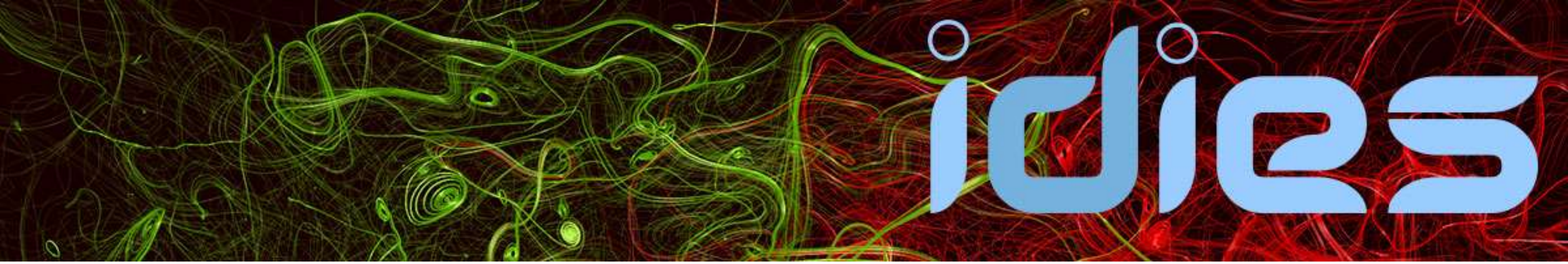


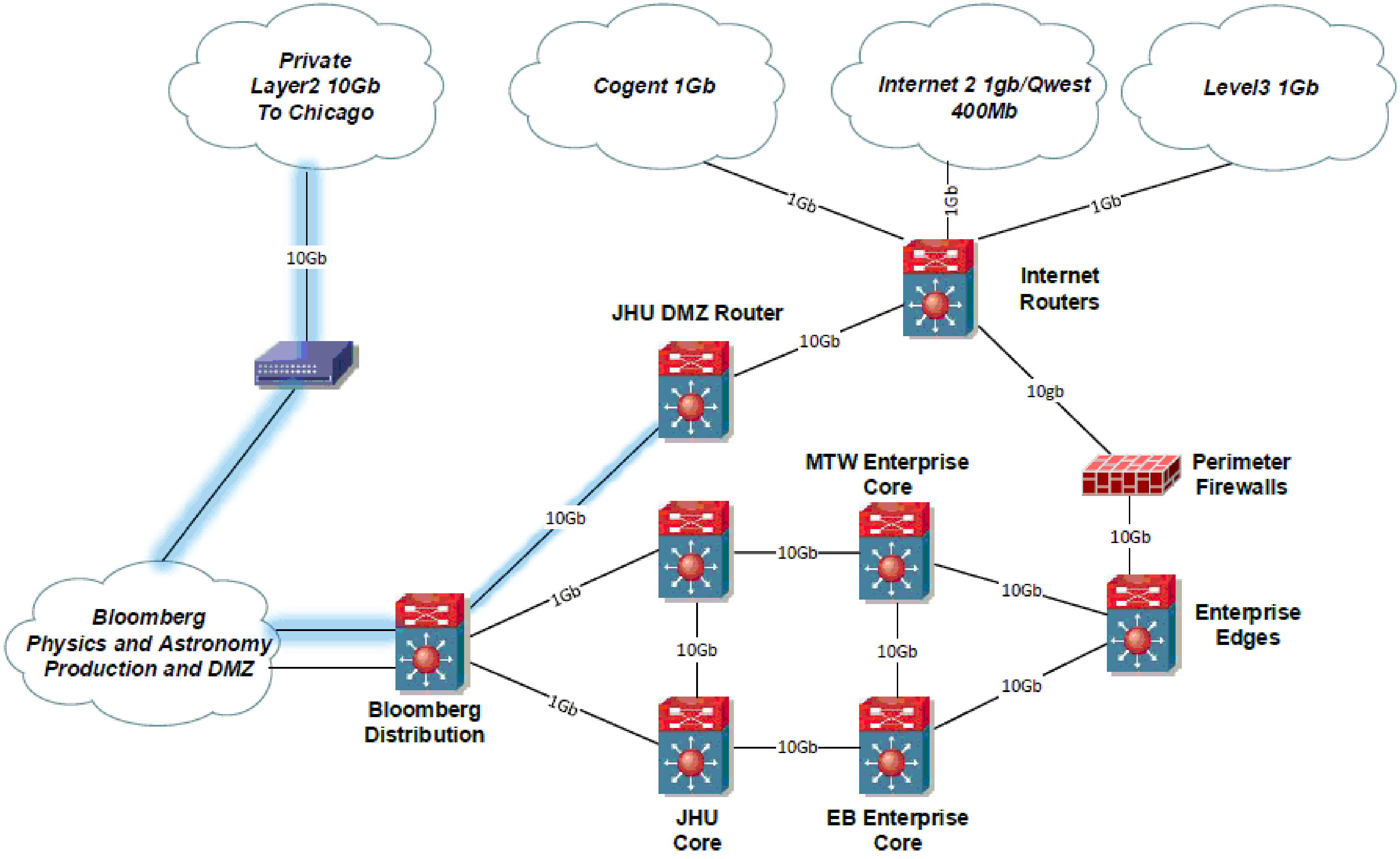
Institute for Data Intensive Engineering and Science Annual Meeting

2015

Infrastructure

- MARCC opening
- High speed networks
- Large data collections

Internet, Campus Core, DMZ and Physics and Astronomy Network Design
Johns Hopkins University
Enterprise Network Architecture and Design
March 2010



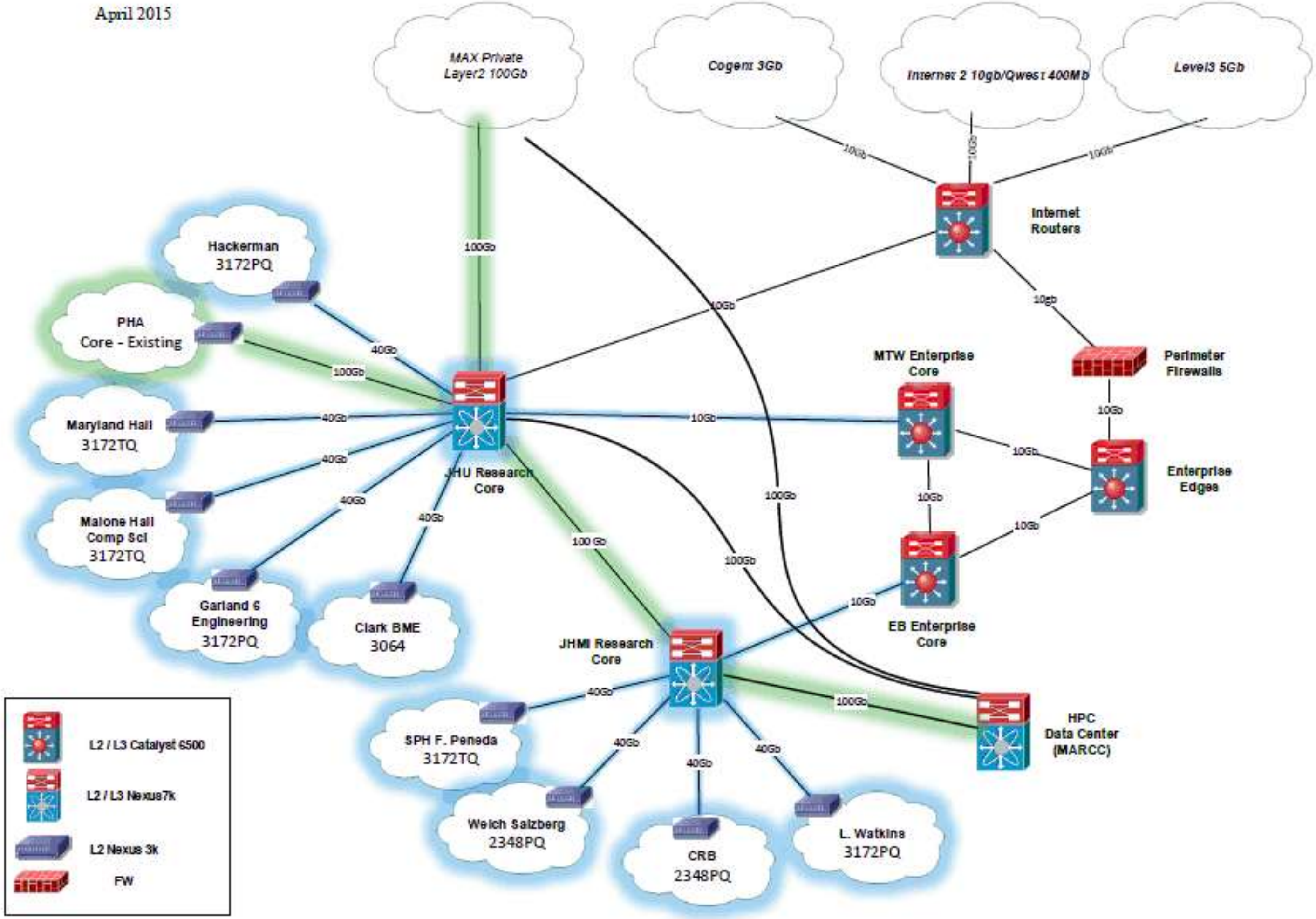
HorNet

HOPKINS RESEARCH NETWORK

JOHNS HOPKINS

UNIVERSITY & MEDICINE

April 2015

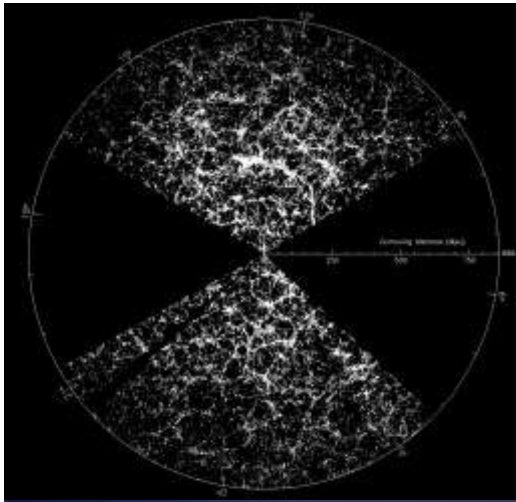


From SkyServer to SciServer

- SDSS
- TurbDB
- Cosmological N-body simulations
- Genomics
- NSF DIBBs => unify these and add new areas
- Work is starting in
    - Ocean Circulation
    - Materials Science
    - Smart Cities

Sloan Digital Sky Survey

“The Cosmic Genome Project”



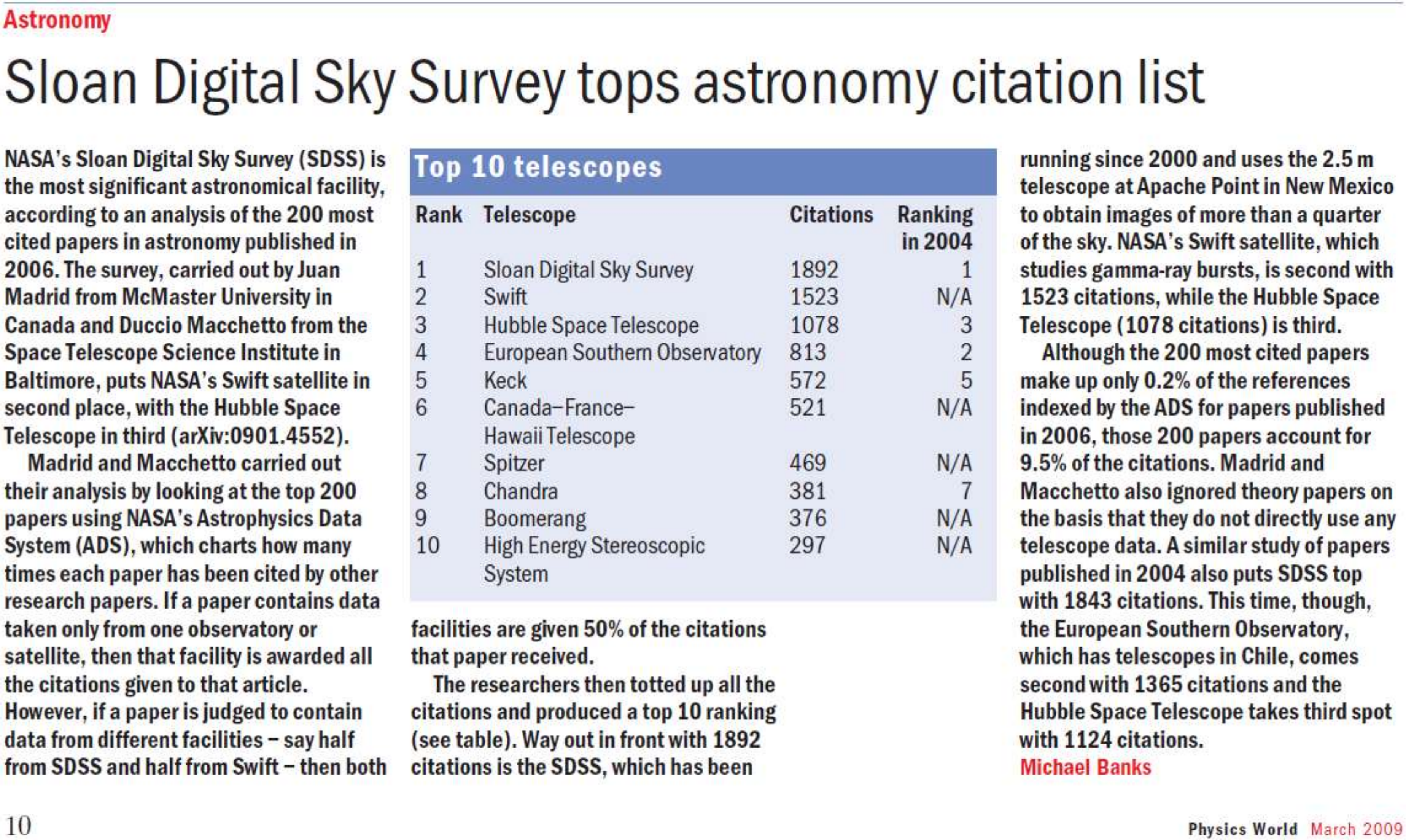
- Started in 1992, finished in 2008
- Data is public
  - 2.5 Terapixels of images => 5 Tpx of sky
  - 10 TB of raw data => 400TB processed
  - 0.5 TB catalogs => 35TB in the end
- DB+spectrograph built @JHU
- SDSS3/4 data served from JHU

Skyserver

Prototype in 21st Century data access

- Database centric computing
- 1.6B web hits in 12 years
- 271M external SQL queries
- 4,000,000 distinct users vs. 15,000 astronomers
- 5,000 refereed publications, 200,000 citations
- The emergence of the “Internet Scientist”
- The world’s most used astronomy facility today
- Collaborative server-side analysis done by 7K astronomers

Impact of the SDSS SkyServer



Astronomy

Sloan Digital Sky Survey tops astronomy citation list

NASA's Sloan Digital Sky Survey (SDSS) is the most significant astronomical facility, according to an analysis of the 200 most cited papers in astronomy published in 2006. The survey, carried out by Juan Madrid from McMaster University in Canada and Duccio Macchetto from the Space Telescope Science Institute in Baltimore, puts NASA's Swift satellite in second place, with the Hubble Space Telescope in third (arXiv:0901.4552).

Madrid and Macchetto carried out their analysis by looking at the top 200 papers using NASA's Astrophysics Data System (ADS), which charts how many times each paper has been cited by other research papers. If a paper contains data taken only from one observatory or satellite, then that facility is awarded all the citations given to that article. However, if a paper is judged to contain data from different facilities – say half from SDSS and half from Swift – then both facilities are given 50% of the citations that paper received.

Top 10 telescopes

Rank	Telescope	Citations	Ranking in 2004
1	Sloan Digital Sky Survey	1892	1
2	Swift	1523	N/A
3	Hubble Space Telescope	1078	3
4	European Southern Observatory	813	2
5	Keck	572	5
6	Canada–France–Hawaii Telescope	521	N/A
7	Spitzer	469	N/A
8	Chandra	381	7
9	Boomerang	376	N/A
10	High Energy Stereoscopic System	297	N/A

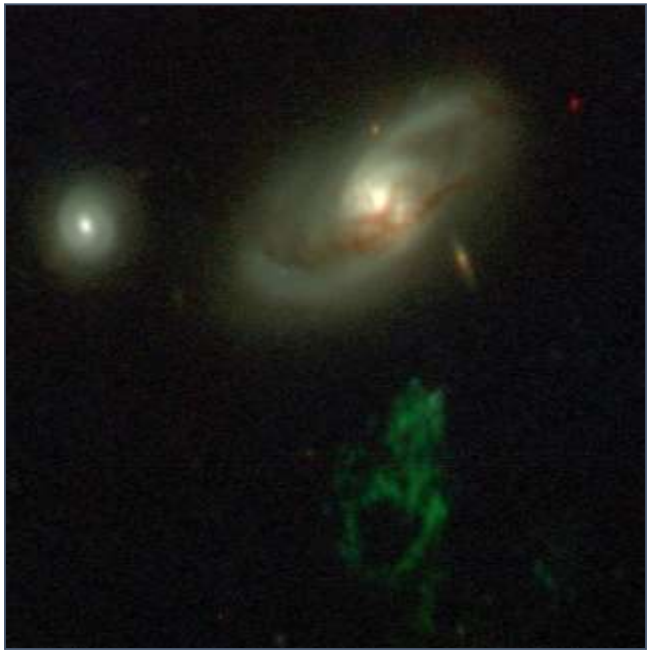
The researchers then totted up all the citations and produced a top 10 ranking (see table). Way out in front with 1892 citations is the SDSS, which has been running since 2000 and uses the 2.5 m telescope at Apache Point in New Mexico to obtain images of more than a quarter of the sky. NASA's Swift satellite, which studies gamma-ray bursts, is second with 1523 citations, while the Hubble Space Telescope (1078 citations) is third.

Although the 200 most cited papers make up only 0.2% of the references indexed by the ADS for papers published in 2006, those 200 papers account for 9.5% of the citations. Madrid and Macchetto also ignored theory papers on the basis that they do not directly use any telescope data. A similar study of papers published in 2004 also puts SDSS top with 1843 citations. This time, though, the European Southern Observatory, which has telescopes in Chile, comes second with 1365 citations and the Hubble Space Telescope takes third spot with 1124 citations.

Michael Banks

GalaxyZoo

- 40 million visual galaxy classifications by the public
- Good publicity (CNN, Times, Washington Post, BBC)
- 300,000 people participating, blogs, poems…
- Original discoveries by the public (Voorwerp, Green Peas)

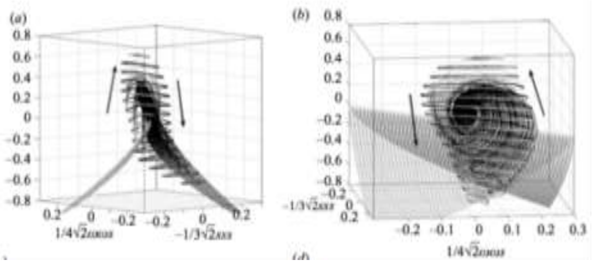


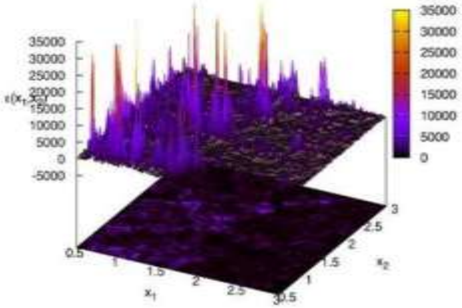
- *Chris Lintott et al*

Immersive Turbulence

"… the last unsolved problem of classical physics…" Feynman

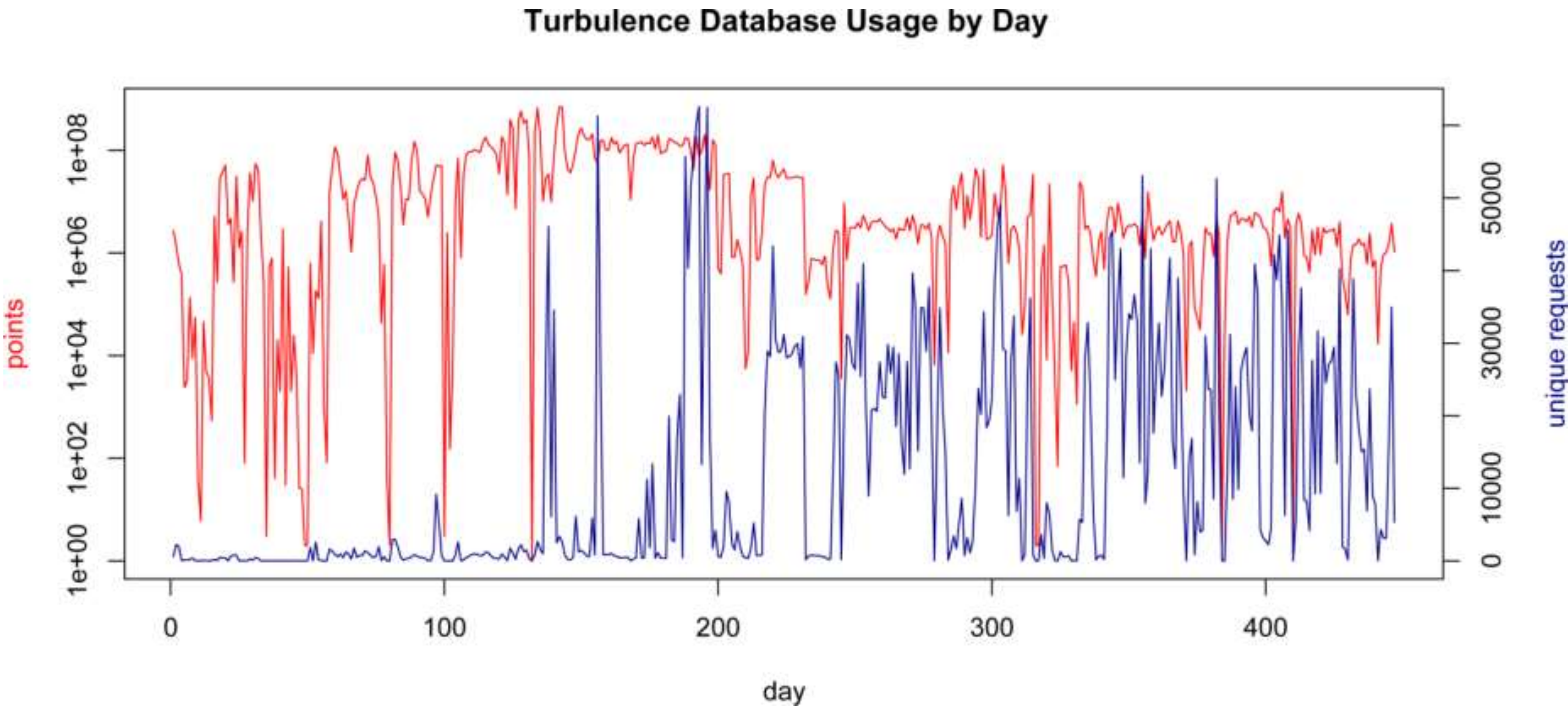
- **Understand the nature of turbulence**
  - Consecutive snapshots of a large simulation of turbulence: 30TB
  - Treat it as an experiment, **play** with the database!
  - **Shoot test particles** (sensors) from your laptop into the simulation, like in the movie Twister
  - 50TB MHD simulation
  - Channel flow 100TB, MHD 256TB
- **New paradigm** for analyzing simulations!



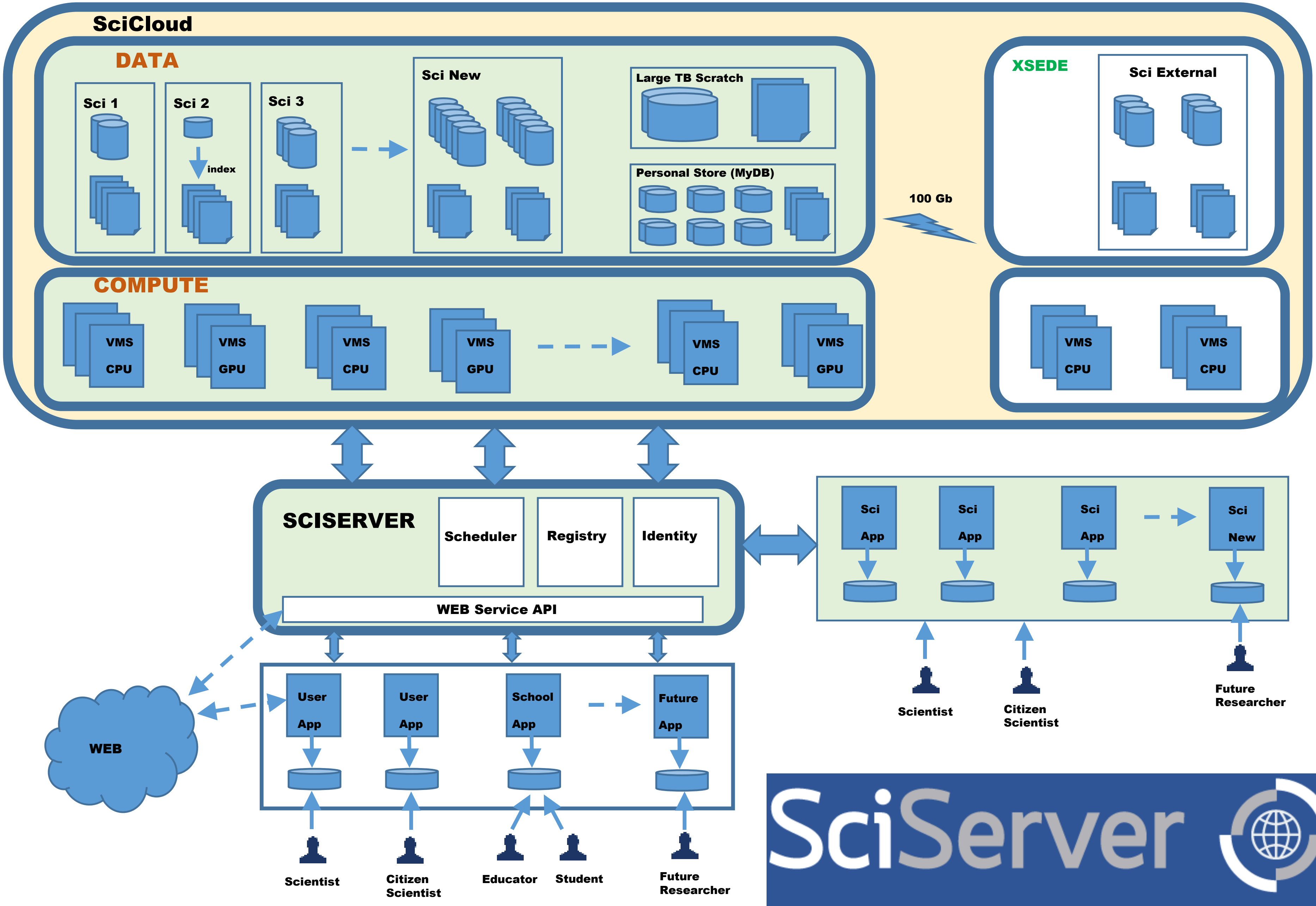


R. Burns, C. Meneveau, T. Zaki, G. Eyink, A. Szalay, E. Vishniac

Daily Usage



2015: exceeded 14T points, delivered publicly idies



SciCloud

DATA

Sci 1

Sci 2

index

Sci 3

Sci New

Large TB Scratch

Personal Store (MyDB)

100 Gb

XSEDE

Sci External

COMPUTE

VMS CPU

VMS GPU

VMS CPU

VMS GPU

VMS CPU

VMS GPU

VMS CPU

VMS CPU

SCISERVER

Scheduler

Registry

Identity

WEB Service API

WEB

User App

User App

School App

Future App

Scientist

Citizen Scientist

Educator

Student

Future Researcher

Sci App

Sci App

Sci App

Sci New

Scientist

Citizen Scientist

Future Researcher

SciServer

iPython and Jupyter

- iPython Notebook
  - Web based interactive scripting
  - Kernel runs on the server -- direct, local access to data
  - Client runs in any browser
- Jupyter
  - Grew out of iPython
  - Can run many types of kernels
  - Python, R, bash, octave, matlab, etc.

New Datasets on the Horizon

- N-body simulations
    - INDRA – now 400TB, soon 1.1PB, running at NERSC
    - JHU now part of the Virgo Consortium
- Turbulence
    - Channel flow
    - RMHD
    - Rotating fluid
- African American Genomes
    - 1000 genomes

Seed Projects

- Several new seed projects
- Program continues, next round due end of Jan 16

New Opportunities

- Space Science
    - Proposals submitted for large archives
    - Pan-STARRS, HSC/PFS
    - WFIRST
    - Strategic partnership with STScI
- Genomics
    - Several new opportunities
- New faculty hires
    - Two new Bloomberg Professors in Big Data arriving in January

“…it takes all the running you can do,
to keep in the same place…”

-- Lewis Carrol