# Using HIPE remotely:
# NHSC Remote Computing
# and running batch jobs

Presenter: Dave Shupe, NHSC

with Jeff Jacobson
and many other contributors at NHSC

# Using HIPE Remotely

- Capabilities of the NHSC service

- Lifecycle of an account request

- Tips for using the remote accounts

# Using HIPE Remotely

- **Capabilities of the NHSC service**
  - Virtual machine accounts
  - Installed software


- Lifecycle of an account request


- Tips for using a remote computer

# The remote computing facilities can supply large resources to many users

- Physical servers with large capacity
  - 3 machines with 64 GB memory
  - 2 machines with 128 GB memory
  - Large server with 384 GB memory

- A virtual machine is created for each request
  - Secure (destroyed after use)
  - Each VM is tailored
  - Memory, disk, # cpus are adjustable

- ssh / scp access

# Key Herschel software is already installed or is available by request

- **Available now**
- HIPE
- Ds9
- Topcat
- Unimap
- SCANAMORPHOS
- IDL

- **Coming soon**
- SANEPIC
- TAMASIS
- SExtractor
- Anaconda Python
- SIMPLE
  - *(see PACS tomorrow am)*

# Using HIPE Remotely

- Capabilities of the NHSC service

- Lifecycle of an account request
  - Helpdesk ticket submission
  - Getting started
  - Data transfer
  - Finishing up

- Tips for using a remote computer

# Submit a ticket to the Remote Computing department of the Helpdesk

Your account request ticket asks for details to help us make the accounts

- Description of datasets
  - Obsids or program names
- Memory and disk space estimates
- HIPE version needed
- Other mapmaking software needed
- Requested starting date and duration

# Your account request will be reviewed and login information sent to you

- Instrument team review
  - Refine memory and disk space estimate
  - Assign a liaison
- Account creation
- Credentials communicated to you
- Login with X-windows enabled
  - Mac OS X: ssh –Y nhscvXX@134.4.YY.XX
  - Linux: ssh –X nhscvXX@134.4.YY.XX

# Choose one of these 3 recommended options for transferring data

- To/from your own computer: *scp* or *sftp*

- From the HSA: Retrieve tarfile
  - Shopping Basket -> Download (*check*)
  - On the remote machine: use *wget* with the link that the HSA emails to you

- Download on-the-fly: Use MyHSA
  - HIPE Edit -> Preferences -> MyHSA -> Advanced -> Save data on-demand
  - In scripts use *getObservation* & *useHsa=True*

When your estimated ending date nears, we will contact you for finishing up

- Ideally, you'll be all done and we can delete your account

- If more time is needed, we can negotiate a new ending date

- If you are taking a break for a week or longer, we can hibernate the account for a limited time

# Using HIPE Remotely

- Capabilities of the NHSC service

- Lifecycle of an account request

- Tips for using a remote computer
  - Virtual Network Computing
  - How to run batch scripts
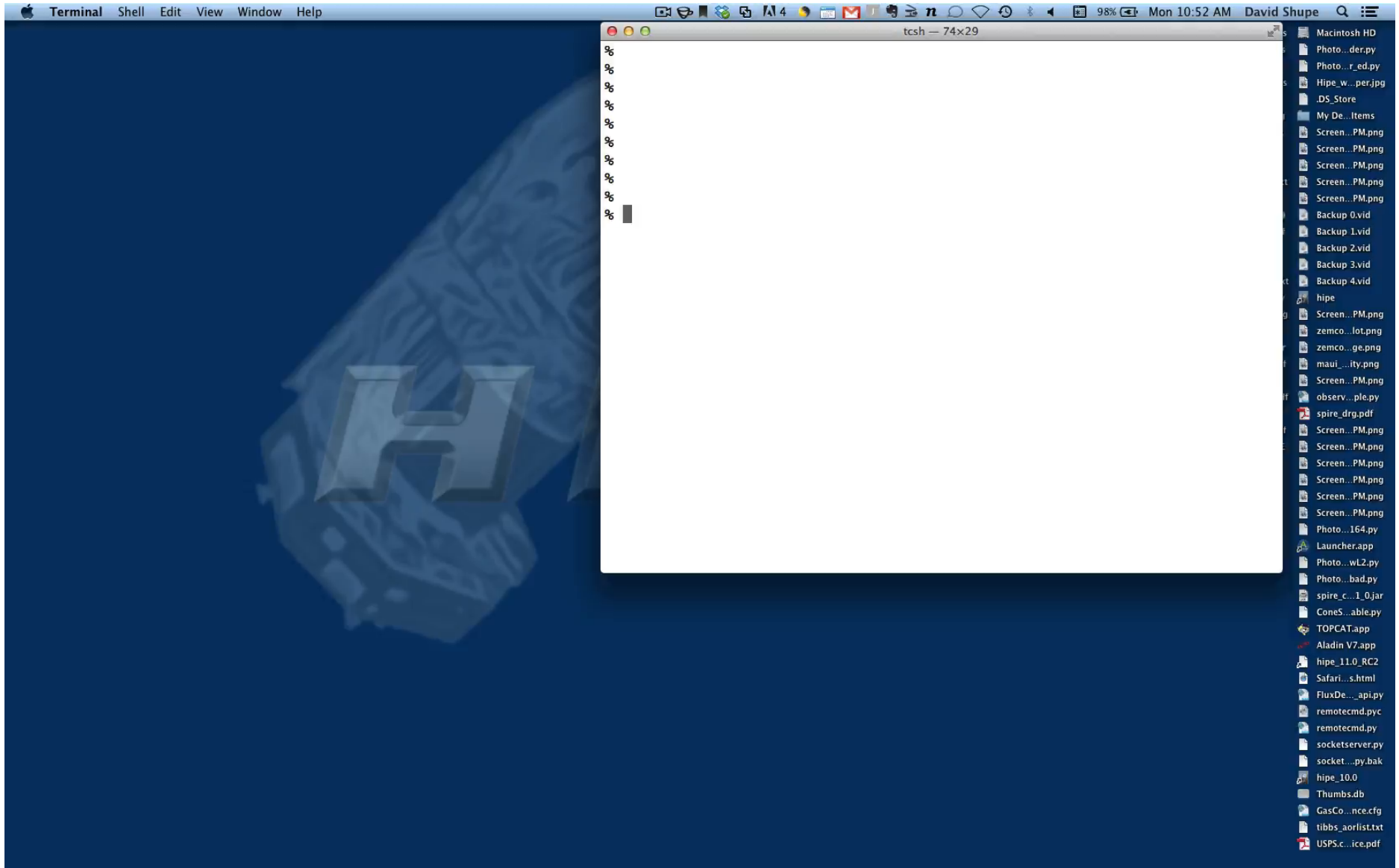  - Best practices for batch scripts

# Virtual Network Computing (VNC) is provided on all accounts

- Virtual Network Computing is now installed and pre-configured on our remote computing accounts
https://nhscsci.ipac.caltech.edu/sc/index.php/ExternalUser/VNC

- VNC provides X-windows sessions that can be closed down and reconnected to later
  - HIPE keeps running,
    even when user is not logged in
  - User can reconnect from a different machine to check on processing progress

# A few tips will help you use VNC

- Vnchipe to start

- Vncjylaunch for batch scripts

- Old window manager (twm)

- Vncremove to clean up prior sessions

- Not needed when you're at IPAC

# VNC demo movie

Batch scripts are run using
"hipe <scriptname.py>"

- The HIPE graphical environment is suppressed in this mode

- Output is streamed from your terminal

- To use VNC, use "vncjylaunch" instead

# Best Practices for remote computing (Babar Ali)

- Avoid GUIs as much as possible
  - Stick to heavy-processing in scripts
- Edit and debug the script on a local machine
  - Using a small test dataset
- Convert images to jpegs for viewing
- Create logfiles
  - Unix: hipe myscript.py >& session.txt
  - Python: redefine sys.stdout to a file
- Process in stages
  - Break at logical points with results saved

# Backup slides

# Common pitfalls and problems

1. Default configuration settings can fail for large jobs
   – Temporary directory can be too small
   – Java PERMSIZE can be too small
2. Plots can stop processing
3. Sessions can time out
   – Firewall timeout of ssh session
   – Internet dropout
4. Network can be slow

*1: Addressed by pre-configuring accounts with a script*

*2, 3: Addressed by careful scripting before, now by Virtual Network Computing*

*4: Still a drawback*

# Remote Services
# Conceptual view