IFIC
INSTITUT DE FÍSICA CORPUSCULAR

EXCELENCIA SEVERO OCHOA

User Support & Physics Analysis at IFIC (Machine Learning)

Presentado por: José F. Salt Caírols
IFIC

12 de Abril de 2018

1ª Reunión Presencial del Proyecto VMGRID (ATLAS TIER2 España)
FPA2016-75141-C2

Universidad Autónoma de Madrid
12 y 13 de Abril de 2018

Contenido:

1.- User Support y contribución al ATLAS Distributed Computing

a) Status report User Support @ IFIC

b) Contribución al DAST

c) Contribución al ADCoS

2.- Physics Analysis usando ML

3.- Conclusiones y Perspectivas

1.-User Support and contribution to ATLAS Core Computing

Hoja de ruta de ATLAS

Run 1 (sqrt(s)=7+8 TeV, Lint = 25 fb-1)

2011 7TeV 5.46 fb-1
2012 8TeV 20.3 fb-1

LS1 (2013-2014)

Run 2 (sqrt(s)=13 TeV, Lint = 75-100 fb-1)

2015 3.21 fb-1
2016 35.6 fb-1
2017 46.9 fb-1
2018 ??

LS2 (2019-2020)

Run 3 (14 TeV) ==> (300 fb-1)

GRID Paradigm

Tutoriales+ consolidación

Ecosystem with different Computing approach

Integrated Luminosity x 4

Alto nivel de automatización

Future computing

Integrated Luminosity x 10

1-a User Support @ IFIC

- Se ha logrado un gran nivel de automatización del Tier-2 y del Tier-3
- Almacenamiento: la opción de lustre está resultando muy positiva
- Se utilizan los UI
  - Las muestras se procesan desde las máquinas "user interfaces" sctui0X.ific.uv.es y las salidas (los "trees" de análisis) –
  - se almacenan en el disco: /lustre/ific.uv.es/grid/atlas/t3/usuario/
- Muy bien valorada la ayuda del grupo Tier-2/Tier-3
- Interés de los usuarios ATLAS/IFIC: Se necesita una granja local donde enviar jobs (en batch) nivel Tier-3
- Actividad creciente en ML en los análisis de física/estudios de los detectores: disponer de una infraestructura con GPU's

- **Documentación: es muy completa y se utiliza por los usuarios consultando los aspectos que necesitan para sus actividades**
  - **Necesita actualización:Hay que actualizarla: DQ2-> Rucio**
  - **Etc**
- **Twiki:**

https://twiki.ific.uv.es/twiki/bin/view/Atlas/GridComputing

Grid Computing for ATLAS

Introducción

Esta página contiene enlaces a las páginas wiki relacionadas con las actividades de computing en el marco de los proyectos ATLAS y EGEE

Enlaces útiles:

Atlas Data Processing @ IFIC	Atlas Data Processing using the IFIC and Grid Resources
EGEE	Actividades del EGEE
GridHowTo	Como usar el GRID
AtlasDataAnalysisTutorial	Tutorial sobre análisis de datos en ATLAS
TipsAndTricks	Recetas útiles para usuarios

1-b Contribución al DAST (Farida)

- **Distributed Analysis Support Team, DAST**
- **Farida Fassi: coordinadora del DAST en EU timezone**
- **Objetivos:**
  - **1) proporcionar ayuda especializada a todos los físicos analistas de ATLAS,**
  - **participar en el desarrollo y en la implementación de nuevas mejoras del sistema,**
  - **3) validar las instalaciones de software necesarias para el procesado de datos in los diferentes centros**
  - **4) comprobar el acceso a los datos y la recuperación de los datos analizados para prevenir posible problemas del sistema de almacenamiento de datos en general**
  - **5) colaborar estrechamente con la persona del soporte de los centros distribuidos para resolver los problemas de los trabajos de análisis que han fallado en dichos centros .**

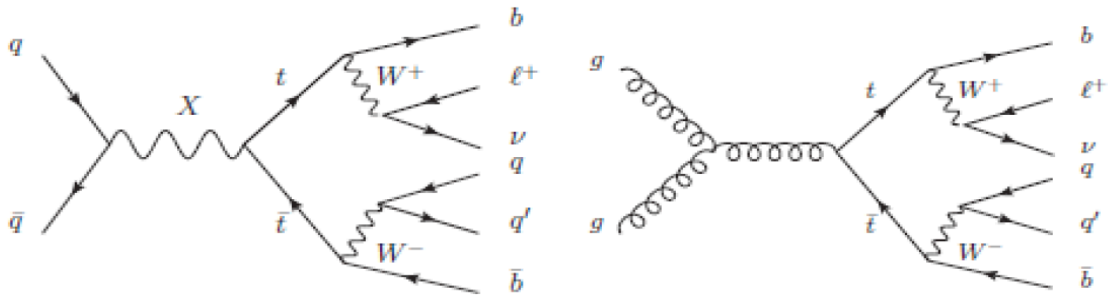
1-c Contribución a los shifts ADCoS

- ATLAS Distributed Computing Operation Shifts: ADCOS
- Link:
  - https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/ADCoS?redirectedfrom=Atlas.ADCoS
- José Salt: ADCoS Senior Shifter
- Objetivos:
  - Control de las transferencias de datos
  - Monitoraje de los ATLAS Central Services
  - BigPanDA
  - Control de Tasks que fallan, borrado de datasets, etc
  - Report de los errores en los sites (abrir GGUS tickets)
  - Control de los Frontier y los Squads
  - etc

- 2-3 shifts al mes (promedio)

2.- Physics Analysis using ML

- **Objetivo: Extracción de la señal de resonancias ttbar de entre el fondo de procesos de SM mediante técnicas de Análisis Multidimensional y Machine Learning**



- **Metodología/Técnicas: incorporar métodos de Machine Learning**
  - **Análisis Multidimensional =>**
  - **ANN**
  - **Deep Learning**
  - **BDT**

TFG => Rubén Rodríguez Casañ -> <u>Samuel Campo Martínez</u>

- Tenemos 27 parámetros.

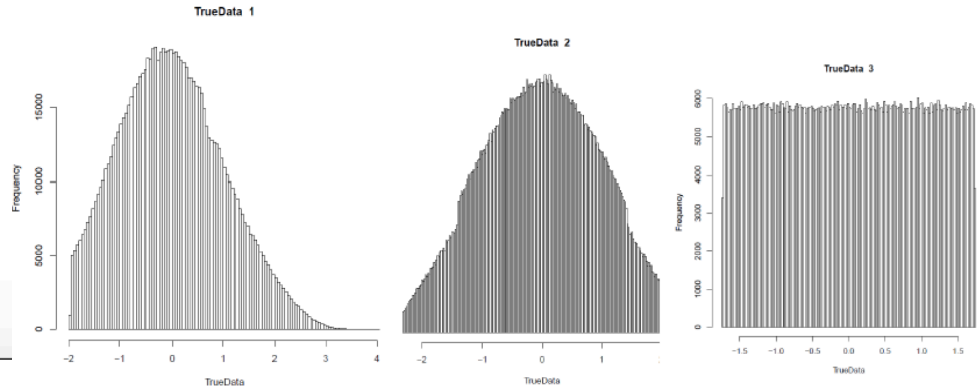
Se dividen en 21 low-level y 5 high-level:

Low-level data:

- **Momento** del lepton principal
- Momento de los cuatro jets principales ($b,\bar{b},q,q'$)
- **Pseudorapidity**.
- **Ángulo azimutal**.
- Información sobre el b-tagging.

High-level data, masa de estos objetos:

- $W \rightarrow l\,\nu$
- $W \rightarrow q\,q'$
- $t \rightarrow W\,b \rightarrow b\,q\,q'$
- $t \rightarrow W\,b \rightarrow l\,\nu\,b'$
- $X \rightarrow t\,\bar{t}$

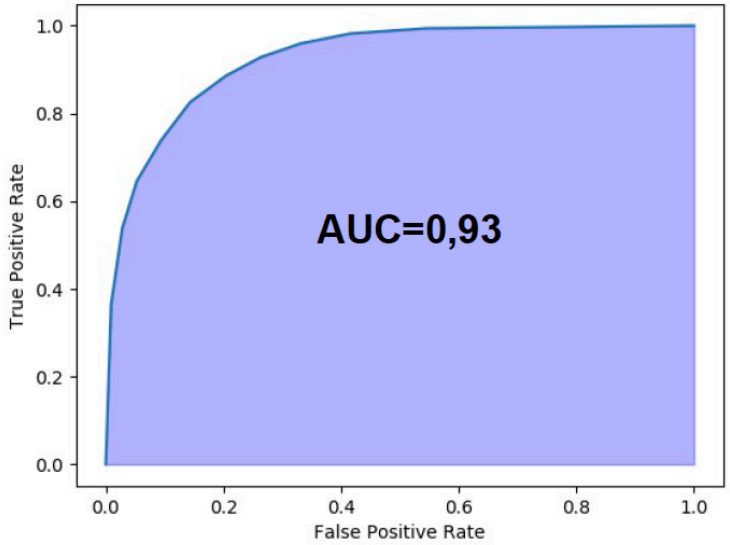


MOMENTO JET PSEUDORAPIDITY ÁNGULO AZIMUTH

- **Caso práctico de aplicación de ML:**
  - **26 features/variables para analizar con dos niveles: 21 de primer nivel y 5 de segundo nivel (masas invariantes)**
  - **Más de 7 millones de sucesos**
  - **2 etiquetas:**
    - **Sucesos del SM**
    - **Sucesos resonancias ttbar**
  - **Clasificación en dos etiquetas (tags) entre 196 millones de números no enteros**

- **Se construye la curva ROC**
  - **(Receiving Operating Characteristic): razón de verdaderos positivos vs razón de falsos positivos**

- **Y se calcula la AUC: Area Under Curve => en ML sirve de comparador de modelos**



Data types – Learning model type

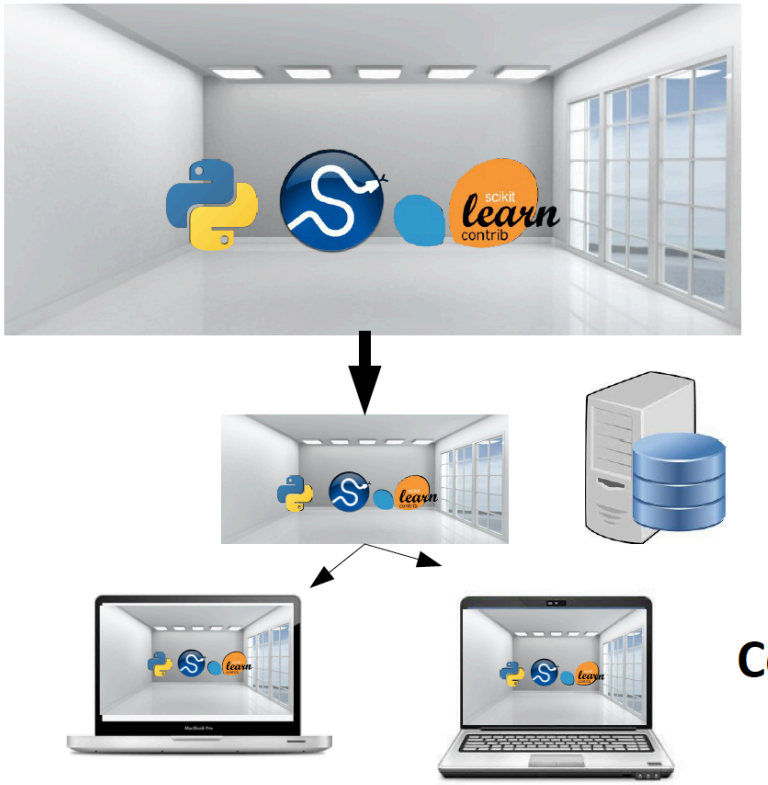
Label data → Supervised learning

Decision Tree from IRIS

Unlabel data → Unsupervised learning

Clustering Iris

Code Laboratories: Enviroments and Dockers



Local computer

Shared enviroment

Companion local computer

Code Laboratories:

- **Anaconda:**

Entornos que pueden ser copiado compartidos pero se necesita instalación
El más sencillo y más común

- **Docker:**

Containers que pueden ser copiados y compartidos sin necesidad de instalación

El más ligero y potente

Speed-upt he obtaining of results sending the execution to GPU'

Improve our analysis on Neural Network using

References to learn more...

Machine Learning explanations and examples:

https://machinelearningmastery.com/
http://scikit-learn.org
https://stackoverflow.com/

Datasets to practise:

https://archive.ics.uci.edu/ml/index.php
https://www.kaggle.com/datasets

3.- Conclusiones y Perspectivas

- **User Support @ IFIC**
  - Razonablemente bien
  - Actualizar algunos aspectos que han quedado obsoletos
  - Incorporar' una vez consolidado, HPC, Cloud, etc
  - Integración con la UAM (?)
- **Contribución DAST y a los shifts de ADCoS en ATLAS**
  - Esfuerzo que cuenta en OTP ATLAS
  - Es una contribución a conservar dentro de ATLAS
- **Análisis de Física aplicando métodos de ML**
  - Todavía en fase inicial tanto en comprensión como en despliegue
  - Perspectiva prometedora: actualización infraestructura, trabajos de estudiantes TFG y TFM
  - Momentum de varios grupos del IFIC en ML

GRACIAS POR VUESTRA ATENCIÓN

PREGUNTAS