JVO portal : VO data search using the cache of VO crawler and Gaia data viewer

Yuji Shirasaki, Christopher Zapart, Masatoshi Ohishi, Yoshihiko Mizumoto (NAOJ)

We present two new features of JVO portal. (1) VO crawler is a crawling system which automatically sends queries to the selected VO services and caches the retrieved data. In order to retrieve all the data (or metadata of image and spectrum) provided by the VO services, it performs queries by changing the region of interest step by step. We registered those data in a distributed database system and made them searchable with a unified interface through the JVO portal. (2) Gaia data viewer is a data visualization service for Gaia source catalog. Gaia source catalog contains more than one billions of records. Due to its large volume, it is hard or time consuming to make a plot to see the distribution of the source parameters satisfying given criteria. By utilizing the distributed database system we provide a quick (quicker than reading from a single HDD) visualization service for Gaia source catalog.

1. VO crawler cache DB

- Most of the existing VO services does not support the search by criteria others than a position.
  - Users may want to search the data by the telescope name, instrument name, or wavelength.
- Queries to the multiple services takes several minutes to perform if they are many.
  - It is inefficient to know what data is available at multiple positions or at a wide area of the sky.
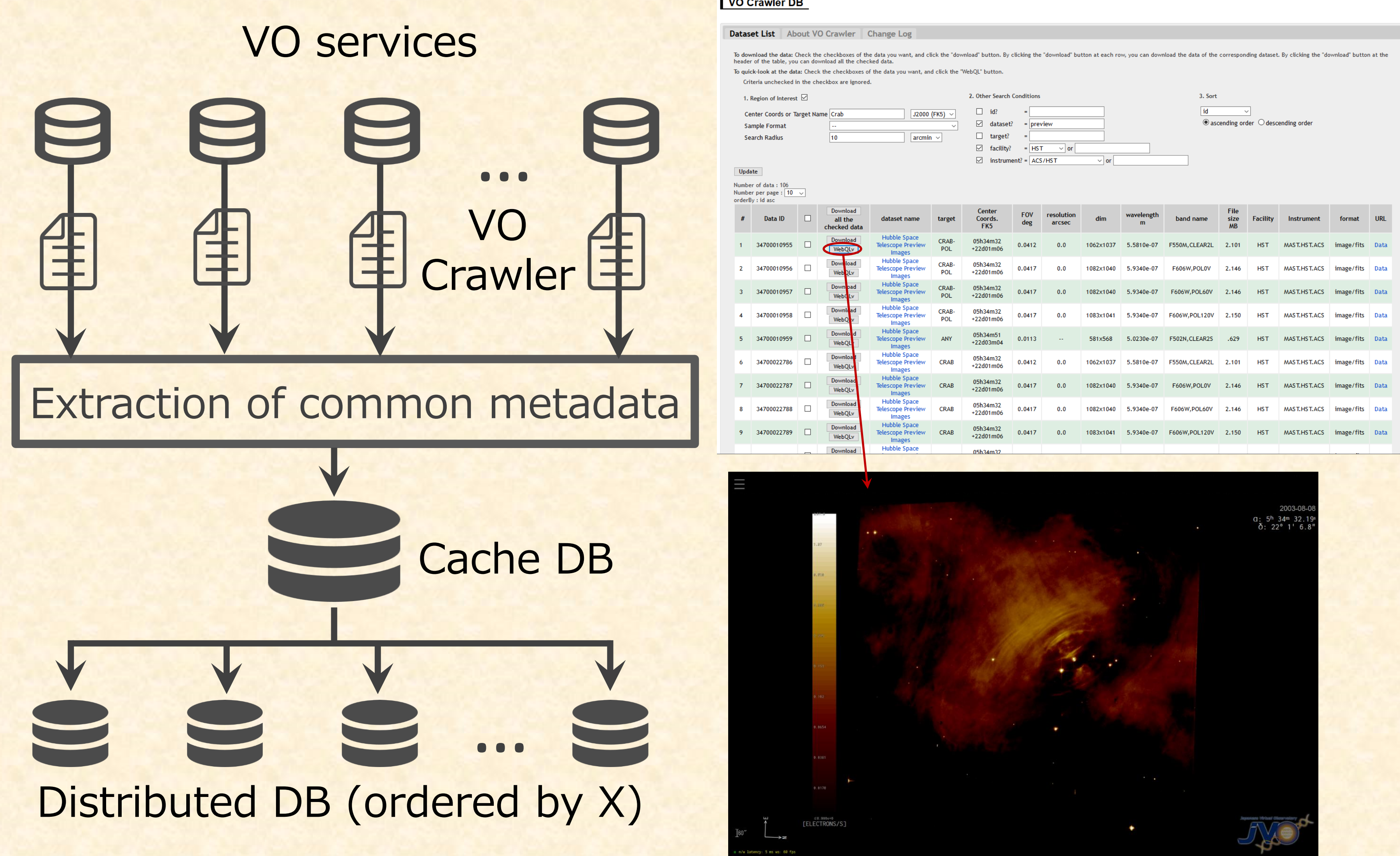- One of the solutions is to create a DB which contains all metadata provided at the VO services.

http://jvo.nao.ac.jp/portal/voc.do

Figure 1. Schematic diagram of the system of VO crawler cache DB and snapshot image of the search interface.

column name	description	column name	description
id	data id	s_xel2	number of y pixel
cid	crawling id	s_resolution	spatial pixel resolution
data_type	service type (SIA|SSA|Cone|TAP)	em_min	lower λ range
obs_collection	title of service	em_max	upper λ range
access_url	data URL	em_mid	middle λ
access_format	data format	em_band_name	band name
access_estsize	data size	em_xel	number of λ pixel
target_name	object name	facility_name	facility name
s_ra	RA of center	instrument_name	instrument name
s_dec	Dec of center	_point	center coordinate
s_fov	size of FOV	_poly	polygon of the data coverage
s_xel1	number of x pixel		

Table 1. Schema of the VO crawler cache DB

2. Gaia source catalog viewer

- NAOJ is one of the affiliate data centers of Gaia. (for Gaia-JASMIN collaboration)
- Gaia source catalog is a big table (> 1billions records) and users tend to be interested in the data of large part of the sky. (to investigate kinematics of the Galactic disk and so on)
- It is too time consuming to download such a lot of data. Whole catalog amounts to 556 GB in gzipped format.
- It is preferable to create a plot at server side with users' specified criteria rather than download the large data. The size of the plot can be very small.
- To speed up for generating the plot, we utilized distributed database system, which is already used for data search of Gaia data at JVO portal.
- Time for creating a plot from all the record is about 10 minutes with 6 servers. We have a plan to increase the multiplicity of the DBs to further speed up the process.
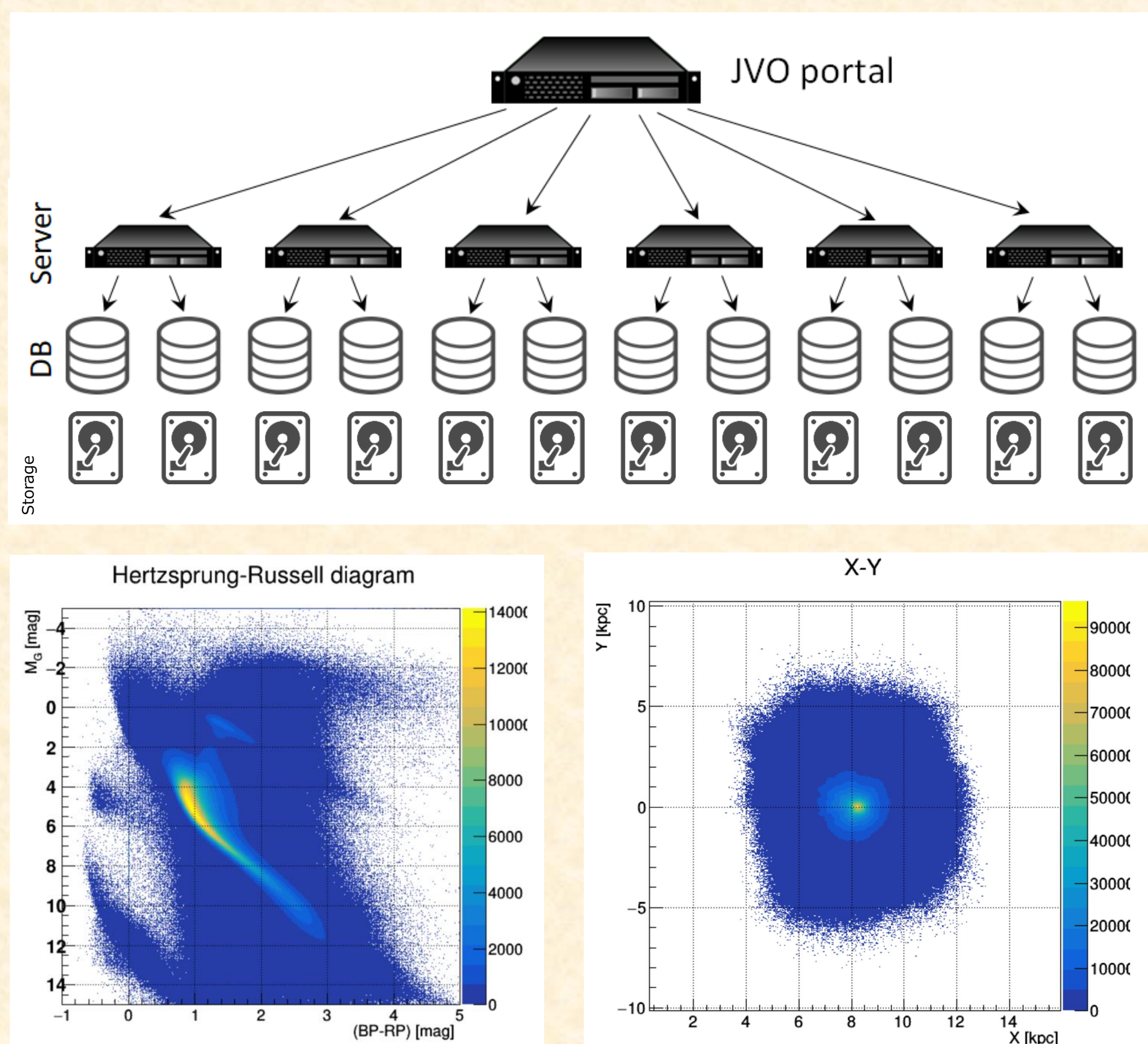

Figure 2. Schematic diagram of the Gaia data viewer system and examples of the plot generated with this system (Left: Color-Mag diagram, Right: position distribution in galactocentric coordinate)