

DATA SHARING USING SCIDRIVE (*FORMER VOBOX*)

Dmitry Mishin, Dmitry Medvedev, Alexander S. Szalay
Johns Hopkins University

Agenda



- Overview
- System components
- Authentication
- Metadata extraction

SciDrive: scientific network storage



(renamed from VOBox)

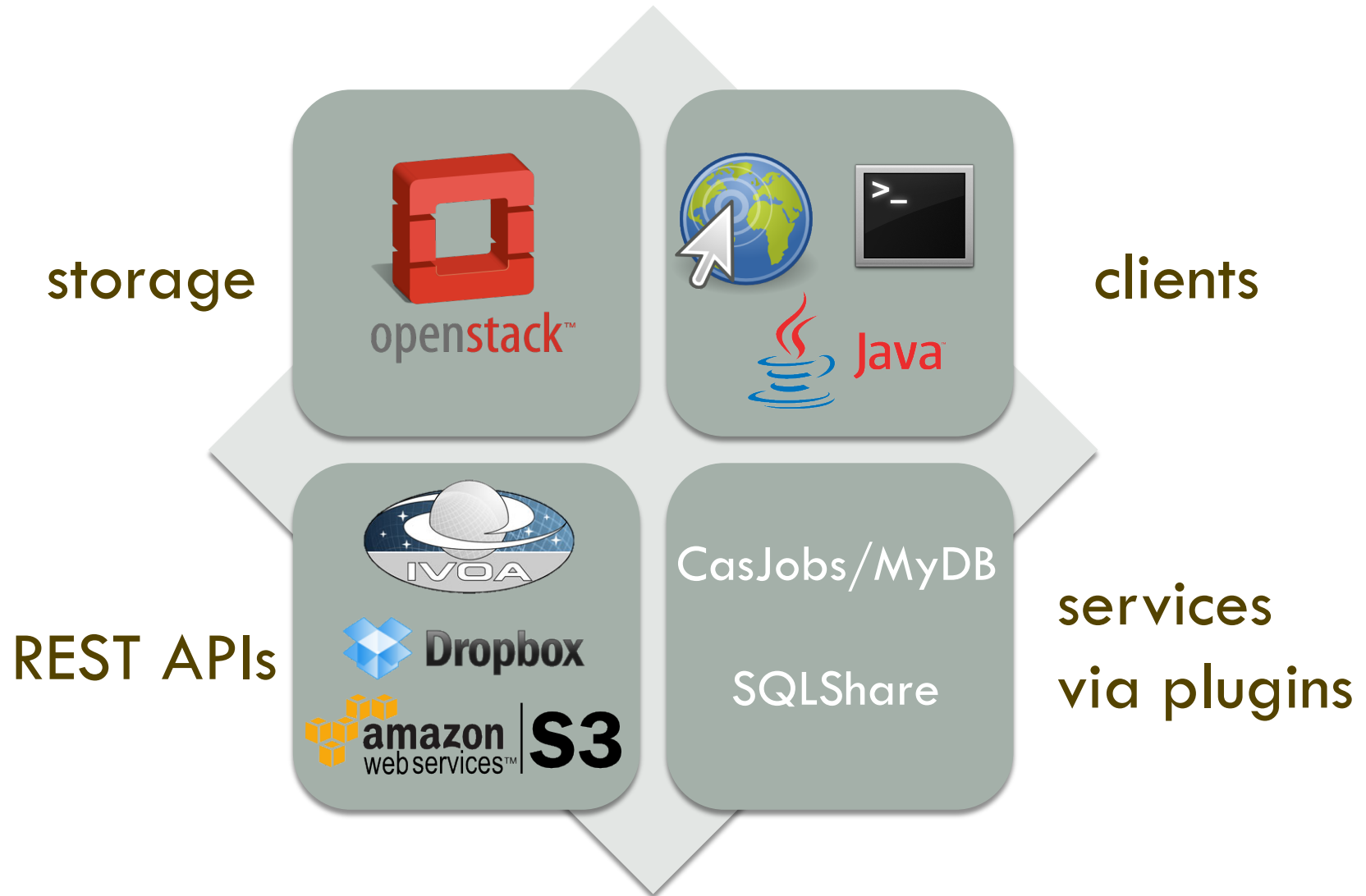
- Data storage REST service
- Data sharing and publication
- Metadata extraction
- Main focus: 'long tail' of datasets power law distribution

Public portal has been launched



<http://www.scidrive.org>

Components



Web portal

The screenshot displays the SciDrive web portal interface. At the top, the SciDrive logo is on the left, and navigation icons (refresh, folder, document, globe, link, calendar, and list) are in the center. On the right, there are 'Help' and 'Log out' buttons. Below the header, the current path is 'Root > test' and a search box contains '* JHU main'. The main area shows a table of files with columns for Name, Size, and Type. The file 'aapl.csv' is selected and highlighted with a blue background and a dashed border.

Name	Size	Type
109px-Tux-G2.png	12.1 KB	image/png
3copies.odt	57.2 KB	application/vnd.oasis.opendocument.text
aapl.csv	292.8 KB	text/csv
autocompleteObjects.xql	1.6 KB	text/plain
autocompleteSearchToJson.xsl	942 B	text/plain
CollectionsManager.java	3.1 KB	text/plain
csv	0 B	...
dataset.csv	14.5 KB	text/csv
laudens.doc	2.5 MB	application/msword
msft_prices.csv	38.9 KB	text/csv
nodes.xlsx	49.2 KB	application/vnd.openxmlformats-officedoc...
nodes_old.xls	31.5 KB	application/vnd.ms-excel
owls.csv	10.9 KB	text/csv
previous.totcloudpct.curr.18z.png	28.9 KB	image/png
ripe_20110605_nodes.csv	943.8 KB	text/plain
Screen Shot 2013-05-01 at 11.31.21 AM.png	51.9 KB	image/png
SL.55.051810.DVD.x86_64.disc2.iso	1.1 GB	application/x-iso9660-image

1 - 21 of 21 items 10 | 25 | 50 | 100 | All 1

Web portal: to use anywhere

- Upload/download data
- Containers sharing
 - ▣ Read-only / Full access
 - ▣ Anonymous / Group of users
- Files sharing
 - ▣ Get direct link to file or folder
- Metadata extraction configuration
 - ▣ Enable plugins
 - ▣ Choose containers for each plugin

Command-line client: to use on server



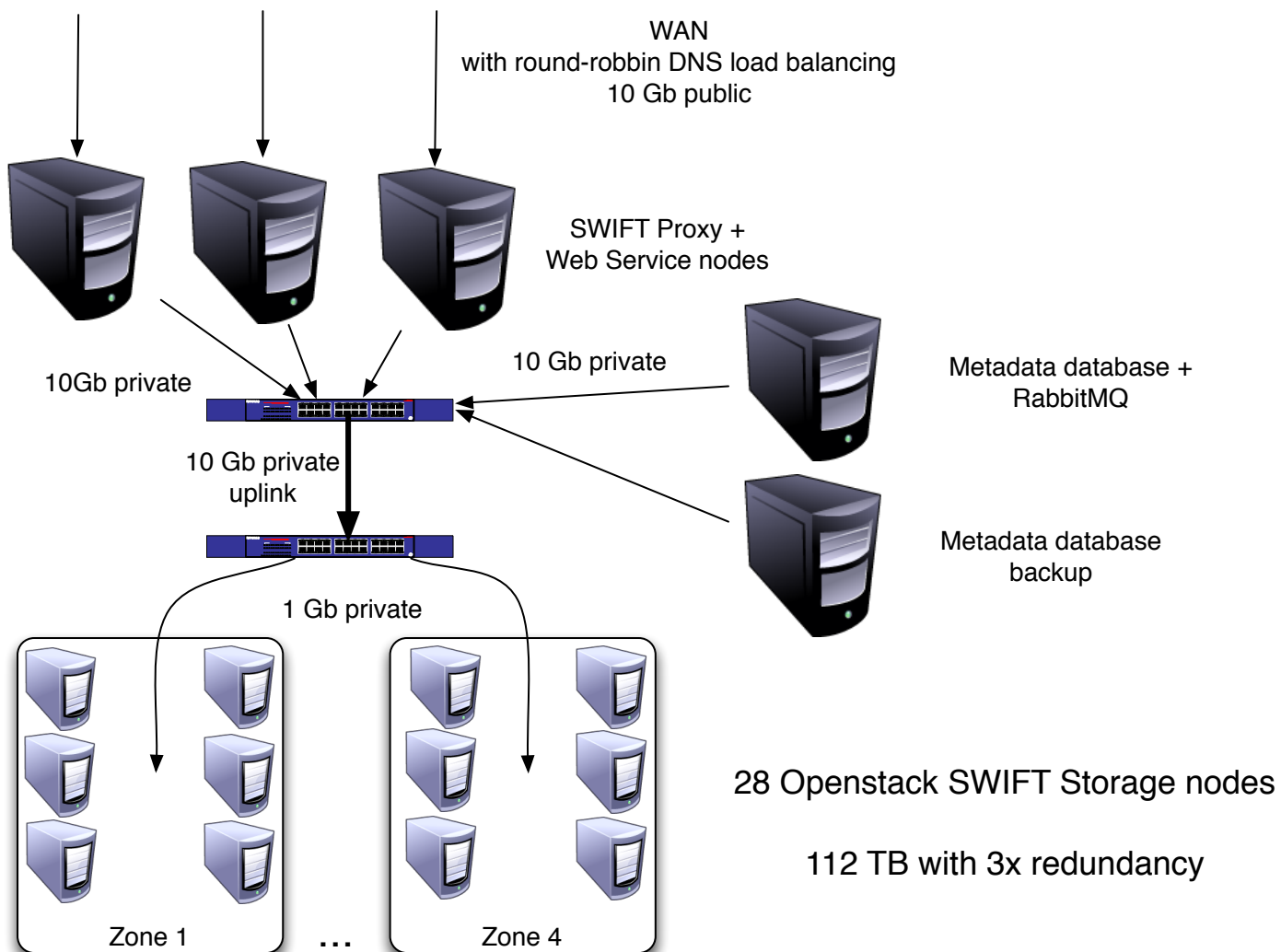
- Available at: <https://github.com/dimm0/scidrive-python-client>
- Built using Dropbox python API
- Can upload chunked data (experimental)
 - for files exceeding 4 GB

Java standalone client: to use on laptop



- Prototype version is released
- Two-way updates are supported
 - ▣ Events from web service
 - ▣ Events from filesystem
- Requires Java 7

Current cluster scheme



Authentication for data sharing



VAO SSO from NCSA

OAuth tokens



- Data access by owner
- Delegate permissions to other users
 - ▣ Provide access to root or container
 - ▣ Provide read-only or full permissions
- Different applications (“consumers”) can automatically get different permissions
- Tokens may have limited lifetime

Data sharing



1. In web portal:

- http://www.scidrive.org/scidrive.html?share=share_id

2. In client to get OAuth token:

- http://www.scidrive.org/vospace-2.0/request_token?share=share_id
- http://www.scidrive.org/vospace-2.0/authorize?share=share_id&provider=vao&oauth_token=token

3. Direct link to file or folder

- http://www.scidrive.org/vospace-2.0/data/random_id

Metadata extraction



- CSV, Excel, FITS (tables & images), TIFF, JPG
 - ▣ CasJobs/MyDB
 - Limited filesize!
 - ▣ SQLShare

- YT simulation archives
 - ▣ Simulation ID to local metadata

Acknowledgements

SciDrive has been funded with support from:

Alfred P. Sloan Foundation



ALFRED P. SLOAN
FOUNDATION

National Science Foundation ACI-1244820

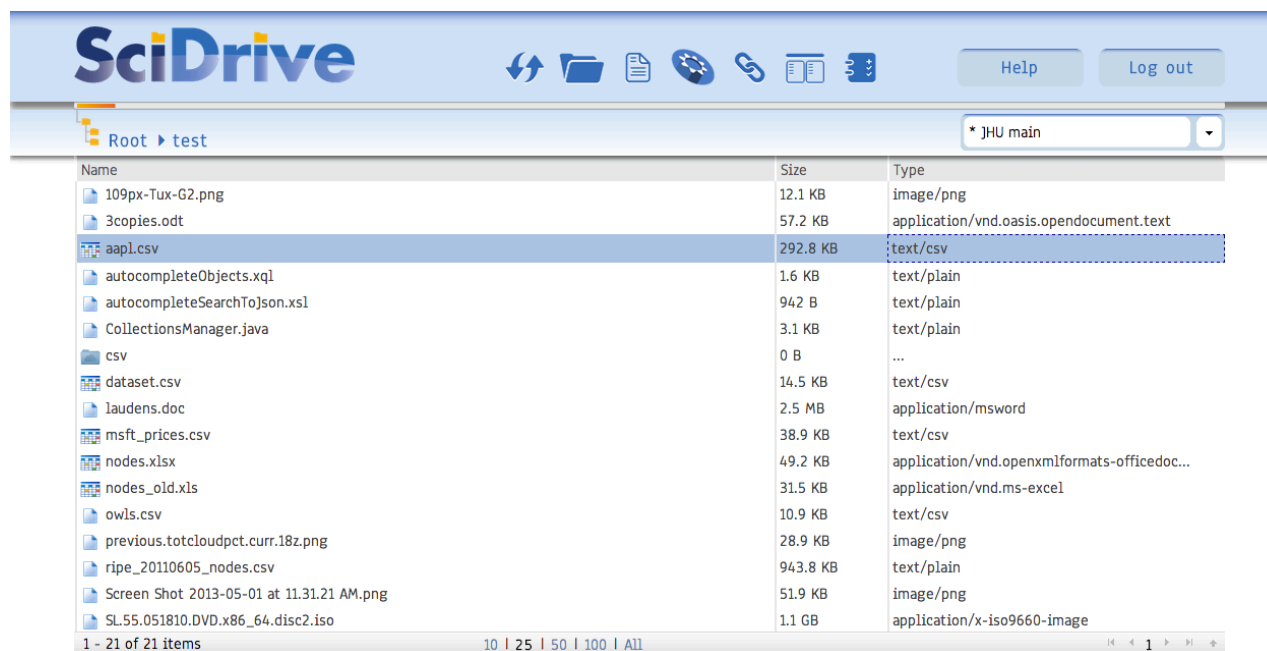


US Virtual Astronomical Observatory



Try it out

- Web portal: <http://www.scidrive.org>



- Source code:
 - <https://github.com/dimm0/scidrive>