



Primer Taller Grid UNAM

aplicación serial paramétrica

Dr. Benjamín Hernández
Instituto de Astronomía Ensenada



Taller Grid UNAM

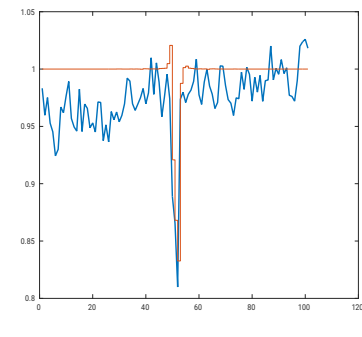
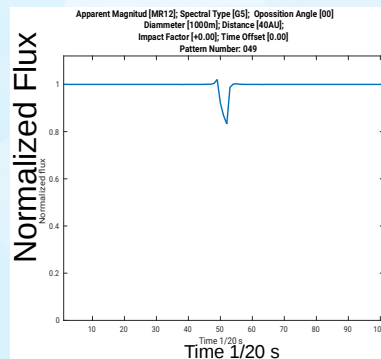
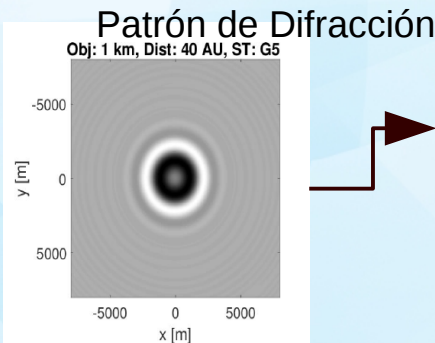
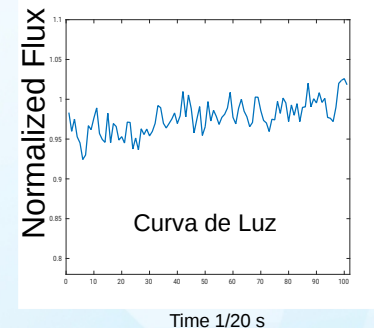
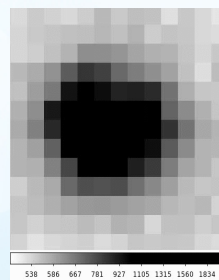
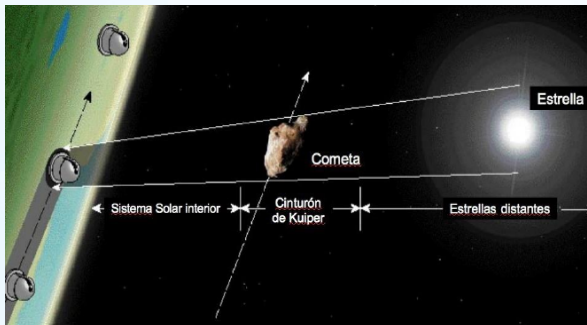
Parte 3.

1. Detección de objetos trans-neptunianos
2. Requerimientos técnicos de ejecución de la aplicación
3. Flujo conceptual de datos de Grid UNAM
4. Estrategias generales de envío

Detección y Clasificación de Objetos Trans-Neptunianos (DeClaTNOs)

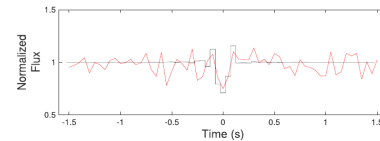
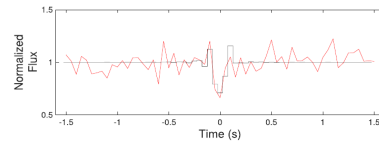
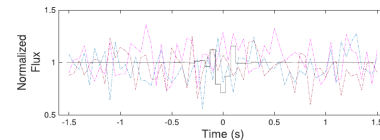
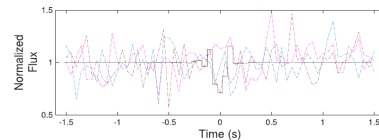
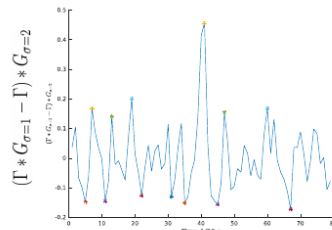
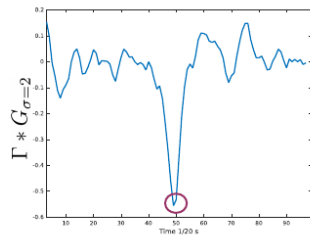
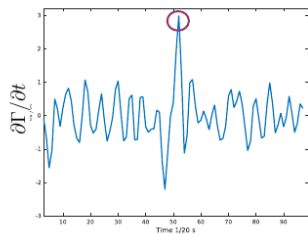
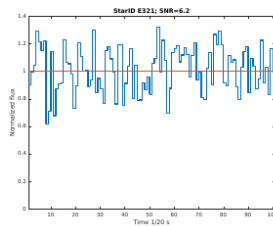
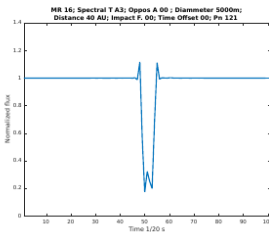
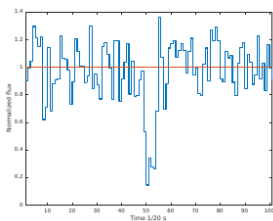
- ✓ Proyecto TAOS II
- ✓ 4,860 perfiles teóricos de los patrones de difracción de las las ocultación
- ✓ Extraen rasgos prominentes:
 - Estadísticas de 1er orden
 - Cambios en la energías de la señal
 - Análisis de espacios Gaussianos
 - Operadores de punto interés producto de un algoritmo evolutivo
- ✓ Clasificas se emplea la Máquina de Vector Soporte

Referencia
B Hernández et al (2022)
<https://doi.org/10.1088/1538-3873/ac7f5c>



DeClaTNOs

$$\Gamma_i = \gamma_i(\cdot) \oplus \eta(\cdot)$$



(a) True positive

(b) False negative

Table 1
Trans-Neptunian Diffraction Profiles Database (TPDB)

| Fields | Values |
|--|--------------------|
| Apparent visual magnitude (V) | 12,14,16 |
| Spectral type (S) | A3V, G5V, M4V |
| Angle from opposition (ϕ) | 0°, 30°, 60° |
| Total of bases | 27 |
| Object diameter (D) | [0.5,1,2,5,10] km |
| Object distance (z) | [30,40,50] au |
| Impact parameter (b) | 0, 0.25, 0.5, 0.75 |
| In fractions of D | |
| Time offset (t_{off}) | 0, 0.25, 0.5 |
| In fractions of 50 ms | |
| Total of patterns per base | 180 |
| Total of classes (y_i) | 4860 |

DeclaTNOs

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

DeClaTNOs

Compilación

```
g++ -g -o DeClaTNOs DeClaTNOs.cpp -L./ -lsvm -lhdf5_hl_cpp -lhdf5_cpp -lhdf5_hl -lhdf5 -lz -lm  
-std=c++0x -Wl,-rpath -Wl,${LIBDIR}/
```

```
export LD_LIBRARY_PATH=./:$LD_LIBRARY_PATH
```

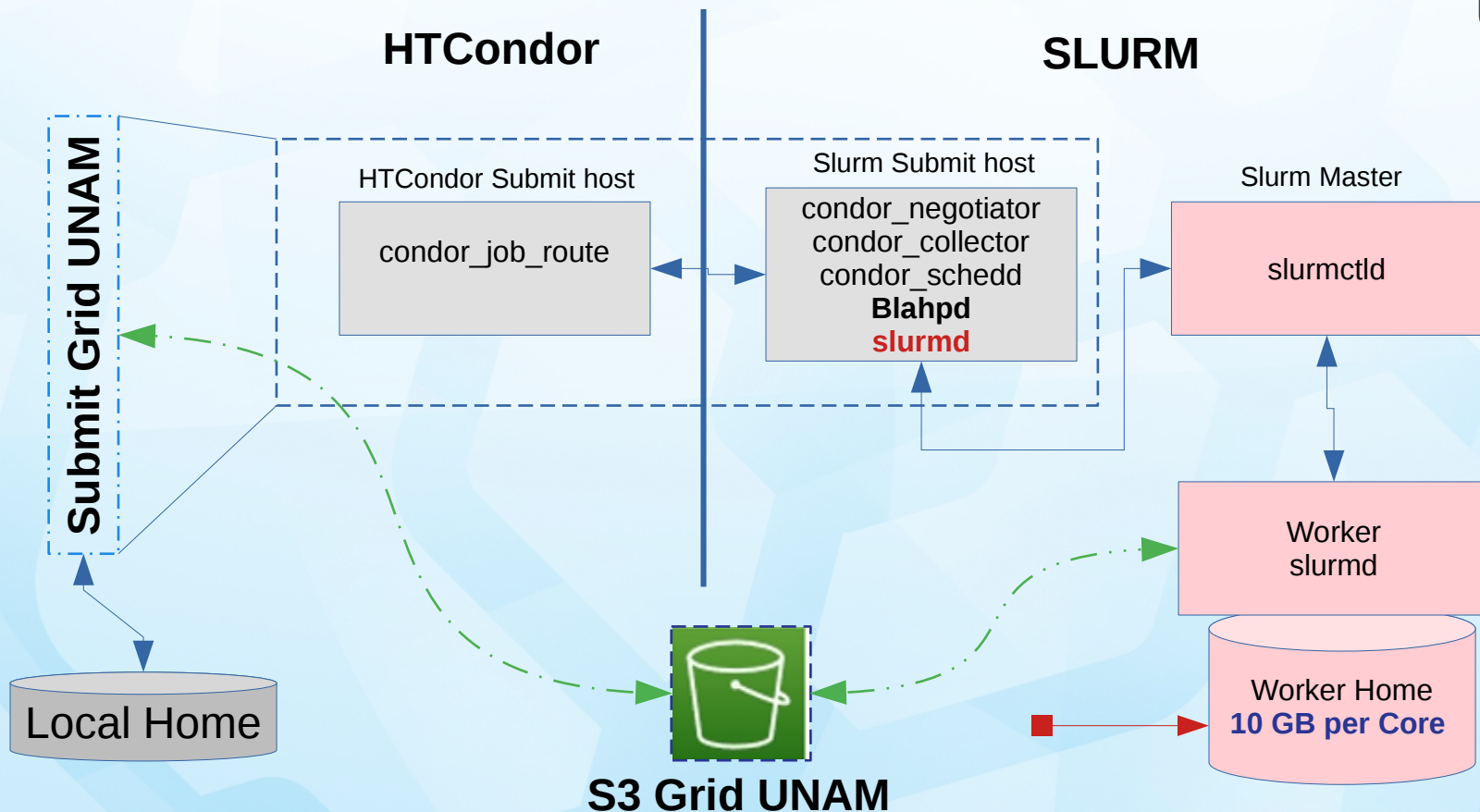
```
$ ./DeClaTNOs --help  
Usage: ./DeClaTNOs [option(s)] VALUES  
Options:  
  -h, --help                      Show this help message  
  -f, --filename HDF5_FILENAME    File Name  
  -A, --telA HDF5_Ligth_Curve     Telescope A (please include the group name)  
  -B, --telB HDF5_Ligth_Curve     Telescope 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, --opposition [0A00|0A30|0A60] 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/
```


- Revisar que su token esté vigente
- Copiar la aplicación DeClaTNOs que está en `gridunam/tallergusrc/bin`
- Copiar las librerías de la maquina de vector de soporte SMV.tar y descomprimirla
- Ejecutar DeClaTNOs sin parámetros

Flujo de Trabajo en Grid UNAM

HTCondor

SLURM



- 100 GB en el almacenamiento S3 por proyecto. Crece dependiendo de los requerimientos técnicos del proyecto.
- 2 GB de RAM por nodo.
- 10 GB de almacenamiento en disco en el espacio “Worker Home”
- 3 días de ejecución.

Recomendaciones:

- Utilizar el almacenamiento S3 de Grid UNAM para los archivos de datos y guardar resultados, ya sea parciales o totales.

Estrategía de envío (gridificación)

Con respecto al tiempo:

- Típicamente, un trabajo llega al nodo de trabajo, tiempo de preparación (TP) en un rango de 2 a 4 min
- Considerar los tiempos del transporte de datos (TT).
- Cuanto tiempo tarda la ejecución de una instancia de mi aplicación (T1A).

$$3d \geq N \times T1A \gg (TP + TT)$$

En el contexto de numero de lotes que debo construir:

- Mientras mas mejor
- Se debe diseñar “scrips” que permita una modificación dinámica para modificar los trabajos que se envían a las diferentes entidades

Estrategía de envío (gridificación)

Con respecto a la ejecución multihebra.

- Con openMP, se recomienda nunca exceder el numero de núcleos del “worker” con menos núcleos

Con respecto al almacenamiento:

- Siempre usar S3 Grid UNAM para datos de entrada y salida
- La aplicación no puede exceder 10 GB en el “Worker Home”, tanto para datos de salida como de entrada.
- Técnicas de “Check Point”. Parar la aplicación, trasferir a S3 y reiniciar en este “punto de quiebre”.

Sitio Grid UNAM

- <https://grid.unam.mx/>
- soporte-grid@unam.mx