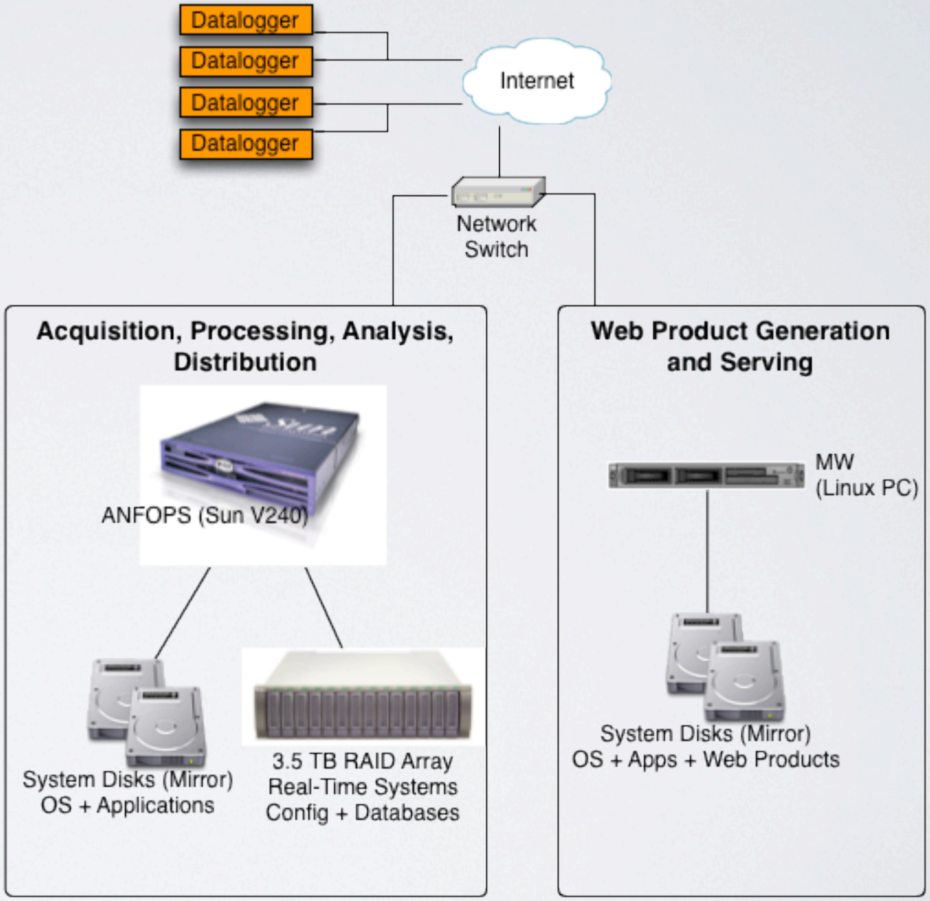
ANF SYSTEMS ARCHITECTURE

Antelope Users Group 2012
University of Nevada, Reno, NV

WHERE WE WERE

- Mid-2005
  - Single Sun V240 with 3.5 TB storage
  - SIngle Linux server for Web site
  - Single Datacenter


Datalogger
Datalogger
Datalogger
Datalogger
Internet
Network Switch
Acquisition, Processing, Analysis, Distribution
ANFOPS (Sun V240)
System Disks (Mirror) OS + Applications
3.5 TB RAID Array Real-Time Systems Config + Databases
Web Product Generation and Serving
MW (Linux PC)
System Disks (Mirror) OS + Apps + Web Products

INFRASTRUCTURE AT MAIN DATACENTER

- 3 Dell R710 x86_64 servers,
    - 192 GB RAM, 512GB mirrored root disks, 2 port Fibre Channel HBA
- Redundant Cisco Switches
- Cisco VPN Gateway, Serial Console Server for Mgmt.
- 3 Apple Xserves
- Redundant QLogic SAN Switches
- Compellent Storage System for Block Storage
- Shared data via NFS
- Legacy Sun Hardware - 3 T5220 systems, 1 T2000 (for network monitoring)

VIRTUAL MACHINES AND ZONES

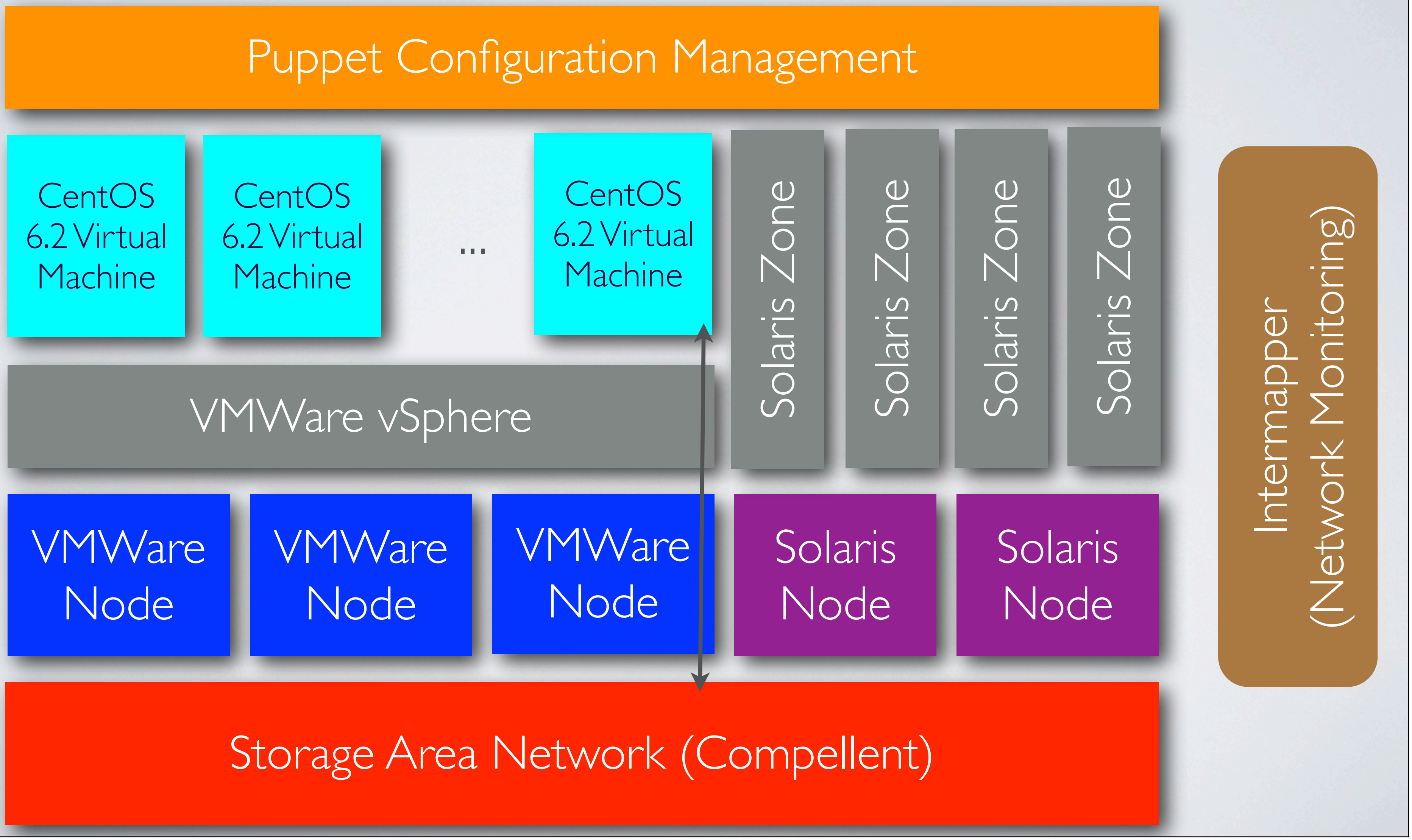
- Live on top of our physical infrastructure
- Primary OS disk and host specific data volumes on SAN storage
- Can be moved between physical servers with varying degrees of ease. VMware is automatic, Zones manual

- 15-20 CentOS Linux VMs (varies)
- ~15 Solaris Zones

BACKUP DATACENTERS

- Main DR site at IGPP
    - 2 Sun T2000 servers
    - 3 Nexsan SATABeast Storage Arrays
    - 1 AC&NC JetStor Array
    - Redundant LAN and SAN
    - Total ~160 TB backup storage
- Scorched earth site at IRIS DMC
    - Single Sun T2000 with 768 GB RAID5 internal disks

CURRENT INFRASTRUCTURE



Puppet Configuration Management

CentOS 6.2 Virtual Machine

CentOS 6.2 Virtual Machine

...

CentOS 6.2 Virtual Machine

Solaris Zone

Solaris Zone

Solaris Zone

Solaris Zone

VMWare vSphere

VMWare Node

VMWare Node

VMWare Node

Solaris Node

Solaris Node

Intermapper (Network Monitoring)

Storage Area Network (Compellent)

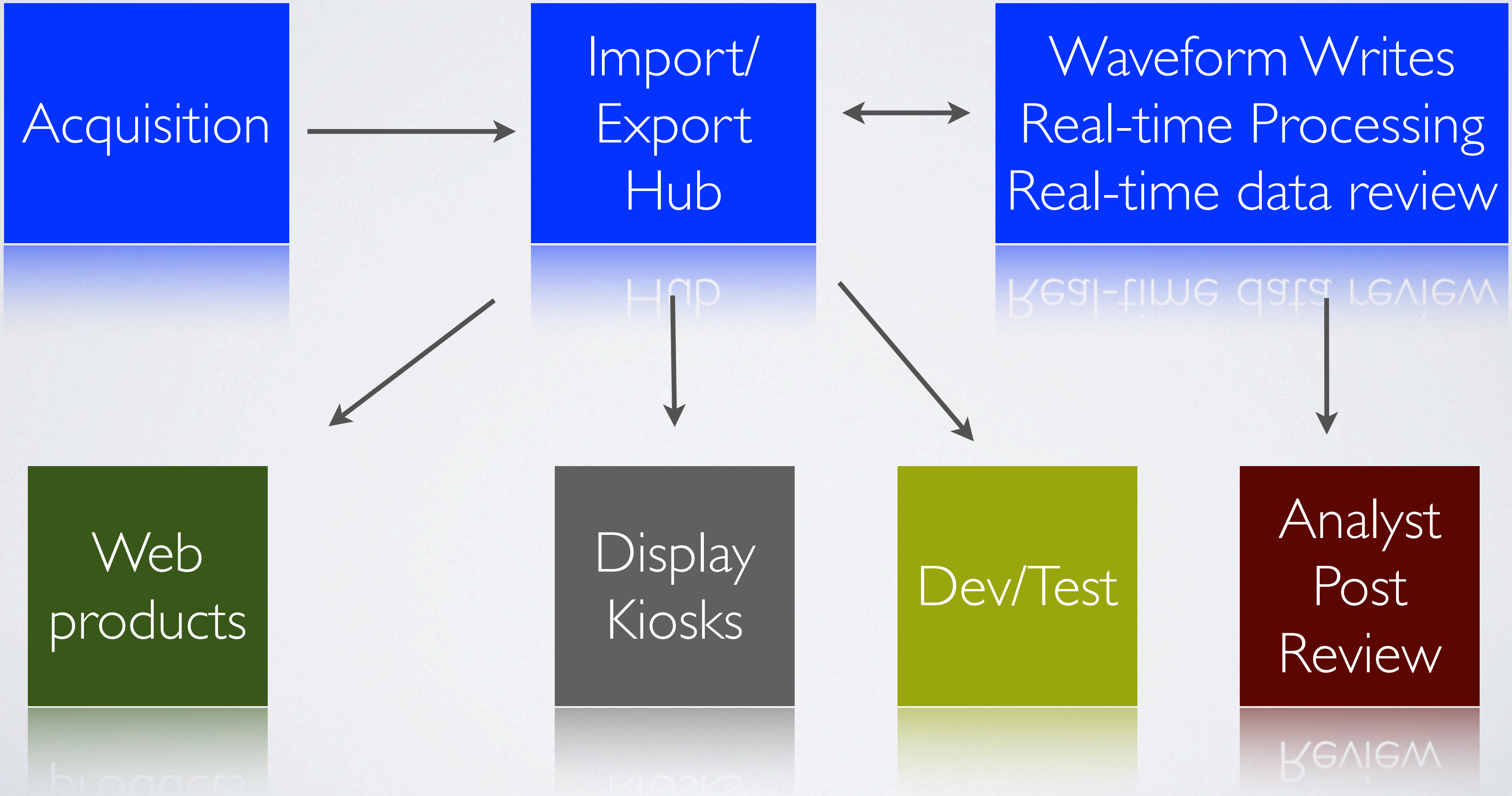
KEY ARCHITECTURE HIGHLIGHTS

- Virtualization
- SAN Storage
- Redundant Network and SAN Connections
- Configuration Management

WHY VIRTUALIZATION?

- Keep intensive processing jobs from interfering with real-time data collection
- Separation of core systems from analyst work and testing
- Easier hardware maintenance
- Easily create clones of existing systems for testing

VIRTUALIZED ANTELOPE SYSTEMS



Acquisition

Import/ Export Hub

Waveform Writes
Real-time Processing
Real-time data review

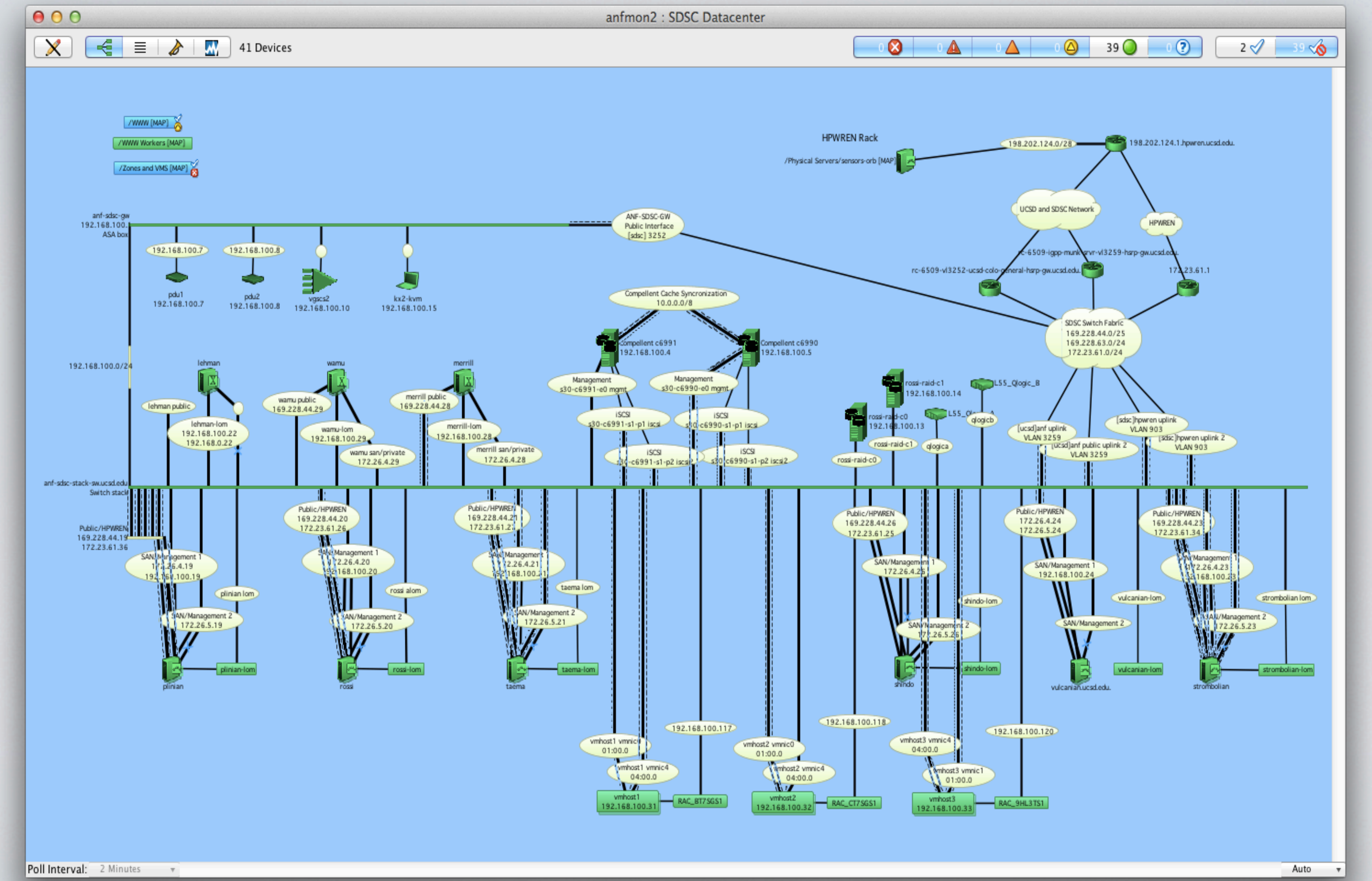
Web products

Display Kiosks

Dev/Test

Analyst Post Review

INTERMAPPER



anfmon2 : SDSC Datacenter

41 Devices

Poll Interval: 2 Minutes

SAN STORAGE

- Compellent system at primary Datacenter
  - 110 TB primary storage
  - no more manually trying to move data to fast disk or slow disk depending on workload
- All of our old disparate storage arrays pressed into service as offsite replication target for Disaster Recovery
- DR data synced via ZFS snapshots

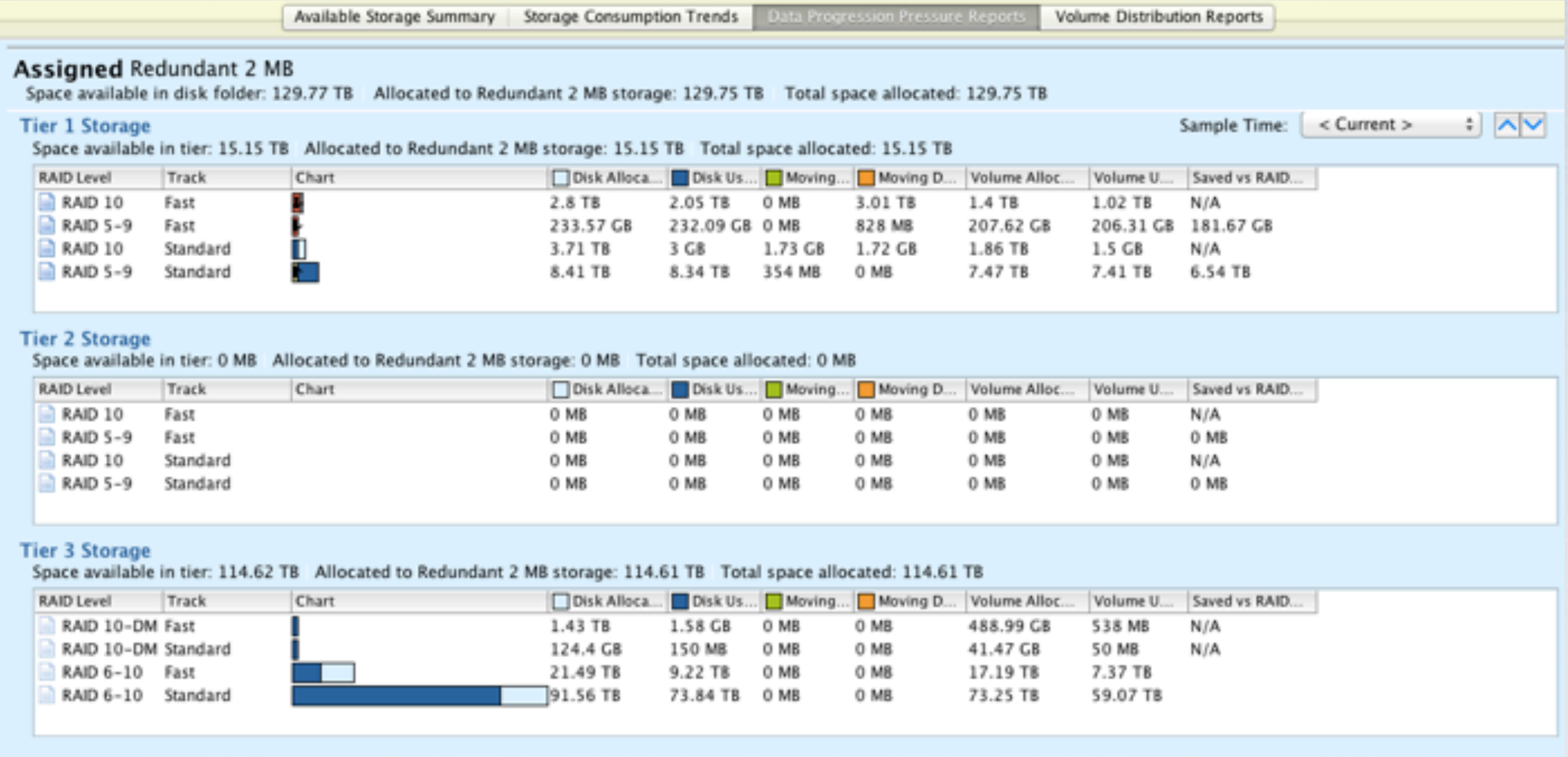
WHY COMPELLENT?

- Antelope and our local extensions generate a lot of different storage workloads
    - Waveform data is write biased, lots of small writes
        - Once it’s written, it’s typically read in large sequential chunks
    - Orbservers can be read or write biased depending on their use
    - RRDs really write biased, horribly inefficient

WHY COMPELLENT?

- Existing storage couldn’t keep up with write biased workload.
- We bought faster disk but it couldn’t fit all of our data there
- Copying data from one volume to another is slow, adds even more workload to overtaxed storage
- We have core systems requirements, but we’re always conducting experiments with data. Hard to predict what will be the most popular data for the day.

COMPELLENT MANAGEMENT


Available Storage Summary
Storage Consumption Trends
Data Progression Pressure Reports
Volume Distribution Reports

Assigned Redundant 2 MB

Space available in disk folder: 129.77 TB Allocated to Redundant 2 MB storage: 129.75 TB Total space allocated: 129.75 TB

Tier 1 Storage

Sample Time: < Current >

Space available in tier: 15.15 TB Allocated to Redundant 2 MB storage: 15.15 TB Total space allocated: 15.15 TB

RAID Level	Track	Chart	□ Disk Alloca...	■ Disk Us...	🟩 Moving...	🟧 Moving D...	Volume Alloc...	Volume U...	Saved vs RAID...
RAID 10	Fast		2.8 TB	2.05 TB	0 MB	3.01 TB	1.4 TB	1.02 TB	N/A
RAID 5-9	Fast		233.57 GB	232.09 GB	0 MB	828 MB	207.62 GB	206.31 GB	181.67 GB
RAID 10	Standard		3.71 TB	3 GB	1.73 GB	1.72 GB	1.86 TB	1.5 GB	N/A
RAID 5-9	Standard		8.41 TB	8.34 TB	354 MB	0 MB	7.47 TB	7.41 TB	6.54 TB

Tier 2 Storage

Space available in tier: 0 MB Allocated to Redundant 2 MB storage: 0 MB Total space allocated: 0 MB

RAID Level	Track	Chart	□ Disk Alloca...	■ Disk Us...	🟩 Moving...	🟧 Moving D...	Volume Alloc...	Volume U...	Saved vs RAID...
RAID 10	Fast		0 MB	0 MB	0 MB	0 MB	0 MB	0 MB	N/A
RAID 5-9	Fast		0 MB	0 MB	0 MB	0 MB	0 MB	0 MB	0 MB
RAID 10	Standard		0 MB	0 MB	0 MB	0 MB	0 MB	0 MB	N/A
RAID 5-9	Standard		0 MB	0 MB	0 MB	0 MB	0 MB	0 MB	0 MB

Tier 3 Storage

Space available in tier: 114.62 TB Allocated to Redundant 2 MB storage: 114.61 TB Total space allocated: 114.61 TB

RAID Level	Track	Chart	□ Disk Alloca...	■ Disk Us...	🟩 Moving...	🟧 Moving D...	Volume Alloc...	Volume U...	Saved vs RAID...
RAID 10-DM	Fast		1.43 TB	1.58 GB	0 MB	0 MB	488.99 GB	538 MB	N/A
RAID 10-DM	Standard		124.4 GB	150 MB	0 MB	0 MB	41.47 GB	50 MB	N/A
RAID 6-10	Fast		21.49 TB	9.22 TB	0 MB	0 MB	17.19 TB	7.37 TB	
RAID 6-10	Standard		91.56 TB	73.84 TB	0 MB	0 MB	73.25 TB	59.07 TB	

Writes go into top tier fast disk, trickle down to slower storage automatically

Least used data trickles to lowest tier of cheap slow disk.

If something suddenly becomes "hot" it will move up to higher tiers

VMWARE AND LINUX MIGRATION

- VMware cluster
  - 3 servers running at 2/3rds capacity, can survive outage of one server with no degradation of performance
- Core TA acquisition, processing, distribution migrated to Linux VMs(CentOS 6.2)
- 110 TB of Compellent storage
- Most Legacy Solaris Systems are being phased out
- TA acquisition, import/export and analyst processing all on Linux
- Web processing partially migrated

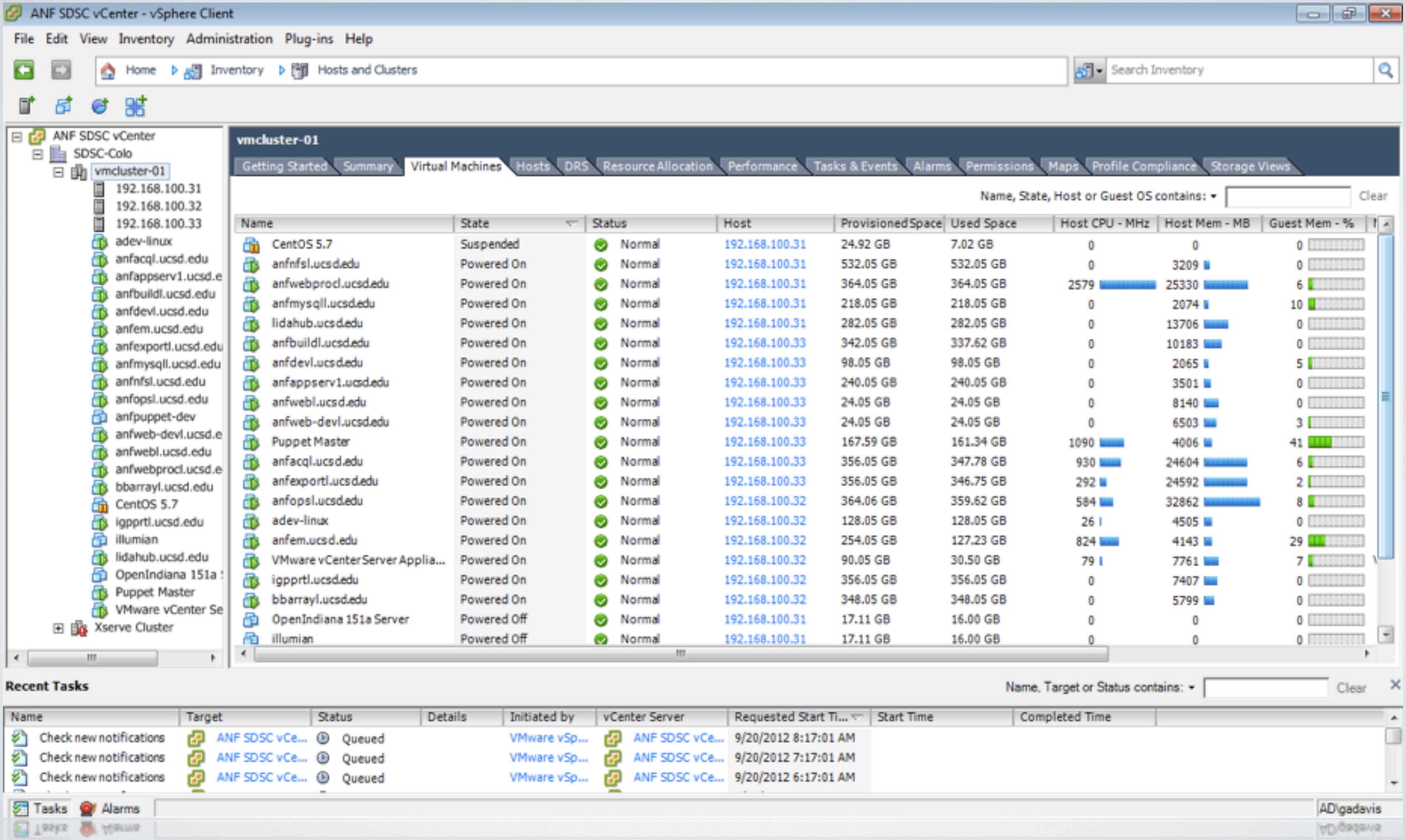
MIGRATION ADVANTAGES AND BENEFITS

- Commodity Intel hardware price/performance significantly better than Oracle (formerly Sun) SPARC, especially post Oracle
- Ongoing software support looks better long term on Linux
  - especially with regards to Antelope
  - Other open source components used in our "stack" have been harder to keep up to date on Solaris
- CPU speeds significantly faster
  - No need to differentiate workloads between floating point and integer
  - Catching up from a maintenance outage is significantly faster
  - Real-time database processing by analysts is significantly faster
    - startup time for dbloc2 went from 10 minutes down to 30 seconds

FURTHER MIGRATION ADVANTAGES AND BENEFITS

- Easier to spin up new virtual machines
- duplicate most of environment for testing
- scaling - add new processing nodes as needed, configured like existing ones
- Made easier by our use of Puppet
  - clone a fresh base system, assign role, puppet does the rest. Installs packages, starts/stops services, etc
  - Infrastructure represented in code. Software, NFS permissions, allowed users, switch configurations.

VMWARE


ANF SDSC vCenter - vSphere Client
File Edit View Inventory Administration Plug-ins Help
Home ▸ Inventory ▸ Hosts and Clusters
Search Inventory
vmcluster-01
Getting Started Summary Virtual Machines Hosts DRS Resource Allocation Performance Tasks & Events Alarms Permissions Maps Profile Compliance Storage Views
Tasks Alarms
ADI\gadavis

- ANF SDSC vCenter
  - SDSC-Colo
    - vmcluster-01
      - 192.168.100.31
      - 192.168.100.32
      - 192.168.100.33
      - adev-linux
      - anfacql.ucsd.edu
      - anfappserv1.ucsd.e
      - anfbuildl.ucsd.edu
      - anfdevl.ucsd.edu
      - anfem.ucsd.edu
      - anfexportl.ucsd.edu
      - anfmysqll.ucsd.edu
      - anfnfsl.ucsd.edu
      - anfopsl.ucsd.edu
      - anfpuppet-dev
      - anfweb-devl.ucsd.e
      - anfwebl.ucsd.edu
      - anfwebprocl.ucsd.e
      - bbarrayl.ucsd.edu
      - CentOS 5.7
      - igpprtl.ucsd.edu
      - illumian
      - lidahub.ucsd.edu
      - OpenIndiana 151a :
      - Puppet Master
      - VMware vCenter Se
    - Xserve Cluster

Name, State, Host or Guest OS contains: ▾ Clear

Name	State	Status	Host	Provisioned Space	Used Space	Host CPU - MHz	Host Mem - MB	Guest Mem - %
CentOS 5.7	Suspended	Normal	192.168.100.31	24.92 GB	7.02 GB	0	0	0
anfnfsl.ucsd.edu	Powered On	Normal	192.168.100.31	532.05 GB	532.05 GB	0	3209	0
anfwebprocl.ucsd.edu	Powered On	Normal	192.168.100.31	364.05 GB	364.05 GB	2579	25330	6
anfmysqll.ucsd.edu	Powered On	Normal	192.168.100.31	218.05 GB	218.05 GB	0	2074	10
lidahub.ucsd.edu	Powered On	Normal	192.168.100.31	282.05 GB	282.05 GB	0	13706	0
anfbuildl.ucsd.edu	Powered On	Normal	192.168.100.33	342.05 GB	337.62 GB	0	10183	0
anfdevl.ucsd.edu	Powered On	Normal	192.168.100.33	98.05 GB	98.05 GB	0	2065	5
anfappserv1.ucsd.edu	Powered On	Normal	192.168.100.33	240.05 GB	240.05 GB	0	3501	0
anfwebl.ucsd.edu	Powered On	Normal	192.168.100.33	24.05 GB	24.05 GB	0	8140	0
anfweb-devl.ucsd.edu	Powered On	Normal	192.168.100.33	24.05 GB	24.05 GB	0	6503	3
Puppet Master	Powered On	Normal	192.168.100.33	167.59 GB	161.34 GB	1090	4006	41
anfacql.ucsd.edu	Powered On	Normal	192.168.100.33	356.05 GB	347.78 GB	930	24604	6
anfexportl.ucsd.edu	Powered On	Normal	192.168.100.33	356.05 GB	346.75 GB	292	24592	2
anfopsl.ucsd.edu	Powered On	Normal	192.168.100.32	364.06 GB	359.62 GB	584	32862	8
adev-linux	Powered On	Normal	192.168.100.32	128.05 GB	128.05 GB	26	4505	0
anfem.ucsd.edu	Powered On	Normal	192.168.100.32	254.05 GB	127.23 GB	824	4143	29
VMware vCenter Server Applia...	Powered On	Normal	192.168.100.32	90.05 GB	30.50 GB	79	7761	7
igpprtl.ucsd.edu	Powered On	Normal	192.168.100.32	356.05 GB	356.05 GB	0	7407	0
bbarrayl.ucsd.edu	Powered On	Normal	192.168.100.32	348.05 GB	348.05 GB	0	5799	0
OpenIndiana 151a Server	Powered Off	Normal	192.168.100.31	17.11 GB	16.00 GB	0	0	0
illumian	Powered Off	Normal	192.168.100.31	17.11 GB	16.00 GB	0	0	0

Recent Tasks

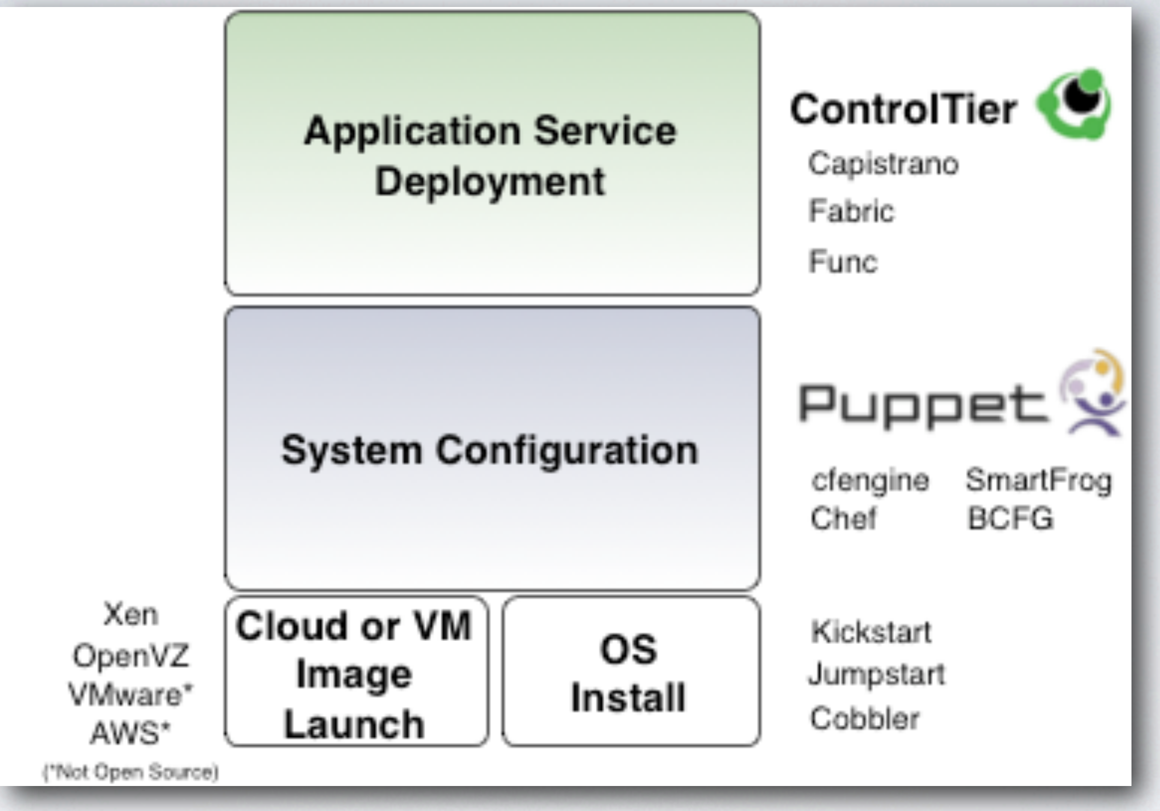
Name, Target or Status contains: ▾ Clear

Name	Target	Status	Details	Initiated by	vCenter Server	Requested Start Ti...	Start Time	Completed Time
Check new notifications	ANF SDSC vCe...	Queued		VMware vSp...	ANF SDSC vCe...	9/20/2012 8:17:01 AM		
Check new notifications	ANF SDSC vCe...	Queued		VMware vSp...	ANF SDSC vCe...	9/20/2012 7:17:01 AM		
Check new notifications	ANF SDSC vCe...	Queued		VMware vSp...	ANF SDSC vCe...	9/20/2012 6:17:01 AM		

VMWARE

- Three physical systems running ESXi 5.1 on bare metal
- Virtual machine that works as a supervisor, command dispatcher
- Automatically monitors load, migrates VMs to less loaded physical nodes as needed, transparent to users.
- Running at 2/3rds capacity so one server can be down for maintenance

SYSTEMS PROVISIONING

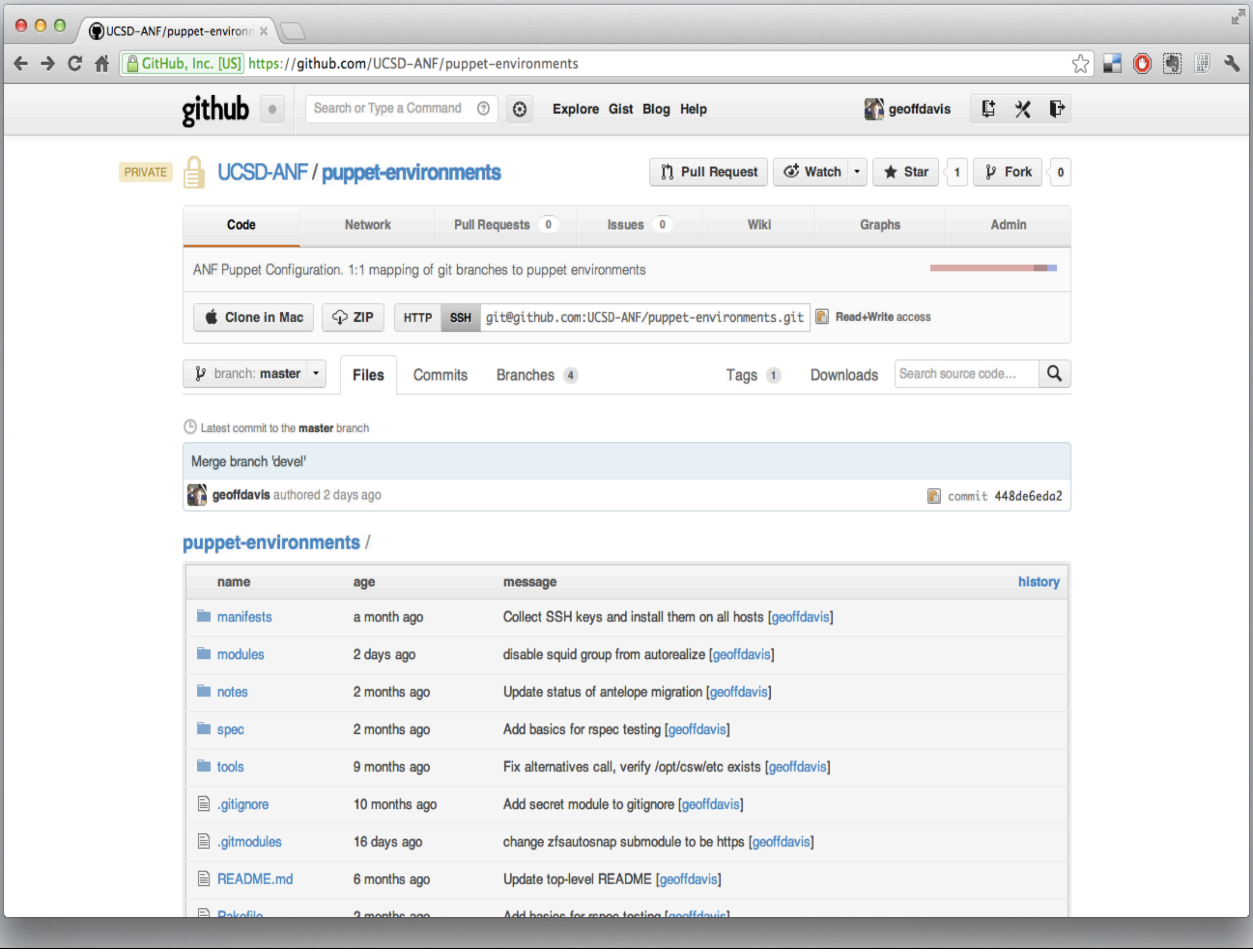


- The cool kids call it "DevOps"
- Something similar in use almost since the beginning of the project (2005), prior to that buzzword being coined
- We aren't using ControlTier but something at that tier will replace my few remaining SSH loops

CONFIGURATION MANAGEMENT WITH PUPPET

- http://puppetlabs.com
- Puppet is a configuration management tool with a declarative syntax. Similar in concept to Makefiles.
- You describe what something should look like, and the interpreter figures out the steps from current configuration to desired configuration.

INFRASTRUCTURE AS CODE. LITERALLY.



ANF Puppet Configuration. 1:1 mapping of git branches to puppet environments

git@github.com:UCSD-ANF/puppet-environments.git

Latest commit to the **master** branch

Merge branch 'devel'

geoffdavis authored 2 days ago commit 448de6eda2

puppet-environments /

name	age	message	history
manifests	a month ago	Collect SSH keys and install them on all hosts [geoffdavis]	
modules	2 days ago	disable squid group from autorealize [geoffdavis]	
notes	2 months ago	Update status of antelope migration [geoffdavis]	
spec	2 months ago	Add basics for rspec testing [geoffdavis]	
tools	9 months ago	Fix alternatives call, verify /opt/csw/etc exists [geoffdavis]	
.gitignore	10 months ago	Add secret module to gitignore [geoffdavis]	
.gitmodules	16 days ago	change zfsautosnap submodule to be https [geoffdavis]	
README.md	6 months ago	Update top-level README [geoffdavis]	
Rakefile	2 months ago	Add basics for rspec testing [geoffdavis]	

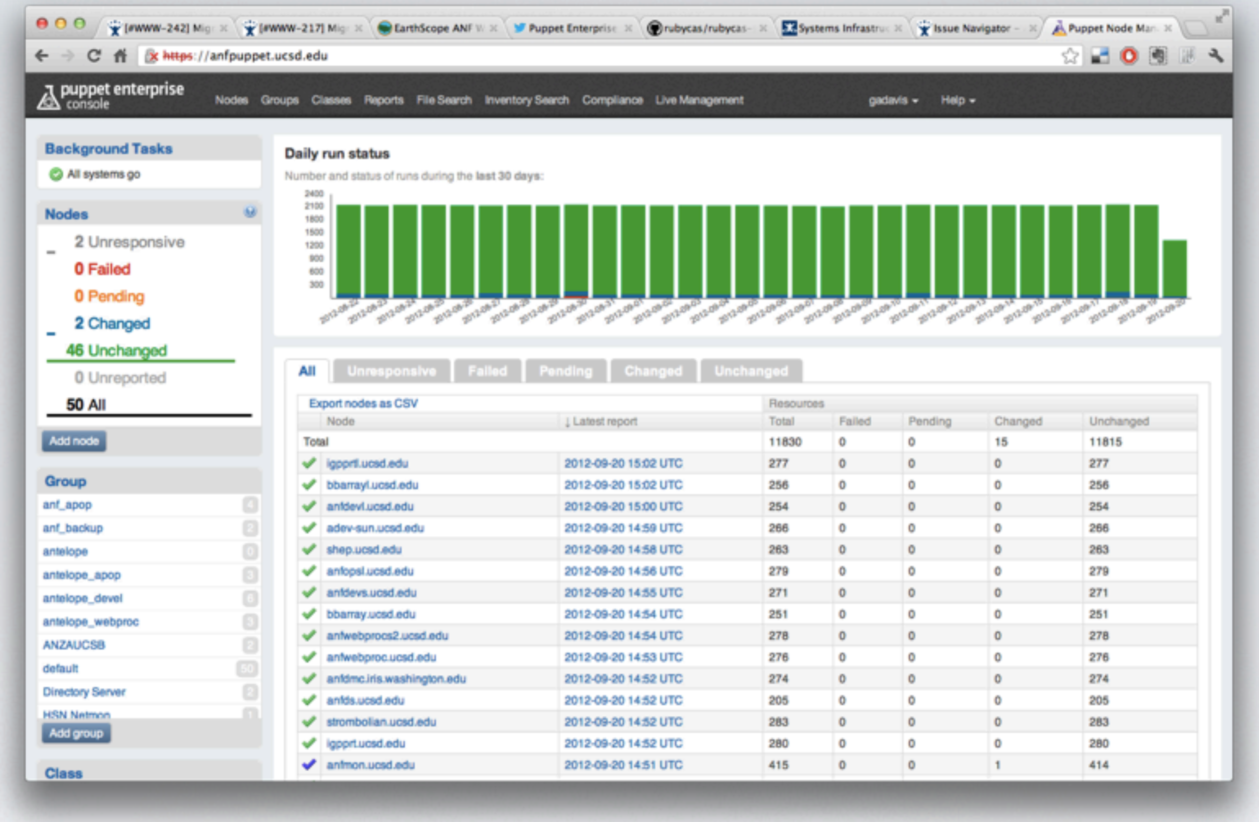
PUPPET CODE EXAMPLE

```
package { 'openssh-server':
    ensure => installed,
}

file { '/etc/ssh/sshd_config':
    source  => 'puppet:///modules/sshd/sshd_config',
    owner   => 'root',
    group   => 'root',
    mode    => '640',
    notify  => Service['sshd'], # sshd will restart whenever you edit this file.
    require => Package['openssh-server'],
}

service { 'sshd':
    ensure => running,
    enable => true,
    hasstatus => true,
    hasrestart => true,
}
```

PUPPET DASHBOARD



Background Tasks

All systems go

Nodes

2 Unresponsive
0 Failed
0 Pending
2 Changed
46 Unchanged
0 Unreported
50 All

Group

anf_apop	4
anf_backup	2
antelope	0
antelope_apop	3
antelope_devel	6
antelope_webproc	3
ANZAUCSB	2
default	50
Directory Server	2
HSN Netmon	1

Class

Daily run status

Number and status of runs during the **last 30 days**:

			Resources				
	Node	↓ Latest report	Total	Failed	Pending	Changed	Unchanged
Total			11830	0	0	15	11815
✓	igpprtl.ucsd.edu	2012-09-20 15:02 UTC	277	0	0	0	277
✓	bbarrayl.ucsd.edu	2012-09-20 15:02 UTC	256	0	0	0	256
✓	anfdevl.ucsd.edu	2012-09-20 15:00 UTC	254	0	0	0	254
✓	adev-sun.ucsd.edu	2012-09-20 14:59 UTC	266	0	0	0	266
✓	shep.ucsd.edu	2012-09-20 14:58 UTC	263	0	0	0	263
✓	anfopsl.ucsd.edu	2012-09-20 14:56 UTC	279	0	0	0	279
✓	anfdevs.ucsd.edu	2012-09-20 14:55 UTC	271	0	0	0	271
✓	bbarray.ucsd.edu	2012-09-20 14:54 UTC	251	0	0	0	251
✓	anfwebprocs2.ucsd.edu	2012-09-20 14:54 UTC	278	0	0	0	278
✓	anfwebproc.ucsd.edu	2012-09-20 14:53 UTC	276	0	0	0	276
✓	anfdmc.iris.washington.edu	2012-09-20 14:52 UTC	274	0	0	0	274
✓	anfds.ucsd.edu	2012-09-20 14:52 UTC	205	0	0	0	205
✓	strombolian.ucsd.edu	2012-09-20 14:52 UTC	283	0	0	0	283
✓	igpprt.ucsd.edu	2012-09-20 14:52 UTC	280	0	0	0	280
✓	anfmon.ucsd.edu	2012-09-20 14:51 UTC	415	0	0	1	414

- Dashboard gives me a quick overview of changes to code.
- Can also monitor other files/services not controlled by Puppet to alert to changes.

PUPPET-ANTELOPE

- This wouldn’t belong at AUG without tying it into Antelope somehow.
- https://github.com/UCSD-ANF/puppet-antelope

```
 antelope::instance { 'antelope' :
   user => 'rt',
   dirs => ['/rtsystems/foo', '/rtsystems/bar'],
}

antelope::instance { 'antelope-baz' :
   user        => 'basil',
   dirs        => '/rtsystems/baz',
   manage_fact => false, # don't create an entry in the antelope_services fact for
this instance.
}
```

ANTELOPE MANAGEMENT

- We use a lot of extensions to Antelope
  - PHP, Perl bindings, Image Magick
- Customized parameter files for ANF
- Instrument responses
- Typically running two or three releases of Antelope
- Want consistent extensions available on all of them

ANTELOPE BUILD PROCESS

- All managed by a series of Makefiles and helper scripts
    - Install Antelope
    - Copy in initial license file for the build host
    - Install $ANTELOPE overlays (localmake_config, some instrument responses)
    - Build $ANF site local tree
- Build contrib

ANTELOPE DISTRIBUTION

- Rsync based
  - We tried packages, but too much flux of contrib code, don't want to redistribute 600+ MB package for 2 MB change during incremental builds
- Deploy to development hosts first
- Stop acquisition, kick analysts off
- Deploy to production
- Deploy to backup sites

POST DISTRIBUTE TASKS

- All managed with Puppet
- System role specific parameter files (Transportable Array versus ANZA)
- Site-specific (and version specific license files)
- Install /etc/init.d/antelope with customized list of rtexec instances