Survey Questions

A study to understand the usefulness of various code review comments

Purpose:

You are being asked to be in a research study to understand the characteristics of useful code reviews, since you have been an active participant in code reviews within the OpenStack community. Your anonymous response will provide researchers with data from which to glean important insights regarding the categories of code review comments that are deemed most useful by developers. These insights may help to improve the effectiveness of code reviews. This study is being conducted at Wayne State University.

Study Procedures

If you take part in the study, you will be asked to complete a web survey that will take approximately 20 minutes.

Benefits

There will be no direct benefits to you for participating in the survey. However, the findings will provide the researchers with data from which to glean important insights to improve the effectiveness of contemporary code reviews.

Risks

There are no risks associated with participating in this study.

Costs

There will be no costs to you for participation in this research study.

Compensation

As an appreciation for your efforts, we will offer \$10 Amazon git cards to 10 randomly selected respondents. If you wish the participate in the random drawing, please provide

your email address at the end of this survey.

Confidentiality:

All data gathered will be anonymous from the time of collection. No identifying information will be associated with the data. There is no risk associated with lack of privacy. Only Dr. Bosu and his graduate students will have access to the data. The data will be stored on Dr. Bosu's password protected computer and secured survey hosting site (Qualtrics). Only summarized data will be presented at meetings or in publications.

Voluntary Participation /Withdrawal:

Taking part in this study is voluntary. You are free to not answer any questions or withdraw at any time. Your decision will not change any present or future relationships with Wayne State University or its affiliates. If you have any questions about this study now or in the future, you may contact Amiangshu Bosu or one of his research team members at the following phone number 313-577-0731. If you have questions or concerns about your rights as a research participant, the Chair of the Institutional Review Board can be contacted at (313) 577-1628. If you are unable to contact the research staff, or if you want to talk to someone other than the research staff, you may also call the Wayne State Research Subject Advocate at (313) 577-1628 to discuss problems, obtain information, or offer input.

Participation

By completing the survey, you are agreeing to participate in this study.

The data that you provide may be collected and used by Qualtrics as per its privacy agreement. Additionally, if you are under the age of 18, please do not complete this survey.

0	I consent, begin the study
0	I do not consent.

Demographics

The questions in this page aims to collect your demographics to better understand how a developer's background may influence his/her perception of the usefulness of

code review comments. However, we do not wish to uniquely identify you by any means.
What is your highest level of education?
High schoolBachelorsMastersPh.D.Other
Which gender do you identify yourself with?
MaleFemaleOtherPrefer not to disclose
How many years of software development experience do you have?
Do you have an open source project that you consider as your primary project (i.e. the project that you have spent most of your time)?
What are your roles in your primary project (please check all appropriate roles below)? □ Developer
☐ Maintainer
☐ Testing
☐ UI Design

☐ Documentation
Requirement analysis
☐ Software Architect
Other (please specify)
How many years have you been practicing tool-based code reviews?
▼
How many hours per week, on average, do you spend reviewing other contributors
code?
~
In your opinion, what makes a code review useful?
Approximately, what proportion of code reviews in your project, do you feel are useful
(i.e., fits the above definition)?
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Percent
Please rate the following categories of code review comments based on their perceived
usefulness to you?

Defects (identification of functional defects)

Coding convention (e.g., white space, blank lines, indentation)

Resource management

Logical mistake (incorrect logic)

API Integration (wrong or unsecured API call)

Naming convention (alternate identifier name)

Refactoring (changing the organization of the code without changing any functionality)

Documentation (adding comments to explain a method or class)

Praise (complement to the author)

Question (asking to clarify a code segment)

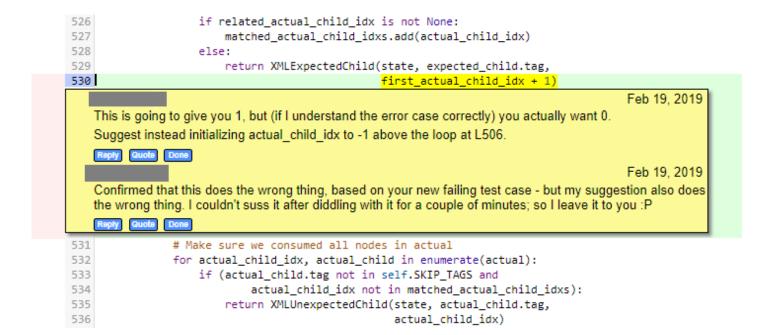
Solution approach (suggestions for an alternate solution approach)

Validation (inadequate validation of inputs)

Alternate output (changing output/ log/ error message)

Thread synchronization

Design discussion (directions to improve current design) In each of the following pages, you will be shown an example code review from the OpenStack Nova project. To better understand the context, code snippets surrounding each review comment will be also included. Please read each code review comment as well as its surrounding code context and rate each code review on a scale of 1 to 10, based on your perceived usefulness of each review comment.



```
890
            # prepare_template() may have cloned the image into a new rbd
            # image already instead of downloading it locally
891
892
            if not self.exists():
893
                 fmt = None
894
                 if not CONF.force_raw_images:
                     image_meta = IMAGE_API.get(kwargs.get('context'),
895
896
                                                   kwargs.get('image_id'))
                     fmt = image_meta.get('disk_format')
897
                                                                                             Feb 27, 2019
   This is user-defined, so this is not a trusted source either. If you have a look at how we handle this in
  fetch to raw we explicitly sanity check the backing file because we can't trust it.
  Please don't duplicate that code here, when we can just use it directly to do the conversion for us, making
  this code redundant. Also please don't add any code which relies on a malicious qcow2 header having been
  previously detected, because there is no strong guarantee that this will always be the case in the future, and
  an innocuous and easy-to-miss mistake would re-introduce a severe security hole.
                                                                                             Feb 28, 2019
  Done
  Reply Quote Done
                 self.driver.import_image(base, self.rbd_name, fmt)
898
899
            self.verify_base_size(base, size)
```

```
1059
1060
             value = filters_utils.validate_num_values(
1061
                 aggregate_vals, multiplier_config, cast_to=float)
1062
         except ValueError as e:
1063
             LOG.warning("Could not decode weight multiplier: '%s'", e)
                                                                                          Jan 19, 2019
   Include the multiplier_name value in this message.
                                                                                          Jan 21, 2019
   Done
1064
             value = multiplier_config
1065
1066
         return value
```

How useful do you find the above code review on a scale of 1

to 10, if you are the author?

```
131
                  if ironic_url is None:
132
                      LOG.warning("Could not discover ironic_url via keystoneauth1: "
133
                                    "Adapter.get_endpoint %s", ironic_url_none_reason)
134
                      # NOTE(eandersson): We pass in region here to make sure
                                                                                             Apr 23 10:49 AM
    This is good as is, but I want to note for posterity some of the background/details.
    This is tightly coupled to the ironicclient implementation, specifically [1]. The **kwargs to the get_client
    method is documented only as "all the other params that are passed to keystoneauth" [2]. This is vague and
    not completely true at least in terms of Adapter kwargs. It's also not a prescriptive/proscriptive interface, so
    ironicclient can change up the internals as it sees fit - as we saw very recently right next door with the
    addition of interface handling [3].
    Basically, it sucks how the ksa pieces of this puzzle are put together. It's brittle and will keep breaking in
    ways similar to this one, and we should fix it.
    [1] https://github.com/openstack/python-
    ironicclient/blob/060029ffcae8bbbd2315947cd059b701b5c31c39/ironicclient/client.py#L121
    [2] https://github.com/openstack/python-
    ironicclient/blob/060029ffcae8bbbd2315947cd059b701b5c31c39/ironicclient/client.py#L66
    [3] https://github.com/openstack/python-ironicclient/commit/d6eea403cbf0d11b06acbecc704f422a7e278462
135
                      # that the Ironic client can make an educated decision when
```

```
7000
7669
7670
                 self.rt.update_available_resource(context, nodename,
7671
                                                    startup=startup)
7672
             except exception.ComputeHostNotFound:
7673
                 # NOTE(comstud): We can get to this case if a node was
                 # marked 'deleted' in the DB and then re-added with a
7674
                 # different auto-increment id. The cached resource
7675
7676
                 # tracker tried to update a deleted record and failed.
                 # Don't add this resource tracker to the new dict, so
7677
7678
                 # that this will resolve itself on the next run.
                                                                                         Jan 14, 2019
   Remove this comment, which refers to a time when we had a ResourceTracker instance per baremetal
   Ironic node.
   Reply Quote Done
                                                                                         Jan 15, 2019
   Done
7679
                 LOG.info("Compute node '%s' not found in "
```

```
3010 if pci_mapping:
3011 for net in network_info:
3012 for old_dev, new_dev in pci_mapping.iteritems():

Use items() instead of iteritems() because iteritems is not supported in python 3.

[Reply] Guots Come

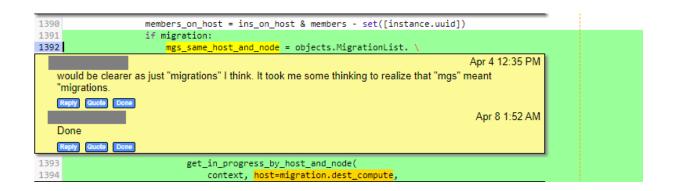
3013 if net['profile']['pci_slot'] == new_dev.address:
```

How useful do you find the above code review on a scale of 1 to 10, if you are the author?

```
# The dict maps domain types such as 'kvm' to the emulator
316
317
            # path for that domain type. Note that these overrides come
           # from <emulator> elements under each <domain>; there is no
318
319
            # <domemulator> element.
320
            self.domemulator = dict()
                                                                                         Mar 5, 2019
  consider renaming this domain_type_emulator_map or domain_type_emulators to make it clearer what it
  contains/does.
  Reply Quote Done
321
322
       def parse_dom(self, xmldoc):
323
           super(LibvirtConfigCapsGuest, self).parse_dom(xmldoc)
324
```

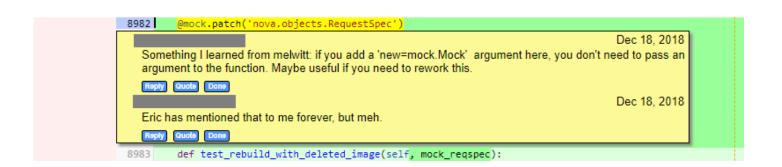
```
696
            if CONF.ironic.partition_key is not None:
                peer_list = set(CONF.ironic.peer_list)
697
                # these configs are mutable; need to check at runtime and init.
698
                # luckily, we run this method from init_host.
699
700
                _check_peer_list()
                LOG.debug('Limiting peer list to %s', peer_list)
701
702
703
            service_list = objects.ServiceList.get_all_computes_by_hv_type(
704
                ctxt, self._get_hypervisor_type())
                                                                                         Feb 26, 2019
  A future improvement might be to only call ServiceList.get_all_computes_by_hv_type() if
  CONF.ironic.peer_list is None. Otherwise, add a new ServiceList.get_compute_by_hosts() method that
  allows you to filter to just the peer_list values.
  Reply Quote Done
                                                                                         Feb 27, 2019
  ++, added a TODO
  Reply Quote Done
705
            services = set()
```

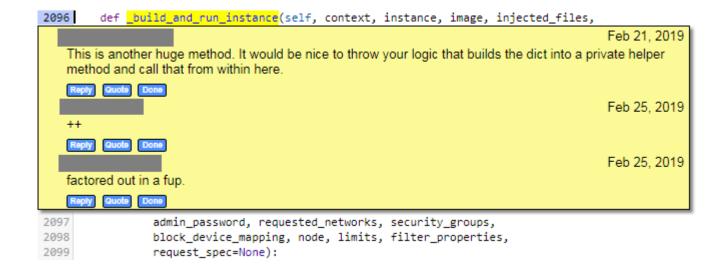
```
303
             try:
304
                 uuids = self._provider_tree.get_provider_uuids(
305
                     rp_uuid, whole_tree=True)
306
             except ValueError:
307
                 # If the provider isn't in the tree, it should also not be in the
                 # timer dict, so nothing to clear.
308
309
                 return
310
311
            # get_provider_uuids returns UUIDs in top-down order, so the first one
312
            # is the root; and .remove() is recursive.
313
             self._provider_tree.remove(uuids[0])
                                                                                              Jan 14, 2019
  if you take my advice about having a separate function on ProviderTree that does the tree-traversal stuff
  here, instead of passing a whole_tree=True argument, you might consider making that new function
  ProviderTree.remove_subtree() and wrap the removal code into the function. You can have
  ProviderTree.remove subtree() return the list of remove UUIDs, as well, so that lines 314-315 will be
  possible.
   That way, you wrap the related tree-fetch-and-removal code together in a single-purpose method on
  ProviderTree and leave ProviderTree.get_provider_uuids() method as-is.
   Reply Quote Done
                                                                                              Jan 14, 2019
   remove already does remove the subtree. The new method would have a different name but do the same.
  thing except for what it returns? May as well modify .remove to return the UUIDs.
   ...which I also considered. Ended up going with this approach as the lesser (or at least a coin flip) of two
  evils. Again, if you feel strongly, I can twiddle it to be different-but-equally-awkward. LMK.
                                                                                              Jan 15, 2019
   Having written get provider uuids in tree, I'd like to keep the deletion non-consolidated to facilitate the test
  cases if that's cool.
```



review on a scale of 1 to 10, if you are the author?

How useful do you find the above code review on a scale of 1 to 10, if you are the author?







```
46 echo '3. test with Ceph for root + ephemeral disks'

47 betOSVersion

Mar 14, 2019

nit: a comment about why this is needed would be good

Reply Quot Done

48 prepare_ceph

49 GLANCE_API_CONF:-/etc/glance/glance-api.conf}

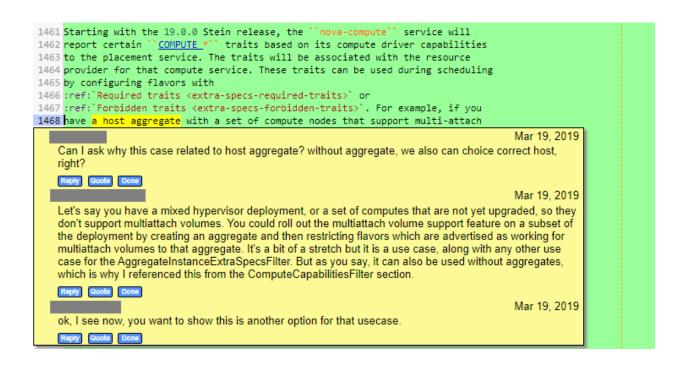
50 configure_and_start_glance

51

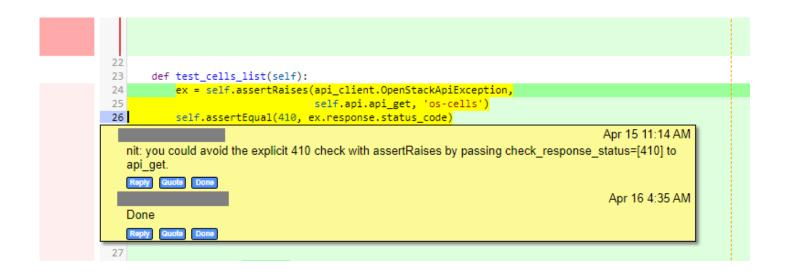
52 configure_and_start_nova

53 run_tempest "Ceph nova&glance test" "^.*test_live_migration(?!.*(test_volume_backed_live_migration))

54
```



```
310
       @property
311
       def compute_rpcapi(self):
          # Lazy load the rpcapi/ComputeAPI upon first use for this connection.
312
           # This way, if we receive a TCP RST, we will not create a ComputeAPI
313
314
           # object we won't use.
315
           if not self._compute_rpcapi:
316
               self._compute_rpcapi = compute_rpcapi.ComputeAPI()
317
            return self._compute_rpcapi
                                                                                       Mar 27, 2019
   oh, that's way nicer than the LazyLoader I would have hacked together.
318
```



```
35 from nova.objects import instance as obj_instance
36 from nova import rc_fields as fields
37 from nova import rpc
38 From nova.scheduler.filters import filters_utils

Jan 11, 2019

wrong import

Repty Quote Done

39
40
41 LOG = logging.getLogger(__name__)
42
```

```
put_call = mock.call(

'/allocations/%s' % uuids.consumer_uuid, updated_allocations,

version='1.28', global_request_id=self.context.global_id)

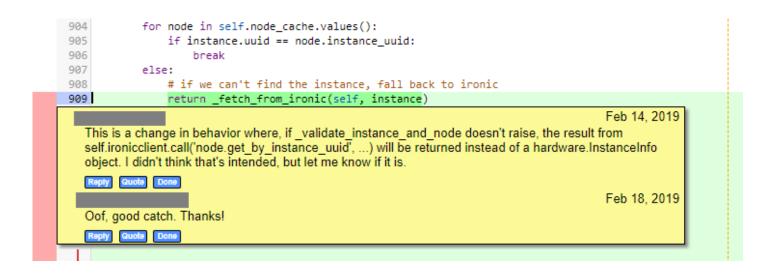
mock_put.assert_has_calls([put_call] * 4)

Feb 28, 2019

excellent unit tests, thank you Gibi.

Repty Guota Dome
```

```
336
            if api_version_request.is_supported(request, min_version="2.63"):
337
                trusted_certs = None
338
                if instance.trusted_certs:
                    trusted_certs = instance.trusted_certs.ids
339
340
                server["server"]["trusted_image_certificates"] = trusted_certs
341
342
            if (api_version_request.is_supported(request, min_version="2.71") and
                                                                                            Feb 25, 2019
    This is kind of weird that we pass in show_server_groups=True but then are also checking the microversion.
   Why not just have the caller pass in show_server_groups based on the microversion? Because when this is
   called from detail() you're going to be checking the microversion for every instance even though
   show server groups will always be False for detail(). Just have the caller figure out the microversion and
   pass in show_server_groups and honor that boolean.
                                                                                            Feb 27, 2019
   Done
                     show_server_groups):
```

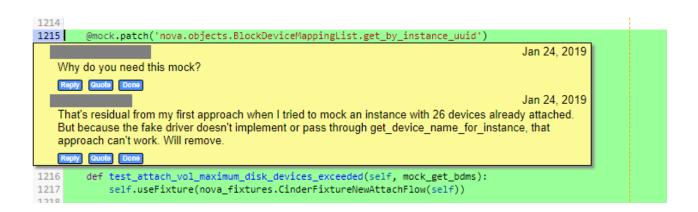


```
30 from oslo_service import fixture as service_fixture
31 from oslo_utils.fixture import uuidsentinel as uuids

Nov 20, 2018
This needs to go away. uuidsentinel was introduced in oslo.utils 3.37.0. This file needs to continue using the nova fixture (see L70).

Reply Quots Done

32 from oslo_utils import timeutils
```



```
def clear_provider_cache(self):

LOG.info("Clearing the report client's provider cache.")

self._provider_tree = provider_tree.ProviderTree()

Nov 6, 2018

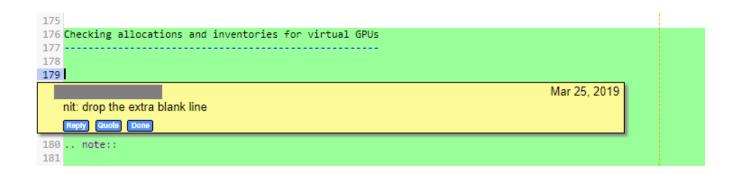
(thinking out loud)

If the existing ProviderTree is being operated on in a greenthread and is holding the lock, that's no biggie because that entire ProviderTree will go out of scope when any references to it do, so there ought not be any issues there.

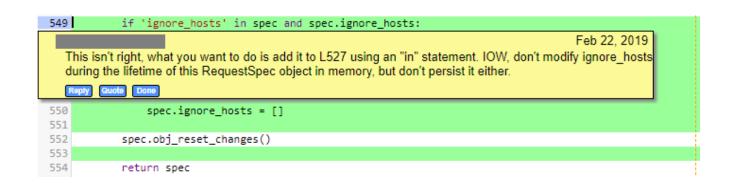
(I'm not even certain if we have potential for that situation, now)

Reply Quot Dono

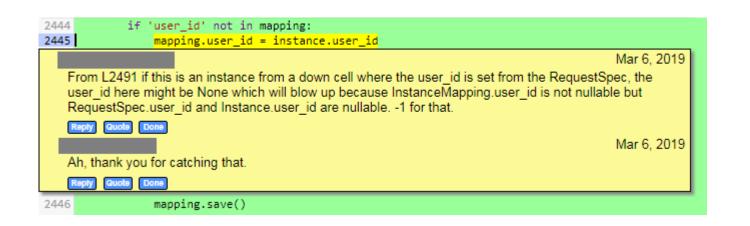
self._association_refresh_time = {}
```



How useful do you find the above code review on a scale of 1 to 10, if you are the author?

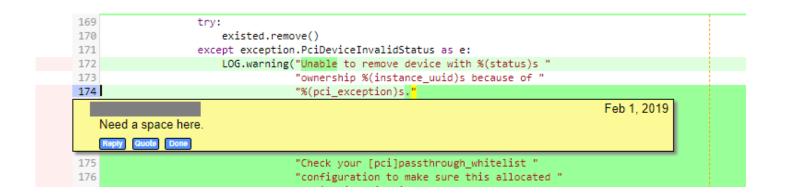


```
1913
             # Now get the resource allocations for the instance based
1914
             # on its embedded flavor.
1915
             resources = scheduler_utils.resources_from_flavor(
1916
                 instance, instance.flavor)
             if dry_run:
1917
                 output(_('[dry-run] Create allocations for instance %(instance)s '
1918
1919
                            on provider %(node_uuid)s: %(resources)s') %
1920
                         {'instance': instance.uuid, 'node_uuid': node_uuid,
1921
                          'resources': resources})
                                                                                          Apr 12, 2019
    So this will return None, and boo(None) is False so num processed are not incremented and therefore
    num processed will stay at 0 and that will result in exit code 4.
    Reply Quote Done
                                                                                          Apr 12, 2019
    Thanks for writing this.
    Reply Quote Done
1922
             else:
1923
                 if placement.put_allocations(
1924
                          ctxt, node_uuid, instance.uuid, resources,
1925
                          instance.project_id, instance.user_id,
1926
                          consumer_generation=None):
1927
                     output(_('Successfully created allocations for '
1928
                               'instance %(instance)s against resource '
1929
                               'provider %(provider)s.') %
                             {'instance': instance.uuid, 'provider': node_uuid})
1930
1931
                      return True
```



```
320
            # RBD volume doesn't support QCOW2 for hosting a virtual machine disk.
321
            # Thus if you want to boot virtual machines in RBD volume, the image
322
            # format must be RAW.
323
            # If you set the force_raw_images to False and the glance image format
324
            # is QCOW2, the format of image file here is QCOW2.
325
            # So that, get and check the format of image file. Then choose the
326
            # right way to import rbd volume from image file.
327
            data = images.qemu_img_info(base)
328
            fmt = data.file_format
                                                                                             Feb 26, 2019
    This is the security hole: you're doing format inspection. It doesn't matter if you do the format inspection
    during a convert or in an info command; it's still the same format inspection.
    In order for this to be secure you would have to pass the input format to gemu-img info explicitly, in which
    case you wouldn't need to run it anyway. But you can't pass it, because we don't have it in this context. IOW:
   you can't do this here.
   If a user can construct a malicious gcow2 header, declare the file as raw, and have it processed by this
   code, then this code will detect it as qcow2 (despite it being raw) and honour the malicious header.
    Reply Quote Done
                                                                                             Feb 27, 2019
    Now, I do not try to get the image from gemu img info. I get it from image meta. And explicitly specify the
    disk format to avoid the 'insecure image format probing'.
            if fmt == "raw":
329
330
                 args = ['--pool', self.pool, base, name]
331
                 # Image format 2 supports cloning,
332
                 # in stable ceph rbd release default is not 2,
333
                 # we need to use it explicitly.
334
                 args += ['--image-format=2']
335
                 args += self.ceph_args()
336
                 processutils.execute('rbd', 'import', *args)
```

```
148
            # Pick a file name for when we upload the media to VIOS
149
            file_name = pvm_util.sanitize_file_name_for_api(
                instance.uuid.replace('-', ''), prefix='cfg_', suffix='.iso',
150
151
                max len=pvm const.MaxLen.VOPT NAME)
152
153
            # Create and upload the media
154
            iso_path = tempfile.mkstemp()[1]
                                                                                         Oct 15, 2018
    Pretty sure we've left the file open here.
   [Later] Yeah: https://www.logilab.org/blogentry/17873
   Since we're Linux, we could get away with:
     with tempfile.NamedTemporaryFile() as fh:
          self._create_cfg_dr_iso(..., fh.name, ...)
         vopt, f_uuid = tsk_stg.upload_vopt(..., fh, fh.name, os.path.getsize(fh.name, os.path.getsize)
    The file deletes itself.
155
            try:
156
                self._create_cfg_dr_iso(instance, injected_files, network_info,
157
                                         iso_path, admin_pass=admin_pass)
```



6663	# Now that we updated the ProviderTree, we want to store it locally	
6664	# so that spawn() or other methods can access it thru a getter	
6665	selfprovider_tree = copy.deepcopy(provider_tree)	
	Feb 14, 2019	
And this also isn't bouncing on that lock, huh?		
Need to get to the bottom of this.		
[Later] Figured it out. eventlet.monkey_patch replaces thread with eventlet.green.thread. The former blows up when you try to copy it; the latter does not. And the latter is doing the right thing by creating a new semaphore. http://paste.openstack.org/show/745119/		
Reply Guote Done		
6666		

Is there anything else you would like to add about this survey?

Thank you so much for your time. As an appreciation for your efforts, please include your email address, if you want to be included in the random drawing for Amazon gift card worth \$10 (We'll randomly select ten respondents).

Powered by Qualtrics