Collaborative Analysis of Petabyte-Scale Scientific Data

Jordan Raddick
Johns Hopkins University
Motivation

- Much of the scientific process is about data
  - collect, clean, archive, organize, analyze, publish, mirror, distribute, curate...

- Jim Gray: “Often it turns out to be more efficient to move the questions than the data.”
Motivation

- Helen Shen: “Brown's students write explanatory text and intersperse it with raw code and the charts and Sitting in the airplane... Brown can interact with the work..”

_Nature, 5 November 2014_
History

- Started with the SDSS SkyServer (2001)
- Goal: instant access to rich content
- 2012: NSF DIBBS to extend/reengineer SkyServer
Collaborative data-driven science

SkyServer in 2001
SkyServer today

Welcome to the new SkyServer!

We are proud to announce that SkyServer now connects to the new SciServer, a science framework. If you have a login through SciServer, you can log in using the control on the top right. Logging in allows you to connect seamlessly with the updated CasJobs and the brand-new SciDrive satellite management tool.


These blog features come from the first release of the SciServer framework, which we call SciServer Altair. For more information on the new features of SciServer Altair, please see our New Features section on the SciServer website.

Data Access
- Navigate
- Quick Look
- Explore
- Finding Chart
- Image List
- Search
- SQL Search
- Download SQL Queries
- Catalogs

Education
- For Educators
- Lesson Plans
- Middle School
- High School
- College Lab Activities
- Instructional Guides
- Student/Research

Links
- Data Release 12
- SDSS-III Science
- Data Archive Server
- About Astronomy
- About the SDSS
- About SkyServer
- FAQ
- Glossary
- Tool User Guides
- Site Traffic
- Privacy Policy
- Sample SQL Queries
- Data Release Papers

Help
- Start here
- FAQ
- Glossary
- Tool User Guides
- Site Traffic
- Privacy Policy
- Sample SQL Queries
- Data Release Papers

News
- The site hosts data from Data Release 12 (DR12): What's new in DR12?
- DR12 What's new on this site?

Site Map
- Home
- Data
- Schema
- Education
- Astronomy
- SDSS
- Contact Us
- Download
- Site Search
- Help
- History

Dr.12
- Name: Mrk 31
- RA: 20h49m57.60s
- Dec: 47°00'32.60"
- Type: LINER
- Distance: 47.7 Mpc
- Size: 4.9 kpc
- Eddington Ratio: 0.01
- Spin Parameter: 0.6
- Black Hole Mass: 4.9 x 10^9 M☉

Select Image Source: SDSS

Drawing options:
- Grid
- Label
- Photometric objects
- Objects with spectra
- Invert Image
- Advanced options
- APOGEE Spectra
- SDSS Outlines
- SDSS Bounding Boxes
- SDSS Fields
- SDSS Masks
- SDSS Plates

Click, hold and drag to navigate!
Data from HPC Simulations

- HPC is an instrument in its own right
- Need public access to the best and latest
  - Largest simulations approach/exceed petabytes
  - Also need ensembles of simulations
- Creates new challenges
  - Access?
  - Data lifecycle?
  - Analysis patterns?
  - Architectural support?
- Our total science data: ~2.5PB
  - Everything is a Big Data problem!
Cosmological Simulations

- Mirror of Millennium DB (universe sim)

  Raw data: Particles → FOF groups Subhalos
  Mock images → Density fields
  Mock catalogs → Synthetic galaxies
  Halo merger trees
Turbulence databases (JHUTB)

http://turbulence.pha.jhu.edu/

Current field $\nabla \times \mathbf{B}$

Vorticity $\nabla \times \mathbf{u}$
Next step: Numerical Laboratories

- From data retrieval to data **analysis**
  - No need to download entire datasets
  - Analysis is server-side through web services

- Use virtual sensor metaphor ("the cow in the tornado")
  - Many access patterns are local in space and/or time

- turbulence.pha.jhu.edu:
  - 19 trillion points delivered!
Hydrostatic and non-hydrostatic simulations of dense waters cascading off a shelf: The East Greenland case

Marcello G. Magaldi, Thomas W.N. Haine
Collaborative data-driven science

Genomics

Building a DB of a trillion short reads from Next Gen Sequencing
Materials Science

Daphalapurkar, Brady, Ramesh, Molinari. JMPS (2011)
SciServer Project Objectives

- Extend infrastructure to support additional science domains
- Host and serve Petabyte datasets
- Support custom user datasets
- Provide access and query services
- Provide scalable compute services
- Support analyses and data sets too large to handle locally
- Provide collaborative tools for shared analysis

*Computations stay CLOSE to the DATA...*
## SciServer Project Components

<table>
<thead>
<tr>
<th>Major Components</th>
<th>Supporting Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Core</strong></td>
<td>Microsoft SQL Server</td>
</tr>
<tr>
<td>• Login Portal</td>
<td>Open Stack</td>
</tr>
<tr>
<td>• CASJobs</td>
<td>Docker</td>
</tr>
<tr>
<td>• SciServer Compute</td>
<td>Jupyter</td>
</tr>
<tr>
<td>• SciDrive</td>
<td></td>
</tr>
<tr>
<td><strong>Applications</strong></td>
<td></td>
</tr>
<tr>
<td>• SkyQuery</td>
<td></td>
</tr>
<tr>
<td>• SkyServer</td>
<td></td>
</tr>
<tr>
<td>• GLUSEEN</td>
<td></td>
</tr>
<tr>
<td>• Turbulence</td>
<td></td>
</tr>
</tbody>
</table>
## SciServer Project Timeline

### Timelines

<table>
<thead>
<tr>
<th>Year 1 (2013-2014)</th>
<th>Project Setup, Scoping, Planning, Begin Refactoring, SDSS Unification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 2 (2013-2014)</td>
<td>Architectural Refactoring – API, Single Sign-on, prototype Compute</td>
</tr>
<tr>
<td><strong>Year 3</strong> (2013-2014)</td>
<td><strong>SciServer System Release, Interactive Compute, Scalable Job Management, Basic Dashboard, Initial Collaborative capabilities</strong></td>
</tr>
<tr>
<td>Year 4 (2013-2014)</td>
<td>Implementation in Science Domains, Educational workbooks</td>
</tr>
<tr>
<td>Year 5 (2013-2014)</td>
<td>System Scale out, Data Analytics, Advanced Deployment Scenarios</td>
</tr>
</tbody>
</table>

NOW
### SciServer Project Current Plans

#### Timelines – Year 3

<table>
<thead>
<tr>
<th>Month</th>
<th>Milestones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr 2016</td>
<td>• SciServer System Release</td>
</tr>
<tr>
<td>May 2016</td>
<td>• Interactive Compute</td>
</tr>
<tr>
<td></td>
<td>• SkyQuery</td>
</tr>
<tr>
<td></td>
<td>• Gluseen</td>
</tr>
<tr>
<td>August 2016</td>
<td>• Prototype Scalable Job Management</td>
</tr>
<tr>
<td></td>
<td>• Basic Dashboard</td>
</tr>
<tr>
<td></td>
<td>• Initial Collaborative capabilities</td>
</tr>
<tr>
<td>October 2016</td>
<td>• Scalable Job Management</td>
</tr>
<tr>
<td></td>
<td>• Turbulence</td>
</tr>
<tr>
<td></td>
<td>• Cosmology</td>
</tr>
<tr>
<td>November 2016</td>
<td>• Project 3 year Review</td>
</tr>
</tbody>
</table>
Login Portal
CasJobs: Query
Collaborative data-driven science

CasJobs: MyDB
SciDrive
What is SciServer Compute?

- Interactive Jupyter notebooks hosted inside Docker containers.
- Pre-configured images to create new containers from (R, Python, MATLAB, ...).
- High-bandwidth, low-latency access to other SciServer services and data sources through the notebooks.
- Users manage their own containers.
What are Docker Containers?

Type 1 Hypervisor

Type 2 Hypervisor

Linux Containers

Collaborative data-driven science
Data Storage Configuration

Collaborative data-driven science

- Docker Container
  - /home/idies
    - workspace
      - persistent
        - sdss_das
  - jupyter Notebook Server

- Persistent Storage
- SDSS DAS

User
Run asynchronous non-interactive jobs in separate Docker containers. It’s meant to be more than just Jupyter notebooks!

Create new VM nodes on-demand to accommodate growing number of users.

Provide scratch (temporary) storage space for working with large amounts of data.

Improve resource management.
Questions?
Architectural Challenges

- Define tradeoffs
  - Data Analytics system is different from supercomputer
  - What is the right balance between I/O and compute?

- Need high bandwidth to big data
  - Computations/visualizations must be on top of the data
  - Must support at least few 100TB per server
  - Petascale: 3 copies for production (or erasure code?)
  - Wide area data movement/backbone is hard

- Lessons from the database world:
  - It is nontrivial to schedule complex I/O patterns
  - For subsets we must use indexing, cache resilient storage