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
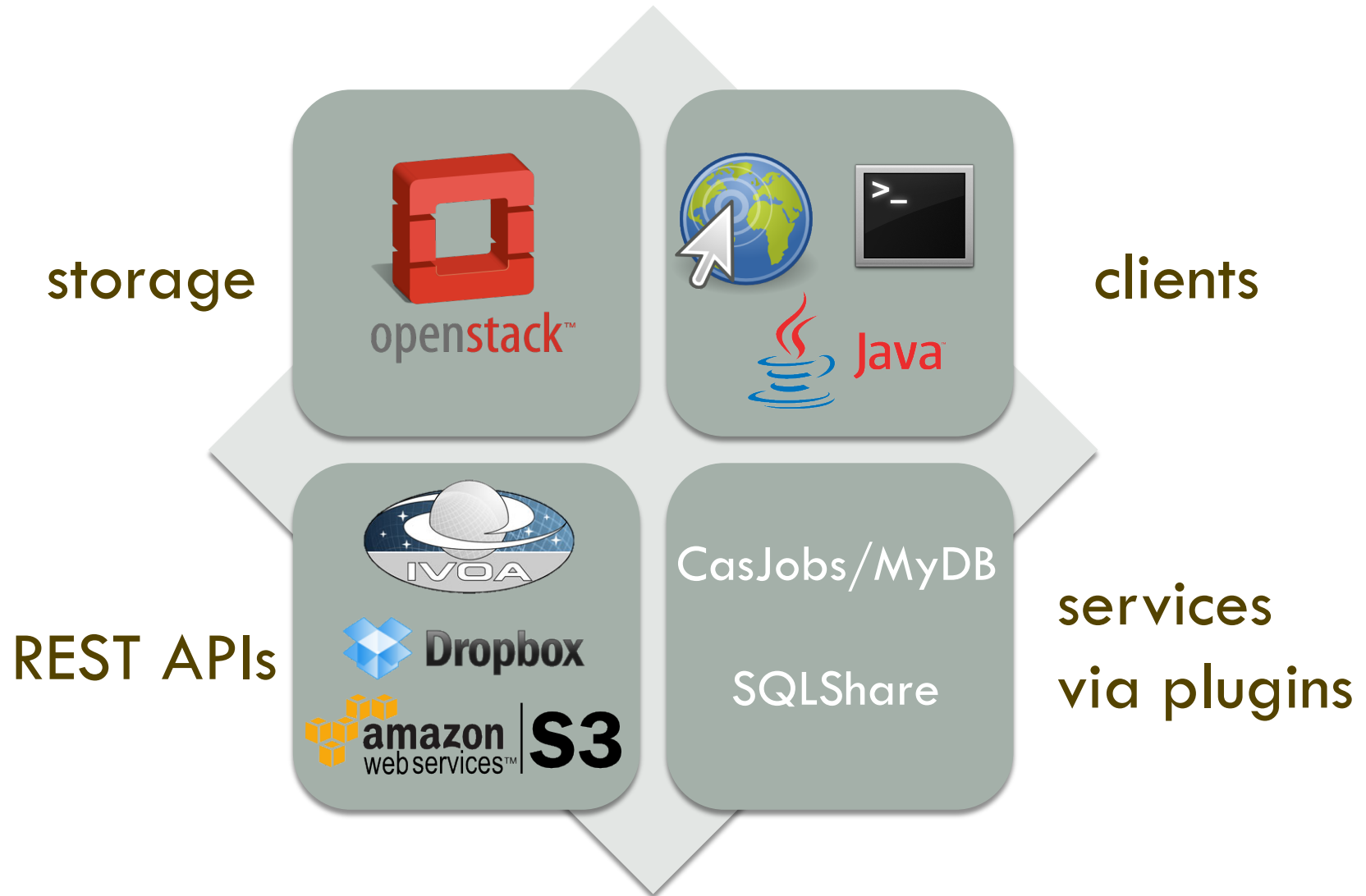
(renamed from VOBox)

- 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

storage

clients

REST APIs

services via plugins

CasJobs/MyDB

SQLShare

Web portal

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
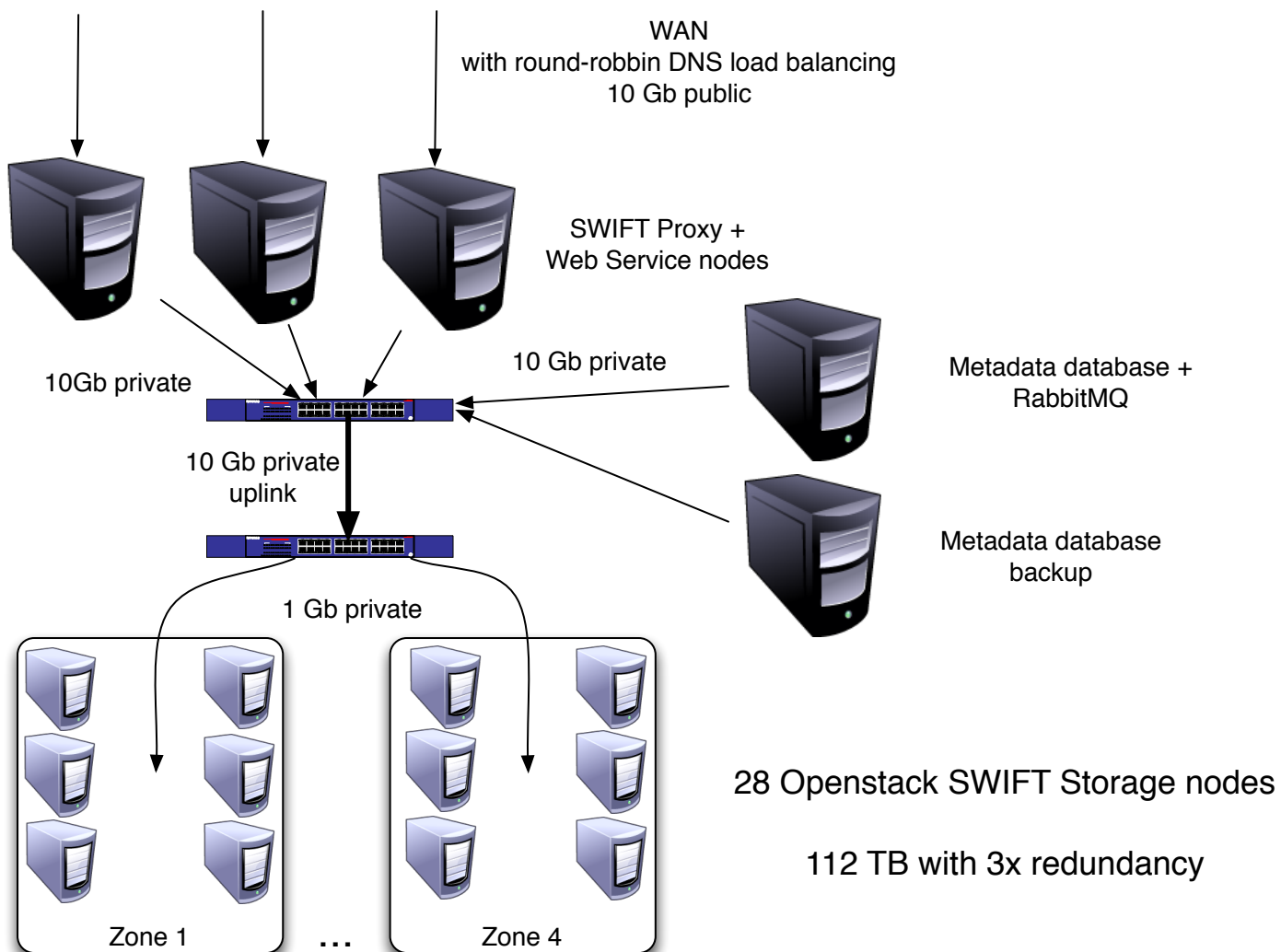
- Available at: https://github.com/dimm0/scidrive-python-client

- 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

WAN
with round-robbin DNS load balancing
10 Gb public

SWIFT Proxy +
Web Service nodes

10 Gb private

Metadata database +
RabbitMQ

10Gb private

10 Gb private
uplink

Metadata database
backup

1 Gb private

28 Openstack SWIFT Storage nodes

112 TB with 3x redundancy

Zone 1   …   Zone 4

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 permssions

- 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:

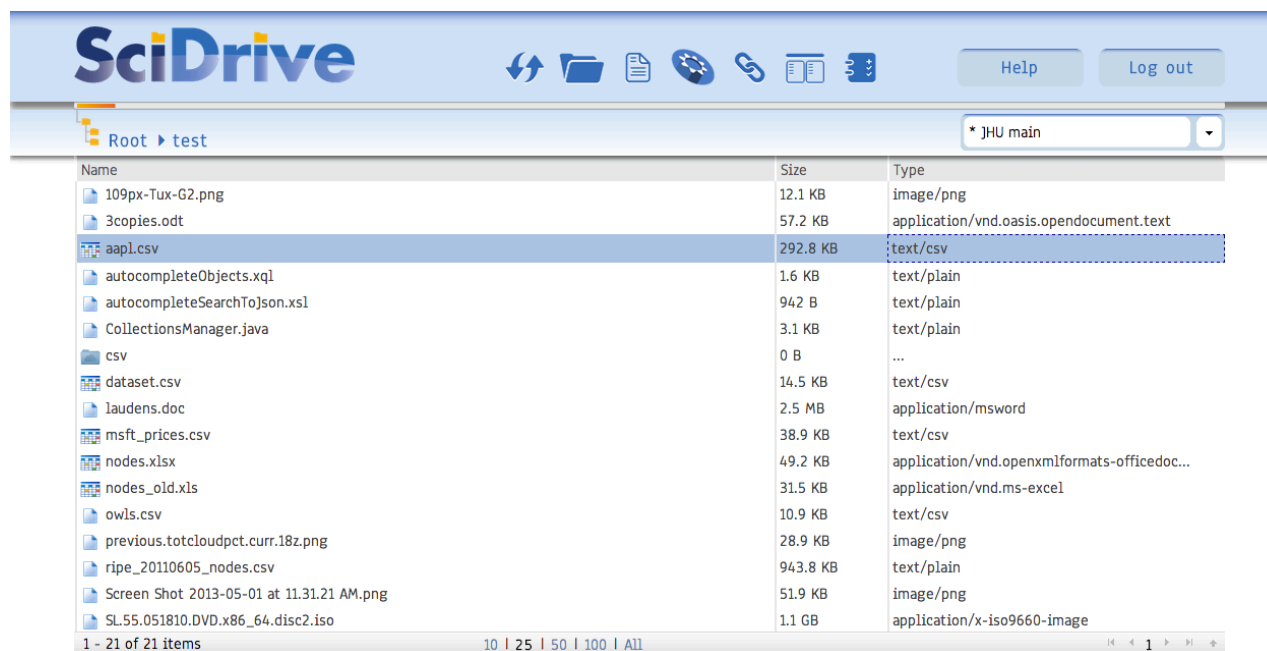Alfred P. Sloan Foundation

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