# *Harry Mangalam*
Research Computing
OIT / UCI

I am a continually Dissatisfied User.

My Drivers

- How to provide the maximum benefit to researchers.

- As Easily as possible (for them).

- As Quickly as possible.

- As Cheaply as possible.

- Using mostly (GRAM) Open Source Software.

E

- BSc & MSc [UBC]
  - DEC MINC-11 La
  - Peak Detection,
    Software in Fort
- PhD [UCSD] Ge
  - Interests in prog
- PostDoc [Salk In
  - Mac, Windows,
    Internet, Gophe
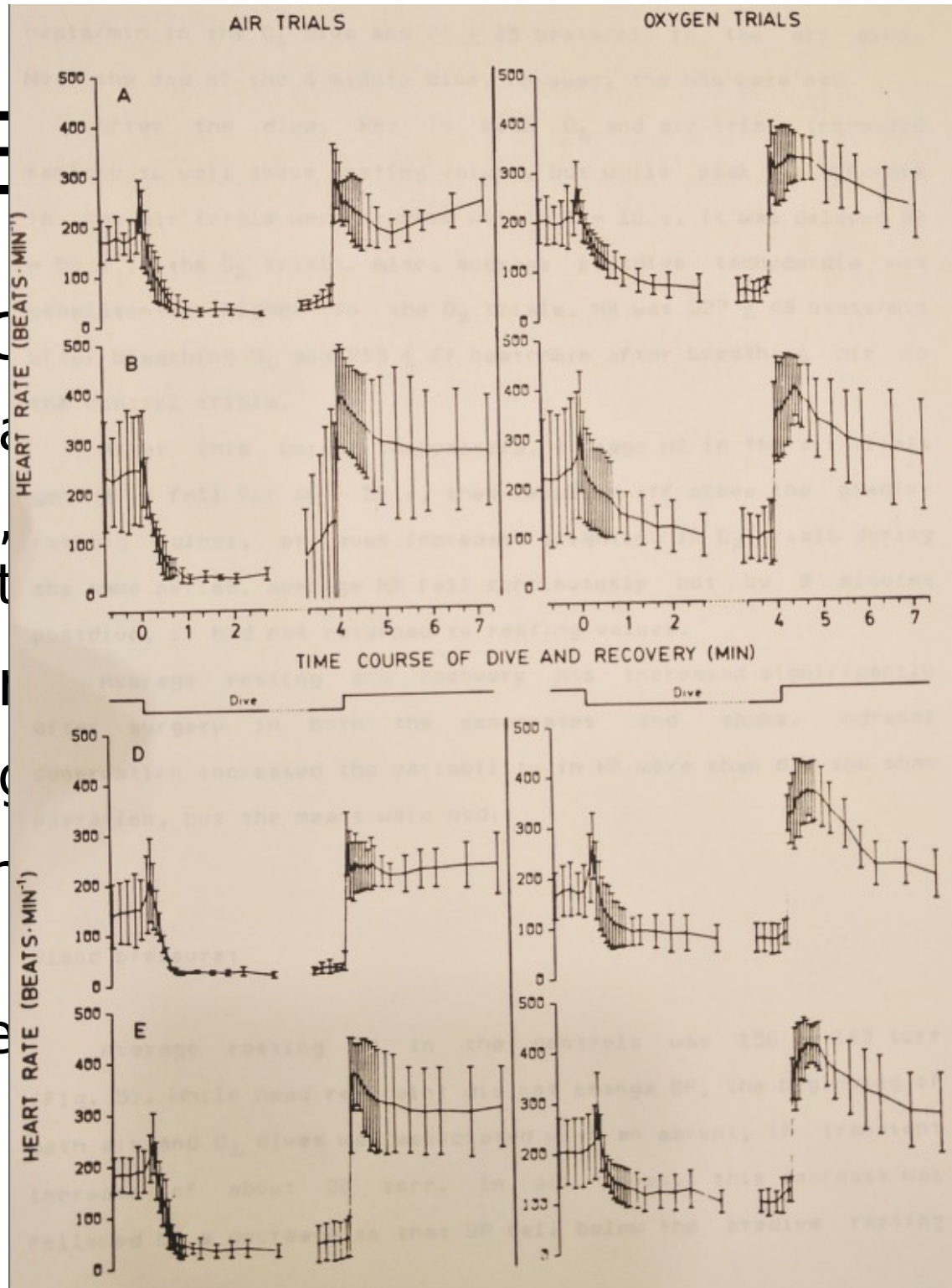
Other Background

- NCGR: GeneX

- Independent Software Developer

- Acero: Commercial Object DB

- UCI/ESS: profiling optimizing code, how SW works.

Sc...

- tacg*

- GeneX*

- nco profiling*

- clusterfork

- scut, cols, sta...

- parsync – self...

- tnc – tar 'n' n...

- katyusha (cur...
  data transfer

http://moo.nac.uci.edu/~hjm/HOWTO_move_data.html#tnc

Harry's HOWTOs

9.7.2. tnc

From the contraction of *tar n netcat*. This is not a novel idea (see Varun Patil's script, but this wrapping tries to be more flexible and comprehensive than most other implementations. Let me know how to make it better.

both push and pull data transfers will be *tar files* (compressed with 'xz' if you use the '--compress' option.)

It can be used to both push and pull data from a remote connection, using only the static IP end.

kdirstat-cache-writer (included in the tarball mentioned above), requires a non-default Perl utility: URI::Escape qw(uri_escape)

```
sudo yum install perl-URI    # CentOS-like

sudo apt-get install liburi-perl  # Debian-like
```

parsync needs to be installed only on the SOURCE end of the transfer and uses whatever *rsync* is available on the TARGET. It uses a number of Linux- specific utilities so if you're transferring between Linux and a FreeBSD host, install parsync on the Linux side. In fact, as currently written, it will *only PUSH data to remote targets*; it will not pull data as rsync ... This will probably in the near future.

http://moo.nac.uci.edu...hjm/parsync/#_download

netCDF OPERATOR (NCO) SCALING

Invited talks

- Basel Life Sciences (2016)
  - Title: Storage for Inforgs

- Supercomputing16
  - Title: BeeGFS in real life (BigData BOF)

Previous Grants

- Salk Institute [MRC]: Postdoctoral Fellowship

- UCI School of Medicine: [Pacific Bell/CalREN]:
  - Telemedicine over ATM
  - 1st MBONE telecast from LBVA.

- NCGR: [NSF] GeneX

OIT Grant & Dev Efforts

- Equipment Donations: [TGMS, HGST]
  - QDR IB enterprise switch, 4 tape robots, multiple large servers, 7 racks of compute servers, NVME cards
- OIT: [NSF] Cyberinfrastructure Engineer
  - Joulien!
- OIT: [UCI] RCIC Proposal

Documentation Examples

- Cyberinfrastructure
  - UC Irvine CyberInfrastructure Plan - 2013
  - A Model Outline for Research Computing
  - How to move data.*
  - The Storage Brick:
    Fast, Cheap, Reliable Terabytes
  - The Perceus Provisioning System
  - Distributed Filesystems: Fraunhofer vs Gluster
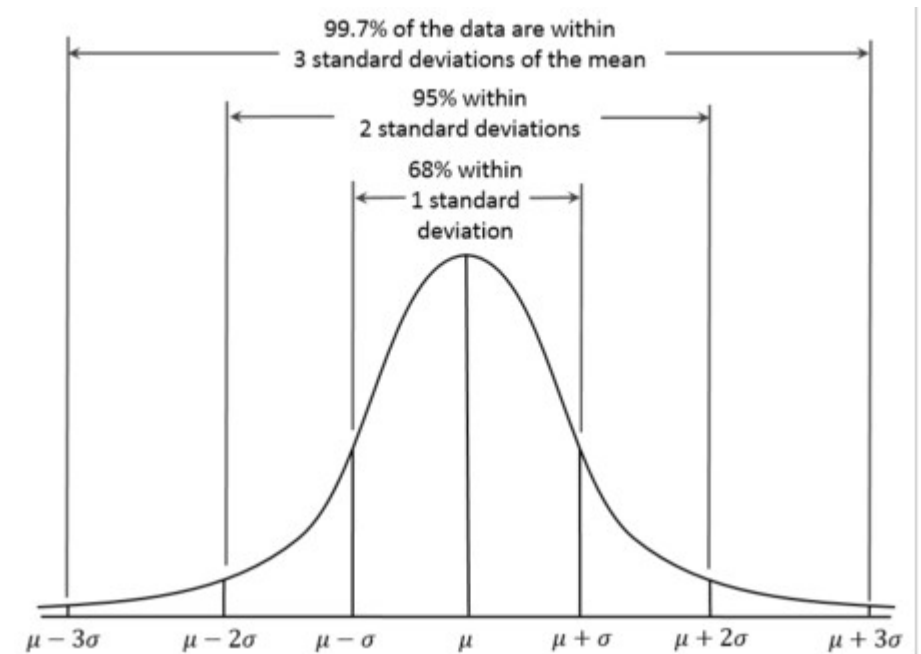
Teaching / Instruction

- BigData Hints for Newbies

- BigData on Linux (Data Science slides)

- Introducing Linux on HPC (PDF Slides)

- A Linux Tutorial for HPC

- Manipulating Data on Linux

Open Source Software

- How to Evaluate Open Source Software
- Open Source and Proprietary approaches in Municipal Information Technology.
- Setting up an LTSP Thin Client System
- Mind Your NegaBit$

Do I fit with UCI?

- Academic, Non-Profit, Solo, & Commercial experience
- Improvements from the User's Perspective.
- '4 Σ' approach vs only the top end.
- 'Catalytic Programming'.
- Some familiarity with UCI.
- Demonstrated strengths in critical areas, especially grants and hardware.

99.7% of the data are within 3 standard deviations of the mean

95% within 2 standard deviations

68% within 1 standard deviation

$\mu - 3\sigma \quad \mu - 2\sigma \quad \mu - \sigma \quad \mu \quad \mu + \sigma \quad \mu + 2\sigma \quad \mu + 3\sigma$

Immediate Priorities

- Hiring good people, esp at PA 1&2, students
- Optimize how the RCIC budget is allocated and spent.
- Change responsibilities; higher PAs addressing appro tasks.
  - re-architecting clusters, schedulers, overall integration
  - assisting with code porting, profiling, optimization
  - addressing research sysadmin problems (w/ EUS)
- Aggressive outreach to UCI Faculty, Depts
  - Meeting with Senior Leaders for 10m intro to RCIC
- Grants applications, coordinated with faculty, Public & Private
- Campus Storage Pool.
- 'Data Days' – 2 headliners, lightning talks, panels, prizes.

Coming Challenges

- Secure Computing
- Continuous review of new technologies:
  - Flash, Xpoint memory
  - Omnipath, >10GbE
  - FPGAs, GPUs, new CPU arch's
  - Filesystems
  - Containers for apps & analysis provenance
  - cloud technologies
- Better Coordination with other UCs

More Challenges

- Assuring and expanding RCIC funding..
- RCIC should expand in the following ways:
  - More **computation**, at least 2x current cores
  - More and faster **storage**, esp hybrid/flash
  - More **usable network services**
  - more **secure** networking via cheaper, faster defenses.
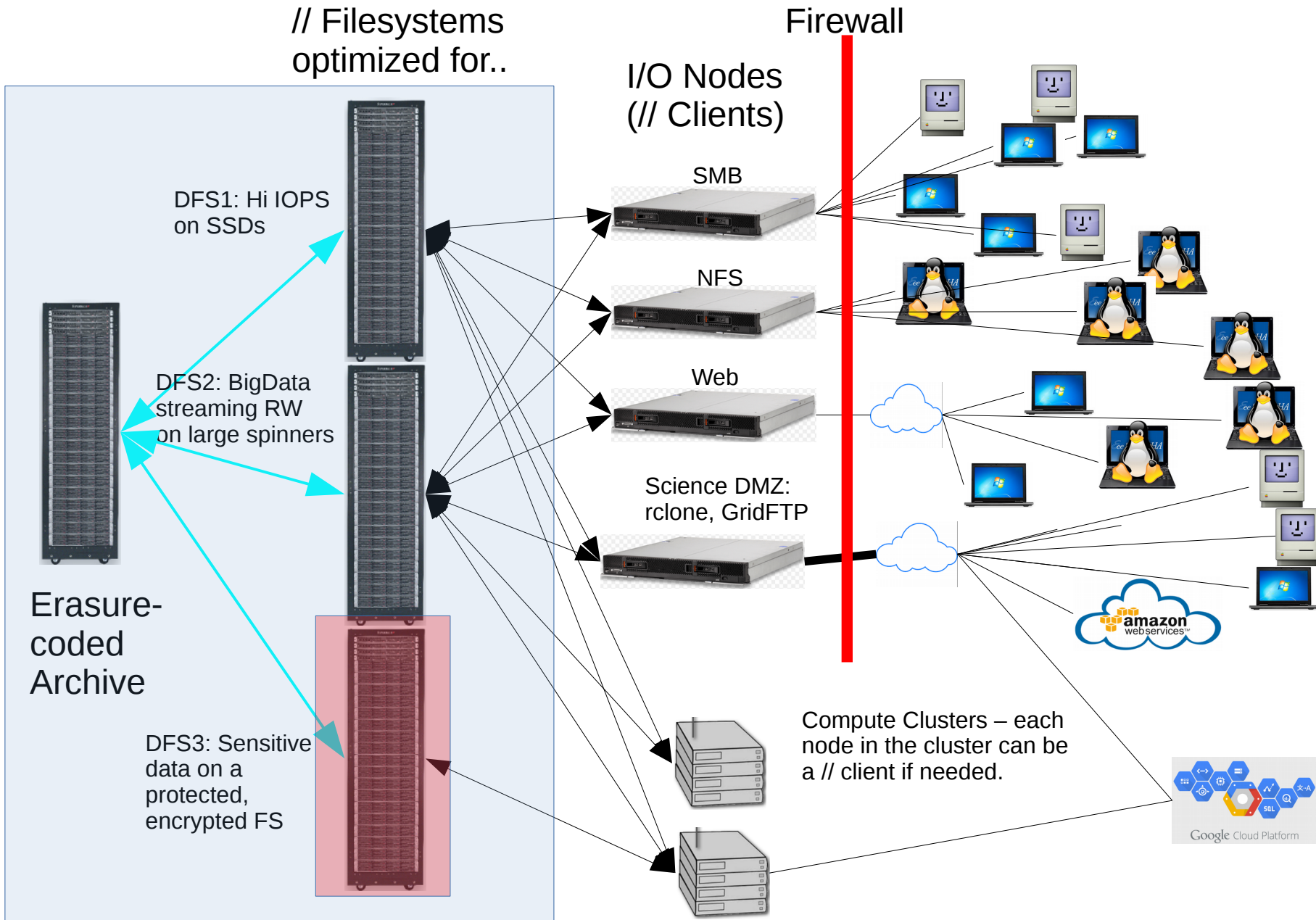 - ***More direct assistance & involvement with researchers***

Good Judgment comes
from Experience.
Experience comes
from Bad Judgment.

Questions?

Appendix Slides

UCI Campus Storage Pool

Back End

// Filesystems optimized for..

rclone, web

**HGST AA?
Ceph?
DDN WOS?
LizardFS?
MozoFS?**

**Erasure-
Coded,
Multi-tenant
Object
Archives**

**DFS1: Hi IOPS
on SSDs**

**DFS2: BigData
streaming RW
on large
spinners**

**DFS3: Sensitive
data on a
protected,
encrypted FS**