

SQL Databases to access cosmological simulation results

CLUES Workshop. Lyon, 2012
Fernando Campos

SQL Databases to access cosmological simulation results

Why??

- Too big data to handle it on files
- Easy access (learn SQL! it's easier than LaTeX!)
- Access available on project's web page

Our bet for DBMS is MySQL

Why??

- Good performance as non-transactional DataBase Management System
- Widely adopted on the community guaranties support and documentation
- AIP is already working on it
- OpenSource, of course ;)

Our bet for CMS is Joomla

Why?

- We want to USE a web system (develop as less as possible), so we decided to use a *Content Management System*
- Widely used, supported and documented
- Provides core functionalities and multiple extensions (Image/Video Galleries, Users Management, etc)
- Allows extensions development with PHP+JavaScript

Of course we found cons

- No extension (already developed) allows you to freely query your database.
- Usually, this is a security issue. In our case, it's "just" a bit dangerous
- Not so easy developing an extension within the Joomla Framework, but worth it!! why???

Once we have a Joomla extension which allows us query our database:

- We will can replicate the system in every project webpage we have
- Look and Feel will be the same for every project
- A new feature in the extension will be available for every project (example queries, helped query creation, explained queries)
- Bad news: we don't have the first release yet, but we will as soon as possible (hopefully this summer)

So... What are you selling to me??

- At the moment we have a temporary open web access in our server (ask me for the URL)
- It's insecure, annoying to use, bad documented and requires knowing the database structure, but if you need the data you can get it!!
- And we are growing our databases (inserting data in the databases is not easy or fast at all)
- If you plan to dump your data into a database in the future, let us know and we will help you to avoid future problems
- You could make our life easier if you keep that in mind when you write your output files. Thanks :)

At this moment (June 2012) we have

- Music, Jubilee and CurieHZ projects' webpages running. We still need time to generate and upload the content
- Stay tuned on:
 - <http://music.ft.uam.es>
 - <http://jubilee-project.org>
 - <http://curiehz.ft.uam.es>

At this moment (June 2012) we have [2]

Music project database with data (AHF) of DM, gas and star on 282 resimulations, structured on 2 tables

- Catalog. Halo catalog with more than 1.4M rows, 7 redshift values, halos and subhalos distinction and luminosity and magnitudes data
- Profile. Halo properties at different r

To do: include merger trees

Maybe: split luminosity, magnitude and particles data in different tables... I need your feedback!!

At this moment (June 2012) we have [3]

Jubilee project database

- FoF: Mvir, position and velocity data in a single table for +542M rows at $z=0$
- SO: One table with Mass, Rad, position, velocity and z data. +16KM rows. Yes, $16e+9$ rows, 1TB!! No way to work with it. We are splitting in one table per redshift, 92 tables. It's that fine for you?? Don't you need different z related data?? Feedback, please!

At this moment (June 2012) we have [4]

CurieHZ project database

- We are dumping the data right now
- At this moment we have created 165 tables for each redshift/snapshot (AHF)

What can I do for you??

(in a couple of months)

- *SQLform* Joomla extension's first release to allow web integrated access to the data from the project web pages
- More simulation data into the database

What can YOU do for me??

- Create your account in every of our project webpages: it's the way I'll contact you with good news!
- Begin using the already working temporary SQLform data access and let me know which problems are you facing during the use of SQL: let me help you!!!

Thanks for listening!!

Any questions??

Fernando Campos
fernando.campos@uam.es