

Introduction to simulation databases with ADQL and Topcat

Kristin Riebe*, GAVO

July 05, 2016

Introduction

Simulation databases like the Millennium Database or CosmoSim contain data sets from cosmological simulations. Each user has access to catalogues of dark matter halos and their history (merger trees), supplemented by (mock) galaxy catalogues and information about the density field or tables with the simulation particles themselves and other data sets. Access to these data is possible via a web interface directly with the browser; for larger data it is more convenient to access the data via [TOPCAT](#), wget and using their TAP/UWS interface, if available. This tutorial shows a few example use cases for extracting the mass function of a simulation, particles for a halo, the mass growth history and progenitors of a halo as well as a slice of the density field.

This tutorial version will use AIP's **DaCHS** instance, which contains only a small part of the data for testing and demonstration purposes. The full data is released at the [CosmoSim^a](#) webpage, tutorials for this web interface can be found at the GAVO tutorials webpage^b; some more demos, examples and tutorials are available at the CosmoSim webpage.

^a<https://www.cosmosim.org/>

^b<http://www.g-vo.org/pmwiki/Documents/Tutorials>

Software: [TOPCAT](#)

1 A first TAP query

- Start [TOPCAT](#), e.g. using `java -jar topcat-*.jar` from the command line in Linux.

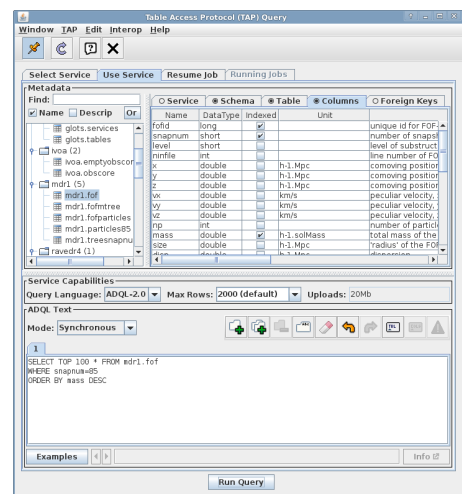
- Go to `VO` → **Table Access Protocol (TAP) query** 

- Write following TAP URL in the bottom line:

`http://gavo.aip.de/tap`

and press **Use Service**. Topcat will switch automatically to the next tab and load all the table metadata.

- For this tutorial, we will only use data from the MDR1 and Bolshoi simulation, located in tables `mdr1` and `bolshoi`. Click on one of their tables (e.g. `mdr1.fof`) in the left tree view and browse the table metadata and columns on the right side. Note, that indexed columns are also marked, which is useful to know since queries on indexed columns will be fast.




- In the bottom text area (**ADQL Text**) enter following ADQL query:

*Leibniz-Institute for Astrophysics Potsdam (AIP); email: kriebe@aip.de

```
SELECT TOP 100 * FROM mdr1.fof
WHERE snapnum=85
ORDER BY mass DESC
```

This query will retrieve the 100 most massive clusters (ordered by mass in descending order) for redshift 0 (= snapshot number 85) from the MDR1-simulation.

- Click the **Run Query** button below the ADQL text area to submit the query.
- In Topcat's main window, a new table appears, which stores the result of your query. Rename the table to **FOF-mostmassive** or similar by editing the **Label** field on the right side at **Current Table Properties**. Confirm by pressing the **Enter** key.
- You can browse the result table at **Views** → **Table data** . Check, if the mass-column is really sorted in descending order.

2 Mass function for FOF groups


In this section we plot the mass function for dark matter clusters of a simulation.

1. Get the mass function

- Switch back to the TAP query window of Topcat.
- A mass function gives the number of objects for a selected mass bin, which can be done directly in ADQL. Let's try this for the MultiDark Run 1 (MDR1) simulation, using the `fof` table. It contains dark matter clusters found with a Friends-of-Friends (FOF) algorithm. Type the following query into the ADQL form:

```
SELECT ROUND(LOG10(mass)) AS logmass,
COUNT(*) AS num
FROM mdr1.fof
WHERE snapnum=85
GROUP BY logmass
ORDER BY logmass
```

This will sort the FOF clusters into logarithmic mass bins using the `FLOOR`-function. The combination of `COUNT` and `GROUP BY` allows to count the number of clusters per bin.

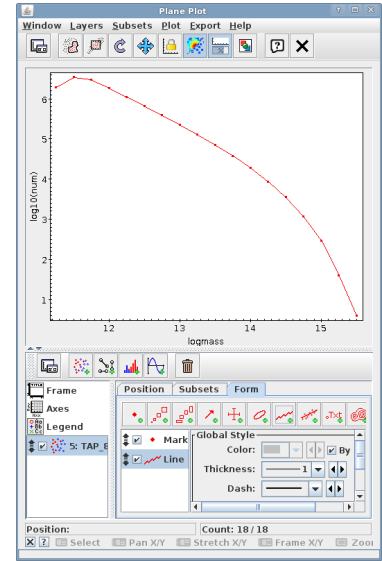
- Change the **Mode** of your ADQL query to **Asynchronous**. This is important since the following queries may take a couple of minutes. using asynchronous mode means that the query will keep running even if the connection to the server times out.
- Click **Run Query**, wait a while and view the resulting table data  after the query finished.
- Refine the bins by using steps of 0.25, i.e. enter following query:

```
SELECT 0.25*(ROUND(LOG10(mass)/0.25)) AS logmass,
COUNT(*) AS num
FROM mdr1.fof
WHERE snapnum=85
GROUP BY logmass
ORDER BY logmass
```

Submit again and check the **Results Table**: There are more rows now (factor 4).

2. View the result

- In Topcat's main window, choose **Graphics** → **Plane Plot** . This opens a new window. In the **Position** tab, choose following axes: **X: logmass**, **Y: num**.
- Select the **Form** tab next to **Position** and **Subsets** and click the button **Add new Line form** for connecting the points.
- Choose **Axes** in the left side to get access to the axes-properties. At tab **Coords** (that's the default) tick **Y Log** to make the y-axis logarithmic.
- Instead of changing axis properties, you can also add the logarithm into your ADQL query or Topcat's Plane Plot window, at Positions, enter `log10(num)` for the Y-axis.
- Now you got your first mass function plot of FOF dark matter clusters for the MDR1 simulation!
- If there are more simulations or halo catalogs available (e.g. from the CosmoSim database), you can load them into Topcat and directly compare the different mass functions with each other.



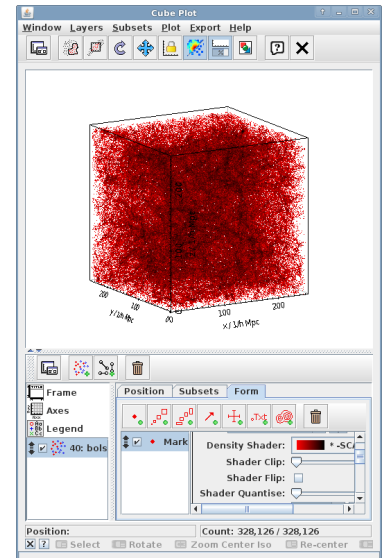
3 3D distribution of dark matter objects

- Go to Topcat's TAP Query interface .
- Change **Max Rows** (just above the ADQL text area) to the maximum, because we will exceed the default limit of 2000 rows.
- Enter following new ADQL query (escape the column name `size` with double quotes, since it is a reserved keyword):

```
SELECT x, y, z, mass, "size" FROM bolshoi.fof
WHERE snapnum=416 AND mass >= 1.e10
AND RAND() < 0.01
ORDER BY mass DESC
```

This randomly selects 1% of all halos above the given mass threshold and sorts them by mass. The query retrieves ~ 3 MB of data and may take a minute to complete. The results will look even nicer, if you change the fraction from 0.01 to 0.10; but this may slow down Topcat's 3D view substantially. If you already have performance problems with Topcat, add `TOP 2000` at the beginning of your query.

- Make sure to select **Mode: Asynchronous**, click **Run Query** and watch the new table being loaded into Topcat.
- In Topcat's main window, select the new table, then choose **Graphics** → **Cube Plot** . This opens a new window for 3D plotting.
- Choose for **positions**: **X: x**, **Y: y**, **Z: z**.
- Switch to the **Form** tab. Increase the window size for better visibility. Select **Shading, Mode: density**.
- Scroll down to **Global Style** and set the dot **Size** to 0 instead of the default 1. Use the left mouse button to rotate the 3D plot. More help on navigation is available by clicking the small question mark at the bottom left of the window. The distribution of FOF objects already gives an impression of the 3D distribution of matter in the universe.




4 Particles of a dark matter halo

For a few simulations, CosmoSim also provides the snapshots of selected timesteps; an extract of the this data is also available with the AIP DaCHS interface which we use here. This makes it possible to extract particles for a given position or a halo. In this section, the particles of an object from the FOF-catalogue of the MDR1-simulation will be extracted and plotted with Topcat. Please note, that we use here the term *halo*, though actually, the FOF catalogue returns only groups of connected particles and makes no statement if they are truly gravitationally bound or not. Thus it would be more correct to speak of *FOF groups* instead, but let's keep it simple.

- In Topcat's TAP Query interface, enter following query:

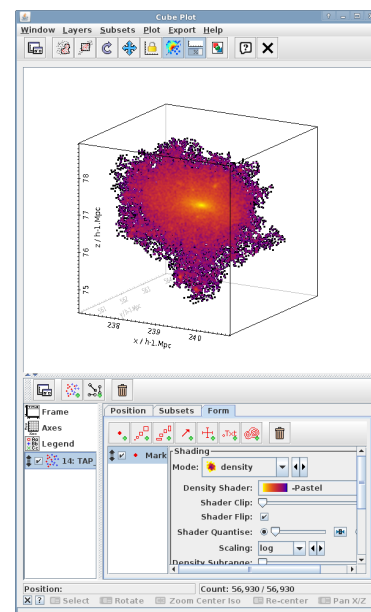
```
SELECT p.x,p.y,p.z
FROM
  mdr1.particles85 AS p,
  (SELECT particleId from mdr1.fofparticles
   WHERE fofId = 85000001014) AS fp
WHERE fp.particleId = p.particleId
```

This query first selects the ids of all particles for the FOF-halo 85000001014 and then uses them to extract their positions from the snapshot stored in table `mdr1.particles85`.

- Click **Run Query** to submit the query.
- In Topcat's main window, select the new table, then choose **Graphics** → **Cube Plot** . This again opens a new window for **3D plotting**.
- Within the **Cube Plot** window of Topcat, choose following axes: **X**: `p.x`, **Y**: `p.y` and **Z**: `p.z`

There are so many particles, that the true shape is somewhat obscured. But we can color the dots by projected density: choose the **Form**-tab (next to **Positions** and **Subsets**) and select at **Shading** the **Mode**: `density`. Select a colormap to your liking as well at **Density Shader**. Use **Pastel**, **log** scaling and check **Flip** to get the same results as on the screenshot.

- You can navigate round the 3D plot using the mouse: drag to rotate, use the mouse wheel to zoom in and out, and right-click (or CTRL-click) to re-center. More help on navigation is available by clicking the small question mark at the bottom left of the window.
- If you look close enough, you can even detect small lumps of higher density that could represent small subhalos. You could check if halo finders have detected them as well by extracting the positions and sizes of dark matter halos in this region (use the `bolshoi.fof` table) and overplotting them here (see more extensive tutorials at the CosmoSim webpage or GAVO Tutorials page).




5 Mass growth of a halo

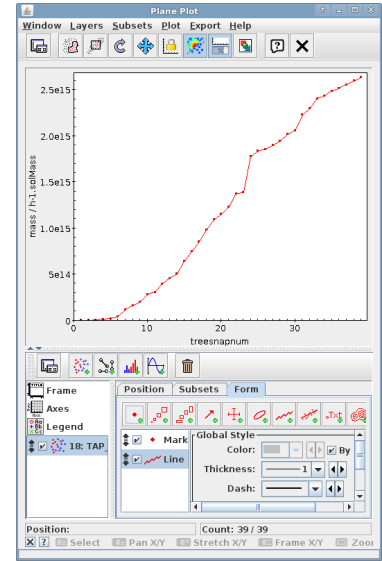
Dark matter halos grow with time, because they accrete loose material ("smooth accretion") as well as other dark matter halos. This can be detected when plotting the mass growth of a halo with time.


- Go to the Topcat TAP Query interface and enter following query:

```
SELECT p.fofTreeId, p.treeSnapnum, p.mass, p.size
FROM mdr1.fofmtree AS p,
(SELECT fofTreeId, mainLeafId FROM mdr1.fofmtree
WHERE fofId = 85000000001) AS m
WHERE p.fofTreeId BETWEEN m.fofTreeId AND m.mainLeafId
ORDER BY p.treeSnapnum
```

This query selects all direct progenitors of the given halo (so-called *main branch*). Adjust the `fofId`, if you want to get the mass growth of a different halo, but note that only values between 85000000000 and 85000000010 will yield results, because the test dataset at AIP's DaCHS instance does not contain the complete data set.

- Click **Run Query** and when the new table is loaded, select table in Topcat's main window.
- In Topcat's main window, choose **Graphics** → **Plane Plot** .
- In the **Position** tab, choose following axes: **X**: `p.treeSnapnum`, **Y**: `p.mass`. The `treeSnapnum` is the time axis here: 0 is a snapshot closer to the beginning and 39 at the end of the simulation.





- Switch to the **Form**-tab and add a new line form  for connecting the points. Steep increases in the mass from one timestep to another indicate the accretion of a lot of mass at once, i.e. merging events with other dark matter halos. Sharp decreasing mass can give a hint that two clusters of dark matter particles are close enough to be considered as one object at a certain timestep, but due to fluctuations in the particle distribution they may be considered as separate objects again in the next timestep.

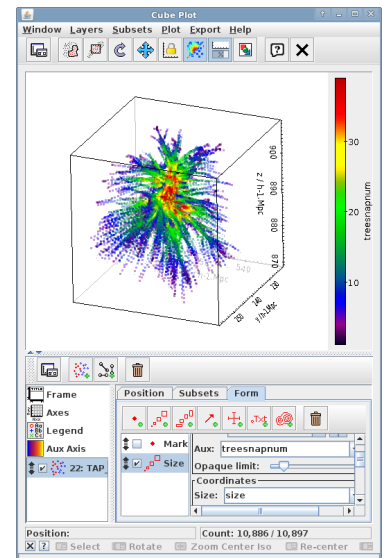
6 Merger tree: progenitors of a halo

Let us now find all the halos which have merged into one dark matter halo at the end of the simulation.

- Go to the Topcat TAP Query interface and enter following query (again adjust the `fofId`, if you want):

```
SELECT p.fofTreeId, p.treeSnapnum, p.x, p.y, p.z, p."size"
FROM mdr1.fofmtree AS p,
(SELECT fofTreeId, lastProgId FROM mdr1.fofmtree
WHERE fofId = 85000000001) AS m
WHERE p.fofTreeId BETWEEN m.fofTreeId AND m.lastProgId
ORDER BY treeSnapnum
```

- Click on **Run Query** and select the result table in Topcat.
- In Topcat's main window, go to **Graphics** → **Cube Plot** .
- Select following axes: **X**: `p.x`, **Y**: `p.y` and **Z**: `p.z`.
- Switch to the **Form**-tab and add a new size form . Set the **Size** to `size` for scaling the dots with the average halo radius.
- Apply more color by selecting at **Shading**, the **Mode**: `aux` and entering **Aux**: `treeSnapnum`. Adjust the **Opaque limit** to your liking.
- Now halos from the beginning of the simulation are colored black-purple, more recent progenitors have yellow-red colors. This nicely shows how the progenitor halos have merged with each other at earlier times to form the final big halo (red) at the center.




7 The density field with TOPCAT

- Go to the Topcat TAP Query interface and enter following query in order to extract a slice of the density from the Bolshoi simulation at the location of one of the most massive dark matter halos:

```
SELECT ix,iy,iz,dens FROM bolshoi.dens256
WHERE snapnum=416 AND iz = 94
ORDER BY ix, iy
```

We use here the table Dens256, which contains a 256^3 grid, and for each grid cell the density value. Since we only fix the iz -value, we will get a 256×256 slice in x - y plane returned.

- Choose **Mode**: **Asynchronous** and click **Run Query**. This may take a while.

- In Topcat's main window, select your table and choose **Graphics** → **Plane Plot**  to open the plane plot window.

- In the **Position** tab, choose following axes: **X**: ix , **Y**: iy .

- Select the **Form** tab next to **Position** and **Subsets**. Increase the size of your window (click and drag with left mouse button) or switch to full screen mode in order to see all the possible options.

- At **Shading** choose the **Mode**: **aux**; and for **Aux**: $dens+1$. This adds density coloring to your dots, but currently your image looks nearly black. We add +1 here, because the density table contains overdensities with minimum values starting at -1, but we want to plot the overdensities here.

- Select the **Aux Axis** from the left menu to adjust its properties. Change here the **Scaling** to **log** instead of linear. This will apply a logarithmic scaling for the colors (since the logarithm is undefined for negative values, we had added +1 to $dens$ beforehand). The web like structure should be visible now.

- At **Aux Shader** choose a color table you like, e.g. **Cold** for blue shades.

- Switch to the **Range** tab to fix the range of your density values: **Minimum Aux**: 0.01, **Maximum Aux**: 100 This will make the structures more prominent.

- Inspect the same slice region at different redshifts, e.g. for snapnum 243 ($z=1$) or 36 ($z=10.9$). Just replace the snapnum in the query above, submit the query and view the results with Topcat. Can you see how the large scale structures formed from small overdensities in the much more homogenous early universe?

