



# Who Built Grizzly

Using `gitdm` to analyze OpenStack contributions

Dan Stangel / April 15, 2013

OpenStack Summit, Portland Oregon

# Using gitdm to measure OpenStack contributions



# Gitdm – the **git data miner**

- Written by Linux kernel gurus Greg Kroah-Hartman and Jonathan Corbet ,  
“to crank out statistics on where kernel patches come from”
- See “The Kernel Report” by Jonathan Corbet
- Forked by Mark McLoughlin (markmc), customized for use with OpenStack
- Simplistic but useful low-level git changeset analysis
- Includes tools to mine relevant data on bugs closed in LaunchPad (lpdm),  
and code reviews done via Gerrit (gerritdm)
- Maps multiple emails/IDs to a single contributor, tracks dates of service  
with organization



# Running gitdm

```
$ for project in cinder glance horizon \  
    keystone nova quantum swift; do  
    git clone https://github.com/openstack/$project.git  
done  
$ git clone https://github.com/markmc/openstack-gitdm.git  
$ cd openstack-gitdm  
$ ./do-it.sh
```



# Gitdm report on Grizzly

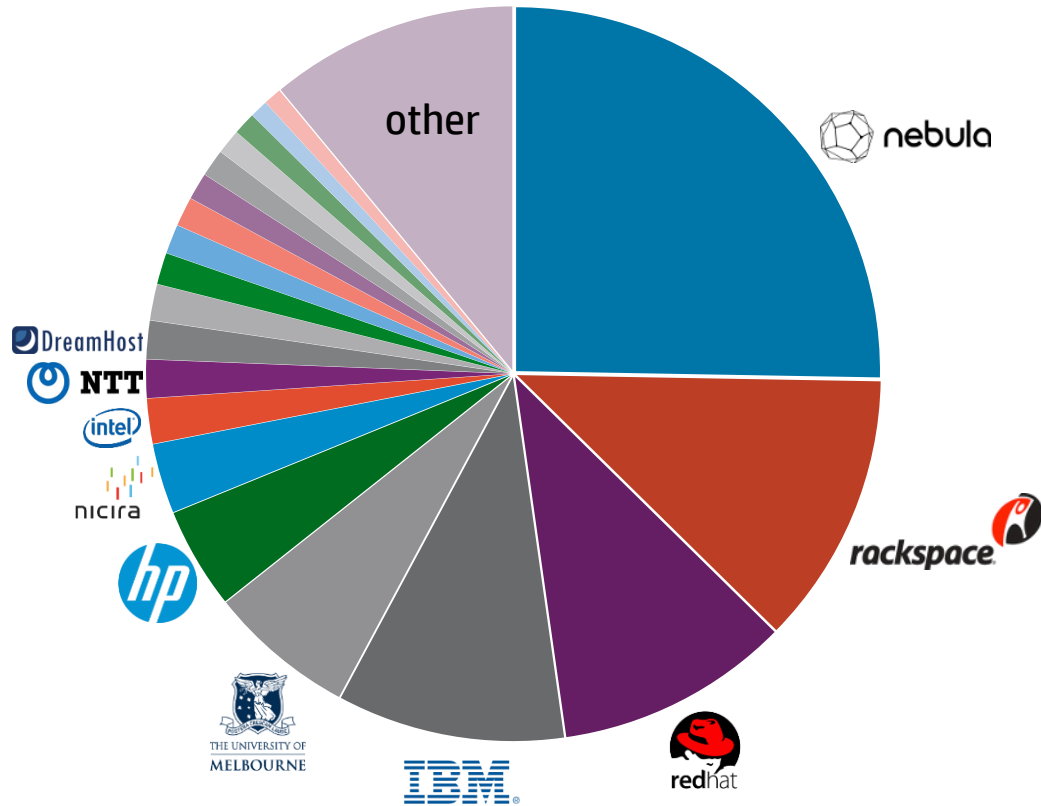
*Published by Mark McLoughlin (markmc) on April 4, 2013*

*Data Set Available Online At*

*<https://github.com/markmc/openstack-gitdm/tree/results/grizzly>*



# Top lines changed for all Grizzly core projects by employer

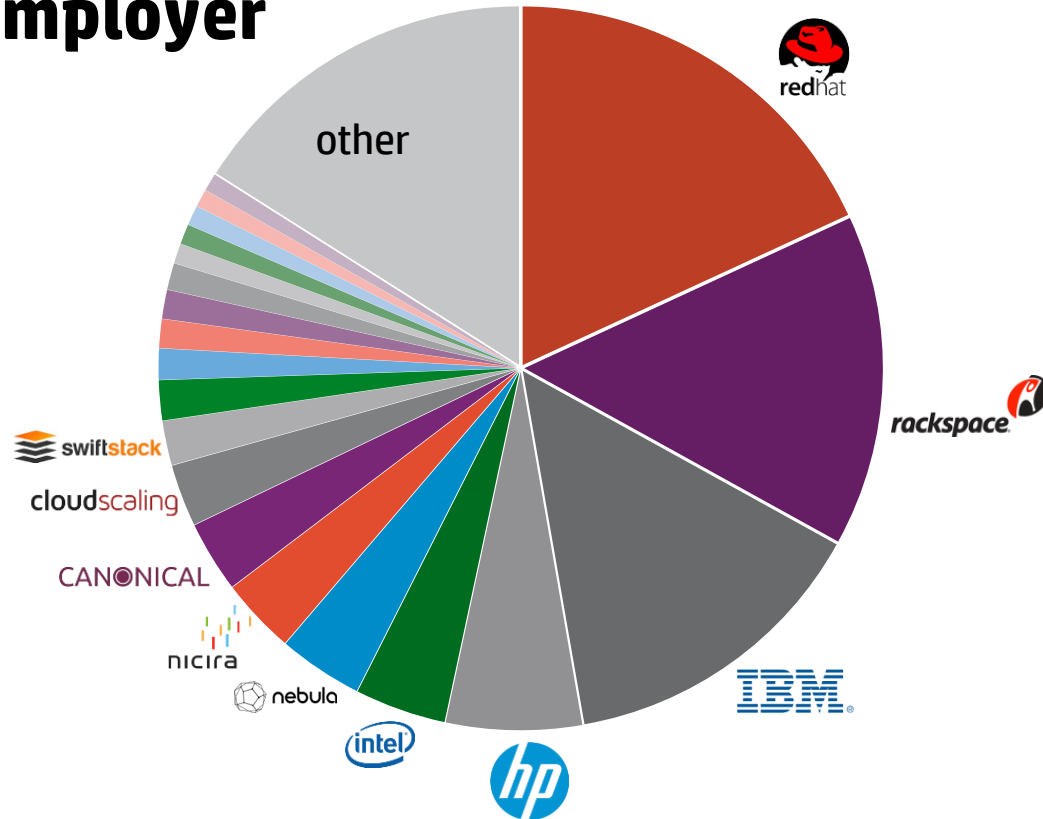


## Core includes

nova, swift, glance,  
horizon, keystone,  
cinder, quantum



# Top changeset contributors for all Grizzly core projects by employer

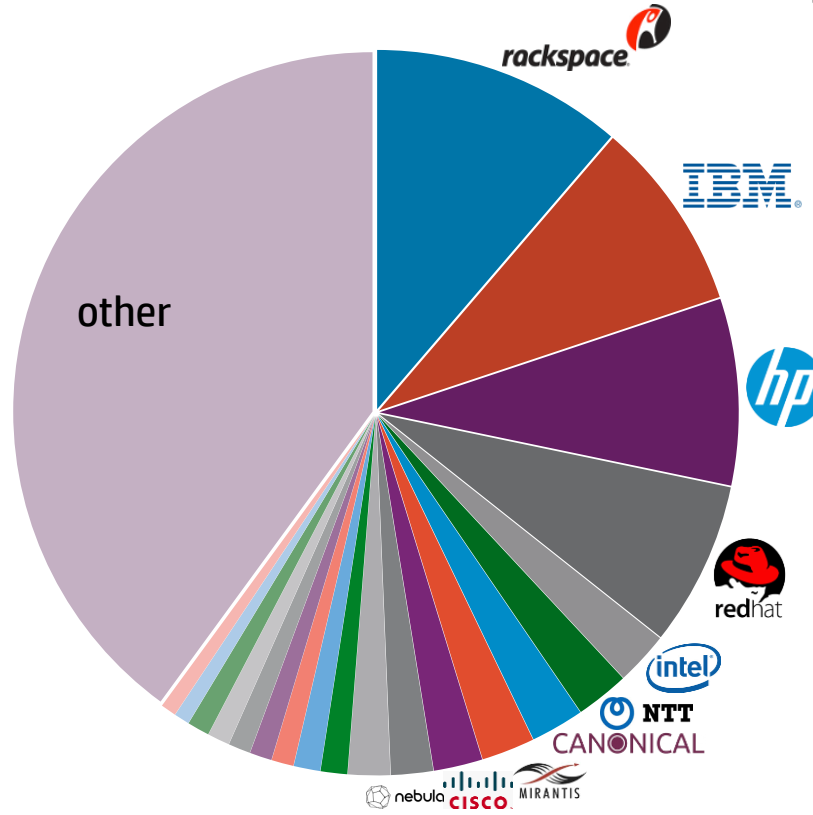


## Core includes

nova, swift, glance,  
horizon, keystone,  
cinder, quantum



# Employers with the most hackers for Grizzly core projects



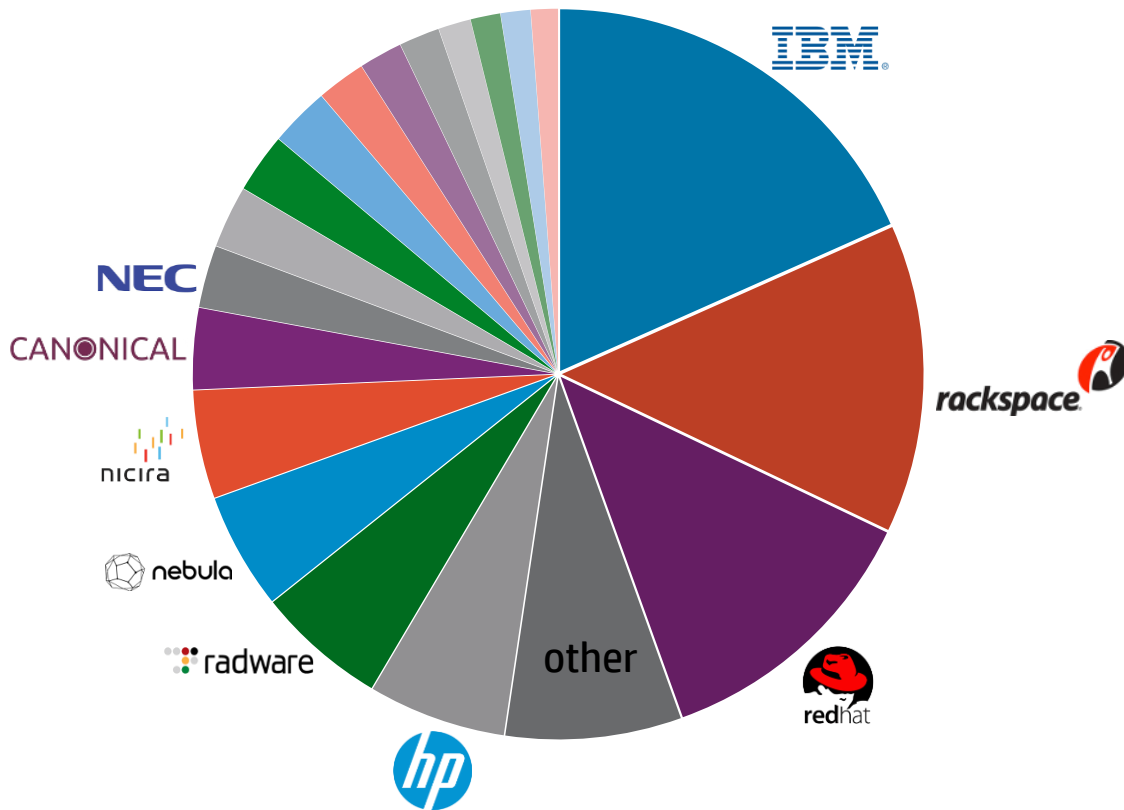
## Core includes

nova, swift, glance,  
horizon, keystone,  
cinder, quantum





# Top bugs fixed by employer for Grizzly core projects

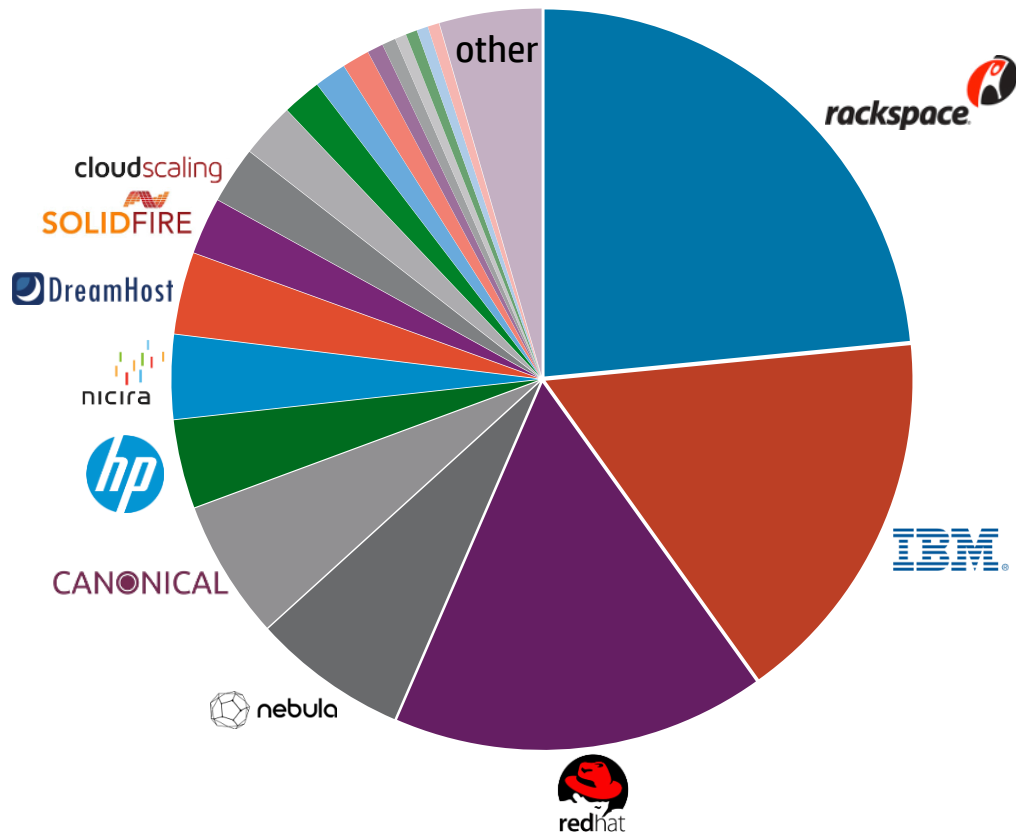


## Core includes

nova, swift, glance,  
horizon, keystone,  
cinder, quantum



# Top code reviewers for Grizzly core projects by employer



## Core includes

nova, swift, glance,  
horizon, keystone,  
cinder, quantum



# Configuring gitdm



# Configuring gitdm – What's in a release

```
$ vi openstack-gitdm/openstack-config/grizzly
```

```
# project revisions  
nova 2012.2..2013.1.rc2 (a701771|..big list of excludes..|5d9a5d1)  
glance 2012.2..2013.1.rc2  
swift 1.7.4..1.8.0.rc2  
keystone 2012.2..2013.1.rc2 (bec8b31|..big list of excludes..|23bb7ec)  
horizon 2012.2..2013.1.rc2  
quantum 2012.2..2013.1.rc2 (9f55912|..big list of excludes..|02fbf7d)  
cinder 2012.2..2013.1.rc3
```



# Configuring gitdm – Who's who

```
$ vi openstack-gitdm/openstack-config/aliases
```

```
#  
# This is the email aliases file, mapping secondary  
# addresses onto a single, canonical address.  
#  
john.doe@example.com john_doe@company_a.com  
johndoe@gmail.com john_doe@company_a.com  
jane-doe@example.ru jane.doe@some_organization.org  
...
```



# Configuring gitdm – Who's who in LaunchPad

```
$ vi openstack-gitdm/openstack-config/launchpad-ids.txt
```

```
johndoe john_doe@company_a.com  
janedoe77 jane.doe@some_organization.org  
...
```



# Configuring gitdm – Who works for who

```
$ vi openstack-gitdm/openstack-config/email-map
```

```
# [user@]domain employer [< yyyy-mm-dd]
johndoe@gmail.com Company A
johndoe@gmail.com Another Company < 2012-07-01
jane.doe@some_organization.org SomeOrg
...
```



# Configuring gitdm – Which orgs are which

```
$ vi openstack-gitdm/openstack-config/domain-map
```

```
# domain employer [< yyyy-mm-dd]
was_company_a.com Company A < 2011-12-31
company_a.com Company A
mail.company_a.com Company A
labs.company_a.com Company A
some_organization.org SomeOrg
...
```





# Opportunities for Improvement



# Measure all of OpenStack – 50 projects in Grizzly

- Including all of the OpenStack projects will provide a complete picture of a rapidly growing community

## Integrated (7)

OpenStack compute  
OpenStack object storage  
OpenStack Image Service  
OpenStack Identity  
OpenStack Dashboard  
OpenStack Networking  
OpenStack Block Storage

## Library (7)

OpenStack common  
python-\*-client

## Gating (4)

devstack  
tempest  
OpenStack-nose  
OpenStack-requirements

## Documentation (8)

OpenStack-manuals  
api-site  
\*-api

## Incubated (4)

Ceilometer  
Heat  
Python ceilometer client  
Python heat client

## Infrastructure (20)

config  
zuul  
jeepyb  
devstack-gate  
gear  
jenkins-job-builder  
gerrit  
git-review  
reviewday

...



# Some Questions to Consider

## When is a defect “done” and who “did” it?

“Fix Committed” seems to be a logical state

- Assignee vs. Opener?  
(maybe count both)

## Can we produce a consistent, unified database mapping contributors and organizations?

Including dates of service

- Centrally maintained
- Open for review

## Should code contributions be weighted?

Lines of code does not tell the whole story

- Difficult to judge the impact of any given commit
- Could weight using sloccount or other code-quality measures
- Could devise a peer-assigned points system via Gerrit



# Final Thought

Providing community-vetted metrics that are generated regularly, and that are based on open source tools and open data sources, can only help us to tell a complete and accurate story about who is working on OpenStack.



# Thank you

