

Pluralismo  
Compromiso  
Inclusión



**Licenciatura en Astronomía**

**Procesamiento de datos/data mining**



# Procesamiento de datos/data mining

I. Fundamentos de la astroinformática

II. Data warehouses y surveys; base de datos

III. Algoritmos de minería de datos

IV. Observatorio Virtual



# Introducción al lenguaje SQL y ADQL

# • Structured Query Language (SQL)



- Español: Lenguaje de consulta estructurado
- Lenguaje de sistemas de gestión de bases de datos relacionales (relational database management systems, RDBMS).
- ¿Qué es una base de datos relacional?  
Conjuntos tabulares de datos donde existen relaciones entre las tablas  
→ mantenimiento y eficiencia del almacenamiento.

# Structured Query Language (SQL)



- No es un lenguaje de programación (e.g., Python, IDL, Fortran)
- Limitaciones a solo lectura para acceder a datos almacenados en RDBMS.
- Define el conjunto de propiedades de los registros que se devolverán desde los datos almacenados



# Astronomical Data Query Language (ADQL)



- Variante del SQL desarrollado para grandes base de datos astronómicos
  - Se ajusta a las necesidades de los astrónomos
- Estandariza los recursos de consulta y amplía la funcionalidad de SQL para su uso en astronomía.
  - Necesidad de estandarización: diferencias en el uso del SQL (e.g., sintaxis).
- Recomendado desde 2005, ADQL 2.1 desde 2017
- Implementado en >100 servicios (e.g., SIMBAD, ESA-Archive, etc)



**Misión:** Determinar la posición de 2 billones de objetos  
→ El más grande y preciso mapa 3D de la Vía Láctea

**Lanzamiento:** 19 de diciembre del 2013

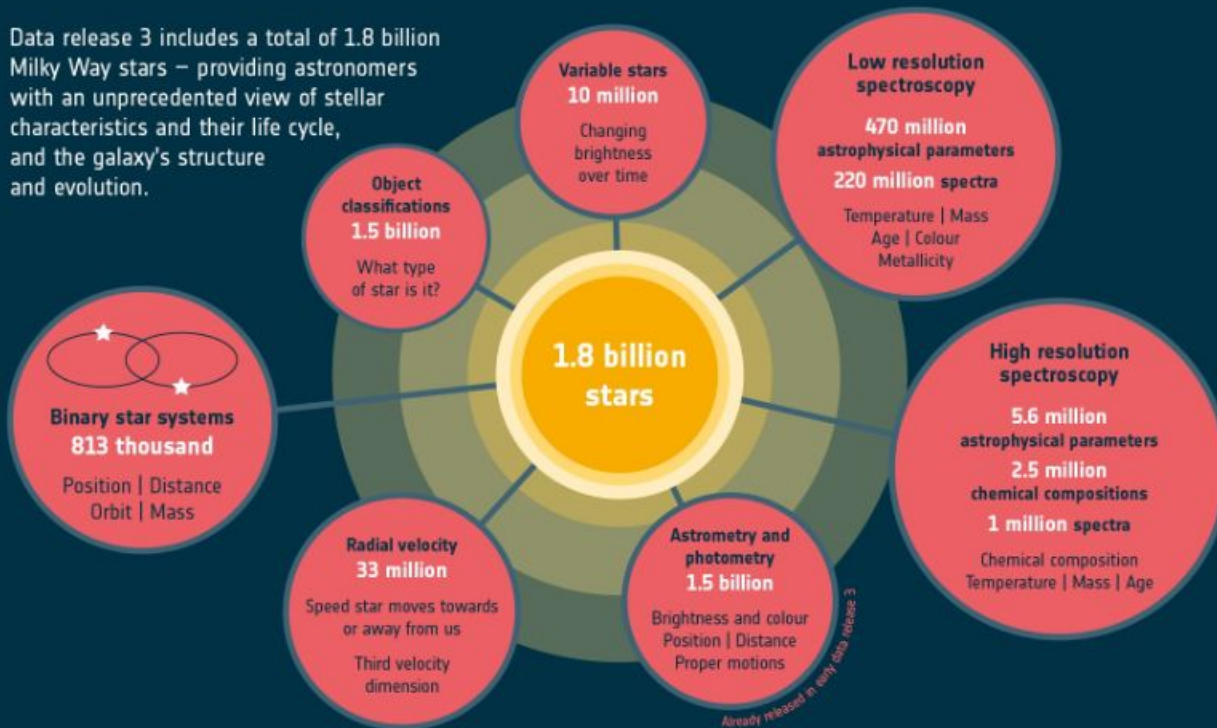




## MILKY WAY STARS



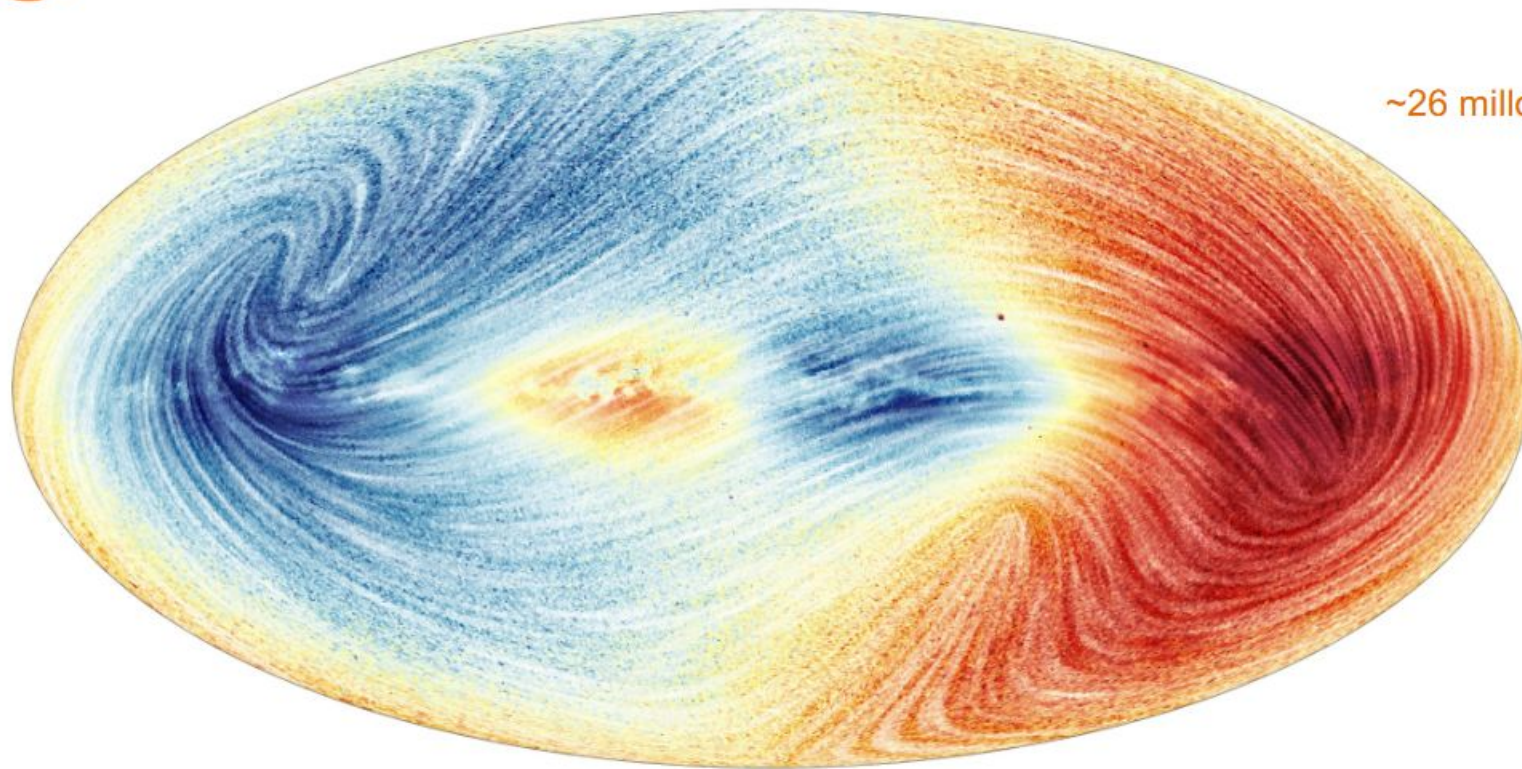
Data release 3 includes a total of 1.8 billion Milky Way stars – providing astronomers with an unprecedented view of stellar characteristics and their life cycle, and the galaxy's structure and evolution.







~26 millones de estrellas



# • **Publicación de datos de Gaia**



Primer lanzamiento : 14 de septiembre 2016

Segundo lanzamiento: 25 de abril 2018

Tercer lanzamiento :

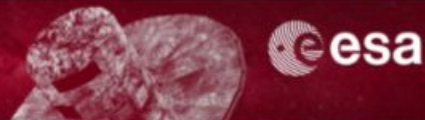
Gaia EDR3: 3 de diciembre 2020

Gaia DR3: 13 de Junio 2022

Gaia FPR (Focused Product Release): Q4 2023

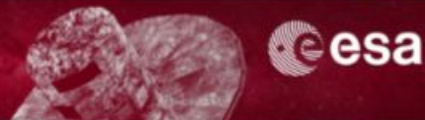
Gaia DR4 (basado en 66 meses de datos) no antes de el final del 2025

Detalles en: <https://www.cosmos.esa.int/web/gaia/release>



## Los datos son públicos!

Ir a: <https://gea.esac.esa.int/archive/>



## Los datos son públicos!

Ir a: <https://gea.esac.esa.int/archive/>

Links útiles para el uso del lenguaje ADQL

<https://www.gaia.ac.uk/data/gaia-data-release-1/adql-cookbook>

<https://www.cosmos.esa.int/web/gaia-users/archive/writing-queries>

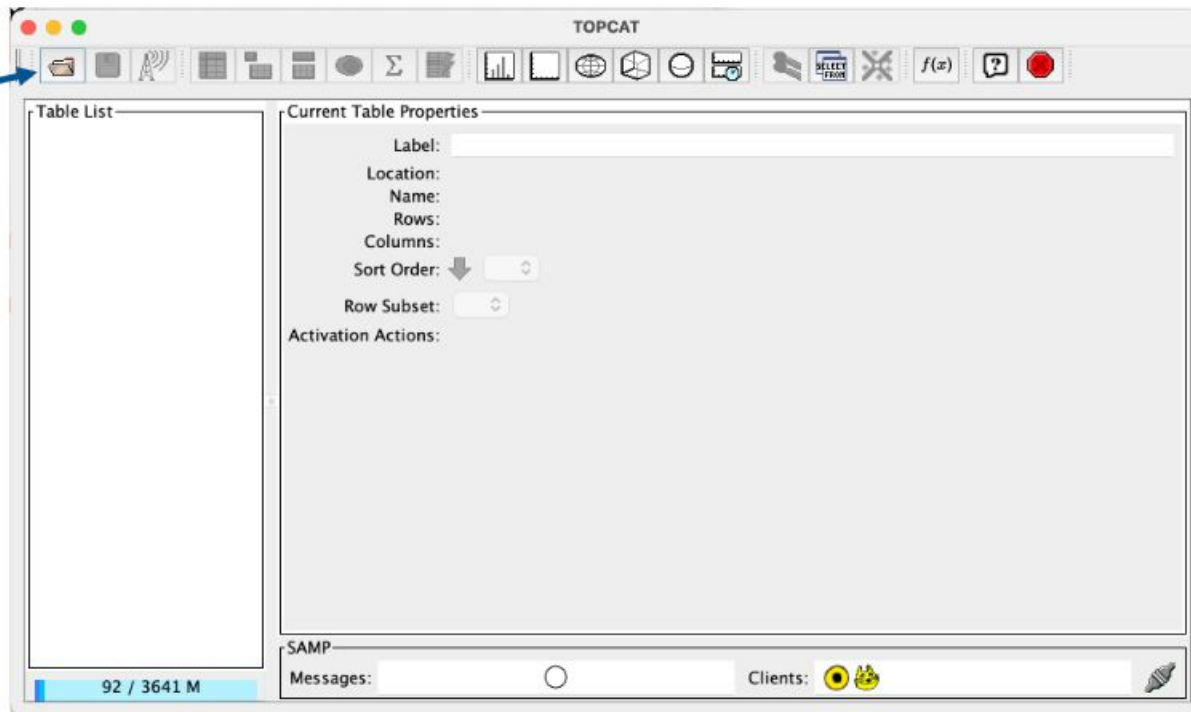


TOPCAT



## Tool for Operations on Catalogues And Tables

Abrir catálogo o buscar tablas





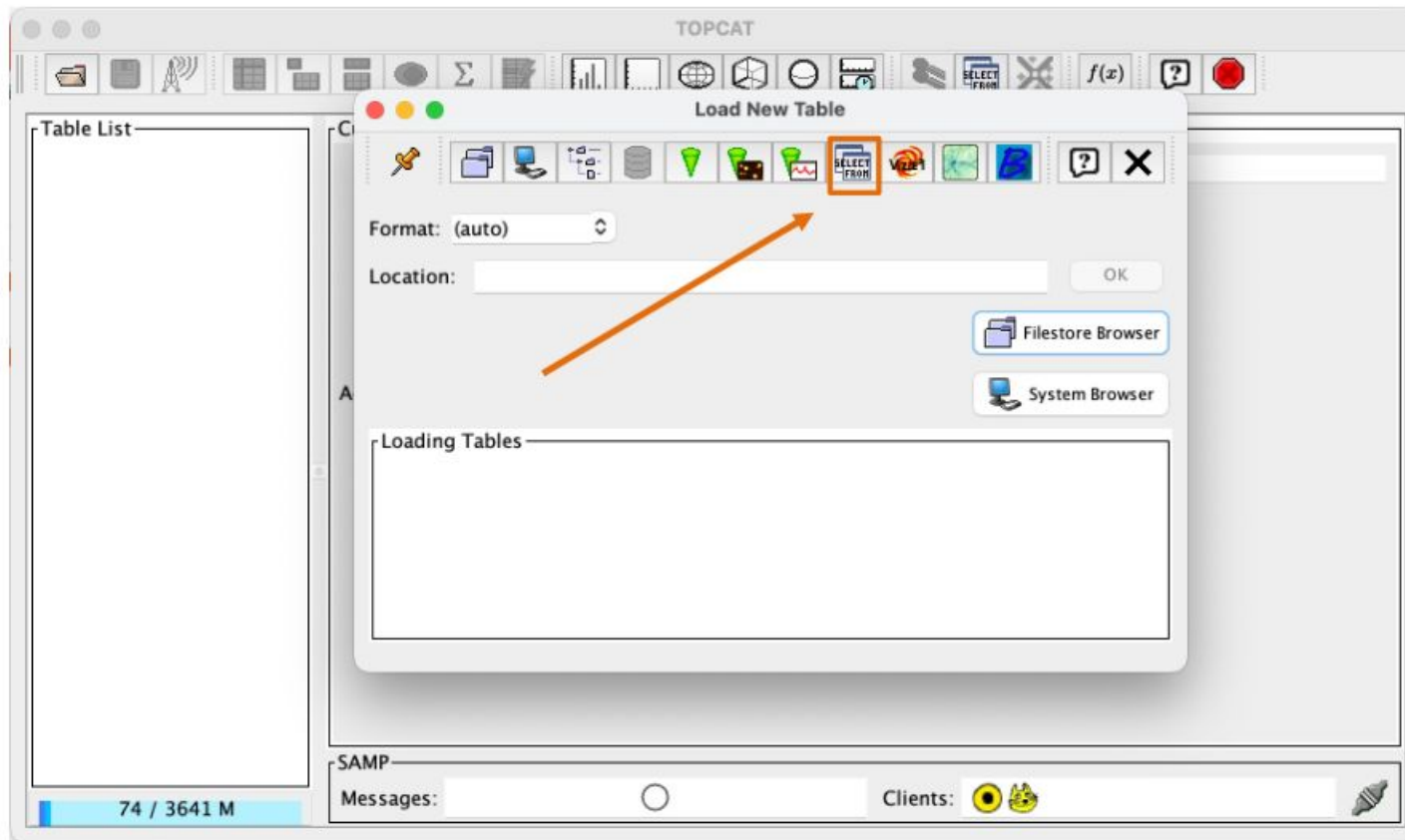




Table Access Protocol (TAP) Query

Select Service Use Service Resume Job Running Jobs

Locate TAP Service

By Table Properties By Service Properties

Keywords:  And

Match Fields: ☒ Table Name ☒ Table Description ☒ Service

Cancel Find Services

All TAP services (127)

- TAPVizieR (53081) - ivo://cds.vizier/tap
- HEASARC (1016) - ivo://nasa.heasarc/services/xamin
- IRSA TAP (927) - ivo://irsa.ipac/tap
- GAIA (230)** - ivo://esavo.gaia/tap
- GAVO DC TAP (225) - ivo://org.gavo.dc/tap
- ARI-Gaia (191) - ivo://uni-heidelberg.de/gaia/tap
- Gaia@AIP TAP Service (188) - ivo://gaia.aip.de/tap
- SDSS DR6 (129) - ivo://wfau.roe.ac.uk/sdssdr6-dsa
- ESASky Legacy (107) - ivo://esavo.esaskylegacy/tap
- ESASky (102) - ivo://esavo.esasky/tap
- ESO TAP\_CAT (100) - ivo://eso.org/tap\_cat
- APPLAUSE - Archives of Photographic Plates for Astronomical USE TAP Service (78) - ivo://www.p
- SkyMapper TAP (70) - ivo://ncl.org.au/skymapper/tap
- PS1DR2 TAP (64) - ivo://archive.stsci.edu/ps1dr2tap
- ATLAS DR1 - VST ATLAS Survey (63) - ivo://wfau.roe.ac.uk/atlasdr1-dsa
- YouCat (61) - ivo://cadn.nrc.ca/youcat

Selected TAP Service

TAP URL:

Use Service

Run Query



Table Access Protocol (TAP) Query

Select Service Use Service Resume Job Running Jobs

Metadata

Find:

☒ Name ☐ Descrip  Or

- GAIA (230)
  - external (17)
  - gaiadr1 (30)
  - gaiadr2 (41)
  - gaiadr3 (91)**
  - gaiadr3 (37)
  - job\_upload (0)
  - public (7)
  - tap\_config (2)
  - tap\_schema (5)
  - tap\_upload (0)

DR: data release

Name: gaiadr3  
Tables: 91  
Description: Gaia Data Release 3

Service Capabilities

Query Language: ADQL-2.1 Max Rows: 3000000 (default) Uploads: 100Mb

ADQL Text

Mode: Synchronous

Examples  Info

Run Query

Hacer búsqueda en ADQL



## El uso de SELECT:

```
SELECT [TOP límite] Lista_seleccionada  
FROM fromClause [WHERE condiciones] [ORDER BY columnas]
```

TOP límite: número entero para especificar cuantas filas se requieren.

ORDER BY columna: lista de nombres de columna

WHERE cláusula

Restricciones a especificar en la búsqueda, con operadores AND, OR, y NOT.

Delimitar una región (todo en grados)

```
1=CONTAINS(  
POINT('ICRS', RA, DEC),  
CIRCLE('ICRS', RAcir, DECcir, radio de búsqueda))
```





## Ejemplos de búsqueda ADQL – SELECT: ORDER BY



TAP 1 (*Table access protocol*, Protocolo de acceso a tabla)

- Ejemplo para seleccionar las estrellas más brillantes
- E.g.: 100 brightest stars in Gaia DR3

```
SELECT TOP 100 source_id, phot_g_mean_mag, parallax, bp_rp  
FROM gaiadr3.gaia_source  
ORDER BY phot_g_mean_mag
```



## Ejemplos de búsqueda ADQL – SELECT: WHERE



### TAP 2

- Condición: estrellas más brillantes que 12, y más cercanas que 50 pc.

```
SELECT TOP 2000 source_id, phot_g_mean_mag, parallax, bp_rp  
FROM gaiadr3.gaia_source  
WHERE phot_g_mean_mag < 15.0 AND parallax > 20.0
```



Universidad  
Central



**4 AÑOS ACREDITADA**  
GESTIÓN INSTITUCIONAL | DESDE DICIEMBRE 2017  
DOCENCIA DE PREGRADO | VINCULACIÓN CON EL MEDIO | HASTA DICIEMBRE 2021

[www.ucentral.cl](http://www.ucentral.cl)