SciServer Compute Brings Analysis to Big Data in the Cloud

Jordan Raddick, Dmitry Medvedev, Gerard Lemson, Barbara Souter
Institute for Data Intensive Engineering and Science, Johns Hopkins University

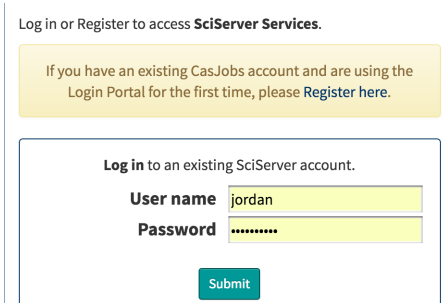
idies
The Institute for Data Intensive Engineering and Science

SciServer Compute is a new resource that lets you run Python and R scripts on big data online in Jupyter notebooks. Compute ***brings the analysis to the data***.

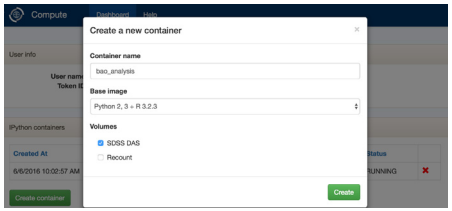
How to compute

1 Go to compute.sciserver.org.

2 Sign in at the login portal.



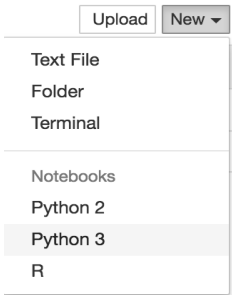
3 Create a container.



4 Click your container name to open.

5 To save data, use the persistent folder.

6 Click New, OR drag any .ipynb file and click Upload.



Data
We offer access to the complete SDSS dataset to the world through SkyServer.
You can store any data you like in a private database in CasJobs, or as files in SciDrive.

SkyServer

Query
You can find SDSS data with SQL queries through SkyServer or CasJobs.
You can run queries through scripts using Compute.

SciDrive

CasJobs

Compute

Collaboration
You can share your data and results with colleagues while you are Logged In.
You can share file data in any format with SciDrive.

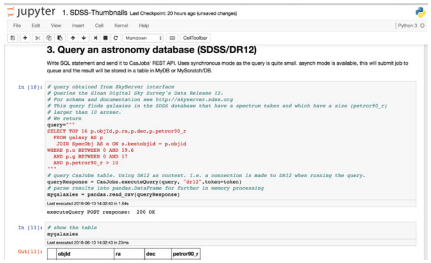
Analysis
You can analyze huge datasets with Python or R scripts through Compute.
You can share scripts and results while Logged In.

Login Portal

Online analysis notebooks

SciServer Compute lets you run computations online, no need to download data. Just go to the Compute website and upload your R or Python (2 or 3) scripts into a Jupyter notebook. You can read and write data from/to your CasJobs MyDB or SciDrive.

Example: SDSS DR12



What you can do

SciServer Compute puts powerful Big Data analysis at your fingertips. Here are some of the things you can do:

- create SDSS thumbnails
- create co-added images
- estimate photometric redshifts
- compare simulations to data
- assign homework and labs to with real data online
- Share your results and scripts

How compute works: SciServer Compute uses Jupyter notebooks running inside a Docker container on a virtual machine at JHU. You can upload scripts into run analyses there in a self-contained, independent and reproducible environment.

Questions? Email us at sciserver-webmaster@jhu.edu!

SciServer is administered by the Institute for Data Intensive Engineering and Science at Johns Hopkins University
SciServer is supported by the National Science Foundation, award ACI-1261715.

NSF

Johns Hopkins University