

# Introduction to simulation databases

**Author:** K. Riebe

**Software:** [MultiDark web interface](#), [TOPCAT](#)



Simulation databases like the Millennium Database (<http://gavo.mpa-garching.mpg.de/Millennium/>) or MultiDark (<http://www.multidark.org>) contain data sets from cosmological simulations. Each user has access to catalogues of dark matter halos, their evolutionary history (merger trees) and the relations between them (substructures), 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*, *IDL* or *R*.

This tutorial gives a quick overview on the MultiDark Database web interface and then shows an example use case for viewing the cosmic web. Two example files (*Bolshoi\*.csv*) are also available on this CD, so that you can follow most of the steps even if you don't have a user account/demo user or no internet connection. More demos and tutorials can be found at the [MultiDark](#) webpage.

## The web interface

1. Open a **web browser** with <http://www.multidark.org>. This webpage provides access to the MultiDark Database, where data sets like dark matter halo catalogues and merger trees are stored for currently three cosmological simulations: Bolshoi, MDR1 and the newly added MDPL simulation, which uses a cosmology consistent with recent results from the Planck satellite.
2. **Registration:**
  - For gaining access to the full data sets, it is required to register at <http://www.multidark.org/MultiDark/pages/RegistrationForm.jsp> beforehand. The registration is free and open to everyone.
  - During tutorials and hands-on sessions a *demo user account* may be enabled for testing purposes. Please ask your tutors for the current demo user's name and password. Some example data are also stored directly on this CD, so that you can follow most steps also without registration.
3. **Query Form:**
  - Navigate to the **Query Form** (<http://www.multidark.org/MultiDark/MyDB>) and login with your username and password (or as the demo user/public user).



- The available simulations and tables are listed at the left side, a **click on the table name** brings you to the corresponding **documentation** page, where all the possible columns are explained.
- SQL queries are written directly into the sql form on the webpage. The dialect used here is called T-SQL and is specific for the Microsoft SQL Server that sits at the backend. Now write

```
select top 100 * from miniMDR1..FOF where snapnum=85 order by mass desc
```

into the sql form. Set the **Maximum number of rows to return** to 100 and click the **Query (browser)** button. After a few seconds the result will be returned as a table directly at the bottom of the web page. You just retrieved the properties of the 100 most massive clusters (ordered by mass) for redshift 0 (=snapshot number 85) from the miniMDR1-simulation database!

- **Query (stream)** has a less strict time out limit and returns data in CSV format, i.e. with commas to separate the values.
- Note that there are **time limits**, so choose carefully which data you really need to download. If you are not familiar with SQL, have a look at some tutorials, e.g. <http://www.multidark.org/MultiDark/Help?page=demo/index> and its subpages.
- In principle, a basic query consists of the following parts:
 

<b>select</b>	followed by a limit ( <b>top 10</b> ) and a list of columns, e.g. <b>x,Rvir,npart</b> or <b>*</b> for all
<b>from</b>	followed by a list of tables (possible join)
<b>Bolshoi..BDMV</b>	database and table name
<b>[where]</b>	followed by a list of restrictions (filter), e.g. <b>snapnum=85</b>
<b>[order]</b>	followed by a (list of) columns, e.g. <b>by mass desc</b>

4. Consult the **documentation** to get to know which table provides which data.

## The Cosmic Web with TOPCAT

### 1 Load a slice (x-y plane) of the cosmic web for the Bolshoi simulation

- Before actually downloading some data, let's first check where the **most massive halo** is located, using the **BDMV halo** catalogue. If you don't have an account, skip this step, proceed to step 2.

- Open a **web browser** with the URL: <http://www.multidark.org>
- Go to the **Query Form**, login and enter following query:


```
select top 1 ix,iy,iz,x,y,z,Mvir from Bolshoi..BDMV
where snapnum=416 order by Mvir desc
```


- **Click Query (browser)** to submit the query.

This selects the most massive halo for snapshot number 416 (corresponds to  $z=0$  for this simulation) and returns its position and mass  $M_{\text{vir}}$  within the virial radius, as well as the indices of a  $1024^3$  grid cell, in which the halo sits ( $ix, iy, iz$ ). The size of the grid is given in the documentation.

- One can **convert these numbers** into the corresponding indices of a coarser grid by dividing them by **factors of 2**. E.g. a grid cell with  $ix=200$  in a  $1024^3$  grid corresponds to  $ix' = \text{floor}(ix/2.) = 100$  in a  $512^3$  grid, and  $ix' = 50$  for a  $256^3$  grid. Let the SQL server do it for you:


```
select top 1 ix,iy,iz, floor(ix/4.) as ix256,
floor(iy/4.) as iy256, floor(iz/4.) as iz256,
x,y,z,Mvir from Bolshoi..BDMV
where snapnum=416 order by Mvir desc
```

- **Click Query (browser)** again to submit the query and check the result. Leave the browser open, since we will need these numbers later on.
- We will now use [TOPCAT](#) to extract and visualize some cosmic web data for the Bolshoi simulation. Start [TOPCAT](#) and click the **Open File** icon .


- Choose **GAVO Millennium Run Query**  (If you don't have an account at hand, load the file from CD instead, see below). In the new dialogue, type following details:


- **Base URL:** <http://wget.multidark.org/MyDB>
- **username and password:** use your own or the demo user credentials (if you have neither, load the example file)
- **SQL Query:** enter a query to retrieve your slice:
  - The required table is **Tweb256**, which contains a  $256^3$  grid, and for each grid cell the eigenvalues and eigenvectors for the gravitational tidal tensor.
  - We select a slice by choosing only rows with a fixed  $iz$ -value:



```
select * from Bolshoi..Tweb256 where iz = 100
```

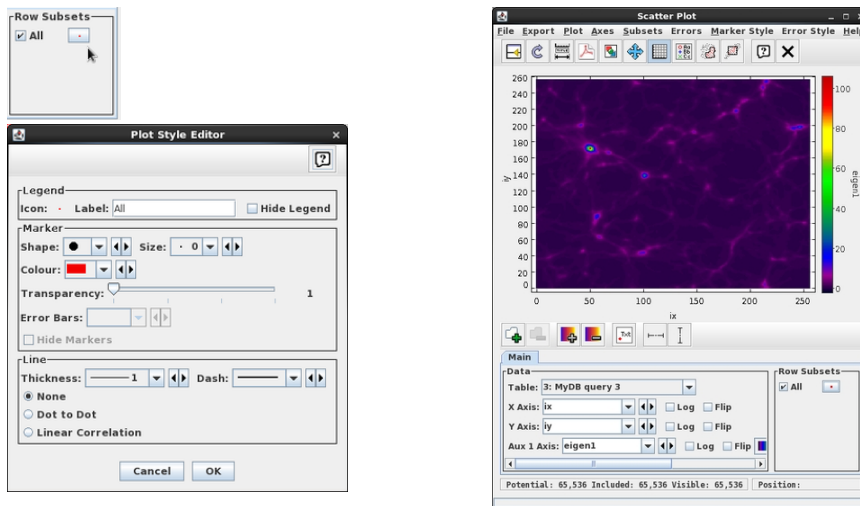
- To make it more interesting, we can also choose the slice which contains the center of the most massive halo, so replace the 100 in the query by the **iz256** value you found with the browser query in the previous step.
- **Click OK** to submit the query.
- In case you don't have an own account and no demo user credentials are available, you can **load a file from the CD**: click the open file icon  and choose the file **Bolshoi-Tweb256-iz94.csv**. Set Table Format to CSV and click OK. Now you loaded a cosmic web slice for  $iz=94$  for the Bolshoi simulation.



## 2 View the Cosmic Web

1. Have a look at the downloaded data with the **Table Browser**  which you can access from [TOPCAT](#)'s main window. It displays your data as a table. The column names have also been imported.

2. **View your data** by clicking on the **Scatter Plot** icon .

- Choose: **X Axis: ix**, **Y Axis: iy**
- Add a **new auxiliary axis** by clicking on the color gradient icon .
- Set **Aux 1 Axis** to one of the eigenvalues (e.g. **eigen1**) for coloring the dots.
- **Increase the point size** by clicking on the colored symbol  in **Row Subsets**, in the box at the lower right of the dialogue window. This opens a new dialogue, where you can also adjust e.g. the transparency of your dots or the marker style. At Marker, Size, choose **at least size 1** for your points.



- Step through **different columns for Aux 1** to get an impression of the other eigenvalues as well. Also look at the pkkeys: These are Peano-Hilbert keys, describing a space-filling curve. They are very important for spatial queries.
3. Just visualising one of the eigenvalues does not yet give much information. Instead, we need to **count the number of eigenvalues** above a certain threshold. This number will allow us to assign each cell with following classification:
- 0 = cosmic void
  - 1 = sheet-like structure
  - 2 = filament
  - 3 = knot (= location of galaxy clusters)
4. We can get this number by adding a **computed column** to our data:
- In [TOPCAT](#)'s main window, click on **Display Column Metadata** .
  - Click on the green plus  to **add a new column** to your table.
  - In the new dialogue, enter a **name for your new column**, e.g. num.th0.4
  - Enter the expression to compute the column:

```
(eigen1>=0.4 ? 1:0) + (eigen2>=0.4 ? 1:0) + (eigen3>=0.4 ? 1:0)
```


This expression checks, if an eigenvalue is greater than or equal to the threshold value (e.g. 0.4) and adds 1, if this is the case, otherwise 0.

- **Click OK.** You can still edit the name and expression of your column by double clicking on it in the Column Metadata dialogue.
- Switch back to the **Scatter Plot Window** and choose the **new computed column** as **Aux 1 axis**
- Now you should see some clustered structures (red knots) and filaments (green). Note that this is just a cut through the z-direction, so perpendicular sheets appear as filaments and filaments may appear as knots etc.
- By the way: instead of calculating the additional column with [TOPCAT](#), you could already do it within the SQL query itself:

```
select *,
  (case when eigen1 >= 0.4 then 1 else 0 end)
+ (case when eigen2 >= 0.4 then 1 else 0 end)
+ (case when eigen3 >= 0.4 then 1 else 0 end)
as num
from Bolshoi..Tweb256 where iz=94
```



This method only has the disadvantage that you would need to query the database again for each new threshold value that you are testing.

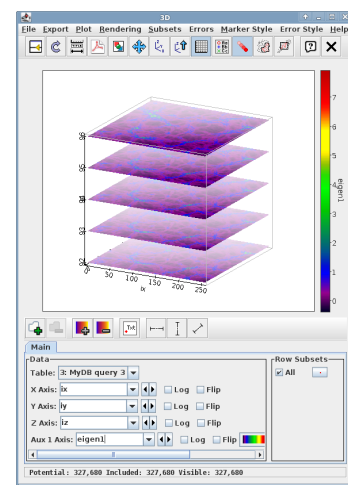
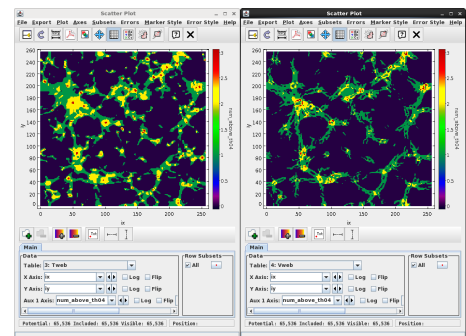
### 3 Explore more data [optional]

1. Repeat the **same steps** for loading a slice of the same region from the **Vweb256** table (you can also load the file Bolshoi-Vweb256-iz92-96.csv from this CD, which actually contains 5 slices in total). **Compare the differences:** the Vweb should give a much better resolution of structures than the Tweb. (Look in the documentation at the webpage and the linked papers to learn about the physical reasons behind this.)
2. Experiment with **different threshold values** by adding more computed columns.
3. Experiment with **slices at different positions**. You can also download several z-slices and plot them in the 3D window . For example, do following query:

```
select * from Bolshoi..Vweb256
where iz between 92 and 96
```



or load the file Bolshoi-Vweb256-iz92-96.csv from the CD.

- Now click on **Three dimensional scatter plot**  and adjust the columns: X: ix, Y: iy, Z: iz.
- Add an auxiliary axis  and choose Aux 1: eigen1 or add again the sum of eigenvalues above a given threshold value as explained above and color by this value.





4. Load a **finer grid** by using the **Vweb512** table and compare with the coarser version. Note, that you need to **convert the iz-value properly** if you want to get matching slices:

```
select * from Bolshoi..Vweb512
where iz = 188
```



We need to multiply the previously used iz-value (e.g. 94) from the 256-grid by 2 to get the matching phkey for the 512 grid. For comparison, open the **Scatter Plot** window  and click on the plus sign  to add another table. Adjust the x,y axis, add again an auxiliary axis for this new table and compare.

## 4 Find halos in cells

1. Open the **Table browser** and a **Scatter plot window** for the same data table (e.g. for Tweb256).
2. **Click on one of the knots** in your plot and check to which cell it belongs: the corresponding **row** in the table browser is **highlighted** now.
3. **Load another table** with TOPCAT by clicking on the **File Open**-icon , then choose **GAVO Millennium query**  again and insert a query for retrieving all halos inside this cell, sorted by mass, e.g.:

```
select * from Bolshoi..BDMV
where snapnum=416
and floor(0.25*ix) = 101
and floor(0.25*iy) = 138
and floor(0.25*iz) = 94
order by Mvir desc
```

Replace 101, 138 and 94 by the values for your cell.<sup>1</sup>

4. **Click OK** to download the data.
5. Switch to the **Scatter plot window** and click on the **+** symbol  for adding a new data set. Choose the **halos-table** you just downloaded.
  - Set **x-axis: ix/4., y-axis: iy/4.**
  - **Adjust the point size** and color by clicking on the colored symbol  at Row Subset.
6. **Zoom** into the cell where the halos are located, by clicking with the left mouse and dragging a rectangle to the lower right corner around the interesting region. To zoom out, click and drag the rectangle in the opposite direction, to the upper left corner.
7. Open the **table browser** for the halos table. Click on the **most massive halo** in the table browser and watch where it is marked in the Scatter Plot window. Zoom in to see if it is lying at the center of a knot region (as would be expected) or farther away.

<sup>1</sup>You can also do this as a public user with the miniMDR1-database, which contains a subvolume of the miniMDR1 simulation. The whole Bolshoi BDMV catalogue would have been much too big for this CD.