Tier-1 centre at NRC «Kurchatov Institute» between LHC Run2 and Run3

Eygene Ryabinkin Alexander Rogovskiy Ilya Lyalin Igor Tkachenko

Evolution of hardware part of Tier-1 site at Kurchatov Institute

- 2013 year prototype at 10% computing power, about 80 nodes;
- 2015 year all hardware is ready, about 400 nodes;
- 2015-2018 upgrading computing and storage facilities, more than 500 nodes.

How does Tier-1 cluster managment system work

- Uses Puppet for cluster managment;
- Syncs puppet modules between master node and all nodes in the cluster;
- Applies local modules via **puppet** apply;
- Uses pdsh for mass modules applying at cluster nodes.
- Every site administrator have own puppet-environment.

Problems with original cluster managment system

- We have to keep modules in syncronizied state: we need special module for modules syncronization and we must not foget to call it before every configuration change at node by updated module
- It is very difficult to notice errors during mass module applying
- High chance to miss node (that was switched off for any reason, for example) and it is very difficult to notice it
- High load of master node during applying modules at all nodes

- **apply**: have to keep modules in syncronizied state (we need special module for modules syncronization and we must not foget to call it before every configuration change at node by updated module)
- **agent**: does not need modules sync at all. We may still use site administrator environment for tests.

- apply: it is very difficult to notice errors during mass module applying
- agent: puppet agent is writing to log-file results of all operations. We may analize log by monitoring tools.

T1.SRVC002: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/group_checkout/group_checkout_ssh.rb]/ensure: defined content as '{md5}62120c2fbdc88c4 8956bb6d9b93fa9d7' T1.SRVC002: Nation (Stage[main]/Sync/File[/sta/suppet]shc/suppet/environments/tia/modules/vcsrepo/spec/ac

T1.SRVC002: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/group_checkout/negative]/ensure: created

lhcb-sdpd14: from /opt/puppetlabs/puppet/lib/ruby/vendor_ruby/puppet/application_support.rb:34:in `push _application_context'

T1.SRVC004: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/shallow_clone/shallow_clone zero depth.rb]/ensure: defined content as '{md5}954d11fe90 296b23263ad35e13570147'

T1.SRVC004: Notice / stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/mounter/ssrepo/spec/ac ceptance/icaker/git/tag_checkout]/ensure: created

lbgf_sdpd14: from /opt/puppetlabs/puppet/lib/ruby/vendor_ruby/puppet/application.rb:363:in `run' 1.SRVC004: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/tag_checkout/negative]/ensure: created

lhcb-sdpd14: from /opt/puppetlabs/puppet/lib/ruby/vendor_ruby/puppet/util/command_line.rb:136:in `run' atlas-sdpd15: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/lib/p uppet/provider/vcsrepo.rb]/ensure: defined content as '{md5}ff406d68236bf7dfe75b0f7e57cf3f7b'

b-sdpd14: from /opt/puppetlabs/puppet/lib/ruby/vendor_ruby/puppet/util/command_line.rb:73:in `execut

T1.SRVC003:iise: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/war__po/spec/ac ceptance/beaker/git/comp.___ien/negative/compression_alpha_checkout.rb]/ensure: difficult content as '{md5}8 edd5a284d10daf4ec3a48fb9cd3701b'

T1.SRVC012: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/user_checkout/user_checkout_git.rb]/ensure: defined content as '{md5}3a57c2f24ade2a45d d17cee0fc483e71'

T1.SRVC001: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/users/manifests /human_users_ppl/created in the interview of the second second second second second second second second second

wites/tia/modules/vcsrepo/examp

/Ctage[main]/Sync/File[/otc/nupnot]abc/numnot/

les/p4/create_client.pp]/ensure: defined content as '{md5}d49fc9c1dd55e01ee9344672ee890b13'
T1.SRVC011: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac
ceptance/beaker/git/group_checkout/group_checkout_git.rb]/ensure: defined content as '{md5}c8c3dfbb63f316f
62670d605696eca23'

T1.SRVC002: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/group_checkout/negative/group_checkout_file_non_existent_group.rb]/ensure: defined con tent as '{md5}a60cd271a7fd12b85acefe916f8ca387'

T1.SRVC002: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/revision_checkout]/ensure: created

T1.SRVC002: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/revision_checkout/negative]/ensure: created

T1.SRVC004: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec/ac ceptance/beaker/git/tag_checkout/negative/tag_checkout_not_exists.rb]/ensure: defined content as '{md5}169 b90ee27574a91822de0ad66b46f30'

- apply: high chance to miss node (that was switched off for any reason, for example) and it is very difficult to notice it
- agent: automatically apply modules at nodes

lhcb-sdpd11: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/Tia/modules/dcache/manifes ts/dumps.pp]/ensure: defined content as '{md5}b6c2981bc15b130028fad156e5f1ebdd' lhcbsdpm: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/batchsys/template s/hosts.equiv.erb]/ensure: defined content as '{md5}1dbad14c24cd7fad551dd2799a19ff20' lhcb-sdpd11: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/dcache/manifes ts/init.pp]/ensure: defined content as '{md5}12a5<u>0930289fa4ad209db91c16b72252</u>' lhcb-sdpd24: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/dcache/manifes ts/config/head.pp]/ensure: defined content as '{md5}b002e3922a6475a750e39f3422b919df' lhcb-sdpd19: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/bonding/templ tes/ifcfg-eth.erb]/ensure: defined content as '{md5}b9cee036e16e46a305da5b505e86f55e' lhcb-sdpd19: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/carp]/ensure: reated lhcb-sdpd9: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/dcache/manifest s/install posol la.pp]/ensure: defined content as '{md5}22f0242ba8114b278f2b191c9f09afde' lhcb-sdpd13: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/dcache/manifes ts/db/billing.pp]/ensure: defined content as '{md5}1cccb29f5e66dc1c155cc92a1b29627f' lhcb-sdpd19: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/carp/manifests /ensure: created lhcb-sdpd10: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/carp/manifests /install.pp]/ensure: defined content as '{md5}26daa4a37b494a913c04219b1d17ea37' lhcb-sdpd13: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/dcache/manifes ts/db/checking.pp]/ensure: defined content as '{md5}d2256cd77cfe39babb03e57afea643a2' lhcb-sdpd9: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/dcache/manifest s/logs.pp]/ensure: defined content as '{md5}7df8c730f8e2a04467e8e7b4fb62d853' lhcb-sdpd9: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/dcache/manifest s/pooll/ensure: created lhcb-sdpd19: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/carp/manifests /config.pp]/ensure: defined content as '{md5}a9547617ddeefcc6521e05ff262cf332' lhcb-sdpd10: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/carp/manifests /install rpm.pp]/ensure: defined content as '{md5}845aafb555d53e892657cc4ef460ae3a' lhcb-sdpd8: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/batchsys/manif€ sts/create users.pp]/ensure: defined content as '{md5}2525991f19fc494b2fb8302e8f93abd4' lhcb-sdpd4: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/bonding/manifes proceeded srm dirs.pp]/ensure: defined content as proceeded and proceded and proceeded and proceeded and proceeded and proc 5}489ddf54aeb03f6cde5fd0d05162b1fb' pdsh@tamer: ch1: ssh exited with exit code 255 atlas-sdpd12: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/vcsrepo/spec, acceptance/beaker/git/compression/negative/compression 7 checkout.rb]/ensure: defined content as '{md5}8bc f4f06b5394d81887013ebee6a0d12' 2018-09-11 T1.SRVC012: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules/yum/manifests/: nit.pp]/ensure: defined content as '{md5}62a14<u>68d09163d098d70573b373080e2'</u> lhcbsdns: Notice: /Stage[main]/Sync/File[/etc/puppetlabs/puppet/environments/tia/modules]/ensure: created

9

- **apply**: high load of master node during applying modules at all nodes; nodes needs syncronization and module applying took a lot of time, despite of local catalog compilation
- agent: automatic modules applying is distributed in time, only compilled catalog goes to nodes.

Compare the efficiency of sync+apply and agent for applying same module at all cluster nodes



Parallel streams

agent apply

Compare the efficiency of sync+apply and agent for applying same module at all cluster nodes



Parallel streams

2018-09-11

agent – apply

12

- **Problem**: not all modules should be applied automatically:
 - We want more control for cluster critical services
 - Services that work with users should be managed in manual mode (no one likes when storage pool with one's data suddenly restarts)
- **Possible solution**: find groups of modules that are cricial for cluster: for now it is modules that are related to storage system, job managment system and WLCG software.
 - Every module is critical before audit: how it may harm critical cluster componens described above. After this audit we may take a decision if this module is safe enough for automatic apply.

- **Problem**: all our modules were written for "oneby-one" mode. In agent-mode we have a lot of duplicate resource declaration and circular dependencies.
- **Solution**: audit of all our puppet code (about 47 modules and 8000 lines of code)

- **Problem**: who watches the watchmen? We have to be sure that agent works in the right way.
- **Possible solution**: use monitoring tools for looking after puppet activity at nodes.

- **Problem**: have to control changes in managment code carefully to avoid massive applying of wrong settings at nodes.
- **Solution**: just use code review!

Our experience of using Phabricator as a code review system

- Tier-1 at KIAE is using code review since prototype stage
- Initially, for code review we used mail list:
 - It is difficult to watch for change status (waiting for update, waiting for review, etc)
 - It is difficult to watch for review progress (what review requests are close to finish)
 - It is difficult to see wthat was changed since previous iterration
- Since the middle of 2016 we are using Phabricator
 - More than 400 changes was reviewed
 - More than 1500 comments
 - Added automatic codestyle checks
- Plans:
 - Unit-tests
 - Isolate puppet master form manual changes: automatically update puppet-master area after commit to version control system

Thank you for your attention!