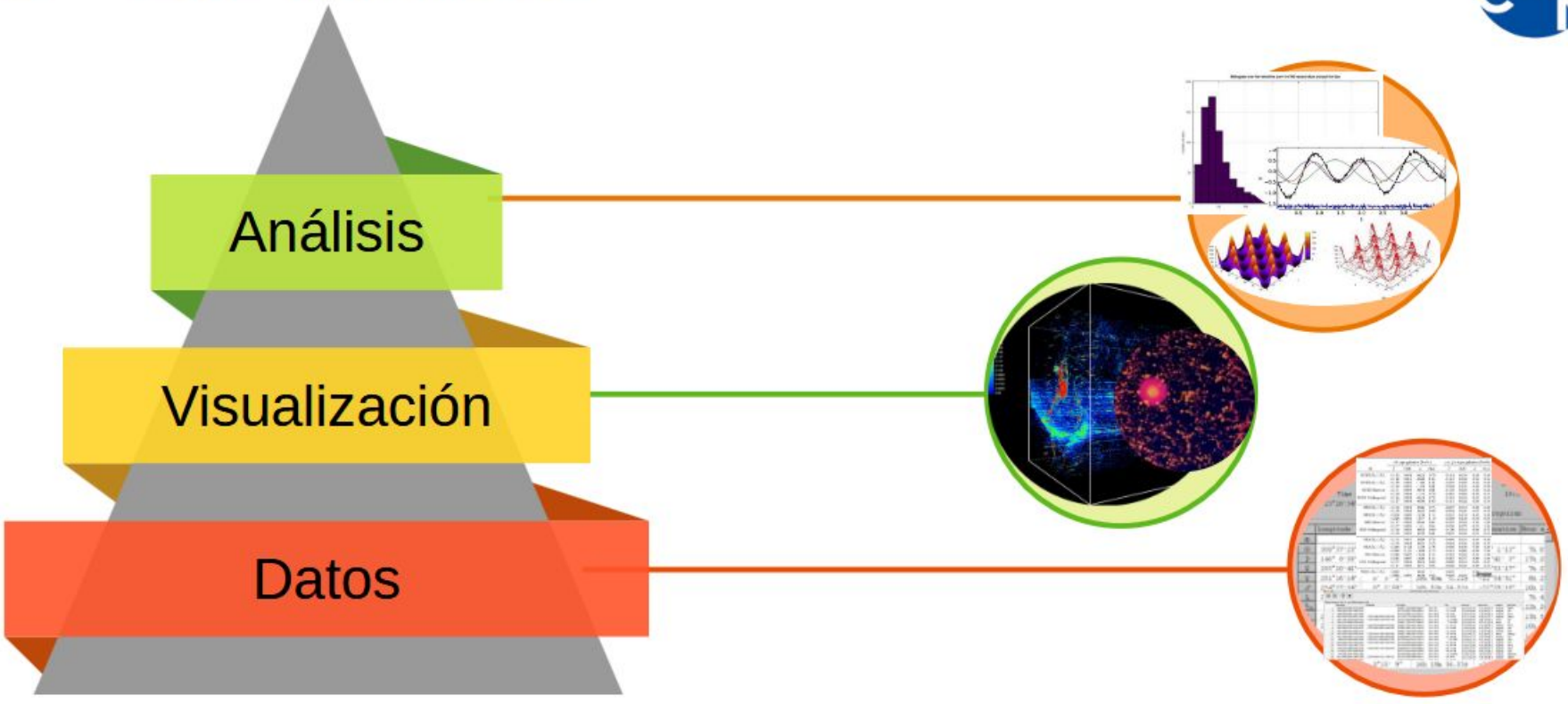
Pluralismo
Compromiso
Inclusión

Procesamiento de datos | **Licenciatura en Astronomía**

Data warehouses & surveys

Data Warehouse



Base de datos centralizada, integrada y consolidada.

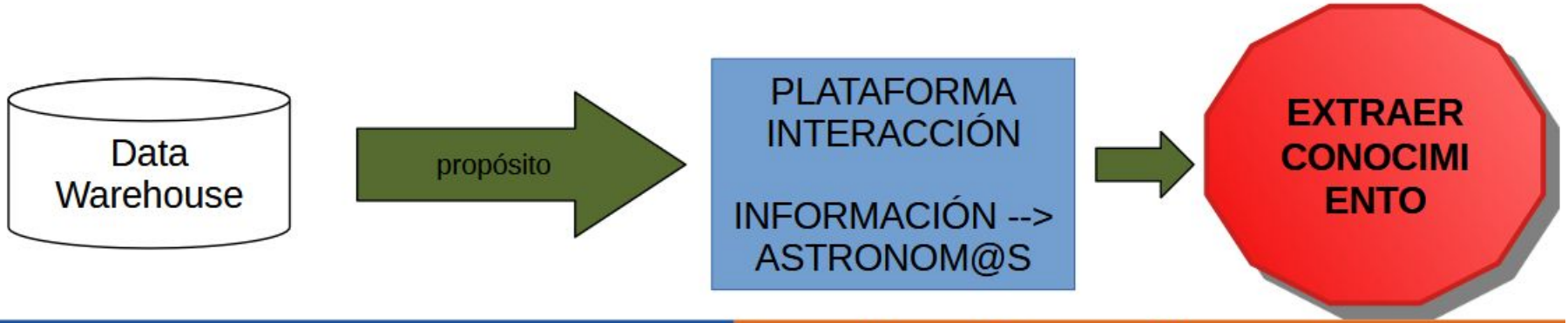
Data Warehouse en astronomía

- Diversos intentos en el pasado.

- Mayores desafíos:

(1) integración de datos provenientes de diferentes fuentes (datos heterogéneos).
(2) preprocesamiento de alto nivel
(3) herramientas de visualización.

Mismo problema tuvo que abordar la industria hace 60 años atras.



Data Warehouse en astronomía

Table 1. Big Data 3V characteristics in astronomical sky surveys.

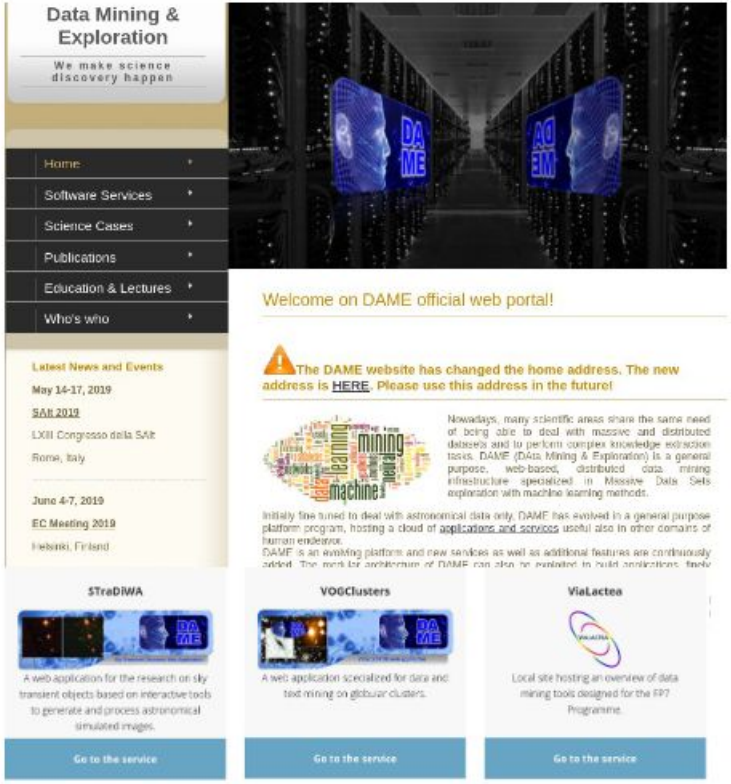
Sky Survey	Volume	Velocity	Variety
SDSS *Sloan Digital Sky Survey*	50 TB	200 GB per day	images, catalogs, redshits
GAIA	100 TB	40 GB per day	more then 100 parameters
Pan-STARRS *Panoramic Survey Telescope and Rapid Response System*	5 PB	5 TB per day	images, catalogs
LSST *Large Synoptic Survey Telescope*	60 PB	10 TB per day	images, catalogs
SKA *Square Kilometer Array*	3 ZB	150 TB per day	images, catalog, redshifts

Notes:
The column Volume refers to raw data produced at the end of the experiment.
Values regarding Pan-STARRS, LSST, and SKA surveys refer to expected Volume and Velocity values.

Garofalo et al. 2016

Problemática actual: Nuestra incapacidad de procesar correctamente toda esta información, ha llevado a un decrecimiento en la generación de nuevo conocimiento en astronomía, a pesar de que cada año el volumen de datos disponibles se duplica! (seguimos apilando datos).

Data Warehouses en astronomía: iniciativas



DAME (DAta mining and exploration):
http://dame.oacn.inaf.it/

Centro de datos CDS:
https://cdsweb.u-strasbg.fr/index-fr.gml
SIMBAD: identificación y bibliografía.
VizieR: catálogos y tablas astronómicas.
Aladin: software de visualización y análisis.

Data Warehouses en astronomía: iniciativas

SPLAT-VO:
http://star-www.dur.ac.uk/~pdraper/splat/splat.html
Herramienta de análisis espectral que puede bajar y analizar espectros desde internet y otras bases de datos.

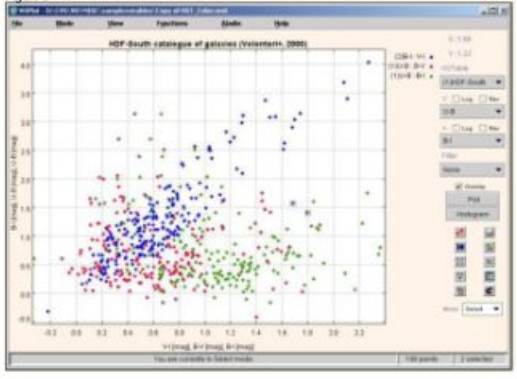
VOPlot – The VOTable plotting utility (Version 1.2.1)

Introduction

VOPlot (VOTable Plotting tool) is an applet for plotting different astronomical graphs using data stored in VOTable format. VOPlot is available in desktop version and a web-based version. The web-based version is integrated with the VizieR Catalogue Service and can be used to plot any catalogue by selecting output layout as "Plot (VOPlot)". The desktop version can be downloaded from http://vo.iucaa.ernet.in/~voi/voplot.htm.

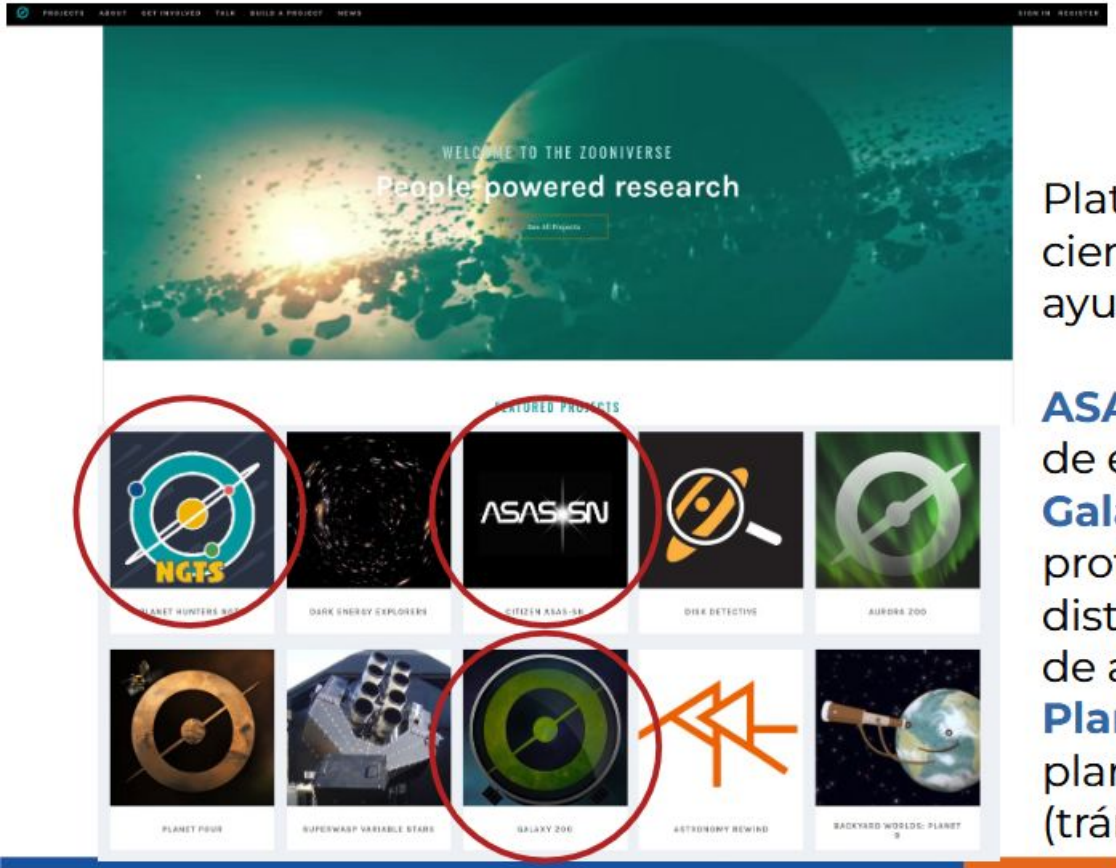
A sample plot with two VOTables loaded is shown in figure 5 below.

Figure 5



http://cdsarc.u-strasbg.fr/vizier/VOPlot/#_Introduction

Data Warehouses en astronomía: iniciativas



https://www.zooniverse.org/

Plataforma que reúne diversos proyectos científicos y que permite que voluntarios ayuden en el análisis de los datos.

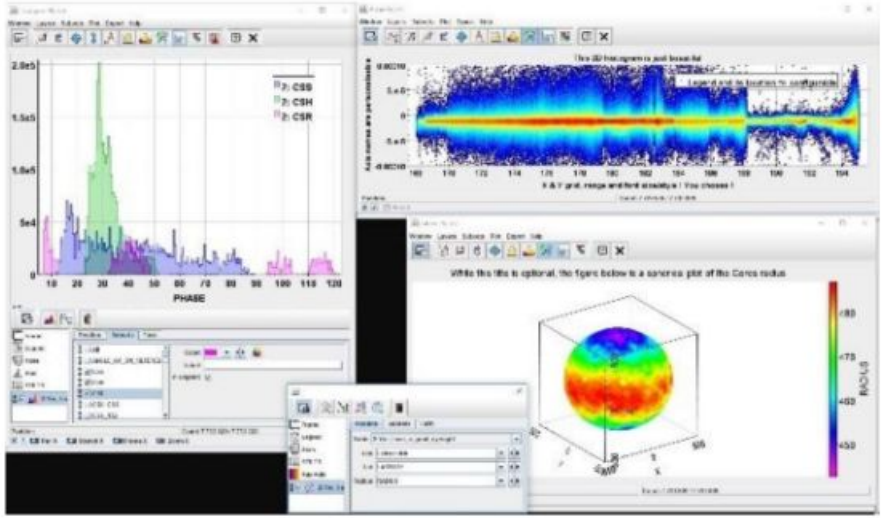
ASAS-SN: Clasificación de curvas de luz de estrellas variables inusuales.
Galaxy-Zoo: Imágenes de campo profundo de diversas galaxias desde distintos catálogos. Se deben clasificar de acuerdo a su morfología.
Planet Hunters: Descubrir nuevos planetas usando datos de surveys (tránsito).

Data Warehouses en astronomía: iniciativas

TOPCAT

What is TOPCAT?

TOPCAT is an interactive graphical viewer and editor for tabular data. Its aim is to provide most of the facilities that astronomers need for analysis and manipulation of source catalogues and other tables, though it can be used for non-astronomical data as well. It understands a number of different astronomically important formats (including FITS, VOTable and CDF) and more formats can be added. It is especially good at *interactive exploration of large (multi-million row, lots of columns) tables.*



Provee diversas herramientas para manipular y analizar grandes volúmenes de datos provenientes de catálogos y tablas.

Diversos tipos de gráficos

Parámetros estadísticos (media, SD, varianza, cuantíles, etc..)

Cross-matching (datos en Vizier o SIMBAD).

Observatorios Virtuales (VO)

Table 2.1. Virtual Observatories Worldwide

Name	Internet Address	Main Focus
International Virtual Observatory Alliance (IVOA)[1]	http://www.ivoa.net/	#Collaboration #Prototyping #Developing Standards #Tools and Services #VOS
Nuevo Observatorio Virtual Argentino (NOVA)	http://nova.iafe.uba.ar/	#Images #Data Center #Visualization #Statistics
Armenia Virtual Observatory (ArVO)	https://www.aras.am/Arvo/main.htm	#Spectroscopic #Photographic #DFBS[2]
AstroGrid	http://www.astrogrid.org/	#Images #Search
Australian Virtual Observatory	http://www.asvo.org.au/	#Theoretical #Queryable Data[3] #SkyMapper[4]
Brazilian Virtual Observatory (BRAVO)	http://bravo.iag.usp.br/	#Promotion
Chinese Virtual Observatory (China-VO)	http://www.china-vo.org/	#Spectra Data[5] #Data Center #Search #Data Integration
Canadian Virtual Observatory (CVO)	http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/cvo/	#Collaboration #Images #Data Center[6]
Chile Virtual Observatory (CHIVO)	http://www.chivo.cl/	#Data Center[7] #FITS search #Jupyter Notebooks[8]
The European Virtual Observatory EURO-VO	http://www.euro-vo.org/	#Education #Images #Spectra (Search and Visualization)
France Virtual Observatory	http://www.france-vo.org/	#Promotion
German Astrophysical Virtual Observatory (GAVO)	http://www.g-vo.org/	#Data Center #Teaching #Services[9]
Hungarian Virtual Observatory (HVO)	http://hvo.elte.hu/	#Collaboration
India Virtual Observatory	http://voi.iucaa.in/voi/	#Morphological Parameters of Galaxy Images #Visualization #Statistical #FITS
Japanese Virtual Observatory (JVO)	http://jvo.nao.ac.jp/	#Promotion
Korean Virtual Observatory	http://kvo.kasi.re.kr/	#Promotion
Russian Virtual Observatory (RVO)	http://www.inasan.ru/eng/rvo/	#Star Evolution #Variable Stars #Experimental Astronomy #Espectro-UV[10]
South African Virtual Observatory	http://www.saao.ac.za/	#Multi-wavelength #Promotion
Spanish Virtual Observatory	http://svo.cab.inta-csic.es/main/index.php	#Data Center - Spectra Photometric and Theoretical[11] #Services #Data Mining (Classification)
Italian Virtual Observatory	http://vobs.astro.it/	#Promotion
Ukrainian Virtual Observatory	http://www.ukr-vo.org/	#Data Center #Services
United States Virtual Observatory Alliance (USVOA)	http://usvoa.cfa.harvard.edu/	#Collaboration #Promotion #Standards #Tools and Services[12]

MODELOS PREDICTIVOS? NUEVO CONOCIMIENTO? EXTRAER

Machin, J. 2017

Spanish Virtual Observatory

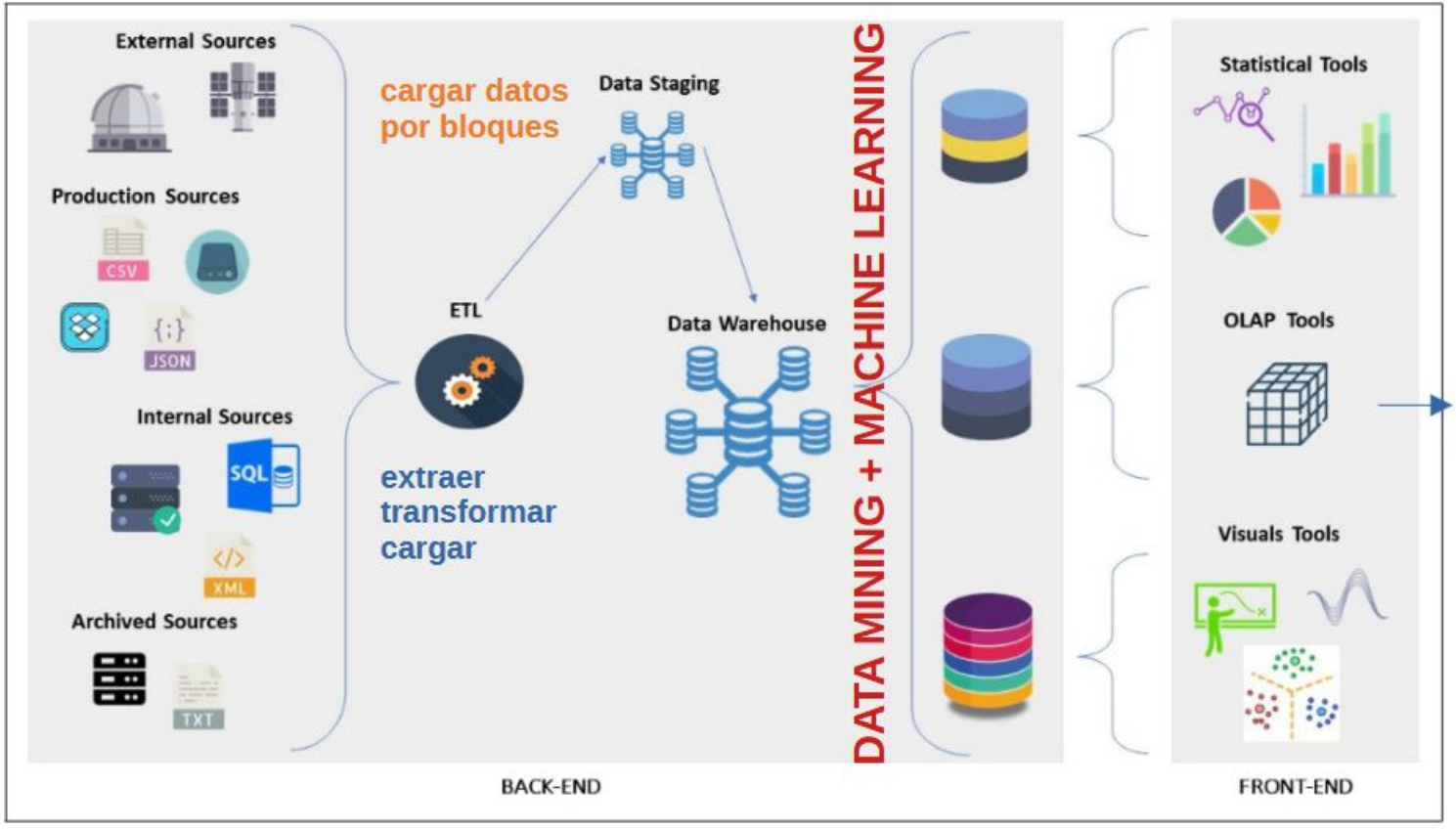
Data Warehouses en astronomía: desafíos

- Integración: Datos 'dispersos' en distintas plataformas.
- Actualización: muchas de las actuales iniciativas no se encuentran actualizadas.
- Plataformas 'poco amigables': interfaces poco intuitivas.
- Falta de herramientas de visualización
- Formato de datos restringido: requieren de habilidades de programación y algoritmos complejos para resolver algunos problemas específicos.
- Librerías de terceras fuentes: requeridas en algunos casos (estas pueden estar descontinuadas o no tienen soporte continuo).

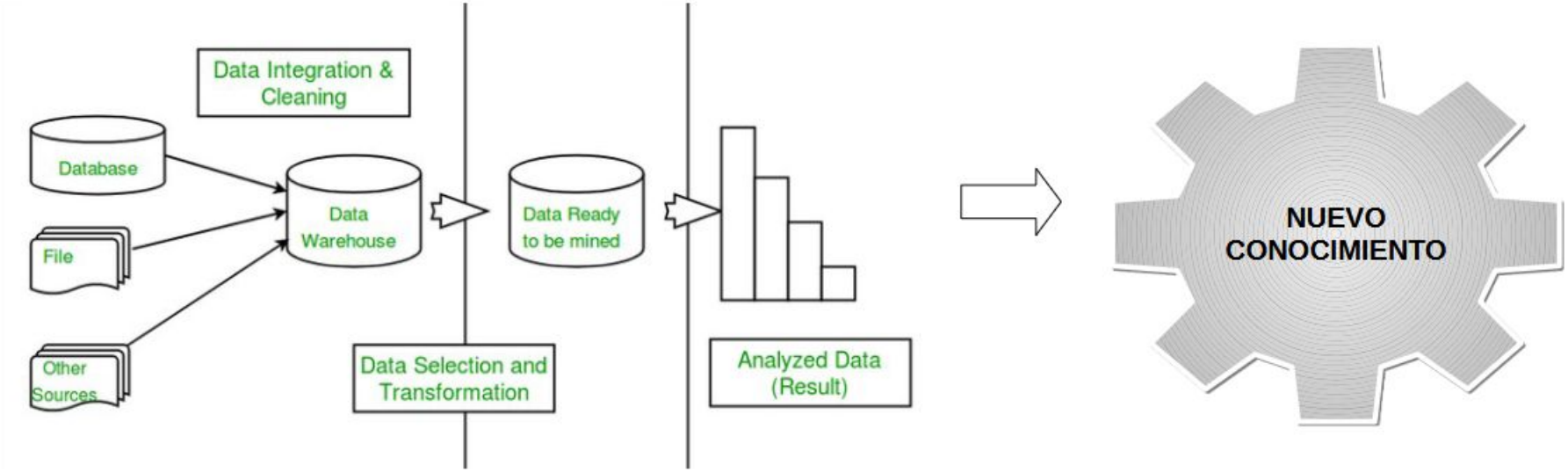
	Aladin	Simbad	Vizer	TopCat	Astropy	VOStat	SkyView	AstroML
Query Search	✓	✓	✓	✓	✓	×	✓	✓
Visualization	✓	×	×	✓	✓	✓	✓	✓
Data Integration	✓	✓	✓	✓	×	×	×	×
Machine Learning	×	×	×	×	✓	✓	×	✓
Data Sets Editions	✓	✓	✓	✓	✓	×	×	×
Features Extraction	×	×	×	×	×	×	×	×
Multi Catalog Feature Search	×	×	×	×	×	×	×	×

Principales diferencias entre algunos software de astronomía (Machin, 2017).

Data Warehouse: ejemplo de arquitectura



Warehouses & Data Mining



Data Mining:

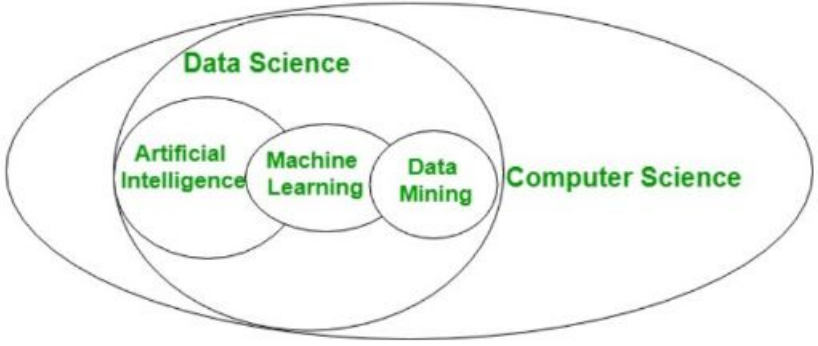
Extraer información útil desde grandes bases de datos y descubrir tendencias y patrones usando técnicas de clustering, clasificación y algoritmos de regresión.

Warehouses & Machine-learning

Machine Learning:

Librerias de algoritmos que permiten al computador ir 'aprendiendo' patrones ó tendencias en al análisis de grandes volúmenes de datos.
En el Warehouse trabaja en conjunto con el data mining.

Propósito: Hacer predicciones respecto a nuevos conjuntos de datos.



Estamos
comprometidos con la
Región de Coquimbo

Comisión Nacional de Acreditación CNA-Chile | 4 AÑOS ACREDITADA
GESTIÓN INSTITUCIONAL
DOCENCIA DE PREGRADO
VINCULACIÓN CON EL MEDIO
DESDE DICIEMBRE 2017
HASTA DICIEMBRE 2021

www.ucentral.cl