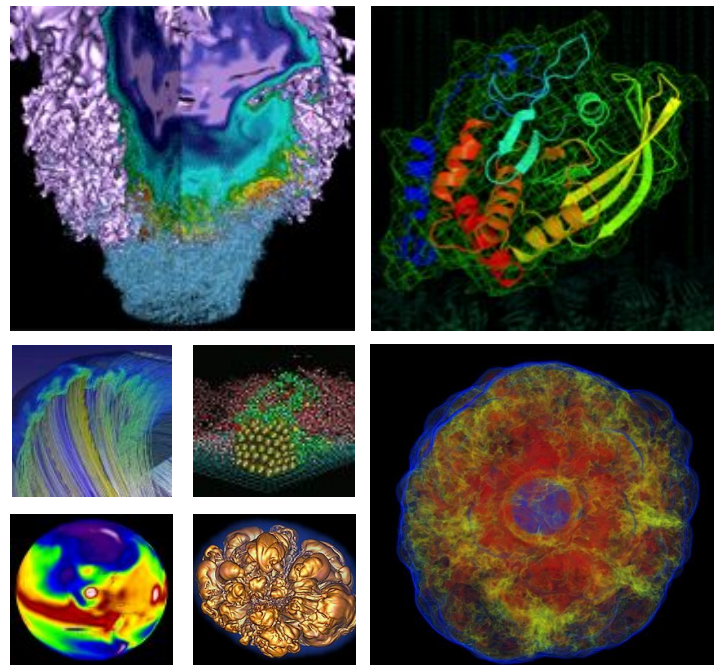


How to make your Cluster look like a Supercomputer (for Fun and Profit)



Georg Rath, Ershaad Basheer, Doug Jacobsen

HEPIX Spring 2019

- **Provide a different option for cluster management**
 - **Personnel changes + complex configuration not working**
 - **Allow for easy integration of new test systems**
- **Leverage existing configuration repositories**
 - **One management infrastructure for all systems**
 - **Two XC systems - Cray Rhine System with Ansible**
 - **One Cray CS system - custom xCAT with cfengine**
- **Provide one environment for our users**
 - **Same (similar) OS**
 - **Same batch system**
 - **Same bugs**

The Mendel System



- 800 Node CS300 Cluster
- Hosts two unique Systems with different requirements
 - PDSF
 - Denovo
- xCAT management software
- ScientificLinux
 - Plus CHOS (<https://github.com/NERSC/chos>)
- CFEngine configuration management
- Slurm workload manager

History - HEPIX 2014



NERSC Cluster Model

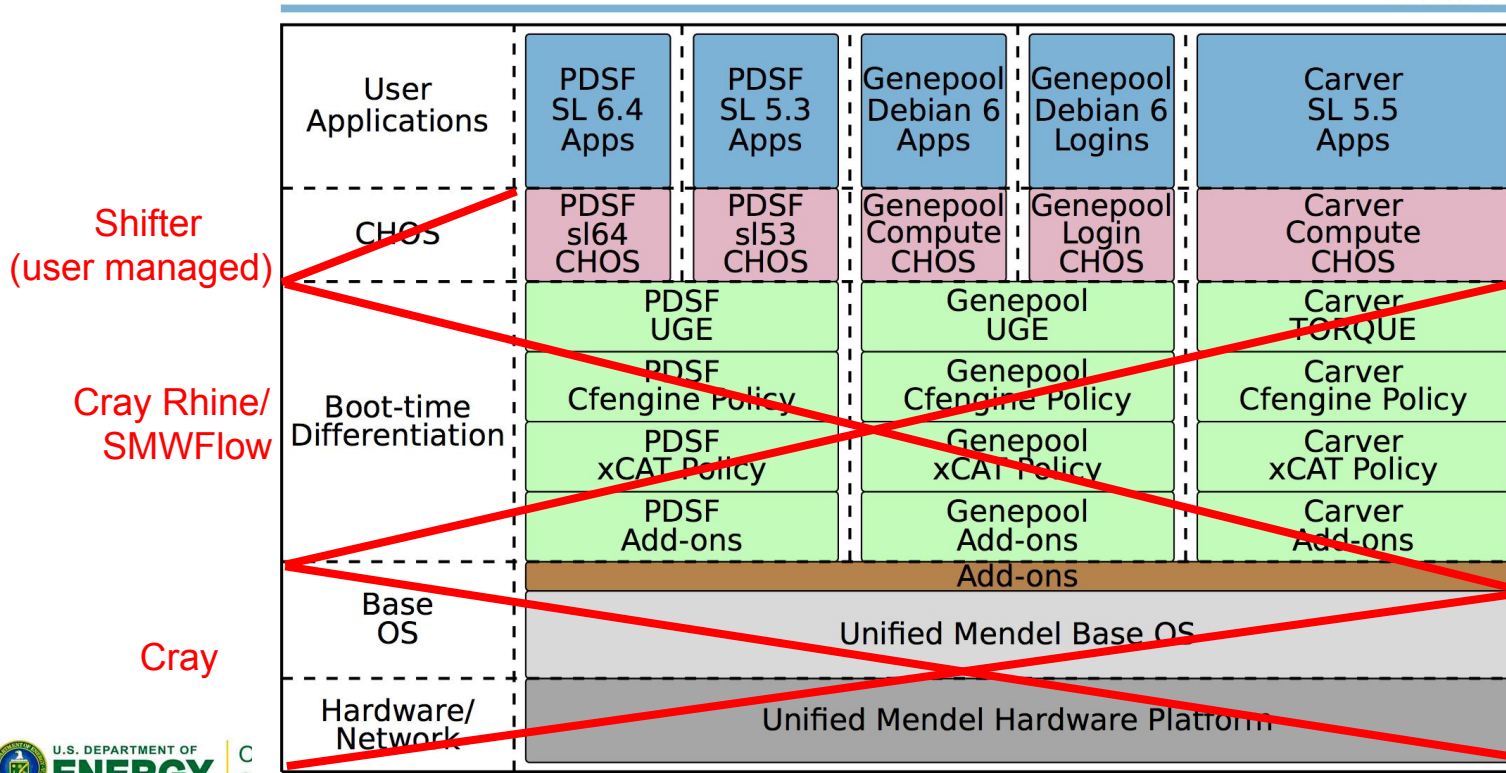


User Applications	PDSF SL 6.4 Apps	PDSF SL 5.3 Apps	Genepool Debian 6 Apps	Genepool Debian 6 Logins	Carver SL 5.5 Apps
CHOS	PDSF sl64 CHOS	PDSF sl53 CHOS	Genepool Compute CHOS	Genepool Login CHOS	Carver Compute CHOS
Boot-time Differentiation	PDSF UGE		Genepool UGE		Carver TORQUE
	PDSF Cfengine Policy		Genepool Cfengine Policy		Carver Cfengine Policy
	PDSF xCAT Policy		Genepool xCAT Policy		Carver xCAT Policy
	PDSF Add-ons		Genepool Add-ons		Carver Add-ons
Base OS	Add-ons				
	Unified Mendel Base OS				
Hardware/ Network	Unified Mendel Hardware Platform				

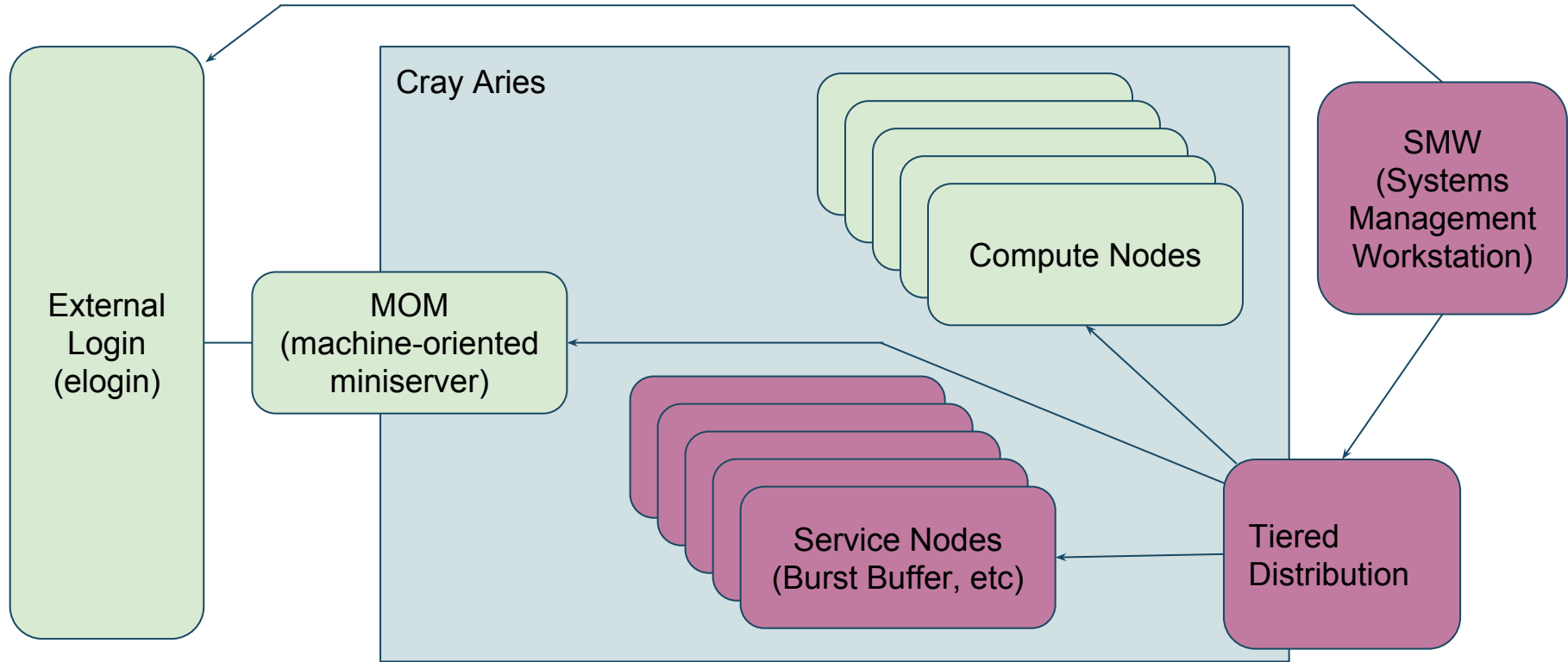
History - now



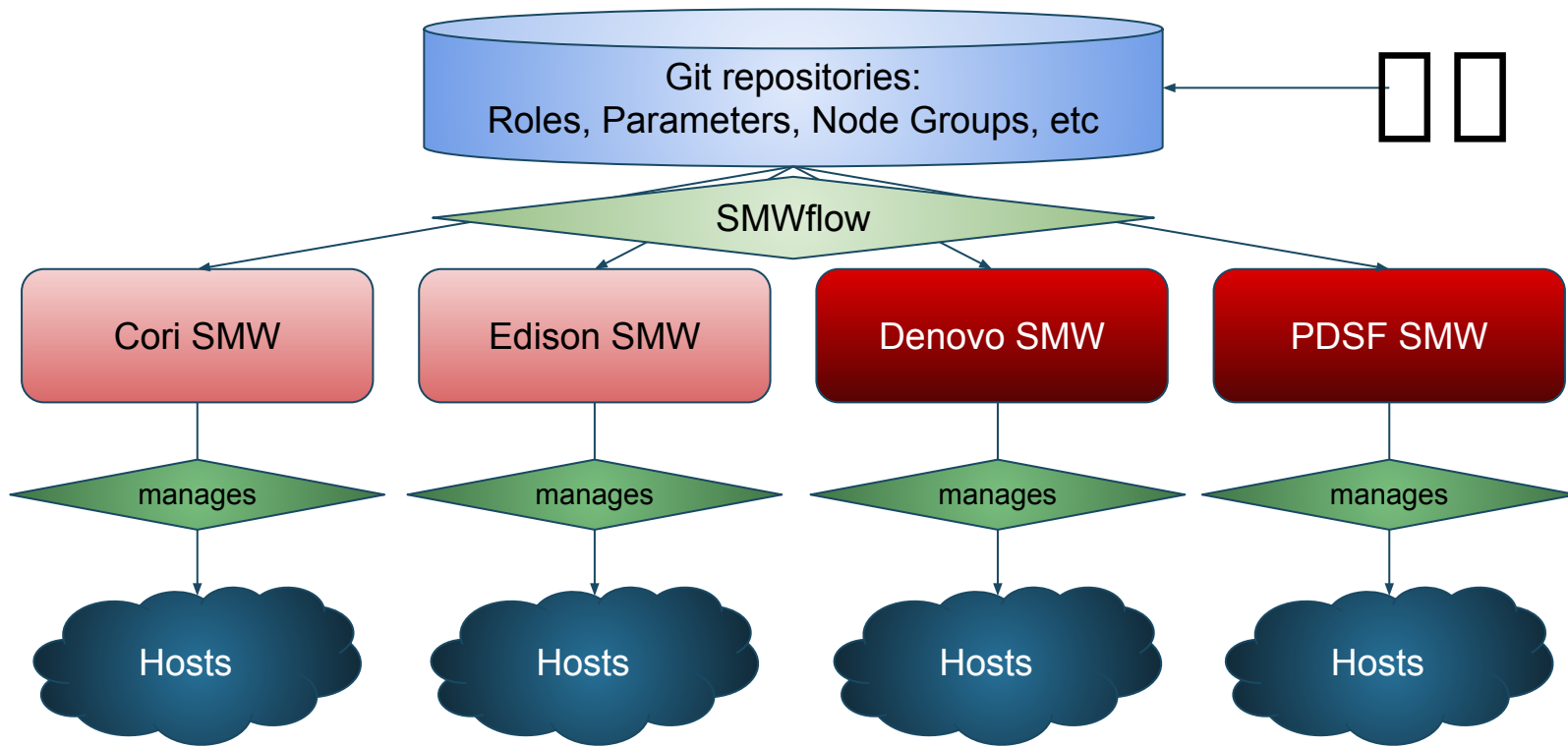
NERSC Cluster Model



Cray System Primer

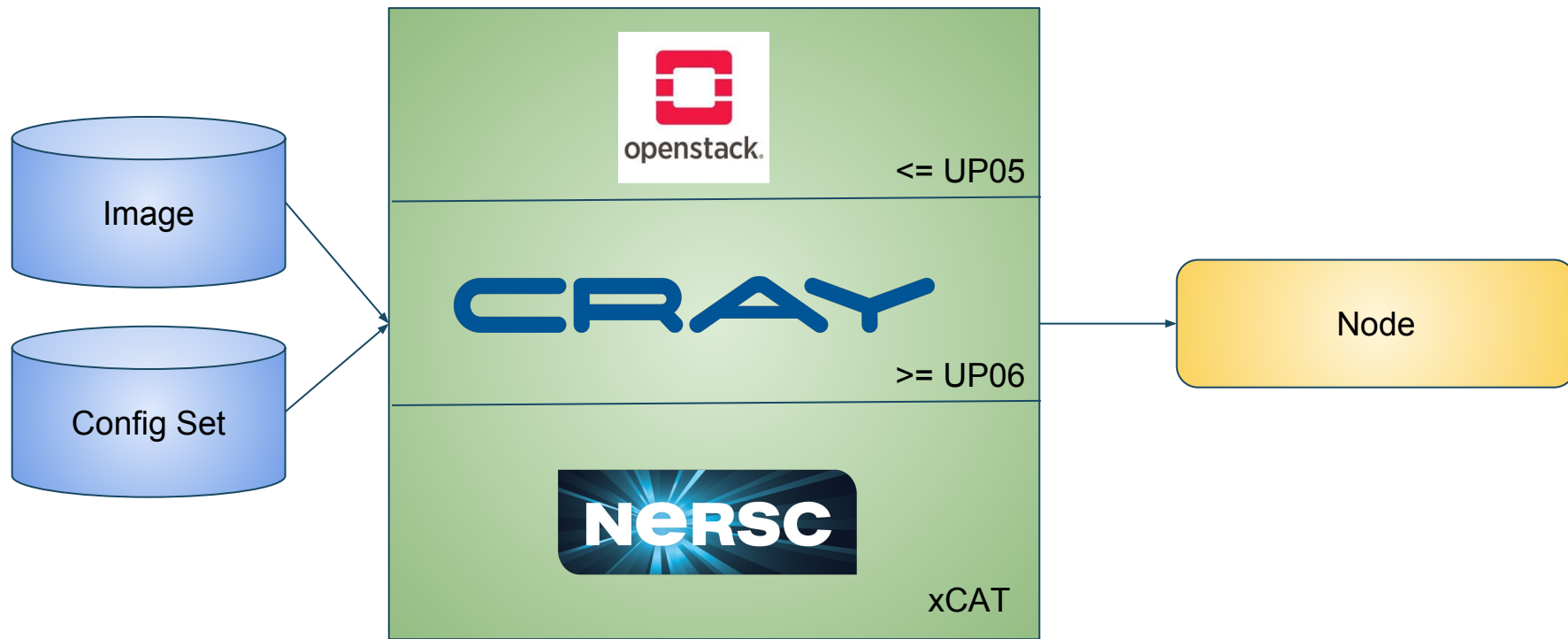


System View



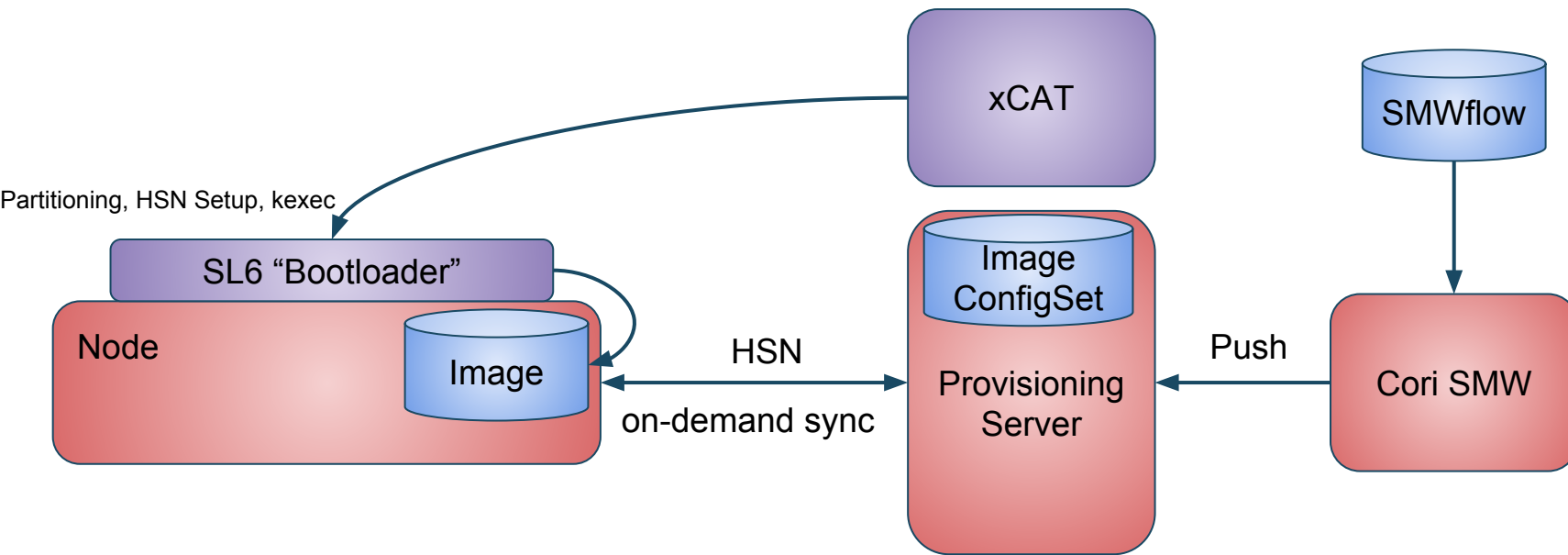
- Images are created by the Cray image management and provisioning system (IMPS)
 - “Standard” image building
- Config sets are created by the Cray configuration management framework (CMF)
 - Cray worksheets converted to Ansible (pull)
- Config set merged into image by the provisioning system
- Elogin provisioning was OpenStack until CLE6-UP06
 - OpenStack was replaced by custom Cray

From SMW to Node



- **Licensing (SuSE Enterprise on XC, programming environment)**
 - **Conversion to OpenSuSE, No PE**
- **Cray Management Stack makes assumptions on hardware configuration**
 - **Remove assumptions**
- **Part of Ansible built into image**
 - **Modify image before transfer**
- **Reuse existing HSN (high speed network) initialization**
 - **Use xCAT to bootstrap**
- **Keep data on the nodes intact (XROOTD)**
 - **xCAT part keeps disk layout intact**

The Boot Process



- **All systems have the same management**
 - **Pooling of efforts**
 - **Consistent experience for users (and admins!)**
- **Move from homegrown to semi-supported**
- **Fix once, deploy everywhere**
- **Template for rapid systems deployment**



Thank You



U.S. DEPARTMENT OF
ENERGY

Office of
Science

