1, José del Peso (UAM), Contratado UAM1

Periodo de ejecución (en trimestres); T1 - T12 (Activity to be done in a continuous way)

H1 - evaluation of current storage systems T4

H2 - evaluation of CPU workload systems T4

H3 - service aggregation strategy path T6

H4 - local network evaluation for LHCONE participation T6

H5 - evaluation of installation and configuration systems and monitoring tools T8

H6 - minimum working virtualization infrastructure T8

E1 - Web Portal T3

E2 - working testbed on each site to be used in aggregation tests T5

E3 - Web portal update T7
E4 - LHCONE connectivity T9
E5 - Web portal update T11
E6 - working true federation testbed T12

C.2.7. Contratación de personal

At least two full time computing engineers are required for the proper system management and operation per site (IFIC+UAM). These persons are needed to keep the correct service level requested by ATLAS which has to cover at least 12 hours a day, 5 days a week including holidays. It is not possible to deliver the expected performance with less people. Two full time employees (FTE) are needed at UAM and three at IFIC. These people will be involved in the following tasks: site and federation operation, infrastructure support including deployment of new hardware and software, convergence towards a true federation and opportunistic use of resources and participation in the ATLAS core computing activities.

The goals presented in previous sections require new R&D activities and software development. IFIC needs to hire an additional person to help in the writing of the Event Index software. The ATLAS Event Index became an important tool for the experiment and we play a key role in its development, deployment and operation. This person will help to ensure our participation and the success in this task.

C.3. IMPACTO ESPERADO DE LOS RESULTADOS

Thanks to this project's contribution to the LHC GRID computing, high energy physicists from Spain and their colleagues around the world will be able to store, process and analyze billion events produced by the LHC proton-proton collider at CERN registered by the ATLAS experiment.

IFIC and UAM have been participating actively during the last 10 years in one of the biggest distributed computing scientific infrastructure around the world composed by more than half a million CPU cores, 300 Pbytes (1 Pbyte = 1000000 Gbytes) of disk storage in around 150 centres.

This project addresses the challenge of the efficient data moving and placement using the internet and the correct management of hundred millions jobs running simultaneously in remote locations launched by thousands of properly authorized scientists which are authenticated into a secured computing environment. Spanish science and technology would benefit from the development and implementation of new solutions to solve this truly challenge that would be useful to the society, not only increasing their fundamental knowledge but also in searching practical solutions to the scientific BIG DATA processing problem.

This infrastructure will help producing scientific results such as, deeply knowledge of the Higgs Boson characteristics which would allow to understand the mechanisms why the particles acquire their mass, searching for candidates to explain the dark matter to clarifying the universe composition and the precise determination of the quark top properties that would allow testing new theories in particle physics like the supersymmetry, among others.

Our dissemination plan to propagate the results will include:

- Publication of results in specialized journals
- Presentations in conferences
- Participation in “Science Weeks” for general public as well as guided tours at the centres
- Creating specific web pages with information related to the project for both, scientific people and general public.
- Joining any other outreach activity that might arise in the University or laboratory environment related to this area

C.4. CAPACIDAD FORMATIVA DEL EQUIPO SOLICITANTE

IFIC: The training capacity in the period 2000-2016 can be classified in two important sections:

1) Training researchers on High Energy Physics and/or GRID Computing Technologies by means of Master Thesis, Ph. D. Thesis, Research Works. We give the following selection of people trained by our group:

- Luis March Ruiz, who started in May 2003 in the LCG-ES project as project fellowship and he presented his PhD thesis at the Department of Atomic, Molecular and Nuclear Physics (FAMN) of the Universitat de Valencia. After 2 postdocs, one in UAM to work at CERN and other in University of Cap Town (South Africa). He has got a as a postdoctoral position at the University of Geneva group to work on ATLAS detector and Grid Facilities
- Alvaro Fernández Casani, started to work at IFIC with project contracts and he got the position of Senior Specialist Technologist (CSIC staff)
- Miguel Villaplana Pérez, after his Master Thesis on July 2008, he obtained his Ph D with the thesis ‘‘Estudio de resonancias exóticas en el experimento ATLAS utilizando

una infraestructura con tecnologías GRID a partir de un prototipo de Tier-3 para el IFIC'' supervised by Santiago González de la Hoz and Marcel Vos. The thesis was presented in December 2013 and the mark was Excellent Cum-Laude. Miguel Villaplana started two year ago a Postdoc in the INFN at Milan to work in ATLAS physics analysis related with extra dimensions and to help in the maintenance of the Tier-2 infrastructure at Milan.

- Elena Oliver, Doctoral student at IFIC. Her PhD 'Study of Jet Substructure in the ATLAS Experiment using Distributed Analysis within the Spanish Tier-2 Infrastructures'.Work Directors: José F. Salt Cairols and Santiago González de la Hoz. The thesis was presented in December 2013 and the mark was Excellent. Elena Oliver is currently working in a private company.
- Victoria Sánchez Martínez, she has finished her PhD thesis and it will be defended by May-June 2016. The thesis title is: 'Estudio del quark top en modelos de nueva física utilizando las infraestructuras Tier2 y Tier3 de Valencia'. Work directors: Santiago González de la Hoz & Marcel Vos. Mark

Moreover, we have supervised Graduate Works and Master Thesis of several students during the mentioned period.

2) To teach courses and tutorials for researchers and/or technicians by means of Courses, Lectures, etc:

- The IFIC team participate in the joined Master and Doctoral Program of three Departments of the Valencia University, namely: FAMN, Theoretical Physics and Astrophysics.
- CSIC Postgraduate Courses “GRID & e-Science” organized by IFIC group, Years: 2005, 2008 and 2010. 3 days course.
- Tutorial of ATLAS Distributed Analysis and Computing. Organized by the Spanish ATLAS Tier-2, IFIC, 2 days tutorial. Years: 2007 and 2009.
- Course of Specialized Training on system Management of an ATLAS T2 Infrastructure. Given by our team during a week in November 2014. Trainees: 4 people of Computing Staff from CERIST (Argel)

IFIC group plans to supervise a new PhD student (FPI grant) who can work in Data Analysis within the Physics Analysis team. This will allow the student to work in ATLAS computing issues and it can be used for his/her authorship qualification for ATLAS publications.

UAM: C. Glasman and J. del Peso are professors at UAM University. They both give lectures in undergraduate and graduate physics studies. New members joining the Group enter a multidisciplinary university environment where regular conferences and lectures are offered. In the last 10 years C. Glasman and J. del Peso have supervised five PhD thesis, namely:

- Measurement of photon plus jets production and identification of boosted top quarks in pp collisions at the LHC using the ATLAS detector. 2015 - Hector de la Torre. Supervisors: C. Glasman and J. Terron
- A study of the dynamics of isolated-photon plus jet production in pp collisions at 7 TeV c.m. energy with the ATLAS detector. 2013 - Josu Cantero. Supervisors: C. Glasman and J. Terron
- Energy Reconstruction in the ATLAS Electromagnetic End-Cap Calorimeter with Calibration Hits. 2010 - Eduardo Nebot. Supervisor: J. del Peso
- Calibration of the Electromagnetic ATLAS Calorimeter and search for the W’ exotic boson. 2010 - Carolina Gabaldon. Supervisor: J. del Peso
- Uniformity of the Electromagnetic EndCap Calorimeter of Atlas. 2006 - Concepcion Oliver. Supervisor: J. del Peso

We plan to supervise a new PhD student (FPI grant) performing the data analysis described in a previous section about “User Support and Data Analysis”. In parallel the student will work in a topic of ATLAS computing, which may be used as their authorship qualification for ATLAS publications.

In addition, the UAM group has trained several generations of computing engineers on system management in a Grid environment who, thereafter, have found a quality job in industry or public institutions. The project gives the opportunity to gain experience with computing tools used in a Grid infrastructure of unparalleled dimensions: about 100 centres all over the world. In particular, people hired by the project can acquire practical knowledge on distributed storage systems, Grid batch system and large scale monitoring. Besides, the UAM group can train on tools for automatic installation and configuration of computer centres, since it has contributed to the development and maintenance of one such system (Quattor) for more than six years.

C.5. IMPLICACIONES ÉTICAS Y/O DE BIOSEGURIDAD