

# Conversión de tablas a HDF5

Valeria Alvarez - HCAI 2023

# Contexto

## GALAXY ZOO



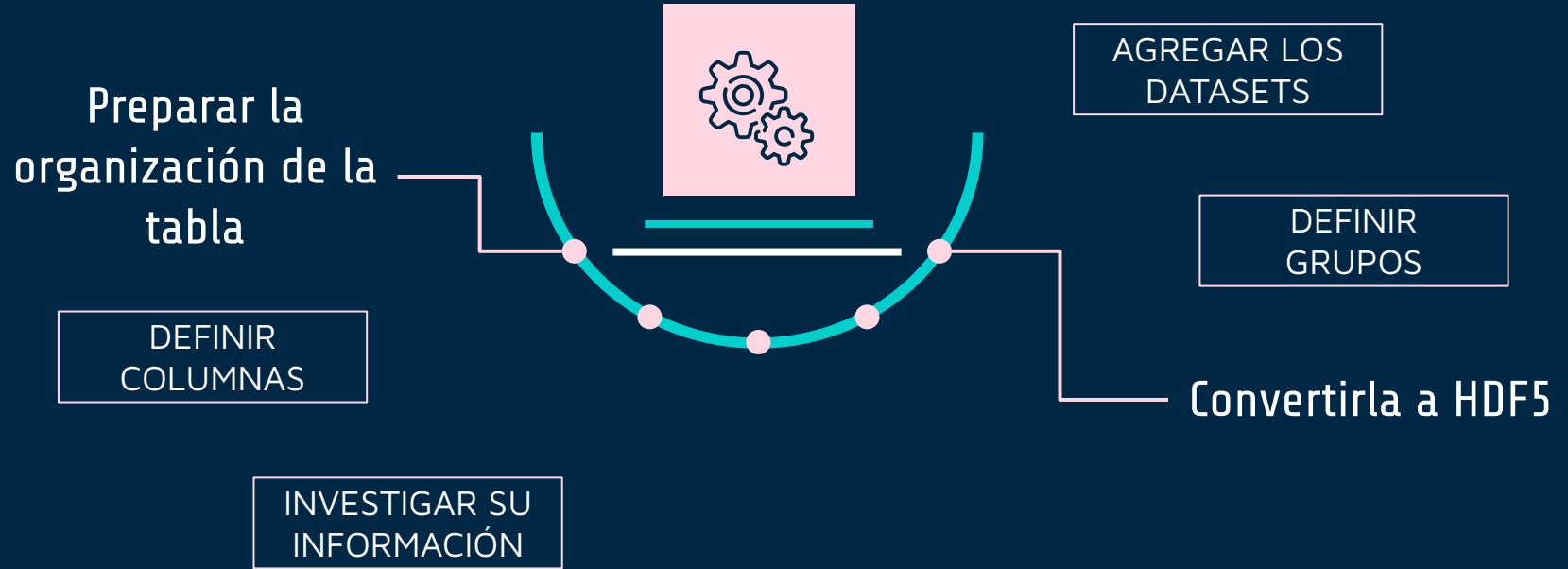
Proyecto en línea de que invita a voluntarios a ayudar a clasificar morfológicamente alrededor de un millón de galaxias.

## HDF5



Formato de archivo diseñado para almacenar y organizar grandes cantidades de datos de manera eficiente

# Objetivo



# ¿Por qué HDF5?



# Conversión de tablas

243.500  
234

Galaxy Zoo 2

MORPHOLOGICAL CLASSIFICATION  
OF MAIN-SAMPLE SPECTROSCOPIC  
GALAXIES FROM GALAXY ZOO 2

313.789  
91

Galaxy Zoo DECaLS

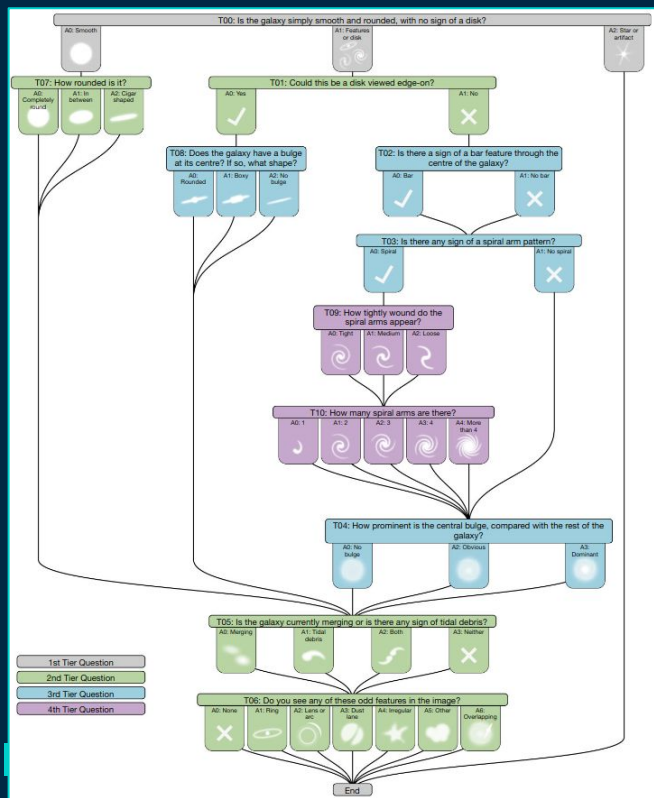
AUTOMATED MORPHOLOGICAL  
CLASSIFICATION WITH PREDICTED  
POSTERIORS

253.286  
127

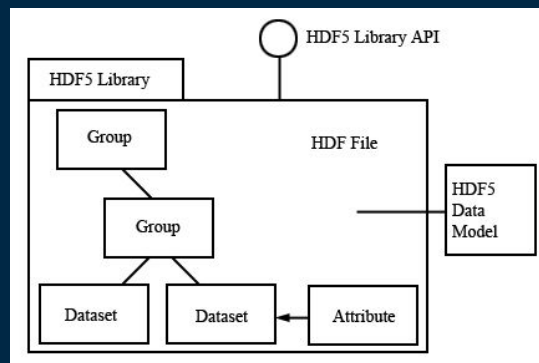
Galaxy Zoo DECaLS

VOLUNTEER CLASSIFICATIONS WITH  
THE IMPROVED DECISION TREE

# Organización



## Árbol de decisión del Galaxy Zoo



SOLUTION

LECTURA POR CAMPO  
REPRODUCIENDO LA  
ESTRUCTURA

# Codigo Python

```
with h5py.File("data.h5", "w") as archivo:

    #Atributos principales
    archivo.attrs["Creator"] = "Valeria Alvarez - HCAI2023"
    archivo.attrs["Fecha"] = "14-12-23"
    archivo.attrs["Tabla"] = "Galaxy Zoo DECaLS 2021 - Voluntarios"

    #Grupos
    merger = archivo.create_group("Merger")
    merger.attrs["Contenido"] = "Clasificación tipo de interacción"

    morfo = archivo.create_group("Morfología")
    morfo.attrs["Contenido"] = "Clasificación morfológica de voluntarios"

    prop = archivo.create_group("Propiedades Generales")
    prop.attrs["Contenido"] = "Propiedades generales básicas de las galaxias de la muestra"

    #Subgrupos
    sof = morfo.create_group("Brazos espirales")
    bar = morfo.create_group("Barras")
    disc = morfo.create_group("Featured")

    #Datasets
    minor_dataset = merger.create_dataset("Minor Merger", data=data["merging_minor-disturbance_debiased"])
    major_dataset = merger.create_dataset("Major Merger", data=data["merging_major-disturbance_debiased"])
    non_dataset = merger.create_dataset("Non Merger", data=data["merging_none_debiased"])
    merging_dataset = merger.create_dataset("Merging Merger", data=data["merging_merger_debiased"])

    sin_dataset = bar.create_dataset("Sin barra", data=data["bar_no_debiased"])
    debil_dataset = bar.create_dataset("Barra débil", data=data["bar_weak_debiased"])
    fuerte_dataset = bar.create_dataset("Barra fuerte", data=data["bar_strong_debiased"])

    si_dataset = sof.create_dataset("Tiene", data=data["has-spiral-arms_yes_debiased"])
    no_dataset = sof.create_dataset("No tiene", data=data["has-spiral-arms_no_debiased"])

    fod_dataset = disc.create_dataset("Featured o Disco", data=data["smooth-or-featured_featured-or-disk_debiased"])
    edgsi_dataset = disc.create_dataset("Disco edge on", data=data["disk-edge-on_yes_debiased"])
    edgno_dataset = disc.create_dataset("Disco face on", data=data["disk-edge-on_no_debiased"])

    ra_dataset = prop.create_dataset("Ra", data=data["ra"])
    dec_dataset = prop.create_dataset("Dec", data=data["dec"])
    z_dataset = prop.create_dataset("Redshift", data=data["redshift"])
```

01

Grupos: División de categorías

02

Subgrupos: Propiedades

03

Dataset: Columnas de datos

04

Atributos: Descripción

# Resultado

Name	Type	Array Size	Value[50](...)
Creator	String, Scalar		Valeria Alvarez - HCAI2023
Fecha	String, Scalar		14-12-23
Tabla	String, Scalar		Galaxy Zoo DECaLS 2021 - Voluntarios

Name	Type	Array Size	Value[50](...)
Creator	String, Scalar		Valeria Alvarez - HCAI2023
Fecha	String, Scalar		14-12-23
Tabla	String, Scalar		Galaxy Zoo 2 - Muestra completa

Name	Type	Array Size	Value[50](...)
Creator	String, Scalar		Valeria Alvarez - HCAI2023
Fecha	String, Scalar		14-12-23
Tabla	String, Scalar		Galaxy Zoo DECaLS 2021 - Clasificación automática



# Resultado

File Window Tools Help

Recent Files: /home/valeria/Astroinformatica/data.h5

data3.h5

- Fracciones
- Merger
  - Feature Disturbed
  - Merger
- Morfología
  - Barras
  - Brazos espirales
  - Disco
- Propiedades Generales
  - Dec
  - Object ID DR8
  - Ra

data.h5

- Fracciones de voluntarios
- Merger
  - Major Merger
  - Merging Merger
  - Minor Merger
  - Non Merger
- Morfología
  - Barras
  - Barra débil**
  - Barra fuerte
  - Sin barra
  - Brazos espirales
  - Featured
- Propiedades Generales

data2.h5

- Merger
- Morfología
- Propiedades Generales

Object Attribute Info General Object Info

Attribute Creation Order: Creation Order NOT Tracked

Number of attributes = 1

Add Attribute Delete Attribute

Name Value[50](...)

Barra La galaxia presenta una barra pequeña, y débil en relación a la galaxia. Puede parecer más ovalada

data.h5

- Fracciones de voluntarios
- Merger
- Morfología
  - Barras
  - Brazos espirales**
  - Featured
    - Disco edge on
    - Disco face on
    - Featured o Disco
- Propiedades Generales
  - Dec
  - Ra
  - Redshift

Object Attribute Info General Object Info

Name: Brazos espirales

Path: /Morfología/

Type: HDF5 Group

Object Ref: 8960

Link Creation Order: Creation Order NOT T

Group Members

Number of members: 2

Name	Type
No tiene	Dataset
Tiene	Dataset

# Errores

## Python code

Converter

H5py :

Thousands of  
datasets can be  
stored in a single file,  
categorized and  
tagged however you  
want.

<https://docs.h5py.org/en/stable/quick.html#quick>

Errors:

Read and write errors

Datatype

```
df1["gz2class"]
```

```
0      SBb?t  
1      Ser  
2      Sc+t  
3      SBc(r)  
4      Ser  
...  
243495    Sc2l  
243496      Sb  
243497    Ei(o)  
243498      Er  
243499      Sen
```

Name: gz2class, Length: 243500, dtype: object

```
0      [1.7009, 2.625, 2.2594, 1.9255, 1.9392, 1.2731...  
1      [2.3321, 2.3202, 2.1054, 2.9906, 3.0344, 2.576...  
2      [2.1725, 2.5291, 2.4078, 2.1751, 2.0782, 2.009...  
3      [2.0856, 1.8337, 1.7132, 1.9778, 2.0041, 1.666...  
4      [4.0723, 4.9291, 2.6826, 4.0653, 3.5179, 4.043...
```

```
313784    [3.8388, 3.8299, 3.1099, 2.9486, 3.1431, 2.569...  
313785    [1.6002, 1.9463, 1.9441, 1.7764, 2.469, 1.4337...  
313786    [2.2018, 1.8569, 1.7185, 2.6965, 2.1947, 1.988...  
313787    [1.4606, 1.8899, 1.7702, 1.5321, 1.6433, 1.518...  
313788    [2.7217, 3.9692, 2.5727, 2.6265, 3.7047, 1.835...
```

Name: bar\_strong\_concentration, Length: 313789, dtype: object

# Portafolio



**HCAI2023** Public

Pin Unwatch 1 Fork 0 Star 0

main 1 Branch 0 Tags

Add file Code

**About**

Documentación curso Herramientas computacionales para la astroinformática  
- Magister en Astronomía ULS 2023

Readme  
 Activity  
 0 stars  
 1 watching  
 0 forks

**Releases**  
No releases published  
[Create a new release](#)

**Packages**  
No packages published  
[Publish your first package](#)

**ValeeMarels** Unidad de simulaciones numéricas 691dd67 - last week 41 Commits

	Codigos	Creada Codigos para C	3 months ago
	CodigosPython	Codigos Python	3 months ago
	Datos	Unidad gestion de datos	last month
	DiseñoSoftware	Unidad de diseño de software	2 months ago
	MetodosNumericos	Unidad de Metodos numericos	3 weeks ago
	Profundizacion	Temas de profundizacion de clases	2 months ago
	Simulaciones	Unidad de simulaciones numéricas	last week
	20231003_120109.jpg	Mapa conceptual POO	2 months ago
	Presentacion Proyecto_Portafolio .pdf	Presentacion Proyecto	2 months ago
	README.md	README portafolios	last week

## README

## Herramientas computacionales para la astroinformatica

### Portafolio: HCAI- Magister en Astronomia 2023

#### -Clases

1. **Codigos**: Codigos en C y Python de ejemplos realizados en clases. Desde las primeras unidades del curso hasta Programacion Orientada a Objetos.
2. **DiseñoSoftware**: Apuntes y ejemplos relacionadas a Principios de Diseños de Softwares.
3. **Datos**: Apuntes y ejemplos relacionados a la unidad de gestion, representación y almacenamiento de datos.
4. **Simulaciones**: Apuntes unidad de simulaciones numéricas (Prof. F. Gomez).
5. **MetodosNumericos**: Apuntes y ejemplos de la unidad de métodos numericos.

#### -Tareas y trabajos

1. **Profundizacion**: Temas de profundizacion autonoma dados clase a clase.
2. **Presentación proyecto**: Presentación mid-term del proyecto relacionado a conversion de tablas a hdf5.
3. **Mapa conceptual**: Definiciones y relación de conceptos de Programación Orientada a Objetos.

