

CADAC

Computational Astrophysics Data Analysis Center

a joint initiative of the
Laboratory for Computational Astrophysics
and
San Diego Supercomputer Center
at the
University of California at San Diego

Outline

- What is **CADAC**?
- **CADAC** archive design
- Why **CADAC**?
- How do I use **CADAC**?
- Current status and future plans

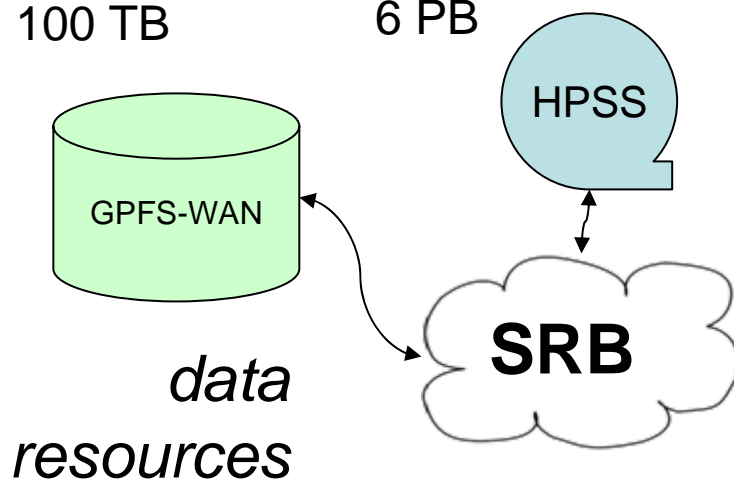
What is CADAC?

- **CADAC** collects and stores results of large astrophysical simulations and provides data analysis resources to researchers worldwide
- hosted by the SDSC, developed by LCA
- *Data resources:*
 - 100 TB shared file space backed up by HPSS
- *Compute resources:*
 - two IBM p690 32-cpu supercomputers with 256 GB of RAM
- *Data management resources:*
 - collection managed by the SRB (Storage Resource Broker)
- *User resources:*
 - CADAC website; SRB client software

CADAC Components

100 TB

6 PB

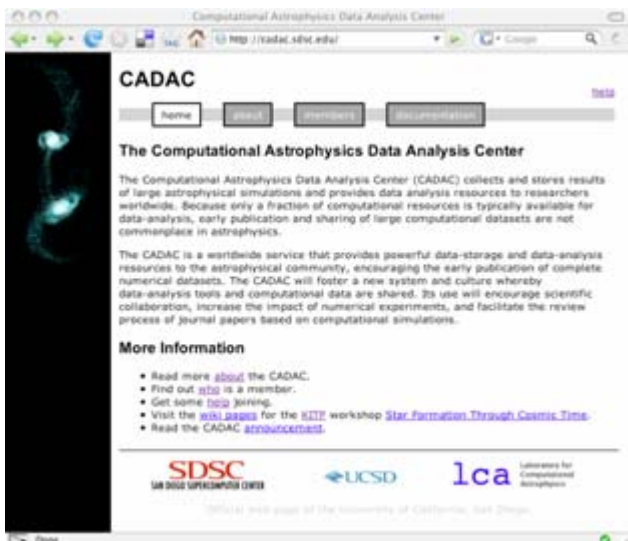


IBM Power4
p690's



*compute
resources*

documentation



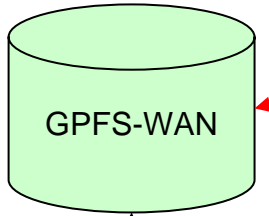
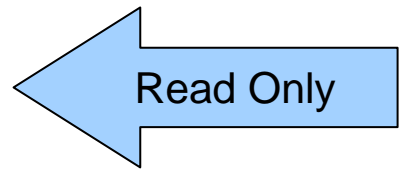
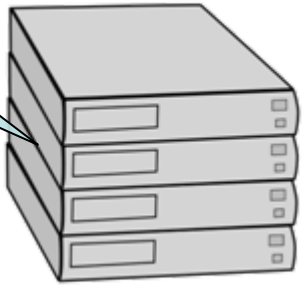
*user
community*



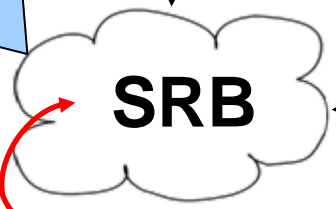
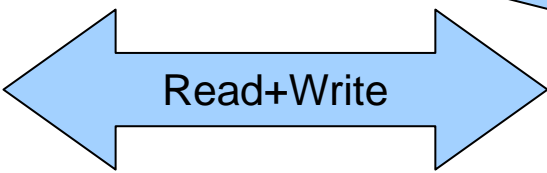
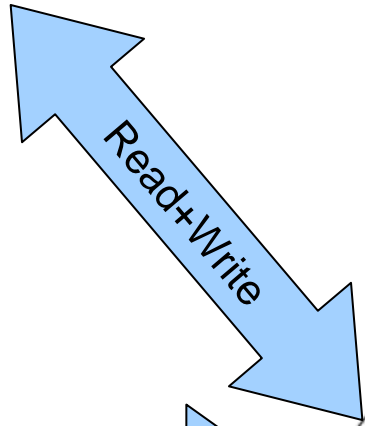
CADAC Data Archive

`/gpfs-wan/project/cadac/srbVault`

dsdirect



workstation



`/CADAC`



Why CADAC?

- Data sharing
 - a place share large CA data sets with your collaborators, your community, the world
- Data analysis
 - data sets are becoming too large to analyze on workstations;
 - data and supercomputers need to be co-located
- Data publication
 - Like an `astro-ph` for publishing the data behind your research publications

How do I use **CADAC**?

- First you must *join* the CADAC project to have access to the resources (see Paolo Padoan or website)
- Read *online documentation* to
 - Download SRB client software
 - Browse the archive
 - Upload data to the archive
 - Upload analysis tools and scripts to archive
 - Run analysis jobs on DataStar

cadac.sdsc.edu

CADAC

[home](#)[about](#)[members](#)[documentation](#)[help](#)

The Computational Astrophysics Data Analysis Center

The Computational Astrophysics Data Analysis Center (CADAC) collects and stores results of large astrophysical simulations and provides data analysis resources to researchers worldwide. Because only a fraction of computational resources is typically available for data-analysis, early publication and sharing of large computational datasets are not commonplace in astrophysics.

The CADAC is a worldwide service that provides powerful data-storage and data-analysis resources to the astrophysical community, encouraging the early publication of complete numerical datasets. The CADAC will foster a new system and culture whereby data-analysis tools and computational data are shared. Its use will encourage scientific collaboration, increase the impact of numerical experiments, and facilitate the review process of journal papers based on computational simulations.

More Information

- Read more [about](#) the CADAC.
- Find out [who](#) is a member.
- Get some [help](#) joining.
- Visit the [wiki pages](#) for the [KITP](#) workshop [Star Formation Through Cosmic Time](#).
- Read the CADAC [announcement](#).

SDSC
SAN DIEGO SUPERCOMPUTER CENTER



lca Laboratory for
Computational
Astrophysics

Official web page of the University of California, San Diego.

Documentation and Guides

Accessing and Using DataStar

For information on logging in to and using [DataStar](#), please visit the [SDSC User Support](#) site.

- [DataStar](#)
- [SDSC User Support](#)

To use DataStar, you'll need to join. See the [help page](#) for instructions.

Contributing Software

CADAC members are welcome to contribute software for data analysis. This is especially useful if you have contributed data with a custom format, since it will help other members to access your results.

- [Instructions on how to contribute software.](#)

Contributing Data

To add or access data in the CADAC, you must first [join](#) to obtain an SRB user name and password.

Accessing Data

To access or browse data in the CADAC archive, please follow the [instructions on setting up an SRB client](#), and then read the [sections on browsing](#) and [getting data](#) in the [instructions on adding data](#).

- [SRB client settings \(connecting\).](#)
- [Browsing the archive.](#)
- [Getting data from the archive.](#)

Communicating

The CADAC maintains an archived mailing list, cadac-users@sdsc.edu.

- [Mailing list information and subscription page.](#)
- [Mailing list archive.](#)

Contributed Software

File Locations

CADAC software should be placed under the GPFS-WAN project directory, `/gpfs-wan/projects/cadac/`, in the `software/` subdirectory. The `software/` directory is analogous to a typical `usr/` directory, with the addition of a `doc/` and `contrib/` directory.

```
/gpfs-wan/projects/cadac/software/  
    bin/  
    contrib/  
    doc/  
    include/  
    lib/  
    share/
```

We would like to reserve the `bin/`, `doc/`, `include/`, `lib/`, `share/` directories for software specific to supporting the CADAC. Contributed software should either be placed in its [own directory](#), or in the `contrib/` directory, [organized by member](#).

Permissions and Ownership

This space is for shared tools, so please ensure the group has read and execute permissions as necessary on any directories or files you add. The group id (`cad116`) should be set automatically, but the permissions will not.

Adding Personal Scripts

Most CADAC members will probably not need to install complete software packages, but will need to share some scripts or other tools. Doing this is simple:

1. Create a directory under `software/contrib`
2. Add your tools as you see fit
3. To help others, put a `README` at the top level

As an example, here is how to create a personal directory, and make sure the group can access it:

```
ds001 % cd /gpfs-wan/projects/cadac/software/contrib/  
ds001 % mkdir rick  
ds001 % chmod g+rx rick
```

SRB Client Settings

Step 0: Remote or Local?

When adding data to the CADAC data archive, the final step is to *put*—like FTP—the data into the SRB. However, you must decide whether you want to do that remotely, or move your data to DataStar first. Accessing the SRB remotely makes adding data a one-step process, but you must install an [SRB client](#). Moving the data to DataStar first, allows you to use your preferred file transfer method, but then you have to put the data into the SRB.

If you choose to use a remote client, this page assumes that you already have the [SRB Scommands](#) installed. Compiled binaries are available for OS X and Windows, users of POSIX operating systems (Unix, Linux, etc.) will need to compile the clients. If you are using another SRB client, such as [inQ](#), the concepts are the same.

Step 1: Check Client Path

Make sure that the SRB client is in your path.

```
$ which Sinit
/usr/local/apps/srb/SRB.v.3.3.1/Sinit
```

On DataStar

If it is not found, add `/usr/local/apps/srb/SRB.v.3.3.1` to your path.

For the bash shell, edit your `~/.bash_profile`, and add

```
export PATH=$PATH:/usr/local/apps/srb/SRB.v.3.3.1
```

If you're using `csh` or `tcsh`, then edit your `.cshrc` file and add

```
set path = ( $path /usr/local/apps/srb/SRB.v.3.3.1 )
```

Remote Users

Following the same steps as if you were on DataStar, replacing `/usr/local/apps/srb/SRB.v.3.3.1` with the location of the `Scommands` on your machine.

Adding Data

Client Configuration

This page assumes you have [configured the SRB client](#) for your system.

SRB `scommands`

From the [Scommands](#) documentation:

`Scommands` are command line utilities that run in the Unix, Windows or Mac OS command shells. Most `Scommand` names are preceded by an "S". These are Unix-like commands for accessing SRB data and metadata.

The `Scommands` may conflict with other conflict with other commands on non-case-sensitive file systems. In particular, the SRB command to remove a file in the SRB is `srm`, which conflict with `rmdir`. Since `rmdir` works on your local file system, you may want to alias it to `rmdir -i` to avoid surprises.

This means that many familiar commands, like `cd`, `pwd` and `ls`, also appear as `Scommands`, but in the form of `scd`, `spwd` and `sls`. In addition, there are a few FTP-like commands, i.e., `sput`, `sget`, which handle data transfers, and more specialized commands for dealing with metadata (not covered here).

Here is a [complete list of the Scommands](#), with links to their documentation.

Browsing

Browsing the SRB is like browsing a file system, you change directories (`scd`), look at stuff (`sls`), and every now and then, forget where you are and check (`spwd`).

Before we do anything, we need to connect to the SRB.

```
$ Sinit
```

When we first connect, we're placed in our home directory in the SRB. You can check this using `spwd`.

```
$ Spwd
/home/rick.cadac
```

Current status and future plans

- KITP workshop
 - CADAC launch
 - Facilitate code comparison project (organizers)
 - Central repository for shared data sets
 - Common set of analysis tools
- Future
 - CADAC catalog (SQL searches)
 - CADAC data grid
 - Publishing mechanisms
 - NSF funding