Instituto de Astronomía Ensenada

Detección y clasificación de Objetos Trans-Neptunianos: Proyecto TAOS II

Por

Benjamin Hernández Valencia
Proyecto Taos-II

Introducción

La detección y clasificación de objetos opacos mas allá de la orbita de Neptuno (TNOs) es un problema de interés en la comunidad astronómica. Una ocultación es detectada, cuando un cuerpo opaco pasa por la linea de visión entre una estrella que se está observando y el telescopio. En este instante, el flujo luminoso total de la estrella, llamada curva de luz, experimenta una modificación en su perfil observado en el espacio tiempo, que depende del tamaño del TNOs, su distancia al observador y propiedades físicas de la estrella, entre otros factores. Estas curvas de luz serán obtenida de la observaciones producidas por el Censo Automático de Objetos Trans-Neptunianos, proyecto TAOS II, ver Figura 1.

TAOS II es una proyecto Internacional en las que participan Taiwan (SINICA), México (IA-UNAM), Canadá (NRC-CNRC & CADC) y Estados Unidos (CFA). Consta de un arreglo de telescopios instalado en la sierra de San Pedro Mártir, Ensenada, B.C. México y detectores de alta candencia (20 Hz). Se espera observar 50,000 estrellas por noche por 5 años, que en su conjunto almacenará del orden de 3.5 PetaBytes de datos.

El algoritmos de **De**tección y **Cla**sificación de **O**bjetos **T**rans-**N**eptuniano**s** (DeClaTNOs) consta de un conjunto de 4,800 perfiles teóricos distintos del comportamiento de los patrones de difracción de las ocultaciones. Extrayendo rasgos estadísticos de primer orden, cambios de energía de la señal, un análisis de espacios Gaussianos a diferentes escalas y utilizando operadores de interés, todos ellos sobre las curvas de luz, se generá un vector de rasgos prominentes de cada uno de estos patrones, que se espera identifiquen unívocamente el fenómeno que los provoca. Utilizando el paradigma de Maquinas de Vector de Soporte, se detecta primero la presencia o ausencias de un evento y seguido la clase de objeto que la produce.

En este proyecto, se utilizará la infraestructura de **Grid UNAM** instalada en la Dirección General de Cómputo y Tecnologías de la Información y Comunicaciones (DGTIC), el proyecto de LAMOD del Instituto de Astronomía (IA) y el Instituto de Ciencias Nucleares (ICN), el Instituto de Astronomía Ensenada (IAE) y el Instituto de Ciencias de la Atmósfera y de Cambio Climático (ICACC). La intensión es evaluar el desempeño de nuestro algoritmo,

así como adquirir experiencia en el uso de esta nueva tecnología, instalada por primera vez en México. DeClaTNOs se ejecutará para un conjunto de aproximadamente 10,000 estrellas, que representa un bloque de observación de 2 hr en los telescopios de TAOS II y aproximadamente 50,000 estrellas, que representan una noche completa de observación.

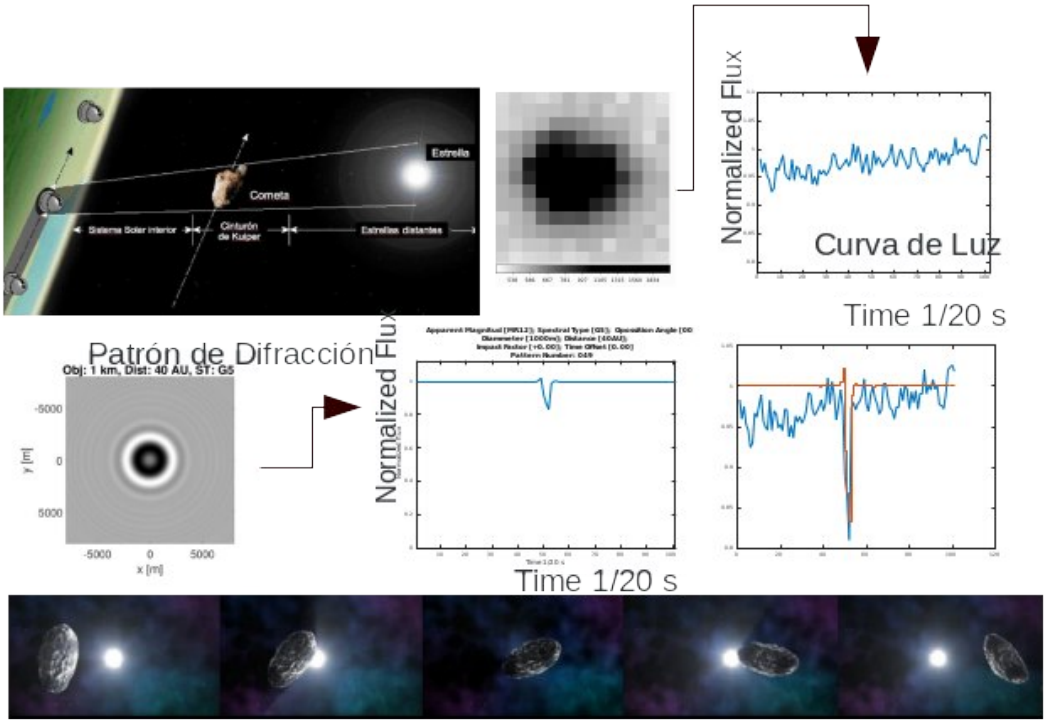

Figura 1: Proceso general de detección de una ocultación.

Responsables y recursos del proyecto

- Responsable Académico: Benjamín Hernández
- Colaboradores:
    - Joel H. Castro
    - Mauricio Reyes
    - Matthew Lehner
    - Carlos Guerrero

- Jose Silva
- J Hernández
- F Alvarez
- Edilberto Sanchez
- Manuel Nuñez
- LT Calvario
- L Figueroa
- C-K Huang
- Shian-Yu Wang
- W-P Chen
- A Granados
- JC Geary
- KH Cook
- JJ Kavelaars
- T Norton
- A Szentgyorgyi
- W-L Yen
- Z-W Zhang

- Responsable Técnico: Benjamín Hernández
- Asesor GridUNAM: Benjamín Hernández
- Entidad: Instituto de Astronomía Ensenada
- Servidor de envío: jamatu.astrosen.unam.mx

Requerimientos Técnicos

Las especificaciones para ejecutar el sistema para una estrella son las siguientes:

- Nombre de ejecutable: DeClaTNOs
- Compilador: g++
- Bibliotecas adicionales: libsvm.so
- Numero de núcleos: 1 (ejecutable serial)
- Memoria RAM: 0.7 GB
- Espacio disco: ~1.5 GB

Para el procesamiento de una noche de observación se requiere:

- Ejecutar del orden de 50,000 veces DeClaTNOs

- Espacio en disco para todos los modelos: ~ 2 GB
- Espacio en disco para las 50,000 estrellas: ~ 190 GB
- Espacio en disco de archivos adicionales: ~ 1 GB
- Total de espacio: ~ 200 GB

En este experimento los archivos de datos contiene unicamente la curva de luz, que son representadas por 3 vector de 144E+3 elementos por estrella. La imagen bidimensional de cada estrella se omite. Asimismo, los vectores de la curva de luz son sintéticos, en donde se les insertó una ocultación en una posición aleatoria, o bien ninguna. Los objetos trans-neptunianos que se insertaron y las características de las estrellas sobre el cual pasa el objeto, son las siguiente:

- Diámetro del objeto: 1 km, 2 km, 5 km, 10 Km
- Distancia del objeto: 30 AU, 40 AU, 50 AU
- El objeto pasa por el ecuador de la estrella: factor de impacto 0
- El muestreo de la ocultación se captura completamente: tiempo de corrimiento 0
- Magnitud aparente de la estrella observada: MR12, MR14, MR16
- Tipo espectral de la estrella: A3, G5, M4
- Ángulo de oposición del campo observado: 0^o, 30^o, 60^o

Estrategia para uso en la Grid UNAM

DeClaTNOs es una aplicación que se ejecuta en forma serial para cada una de las estrellas. Consta de 12 parámetros, que describen el metadato de la observación y el nombre del modelo que se aplicará, ver Recuadro 1. Tiene un tiempo de ejecución promedio 1m30s con una máximo de 4m30s y un mínimo de 30s, en un servidor Xeon E5-2630 @ 2.2 GHz, llamado arrokoth. Esta variación depende de la magnitud aparente de la estrella observada y el

```
$ ./DeClaTNOs --help
Usage: ./DeClaTNOs [option(s)] VALUES
Options:
        -h,--help               Show this help message
        -f,--filename  HDF5_FILENAME     File Name
        -A,--telA  HDF5_Ligth_Curve      Telescope A (please include the group name)
        -B,--telB  HDF5_Ligth_Curve      Telescope B (please include the group name)
        -C,--telC  HDF5_Ligth_Curve      Telescope C (please include the group name)
        -d,--diffraction  DifraccionPattern_FILENAME Database of the Theoretical Diffraction
                                            File Name
        -M,--mr [MR12|MR14|MR16]         Apparent magnitude
        -S,--spectral [A3|G5|M4]         Start Spectral Type
        -O,--opossition [OA00|OA30|OA60] Opposition Angle of the Field
        -p,--overlap   overlap_factor    Overlap factor for windows analysis
        -m,--svm BSVM_model_FILENAME Prefix  SVM Model wiht postfix
                                                [.3T.2C.svm*|.1T.2C.svm*|.3T.180C.svm*]
        -o,--output OUTPUT_FILENAME      File name of Posible TNOs in plain txt format
        -g, --debug [0|1|2]               verbosity info 0=no Info; 1 +Info 2 ++Info


        ---------
        TAOS II Team  - UNAM, Mexico; SINICA, Taiwan; CFA, USA -
        http://taos2.astrosen.unam.mx/
        http://taos2.asiaa.sinica.edu.tw/
```

Recuadro 1: Parámetros para la ejecución de DeClaTNOs

Tomando en cuenta estas características , la estrategia que se utilizó es la siguiente, vea Figura 2:

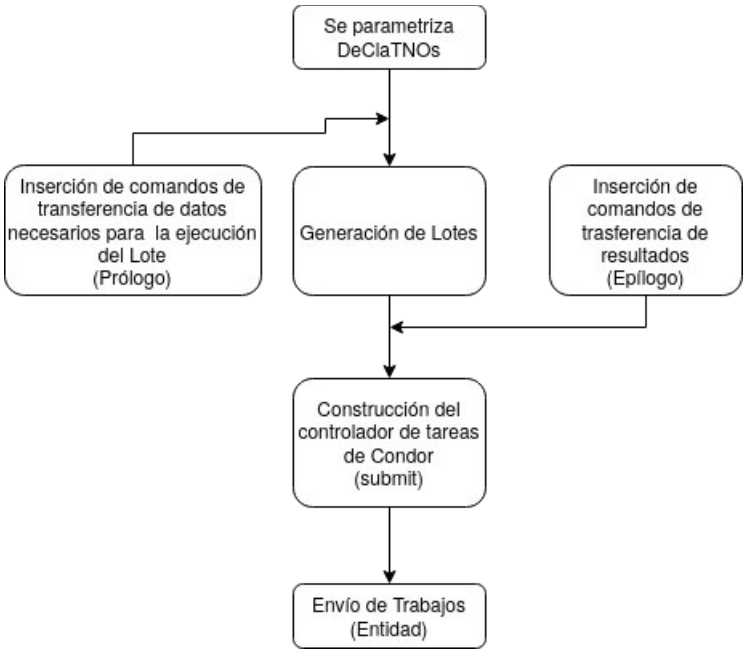

Figura 2: Diagrama de flujo de la estrategia para la ejecución de DeClaTNOs en Grid UNAM. Vea texto.

1. **Se parametriza DeClaTNOs**: Se asigna el metadato de la estrella observada y los parámetros intrínsecos de la observación. Por ejemplo, en el Recuadro 2, la primera linea corresponde al nombre del archivo, en formato HDF5, que contiene la estrella que será analizada. Para este caso el valor de la variable es

theFileHDF5="20230525T204000.D002.h5" . Las siguiente linea corresponden a una ejecución del algoritmo para una estrella. Las dos lineas restantes “theDate” sirven para obtener el tiempo Juliano (época 1970) en el que terminó de ejecutarse esa instancia. Esta información servirá para mostrar el rendimiento en el entorno **Grid UNAM**.

```
theFileHDF5="20230525T204000.D002.h5"
…….
…….
./DeClaTNOs -f 20230525T204000.D002.h5 -A /E03a9b/TA/LC -B /E03a9b/TB/LC -C /E03a9b/TC/LC -d
DiffractionPattern.h5 -M MR16 -S A3 -O OA00 -m M001_MR16_A3_OA00 -o 20230525T204000_E03a9b.txt -p 2 -g
0
theDate=`date +%s`
echo "The Date: ${theDate}"
…….
…….
```

Recuadro 2: Bloque de parámetros

Para este caso, el metadato de la observación corresponde a: "RunID"="20230525T204000"; "StarID"="/E03a9d"; "StarName"="Artificial Star"; “SNR”=6.2; “MR”= “MR16”; “SpecType”= “A3” y “OpposAngle”= “00”.

2. **Generación de lotes de trabajo**: Se decidió agrupar la ejecución en lotes de trabajo de 351 estrellas, que resulta de insertar un patrón de difracción distinto por tipo de objeto (4 diámetros de objetos, 3 distancias, 3 magnitudes aparentes de estrellas, 3 tipos espectrales de estrellas y 3 ángulos de oposición) y 27 curvas de luz sin inserción de patrón (solo ruido). Se generaron 150 lotes distintos para ser enviados a Grid UNAM.

   Es así, que cada lote contiene 351 bloques de linea similares a la mostrada en el Recuadro 2.

3. **Prólogo**: Se diseñó un archivo que contiene las instrucciones necesarias para transferir los archivos necesarios para ejecutar cada lote. Estos archivos se transfieren previamente al sistema de almacenamiento S3 de Grid UNAM. En particular se utilizó el Instituto de Astronomía Ensenada (IAE). Cabe hacer notar, que la ejecución no dependerá de donde se almacenen los datos de entrada.

El archivo de prólogo se insertó al inicio de cada lote, ver el Recuadro 3. Consta de 4 secciones que se describen a continuación:

1. Información general. Consta de la fecha de inicio del trabajo y nombre del nodo en donde se está ejecutado el lote.
2. Los comandos para transferir los datos necesarios para su ejecución, desde el almacenamiento S3 hacia el espacio de almacenamiento local del nodo de trabajo. Note que los modelos de la máquina de vector de soporte previamente entrenados (Modelos.tar), las bibliotecas especiales (SVM.tar) y los archivos HDF5 comunes (H5_CommonFiles.tar), están empaquetados en un solo archivo utilizando el comando “tar”de linux. Asimismo, el desdoblamiento de estos archivos se ejecuta en en mismo instante de la transferencia. Finalmente, se copia solamente el archivo que contiene las curvas de luz de ese bloque ($theFileHDF5). Existen 150 archivos $theFileHDF5 distintos, uno por bloque.
3. Se modifican las variables de ambiente para que el sistema pueda leer las bibliotecas especiale en el momento de ejecución.
4. Finalmente, se captura el tiempo en que terminó la transferencia de datos. Esto es equivalente también, al tiempo de inicio de la ejecución del lote de trabajo.

```
####################
## General Info
####################
echo "Command: $0 $@ "
theDateH=`date `
echo "RunDate: ${theDateH}"
theHost=`hostname`
echo "Hostname: ${theHost}"

#################
#Transfer Files
#################

echo "Estoy por default en: `pwd` "
echo "El home: ${HOME}"
echo "El Grid Work: ${GRIDUNAM_WORKDIR}"

cd  ${GRIDUNAM_WORKDIR}
echo "Ahora debo estar en Grid Work: `pwd`"
cp  ${HOME}/../DeClaTNOs .

mcli cat  gridunam/tnos/SVM.tar | tar xf -
mcli cat  gridunam/tnos/Modelos.tar | tar xf -
mcli cat  gridunam/tnos/H5_CommonFiles.tar | tar xf -
mcli cp  gridunam/tnos/$theFileHDF5 .

###################
## Env Variables #
###################
export LIBDIR=./
export LD_LIBRARY_PATH=$LIBDIR:$LD_LIBRARY_PATH

###################
# Execution Pool #
###################

theDate=`date +%s`
echo "The Date: ${theDate}"
```

Recuadro 3: Contenido del Prólogo de la ejecución

4. **Epílogo**. Se diseñó un archivo que transfiere los resultados al final de cada lote, hacia un directorio de resultados en el almacenamiento S3, *gridunam/tnos/Res01*. Al terminar libera el espacio de almacenamiento local (${GRIDUNAM_WORKDIR}), ver Recuadro 4.

Los 4 procesos anteriores definen completamente un lote de trabajo que se almacena en un archivo de texto con el nombre run_RunID.DXXX.sh. En donde RunID es el descriptor de la fecha de la observación, equivalente a 10,000 estrellas observadas por un periodo de tiempo de 2 horas, a una cadencia de 20 imágenes por segundo. En una noche de observación se observan 5 RunID. D es un identificados de bloque que contiene 351 estrellas. XXX es un numero consecutivo en el rango de 1 a 30, un lote de trabajo, que en su conjunto agrupan a 10,500 estrellas. Entonces, 30 archivos con el mismo RunID es una observación y una noche de observación está compuesta, para este experimento, del 150 lotes.

Finalmente, todos estos archivos se convierten en ejecutables. Se emplea el comando chmod a+x run_RunID.DXXX.sh, y se transfieren al almacenamiento S3 de Grid UNAM.

```
###################
#Transfer Results
###############

mcli cp *.txt gridunam/tnos/Res01
mcli cp $HOME/../*.log  gridunam/tnos/Res01
mcli cp $HOME/../*.out  gridunam/tnos/Res01
mcli cp $HOME/../*.err  gridunam/tnos/Res01


##############
# Clear space#
##############

rm -f *

#########
```

Recuadro 4: Epílogo del lote de trabajo. Vea texto.

5. **Construcción del controlador tareas, basado en HT-condor**. La idea principal es controlar el numero de lotes de ejecución en alguna entidad, simplemente generando un archivo con los nombres de los bloques. Por ejemplo, si se genera un archivo de texto con los nombres de 10 lotes, implica que HT-Condor enviará 10 tareas y utilizara 10 núcleos de proceso en alguna de las entidades. Para el caso de DGTIC, se envió un archivo con los nombres de los 150 bloques y para el caso de IAE, se envió un archivo con los nombres de 60 bloques. Para lograr ésto, el procedimiento es es

siguiente:

- **Construcción del archivo "submit"**. El despachador de tareas y los comandos de control están basados en HT-Condor. En forma breve se explicará cada linea utilizada, sin embargo hay que tener en consideración que existen muchas mas directivas, dependiendo del tipo de trabajo que se quiera enviar. En el Recuadro 5 se muestra el prototipo de este archivo, en donde la palabra ENTIDAD corresponde a una dependencia que pertenece a Grid UNAM. La descripción de cada línea es la siguiente:

```
# Required for remote HTCondor-CE submission
universe = vanilla

# Necessary for S3 storage
environment = "MC_HOST_gridunam=$ENV(MC_HOST_gridunam)"

# Files
executable = ./theMasterTNOS.sh
arguments = $(losBloques)
output = tnos.ENTIDAD.vanilla.$(Cluster).$(Process).out
error = tnos.ENTIDAD.vanilla.$(Cluster).$(Process).err
log = tnos.ENTIDAD.vanilla.$(Cluster).$(Process).log

# File transfer behavior
ShouldTransferFiles = YES
WhenToTransferOutput = ON_EXIT
transfer_input_files = losBloques_ENTIDAD.txt, DeClaTNOs

#Run jobs as line number contains  losBloques_ENTIDAD.txt
queue losBloques from losBloques_ENTIDAD.txt
```

Recuadro 5: Construcción del archivo "submit"

- *universe=vanilla* : todos trabajos serán de este tipo para la versión actual de Grid UNAM.
- *environment = "MC_HOST_gridunam=$ENV(MC_HOST_gridunam)"* : Definición del token para el acceso al almacenamiento S3. Este token es generado por el usuario en el nodo submit antes de enviar un trabajo que requiera almacenamiento S3.
- *executable = ./theMasterTNOS.sh* : Es el nombre del proceso maestro de

envío. Vea la sección siguiente.

- *arguments = $(losBloques)* : El proceso maestro se envía n-veces como el numero de lineas en la variable $(losBloques).
- *output = tnos.ENTIDAD.vanilla.$(Cluster).$(Process).out*
  error = tnos.ENTIDAD.vanilla.$(Cluster).$(Process).err
  log = tnos.ENTIDAD.vanilla.$(Cluster).$(Process).log : nombre de lo archivos de salida, error y estado de la ejecución.
- *ShouldTransferFiles = YES*
  WhenToTransferOutput = ON_EXIT : activar la transferencia de archivos y la recuperación al terminar el proceso.
- *transfer_input_files = losBloques_dgtic.txt, DeClaTNOs* : Se transfieren el archivo con los nombres de los lotes y el ejecutable para la detección y clasificación de trans-neptunianos.
- *queue losBloques from losBloques_ENTIDAD.txt* : la variable losBloques es un vector que contiene el nombre de los bloques que están definidos en el archivo losBloques_ENTIDAD.txt

- **Construcción del archivo maestro (theMaterTNOS.sh)**. El archivo maestro es simple, ver Recuadro 6.
  - *TheB=$1* : Un valor de la variable losBloques es el primer argumento que se asigna a TheB.
  - c*d $HOME/../* : Es ubica en el directorio por omisión en donde HT-Condor trasfiere los archivos.
  - *mcli cp --preserve gridunam/tnos/$theB .* : se transfiere del almacenamiento S3 el nombre de bloque. Es importante utilizar la directiva "*--preserve*" para que la propiedad de ejecución se conserve.
  - *./$theB* : Finalmente se ejecuta el bloque.

```
#!/bin/bash

theB=$1

cd $HOME/../

mcli cp --preserve gridunam/tnos/$theB .

echo "The Run: $theB"

./$theB
```

Recuadro 6: Programa maestro de envío de lotes.

6. **Envío de Trabajos.** Para el envío de trabajos se empleó el sistema desarrollado por el grupo de desarrolladores de Grid UNAM, tanto para la generación de los token de procesamiento como los de almacenamiento. En el siguiente ejemplo se reporta el estado de DGTIC y el envío de los 150 lotes.
    - Se inspecciona el estado actual de la entidad DGTIC, utilizando el comando “gu_show_cpu_cluster $dgtic”. En la Figura 3 se muestra una captura de pantalla de ese comando, desde el nodo submit de la entidad IAE (desde Ensenada). Se observa que existen 224 núcleos disponibles y es posible enviar los 150 lotes para procesar del orden de 50,000 estrellas.

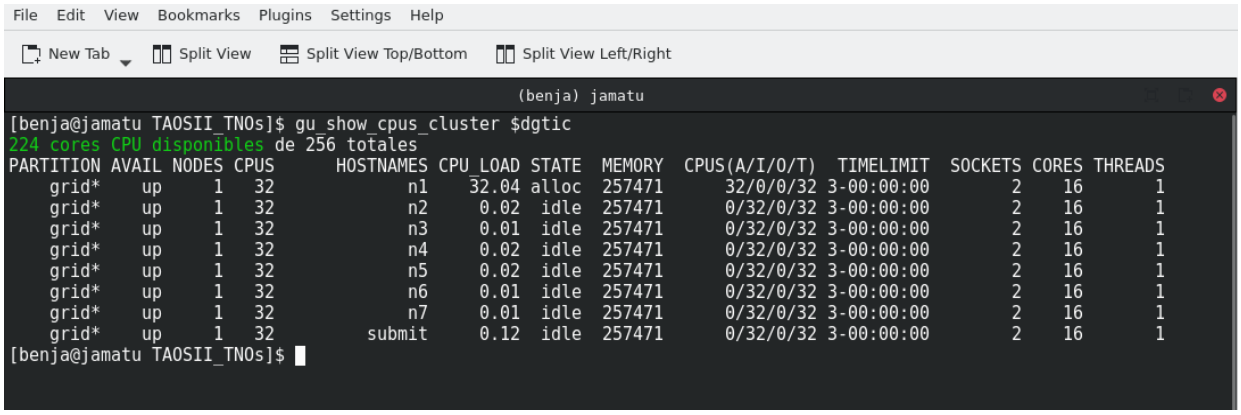

Figura 3: Inspección del estado actual de la Entidad DGTIC, utilizando los comandos de **Grid UNAM**.

- Se envía el trabajo con el comando “gu_enviar_jobs tnos.dgtic.vanilla.sub $dgtic”, ver Figura 4. El trabajo asignado tiene el numero 3889 y se enviaron 150 lotes.

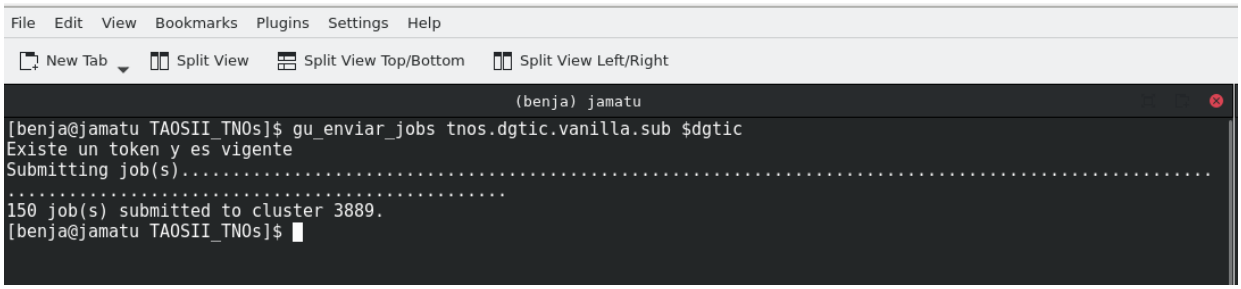

Figura 4: Envío de un trabajo desde IAE (Ensenada) utilizando los procesadores de DGTIC.

- Finalmente se puede inspeccionar el estado del trabajo en cualquier momento posterior con el comando “gu_consltar_jobs $dgitc”. En la Figura 5 se muestra que que el trabajo 3889 está ejecutandose en la entidad DGTIC. Hay que hacer notar, que las tomas de pantalla se realizaron días después de la obtención de resultados. La intensión es mostrar el uso de los comandos en este documento.

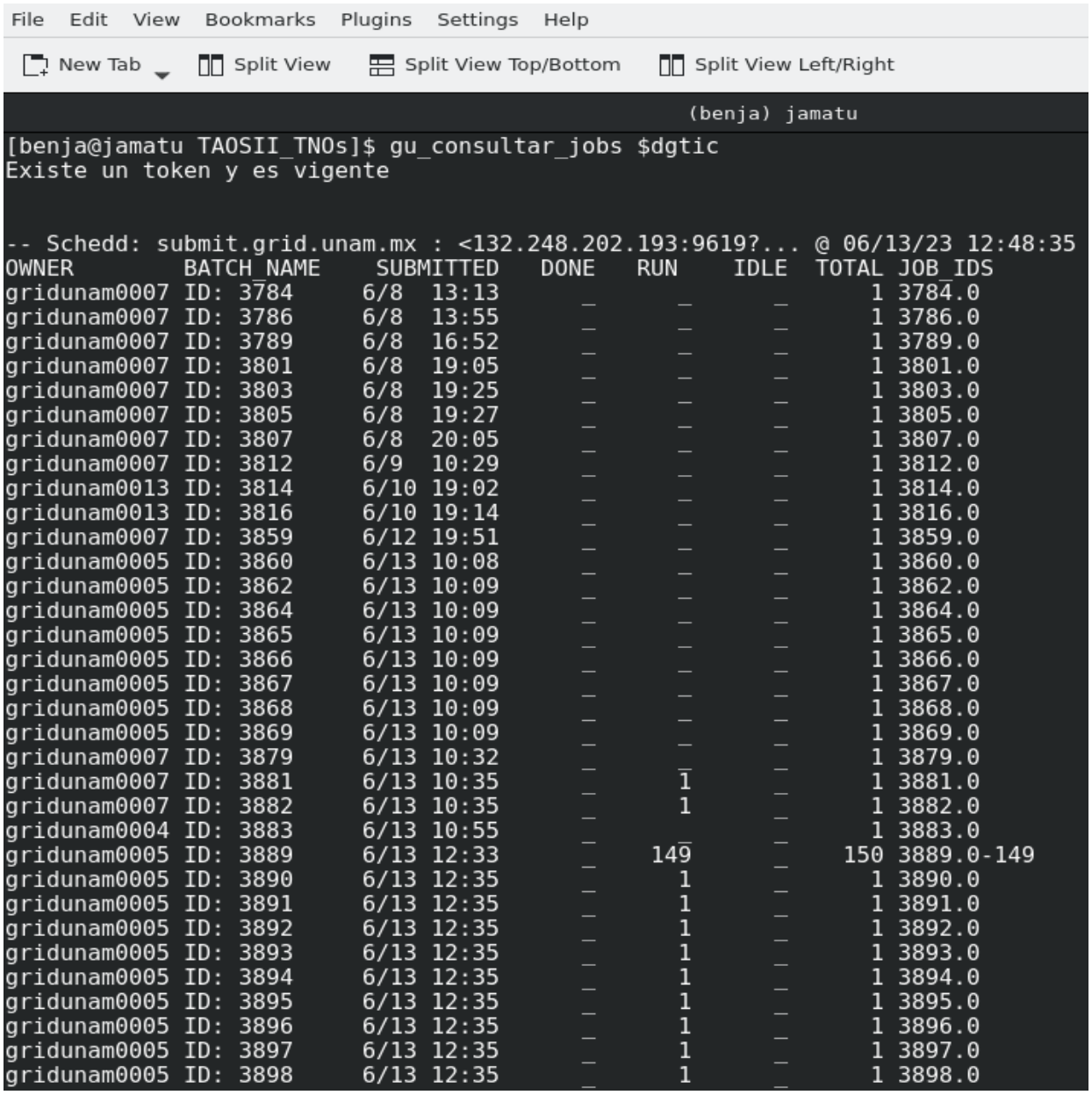

Figura 5: Consulta del estado de un trabajo

Resultados

Se realizaron 2 experimentos, en donde el el nodo de envío, jamatu.astrosen.unam.mx, y el nodo de almacenamiento S3, atenea.astrosen.unam.mx, se encuentran instalados en el Instituto de Astronomía Ensenada ubicado en Ensenada. B. C., México.

Todos los archivos necesarios para la ejecución se almacenan en S3 y fueron transferidos previamente al directorio ***gridunam/tnos*** desde un servidor que no forma parte de Grid UNAM. Estos ocupan del orden de ~200 GB de espacio. Su descripción se muestra a continuación:

- Se utilizaron 162 modelos previamente entrenados utilizando el paradigma de maquina de vector de soporte. Estos se empacaron en un solo archivo utilizando el comando “tar”, llamado Modelos.tar, con un tamaño de ~1.5 GB
- Los 2 archivos HDF5 comunes en toda ejecución, H5_CommonFiles.tar, tiene un tamaño de ~0.4 GB.
- Las bibliotecas especiales de la maquina de vector de soporte, SVM.tar, tiene un tamaño de 0.04 GB.
- Los 150 archivos HDF5 de cada lote, llamados RunID.DXXX.h5, ocupan ~ 180 GB. Estos archivos no se empaquetan dado que cada núcleo o lote solo requiere de uno de ellos para su ejecución.
- Los 150 archivos de ejecución, llamados run_RunID.DXXX.sh, ocupan ~0.02 GB.
- Los archivos de resultados de las 50,000 estrellas, llamados RunID_StarID.txt, se almacenan en ***gridunam/tnos/Res01,*** los cuales ocupan ~20 GB.

El algoritmo DeClaTNOs y el programa maestro theMasterTNOS.sh son enviados utilizando la sintaxis de HT-Condor.

Se hicieron 2 experimentos. En el primero se enviaron los 150 lotes a la entidad DGTIC y en el segundo se enviaron 60 lotes a la entidad IAE. En ambos casos se utilizo el almacenamiento de Ensenada. La Table 1 muestra los resultados de los tiempos de ejecución, de DeClaTNOs, los tiempos de transporte de datos y el tiempo total del proceso completo.

Tabla 1: Tiempos de los lotes procesados en la entidad DGTIC e IAE.

	DGTIC	IAE
Numero de lotes enviados	150	60
Numero de lotes exitosos	140	60
Numero de estrellas procesadas	49140	21660
Tiempo de Ejecución máximo	4.8 hr	13.24 hr
Tiempo de Ejecución mínimo	2.3 hr	13.2 hr
Tiempo de Ejecución promedio	4.2 hr	13.2 hr
Desviación estándar del tipo de ejecución	0	0.019 hr
Tiempo de transporte de datos	2.8 hr (200 GB)	1.1 hr (80 GB)
Tiempo total promedio	**7 hr**	**14.3 hr**

Asimismo, la Figura 6 muestra el tiempo que tardó cada lote, tanto para su ejecución como en su transporte de datos, en ambas entidades. El círculo representa el tiempo de trasporte de datos. El área entre la curva de círculos y los diamantes es el tiempo de ejecución. Los diamantes son el tiempo total de proceso, en donde el tiempo es el tiempo real conocido como “wall time”.

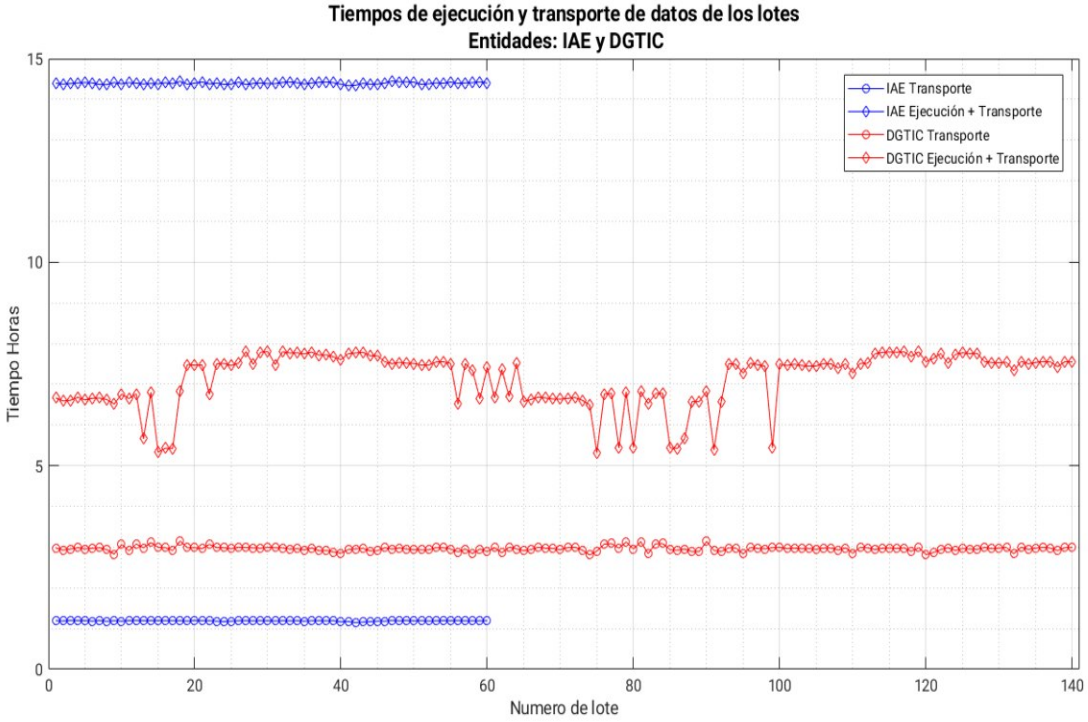

Figura 6: Tiempos de ejecución y transporte de DeClaTNOs ejecutados en las entidades DGTIC e IAE.

De la observación del comportamiento de la figura 6 se observa que:

- El tiempo de transporte de datos no impacta significativamente en el tiempo total de la ejecución de un algoritmo en el entorno Grid UNAM. En este experimento el almacenamiento está ubicado en el S3 de Ensenada y es el mismo para ambos experimentos. Esto se concluye observando la razón tiempo-transporte/Numero-lotes. Para IAE esta razón es *1.1/60=0.018* y para el caso de DGTIC es *2.8/140=0.02*, entonces el tiempo de trasporte de datos IAE $\simeq$ DGTIC. y no importa la distancia geográfica entre los nodos de ejecución y almacenamiento.
- En este mismo sentido, la Figura 6 muestra que los tiempos de transporte son constantes y sostenidos.
- En la entidad DGTIC, 10 de los 150 lotes no se ejecutaron, debido a que falló el transporte inicial de datos. En un entorno distribuido y heterogéneo, como es el Grid UNAM, es común este fenómeno. En la literatura especializada, se recomienda adicionar a los “scripts” de ejecución, la validación de los datos de entrada antes del momento de la ejecución. Si falla,

emitir un aviso al usuario para que reenvíe dicho trabajo.

- La variación de los tiempos de ejecución de diferentes lotes y entre las diferentes entidades, obedece básicamente a la inherente heterogeneidad del hardware disponible en el momento de la ejecución. Es evidente que los nodos de las entidad IAE y DGTIC son dispares, Opteron 6376 y Xeon Gold 6346 respectivamente, por tanto se esperaba ese comportamiento.

DeClaTNOs se ejecutó también en un servidor local (arrokoth) con 20 núcleos disponibles. El tiempo de ejecución total fué de 3.7 hr para 20 lotes. Es así que para ejecutar los 60 lotes reportados en la Entidad IAE se requieren de 11.1 hr. En el caso de la Entidad DGTIC, se requerirían de 26 hr. Es importante enfatizar que el tiempo de transporte en el servidor local no se contabiliza ya que los datos se encuentran en la misma máquina. En un ambiente de producción arrokoth es insuficiente para analizar una noche completa de observación.

Finalmente, se ejecutó con éxito 50,000 veces un algoritmo paramétrico serial que requiere de una entrada moderada de datos, del orden de 3 GB, en un entorno de recursos distribuidos, como es Grid UNAM y geográficamente distantes, en un tiempo suficiente para una fase de producción.

Conclusiones

Uno de los requerimientos básicos para la detección y clasificación de objetos trans-neptunianos, es el análisis de las 50,000 estrellas que se producirán diariamente en los telescopios del proyecto TAOS-II. La ejecución de nuestro algoritmo de detección es posible realizarla en un entorno de computación distribuida. Grid UNAM nos ofrece este vehículo para realizarla.

Para el caso de analizar las 50,000 estrellas en un servidor local con 20 núcleos, el proceso tardaría más de 24 hr. Ahora bien, dado el hecho de que el proyecto TAOS II capturará diariamente esta cantidad de estrellas, los algoritmos de detección deben analizarse en menos de 24 hr y este recurso local no es suficiente para este proceso. Entonces, el uso de Grid UNAM, es una alternativa viable para el proceso de detección y clasificación de objetos trans-Neptunianos, del proyecto TAOS II.

Referencias

1. https://taos2.astrosen.unam.mx/
2. https://taos2.asiaa.sinica.edu.tw/
3. B Hernández-Valencia, JH Castro-Chacón, M Reyes-Ruiz, MJ Lehner, CA Guerrero, JS Silva, JB Hernández-Águila, FI Alvarez-Santana, E Sánchez, JM Nuñez, LT Calvario-Velásquez, Liliana Figueroa, C-K Huang, Shiang-Yu Wang, C Alcock, W-P Chen, Agueda Paula Granados Contreras, JC Geary, KH Cook, JJ Kavelaars, T Norton, A Szentgyorgyi, W-L Yen, Z-W Zhang, G Olague (2022) “Pattern Recognition Using SVM for the Classification of the Size and Distance of Trans-Neptunian Objects Detected by Serendipitous Stellar Occultations”. https://doi.org/10.1088/1538-3873/ac7f5c